

Министерство сельского хозяйства Российской Федерации
ФГБОУ ВО «Кубанский государственный аграрный университет
имени И. Т. Трубилина»

А. И. Орлов, Е. В. Луценко, В. И. Лойко

ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ,
МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ
ОБЕСПЕЧЕНИЕ КОНТРОЛЛИНГА,
ИННОВАЦИЙ И МЕНЕДЖМЕНТА

Монография

Под общей редакцией С. Г. Фалько

Краснодар
КубГАУ
2016

УДК 004.8:005.521
ББК 32.813
О-66

Рецензенты:

М. Н. Павленков – доктор экономических наук, профессор
(Дзержинский филиал Национального исследовательского Нижегородского
государственного университета им. Н. И. Лобачевского);

Н. П. Резникова – доктор экономических наук, профессор
(Московский технический университет связи и информатики)

Орлов А. И.

О-66 Организационно-экономическое, математическое и программное обеспечение контроллинга, инноваций и менеджмента: монография / А. И. Орлов, Е. В. Луценко, В. И. Лойко ; под общ. ред. С. Г. Фалько. – Краснодар : КубГАУ, 2016. – 600 с.

ISBN 978-5-00097-154-3

Монография посвящена организационно-экономическому, математическому и программному обеспечению контроллинга, инноваций и менеджмента и продолжает исследование подходов к решению проблем контроллинга путем применения методов, описанных в ранее выпущенных книг «Системная нечеткая интервальная математика» (Орлов А. И., Луценко Е. В., 2014) и «Перспективные математические и инструментальные методы контроллинга» (Орлов А. И., Луценко Е. В., Лойко В. И., 2015). Некоторые мысли, излагаемые в монографии, носят спорный и дискуссионный характер и высказаны в порядке научного обсуждения.

Предназначено для всех, кто интересуется контроллингом, его инструментами и применением.

УДК 004.8:005.521
ББК 32.813

ISBN 978-5-00097-154-3

© Орлов А. И., Луценко Е. В.,
Лойко В. И., 2016
© ФГБОУ ВО «Кубанский
государственный аграрный
университет имени И. Т. Трубилина», 2016

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	5
ГЛАВА 1. ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРИ РЕШЕНИИ ЗАДАЧ КОНТРОЛЛИНГА.....	6
1.1. Несколько слов о терминах и развитии науки	6
1.2. Организационно-экономическое моделирование и эконометрика как научно-практические и учебные дисциплины.....	8
1.3. Прогнозирование как одна из функций управления промышленными предприятиями	10
1.4. Проблемы устойчивости в моделях и методах разработки стратегии предприятия	10
1.5. После кризисов - солидарная информационная экономика.....	12
1.6. Заключительные замечания.....	12
ГЛАВА 2.РАЗВИТИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ В НАШЕЙ СТРАНЕ.....	13
2.1. Основные этапы становления статистических методов	13
2.2. Вероятностно-статистические методы в работах А. Н. Колмогорова.....	23
2.3. Вероятностно-статистические методы в работах Б. В. Гнеденко.....	32
2.4. Непараметрическая и прикладная статистика в нашей стране	41
2.5. Состояние и перспективы развития прикладной и теоретической статистики.....	53
ГЛАВА 3. НЕКОТОРЫЕ РЕЗУЛЬТАТЫ ПРИКЛАДНОЙ СТАТИСТИКИ.....	62
3.1. Структура прикладной статистики.....	62
3.2. Теоретические инструменты статистических методов	73
3.3. Распределения реальных статистических данных не являются нормальными.....	83
3.4. Выборочные исследования.....	90
3.5. Проверка однородности	110
3.6. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей	142
ГЛАВА 4. ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ КОНТРОЛЛИНГА, ИННОВАЦИЙ И МЕНЕДЖМЕНТА В РАКЕТНО-КОСМИЧЕСКОЙ ПРОМЫШЛЕННОСТИ.....	158
4.1. Организационно-экономическое обеспечение ракетно-космической промышленности.....	158
4.2. Основные идеи солидарной информационной экономики - базовой организационно-экономической теории	167
4.3. Организационно-экономические подходы к оценке реализуемости инновационно- инвестиционных проектов.....	176
4.4. Управление требованиями и оценка реализуемости проектов создания изделий ракетно-космической техники.....	186
4.5. Аддитивно-мультипликативная модель оценки рисков при создании ракетно-космической техники.....	193
4.6. Многообразие методов прогнозирования	209
4.7. Оценка инфляции по независимой информации	221
4.8. Прогноз развития информационно-коммуникационных технологий	234

4.9. ПРОФЕССИОНАЛЬНЫЕ СТАНДАРТЫ, ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ И УПРАВЛЕНИЕ В РАКЕТНО-КОСМИЧЕСКОЙ ПРОМЫШЛЕННОСТИ	245
ГЛАВА 5. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ В ЭКОЛОГИИ И ЭКОНОМИКЕ	255
5.1. АСК-анализ влияния экологических факторов на качество жизни населения региона	255
5.2. Системно-когнитивные модели прогнозирования развития многоотраслевой агропромышленной корпорации	280
ГЛАВА 6. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ В ОЦЕНКЕ РЕЗУЛЬТАТОВ НАУЧНОЙ И ПРЕПОДАВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ.....	291
6.1. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос», системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине	291
6.2. Хиршамания при оценке результатов научной деятельности, ее негативные последствия и попытка их преодоления с применением многокритериального подхода и теории информации	340
6.3. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию.....	359
6.4. Синтез и верификация многокритериальной системно-когнитивной модели университетского рейтинга Гардиан и ее применение для сопоставимой оценки эффективности российских вузов с учетом направления подготовки	376
ГЛАВА 7. ПРИМЕНЕНИЕ АСК-АНАЛИЗА И ТЕОРИИ ИНФОРМАЦИИ ДЛЯ РЕШЕНИЯ НЕКОТОРЫХ ЗАДАЧ СТАТИСТИКИ	425
7.1. Исследование статистических распределений методами теории информации	425
7.2. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в них.....	454
7.3. Асимптотический информационный критерий качества шума	508
ЗАКЛЮЧЕНИЕ	539
ЛИТЕРАТУРА	540

ПРЕДИСЛОВИЕ

Организационно-экономическое, математическое и программное обеспечение (короче, инструментальное обеспечение) контроллинга, инноваций и менеджмента - предмет настоящей книги. Она состоит из 7 глав.

Небольшая вводная глава 1 посвящена основным вопросам применения организационно-экономического моделирования при решении задач контроллинга. В главе 2 впервые в монографической литературе дан исторический анализ развития математико-статистических методов в нашей стране. Некоторые результаты прикладной математической статистики рассмотрены в главе 3, в том числе теоретические инструменты прикладной статистики, непараметрические методы проверки однородности и восстановления зависимости. В главе 4 обсуждаются на примере ракетно-космической промышленности недавние научные разработки (2013–2016 гг.) в области организационно-экономического обеспечения контроллинга, инноваций и менеджмента.

Следующие три главы посвящены различным применениям автоматизированного системно-когнитивного анализа (АСК-анализа). Плодотворное использование АСК-анализа в экономике представлено в главе 5. АСК-анализ в оценке результатов научной и преподавательской деятельности - предмет главы 6. В заключительной главе 7 АСК-анализ и теория информации применяются для решения некоторых задач прикладной статистики.

Настоящая монография продолжает ранее выпущенные тем же издательством книги "Системная нечеткая интервальная математика" (Орлов А.И., Луценко Е.В., 2014) и "Перспективные математические и инструментальные методы контроллинга" (Орлов А.И., Луценко Е.В., Лойко В.И., 2015).

Для удобства читателей каждая глава книги имеет свою нумерацию литературных источников, таблиц, формул и т.п.

Монография предназначена для научных работников в области экономики и управления, прежде всего для исследователей, развивающих математические и инструментальные методы экономики. Она может быть использована при преподавании дисциплин, связанных с организационно-экономическим, математическим и программным обеспечением контроллинга, инноваций и менеджмента.

Монография написана известными российскими учеными, внесшими большой вклад в развитие организационно-экономического моделирования, математических и инструментальных методов экономики, прежде всего в области контроллинга, инноваций и менеджмента. Она имеет высокую степень научной новизны. Поэтому естественно, что некоторые мысли, излагаемые в монографии, носят спорный и дискуссионный характер и высказаны авторами в порядке научного обсуждения.

Научный редактор монографии:

*Исполнительный директор
Объединения контроллеров,
д.э.н., профессор С. Г. Фалько*

ГЛАВА 1. ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРИ РЕШЕНИИ ЗАДАЧ КОНТРОЛЛИНГА

На кафедре «Экономика и организация производства» научно-учебного комплекса «Инженерный бизнес и менеджмент» МГТУ им. Н.Э.Баумана в конце XX – начале XXI вв. создана научная школа в области организационно-экономического моделирования, эконометрики и статистики [1]. Преподавание соответствующих дисциплин курирует одноименная секция кафедры, научные исследования ведет Лаборатория экономико-математических методов в контроллинге Научно-образовательного центра "Контроллинг и управленческие инновации" МГТУ им. Н.Э. Баумана. Глава 1 посвящена деятельности отечественной научной школы в области организационно-экономического моделирования, эконометрики и статистики, ведущимся научным исследованиям и некоторым полученным результатам.

1.1. Несколько слов о терминах и развитии науки

Любое широко используемое понятие – «контроллинг», «маркетинг», «статистика» и т.п. – имеет сотни и тысячи определений. Поэтому поясним понимание используемых нами терминов.

Организационно-экономическое моделирование (ОЭМ) – научная, практическая и учебная дисциплина, посвященная разработке, изучению и применению математических и статистических методов и моделей в экономике и управлении народным хозяйством, прежде всего промышленными предприятиями и их объединениями. Термин «экономико-математические методы и модели» (ЭММиМ) имеет близкое содержание. Отличаются эти термины акцентами. ОЭМ нацелено на решение задач организации (управления, менеджмента) и экономики с помощью моделирования. В ЭММиМ сильна математическая составляющая, не связанная непосредственно с проблемами реального мира.

Статистические методы в экономике – предмет эконометрики, базой которой является прикладная статистика. Развитие представлений в этой области и соответствующей терминологии рассмотрим подробнее.

В нашей стране к 60-м годам XX в. сформировалась научно-практическая дисциплина, которую называем классической математической статистикой. Специалисты-статистики учились теории по книге известного шведского математика Г. Крамера [2], написанной в военные годы и впервые изданной у нас в 1948 г. Из прикладных руководств назовем учебник [3] и таблицы с комментариями [4].

Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Примером таких работ является монография [5]. В ней получены продвинутые математические результаты, но трудно выделить рекомендации для статистика, анализирующего конкретные данные.

Как реакция на уход в математику выделилась новая научная дисциплина - прикладная статистика. В учебнике [6] в качестве рубежа, когда это стало очевидным, мы указали 1981 г. – дату выхода массовым тиражом (33 940 экз.) сборника [7], в названии которого использован термин «прикладная статистика». С этого времени линии развития математической статистики и прикладной статистики разошлись. Первая из этих дисциплин полностью ушла в математику, перестав интересоваться практическими делами. Вторая позиционировала себя в качестве науки об обработке данных – результатов наблюдений, измерений, испытаний, анализов, опытов.

Вполне естественно, что в прикладной статистике стали развиваться математические методы и модели. Необходимость их развития вытекает из потребностей конкрет-

ных прикладных исследований. Это математизированное ядро прикладной статистики хочется назвать теоретической статистикой. Тогда под собственно прикладной статистикой следует понимать обширную промежуточную область между теоретической статистикой и применением статистических методов в конкретных областях. В нее входят, в частности, вопросы формирования вероятностно-статистических моделей и выбора конкретных методов анализа данных (т.е. методология прикладной статистики и других статистических методов), проблемы разработки и применения информационных статистических технологий, организации сбора и анализа данных, т.е. разработки статистических технологий.

Таким образом, общая схема современной статистической науки выглядит следующим образом (от абстрактного к конкретному):

1. Математическая статистика – часть математики, изучающая статистические структуры. Сама по себе не дает рецептов анализа статистических данных, однако разрабатывает методы, полезные для использования в теоретической статистике.

2. Теоретическая статистика – наука, посвященная моделям и методам анализа конкретных статистических данных.

3. Прикладная статистика (в узком смысле) посвящена статистическим технологиям сбора и обработки данных. Она включает в себя методологию статистических методов, вопросы организации выборочных исследований, разработки статистических технологий, создания и использования статистических программных продуктов.

4. Применение статистических методов в конкретных областях (в экономике и менеджменте – эконометрика, в биологии – биометрика, в химии – хеометрия, в технических исследованиях – технометрика, в геологии, демографии, социологии, медицине, истории, и т.д.).

Часто позиции 2 и 3 вместе называют прикладной статистикой. Иногда позицию 1 именуют теоретической статистикой. Эти терминологические расхождения связаны с тем, что описанное выше развитие рассматриваемой научно-прикладной области не сразу, не полностью и не всегда адекватно отражается в сознании специалистов. Так, до сих пор выпускают учебники, соответствующие уровню представлений середины XX века.

Примечание. Здесь мы уточнили схему внутреннего деления статистической теории, предложенную нами в 1990 г. в [8]. Естественный смысл приобрели термины «теоретическая статистика» и «прикладная статистика» (в узком смысле). Однако необходимо иметь в виду, что в недавнем нашем учебнике [6] прикладная статистика понимается в широком смысле, т.е. как объединение позиций 2 и 3. К сожалению, в настоящее время невозможно отождествить теоретическую статистику с математической, поскольку последняя (как часть математики - научной специальности «теория вероятностей и математическая статистика») заметно оторвалась от задач практики.

Отметим, что математическая статистика, как и теоретическая с прикладной, заметно отличается от ведомственной науки органов официальной государственной статистики. ЦСУ, Госкомстат, Росстат применяли и применяют лишь проверенные временем приемы позапрошлого (19-го) века. Возможно, следовало бы от этого ведомства полностью отмежеваться и сменить название дисциплины, например, на «Анализ данных». В настоящее время компромиссным самоназванием рассматриваемой научно-практической дисциплины является термин «статистические методы».

В последние пять лет стало проявляться переосмысление терминов. Поскольку поток научных публикаций по математической статистике практически иссяк (видимо, отрыв от практики анализа реальных данных оказался фатальным), описанное выше разделение прикладной и математической статистики перестало быть актуальным, использование самого термина "прикладная статистика" в научном плане уже не является

необходимым. Однако в организационном плане этот термин продолжает оставаться полезным для отмежевания от учебной дисциплины "теория вероятностей и математическая статистика". Дело в том, что "прикладная статистика" преподается в соответствии с новой парадигмой математических методов экономики [1], прикладной статистики [12] и математической статистики [13], в то время как преподавание учебной дисциплины "теория вероятностей и математическая статистика" обычно ведется в соответствии с устаревшей парадигмой середины XX в.

Во второй половине 80-х годов развернулось общественное движение, имеющее целью создание профессионального объединения статистиков. Аналогами являются британское Королевское статистическое общество (основано в 1834 г.) и Американская статистическая ассоциация (создана в 1839 г.). К сожалению, деятельность учрежденной в 1990 г. Всесоюзной статистической ассоциации (ВСА) [9] оказалась парализованной в результате развала СССР. Некоторую активность проявили Российская ассоциация статистических методов, Российская академия статистических методов, Белорусская статистическая ассоциация, созданные на базе ВСА. Международные конференции "Статистика и ее применения" регулярно проводятся в Ташкенте.

В ходе создания ВСА проанализировано состояние и перспективы развития теоретической и прикладной статистики. Создана новая парадигма организационно-экономического моделирования, эконометрики и статистики, основанная, в частности, на переходе от использования параметрических семейств распределений к непараметрической и нечисловой статистике. Выявлена необходимость создания нового поколения учебной литературы, которая должна сменить издания на основе идей середины XX в.

Реализация этой задачи – создание системы учебных дисциплин и учебников нового поколения, отражающих современную научную парадигму в рассматриваемой области – основное достижение научной школы МГТУ им. Н.Э. Баумана в области организационно-экономического моделирования, эконометрики и статистики.

1.2. Организационно-экономическое моделирование и эконометрика как научно-практические и учебные дисциплины

Для достижения успеха при стратегическом планировании на российских предприятиях должны использоваться современные методы организационно-экономического моделирования, основанные на интенсивном использовании математических и статистических методов. Рассмотрим основные подходы к преподаванию дисциплин "Эконометрика" и «Организационно-экономическое моделирование» (ОЭМ) кафедрой «Экономика и организация производства» научно-учебного комплекса «Инженерный бизнес и менеджмент» (НУК ИБМ) МГТУ им. Н. Э. Баумана.

Развитие научно-технического прогресса предъявляет все большие требования к современным инженерам, управленцам и экономистам. В специалистах такого уровня особо остро нуждаются высокотехнологичные предприятия оборонно-промышленного комплекса. Для их подготовки в 2005 г. была разработана новая специальность «Менеджмент высоких технологий», выпущен одноименный учебник [10].

Для повышения эффективности управления промышленными предприятиями и интегрированными производственно-корпоративными структурами необходимо применять современные организационно-экономические методы и модели. В соответствии с этим в программу обучения по специальности «Менеджмент высоких технологий» и была введена дисциплина ОЭМ. В настоящее время "Эконометрика" преподается бакалаврам специальностей "Инноватика", "Экономика", "Менеджмент", а ОЭМ - магистрантам специальности "Организация и управление наукоемким производством".

Изучаемые методы признаны вооружить будущих специалистов, т.е. выпускников нашего Университета, современным статистическим и математическим инструментарием, необходимым для профессиональной деятельности, с целью повышения конкурентоспособности на рынке труда.

В рассматриваемых дисциплинах изучаются методы разработки и принятия управленческих решений. Значительная часть курсов посвящена выборочным исследованиям. Рассматриваются примеры применения случайных выборок при оценивании функции спроса и изучении предложения потребителей, а также в связи с контролем качества продукции и экологической обстановки. Обсуждается применение статистики нечисловых данных в теории и практике экспертных оценок. Рассматриваются оптимальные методы в экономике и управлении на примере управления запасами. Для целей прогнозирования изучается линейный регрессионный анализ, который рассматривается в основном на примере восстановления линейной зависимости между двумя переменными. Ежегодно в качестве лабораторной работы студенты собирают информацию о ценах для последующего расчета индексов инфляции.

Разработаны учебники и учебные пособия по организационно-экономическому моделированию, эконометрике и статистике. Эти дисциплины опираются на теорию вероятностей и математическую статистику. В интересах студентов, изучающих наш цикл дисциплин, выпущен справочник, содержащий основные факты в области теории вероятностей и прикладной статистики [11].

Первый этап в триаде наших дисциплин - курс "Прикладная статистика" на втором году обучения. В соответствии с новой парадигмой прикладной статистики [12, 13] выпущен учебник [6], соответствующий современным представлениям об этой научно-прикладной области.

На четвертом году обучения бакалавров преподаем годовой курс "Эконометрика". Представления отечественной научной школы в области эконометрики отражены в одноименном учебнике. Первые его три издания [14–16] отличаются сравнительно мало. Четвертое издание [17], соответствующее осеннему семестру, полностью переработано.

Для магистрантов второго года обучения предназначен курс "Организационно-экономическое моделирование". Выпущенный в трех частях одноименный учебник [18–20] включает также материал ранее рассмотренных дисциплин.

С точки зрения теории и практики разработки и принятия управленческих решений методы прикладной статистики, эконометрики и организационно-экономического моделирования рассмотрены в [21–24]. Взгляд со стороны менеджмента проработан в [25, 26].

В рамках научно-учебного комплекса (факультета) "Инженерный бизнес и менеджмент" ведутся теоретические исследования и прикладные разработки в области организационно-экономического моделирования, прежде всего в Лаборатории экономико-математических методов в контроллинге Научно-образовательного центра "Управленческие инновации". Основные места публикаций – журналы «Контроллинг», «Заводская лаборатория. Диагностика материалов», Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета, "Инновации в менеджменте". Рассмотрим примеры ведущихся исследований и некоторые полученные результаты.

1.3. Прогнозирование как одна из функций управления промышленными предприятиями

На современном этапе к системе управления промышленными предприятиями предъявляются всё новые требования. В частности, необходима разработка эффективных систем прогнозирования для нужд управления промышленными предприятиями.

Основные функции управления были сформулированы еще Анри Файолем: «Управлять – значит прогнозировать и планировать, организовывать, руководить командой, координировать и контролировать» [27]. Результаты прогнозирования необходимы для планирования [28].

Прогнозирование, как функция управления промышленными предприятиями, в настоящее время недостаточно используется на практике. Есть ряд нерешенных проблем.

1. Часто используются устаревшие методы, не соответствующие современным требованиям.

2. Прогнозируются лишь отдельные характеристики (показатели) промышленных предприятий. В то время как решение проблемы прогнозирования заключается в рассмотрении предприятия как единой системы [29].

3. Предприятие в процессе своей производственно-хозяйственной деятельности сталкивается с проблемой учета инфляции. Существуют два способа учета инфляции – обратиться к внешним источникам или рассчитать самостоятельно [17, 30]. Последнее предполагает разработку организационно-экономического моделирования инфляции.

Как правило, выделяют два вида прогнозирования – поисковое (эксплораторное, эксплоративное) и нормированное (нормативное). Основываясь на идеях метода когнитивных карт ЖОК [6, 20, 22], вводим еще один – вариативный, отвечающий на вопрос: какие будут изменения при варьировании управляющими факторами.

В большинстве случаев распределение погрешностей измерений отличается от нормального [6, 17, 31]. Следовательно, необходимо разрабатывать методы, не требующие знания закона распределения наблюдаемой случайной величины, т.е. непараметрические [32, 33]. В работах [34, 35] в рамках непараметрической вероятностно-статистической модели получено асимптотическое распределение точки пересечения, уровня качества и временного лага двух регрессионных линейных зависимостей. На основе метода линеаризации найдены асимптотические дисперсии и доверительные интервалы для рассматриваемых характеристик [36].

Внедрение методов прогнозирования [37] целесообразно осуществлять в духе системы «Шесть сигм». Эта система может быть использована не только для повышения качества продукции и услуг. Она решает задачи совершенствования бизнеса, в том числе организации внедрения современных организационно-экономических методов и моделей [38–40].

1.4. Проблемы устойчивости в моделях и методах разработки стратегии предприятия

Процессы стратегического планирования на промышленных предприятиях реализуются в реальных ситуациях с достаточно высоким уровнем неопределенности. Велика роль нечисловой информации как на «входе», так и на «выходе» процесса принятия управленческого решения. Неопределенность и нечисловая природа управленческой информации должны быть отражены при анализе устойчивости экономико-математических методов и моделей.

Для обоснованного практического применения математические модели процессов управления промышленными предприятиями и основанных на них экономико-математических методов должна быть изучена их устойчивость по отношению к допустимым отклонениям исходных данных и предпосылок моделей. Возможные применения результатов подобного исследования:

- заказчик научно-исследовательской работы получает представление о точности предлагаемого решения;
- удается выбрать из многих моделей наиболее адекватную;
- по известной точности определения отдельных параметров модели удается указать необходимую точность нахождения остальных параметров;
- переход к случаю «общего положения» позволяет получать более сильные с математической точки зрения результаты.

Следовательно, необходима разработка и развитие теоретических основ и методологии обоснования, выбора и создания новых математических методов и моделей, направленных на рационализацию и оптимизацию управления экономической составляющей производственно-хозяйственной деятельности промышленных предприятий на основе изучения их устойчивости по отношению к допустимым отклонениям исходных данных и предпосылок моделей.

Реализованную нами исследовательскую программу [41, 42] целесообразно разбить на шесть направлений, посвященных решению следующих задач:

1. Развить методологию разработки математических методов и моделей процессов управления промышленными предприятиями, разработать общий подход к изучению устойчивости (общую схему устойчивости) таких моделей и методов и выделить частные постановки проблем устойчивости, в том числе устойчивость к изменению данных, их объемов и распределений, по отношению к временным характеристикам. Один из подходов к построению устойчивых методов и моделей - моделирование с помощью нечисловых объектов.

2. Разработать непараметрические (устойчивые к изменению распределения) статистические методы для решения конкретных задач управления промышленными предприятиями – для оценки характеристик, прогнозирования, сегментации рынка и др.

3. Установить связи между различными видами объектов нечисловой природы, построить вероятностные модели их порождения. На основе расстояний (показателей различия, мер близости) и задач оптимизации развить статистическую теорию в пространствах общей природы, Разработать методы моделирования конкретных нечисловых объектов.

4. Разработать асимптотическую статистику интервальных данных на основе понятий нотны и рационального объема выборки, развить интервальные аналоги основных областей прикладной статистики.

5. На основе концепции устойчивости по отношению к временным характеристикам (моменту начала реализации проекта, горизонту планирования) провести экономико-математическое моделирование ряда процессов стратегического управления промышленными предприятиями: обосновать применение асимптотически оптимальных планов, дать характеристику моделей с дисконтированием.

6. Разработать устойчивые экономико-математические методы и модели процессов управления экономикой в функциональных областях производственно-хозяйственной деятельности промышленных предприятий, в частности, при использовании экспертных методов, в инновационном и инвестиционном менеджменте, при управлении качеством промышленной продукции, при выявлении предпочтений потребителей, при управлении материальными ресурсами предприятия.

1.5. После кризисов - солидарная информационная экономика

Экономические кризисы последних лет выявил необходимость разработки новых организационно-экономических механизмов стратегического планирования и развития предприятий. Организация производства должна быть основана на адекватной экономической теории.

Перспективные организационно-экономические механизмы управления производственно-хозяйственной деятельностью предлагаем конструировать на основе солидарной информационной экономики (ранее использовался термин "неформальная информационная экономика будущего"), разрабатываемой как методологическая основа конкретных исследований в области организационно-экономического моделирования [43–48]. Отметим некоторые ее черты.

1. При принятии решений необходимо учитывать не только экономические, но и технологические, социальные, экологические, политические факторы. Экономика в целом – служанка общества, выполняет его требования. Цели общества первичны, экономические механизмы вторичны.

2. На основе современных информационных технологий и методов разработки и принятия управленческих решений [21, 22, 24] необходимо реализовать идеи В. М. Глушкова и Ст. Бира. В 1960-х В. М. Глушков предложил создать Общегосударственную автоматизированную систему управления экономикой страны (ОГАС), а Ст. Бир разработал автоматизированную систему управления национализированными предприятиями Чили «Киберсин».

3. Новым по сравнению с временами В. М. Глушкова и С. Бира является широкое распространение Интернет-технологий, позволяющее аппаратно реализовать право граждан на участие в принятии касающихся их решений. Открытый процесс создания реальных организационных модулей системы, привлечения участников, прокладывания горизонтальных связей, осуществления операций с ее помощью можно назвать Open P2P Society - «Открытое сетевое общество». Неформальность – важнейшая черта, обеспечивающая свободу передачи информации и участие всех заинтересованных лиц и организаций в выработке и реализации решений. Удастся реализовать основные идеи анархизма.

4. В перспективе путем предварительного обсуждения и планирования можно будет снять проблему нерационального производства товаров и услуг. Удастся снять противоречие между «планом» и «рынком», избавиться от недостатков, но сохранить достоинства каждого из этих подходов к организации хозяйственной жизни.

5. Экономическую теорию надо избавить от крена в сторону хрестоматистики. Экономика – это наука о том, как производить, а не о том, как делить прибыль. Основное ядро современной экономической теории – это экономика предприятия.

1.6. Заключительные замечания

Организационно-экономическое моделирование (включая эконометрику, прикладную статистику, теорию принятия решений и другие экономико-математические методы и модели) разрабатывает интеллектуальные инструменты, позволяющие контроллеру успешно решать стоящие перед ним задачи модернизации систем управления предприятиями и организациями. В журнале «Контроллинг» опубликован ряд разработок Лаборатории экономико-математических методов в контроллинге Научно-образовательного центра "Контроллинг и управленческие инновации" МГТУ им. Н.Э. Баумана, из которых отметим первую по времени программную статью [49] и связан-

ные с тематикой настоящей главы работы [50, 51]. Итоги проведенных исследований отражены в монографиях [52, 53].

ГЛАВА 2. РАЗВИТИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ В НАШЕЙ СТРАНЕ

2.1. Основные этапы становления статистических методов

Статистические методы широко и успешно применяются практически во всех отраслях народного хозяйства, почти во всех областях научных исследований. Литература по статистическим методам необозрима – на русском языке десятки тысяч книг и статей. Однако есть пробел – практически нет работ по истории статистических методов. Вслед за выдающимся математиком академиком АН УССР Б. В. Гнеденко полагаем, что научный работник и тем более преподаватель должен знать историю своей дисциплины, чтобы понимать настоящее ее состояние и быть способным прогнозировать дальнейшее развитие, опираться на эти прогнозы в своей научной деятельности. Поэтому представляется необходимым включение в настоящую монографию главы, посвященной истории прикладной математической статистики, развитию статистических методов.

2.1.1. Моисей – первый статистик

Самая ранняя из известных нам статистических работ входит в Библию. В Ветхий Завет включена Четвертая книга Моисеева под названием «Числа». Глава 1 этой книги посвящена переписи военнообязанных. Она начинается так (цитируем по синодальному изданию Библии):

«1. И сказал Господь Моисею в пустыне Синайской, в скинии собрания, в первый день второго месяца, во второй год по выходе их из земли Египетской, говоря:

2. Исчислите всё общество сынов Израилевых по родам их, по семействам их, по числу имен, всех мужского пола поголовно,

3. От двадцати лет и выше, всех годных для войны у Израиля, по ополчениям их исчислите их – ты и Аарон.

4. С вами должны быть из каждого колена по одному человеку, который в роде своем есть главный.

** *

21. Исчислено в колене Рувимовом сорок шесть тысяч пятьсот.

** *

23. Исчислено в колене Симеоновом пятьдесят девять тысяч триста.

** *

46. И было всех вошедших в исчисление шестьсот три тысячи пятьсот пятьдесят».

Практическая направленность этого статистического исследования вполне очевидна. Обратите внимание, что оно предпринято по решению руководства страны (в библейских терминах – «общества сынов Израилевых»), причем к работам привлечены региональные начальники (главные по коленам, на которые делилось государство). Четко указана совокупность, подлежащая переписи - мужчины от 20 лет и старше, годные для войны (военнообязанные).

Древность исследования проявляется только в том, что стандартные описания результатов учета военнообязанных по коленам выражены словами. Сейчас мы представили бы результаты в виде таблицы (табл.1). Таблицы такого типа постоянно состав-

ляют органы государственной статистики и в настоящее время (см. портал <http://www.gks.ru/wps/portal> Федеральной службы государственной статистики РФ (краткое название – Росстат)).

Итак, при сравнении с деятельностью Росстата описанное в Библии исследование, выполненное под руководством Моисея, является вполне современным по своим задачам и методам.

Таблица 1 – Число всех годных для войны у Израиля

№ п/п	Родоначальник колена	Число военнообязанных
1	Рувим	46 500
2	Симеон	59 300
3	Гад	45 650
4	Иуда	74 600
5	Иссахар	54 400
6	Завулон	57 400
7	Ефрем	40 500
8	Манассия	32 200
9	Вениамин	35 400
10	Дан	62 700
11	Асир	41 500
12	Неффалим	53 400
	Всего	603 550

2.1.2. Развитие представлений о статистике

В Библии не было терминов «статистика» или «статистик». Согласно [1] впервые термин «статистик» мы находим в художественной литературе – в «Гамлете» Шекспира (1602 г., акт 5, сцена 2). Смысл этого слова у Шекспира – знать, придворные. По-видимому, оно происходит от латинского слова *status*, что в оригинале означает «состояние» или «политическое состояние».

В течении следующих 400 с небольшим лет термин «статистика» понимали и понимают по-разному. В работе [1] сотрудниками Межфакультетской лаборатории статистических методов МГУ им. М. В. Ломоносова собрано более 200 определений этого термина, некоторые из которых обсуждаются ниже.

Вначале под статистикой понимали описание экономического и политического состояния государства или его части. Например, к 1792 г. относится определение: «Статистика описывает состояние государства в настоящее время или в некоторый известный момент в прошлом». И в настоящее время деятельность государственных статистических служб достаточно хорошо соответствует этому определению.

Однако постепенно термин «статистика» стал использоваться более широко. По Наполеону Бонапарту: «Статистика – это бюджет вещей». Тем самым статистические методы были признаны полезными не только для административного управления, но и для управления на уровне отдельного предприятия. Согласно формулировке 1833 г. «цель статистики заключается в представлении фактов в наиболее сжатой форме». Здесь статистика уже не связывается ни с государствоведением, ни с социально-экономическими проблемами вообще.

Приведем еще два высказывания. «Статистика состоит в наблюдении явлений, которые могут быть подсчитаны или выражены посредством чисел» (1895). «Статистика – это численное представление фактов из любой области исследования в их взаимосвязи» (1909).

В XX в. статистику часто рассматривают прежде всего как самостоятельную научную дисциплину. «Статистика есть совокупность методов и принципов, согласно которым проводится сбор, анализ, сравнение, представление и интерпретация числовых данных» (1925). В 1954 г. академик АН УССР Б. В. Гнеденко дал следующее определение: «Статистика состоит из трех разделов:

- 1) сбор статистических сведений, т.е. сведений, характеризующих отдельные единицы каких-либо массовых совокупностей;
- 2) статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе данных массового наблюдения;
- 3) разработка приемов статистического наблюдения и анализа статистических данных. Последний раздел, собственно, и составляет содержание математической статистики».

Термин «статистика» употребляют еще в двух смыслах. Во-первых, в обиходе под «статистикой» часто понимают набор количественных данных о каком-либо явлении или процессе. Во-вторых, специалисты в области статистических методов называют «статистикой» функцию от результатов наблюдений, используемую для оценивания характеристик и параметров распределений и проверки гипотез.

Чтобы подойти к современному состоянию, в частности, разъяснить широко используемый термин «прикладная статистика», кратко рассмотрим историю реальных статистических работ.

2.1.3. Краткая история статистических методов

Как уже отмечалось, типовые примеры раннего этапа применения статистических методов описаны в Ветхом Завете. Там, в частности, приводится число воинов в различных племенах («коленах»). С математической точки зрения дело сводилось к подсчету числа попаданий значений наблюдаемых признаков в определенные градации.

В дальнейшем результаты обработки статистических данных стали представлять в виде таблиц и диаграмм, как это и сейчас делает Росстат. Надо признать, что по сравнению с Ветхим Заветом есть прогресс - в Библии не было таблиц и диаграмм. Однако нет продвижения по сравнению с работами российских статистиков конца девятнадцатого - начала двадцатого века (типовой монографией тех времен можно считать книгу [2], которая в настоящее время ещё легко доступна).

Сразу после возникновения теории вероятностей (Паскаль, Ферма, 17 век) вероятностные модели стали использоваться при обработке статистических данных. Например, изучалась частота рождения мальчиков и девочек, было установлено отличие вероятности рождения мальчика от $1/2$, анализировались причины того, что в парижских приютах эта вероятность не та, что в самом Париже, и т.д. Имеется достаточно много публикаций по истории теории вероятностей с описанием раннего этапа развития статистических методов исследований, к лучшим из них относится очерк [3]. Отметим, что основатель современного бухгалтерского учета Лука Пачолли (1445–1517) хорошо известен и историкам теории вероятностей. Это символично, поскольку вопросы учета и статистики тесно переплетаются в деятельности современного инженера, экономиста и менеджера.

В 1794 г. (по другим данным - в 1795 г.) великий немецкий математик и физик Карл Фридрих Гаусс разработал метод наименьших квадратов, один из наиболее популярных ныне статистических методов, и применил его при расчете орбиты астероида (в современной терминологии – малой планеты) Церера - для борьбы с ошибками астрономических наблюдений [4]. В XIX веке заметный вклад в развитие практической ста-

тики внес бельгиец Ламбер Адольф Жак Кетле, на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей, таких, как доля самоубийств среди всех смертей [5]. Интересно, что основные идеи статистического приемочного контроля и сертификации продукции обсуждались академиком Петербургской АН М. В. Остроградским (1801–1862) в 1846 г. и применялись в российской армии ещё в середине XIX в. [3]. Статистические методы управления качеством и сертификации продукции сейчас весьма актуальны.

2.1.4. Параметрическая статистика

Современный этап развития статистических методов можно отсчитывать с 1900 г., когда английский математик, статистик, биолог и философ Карл Пирсон основал журнал «*Biometrika*». Первая треть XX в. прошла под знаком параметрической статистики. Изучались методы, основанные на анализе данных из параметрических семейств распределений, описываемых кривыми семейства Пирсона. Наиболее популярным было нормальное (гауссово) распределение. Для проверки гипотез использовались критерии Пирсона, Стьюдента, Фишера, основанные на вероятностно-статистических моделях, в которых результаты измерений (наблюдений, испытаний, опытов, анализов) имели нормальное распределение. В те годы были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

Разработанную в первой трети XX в. теорию анализа данных называем параметрической статистикой, поскольку ее основной объект изучения - это выборки из распределений, описываемых одним параметром или небольшим числом параметров (2–4). Наиболее общим является семейство кривых Пирсона, задаваемых четырьмя параметрами.

С математической точки зрения параметрическая статистика дает интересные теоретические схемы, на основе которых удастся построить развитую теорию. Для профессионалов укажем на теорию достаточных статистик, неравенство Рао - Крамера, теорию оптимального оценивания и другие внутриматематические достижения.

Критика параметрической статистики вытекает из ее оторванности от практики статистической работы. Как правило, нельзя указать каких-либо веских причин, по которым распределение результатов конкретных наблюдений непременно должно входить в то или иное параметрическое семейство. Исключения есть, и они хорошо известны: если вероятностная модель предусматривает суммирование независимых случайных величин, то сумму естественно описывать нормальным распределением; если же в модели рассматривается произведение таких величин, то итог, видимо, приближается логарифмически нормальным распределением, и т.д. Однако подобных моделей нет в подавляющем большинстве реальных ситуаций, и приближение реального распределения с помощью кривых из семейства Пирсона или его подсемейств - чисто формальная операция.

Именно из таких соображений критиковал параметрическую статистику академик АН СССР С. Н. Бернштейн в 1927 г. в своем докладе на Всероссийском съезде математиков [6]. Однако эта теория, к сожалению, до сих пор остается основой преподавания статистических методов и продолжает использоваться основной массой прикладников, использующих статистические методы в различных отраслях народного хозяйства и областях науки, но далеких от новых достижений в статистической науке. Почему так происходит? Чтобы попытаться ответить на этот вопрос, обратимся к наукометрии, т.е. к статистическим методам в науковедении.

2.1.5. Наукометрия статистических исследований

В рамках движения за создание Всесоюзной статистической ассоциации (учреждена в 1990 г.) был проведен анализ статистики как области научно-практической деятельности. Он показал, в частности, что актуальными для специалистов в настоящее время являются не менее чем 100 тысяч публикаций (подробнее см. статьи [7, 8]). Реально же каждый из нас знаком с существенно меньшим количеством книг и статей. Так, в известном трехтомнике М. Кендалла и А. Стьюарта [9–11] – наиболее полном на русском языке издании по статистическим методам - всего около 2 тысяч литературных ссылок. При всей очевидности соображений о многократном дублировании в публикациях ценных идей приходится признать, что каждый специалист по статистическим методам владеет лишь небольшой частью накопленных в этой области знаний. Поэтому нет ничего удивительного в том, что приходится постоянно сталкиваться с игнорированием или повторением ранее полученных результатов, с уходом в тупиковые (с точки зрения практики) направления исследований, с беспомощностью при обращении к реальным данным, и т.д. Все это - одно из проявлений адапционного механизма торможения развития науки, вызванного ее быстрым ростом, о котором еще 45 лет назад писали В. В. Налимов и другие науковеды (см., например, [12]).

Традиционный предрассудок состоит в том, что каждый новый результат, полученный исследователем - это кирпич в непрерывно растущее здание науки, который непременно будет проанализирован и использован научным сообществом, а затем и при решении практических задач. Реальная ситуация - совсем иная. Основа профессиональных знаний исследователя, инженера, экономиста, менеджера, социолога, историка, геолога, медика закладывается в период обучения. Затем знания пополняются в том узком направлении, в котором работает специалист. Следующий этап – тиражирование знаний при обучении нового поколения. В результате вузовские учебники отстают от современного развития на десятки лет. Так, учебники по математической статистике, согласно мнению экспертов, по научному уровню в основном соответствуют 40–60-м годам XX в. А потому середине XX в. соответствует большинство вновь публикуемых исследований и тем более – прикладных работ. Одновременно приходится признать, что результаты, не вошедшие в учебники, независимо от их ценности почти все забываются. Достаточно взглянуть на длинные ряды библиотечных полок с номерами научных журналов за последние сто – двести лет. Сколько из них были хотя бы раз открыты в текущем веке? Кроме того, сейчас все популярнее поиск информации в Интернете – вплоть до того, что кое-кто из молодых даже забывает о существовании библиотек. А ведь в Интернете можно найти лишь небольшую часть опубликованных в XX (и даже в XXI) веке научных работ. Например, статьи, опубликованные в журнале «Заводская лаборатория. Диагностика материалов» до 2006 г., в Интернете отсутствуют, за исключением единичных работ, которые самостоятельно размещены авторами на тех или иных ресурсах.

Активно продолжается развитие тупиковых направлений. Психологически это понятно. Приведем пример из опыта первого из авторов этой книги. В свое время по заказу Госстандарта им были разработаны методы оценки параметров гамма-распределения [13]. Поэтому рассматриваемому исследователю близки и интересны работы по оцениванию параметров по выборкам из распределений, принадлежащих тем или иным параметрическим семействам, понятия функции максимального правдоподобия, эффективности оценок, использование неравенства Рао - Крамера и т.д. К сожалению, он знает, что это – тупиковая ветвь теории статистики, поскольку реальные данные не подчиняются каким-либо параметрическим семействам, надо применять иные статистические методы, о которых речь пойдет ниже. Понятно, что специалистам по

параметрической статистике, потратившим многие годы на совершенствование в своей области, психологически трудно согласиться с этим утверждением. В том числе и рассматриваемому здесь исследователю. Но необходимо идти вперед. Поэтому написанные им учебники во многом очищены от тупиковых подходов. В том числе и от неравенства Рао - Крамера. Однако включены разделы, посвященные оцениванию параметров распределений, поскольку эта тематика часто обсуждается в литературе, причем с устаревших позиций. Например, вместо уходящих в прошлое оценок максимального правдоподобия в настоящее время рекомендуют использовать одношаговые оценки.

2.1.6. Непараметрическая статистика

Статистические методы, которые не основаны на нереалистическом предположении о том, что рассматриваемые выборки взяты из распределений, описываемых одним параметром или небольшим числом параметров (2–4), называют *непараметрическими*. При математическом обосновании непараметрических статистических методов обычно вводят те или иные условия регулярности, например, требуют непрерывности функции распределения результатов наблюдений или существования математического ожидания и дисперсии. Как правило, подобные условия регулярности носят внутриматематический характер и не ограничивают прикладные возможности непараметрических методов.

Примерами являются критерии Колмогорова, Смирнова, Реньи, Вилкоксона, омега-квадрат (Крамера – Мизеса – Смирнова) [14], предназначенные для проверки гипотез согласия и однородности и разработанные в 30-е – 40-е годах XX в. История непараметрических коэффициентов корреляции Спирмена и Кендалла уходит корнями в работы начала XX в. В 50-х годах с известной работы Н. В. Смирнова [15] началась разработка методов непараметрического оценивания плотности. Непараметрическая статистика активно развивается и в XXI веке.

Во второй половине XX в. появились новые области статистических методов – робастная статистика, компьютерное статистическое моделирование (методы статистических испытаний (Монте-Карло), бутстреп-методы), статистика нечисловых и интервальных данных. Эти области активно развиваются и в настоящее время.

Иные причины привели к появлению и распространению прикладной статистики. Что означает этот термин? Вполне естественно, что математическая статистика выступает как метатеория по отношению к статистическим методам в той или иной области применения – к эконометрике, т.е. статистическим методам в экономике [16], к наукометрии [12], к биометрике и другим «метрикам». По цитированному выше определению Б. В. Гнеденко: «разработка приемов статистического наблюдения и анализа статистических данных составляет содержание математической статистики». Почему понадобилась новая научная область – прикладная статистика – между математической статистикой и статистическими методами в конкретных областях применений? Для ответа на этот вопрос необходимо обсудить внутреннюю логику развития статистических методов как научно-прикладной дисциплины.

2.1.7. Появление прикладной статистики

В нашей стране термин «прикладная статистика» вошел в широкое употребление в 1981 г. после выхода массовым тиражом (33940 экз.) сборника «Современные проблемы кибернетики (прикладная статистика)». В этом сборнике обосновывалась трехкомпонентная структура прикладной статистики [17]. Во-первых, в нее входят ориентированные на прикладную деятельность статистические методы анализа данных (эту

область можно назвать прикладной математической статистикой и включать также и в прикладную математику). Однако прикладную статистику нельзя целиком относить к математике. Она включает в себя две нематематические области. Во-первых, методологию организации статистического исследования: как планировать исследование, как собирать данные, как подготавливать данные к обработке, какие вероятностно-статистические модели использовать, какие статистические методы выбирать для обработки данных, как представлять результаты. Во-вторых, организацию компьютерной обработки данных, в том числе разработку и использование баз данных и электронных таблиц, статистических программных продуктов, например, диалоговых систем анализа данных. В нашей стране термин «прикладная статистика» использовался и ранее 1981 г., но лишь внутри сравнительно небольших и замкнутых групп специалистов [17].

Прикладная статистика и математическая статистика – это две разные научные дисциплины. Различие четко проявляется не только в исследованиях, но и при преподавании. Курс математической статистики состоит в основном из доказательств теорем, как и соответствующие учебники и учебные пособия. В курсах прикладной статистики основное – методология анализа данных и алгоритмы расчетов, а теоремы приводятся как обоснования этих алгоритмов, доказательства же, как правило, опускаются (их можно найти в научной литературе).

К настоящему времени беспристрастному наблюдателю очевидно четко выраженное размежевание этих двух научных дисциплин. Математическая статистика исходит из сформулированных в 1930–1950 гг. постановок математических задач, происхождение которых связано с рассматриваемыми в те времена проблемами анализа статистических данных. Начиная с 70-х годов XX в. исследования по математической статистике посвящены обобщению и дальнейшему математическому изучению этих старых задач. Поток новых математических результатов (теорем) не ослабевает, но новые практические рекомендации по обработке статистических данных при этом почти не появляются. Можно сказать, что математическая статистика как научное направление замкнулась внутри себя.

Сам термин «прикладная статистика» возник как реакция на описанную выше тенденцию. Прикладная статистика нацелена на решение реальных задач. Поэтому в ней возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится математическими методами, т.е. путем доказательства теорем. Большую роль играет методологическая составляющая – как именно ставить задачи, какие предположения принять с целью дальнейшего математического изучения. Велика роль современных информационных технологий, в частности, компьютерного эксперимента.

Рассматриваемое соотношение математической и прикладной статистик отнюдь не являются исключением. Как правило, математические дисциплины проходят в своем развитии ряд этапов. Вначале в какой-либо прикладной области возникает необходимость в применении математических методов и накапливаются соответствующие эмпирические приемы (для геометрии это – «измерение земли», т.е. землемерие, в Древнем Египте). Затем возникает математическая дисциплина со своей аксиоматикой (для геометрии это – время Евклида). Затем идет внутриматематическое развитие и преподавание (известно, что большинство результатов элементарной геометрии получено учителями гимназий в XIX в.). При этом на запросы исходной прикладной области перестают обращать внимание, и та для решения своих задач порождает новые научные дисциплины (сейчас «измерением земли» занимается не геометрия, а геодезия и картография). Затем научный интерес к исходной дисциплине иссякает, но преподавание по традиции продолжается (элементарная геометрия «ушла» из вузов, но до сих пор изучается в средней школе, хотя трудно понять, в каких практических задачах может по-

надобиться, например, теорема о том, что высоты треугольника пересекаются в одной точке). Следующий этап – окончательное вытеснение дисциплины из реальной жизни в историю науки (объем преподавания элементарной геометрии в настоящее время постепенно сокращается, в частности, ей все меньше уделяется внимания на вступительных экзаменах в вузах). К интеллектуальным дисциплинам, закончившим свой жизненный путь, относится средневековая схоластика. Как справедливо отмечает профессор МГУ им. М. В. Ломоносова В. Н. Тутубалин [18], теория вероятностей и математическая статистика успешно двигаются по ее пути – вслед за элементарной геометрией.

Резюмируем сказанное. Хотя статистические данные собираются и анализируются с незапамятных времен (см., например, Книгу Чисел в Ветхом Завете), современная математическая статистика как наука была создана, по общему мнению специалистов, сравнительно недавно – в первой половине XX в. Именно тогда были разработаны основные идеи и получены результаты, излагаемые ныне в учебных курсах математической статистики. После чего специалисты по математической статистике занялись внутриматематическими проблемами, а для теоретического обслуживания проблем практического анализа статистических данных стала формироваться новая дисциплина – прикладная статистика.

В настоящее время статистическая обработка данных проводится, как правило, с помощью соответствующих программных продуктов. Разрыв между математической и прикладной статистикой проявляется, в частности, в том, что большинство методов, включенных в статистические пакеты программ (например, в заслуженные *Statgraphics* и *SPSS* или в более новую систему *Statistica*), даже не упоминается в учебниках по математической статистике. В результате специалист по математической статистике оказывается зачастую беспомощным при обработке реальных данных, а пакеты программ применяют (что еще хуже – и разрабатывают) лица, не имеющие необходимой теоретической подготовки. Естественно, что они допускают разнообразные ошибки, в том числе в таких ответственных документах, как государственные стандарты по статистическим методам (о грубых ошибках в ГОСТах рассказано в статье [19]).

2.1.8. Что дает прикладная статистика народному хозяйству?

Так называлась наша статья [20], в которой приводились многочисленные примеры успешного использования прикладной статистики и других статистических методов при решении практических задач. Перечень примеров можно продолжать практически безгранично (см., например, сводку [21]).

Методы прикладной статистики используются в зарубежных и отечественных экономических и технических исследованиях, работах по управлению (менеджменту), в медицине, социологии, психологии, истории, геологии и других областях. Их применение дает заметный экономический эффект. Например, в США – не менее 20 миллиардов долларов ежегодно только в области статистического контроля качества. В 1988 г. затраты на статистический анализ данных в нашей стране оценивались в 2 миллиарда рублей ежегодно [22]. Согласно расчетам сравнительной стоимости валют на основе потребительских паритетов [16], эту величину можно сопоставить с 6 миллиардами долларов США. Следовательно, объем отечественного «рынка статистических услуг» был заметно меньше, чем в США, что совпадает с оценками и по другим показателям, например, по числу специалистов.

Своеобразие исторического пути России привело к тому, что в нашей стране нет специализированного научного журнала по статистическим методам. Публикации по новым статистическим методам, по их применениям в технико-экономических исследованиях, в инженерном деле постоянно появляются, прежде всего, в журнале «Заво-

дская лаборатория. Диагностика материалов», в секции «Математические методы исследования». Надо назвать также журналы «Автоматика и телемеханика» (издается Институтом проблем управления Российской академии наук), «Экономика и математические методы» (издается Центральным экономико-математическим институтом РАН), Научный журнал КубГАУ (издается Кубанским государственным аграрным университетом, г. Краснодар) и др.

Однако необходимо констатировать, что для большинства менеджеров, экономистов и инженеров прикладная статистика и другие статистические методы являются пока экзотикой. Это объясняется тем, что в вузах современным статистическим методам почти не учат. Во всяком случае, по состоянию на 2014 г. каждый квалифицированный специалист в этой области – самоучка.

Этому выводу не мешает то, что в вузовских программах обычно есть два курса, связанных со статистическими методами. Один из них – «Теория вероятностей и математическая статистика». Этот небольшой курс обычно читают специалисты с математических кафедр. Они успевают дать лишь общее представление об основных понятиях математической статистики первой половины XX в. Кроме того, внимание математиков обычно сосредоточено на внутриматематических проблемах, их больше интересует доказательство теорем, а не применение современных статистических методов в задачах экономики и менеджмента. Другой курс – «Статистика» или «Общая теория статистики», входящий в стандартный блок экономических дисциплин. Фактически он является введением в прикладную статистику и содержит первые начала эконометрических методов (по состоянию на 1900 г.).

Статистические методы как учебный предмет опираются на два названных вводных курса. Она призвана вооружить специалиста современным статистическим инструментарием. Специалист – это инженер, экономист, менеджер, геолог, медик, социолог, психолог, историк, химик, физик и т.д. Во многих странах мира – Японии и США, Франции и Швейцарии, Перу и Ботсване и др. – статистическим методам обучают в средней школе. ЮНЕСКО постоянно проводят конференции по вопросам такого обучения [23]. В СССР и СЭВ, а теперь – по плохой традиции – и в России игнорируют этот предмет в средней школе (в последние годы ситуация начинает меняться) и лишь слегка затрагивают его в высшей. Результат на рынке труда очевиден – снижение конкурентоспособности специалистов.

Проблемы прикладной статистики и других статистических методов постоянно обсуждаются специалистами. Широкий интерес вызвала дискуссия в журнале «Вестник статистики», в рамках которой были, в частности, опубликованы статьи [8, 20]. На появление в нашей стране прикладной статистики отреагировали и в США [24].

Контрастом к сказанному является тот неоспоримый факт, что в нашей стране получены многие фундаментальные результаты прикладной статистики. Огромное значение имеют работы академика РАН А. Н. Колмогорова [25]. Во многих случаях именно его работы дали первоначальный толчок дальнейшему развитию ряда направлений прикладной статистики. Зачастую еще 60–80 лет назад А. Н. Колмогоров рассматривал те проблемы, которые только сейчас начинают широко обсуждаться. Как правило, его работы не устарели и сейчас. Свою жизнь посвятили прикладной статистике члены-корреспонденты АН СССР Н. В. Смирнов и Л. Н. Большев. В наших учебниках постоянно встречаются ссылки на лучшую публикацию XX в. по статистическим методам – составленные ими подробно откомментированные «Таблицы ...» [14].

2.1.9. Статистические методы в России

Специалисты по истории официальной государственной статистики установили [5], что в России, как и в других странах, статистические исследования проводились с момента возникновения государств. Цели этих исследований, как и описанных в Библии работ под руководством Моисея, вытекали из потребностей государственного управления, прежде всего налогообложения и обороны страны. С XII века (в традиционной хронологии) на Руси проводились переписи населения [5]. Развитие статистической науки началось в России сразу же с выделением в начале XVIII века исследовательской деятельности как необходимой составляющей забот государства. Проще говоря, сразу же с организацией первого научного учреждения – Академии наук.

Первое статистико-экономическое обозрение России было составлено Иваном Кириловичем Кириловым (1689 – 1737), обер-секретарем Сената (в написании отчества и фамилии И. К. Кирилова, названий трудов сохранена исходная орфография), под названием «Цветущее состояние Всероссийского государства...». Первый в России научный труд по вопросам организации учета населения – «Разсуждение о ревизии поголовной и касающемся до оной» – был написан в 1747 г. Василием Никитичем Татищевым (1686–1750), известным государственным деятелем той эпохи. Он, в частности, одним из первых применял анкеты для сбора статистических данных. Большой вклад в теорию и практику отечественной статистики внес Михаил Васильевич Ломоносов (1711–1765).

Подробное описание развития статистической науки и практики в России можно найти в трудах по истории социально-экономической ветви статистики (см., например, [5, 26]). К сожалению, в этих работах обычно не рассматривается развитие отечественной вероятностно-статистической научной школы (о ней см., например, [3]).

Реформы императора Александра Второго, прежде всего создание земств (органов местного самоуправления), дали мощный стимул развитию статистики. Связано это было прежде всего с тем, что штатное расписание губернских и уездных земств, как правило, включало должность статистика. Так, к концу 1894 г. за 15 лет активной статистической деятельности были собраны, разработаны и опубликованы земствами материалы крестьянских подворных переписей по 172 уездам, охватившим около 4 миллионов крестьянских дворов – примерно четвертую часть всего населения России [5, стр.109].

Проведение статистических исследований было делом чести для отечественной интеллигенции. Например, Антон Павлович Чехов по собственной инициативе провел в 1890 г. перепись на Сахалине, лично опросив несколько тысяч каторжников [27].

Расцвет статистики в конце XIX века проявился в появлении большого числа оригинальных исследований, выполненных на высоком профессиональном уровне. Одна из них хорошо известна и в настоящее время, что объясняется личностью автора. Речь идет о книге В. И. Ульянова (Ленина) «Развитие капитализма в России. Процесс образования внутреннего рынка для крупной промышленности» [2]. Она была издана в 1899 г., когда автору было 29 лет. По современным критериям за эту монографию автору можно было бы присудить ученую степень доктора экономических наук. Это утверждение свидетельствует не только о высоком профессиональном уровне В. И. Ульянова как исследователя, но и об известной деградации социально-экономической статистики за последние сто лет.

В области статистической теории наибольшие достижения в XX веке были получены в России в математической статистике. Упомянем работы проф. А. А. Чупрова (1874–1926) по теории корреляции. Несколько позже началась деятельность А. Н. Колмогорова.

Среди математиков XX столетия академик АН СССР А. Н. Колмогоров (1903–1987) должен быть назван первым. Именно его работы дали первоначальный толчок дальнейшему развитию ряда направлений, важных для современных статистических методов. Зачастую еще 60–80 лет назад А.Н. Колмогоров рассматривал те проблемы, которые только сейчас начинают широко обсуждаться.

Отечественным исследованиям XX в. в области прикладной статистики и других статистических методов посвящены дальнейшие разделы настоящей главы.

Отметим здесь, что развитие статистических методов активно продолжается и в настоящее время. В XXI в. выявлена и сформулирована новая парадигма математической статистики [28], развивается статистика нечисловых данных [29], включая теорию классификации [30] и статистику интервальных данных [31], развита системная нечеткая интервальная математика [32, 33] и на ее основе – теория когнитивных функций [34], и т.д.

2.2. Вероятностно-статистические методы в работах А. Н. Колмогорова

В нашей стране развитием и применением статистических методов в XX–XXI вв. занимались десятки тысяч специалистов. Среди них выделяется гигантская фигура Андрея Николаевича Колмогорова. В развитии теории вероятностей и математической статистики четко виден переход от предыстории к современности, и этот переход – работы академика АН СССР А. Н. Колмогорова (1903–1987). Именно его работы дали первоначальный толчок дальнейшему развитию ряда направлений, важных для современных статистических методов. Зачастую еще 60–80 лет назад А.Н. Колмогоров рассматривал те проблемы, которые только сейчас начинают широко обсуждаться. Поразительно, но длинный ряд современных публикаций с научной точки зрения являются устаревшими по сравнению с работами А. Н. Колмогорова 1930-х годов.

Предыстории статистических методов посвящена статья [1]. В настоящей публикации с современной точки зрения [2] рассмотрим вероятностно-статистические методы исследования, созданные А. Н. Колмогоровым. Обсудим работы А.Н. Колмогорова по аксиоматическому подходу к теории вероятностей, критерию согласия эмпирического распределения с теоретическим, свойствам медианы как оценки центра распределения, эффекту «вздувания» коэффициента корреляции, теории средних величин, статистической теории кристаллизации металлов, методу наименьших квадратов, свойствам сумм случайного числа случайных слагаемых, статистическому контролю, несмещенным оценкам, аксиоматическому получению логарифмически нормального закона распределения при дроблении, методам обнаружения различий при экспериментах типа погодных.

Факты жизни и творчества А. Н. Колмогорова подробно рассмотрены в сборнике [3]. Его основные работы изданы в трех томах [4–6]. Работы отобраны им самим и прокомментированы его учениками. Андрей Николаевич считал, что хорошая математическая работа должна содержать простую идею (желательно геометрического характера) и использовать «тонкую» аналитику, а хорошая и полезная прикладная работа должна опираться на фундаментальные теоретические основы.

2.2.1. Аксиоматический подход к теории вероятностей

Аксиоматический подход позволил рассматривать теорию вероятностей и математическую статистику как часть математики, проводить рассуждения на математическом уровне строгости. В частности, было введено четкое различие между частотой и вероятностью, случайная величина стала рассматриваться как функция от элементарно-

го исхода, и т.д. За основу методов статистического анализа данных стало возможным брать вероятностно-статистические модели, сформулированные в математических терминах. В результате удалось четко отделить строгие утверждения от обсуждения философских вопросов случайности, преодолеть подход на основе понятия равновозможности, имеющий ограниченное практическое значение. Наиболее существенно, что после работ А. Н. Колмогорова нет необходимости связывать вероятности тех или иных событий с пределами частот или иными методами оценок вероятностей. В частности, так называемые «субъективные вероятности» имеют смысл экспертных оценок вероятностей.

После выхода (в 1933 г. на немецком языке и в 1936 г. – на русском) основополагающей монографии [7] аксиоматический подход к теории вероятностей стал общепринятым в научных исследованиях в этой области. Во многом перестроилось преподавание. Повысился научный уровень многих прикладных работ. Однако традиционный подход оказался на удивление живучим. С целью повышения строгости формулировок приходится помещать в наших учебниках ([8], [9] и др.) сводки терминов и определений в области вероятностно-статистических методов, опирающаяся на аксиоматику [7].

В послевоенные годы А. Н. Колмогоров формализовал понятие случайности на основе теории информации [6]. Грубо говоря, числовая последовательность является случайной, если ее нельзя заметно сжать (т.е. описать существенно короче) без потери информации. Однако этот подход не был предназначен для использования в прикладных работах и преподавании. Он представляет собой важное методологическое и теоретическое продвижение.

2.2.2. Критерии согласия

В работе 1933 г. «Об эмпирическом определении закона распределения» [5, с.134-141] А. Н. Колмогоров предложил и изучил «критерий Колмогорова», предназначенный для проверки согласия эмпирического распределения с полностью известным теоретическим. Пусть элементы выборки (независимые случайные величины) объема n имеют непрерывную функцию распределения $F(x)$. Эмпирической функцией распределения $F_n(x)$ называется доля элементов выборки, не превосходящих x . Критерий Колмогорова предназначен для проверки гипотезы

$$H_0 : F(x) \equiv F_0(x),$$

где $F_0(x)$ – заданная функция распределения. Его статистика имеет вид

$$D_n = \sqrt{n} \sup_x | F_n(x) - F_0(x) |.$$

В [5, с.134-141] показано, что функция распределения статистики D_n имеет предел,

$$\lim_{n \rightarrow \infty} P\{D_n \leq \lambda\} = K(\lambda),$$

и рассчитана первая в истории таблица функции распределения Колмогорова $K(\lambda)$.

Работа [5, с.134-141] породила одно из основных направлений непараметрической статистики. И в настоящее время непараметрические критерии согласия (Колмогорова, Смирнова, омега-квадрат и др.) широко используются. Они были разработаны для проверки согласия с *полностью известным* теоретическим распределением. Основная идея критериев Колмогорова, омега-квадрат и аналогичных им состоит в измерении расстояния между функцией эмпирического распределения и функцией теоретического распределения. Различаются эти критерии видом расстояний в пространстве функций распределения. Расчетные формулы, таблицы распределений и критических значений

широко распространены (см., например, лучший сборник таблиц математической статистики на русском языке [10]).

Часто возникает задача проверки гипотезы согласия эмпирического распределения с параметрическим семейством, например, с семейством нормальных, Вейбулла – Гнеденко или гамма-распределений. Представляется естественным оценить параметры распределения по выборке, а затем формально воспользоваться критериями согласия Колмогорова, Смирнова или омега-квадрат. При этом вместо фиксированной теоретической функции распределения подставляют функцию из параметрического семейства, в которой параметры заменены на их выборочные оценки. В отличие от классических критериев, при этом измеряются расстояния от эмпирической функции распределения до многообразий (в пространстве функций распределения), соответствующих параметрическим семействам. Развита [11] математическая техника проектирования в функциональных пространствах, которая позволяет строить методы проверки рассматриваемых гипотез.

Однако распределения таких критериев (как предельные, так и при конечных объемах выборок) *существенно отличаются* от распределений классических критериев согласия Колмогорова, Смирнова или омега-квадрат (подробнее см. [12]). Такие критерии в отличие от классических обычно называют «критериями согласия с параметрическим семейством типа Колмогорова – Смирнова и типа омега-квадрат». (Как показано в [12, 13] на основе анализа исходных публикаций, корректно употреблять термины «критерий Колмогорова», «критерий Смирнова», «критерий типа Колмогорова – Смирнова», но нельзя говорить о несуществующем «критерии Колмогорова – Смирнова».) В [14] собраны основные факты о критериях согласия с параметрическими семействами типа Колмогорова – Смирнова и типа омега-квадрат и необходимые краткие таблицы. Современное положение дел в этой области отражено в [2, 12]. Наиболее существенное продвижение в изучении критериев типа Колмогорова – Смирнова достигнуто профессором МГУ им. М. В. Ломоносова Ю. Н. Тюриным [15] и его научной школой.

2.2.3. «Вздувание» коэффициента корреляции

Интересное явление обнаружено А. Н. Колмогоровым в работе 1933 г. «К вопросу о пригодности найденных статистическим путем формул прогноза» [5, с. 161-167]. Предположим, что имеется много наборов предикторов (факторов, признаков, переменных, регрессоров). Для каждого из них строится наилучшее приближение отклика с помощью линейной функции от предикторов. Показателем качества приближения служит коэффициент корреляции между откликом и наилучшей линейной функцией от предикторов (в настоящее время чаще используют его квадрат, называемый коэффициентом детерминации). Эффект «вздувания» коэффициента корреляции состоит в том, что при увеличении числа проанализированных наборов предикторов заметно растет максимальный из соответствующих коэффициентов корреляции - показателей качества приближения. Создается впечатление, что тот набор предикторов, на котором достигается рассматриваемый максимум, дает хорошее приближение для отклика. Однако это впечатление развеивается при попытке использовать соответствующую зависимость для прогноза – по новым данным коэффициент корреляции между откликом и ранее найденной линейной функцией от предикторов оказывается значительно меньшим.

В настоящее время весьма популярны методы поиска «наиболее информативного множества признаков» в регрессионном и дискриминантном анализе. Соответствующие алгоритмы, как правило, основаны на переборе большого числа наборов признаков. Поэтому, как показано в [16], актуальность работы А. Н. Колмогорова [5, с. 161-167] в настоящее время существенно повысилась. Эффект «вздувания» коэффициента

корреляции является одним из проявлений неклассического поведения статистических характеристик в ситуации, когда одна и та же статистическая процедура осуществляется многократно, например, при множественных проверках статистических гипотез (см. [17, разд. 4.3]).

В течение полувека А. Н. Колмогоров интересовался статистическими постановками, в которых число неизвестных параметров растет вместе с объемом данных. К ним относится и работа [5, с. 161-167]. А в 1970-х годах он стимулировал исследования по т.н. «асимптотике Колмогорова» (в современной терминологии)

$$p \rightarrow \infty, n \rightarrow \infty, p/n \rightarrow \lambda > 0$$

где p - число параметров, n - объем выборки. Эта асимптотика весьма актуальна как для многомерного статистического анализа, так и для статистики нечисловых данных [18], а также для задач статистического приемочного контроля [8, раздел 13.5] и анализа социологических данных (см. [17, гл. 13]).

В настоящее время в асимптотике Колмогорова вместо от третьего предельного перехода обычно отказываются, но требуют отделенности дроби p/λ от 0 и ∞ , т.е. требуют существования числа $\varepsilon > 0$ такого, что

$$p \rightarrow \infty, n \rightarrow \infty, p/n > \varepsilon, p/n < 1/\varepsilon.$$

2.2.4. Метод медианы в теории оценивания

Пусть X_1, X_2, \dots, X_n - независимые одинаково распределенные случайные величины с функцией распределения F и непрерывной плотностью f . Пусть μ и σ^2 - соответственно математическое ожидание и дисперсия, а m - медиана распределения F (т.е. $P\{X_1 \geq m\} \geq 1/2$ и $P\{X_1 \leq m\} \geq 1/2$). Медиана всегда существует, но не всегда определяется однозначно. Обычно в качестве оценки для μ используют (в случае нормального закона, прежде всего) выборочное среднее арифметическое

$$\bar{X}_n = n^{-1} \sum_{k=1}^n X_k,$$

обладающее при условии нормальности F оптимальными свойствами. Что делать, если распределение F отлично от нормального? В работе 1931 г. «Метод медианы в теории ошибок» [5, с.111-114] А. Н. Колмогоров предлагает в этом случае оценивать по выборке другую среднюю характеристику распределения - медиану m (для симметричных распределений эти две характеристики совпадают). Пусть $X_n(k)$ - k -ая порядковая статистика, построенная по рассматриваемой выборке. Если n чётно, то в качестве оценки m_n медианы m возьмем $X_n(n/2)$; если же $n = 2k+1$, то в качестве оценки m возьмем $X_n(k)$. С целью сравнения оценок \bar{X}_n и m_n рассмотрим преобразованные величины

$$\alpha_n = \sqrt{n}(\bar{X}_n - \mu), \beta_n = \sqrt{n}(m_n - m).$$

Согласно центральной предельной теореме предельное (при $n \rightarrow \infty$) распределение величины α_n является асимптотически нормальным с нулевым средним и дисперсией σ^2 . Можно показать [5, с.111-114], что распределение величины β_n является асимптотически нормальным с нулевым средним и дисперсией $\sigma_m = (1/2)/f(m)$, если $f(m)$ отлично от 0. Мерой сравнительной точности обоих методов является отношение $\lambda = \sigma_m/\sigma = (1/2)/[\sigma f(m)]$. В случае нормальной плотности f имеем $\lambda = (\pi/2)^{1/2} \approx 5/4$. Как показал А. Н. Колмогоров [5, с.111-114], для унимодальных распределений отношение λ может принимать любое значение из интервала $(0; \sqrt{3})$, но не может превосходить $\sqrt{3}$.

Дискуссия о том, какую выборочную характеристику использовать для оценивания среднего значения совокупности (понимаемого как типичное значение), продолжается. В силу закона больших чисел выборочное среднее арифметическое при росте объема выборки приближается к математическому ожиданию (если оно существует). Но оно весьма неустойчиво по отношению к отклонениям «хвостов» распределения [9, разд. 4.7]. Как говорят, эта оценка не является робастной, в отличие от медианы. Медиана хороша тем, что является допустимым средним при измерениях в порядковой шкале, в то время как выборочное среднее арифметическое не обладает этим свойством [19, разд. 3.1], зато в ряде случаев имеет меньшую дисперсию (как для выборки из нормального распределения) и хорошо известно среди неспециалистов. Заслуга А. Н. Колмогорова в том, что он выявил рассматриваемую спорную точку в статистической теории и исчерпывающим образом сравнил асимптотические распределения двух видов средних величин.

2.2.5. Средние по Колмогорову

Естественная система аксиом приводит к так называемым ассоциативным средним. Их общий вид нашел в 1930 г. А. Н. Колмогоров [4, с.136-138]. Теперь их называют «средними по Колмогорову» (или «средними Колмогорова»). Для чисел X_1, X_2, \dots, X_n среднее по Колмогорову вычисляется как

$$G\{(F(X_1)+F(X_2)+\dots+F(X_n))/n\},$$

где F - строго монотонная функция (т.е. строго возрастающая или строго убывающая), G - функция, обратная к F . Среди средних по Колмогорову - много хорошо известных средних величин. Так, если $F(x) = x$, то среднее по Колмогорову - это среднее арифметическое. Для положительных X_1, X_2, \dots, X_n : если $F(x) = \ln x$, то среднее по Колмогорову - это среднее геометрическое, если $F(x) = 1/x$, то среднее гармоническое, если $F(x) = x^2$, то среднее квадратическое, и т.д. Однако такие популярные средние, как медиана и мода, нельзя представить в виде средних по Колмогорову. В прикладной статистике к средним по Колмогорову обращаются в связи с задачей выбора алгоритмов для анализа данных, измеренных в той или иной шкале (см. [17, разд. 10.3]). Так, для алгоритмов усреднения установлено, что в шкале интервалов из всех средних по Колмогорову допустимым является только среднее арифметическое, а в шкале отношений - только степенные средние с $F(x) = x^c$, (при c , отличном от 0) и среднее геометрическое. Таким образом, среднее геометрическое или среднее квадратическое температур (в шкале Цельсия) или расстояний не имеют смысла. В качестве среднего в шкале интервалов надо применять среднее арифметическое. А также можно использовать медиану или моду.

В теории средних величин рассматриваемая работа А. Н. Колмогорова - самое значительное продвижение за последние 200 лет, после того, как О. Коши ввел общее понятие среднего.

2.2.6. Статистическая теория кристаллизации металлов.

В работе 1937 г. [5, с. 178-182] А. Н. Колмогоров разработал модель возникновения центров кристаллизации и нарастания закристаллизованной массы. При широких допущениях им была найдена точная формула для вероятности $p(t)$, с которой наудачу выбранная точка P из объема, заполненного подлежащим кристаллизации веществом, попадет в течение промежутка кристаллизации t внутрь уже закристаллизованной массы. С достаточным приближением можно считать, что доля вещества, закристаллизованного за время t , также равно $p(t)$. Рассчитано число центров кристаллизации, об-

разующихся в течение всего процесса кристаллизации. Полученные в работе [5, с. 178-182] результаты до сих пор представляют интерес для всех специалистов, связанных с изучением и использованием процессов кристаллизации металлов и иных веществ.

2.2.7. Метод наименьших квадратов

В двух работах А. Н. Колмогорова [5, с.267-283, с. 283-288], опубликованных в 1946-1947 гг., построена геометрическая теория метода наименьших квадратов, выявляющая роль ортогонального проектирования на подпространства конечномерного евклидова пространства с целью получения оценок параметров. Эта идея затем широко использовалась как в научных исследованиях, так и при преподавании, в частности, в работах В. Н. Тутубалина, Ю. Н. Тюрина.

Другая составляющая этих работ А. Н. Колмогорова - построение алгоритмов доверительного оценивания и проверки гипотез на основе предположения о нормальности распределения погрешностей измерения. К настоящему времени эти результаты ушли в прошлое, поскольку установлено, что в подавляющем большинстве случаев распределение погрешностей заметно отличается от нормального (см. [20] и [17, разд. 2.1]). Поэтому современный подход (см., например, [21, 22] и [17, гл. 6]) к методу наименьших квадратов является непараметрическим, т.е. в определенном смысле наблюдается возврат к доколмогоровским взглядам.

2.2.8. Суммы случайного числа случайных слагаемых

Они глубоко изучены в работе 1949 г. [5, с.308-313], выполненной совместно с Ю. В. Прохоровым (1929–2013), в дальнейшем академиком АН СССР. Эта статья стимулировала исследования по важному для приложений виду предельных теорем (см. [23, с.300-312], [24, с.223-228]). Речь идет прежде всего о статистическом последовательном анализе [25], в частности, об изучении времени наблюдения в задаче последовательного различения двух простых гипотез. Предельные теоремы [26, 27] о суммах случайного числа случайных слагаемых находят применения в задачах статистического контроля качества и надежности по Вальду, в моделях управления запасами в логистике и других прикладных областях. Нами получен ряд результатов о суммах случайного числа случайных слагаемых при изучении двухуровневой модели управления запасами и асимптотики квантования в связи с выбором числа градаций в социологических анкетах (см. [17, разд. 8.4 и 12.3] и монографию [28]).

2.2.9. Статистический контроль

А. Н. Колмогоров – основоположник современной теории статистического приемочного контроля в нашей стране. Около 150 лет статистические методы применяются в России для проверки соответствия продукции установленным требованиям, т.е. для сертификации. Так, еще в 1846 г. действительный член Петербургской академии наук М. В. Остроградский рассматривал задачу статистического контроля партий мешков муки или штук сукна армейскими поставщиками [29]. Однако современный этап начался в 1951 г. с брошюры А. Н. Колмогорова [30]. С тех пор в России в статистическом контроле качества было сделано многое, особенно в области теории [31 - 33]. (К сожалению, до сих пор нет доступной информации об отечественных разработках и стандартах в области статистического контроля качества в оборонных отраслях промышленности до, во время и после Великой Отечественной Войны.) Вопросы статистического контроля постоянно рассматриваются на страницах журнала «Заводская лабора-

тория» (с 1994 г. – «Заводская лаборатория. Диагностика материалов») – основного места публикации отечественных работ по статистическим методам [34, 35].

Большое значение для развития статистических методов управления качеством имеют статья А. Н. Колмогорова 1933 г. [5, с.134-141] о критерии согласия эмпирического распределения с теоретическим и статья 1950 г. о несмещенных оценках [5, с. 340-363]. Актуальность первой из них определяется недостатками в используемых до сих пор статистических методах управления качеством. Широко распространенные ошибки состоят в том, что для критериев согласия с параметрическими семействами используют критические значения классических критериев. При этом, например, гипотеза нормальности принимается гораздо чаще, чем следует [12]. Поскольку в действующей нормативно-технической документации дальнейшие этапы анализа данных часто зависят от того, принимается нормальность или нет, то ошибки при такой проверке могут иметь далеко идущие последствия. Так, при анализе характеристик эластомерных материалов при ошибочном подходе из 30 выборок нормальность была отвергнута лишь для 2, а при правильном - для 26, т.е. в подавляющем большинстве случаев. Указанные ошибки встречаются в массе публикаций (хотя специалистам суть дела хорошо известна уже почти 50 лет [36]). Наиболее известным примером является полностью ошибочный ГОСТ 11.006-74 (СТ СЭВ 1190-78) «Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим» (хотя он формально отменен в 1987 г., но продолжает использоваться неспециалистами как методический материал; об ошибочных стандартах по статистическим методам см. [35]).

Популярны и другие ошибки при применении рассматриваемых критериев согласия. Некоторые пытаются их использовать для сгруппированных данных, что приводит к излишне частому принятию гипотез [37]. Другие вместо эмпирической функции распределения рассматривают иные оценки теоретической функции распределения. Например, при использовании вероятностной бумаги удобно ординату точки, соответствующей i -ой порядковой статистике, установить равной $(i-0.5)/n$, а не i/n , как в классической эмпирической функции распределения. Возникает искушение построенную таким методом оценку использовать в критериях согласия вместо эмпирической функции распределения. Увы, распределение изменится (впрочем, в данном случае при росте объема выборки различие будет исчезать). Ряд ошибок рассмотрен в пояснительной части таблиц [10].

2.2.10. Несмещенные оценки

При оценивании по выборке параметров распределений (либо функций от них) рекомендуют использовать метод максимального правдоподобия, дающий при выполнении условий регулярности асимптотически оптимальные оценки. Однако часто возникают трудности с решением уравнений правдоподобия. Поэтому вместо оценок максимального правдоподобия применяют асимптотически им эквивалентные одношаговые оценки (см. [17, разд. 3.2]) или оценки иных видов. Среди последних популярными (см., например, [38, гл.2]) являются несмещенные оценки. При конечном объеме выборки оценки максимального правдоподобия в ряде случаев хуже несмещенных оценок, как показал проф. Я. П. Лумельский в статье [39].

Основная идея использования несмещенных оценок состоит по А. Н. Колмогорову [5, с.340-363] в следующем. Во многих важных случаях эти оценки существуют. С другой стороны, чрезмерное разнообразие несмещенных оценок может быть значительно сокращено, если воспользоваться несмещенными оценками, которые выражаются через надлежащим образом выбранные достаточные статистики. Надо употреблять только несмещенные оценки, выражающиеся через достаточные статистики: ока-

зывается, что при этом мы не суживаем круг задач, в котором несмещенные оценки существуют, и при переходе от произвольной (даже плохой) несмещенной оценки к осредненной оценке, выражающейся через достаточную статистику, мы можем только уменьшить дисперсию оценки. Имеет место [38, гл.2] теорема Рао – Блекуэлла – Колмогорова: оптимальная оценка, если она существует, является функцией от достаточной статистики.

А. Н. Колмогоров первым ([5, с.340-363], [30]) применил несмещенные оценки в задачах статистического контроля. Он впервые использовал несмещенные оценки для определения эффективности реально используемых планов выборочного контроля по альтернативному признаку. На основе идей А. Н. Колмогорова рядом авторов были построены несмещенные оценки для предъявленного и пропущенного брака, для априорного распределения числа дефектных изделий в контролируемых партиях, а также получены несмещенные оценки при контроле по альтернативному и количественному признакам (см. [33], а также комментарии проф. Ю. К. Беляева и проф. Я. П. Лумельского в [5, с.522-523]). Несмещенные оценки основных показателей контроля включены в некоторые государственные стандарты (например, в ГОСТ 24660-81 «Статистический приемочный контроль по альтернативному признаку на основе экономических показателей»).

Полученная А. Н. Колмогоровым несмещенная оценка плотности нормального распределения нашла широкое применение в задачах контроля по количественному признаку. В дальнейшем этот результат был перенесен на многомерное нормальное распределение, а также применен для задач статистической классификации. Метод проверки гипотез по совокупности малых выборок, разработанный нами в [28], также основан на использовании несмещенных оценок. Этот метод применяется при статистическом приемочном контроле по нескольким альтернативным признакам [8, раздел 13.5]. Отметим, что в этом случае, как и в теории люсианов [19], оказывается нецелесообразным переход к осредненной оценке, выражающейся через достаточную статистику.

Введенные А. Н. Колмогоровым верхние и нижние оценки могут быть использованы и в тех случаях, когда несмещенные оценки не существуют. Именно так обстоит дело при оценивании пропущенного брака при биномиальном распределении и плане одноступенчатого контроля. Рядом авторов были получены верхние и нижние оценки функций неизвестных параметров, а также оценки с минимальным смещением.

2.2.11. О логнормальном законе распределения

В 1940 г. проф. Н. К. Разумовский привел много случаев, в которых логарифмы размеров частиц (золотин в золотоносных россыпях, частиц горных пород при их дроблении и т.п.) приближенно подчиняются нормальному закону распределения. В 1941 г. А. Н. Колмогоров указал общую схему случайного процесса последовательного дробления частиц, при которой в пределе, при неограниченном продолжении дробления, нормальный закон для логарифмов размеров частиц может быть установлен теоретически [5, с.264-266]. (Напомним, что положительная случайная величина X имеет логнормальный закон распределения, если логарифм величины X имеет нормальный закон распределения; условия, при которых вероятностная модель приводит к нормальному закону, хорошо известны.)

Идея А. Н. Колмогорова о выводе вида распределения случайной величины в конкретной прикладной задаче из некоторой системы аксиом нашла свое развитие, в частности, в экономике. Установлено, что распределения различных видов доходов (индивидуальных, подушевых, заработной платы и т.п.) хорошо приближаются с по-

мощью логарифмически нормальных функций распределения. Исходя из этого факта, Росстат для сверхбольших и сверхмалых доходов вместо наблюдаемых значений доходов использует расчетные на основе логарифмически нормальных функций распределения. Это объясняется большими трудностями в замере величин сверхбольших и сверхмалых доходов, необходимостью привлечения для таких замеров правоохранительных органов. Сам же логарифмически нормальный закон выводят аксиоматически, полагая, что доход того или иного вида можно считать полученным как произведение большого числа независимых случайных величин.

Если же принять, что практически значимый результат определяется с помощью максимального значения из ряда независимых одинаково распределенных случайных величин, как это делают в теории рекордов [40], то приходим к возможности описать распределение максимума одним из трех известных законов, исходя из предельных теорем.

2.2.12. Обнаружение различий

В семидесятых – восьмидесятых годах XX в. под научным руководством А. Н. Колмогорова на механико-математическом факультете МГУ им. М.В. Ломоносова работала группа исследователей, занимавшаяся статистическим анализом эффективности экспериментальных методов управления погодой. Речь идет об изменении количества выпавших осадков, борьбе с градом и рассеянии туманов. Среди прочих [41] вероятностных моделей использовалась и следующая.

Имеется n объектов U_1, U_2, \dots, U_n и с каждым объектом U_k связана пара чисел a_k и b_k , $k = 1, 2, \dots, n$. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ – последовательность независимых случайных величин, причем величина ε_k принимает значение 1 (считаем, что имеет место воздействие) с вероятностью p_k и значение 0 (воздействие отсутствует) с вероятностью $q_k = 1 - p_k$, $k = 1, 2, \dots, n$. В результате наблюдений над объектами нам известны значения случайных пар (ε_k, X_k) , $k = 1, 2, \dots, n$, где $X_k = a_k$ при $\varepsilon_k = 0$ и $X_k = b_k$ при $\varepsilon_k = 1$. Задача состоит в сравнении двух последовательностей $a^{(n)} = (a_1, a_2, \dots, a_n)$ и $b^{(n)} = (b_1, b_2, \dots, b_n)$. Тем самым в этой модели (могущей быть использованной и в других случаях, когда необходимо установить наличие или отсутствие эффекта воздействия) предполагается, что числа a_k и b_k неслучайны и вся случайность связана с процессом рандомизации. С помощью оценок Горвица – Томпсона и их обобщений [42] можно построить [41] ряд статистических критериев для проверки гипотезы

$$H(n): A(n) = B(n),$$

где

$$A(n) = \sum_{k=1}^{k=n} a_k, \quad B(n) = \sum_{k=1}^{k=n} b_k.$$

А. Н. Колмогоров заметил, что дисперсии оценок в критериях могут быть заметно уменьшены, если имеются хорошие методы прогноза, позволяющие до начала наблюдений указывать оценки a_k^* и b_k^* для a_k и b_k соответственно. Полагая $a_k = a_k^* + \Delta a_k$ и $b_k = b_k^* + \Delta b_k$, мы можем упомянутые выше процедуры применить не к a_k и b_k , а к Δa_k и Δb_k . При этом получают оценки, правильные независимо от качества прогноза, но они будут лучше оценок без обращения к прогнозам лишь в случае хороших прогнозов, когда величины $|\Delta a_k|$ и $|\Delta b_k|$ значительно меньше $|a_k|$ и $|b_k|$ соответственно.

Эти модели и методы нашли ряд применений, в частности, при разработке рандомизированной процедуры оценки айсберговой угрозы, связанной с добычей нефти и газа на морском арктическом шельфе [43].

Упомянем также работы А. Н. Колмогорова по теории стрельбы, выполненные в военные годы, по генетике и биологии, по лингвистике. Особенно велик вклад, сделан-

ный А. Н. Колмогоровым и его учениками (М. Д. Миллионщиков, А. С. Монин, А. М. Обухов, А. М. Яглом и др.) в теорию турбулентности. Здесь прежде всего следует упомянуть знаменитый Колмогоровский «закон двух третей» о распределении энергии в спектре турбулентности, полученный из простых соображений размерности (подробнее см. [3, с.445, 475], [4]).

Бесспорно, что многие работы А. Н. Колмогорова [4–6] представляют несомненный интерес для всех, кто разрабатывает или применяет статистические методы. Его мысли еще долго будут приносить нашей стране и всему миру всем практическую пользу. Отечественная вероятностно-статистическая научная школа порождена идеями А. Н. Колмогорова. Это хорошо видно на примере работ его ученика академика АН УССР Б. В. Гнеденко, которым посвящен следующий раздел настоящей главы.

2.3. Вероятностно-статистические методы в работах Б. В. Гнеденко

При обсуждении основных этапов становления статистических методов [1] была выделена гигантская фигура А. Н. Колмогорова, заложившего основы современной научной дисциплины «теория вероятностей и математическая статистика» и решившего ряд фундаментальных научных задач. Многие работы А. Н. Колмогорова актуальны и в настоящее время [2], как и работы его ученика и сотрудника Бориса Владимировича Гнеденко. Материал данного раздела продолжает цикл публикаций по истории статистических методов в нашей стране (начатый статьями [1] и [2]) и посвящен основным научным результатам, полученным Б. В. Гнеденко.

При анализе актуальных для XXI в. научных результатов академика АН УССР Б. В. Гнеденко (1912–1995) основное внимание уделим предельным теоремам теории вероятностей, математической статистике, теории надежности, статистическим методам управления качеством и теории массового обслуживания. Одна из основных научных заслуг Б.В. Гнеденко – обоснование необходимости развития статистических методов как самостоятельного научного направления, подробное рассмотрение ряда проблем, относящихся к этому направлению.

В XXI веке наиболее ценным для нас является удивительное умение Б. В. Гнеденко объединить в своем творчестве глубокие теоретические изыскания и практические разработки. В настоящее время всё глубже становится разрыв между внутриматематическими изысканиями, от которых в обозримом будущем нельзя ждать практической пользы, и попытками решения прикладных задач методами, устаревшими на полвека. Уникальность Б. В. Гнеденко и состоит в том, что он своей личностью устранял этот пагубный разрыв. Он был одновременно великим теоретиком и великим прикладником. Чем больше проходит времени с того печального для отечественной науки момента, когда Б. В. Гнеденко завершил свои земные труды, тем яснее становится основополагающая роль его идей, его методологического подхода в нашей нынешней работе. Научный путь Б. В. Гнеденко заслуживает подробного осмысления.

Из теоретических исследований Б. В. Гнеденко больше всего известны работы по предельным теоремам теории вероятностей, в том числе классическая монография о суммах независимых случайных величин 1949 г., написанная совместно с А. Н. Колмогоровым, статьи по предельным распределениям крайних членов вариационного ряда. Основополагающие результаты получены им в математической статистике, например, в задаче проверки однородности двух выборок. Для прикладников Б. В. Гнеденко - лидер в области теории надежности, массового обслуживания, статистических методов управления качеством продукции. По его «Курсу теории вероятностей» учились многие поколения специалистов. Большое значение имеют работы по истории науки и по

другим направлениям, среди которых особенно выделяется методология научных исследований.

2.3.1. От практики – к теории, от теории – к практике (четыре этапа научного пути)

Научный путь Б. В. Гнеденко можно разбить на четыре этапа [3]. Первый (1930–1934) прошел на кафедре математики текстильного института в г. Иваново, куда он был направлен в 1930 г. после окончания Саратовского университета. Именно там Б. В. Гнеденко пришел к глубокому убеждению, что полноценная творческая жизнь математика связана с широким использованием математических методов в решении задач практики и одновременном развитии самих математических методов, без чего невозможно глубокое изучение и удовлетворение потребностей практики. В ивановский период он увлекся теорией вероятностей.

Второй этап (1934–1945) - исследовательская работа в Москве. В 1934 г. Б. В. Гнеденко поступил в аспирантуру Московского государственного университета им. М. В. Ломоносова. Его научными руководителями стали А. Я. Хинчин и А. Н. Колмогоров. Еженедельно собирался общегородской семинар по теории вероятностей, где с новыми результатами выступали известные ученые А. Н. Колмогоров, Е. Е. Слуцкий, Н. В. Смирнов, А. Я. Хинчин, а также аспиранты, молодые математики, физики, биологи и инженеры. Б. В. Гнеденко увлекся предельными теоремами для сумм независимых случайных величин. В июне 1937 г. он защитил кандидатскую диссертацию «О некоторых результатах по теории безгранично-делимых распределений», а в начале июня 1941 г. - докторскую диссертацию, состоящую из двух частей: теории суммирования независимых случайных величин и теории распределения максимального члена вариационного ряда. В годы Великой Отечественной войны Б. В. Гнеденко принимал активное участие в решении многочисленных задач, связанных с обороной страны.

Третий этап научного пути Б. В. – украинский (1945–1960). В 1945 г. Академия наук Украинской ССР избрала Б. В. Гнеденко своим членом-корреспондентом и направила во Львов, где он восстанавливал после войны Львовский университет и организовывал учреждения Академии наук УССР. Во Львове Б. В. Гнеденко читал разнообразные курсы: математический анализ, вариационное исчисление, теорию аналитических функций, теорию вероятностей, математическую статистику и др. Его научная работа в этот период также была весьма разнообразна. Ему удалось доказать в окончательной формулировке локальную предельную теорему для независимых, одинаково распределенных решетчатых слагаемых (1948 г.). Здесь начались исследования по непараметрическим методам статистики. Но, по нашему мнению, основное значение имела работа Б. В. Гнеденко над учебником «Курс теории вероятностей» [4] (первое издание - 1949 г.) и монографией «Предельные распределения для сумм независимых случайных величин» [5], подготовленной совместно с А. Н. Колмогоровым.

В 1950 г. Президиум АН УССР перевел Б. В. Гнеденко в Киев, где в Институте математики АН УССР был организован отдел теории вероятностей и математической статистики. Одновременно Б. В. Гнеденко заведовал кафедрой математического анализа в Киевском университете.

Естественно, что очень скоро вокруг него образовалась группа молодых ученых, увлекавшаяся теорией вероятностей и задачами математической статистики. Первыми киевскими учениками Б. В. Гнеденко были В. С. Королук и В. С. Михалевич, впоследствии известные ученые. Характерно для Б. В. Гнеденко, что в Киеве он организовал городской семинар по истории математики при Институте математики АН УССР. Этот семинар объединил многих ученых, работающих в области истории науки.

В 1953 – 1954 гг. Б. В. Гнеденко работал в ГДР, а по возвращении Президиум АН УССР поручил ему возглавить работу по организации Вычислительного центра. Ядром группы ученых были сотрудники академика АН СССР С. А. Лебедева, разработчика первой в Европе ЭВМ, получившей название МЭСМ (малая электронная счетная машина). Одновременно Б. В. Гнеденко возглавил работу по созданию курса программирования для ЭВМ, который начал читать студентам Киевского университета – будущим сотрудникам Вычислительного центра. Этот курс [6] - первая в СССР книга по программированию. Начались работы по проектированию универсальной машины «Киев» и специализированной машины для решения систем линейных алгебраических уравнений. В этот период Президиум АН УССР возложил на Б. В. Гнеденко обязанности директора Института математики АН УССР и председателя бюро физико-математического отделения.

Широкая организационная деятельность не ослабила научной и педагогической деятельности Б. В. Гнеденко. Именно к этому периоду относится начало разработки им двух новых направлений прикладных научных исследований - теории массового обслуживания и вопросов использования математических методов в современной медицине.

Четвертый этап научного пути (1960–1995) – снова Москва. В 1960 г. Б. В. Гнеденко переехал в Москву и возобновил работу в Московском государственном университете им. М. В. Ломоносова (МГУ). Сразу же Б. В. Гнеденко организовал московский семинар по математической теории надежности и теории массового обслуживания, привлекая многочисленных участников. Большое внимание Б. В. Гнеденко уделял разработке основ теории надежности, решению задач теории резервирования с восстановлением, оптимальной профилактики, управлению качеством промышленной продукции в процессе производства.

В 1965 г. А. Н. Колмогоров передает Б. В. Гнеденко руководство кафедрой теории вероятностей механико-математического факультета МГУ им. М. В. Ломоносова, которой Борис Владимирович заведовал до своих последних дней.

Методологическими проблемами математики Б. В. Гнеденко систематически интересовался с конца 1950-х годов. Он – член научного совета при Президиуме АН СССР по философским проблемам естествознания. С первых дней Общества по распространению научных и политических знаний (общество «Знание») он принимает активное участие в его работе. Жизненному и научному пути Б. В. Гнеденко посвящены статьи [3], [7] и другие публикации.

Общее количество опубликованных научных трудов Б. В. Гнеденко – около тысячи. Рассмотрим подробнее основные направления его научной деятельности.

2.3.2. Суммирование независимых случайных величин

В 30-е годы внимание Б. В. Гнеденко привлекли задачи, связанные с суммированием независимых случайных величин (с.в.). Интерес к таким задачам появился в математике еще в XVII в. Невозможность прямых вычислений распределений сумм независимых с.в. приводит к необходимости получения и изучения асимптотических формул для них, т.е. таких формул, которые позволяют находить с нужной точностью требующиеся нам вероятности, связанные с суммами с.в. Эти формулы даются предельными теоремами теории вероятностей. Таким образом, аппроксимация многократных сверток распределений потребовала развития глубокой математической теории, которая называется теорией предельных теорем для сумм независимых с.в. или теорией суммирования.

Начало развития этой теории связано с работами Я. Бернулли и А. Муавра начала XVIII в., в которых были доказаны закон больших чисел (ЗБЧ) и центральная предельная теорема (ЦПТ) для независимых с.в., принимающих два значения. Эти исследования были продолжены в XIX в. П. Лапласом, С. Пуассоном, К. Гауссом и другими учеными, но вплоть до 1860-х гг. рассматривались лишь с.в., принимающие два значения. Лишь в 1867 г. П.Л. Чебышев получил ЗБЧ в общем виде, а достаточно общая форма ЦПТ была найдена лишь в работах А.М. Ляпунова и А.А. Маркова на рубеже XIX и XX вв. Наиболее бурное развитие теории суммирования пришлось на 20 - 40 гг. XX в. и связано с именами А. Н. Колмогорова, Б. В. Гнеденко, А. Я. Хинчина, П. Леви, В. Феллера и Дж. Линдеберга.

Класс возможных предельных распределений для сумм независимых случайных величин, как показали А. Я. Хинчин и Г. М. Бавли, совпадает с классом безгранично-делимых распределений. Оставалось выяснить условия существования предельных распределений и условия сходимости к каждому возможному предельному распределению. Заслуга постановки этих задач и их решения принадлежит Б. В. Гнеденко. Он в 1937 г. предложил оригинальный метод, получивший название метода сопровождающих безгранично-делимых законов. Единым приемом удалось получить все ранее найденные в этой области результаты, а также и ряд новых.

В теории суммирования доказывались как интегральные предельные теоремы, то есть теоремы о сходимости функций распределения, так и локальные теоремы, то есть теоремы о сходимости плотностей (для гладких распределений) и об асимптотическом поведении вероятностей отдельных значений для решетчатых распределений. В 20–40 гг. XX в. получены исчерпывающие результаты о ЗБЧ в классической формулировке. Отметим, что законы больших чисел в пространствах нечисловой природы, найденные в последней четверти XX в., формулировались и доказывались исходя из совсем иных подходов - не на основе суммирования, а на основе решения оптимизационных задач (см., например, [8–10]).

Во всех разделах теории суммирования Б. В. Гнеденко получил фундаментальные результаты, пролившие свет на существо дела. Итогом развития классической теории суммирования явилась публикация в 1949 г. монографии Б. В. Гнеденко и А. Н. Колмогорова [5], которую можно назвать монументом создателям этой теории. Методы и результаты теории суммирования применяются в различных разделах теории вероятностей, статистических методов и их применений, а книга [5] остается источником новых идей для многих исследователей. Эта книга - одно из наиболее замечательных достижений математики XX века.

2.3.3. Предельные теоремы для крайних порядковых и делимых статистик

Работы по предельным теоремам для крайних порядковых статистик публикуются уже в течение почти сотни лет, начиная с двадцатых годов XX в. Среди авторов таких публикаций: Додж, фон Мизес, Фреше, Фишер, Типпет, Б. де Финетти, Гумбель, В. Б. Невзоров и другие. Здесь наиболее полные и глубокие результаты в этой области получены Б. В. Гнеденко [11].

Пусть x_1, \dots, x_n - независимые одинаково распределенные с функцией распределения F случайные величины; тогда величины $x_n(1) = \min x_k$ и $x_n(n) = \max x_k$ называются крайними (или экстремальными) порядковыми статистиками, а также крайними членами вариационного ряда. Предположим, что для функции распределения F найдутся последовательности констант $\{a_n > 0\}, \{b_n\}$, для которых существуют невырожденные предельные (с ростом n) функции распределения G крайних членов преобразованной

выборки $\{a_n^{-1}(x_k - b_n)\}, k = 1, 2, \dots, n$. Тогда согласно общей теории, построенной Б. В. Гнеденко, функция G имеет один из трех типов. Среди них широко используемое на практике распределение Вейбулла – Гнеденко [12] (в устаревшей литературе именуется распределением Вейбулла). Б. В. Гнеденко нашел необходимые и достаточные условия, относящиеся к F , позволяющие получить тот или иной тип G .

Являясь выдающимся специалистом по теории суммирования независимых случайных величин, Борис Владимирович решил результаты этой теории применить к суммированию зависимых случайных величин. Поэтому он проявил интерес [13] к таким случайным величинам w_1, \dots, w_N , совместное распределение которых совпадает с условным совместным распределением некоторых независимых случайных величин $\theta_1, \dots, \theta_N$ при условии фиксации суммы последних в некоторой точке. Отправляясь от величин w_1, \dots, w_N , можно построить [13] класс сумм зависимых случайных величин, называемых в отечественной литературе разделимыми статистиками. Распределения последних известным образом выражаются через распределения сумм соответствующих независимых случайных величин (векторов). Тем самым, для получения предельных (с ростом числа слагаемых) теорем для разделимых статистик надо воспользоваться результатами суммирования независимых величин или их многомерными аналогами — в случае векторов.

2.3.4. Теория массового обслуживания

Большим и весьма практически важным разделом современных статистических методов, в становление и развитие которого Б. В. Гнеденко внес неоценимый вклад, является теория массового обслуживания (ТМО). Первый цикл работ в этом направлении он выполнил на первом этапе своего научного пути — на кафедре математики текстильного института в г. Иваново. В частности, он занимался изучением связи неровности пряжи по номеру и весу, выяснением эффективности перехода от обслуживания одного станка к обслуживанию нескольких станков, оценкой длины среднего перехода между станками, который выполняет ткачиха в процессе обслуживания ткацких станков, выявлением особенностей метода станкообходов для нормирования рабочего времени станка и рабочего. Этой тематике посвящена первая книга Б. В. Гнеденко [14].

В опубликованной перед самой войной работе [15] Б. В. Гнеденко решает задачу определения среднего числа зарегистрированных счетчиком Гейгера-Мюллера частиц (известно, что в силу наличия «мертвой зоны» счетчик Гейгера-Мюллера регистрирует не все попадающие в него частицы). В терминах ТМО рассматриваемая модель может быть описана как однолинейная система массового обслуживания (СМО) с потерями, нестационарным пуассоновским входящим потоком и постоянным временем обслуживания. Заметим, что и к настоящему времени системы массового обслуживания с нестационарным входящим потоком исследованы крайне мало.

К задачам ТМО Б. В. Гнеденко возвращается в 50-е годы, хотя, по собственному признанию, уже во время войны он не раз размышлял над ними. И до последних дней жизни это направление, наряду с теорией суммирования и математической теорией надежности, становится одним из основных в его научной деятельности. Борис Владимирович обобщает формулы Эрланга на системы с ненадежными восстанавливаемыми приборами, рассматривая как случай с потерей требования при отказе прибора, так и случай перехода недообслуженного требования на другой свободный прибор, и т.д.

В 1956 г. Б. В. Гнеденко прочитал первый в СССР спецкурс по ТМО. В 1958 г. цикл его лекций по теории массового обслуживания был опубликован, а затем послу-

жил основой для широко известной монографии [16], выпущенной в 1966 г. Эта книга и до сих пор остается одной из основополагающих при подготовке специалистов по ТМО не только в нашей стране, но и за рубежом. Отметим еще две его монографии ([17, 18]), оказавших значительное влияние на развитие ТМО.

В последующие годы Б. В. Гнеденко опубликовал еще более 30 статей, относящихся к ТМО. В этих статьях, наряду с решением отдельных задач по ТМО, он дает детальные обзоры существующих методов исследования, формулирует новые проблемные направления. Важнейшей задачей Б. В. Гнеденко считал пропаганду на всех уровнях, начиная от школьников и кончая профессиональными математиками и управленцами высокого уровня, широчайшего внедрения методов ТМО в инженерную практику.

2.3.5. О работах Б. В. Гнеденко в области математической статистики, теории надежности и контроля качества

Статистические методы были в центре научных и педагогических интересов Б. В. Гнеденко на протяжении всей его творческой жизни. «Каждому специалисту нужно знать математическую статистику» – так называется одна из его статей 1961 г. [19]. Уже в первых его публикациях, посвященных математическому анализу проблем текстильного производства, проявился живой интерес и умение Бориса Владимировича работать с реальными данными.

Мировую известность Б. В. Гнеденко как статистику принес цикл работ, выполненный им вместе со своими учениками и сотрудниками в конце 40-х – первой половине 50-х годов. Он изучал проблему проверки гипотезы однородности двух независимых выборок с помощью статистики, равной максимуму разности соответствующих эмпирических функций распределения (т.н. двухвыборочная односторонняя статистика Н. В. Смирнова). Б. В. Гнеденко предложил метод вычисления точного распределения статистики критерия для конечных выборок равного объема, позволивший получить простое доказательство найденных ранее Н. В. Смирновым предельных теорем и достаточно точные асимптотические разложения (эти результаты и их дальнейшее развитие рассмотрены в статье [20]). А. Н. Колмогоров высоко оценил исследования Б. В. Гнеденко по непараметрической статистике [21]. И сейчас, через 50 лет, эти результаты Б. В. Гнеденко по-прежнему актуальны для применения математических методов исследования.

По статистике Б. В. Гнеденко опубликовал более 50 работ. Среди них есть посвященные проблемам статистического образования, а также приложениям статистических методов в технических исследованиях, теории надежности и контроле качества, экономике и социальных науках, биологии и медицине, во многих других областях.

Б. В. Гнеденко всегда был среди тех ученых, которые, с одной стороны, глубоко понимали необходимость развития вычислительной техники как основы и предпосылки внедрения результатов теоретических (и в том числе математико-статистических) исследований в практику; а с другой - предвидели широкие горизонты новых исследований, которые представляли высокопроизводительные компьютеры. Он не только руководил созданием Вычислительного центра АН УССР, но и был у истоков создания Института кибернетики АН УССР. Как уже отмечалось, Б. В. Гнеденко был написан первый в СССР учебник по программированию [6]. Начатые Б. В. Гнеденко в сотрудничестве с Н. М. Амосовым работы по машинной диагностике сердечных заболеваний во многих своих аспектах являются примером высококлассного прикладного статистического исследования, по своей тематике относящегося к проблемам классификации. К сожалению, Б. В. Гнеденко не дали завершить эти исследования. Являясь одним из

виднейших математиков, работавших в то время на Украине, он был вынужден покинуть Киев и переехать в 1960 г. в Москву.

Вопросами теории надежности и проблемами управления (а значит, и контроля) качества Б. В. Гнеденко начал заниматься еще во второй половине 50-х годов. По мере знакомства с уровнем качества продукции промышленных предприятий в нем крепла уверенность в необходимости использования математических методов для объективной оценки качества и прогноза надежности изделий. К разработке математической теории надежности он привлек своих учеников И. Н. Коваленко, В. С. Королюка, Т. П. Марьяновича. Сам Б. В. Гнеденко в это время выполнил ряд прикладных работ, связанных с анализом надежности и методикой расчета нагрузки электрических сетей промышленных предприятий.

В Москве, будучи одним из создателей и признанным лидером советской школы математической теории надежности, Б. В. Гнеденко приобрел огромное неформальное влияние на развитие этой теории не только на всей территории СССР, но и далеко за ее пределами. Другой мощной школой в теории надежности является североамериканская. Две школы отличались по тематике исследований и во многом дополняли друг друга. Достижения этих школ 60–80-х годов до сих пор определяют мировое развитие теории надежности.

Продвижению результатов математической теории надежности в практику Б. В. Гнеденко придавал не меньшее значение, чем развитию самой математической теории. По его мнению, важнейшими аспектами востребованности и успешного применения практикой являются

(а) наличие в теории богатого набора математических моделей, отражающих разнообразные явления предметной области;

(б) наличие в предметной области специалистов, способных понять математические модели и превратить их в «руководящие указания» на производстве;

(в) наличие литературы самого разного уровня, отражающей достижения теории и практику ее применения;

(г) возможность прямого контакта между создателями теории и специалистами предметной области для взаимной корректировки задач теории и методов ее приложения в предметной области.

Все перечисленные выше моменты нашли счастливое сочетание в работе огромного незримого коллектива ученых и практиков, имевших отношение к созданию и приложению теории надежности и к управлению качеством в СССР. Усилиями Б. В. Гнеденко, его сотрудников и учеников с 1960 по 1985 гг. была разработана весьма разветвленная математическая теория надежности и математическая теория контроля качества. Была налажена широкая пропаганда необходимости практического использования теоретических результатов, в том числе по линии общества «Знание». Организованы семинары и лекционные курсы в Политехническом музее (Москва), в МГУ им. М. В. Ломоносова, а затем и во многих городах СССР, где инженерный состав получал необходимую математическую подготовку для понимания и применения методов теории надежности и контроля качества. В кабинете надежности при Политехническом музее все заинтересованные лица могли получить консультации у ведущих специалистов, включая и самого Б. В. Гнеденко. Издательства «Советское радио» и «Знание» выпустили серию книг, посвященных различным аспектам теории надежности и контроля качества. Огромное влияние оказала основополагающая монография [22], а также ряд других монографий с участием Б. В. Гнеденко, в частности, небольшая яркая книга [23].

Была развернута большая работа по подготовке специалистов высшей категории в области теории надежности. В руководстве ряда отраслей промышленности оказались

специалисты, хорошо понимающие необходимость внедрения современных методов теории надежности и контроля качества. И во всем этом самое непосредственное участие принимал Б. В. Гнеденко. В результате достижения математической теории надежности и контроля качества нашли широкое признание, как в научных кругах, так и среди прикладников. Правда, с сожалением приходится констатировать, что в целом на реальный подъем качества продукции в стране, за исключением предприятий оборонно-промышленного комплекса, эти достижения повлияли мало.

Развитие теории управления качеством и надежностью активно продолжается и в настоящее время. В частности, в журнале «Заводская лаборатория. Диагностика материалов» постоянно обсуждаются различные прикладные и теоретические проблемы управления качеством [24, 25]. В современных условиях реализация накопленного научного потенциала может дать значительное ускорение экономического роста как отдельных предприятий, так и страны в целом.

Конечно, нельзя не отметить и огромный личный вклад Б. В. Гнеденко в математическую теорию надежности. Предметом его наибольшего интереса была теория резервированных систем с восстановлением. Здесь им была поставлена задача, которая имела многочисленные продолжения в работах других математиков, а именно – задача об асимптотическом распределении момента первого отказа резервной группы с быстрым восстановлением. Б.В. удалось установить связь с асимптотической теорией суммирования случайного числа случайных слагаемых. И эта задача была им с блеском решена. Отметим, что подобные суммы используются не только в теории надежности, но и в различных иных прикладных областях, в частности, в логистике, т.е. науке о движении материальных, финансовых и информационных потоков (см., например, монографии [26, 27]).

И как здесь не вспомнить слова Б. В. Гнеденко о взаимообогащении фундаментальных и прикладных наук: «Я глубоко убежден в том, что прикладные проблемы не только дают возможность демонстрации силы математических методов и решения множества задач, необходимых для жизненной практики, но имеют огромное значение для развития самой математики. Дело в том, что в прикладных задачах часто приходится сталкиваться с совсем новыми ситуациями, о которых математик-теоретик не может догадаться. Традиционные методы математики недостаточны для решения возникающих вопросов, требуется разработка новых методов исследования и, возможно, – даже новых ветвей математики. Но практика важна для науки и тем, что именно практика выясняет возможности той или иной области математики для решения актуальных проблем других научных дисциплин и повседневных нужд общества. И, в конечном счете, ценность исследований математика будет определяться по тому, насколько широко и глубоко развиваемые им теории позволяют проникнуть в проблемы познания законов окружающего мира, помогают решению житейских проблем, касающихся всего общества. Чем теснее связана та или иная ветвь математики с практикой жизни, тем разнообразнее ее проблемы, тем быстрее она развивается. Так было, так есть и так будет» [28].

2.3.6. История математики и преподавание

Вскоре после создания Академии педагогических наук РСФСР (основана в 1943 г.) Б. В. Гнеденко был приглашен в Институт методов обучения. Итог его работы – книга [29], адресованная в первую очередь учителям и школьникам. Эта замечательная книга была первым достаточно полным исследованием истории математики в нашей стране.

Несомненной заслугой Б. В. Гнеденко является то, что он показал, что история математики необходима действующему математику. На Третьем Всесоюзном математическом съезде (1956) Б. В. Гнеденко перечислил магистральные направления историко-научных исследований в этой области. Он подчеркнул значение истории математики

- «а) для целей выяснения общих закономерностей развития математики,
- б) для выявления общих перспектив ее последующего развития,
- в) для выявления методологических установок науки,
- г) для выяснения связей с другими науками и роли математики в истории культуры,
- д) для целей преподавания и воспитания» [30, с.100].

Эти задачи Б. В. Гнеденко реализовывал на протяжении пятидесяти лет, написав более 180 работ по истории математики. Среди них - более 32 биографических статей, посвященных Н. И. Лобачевскому, П. Л. Чебышеву, М. В. Остроградскому, А. Н. Колмогорову и др. В фундаментальной работе [31] он прослеживает предысторию теории вероятностей, анализируя труды ученых, стоящих у истоков этой науки: Л. Пачолли (основатель бухгалтерского учета), Дж. Кардано, Н. Тарталья, Г. Галилея, Б. Паскаля, П. Ферма, Х. Гюйгенса. Интересен раздел «Статистический контроль качества продукции» [31, с.47 – 50], в котором Б. В. Гнеденко прослеживает истоки этой важной прикладной области вплоть до 1740 г. Мастерски умел Б. В. Гнеденко показать в элементарных рассуждениях предшественников зерна более широких идей. Изложение столь понятно и интересно, что хочется заглянуть в первоисточники – труды Я. Бернулли, П. Л. Чебышева, П. Леви и других.

Наиболее известной книгой Б. В. Гнеденко – учебником «Курс теории вероятностей» [4] – пользуются студенты математических специальностей университетов уже свыше полувека. Он выдержал несколько десятков изданий в СССР, США, ГДР, Японии и многих других странах. Совместно с А. Я. Хинчиным Б. В. Гнеденко написал научно-популярную книгу [32], которая также вот уже более пятидесяти лет пользуется огромной популярностью и выдержала множество изданий в СССР и за рубежом.

Б. В. Гнеденко уделял большое внимание вопросам преподавания. Он руководил научно-исследовательскими семинарами по программированному обучению, по вопросам преподавания в средней школе, был председателем секции теории вероятностей и математической статистики и секции средней школы Московского математического общества. Большое число статей было им опубликовано в журналах «Вестник высшей школы», «Математика в школе», в сборниках научно-методического совета Минвуза СССР.

Лекции Б. В. Гнеденко пользовались большим успехом в любой аудитории. Естественна попытка проанализировать те средства, которые использовал Б. В. Гнеденко для воздействия на слушателей во время лекций. Суть их в простоте, в уважении своих слушателей, в желании передать им те сведения, которые им необходимы; в демонстрации на ярких и доступных примерах важности того, о чем идет речь; в умении связывать общие идеи с различными частными задачами, которые близки интересам слушателей; в ненавязчивом, постоянном воспитании научного мировоззрения. И все это вместе взятое высказывалось Б. В. Гнеденко на лекциях так, что в каждый момент звучало нужное слово с нужной интонацией. И не только в лекциях для студентов, но и в выступлениях для школьников средних классов [33, 34].

Охватывая в своем творчестве весь диапазон, который может попасть в поле зрения математика - от исходной практической проблемы до теоретической чисто математической задачи и затем от решения этой задачи обратно к практической проблеме - Б. В. Гнеденко вполне естественно обращался к осмыслению своего пути исследователя. Он посвящал методологическим исследованиям отдельные работы, постоянно об-

рашался к проблемам таких исследований в книгах более общего характера [35]. Методологические вопросы постоянно обсуждались также в публикациях, посвященных роли математических методов исследования в научно-техническом прогрессе [36] или применению современных статистических методов в управлении качеством продукции [23, 37].

Много сил отдавал Б. В. Гнеденко постоянному общению со специалистами самых разных отраслей народного хозяйства, областей науки. Неоценимую помощь оказал он своими консультациями тысячам людей.

Своей личностью, своей собственной научной, педагогической и организационной работой Б. В. Гнеденко показывал пример плодотворного единения теории и практики. И символично, что именно он в 1961 г. создал (вместе с проф. В. В. Налимовым) раздел «Математические методы исследования» в журнале «Заводская лаборатория» (в настоящее время – «Заводская лаборатория. Диагностика материалов») и возглавлял его более 30 лет. И в настоящее время для нас важны его методологические выступления на страницах этого журнала [36, 38], в котором публикуются основные отечественные работы по статистическим методам.

Авторы искренне благодарны В. М. Золотареву, Д. Б. Гнеденко, Н. К. Добровольской, Э. М. Кудлаеву, А. В. Печинкину, Н. Х. Розову, В. В. Сенатову, Е. В. Чепурину, В. Н. Чиненовой за предоставленные материалы, использованные при составлении раздела 2.3.

2.4. Непараметрическая и прикладная статистика в нашей стране

Продолжим обсуждение истории отечественной статистики. В разделе 2.1 и статье [1] рассмотрены основные этапы становления статистических методов. Работам выдающихся отечественных исследователей А. Н. Колмогорова и Б. В. Гнеденко посвящены разделы 2.2 и 2.3 выше, статьи [2] и [3] соответственно. В настоящем разделе обсудим развитие непараметрической и прикладной статистики в нашей стране в 1930–1980 гг.

В довоенный период отечественная вероятностно-статистическая наука прославилась двумя достижениями. Об одном – построении А. Н. Колмогоровым теории вероятностей на основе теории меры и интеграла Лебега – уже говорилось [2]. Второе – разработка непараметрических критериев проверки согласия и однородности. Сначала фундаментальный результат – критерий согласия эмпирического с распределением с теоретическим (критерий Колмогорова) – был получен А. Н. Колмогоровым [4, с.134–141], затем дело взял в свои руки член-корреспондент АН СССР Николай Васильевич Смирнов (1900–1966).

2.4.1. О работах Н. В. Смирнова

Его основные научные труды опубликованы в сборнике [5], на который и будем ссылаться. Наиболее ценная книга XX в. по статистическим методам, на наш взгляд, подготовлена членами-корреспондентами АН СССР Л. Н. Большевым и Н. В. Смирновым. Это – «Таблицы математической статистики» [6]. Название не должно обманывать – весьма полезна начинающая книгу пояснительная часть (разделы с кратким и строжайше выверенным описанием классических статистических методов, примерами

их применения, комментариями к таблицам). Учебники Н. В. Смирнова по статистическим методам и по сей день остаются среди лучших [7, 8].

Как ученый Н. В. Смирнов получил много глубоких результатов. Так, с его работы 1951 г. «О приближении плотностей распределения случайных величин» (см. [9] и [5, с.205-223]) началось развитие такого перспективного, в том числе в статистике нечисловых данных [10, гл.11], направления, как непараметрические оценки плотности. Однако с его именем связывают прежде всего «критерии Смирнова». Пусть $F_n(t)$ – эмпирическая функция распределения, построенная по выборке объема n из непрерывной функции распределения $F(t)$. Напомним, что согласно Л.Н. Большеву и Н.В. Смирнову [6] значение эмпирической функции распределения в точке x равно доле результатов наблюдений в выборке, меньших x . Одновыборочные критерии Смирнова, введенные в статье 1939 г. «Об отклонениях эмпирической функции распределения» [5, с.88-107], основаны на статистиках

$$D_n^- = \inf_{-\infty < t < +\infty} (F_n(t) - F(t)), \quad D_n^+ = \sup_{-\infty < t < +\infty} (F_n(t) - F(t))$$

Очевидно, критерий Колмогорова есть максимум этих двух статистик. Поэтому возникает желание объединить все три критерия в одну группу – группу критериев Колмогорова-Смирнова. Однако разработанные Н. В. Смирновым методы рассуждений, использованные для получения распределений рассматриваемых статистик, совершенно оригинальны. Они не имеют ничего общего с подходом А. Н. Колмогорова. Поэтому мы считаем, что надо говорить отдельно о критерии Колмогорова и отдельно о критериях Смирнова, а если уж надо объединить их вместе, то говорить о критериях *типа* Колмогорова-Смирнова, но не о критериях Колмогорова-Смирнова, поскольку употребление последнего выражения приводит к искажению исторической правды [11, 12].

Двухвыборочные критерии Смирнова однородности двух независимых выборок были им предложены и изучены в 1939 г. (см. [5, с.117-127]). Единственное ограничение – функции распределения $F(x)$ и $G(x)$ должны быть непрерывными. Критерии Смирнова основан на использовании эмпирических функций распределения $F_m(x)$ и $G_n(x)$, построенных по первой и второй выборкам соответственно. Значение двухвыборочной двухсторонней статистики Смирнова

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|$$

сравнивают с соответствующим критическим значением и по результатам сравнения принимают или отклоняют гипотезу H_0 о совпадении (однородности) функций распределения (подробнее – см., например, [6] или [12]). Практически значение статистики $D_{m,n}$ рекомендуется согласно [6] вычислять по формулам

$$D_{m,n}^+ = \max_{1 \leq r \leq m} \left(\frac{r}{m} - F_n(y_r) \right) = \max_{1 \leq s \leq n} \left(G_n(x_s) - \frac{s-1}{n} \right),$$

$$D_{m,n}^- = \max_{1 \leq r \leq m} \left(F_n(y_r) - \frac{r-1}{m} \right) = \max_{1 \leq s \leq n} \left(\frac{s}{n} - G_n(x_s) \right),$$

$$D_{m,n} = \max(D_{m,n}^+, D_{m,n}^-),$$

где $x'_1 < x'_2 < \dots < x'_n$ – элементы первой выборки x_1, x_2, \dots, x_n , переставленные в порядке возрастания, а $y'_1 < y'_2 < \dots < y'_m$ – элементы второй выборки y_1, y_2, \dots, y_m , также переставленные в порядке возрастания. Поскольку функции распределения $F(x)$ и $G(x)$ предполагаются непрерывными, то вероятность совпадения каких-либо выборочных значений

равна 0. Статистики $D_{m,n}^-, D_{m,n}^+$ также могут быть использованы для проверки однород-

ности двух независимых выборок. Их называют двухвыборочными односторонними статистиками Смирнова.

Статистика омега-квадрат (подробнее см. о ней в [12] и [13, гл.2.3])

$$\omega_n^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 dF(x)$$

также используется для проверки согласия эмпирического распределения с фиксированным теоретическим. Эту статистику в 1928–1931 гг. предлагали использовать Г. Крамер и Р. фон Мизес, однако ее предельное распределение вычислил в 1937 г. Н. В. Смирнов в статье «О распределении ω^2 – критерия Мизеса» [5, с. 60 – 78], что и позволило использовать эту статистику в практических расчетах. Поэтому статистику ω^2 обычно называют также статистикой Крамера – Мизеса – Смирнова. Имеющаяся в указанной статье [5, с. 60 – 78] погрешность в формулировке леммы 6 (с. 75, формула (97)) (пропущен множитель $(-1)^k$ из-за неадекватного применения теории функций комплексного переменного) исправлена нами в статье [14].

Двухвыборочная двухсторонняя статистика Смирнова и аналогичная статистике Крамера – Мизеса – Смирнова двухвыборочная статистика типа омега-квадрат (статистика Лемана – Розенблатта) позволяют построить состоятельные критерии проверки абсолютной однородности независимых выборок, позволяющие обнаруживать любые альтернативы гипотезе тождественного совпадения функций распределения двух выборок. Алгоритмы расчета этих статистик приведены в [15]. Наиболее продвинутые таблицы распределения двухвыборочной двухсторонней статистики Смирнова подготовлены нашим коллективом [16].

Как следует из сказанного выше, А. Н. Колмогоров и Б. В. Гнеденко внесли огромный вклад в развитие статистических методов. Однако они занимались и многими другими проблемами (особенно А. Н. Колмогоров). Полностью посвятили себя статистическим методам в XX в. только два исследователя с академическими званиями – члены-корреспонденты АН СССР Н. В. Смирнов и Л. Н. Большев.

2.4.2. О Л.Н. Большеве

Логин Николаевич Большев (1922–1978) до конца Великой Отечественной войны участвовал в боевых действиях как летчик-истребитель. В 1951 г. окончил механико-математический факультет МГУ им. М.В. Ломоносова, будучи учеником А. Н. Колмогорова. Затем стал сотрудником Математического института АН СССР, в котором работал бок о бок с Н. В. Смирновым, которого и сменил в 1966 г. на посту руководителя отдела математической статистики. Для работ Л. Н. Большева [17] характерно сочетание высокого математического уровня с направленностью на практические приложения статистических методов. Его безвременная кончина обозначила рубеж, после которого разрыв между математической статистикой и статистическими методами (включая прикладную статистику) стал в сложившихся отечественных условиях неизбежным.

Постоянно используются в практической работе «Таблицы математической статистики» Л. Н. Большева и Н. В. Смирнова [6]. Будут применяться и в дальнейшем, поскольку распространенные статистические программные продукты имеют существенно более низкий научный уровень по сравнению с этой книгой [18].

В статье [19], провозгласившей выделение прикладной статистики как самостоятельного научного направления, отмечено, что Л. Н. Большев в беседе с одним из авторов настоящей книги активно возражал против термина «прикладная статистика», поскольку «Вся статистика является прикладной». При этом он отметил, что этот термин

– реакция на развитие т.н. «аналитической статистики», которая занимается внутриматематическими вопросами.

2.4.3. В. В. Налимов как организатор науки

Профессор МГУ им Ломоносова, доктор технических наук Василий Васильевич Налимов (1910 – 1997) — создатель и руководитель нескольких новых научных направлений: метрологии количественного анализа, химической кибернетики, математической теории эксперимента и наукометрии. Занимался проблемами математизации биологии, анализом оснований экологического прогноза, вероятностными аспектами эволюции, проблемами языка и мышления, философией и методологией науки, проблемами человека в современной науке, вероятностной теорией смыслов. Свой жизненный путь описал в книге [20].

Известность пришла к В. В. Налимову после выхода книги «Применение математической статистики при анализе вещества» [21] – справочника по применению классических статистических методов в работе химиков-аналитиков. Поскольку В. В. Налимов пришел в статистические методы не из математики, а из практической деятельности в заводских лабораториях, то и книга его была ориентирована на потребности практики.

Следующим шагом было создание секции «Математические методы исследования» редколлегии журнала «Заводская лаборатория». Сейчас под названием журнала стоит: «Ежемесячный научно-технический журнал по аналитической химии, физическим, математическим и механическим методам исследования, а также сертификации материалов». У истоков секции стояли Б.В. Гнеденко и В. В. Налимов, однако реально работу секции организовывал В. В. Налимов. Под его руководством она стала и остается поныне штабом развертывания исследований по статистическим методам в нашей стране.

В соответствии с тематикой журнала публикации секции посвящены в основном статистическим методам анализа данных измерений, наблюдений, испытаний, анализов, опытов. Большое значение придается математическим методам планирования экспериментов. В частности, при оптимизации технологических процессов в металлургической, химико-технологической, фармацевтической и иных отраслях промышленности применение методов экстремального планирования экспериментов позволяет заметно повысить выход продукта, обычно на 30 – 300%.

Основные направления работы секции – прикладная статистика и планирование эксперимента. В первом из них принимается, что экспериментатор не может выбирать точки (значения факторов), в которых проводятся измерения, во втором, напротив, выбор возможен, и основная задача - оптимальный подбор таких точек. Большое внимание уделяется вопросам оптимального управления технологическими процессами, в частности, статистическим методам управления качеством продукции. Рассматриваются также теория и практика экспертных оценок, применение нечетких множеств и др.

Заслугой В. В. Налимова является то, что в 60-е – 70-е годы XX в. в нашей стране была создана мощная научно-практическая школа в области планирования эксперимента. Перу В. В. Налимова принадлежит длинный ряд статей и книг, посвященный развитию теории и практики планирования эксперимента [22–24]. Итоги развития этой области статистических методов подведены учениками В. В. Налимова в статье [25], математическим основам планирования эксперимента посвящен справочник [26].

В 1961 г. была создана секция «Химическая кибернетика» (под председательством В. В. Налимова) в Научном совете по комплексной проблеме «Кибернетика» при Президиуме АН СССР. С 1971 г. В. В. Налимов возглавлял секцию «Математическая

теория эксперимента». Она объединяла более 500 активно действующих специалистов, работавших в академических и отраслевых институтах, вузах и на промышленных предприятиях. Развитие новой отрасли науки отслеживалось методами наукометрии [27], во многом созданной трудами В. В. Налимова.

В 1965 г. А. Н. Колмогоров организовал в МГУ им. М.В. Ломоносова межфакультетскую Лабораторию статистических методов и пригласил В. В. Налимова стать его первым заместителем. Задачи, поставленные перед Лабораторией, формулировались примерно так: изучение и дальнейшая разработка вероятностно-статистических методов; их пропаганда и широкое внедрение в научную, инженерную и медицинскую практику; хозяйственная деятельность; педагогическая и издательская деятельность; проведение общемосковских семинаров, летних научных школ, участие в конференциях [20, с.272]. Штатный состав достигал 130 человек. Такого мощного научного института – лидера не было в нашей стране. Нет и сейчас.

Организационным структурам, занимавшимся развитием статистических методов в нашей стране, не удалось укрепиться.

Большим успехом было введение в начале 70-х годов преподавания в вузах химической кибернетики и создание соответствующих кафедр. Однако через год последовало решение о сокращении штатов, и эти вновь введенные кафедры перестали существовать.

Ректор МГУ им. М.В. Ломоносова академик И.Г. Петровский поддерживал создание и развитие межфакультетской Лаборатории статистических методов А. Н. Колмогорова. Однако после его смерти выяснилось, что эта Лаборатория существует, так сказать, «нелегально», т.к. не входит в структуру университета. И в 1975 г. Лаборатория была расформирована. Ее сотрудники были распределены между пятью факультетами университета. Оказался уничтоженным единственный в нашей стране центр, занимавшийся методологическими аспектами вероятностно-статистического моделирования [20, с.291]. И это резко отрицательно сказалось на уровне отечественных прикладных работ.

В июле 1959 г. при Президиуме АН СССР был создан Совет по кибернетике, который возглавил академик А. И. Берг. Инженер-адмирал (высшее флотское звание) Аксель Иванович Берг (1893–1979) работал в области создания, развития и применения радиолокации и современных систем радионавигации, над проблемами кибернетики, став крупнейшим специалистом в основных областях этой отрасли науки. Как уже отмечалось, около 20 лет А. И. Берг поддерживал развитие статистических методов. А после его смерти новое руководство Совета «перекрыло кислород» этой тематике.

После смерти в 1978 г. члена-корреспондента АН СССР Л. Н. Большева резко сократилось сотрудничество между математиками и статистиками, разошлись пути математической и прикладной статистики.

Все эти события второй половины 70-х годов способствовали тому, что интересы В. В. Налимова сместились из научно-организационной деятельности в сферу его личных научных интересов. В книге «Вероятностная модель языка» [28] В. В. Налимов развивает мысль о нечеткости слов в естественном языке (ср. с констатацией «Мы мыслим нечетко» в статье [29]). Затем в длинной серии публикаций В. В. Налимов разрабатывает *вероятностно ориентированную философию*, включая вероятностное исчисление смыслов [30]. Последняя научная книга В. В. Налимова «В поисках иных смыслов» [31] начинается так: «Основная задача автора состоит в том, чтобы показать, что в наше время – в век утраты фундаментальных смыслов и всеобщей разбросанности знаний по отдельным законам многоликой культуры – все же возможно построение единых, по-прежнему целостно звучащих метафизических систем».

В настоящем разделе, а также в статьях [1–3] обсуждаются основные достижения пяти выдающихся исследователей советского периода – А. Н. Колмогорова, Б. В. Гнеденко, Н. В. Смирнова, Л. Н. Большева, В. В. Налимова. Вместе с ними работали тысячи специалистов. Нельзя не назвать А. Я. Хинчина, С. Н. Бернштейна, Е. Е. Слуцкого, В. С. Немчинова, В. И. Романовского, Г. К. Круга, А. А. Любишева. И вспомнить многих, многих других. История русской и советской статистики требует дальнейшего изучения, прежде всего потому, что старые дискуссии продолжаются и сейчас. Так, в недавних монографиях [32, 33] обсуждаются многие из тех проблем, которые волновали В. В. Налимова [20].

2.4.4. Дискуссия о прикладной статистике

Глазами американцев: российская дискуссия о прикладной статистике. Развитие прикладной статистики в нашей стране сопровождалось бурными дискуссиями. Объективный анализ их начального этапа был дан на страницах органа Американской статистической ассоциации. Статья Сэмюэля Котца и Кэтлин Смит «Пространство Хаусдорфа и прикладная статистика: точка зрения ученых СССР» [34] описывает различные взгляды, имеющие распространение и в XXI веке. Чтобы «взглянуть со стороны» на споры 80-х, используем эту статью.

Статья [34] посвящена дискуссии, развернувшейся на страницах советского статистического журнала «Вестник статистики» по вопросам существования и релевантности (уместности) прикладной статистики как самостоятельной научной дисциплины. В ней анализируется содержание четырех писем редактору и редакционных комментариев к ним, которые были опубликованы в этом журнале в период с октября 1985 г. по июнь 1987 г. Основная задача статьи состоит в том, чтобы осветить длительную (продолжающуюся по крайней мере 40 лет) полемику в советской (и российской) статистике между «идеологическими пуристами» и «прагматиками», которая в 80-е годы значительно усилилась. Существование разногласий, безусловно, не является новым явлением среди статистиков и в определенной степени оно носит здоровый характер, способствуя выработке критического отношения к предмету. Poleмика в 80-х затрагивает суть предмета в отличие от более ранних этапов, когда она отличалась идеологической направленностью. В 50–60-е, в период хрущевской оттепели, когда в СССР более свободно начали публиковать статистические данные, в журнале «*The American Statistician*» («Американский статистик») – органе Американской статистической ассоциации – было опубликовано несколько статей, посвященных различным аспектам советской статистики, как организационным, так и затрагивающим существо предмета.

Советская статистика: 1917 – 1964. Вопросы развития статистики в СССР с 1917 по 1964 г. довольно подробно освещены в статьях С. Котца [35, 36], прежде всего борьба двух противоположных мнений по вопросу о роли и содержании статистической науки в СССР. Между официальными статистиками Центрального статистического управления (ЦСУ, затем Госкомстат, ныне Росстат) и статистиками – экономистами математической направленности во главе с В. С. Немчиновым (1890–1964) возникли разногласия.

Официальные статистики считали, что статистика представляет собой описательную науку, в задачи которой входит сбор данных по плановой экономике, и что в условиях коммунизма статистику в конечном счете заменит простая бухгалтерия. Противоположных взглядов придерживались практики и статистики теоретической направленности. Они считали, что статистика и теория вероятностей важны в любой области. В 1954 г. на Всесоюзном научном совещании по теоретическим вопросам статистики (см. о нем в [37, с.243-247]), в работе которой приняли участие ведущие ученые, известный

советский математик А. Н. Колмогоров (1903–1987) помог представителям этих двух противостоящих школ прийти к прагматическому компромиссу. На совещании 1954 г. было заявлено, что статистика является самостоятельной общественной наукой и что «она изучает количественный аспект массовых социальных явлений в неразрывном единстве с их качественным аспектом» (см. Котц, [36, с.136]). Был сделан вывод, что советскую статистику от «буржуазной» статистики отличает акцент на качественном аспекте явлений. Для «буржуазной» статистики, согласно официальной оценке в Советском Союзе, характерен формальный, чисто математический подход к изучению социальных явлений, при котором количественный показатель рассматривается отдельно от качественной основы.

Однако на математическую статистику как часть математики «официальные статистики» покушаться не решились, поскольку математическая статистика использовалась для решения задач обороноспособности страны. Вместе с тем статистические методы в промышленности и технических исследованиях, статистические методы в медицине, химии, геологии, экономике, социологии, психологии, истории и в других конкретных областях остались вне официальной структуры науки и образования. В результате решений совещания 1954 г. работы по этим направлениям шли под иными именами. Использовались термины типа «экономическая кибернетика», «математическое моделирование в медицине» и др. Недаром сборник «Современные проблемы кибернетики (прикладная статистика)» [19], с момента выхода которого (1981) говорим о самостоятельности прикладной статистики, имеет название, свидетельствующее о «крыше» нашей науки (в данном случае «крыша» - это кибернетика, а в организационном плане – Научный совет АН СССР по комплексной проблеме «Кибернетика»). Соответственно в вузах не велась подготовка специалистов по статистическим методам в конкретных областях. В результате констатируем отставание на порядок от англоговорящих стран по числу специалистов. В США статистиков больше, чем математиков, а у нас создание Лаборатории статистических методов под руководством А. Н. Колмогорова со штатом в 130 человек рассматривалось как большой успех (в то время в Индии, в институте, которым руководил Махаланобис, работало около 2000 человек [20, с.271]).

Разногласия в 80-е годы. Появление статьи [36] совпало с началом публикации (с 1965 г.) полупериодического журнала «Ученые записки по статистике» под редакцией Немчинова (т.е. серии сборников статей, выпускавшихся издательствами «Наука», «Статистика», «Финансы и статистика»).

В 1986 году вышел юбилейный 50-й выпуск «Ученых записок по статистике». В нем опубликовали свои статьи статистики математической ориентации. Многие из них - выпускники вузов и кандидаты наук престижной школы теории вероятностей и математической статистики при МГУ им. М. В. Ломоносова, которую первоначально возглавлял А. Н. Колмогоров, и такой же школы при Ленинградском университете, во главе которой некоторое время стоял академик АН СССР Ю. В. Линник. Эти ученые работали в больших городах, в различных институтах, занимающихся, в том числе, вопросами применения прикладной статистики. Ученые выполняли ориентированные на практическое применение работы в прикладном многомерном статистическом анализе, теории управления запасами и т.д., однако создается впечатление, что они испытывали необходимость и желание заниматься вопросами, носящими более математический характер. Эта тенденция нашла свое отражение на страницах сборника «Ученые записки по статистике», в котором постепенно, но постоянно начали публиковать статьи математического и абстрактного характера, что вызвало недовольство среди статистиков различных научно-исследовательских институтов, связанных с органами официальной государственной статистики (в то время - ЦСУ).

В 1983 году в издательстве «Наука» вышел в свет 45-й том «Ученых записок по статистике», который был скромно озаглавлен «Прикладная статистика» [38], и разразился скандал. Опишем ход развития полемики, проанализировав содержание четырех писем редактору, которые были опубликованы с октября 1985 г. по июль 1987 г. в ежемесячном журнале «Вестник статистики» - органе ЦСУ.

В ответ на публикацию в сборнике «Ученые записки по статистике» многочисленных математических статей абстрактного характера К. Тимофеев (псевдоним) написал сердитое письмо под заголовком «Что же такое прикладная статистика?» [39]. Он утверждал, что термин «прикладная статистика» является абсурдным, так как то, что она якобы описывает, является одной из областей статистической науки, а не новым направлением. Тимофеев заявил: «Из содержания представленных в 45-м томе статей становится совершенно очевидным: название «Прикладная статистика» использовано для того, чтобы в «Ученых записках по статистике» опубликовать материалы, которые к ней (т.е. к статистике) не имеют ни прямого, ни даже косвенного отношения» [39, с.66]. Кроме этого, он выразил несогласие с рядом приведенных в сборнике математических формул и абстрактных концепций. В частности, он привел цитату из статьи, в которой говорится, что статья посвящена «измеримым отображениям произвольного вероятностного пространства в множество непустых компактов плоскости, снабженное метрикой Хаусдорфа» (метрика Хаусдорфа – одно из расстояний между множествами; критикуемая статья была озаглавлена «Статистика случайных множеств»). Тимофеев не только не захотел перенестись «в другое измерение», он подверг автора критике за то, что он в своей статье сослался на работы зарубежных ученых, а не на работы классиков марксизма-ленинизма и советские статистические источники, а также за то, что он написал работу, не связанную с реальной жизнью. Он с неодобрением указал, что авторы статей, публикуемых в «Ученых записках по статистике», часто ссылаются на свои собственные работы. Он написал: «Создается впечатление, что книга «Прикладная статистика» использована не только для публикации не относящихся к статистике материалов, но и для рекламы и саморекламы некоторых математиков, решивших снять себе славу в области экономики и статистики» [39, с.67]. Тимофеев признал, что эти статьи могут представлять определенный интерес для математиков, однако он полагал, что они вряд ли будут полезны в практической работе тем специалистам, на службе у которых, по его мнению, должна быть статистическая наука, а именно статистикам, экономистам и социологам.

Через десять месяцев журнал «Вестник статистики» опубликовал ответ [40] на выступление Тимофеева. Один из авторов, которых критиковал Тимофеев, А. Орлов, написал ответ в таком же резком тоне, и он был опубликован в официальном органе ЦСУ. В своей статье, перед которой было напечатано вступление от редакции, Орлов пункт за пунктом критиковал позицию Тимофеева. Орлов представил себя, как современного статистика. Он написал, что Тимофеев запутался и не знаком с переменами, которые произошли в статистике. Он отметил, что термин «прикладная статистика» не является ни новым, ни редко употребляемым. Он используется специалистами различных учреждений по всей стране. Он провел грань между математической статистикой и прикладной статистикой, добавив, что прикладная математическая статистика является «неотъемлемой частью» прикладной статистики, а прикладная математическая статистика наряду с аналитической статистикой (т.е. частью математической статистики, ориентированной на решение внутриматематических проблем, например, характеристических задач [41]) составляют математическую статистику, которая является одной из областей математики. Однако Орлов подчеркнул, что прикладная статистика включает и нематематические области, такие, как «методология организации и проведения прикладного статистического исследования и применения его результатов (как плани-

ровать исследование, как выбирать вероятностно-статистическую модель, как собирать данные, как подготавливать их к обработке, как представлять результаты обработки и т.д.), а также соответствующее программное обеспечение» [40, с.54].

Далее он указал, что интенсивное использование компьютеров в прикладной статистике свидетельствует о том, что в действительности ее можно рассматривать как часть кибернетики.

Орлов привел много примеров использования прикладной статистики в народном хозяйстве, сделав акцент на планировании эксперимента и контроле качества. Он отметил, что благодаря прикладной статистике была получена большая экономия финансовых средств: «Высокая эффективность прикладной статистики естественна - она родилась из практических нужд» [40, с.54]. Он охарактеризовал большой вклад в практическую работу, который внесли многие из тех статей, которые Тимофеев высмеял за абстрактные заголовки. В заключение статьи он привел таблицу, из которой видно, что ученые, публикующие свои работы в «Ученых записках по статистике», чаще ссылаются на работы советских авторов, чем зарубежных, и он подчеркнул, что эти авторы опираются на опыт своей практической работы, а не повторяют ранее опубликованный материал. Он составил эту таблицу на основе советского реферативного журнала «Математика», в котором «советские публикации составляют 1/6 мировых публикаций по прикладной статистике, реферируемых за год» [40, с.56].

Однако, по-видимому, редакторов журнала «Вестник статистики» не убедили доводы Орлова. В дополнение к его письму они напечатали свое заявление о том, что письмо Тимофеева было опубликовано для того, чтобы показать, что сборник «Ученые записки по статистике» перестал отвечать своей цели и превратился в математический журнал и что содержание статей в «Прикладной статистике» (том 45 «Ученых записок по статистике») не отвечает названию сборника. Более того, редакторы добавили, что находят доводы Тимофеева убедительными. Выступив с критикой письма Орлова, они упрекнули его за то, что он пытается «опровергнуть содержание письма К. Тимофеева, а заодно изобразить его автора как человека, не сведущего в делах, которыми занимается А. Орлов, а с ним и ряд других математиков» [40, с.57]. Они продолжали утверждать, что многие леммы и теоремы, которыми оперирует Орлов и его коллеги, не используются в практической работе. В частности, они проявили упорное желание узнать, «каков экономический эффект (в миллионах рублей), который удалось извлечь из шума при помощи измеримых отображений «произвольного вероятностного пространства в множество непустых компактов плоскости, снабженное метрикой Хаусдорфа» [40, с.57]. Касаясь ссылок на работы зарубежных авторов, редакторы отметили, что из таблицы Орлова видно, что ученые действительно ссылаются на зарубежные источники, и таким образом они приходят к выводу, что их утверждение верно. Обширные политизированные тексты «редакторов», весьма враждебные, но не подписанные, демонстрируют распространенные в то время – да и сейчас – приемы борьбы, используемые врагами современной науки.

Подтверждением того, что спорные вопросы еще не решены, по крайней мере в умах читателей, явилась публикация третьего письма, написанного Н. Шереметом [42]. Шеремет, доцент Московского института инженеров железнодорожного транспорта (МИИТ), придерживается умеренных взглядов по вопросу об определении прикладной статистики и ее роли. В начале своего письма он отметил, что Тимофеев не ответил на свой собственный вопрос: «Что же такое прикладная статистика?» По мнению Шеремета, прикладные науки являются связующим звеном между чисто «инженерными» работами и научными исследованиями или чистой наукой. Он выступил в защиту необходимости стадии «корректировки» или «подстройки» между стадиями научных изысканий и применением научных теорий на практике. Затем он привел хорошо известное

мнение Большева о том, что вся статистика является прикладной (Л. Н. Большев высказал это мнение в личной беседе с А. И. Орловым, цитата была включена в статью [40]), но не поддержал это утверждение, так как оно является слишком широким обобщением. Затем Шеремет проанализировал точку зрения, что каждая наука имеет свою собственную статистику (например, физическая статистика и биологическая статистика), но отверг ее, так как она противоречит мнению Ф. Энгельса, высказанному при подобных обстоятельствах в связи с механикой, физикой и химией. Шеремет критиковал Орлова за примеры из области экономики, так как эти примеры могли привести к ошибочному – по мнению Шеремета - предположению, что прикладная статистика является универсальной наукой.

Шеремет настаивал на определении статистики как общественной науки, однако признает возможность использования прикладной статистики в своей собственной области. Шеремет написал в свойственных ему неопределенных выражениях:

«Можно предположить, что предметом данной научной дисциплины являются «статистические данные»... Здесь уже не важно, от какого реального явления отвлечены данные абстрактные понятия... Математическая идеализация «статистических данных» и операций над ними дает возможность сводить известное разнообразие связей и закономерностей конкретной практической области к их определенному классу, производить необходимые расчеты» [42, с.69].

Он заявил, что прикладная статистика пока еще не является – по его мнению - четко определенной областью, и в заключение написал, что «прикладной статистике» в большей степени присущи черты междисциплинарных исследований, чем исследований, проводимых в рамках самостоятельной дисциплины [42, с.71].

В письме Шеремет допустил несколько неточностей, граничащих с дезинформацией. Он, кажется, не знает, что с 1973 г. журнал «Анналы статистики» («The Annals of Statistics») – один из основных западных статистических журналов) является непосредственным продолжением журнала «Анналы математической статистики» («Annals of Mathematical Statistics») и не делает разницы между узким техническим термином «статистика» (как функция от результатов наблюдений) и термином «статистика» (как наука и методология). Ссылка на элементарный учебник Вайнберга и Шумахера 1969 г. [43] как на образцовую современную монографию по прикладной статистике в лучшем случае вызывает сомнение.

Показательным является сам факт публикации подобного письма без редакционного комментария в советском консервативном журнале по статистическим наукам - в журнале, который со времени своего возрождения в 1949 году стал выразителем позиций официальных статистиков (многие из них строго придерживаются марксистско-ленинской ориентации), рассматривающих статистику только как описательную науку.

На страницах «Вестника статистики» письмо Шеремета было не единственным откликом на полемику между Тимофеевым и Орловым. По всей видимости независимо от письма Шеремета, в июле 1987 года «Вестник статистики» опубликовал письмо И. Манделя [44], доцента института Народного хозяйства в Алма-Ате (Казахстан). В качестве комментария на письма Тимофеева и Орлова Мандель составил развернутую схему, отражающую взаимосвязь теории статистики, прикладной статистики и математической статистики. Эта схема была представлена наряду с шестью другими методологическими приемами, чтобы показать, какое влияние оказывают теория статистики, прикладная статистика и математическая статистика на методы исследования массовых явлений. Главным в его доводах является положение о том, что в то время, как «теория статистики» в основном отражает «социальную сферу» массовых процессов, прикладная статистика должна быть направлена на отражение массовых явлений любого характера. Таким образом, прикладная статистика должна являться своего рода «буферной

наукой», которая переводит результаты математической статистики на язык, понятный исследователям в различных областях науки и практики. Он высказал сожаление по поводу существующих расхождений во взглядах между чистыми математическими статистиками и чистыми «прикладными» и обратил внимание на многочисленные примеры неправильного использования статистической методологии. Он приветствовал усилия математиков (в СССР и за рубежом), направленные на ликвидацию разрыва между математикой и реальным миром. В заключение он посоветовал называть прикладную статистику в значении «буферной науки» «прикладной математической статистикой». На конкретный вопрос о том, является ли сборник «Ученые записки по статистике» подходящим изданием для публикации статей по прикладной (математической?) статистике, он дает категорический отрицательный ответ, полностью совпадающий с мнением Тимофеева по этому вопросу. Мандель составил таблицу, согласно которой в 4 выпусках «Записок» (1978 - 1985), подготовленных прикладными статистиками, опубликовано 85 статей (общий объем -1092 стр.). Из них 62 статьи (787 стр.), т.е. почти три четверти, по его мнению, по своему содержанию больше подходили для публикации в известном советском журнале «Теория вероятностей и ее применения», так как были посвящены чисто математическим результатам и написаны в виде теорем и доказательств. Мандель, увы, не знал, что к тому времени редколлегия и авторы журнала «Теория вероятностей и ее применения» уже полностью оторвались от практики анализа статистических данных. По мнению Манделя, отличительной чертой прикладной статистики является отсутствие доказательств; для нее характерны только ссылки на теоремы и обсуждение вопросов «истинно» прикладного характера.

Обсуждение было продолжено в феврале 1988 г., когда в очередном выпуске «Вестника статистики» было опубликовано письмо болгарского профессора, специалиста по статистике, В. Цонева [45]. Он предлагает коренным образом изменить терминологию, связанную со всей статистической наукой.

Перестройка в области статистики. «Перестройка» в политике отразилась и в области статистики. Это проявилось не только в публикации новых статистических данных по промышленному травматизму, алкоголизму, преступности и т.д., но также и в координации работы многочисленных учреждений, занимающихся обработкой статистических данных. Реорганизация ЦСУ явилась еще одним свидетельством озабоченности правительства недостатками в данной области. К примеру, статистические данные, связанные с производством черных металлов, собирались и обрабатывались тремя учреждениями - Госпланом, ЦСУ и Институтом экономики министерства черной металлургии. На Всесоюзной конференции статистиков в мае 1985 г. выяснилось, что данные по прокату черных металлов, поступающие из этих трех источников, «совершенно разные» [46]. В феврале 1987 г. литературно-художественный журнал «Новый мир» выступил с открытой и резкой критикой отсутствия достоверных статистических данных. Несколько статистиков, среди них – Н. Шеремет и Т. Козлов, заведующий кафедрой статистики МИИТ - выступили с резким опровержением. Дискуссии продолжаются, поскольку ситуация в XXI в. аналогична – статистические данные разных ведомств и организаций не совпадают.

Как отмечено в [34], разногласия между учеными, о которых говорилось выше, характерны не только для Советского Союза. Американские и другие западные статистики также сталкиваются с проблемой определения роли прикладной статистики и, в более широком плане, с проблемой определения статистики как науки.

По нашему мнению, статистику следует рассматривать как одну из наук верхнего уровня иерархии, т.е. в одном ряду с математикой, физикой, экономикой. Так и сделано в РИНЦ - в авторском указателе тематика "Статистика" стоит рядом с тематиками "Математика", "Физика", "Экономика. Экономические науки".

2.4.5. Попытки объединения отечественных статистиков

В марте 1989 г. в Центральном экономико-математическом институте АН СССР состоялся Всесоюзный круглый стол «Статистика и перестройка», на котором собрались представители различных направлений в статистике – впервые в отечественной истории! Выступления были опубликованы в виде 55-го тома «Ученых записок по статистике» [47].

Высшей точкой общественного движения, ставящего целью объединение отечественных статистиков, было создание в 1990 г. Всесоюзной статистической ассоциации (ВСА), объединившей статистиков всех направлений – специалистов по прикладной и математической статистике, по надежности (в основном представителей оборонно-промышленного комплекса), преподавателей экономико-статистических дисциплин, работников официальной государственной статистики (см., например, [48, 49]). Ведущую роль в создании ВСА сыграли работники Всесоюзного центра статистических методов и информатики, созданного автором настоящей статьи в 1989 г. Наша платформа была изложена в статье [50], опубликованной, несмотря на ее весьма резкую форму, в «Вестнике статистики». Устав ВСА, решения Учредительного съезда и Пленума управления ВСА предусматривали различные формы работы [51].

В ходе дискуссий при создании ВСА были выработаны основные положения новой парадигмы прикладной статистики [52]. В рамках этой парадигмы в течение дальнейшей четверти века выполнены многочисленные исследования, в частности, развивались статистика объектов нечисловой природы [53], статистика интервальных данных [54], недавно была сформирована системная нечеткая интервальная математика [33].

Однако в 1991 г. СССР прекратил свое существование. ВСА, как и другие союзные организации, перестала действовать. И наметившееся единство статистиков распалось. Госкомстат РФ полностью «закрылся» от статистической науки, перестал даже отвечать на обращения профессиональных статистических организаций. Одновременно произошел окончательный отрыв специалистов математической статистики от практики. В настоящее время журнал «Теория вероятностей и ее применения» не представляет никакого интереса для тех, кто обрабатывает конкретные данные. При этом публикации работ по математической статистике почти прекратились. Поэтому целесообразно вспомнить мнение Л.Н. Большева и принять, что прикладная статистика – это и есть математическая статистика [55]. Отметим выпуск энциклопедии «Вероятность и математическая статистика» [56], содержащей массу полезной информации для специалистов по статистическим методам.

Работы по прикладной статистике и другим статистическим методам продолжались в рамках Российской ассоциации статистических методов (созданной на базе одноименной секции ВСА) и Российской академии статистических методов, а также в рамках Белорусской статистической ассоциации. В Узбекистане проводились международные конференции "Статистика и ее применения" (в 2012, 2013, 2015 гг.). Основным местом публикации отечественных работ по статистическим методам является раздел «Математические методы исследования» журнала «Заводская лаборатория. Диагностика материалов», созданный в 1961 г. Б. В. Гнеденко и В. В. Налимовым. В ней за 55 лет помещено около 1000 статей по различным направлениям прикладной статистики, прежде всего по статистическому анализу числовых величин, статистике нечисловых данных, многомерному статистическому анализу, статистике случайных процессов и временных рядов, планированию эксперимента, опыту применения статистических методов при решении конкретных прикладных задач.

2.5. Состояние и перспективы развития прикладной и теоретической статистики

Математические результаты объективны. Теорема либо доказана, либо нет. А вот о значении тех или иных результатов споры возникают. Тем более трудно настаивать на полной объективности выводов, рассуждая о развитии всей статистической науки. Поэтому жанр этого раздела монографии – субъективный анализ состояния и перспектив развития нашей научной области.

Интерес к тематике настоящего раздела возник в 1970-х гг., проявившись в создании одним из авторов монографии подкомиссии "Статистика объектов нечисловой природы" Научного Совета АН СССР по комплексной проблеме "Кибернетика". Позже на Учредительном съезде Всесоюзной статистической ассоциации (ВСА) в октябре 1990 г. коллеги выбрали А. И. Орлова вице-президентом ВСА (руководителем секции статистических методов).

С 1970-х годов приходилось обдумывать ситуацию в статистике. Конкретные результаты собраны в монографиях [1–3] и др., которые можно рассматривать и как учебники (см. также книги и статьи на сайте «Высокие статистические технологии» <http://orlovs.pp.ru>). Доказательства многих конкретных результатов опубликованы прежде всего в научных журналах "Заводская лаборатория. Диагностика материалов", "Научном журнале КубГАУ", периодических сборниках научных трудов "Статистические методы оценивания и проверки гипотез", "Ученые записки по статистике". А общие соображения – методологическая основа ведущихся работ – рассмотрены ниже в этом разделе. Для обеспечения связного и последовательного изложения повторно сформулируем ряд положений, которые обсуждались ранее, в частности, в главе 1 настоящей монографии.

2.5.1. Послевоенное развитие отечественной статистики

К 1960-м гг. в нашей стране сформировалась научно-практическая дисциплина, которую называем *классической математической статистикой*. Статистики старшего поколения учились теории по замечательной книге шведского математика Г. Крамера [4], написанной в военные годы и впервые изданной у нас в 1948 г. Из прикладных руководств назовем учебник [5] и таблицы с комментариями [6].

Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Примером таких работ является монография [7]. В ней получены продвинутое математические результаты, но нет никаких рекомендаций, которые мог бы использовать статистик, анализирующий конкретные данные наблюдений, измерений, испытаний, анализов, опытов, обследований.

Как реакция на уход в чистую математику значительной части специалистов была выделена новая научная дисциплина - *прикладная статистика*. В учебнике [1] в качестве рубежа, когда это стало очевидным, указан 1981 г. – дата выхода массовым тиражом (33 940 экз.) сборника [8], в названии которого использован термин «прикладная статистика». С этого времени линии развития математической статистики и прикладной статистики разошлись. Первая из этих дисциплин полностью ушла в чистую математику, перестав интересоваться практическими делами. Вторая позиционировала себя в качестве науки об обработке данных – результатов наблюдений, измерений, испытаний, анализов, опытов.

Вполне естественно, что в прикладной статистике стали развиваться свои собственные новые математические методы и модели. Необходимость их разработки вытекает из потребностей конкретных прикладных исследований. Это математизированное

ядро прикладной статистики целесообразно назвать *теоретической статистикой*. Тогда под собственно прикладной статистикой следует понимать обширную промежуточную область между теоретической статистикой и применением статистических методов в конкретных областях. В нее входят, в частности, вопросы формирования вероятностно-статистических моделей и выбора конкретных методов анализа данных (т.е. *методология* прикладной статистики и других статистических методов), проблемы разработки и применения информационных статистических технологий, организации сбора и анализа данных, т.е. разработки статистических технологий.

Таким образом, общая схема современной статистической науки выглядит следующим образом (от абстрактного к конкретному):

1. Математическая статистика – часть математики, изучающая статистические структуры. Сама по себе не дает рецептов анализа статистических данных, однако разрабатывает методы, полезные для использования в теоретической статистике.

2. Теоретическая статистика – наука, посвященная моделям и методам анализа конкретных статистических данных.

3. Прикладная статистика (в узком смысле) посвящена статистическим технологиям сбора и обработки данных. Она включает в себя методологию статистических методов, вопросы организации выборочных исследований, разработки статистических технологий, создания и использования статистических программных продуктов.

4. Применения статистических методов в конкретных областях (в экономике и менеджменте – эконометрика, в биологии – биометрика, в химии – хемометрия, в технических исследованиях – технометрика, в геологии, демографии, социологии, медицине, истории, и т.д.).

Часто позиции 2 и 3 вместе называют прикладной статистикой. Иногда позицию 1 именуют теоретической статистикой. Эти терминологические расхождения связаны с тем, что описанное выше развитие рассматриваемой научно-прикладной области не сразу, не полностью и не всегда адекватно отражается в сознании специалистов. Так, до сих пор выпускают учебники, соответствующие уровню представлений середины XX века.

Примечание. Здесь уточнена схема внутреннего деления статистической теории, предложенная в [9]. Естественный смысл приобрели термины «теоретическая статистика» и «прикладная статистика» (в узком смысле). Однако необходимо иметь в виду, что в сравнительно недавнем учебнике [1] прикладная статистика понимается в широком смысле, т.е. как объединение позиций 2 и 3. К сожалению, в настоящее время невозможно отождествить теоретическую статистику с математической, поскольку последняя (как часть математики - научной специальности «теория вероятностей и математическая статистика») полностью оторвалась от задач практики.

Отметим, что математическая статистика, как и теоретическая с прикладной, не имеет ничего общего с ведомственной наукой органов официальной государственной статистики. Эти органы, именовавшиеся последовательно ЦСУ, Госкомстат, Росстат, применяли и применяют лишь проверенные временем приемы позапрошлого века. Возможно, следовало бы от этого ведомства отмежеваться полностью и сменить название дисциплины, например, на «Анализ данных». Такое кардинальное решение, к сожалению, сделано бы неясным положение сотен тысяч публикаций, в названиях которых есть слова "статистика" или "статистический". В настоящее время компромиссным самоназванием нашей научно-практической дисциплины является термин «статистические методы».

Во второй половине 80-х годов развернулось общественное движение, имеющее целью создание профессионального объединения статистиков (об этом движении подробнее рассказано в статье [10]). Аналогами являются британское Королевское стати-

стическое общество (основано в 1834 г.) и Американская статистическая ассоциация (создана в 1839 г.). К сожалению, деятельность учрежденной в 1990 г. Всесоюзной статистической ассоциации оказалась парализованной в результате развала СССР. Некоторую активность проявили созданные на базе ВСА Российская ассоциация статистических методов, Российская академия статистических методов, Белорусская статистическая ассоциация. Пришло время оживить их деятельность. В последние годы большие ежегодные конференции проводятся в Ташкенте (см., например, материалы III научно-практической конференции "Статистика и ее применения" [11]).

В ходе создания ВСА было проанализировано состояние и перспективы развития теоретической и прикладной статистики. Обсудим их.

2.5.2. Новые идеи последних десятилетий: точки роста

В работе [12] впервые выделены пять актуальных направлений, в которых развивается современная прикладная статистика, т.е. пять «точек роста» статистической науки: непараметрика, робастность, бутстреп, интервальная статистика, статистика объектов нечисловой природы (см. также [13, 14]). Кратко обсудим эти актуальные направления.

Непараметрика, или непараметрическая статистика, позволяет делать статистические выводы, оценивать характеристики и плотность распределения, проверять статистические гипотезы без слабо обоснованных предположений о том, что функция распределения элементов выборки входит в то или иное параметрическое семейство. Например, широко распространена вера в то, что статистические данные часто подчиняются нормальному распределению. Математики думают, что это - экспериментальный факт, установленный в прикладных исследованиях. Прикладники уверены, что математики доказали нормальность результатов наблюдений. Между тем анализ конкретных результатов наблюдений, в частности, погрешностей измерений, приводит всегда к одному и тому же выводу - в подавляющем большинстве случаев реальные распределения существенно отличаются от нормальных [2]. Некритическое использование гипотезы нормальности часто приводит к значительным ошибкам, например, при отбраковке резко выделяющихся результатов наблюдений (выбросов), при статистическом контроле качества и в других случаях [1]. Поэтому целесообразно использовать непараметрические методы, в которых на функции распределения результатов наблюдений наложены лишь весьма слабые требования. Обычно предполагается лишь их непрерывность. К настоящему времени с помощью непараметрических методов можно решать практически тот же круг задач, что ранее решался параметрическими методами [15, 16]. Однако эта информация еще не вошла в массовое сознание. До сих пор тупиковой тематике параметрической статистики посвящены обширные разделы учебников и программных продуктов.

Основная идея работ по *робастности*, или *устойчивости*, состоит в том, что выводы, полученные на основе математических методов исследования, должны мало меняться при небольших изменениях исходных данных и отклонениях от предпосылок модели [17, 18]. Здесь есть два круга задач [19, 20]. Один - это изучение устойчивости распространенных алгоритмов анализа данных. Второй - поиск робастных алгоритмов для решения тех или иных задач. Отметим, что сам по себе термин "робастность" не имеет точно определенного смысла. Всегда необходимо указывать конкретную вероятностно-статистическую модель. При этом модель «засорения» Тьюки - Хубера - Хампеля обычно не является практически полезной. Дело в том, что она ориентирована на «утяжеление хвостов», а в реальных ситуациях «хвосты» обрезаются априорными ог-

раничениями на результаты наблюдений, связанными, например, с ограниченностью шкал используемых средств измерения.

Бутстреп - направление непараметрической статистики, опирающееся на интенсивное использование информационных технологий [21]. Основная идея состоит в «размножении выборок», т.е. в получении набора из многих выборок, напоминающих полученную в эксперименте. По такому набору можно непосредственно оценить свойства различных статистических процедур, не прибегая к излишне обременительным (а чаще и необоснованным) семействам вероятностно-статистических моделей. Простейший способ «размножения выборки» состоит в исключении из нее одного результата наблюдения. Исключаем первое наблюдение, получаем выборку, похожую на исходную, но с объемом, уменьшенным на 1. Затем возвращаем исключенный результат первого наблюдения, но исключаем второе наблюдение. Получаем вторую выборку, похожую на исходную. Затем возвращаем результат второго наблюдения, и т.д. Есть и иные способы «размножения выборок». Например, можно по исходной выборке построить ту или иную оценку функции распределения, а затем методом статистических испытаний смоделировать ряд выборок из элементов, функция распределения которых совпадает с этой оценкой. Обобщая, можно сказать, что к настоящему времени в дополнение к классическим инструментам прикладной статистики – предельным теоремам теории вероятностей – добавились новые, основанные на интенсивном использовании компьютеров [22]. Бутстреп – лишь один из таких инструментов. Метод статистических испытаний (Монте-Карло) – вот партнер и конкурент асимптотическим методам математической статистики.

Интервальная статистика - это совокупность методов анализа интервальных статистических данных. Вполне очевидно, что все средства измерения имеют погрешности. Однако до недавнего времени это очевидное обстоятельство никак не учитывалось в статистических процедурах. Только недавно начала развиваться теория интервальной статистики, в которой предполагается, что исходные данные - это не числа, а интервалы. Интервальную статистику можно рассматривать как часть интервальной математики. Выводы в ней часто принципиально отличны от классических. Основные результаты статистики интервальных данных рассмотрены в статьях [23, 24], а подробные изложения включены в учебники [1, 3, 25], а также в монографию [26].

2.5.3. Статистика объектов нечисловой природы

Перейдем к статистике объектов нечисловой природы (она же - статистика нечисловых данных, или нечисловая статистика) [25, 27]. Сначала напомним, что типичный исходный объект в прикладной статистике - это выборка, т.е. совокупность независимых одинаково распределенных случайных элементов. Какова природа этих элементов? В классической математической статистике элементы выборки - это числа. В многомерном статистическом анализе - вектора. А в нечисловой статистике элементы выборки - это объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы лежат в пространствах, не имеющих векторной структуры.

Примерами объектов нечисловой природы являются:

- значения качественных признаков, в том числе результаты кодировки объектов с помощью заданного перечня категорий (градаций);
- упорядочения (ранжировки) экспертами образцов продукции (при оценке её технического уровня, качества и конкурентоспособности) или заявок на проведение научных работ (при проведении конкурсов на выделение грантов);

- классификации, т.е. разбиения объектов на группы сходных между собой (кластеры);
- толерантности, т.е. бинарные отношения, описывающие сходство объектов между собой, например, сходства тематики научных работ, оцениваемого экспертами с целью рационального формирования экспертных советов внутри определенной области науки;
- результаты парных сравнений или контроля качества продукции по альтернативному признаку ("годен" - "брак"), т.е. последовательности из 0 и 1;
- графы различных видов (неориентированные, ориентированные, с весами, иерархические структуры и т.п.);
- множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга;
- слова, предложения, тексты;
- вектора, координаты которых - совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности организации или анкета эксперта, в которой ответы на часть вопросов носят качественный характер, а на часть - количественный;
- ответы на вопросы экспертной, медицинской, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возможно, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты; и т.д.

Рассмотренные выше интервальные данные тоже можно рассматривать как пример объектов нечисловой природы, а именно, как частный случай нечетких множеств. Если характеристическая функция нечеткого множества равна 1 на некотором интервале и равна 0 вне этого интервала, то задание такого нечеткого множества эквивалентно заданию интервала. Напомним, что *теория нечетких множеств в определенном смысле сводится к теории случайных множеств*. Цикл соответствующих теорем приведен в работах [17, 28], а также в учебниках [1–3], монографии [26].

С 70-х годов в основном на основе запросов теории экспертных оценок [29–31] (а также технических исследований, экономики, социологии и медицины) развивались различные направления статистики объектов нечисловой природы. Были установлены основные связи между конкретными видами таких объектов, разработаны для них базовые вероятностные модели. Сводка дана в монографии [17].

Следующий этап (80-е годы) - выделение статистики объектов нечисловой природы в качестве самостоятельной дисциплины в рамках математических методов исследования, ядром которого являются методы статистического анализа данных произвольной природы. Для работ этого периода характерна сосредоточенность на внутренних проблемах нечисловой статистики.

К 90-м годам статистика объектов нечисловой природы с теоретической точки зрения была достаточно хорошо развита, основные идеи, подходы и методы были разработаны и изучены математически, в частности, доказано достаточно много теорем. Однако она оставалась недостаточно апробированной на практике. И в 90-е годы наступило время перейти от теоретико-статистических исследований к применению полученных результатов на практике и включить их в учебный процесс, что и было сделано (см., например, учебники [1–3]).

Следует отметить, что в статистике объектов нечисловой природы одна и та же математическая схема может с успехом применяться во многих областях, для анализа данных различных типов, а потому ее лучше всего формулировать и изучать в наиболее общем виде, для объектов произвольной природы.

2.5.4. Основные идеи статистики объектов нечисловой природы

В чем принципиальная новизна нечисловой статистики? Для классической математической статистики характерна операция сложения. При расчете выборочных характеристик распределения (выборочное среднее арифметическое, выборочная дисперсия и др.), в регрессионном анализе и других областях этой научной дисциплины постоянно используются суммы. Математический аппарат - законы больших чисел, Центральная предельная теорема и другие теоремы - нацелены на изучение сумм. В нечисловой же статистике нельзя использовать операцию сложения, поскольку элементы выборки лежат в пространствах, где нет операции сложения. Методы обработки нечисловых данных основаны на принципиально ином математическом аппарате - на применении различных расстояний в пространствах объектов нечисловой природы.

Кратко рассмотрим несколько идей, развиваемых в статистике объектов нечисловой природы для данных, лежащих в пространствах произвольного вида. Они нацелены на решение классических задач описания данных, оценивания, проверки гипотез - но для неклассических данных, а потому неклассическими методами.

Первой обсудим проблему определения средних величин. В рамках теории измерений удастся указать вид средних величин, соответствующих тем или иным шкалам измерения [17]. В классической математической статистике средние величины вводят с помощью операций сложения (выборочное среднее арифметическое, математическое ожидание) или упорядочения (выборочная и теоретическая медианы). В пространствах произвольной природы средние значения нельзя определить с помощью операций сложения или упорядочения. Теоретические и эмпирические средние приходится вводить как решения экстремальных задач. Теоретическое среднее определяется как решение задачи минимизации математического ожидания (в классическом смысле) расстояния от случайного элемента со значениями в рассматриваемом пространстве до фиксированной точки этого пространства (минимизируется указанная функция от этой точки). Для эмпирического среднего математическое ожидание берется по эмпирическому распределению, т.е. берется сумма расстояний от некоторой точки до элементов выборки и затем минимизируется по этой точке [32]. При этом как эмпирическое, так и теоретическое средние как решения экстремальных задач могут быть не единственными элементами рассматриваемого пространства, а являться некоторыми множествами таких элементов, которые могут оказаться и пустыми. Тем не менее удалось сформулировать и доказать законы больших чисел для средних величин, определенных указанным образом, т.е. установить сходимость (в специально определенном смысле) эмпирических средних к теоретическим [1, 2].

Оказалось, что методы доказательства законов больших чисел допускают существенно более широкую область применения, чем та, для которой они были разработаны. А именно, удалось изучить асимптотику решений экстремальных статистических задач, к которым, как известно, сводится большинство постановок прикладной статистики. В частности, кроме законов больших чисел установлена и состоятельность оценок минимального контраста, в том числе оценок максимального правдоподобия и робастных оценок [25]. К настоящему времени подобные оценки изучены также и в интервальной статистике.

В статистике в пространствах произвольной природы большую роль играют непараметрические оценки плотности, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализов. В нечисловой статистике предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в том числе в дискретных пространствах. В частности, доказа-

на их состоятельность, изучена скорость сходимости и установлен примечательный факт совпадения наилучшей скорости сходимости в произвольном пространстве с той, которая имеет быть в классической теории для числовых случайных величин [33, 34].

Дискриминантный, кластерный, регрессионный анализы в пространствах произвольной природы основаны либо на параметрической теории - и тогда применяется подход, связанный с асимптотикой решения экстремальных статистических задач - либо на непараметрической теории - и тогда используются алгоритмы на основе непараметрических оценок плотности.

Для проверки гипотез могут быть использованы статистики интегрального типа, в частности, типа омега-квадрат. Любопытно, что предельная теория таких статистик, построенная первоначально в классической постановке, приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида, поскольку при этом удалось провести рассуждения, опираясь на базовые математические соотношения, а не на те частные (с общей точки зрения), что были связаны с конечномерным пространством [35].

Представляют практический интерес результаты, связанные с конкретными областями статистики объектов нечисловой природы [36] и соответствующими математико-статистическими моделями порождения нечисловых данных [37], в частности, со статистикой нечетких множеств и со статистикой случайных множеств (напомним, что теория нечетких множеств в определенном смысле сводится к теории случайных множеств), с непараметрической теорией парных сравнений и люсианов (бернуллиевских бинарных векторов) [38], с аксиоматическим введением метрик в конкретных пространствах объектов нечисловой природы [39], а также с рядом других конкретных постановок.

Для анализа нечисловых, в частности, экспертных данных весьма важны методы классификации [40]. Констатируем, что наиболее естественно ставить и решать задачи классификации, основанные на использовании расстояний или показателей различия, в рамках статистики объектов нечисловой природы. Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа) [41, 42].

2.5.5. О нерешенных проблемах теоретической и прикладной статистики

За каждым новым научным результатом открывается многообразие неизвестного. Рассмотрим несколько конкретных постановок.

В статистике в пространствах общей природы получены аналоги классического закона больших чисел. Но нет аналога центральной предельной теоремы. Какова скорость сходимости эмпирических средних к теоретическим? Как сравнить различные способы усреднения? В частности, что лучше применять для усреднения упорядочений – медиану Кемени или среднее по Кемени (среднее отличается от медианы тем, что в качестве показателя различия берется не расстояние Кемени, а его квадрат)? Какие конкретные представители различных классов непараметрических оценок плотности достойны рекомендации для использования в нацеленных на практическое применение алгоритмах анализа нечисловых данных?

До сих пор не проведена классификация классических статистических методов с точки зрения теории измерений. Законченные результаты получены только для теории средних [1–3]. Установлено, что для измерений в порядковой шкале в качестве средних можно использовать только порядковые статистики, например, медиану (при нечетном объеме выборки). Среднее арифметическое, столь любимое профанами, применять нельзя. Однако многочисленные эксперименты показывают, что упорядочения объек-

тов по средним арифметическим рангов и по медианам рангов в подавляющем большинстве случаев совпадают. Нужна теория, объясняющая этот экспериментальный факт.

Все более широкое распространение получает теория нечеткости. Давно установлено, что она в определенном смысле сводится к теории случайных множеств [1–3]. Требуется на основе этого сведения проанализировать различные теоретические и прикладные постановки теории нечеткости и рассмотреть их в рамках вероятностно-статистического моделирования.

Перейдем к классическим областям статистики. Начнем с обсуждения влияния отклонений от традиционных предпосылок. В вероятностной теории статистических методов выборка обычно моделируется как конечная последовательность независимых одинаково распределенных случайных величин или векторов. В парадигме середины XX в. часто предполагают, что эти величины (вектора) имеют нормальное распределение.

При внимательном взгляде совершенно ясна нереалистичность приведенных классических предпосылок. Независимость результатов измерений обычно принимается "из общих предположений", между тем во многих случаях очевидна их коррелированность. Одинаковая распределенность также вызывает сомнения из-за изменения во времени свойств измеряемых образцов, средств измерения и психофизического состояния специалистов, проводящих измерения (испытания, анализы, опыты). Даже обоснованность самого применения вероятностных моделей иногда вызывает сомнения, например, при моделировании уникальных измерений (согласно классическим воззрениям, теорию вероятностей обычно привлекают при изучении массовых явлений). И уж совсем редко распределения результатов измерений можно считать нормальными [1, 2].

Итак, методы классической математической статистики обычно используют вне сферы их обоснованной применимости. Каково влияние отклонений от традиционных предпосылок на статистические выводы? В настоящее время об этом имеются лишь отрывочные сведения. Приведем три примера.

Пример 1. Построение доверительного интервала для математического ожидания обычно проводят с использованием распределения Стьюдента (при справедливости гипотезы нормальности). Как следует из Центральной предельной теоремы (ЦПТ) теории вероятностей, в асимптотике (при большом объеме выборки) такие расчетные методы дают правильные результаты (из ЦПТ вытекает использование квантилей нормального распределения, а из классической теории - квантилей распределения Стьюдента, но при росте объема выборки квантили распределения Стьюдента стремятся к соответствующим квантилям нормального распределения).

Пример 2. Для проверки однородности двух независимых выборок (на самом деле - для проверки равенства математических ожиданий) обычно рекомендуют использовать двухвыборочный критерий Стьюдента. Предпосылки его использования – это нормальность распределений, соответствующих выборкам, и равенство их дисперсий. Что будет при отклонении от нормальности распределений, из которых взяты выборки, от нормальности? Если объемы выборок равны или если дисперсии совпадают, то в асимптотике (когда объемы выборок безгранично возрастают) классический метод является корректным. Если же объемы выборок существенно отличаются или дисперсии различны, то критерий Стьюдента проверки гипотезы однородности применять нельзя, поскольку распределение двухвыборочной статистики Стьюдента будет существенно отличаться от классического. Поскольку проверка равенства дисперсий - более сложная задача, чем проверка равенства математических ожиданий, то для выборок разного объема использовать двухвыборочную статистику Стьюдента не следует, целесообразно применять критерий Крамера-Уэлча [1, 2, 43].

Пример 3. В задаче отбраковки (исключения) резко выделяющихся наблюдений (выбросов) расчетные методы, основанные на нормальности, являются крайне неустойчивыми по отношению к отклонениям от нормальности, что полностью лишает эти методы научной обоснованности [1, 2].

Примеры 1 - 3 показывают весь спектр возможных свойств классических расчетных методов в случае отклонения от нормальности. Методы примера 1 оказываются вполне пригодными при таких отклонениях, примера 2 - пригодными в некоторых случаях, примера 3 - полностью непригодными.

Итак, имеется *необходимость изучения свойств расчетных методов классической математической статистики, опирающихся на предположение нормальности, в ситуациях, когда это предположение не выполнено*. Аппаратом для такого изучения наряду с методом Монте-Карло могут послужить предельные теоремы теории вероятностей, прежде всего ЦПТ, поскольку интересующие нас расчетные методы обычно используют разнообразные суммы. Пока подобное изучение не проведено, остается неясной научная ценность, например, применения основанного на предположении многомерной нормальности факторного анализа к векторам из переменных, принимающих небольшое число градаций и к тому же измеренных в порядковой шкале.

Почему необходимо изучение классических алгоритмов, а не построение новых, специально предназначенных для работы в условиях отклонения от классических предположений?

Во-первых, потому, что классические алгоритмы в настоящее время наиболее распространены (благодаря сложившейся системе образования прикладников). Например, в научных медицинских исследованиях для проверки однородности двух независимых выборок традиционно используют критерий Стьюдента, при этом условия его применимости не проверяют. Насколько обоснованными являются выводы? Как следует из примера 2, во многих случаях выводы нет оснований подвергать сомнению, хотя они получены с помощью некорректной процедуры.

Во-вторых, более новые подходы зачастую методологически уязвимы. Так, известная робастная модель засорения Тьюки - Хубера нацелена на борьбу с большими выбросами, которые зачастую физически невозможны из-за ограниченности интервала значений измеряемой характеристики, в котором работает конкретное средство измерения. Следовательно, модель Тьюки - Хубера - Хампеля имеет скорее теоретическое значение, чем практическое. Сказанное, конечно, не обозначает, что следует прекратить разработку, изучение и внедрение непараметрических и устойчивых методов, выделенных выше как «точки роста» современной прикладной статистики.

Нерешенным проблемам теоретической и прикладной статистики посвящены статьи [44, 45]. Одна из важных проблем - использование асимптотических результатов при конечных объемах выборок. Конечно, естественно изучить свойства алгоритма с помощью метода Монте-Карло. Однако из какого конкретного распределения брать выборки при моделировании? От выбора распределения зависит результат. Кроме того, датчики псевдослучайных чисел лишь имитируют случайность. До сих пор неизвестно, каким датчиком целесообразно пользоваться в случае возможного безграничного роста размерности пространства.

Другая проблема – обоснование выбор одного из многих критериев для проверки конкретной гипотезы. Например, для проверки однородности двух независимых выборок можно предложить критерии Стьюдента, Крамера - Уэлча, Лорда, хи-квадрат, Вилкоксона (Манна-Уитни), Ван-дер-Вардена, Сэвиджа, Н. В. Смирнова, типа омега-квадрат (Лемана-Розенблатта), Реньи, Г. В. Мартынова и др. Какой выбрать?

Критерии однородности проанализированы в [46]. Естественных подходов к сравнению критериев несколько - на основе асимптотической относительной эффек-

тивности по Бахадуру, Ходжесу - Леману, Питмену. И каждый критерий является оптимальным при соответствующей альтернативе или подходящем распределении на множестве альтернатив. При этом математические выкладки обычно используют альтернативу сдвига, сравнительно редко встречающуюся в практике анализа реальных статистических данных. Итог печален - блестящая математическая техника, продемонстрированная в [46], не позволяет дать рекомендации для выбора критерия проверки однородности при анализе реальных данных.

Проблемы разработки высоких статистических технологий поставлены в [47, 48] (см. также одноименный сайт <http://orlovs.pp.ru>). Используемые при обработке реальных данных статистические технологии состоят из последовательности операций, каждая из которых, как правило, хорошо изучена, поскольку сводится к оцениванию (параметров, характеристик, распределений) или проверке той или иной гипотезы. Однако статистические свойства результатов обработки, полученных в результате последовательного применения таких операций, мало изучены. Необходима теория, позволяющая изучать свойства статистических технологий и так их конструировать, чтобы обеспечить высокое качество обработки данных.

ГЛАВА 3. НЕКОТОРЫЕ РЕЗУЛЬТАТЫ ПРИКЛАДНОЙ СТАТИСТИКИ

3.1. Структура прикладной статистики

Прикладная статистика - наука о том, как обрабатывать статистические данные. Как самостоятельная научно-практическая область она развивается весьма быстро. В ее состав входят многочисленные широко и глубоко развитые научные направления. Те, кто применяет прикладную статистику и другие статистические методы, обычно ориентированы на конкретные области исследования, т.е. не являются специалистами по прикладной статистике. Поэтому представляется полезным провести критический анализ современного состояния прикладной статистики и обсудить тенденции развития статистических методов.

3.1.1. Что дает прикладная статистика народному хозяйству?

Такой вопрос часто задают специалисты различных областей науки, отраслей народного хозяйства, не владеющие методами прикладной статистики. В ответ нами была написана статья [1], в которой приводились многочисленные примеры успешного использования методов прикладной математической статистики при решении практических задач. Перечень примеров можно продолжать практически безгранично. Например, можно сослаться на обобщающую монографию В. Г. Горского [2], на диссертацию А. Н. Гуды [3]. По данным Института информации Гарфилда (США) каждая из основополагающих книг В. В. Налимова [4, 5] цитировалась не менее 1000 раз (см. также монографию [6, с.270, 274, 373]). Практически в любом номере журнала "Заводская лаборатория. Диагностика материалов" есть работы, в которых те или иные методы прикладной статистики применяются для решения прикладных задач. Не раз публиковались в этом журнале и обобщающие статьи по вопросам применения прикладной статистики [7–10].

Итак, бесспорно совершенно, что методы прикладной статистики успешно применяются в различных отраслях народного хозяйства, практически во всех областях

науки. Согласно докладу [11, с.157-158], в 1988 г. затраты в СССР на статистический анализ данных оценивались в 2 миллиарда рублей ежегодно.

Большая практическая значимость прикладной статистики оправдывает целесообразность проведения работ по развитию ее методологии, в которых эта область научной и прикладной деятельности рассматривалась бы как целое, "с высоты птичьего полета". Чтобы иметь возможность обсуждения тенденций развития прикладной статистики и других статистических методов, кратко рассмотрим их историю.

3.1.2. Об истории прикладной статистики

Типовые примеры раннего этапа применения статистических методов описаны в Ветхом Завете (см., например, Книгу Чисел - четвертую книгу Моисееву). С математической точки зрения они сводились к подсчетам числа попаданий значений наблюдаемых признаков в определенные градации. В дальнейшем результаты стали представлять в виде таблиц и диаграмм, как это и сейчас делает Федеральная служба государственной статистики (Росстат). Надо признать, что по сравнению с Ветхим Заветом есть прогресс - в Библии не было таблиц. Однако в работах Росстата нет продвижения по сравнению с работами российских статистиков конца девятнадцатого - начала двадцатого века (типовой монографией тех времен можно считать книгу [12], которая в настоящее время ещё легко доступна).

Сразу после возникновения теории вероятностей (Паскаль, Ферма, XXVII век) вероятностные модели стали использоваться при обработке статистических данных. Например, изучалась частота рождения мальчиков и девочек, было установлено отличие вероятности рождения мальчика от 0.5, анализировались причины того, что в парижских приютах эта вероятность не та, что в самом Париже, и т.д. Имеется достаточно много публикаций по истории теории вероятностей, однако в некоторых из них имеются неточные утверждения, что заставило академика Украинской АН Б. В. Гнеденко включить в шестое издание своего знаменитого курса [13] главу по истории математики случайного, выпущенную затем отдельным изданием [14].

Как установил Ф. Клейн, анализируя записные книжки великого немецкого математика и физика К. Гаусса, тот в 1794 г. разработал метод наименьших квадратов, один из наиболее популярных ныне статистических методов, и применил его при расчете орбиты астероида Церера - для борьбы с ошибками астрономических наблюдений [15]. Математические публикации К. Гаусса по методу наименьших квадратов появились лишь через 12 лет, когда близкие вопросы стали обсуждаться в литературе. В XIX веке заметный вклад в развитие практической статистики внес в развитие практической статистики внес бельгиец Ламбер Адольф Жак Кетле, на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей, таких, как доля самоубийств среди всех смертей [16]. Важно отметить, что основные идеи статистического приемочного контроля и сертификации продукции глубоко анализировались российским академиком М. В. Остроградским и применялись в российской армии ещё в середине XIX в. [17, 18]. Статистические методы управления качеством, сертификации и классификации продукции сейчас весьма актуальны как с теоретической, так и с прикладной точек зрения [19, 20].

Современный этап развития прикладной статистики можно отсчитывать с 1900 г., когда англичанин К. Пирсон основал журнал "*Biometrika*". Мы полагаем, что именно 1900 г. следует считать началом развития прикладной математической статистики как самостоятельной науки, хотя вначале она выступала под другими именами (биометрика, актуарные исследования (т.е. исследования в области страхового дела) и др.). Первая треть XX в. прошла под знаком параметрической статистики. Изучались методы,

основанные на анализе данных из параметрических семейств распределений, описываемых кривыми семейства Пирсона. Наиболее популярным было нормальное (гауссово) распределение. Для проверки гипотез использовались критерии Пирсона, Стьюдента, Фишера. Были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

Разработанную в первой трети XX в. теорию будем называть параметрической статистикой, поскольку ее основной объект изучения - это выборки из распределений, описываемых одним или небольшим числом параметров. Наиболее общим является семейство кривых Пирсона, задаваемых четырьмя параметрами. Как правило, нельзя указать каких-либо веских причин, по которым конкретное распределение результатов наблюдений должно входить в то или иное параметрическое семейство. Исключения хорошо известны: если вероятностная модель предусматривает суммирование независимых случайных величин, то сумму естественно описывать нормальным распределением; если же в модели рассматривается произведение таких величин, то итог, видимо, приближается логарифмически нормальным распределением, и т.д. Однако в подавляющем большинстве реальных ситуаций подобных моделей нет, и приближение реального распределения с помощью кривых из семейства Пирсона или его подсемейств - чисто формальная операция. Именно из таких соображений критиковал параметрическую статистику академик С. Н. Бернштейн в 1927 г. в своем докладе на Всероссийском съезде математиков [21]. Однако эта теория, к сожалению, до сих пор остается основой преподавания статистических методов и продолжает использоваться основной массой прикладников, остающихся далекими от новых веяний в статистике. Почему так происходит? Чтобы попытаться ответить на этот вопрос, обратимся к наукометрии.

3.1.3. Наукометрия прикладной статистики

Проведенный в 1980-е годы (в ходе работ по созданию Всесоюзной статистической ассоциации [22]) анализ прикладной статистики как области научно-практической деятельности показал, в частности, что актуальными для специалистов в настоящее время являются не менее чем 100 тысяч публикаций (подробнее см. статьи [8, 23, 24]). Реально же каждый из нас знаком с существенно меньшим количеством книг и статей. Так, в наиболее обширном на русском языке сочинении по прикладной статистике - известном трехтомнике Кендалла и Стьюарта [25–27] - приведено около 2 тысяч литературных ссылок. При всей очевидности соображений о многократном дублировании в публикациях ценных идей приходится признать, что каждый специалист по прикладной статистике владеет лишь небольшой частью накопленных в этой области знаний (это утверждение в полной мере относится и к специалистам в других областях). Неудивительно, что приходится постоянно сталкиваться с игнорированием или повторением ранее полученных результатов, с уходом в тупиковые (с точки зрения практики) направления исследований, с беспомощностью при обращении к реальным данным, и т.д. Все это - одно из проявлений адапционного механизма торможения развития науки, о котором еще около 50 лет назад писали В. В. Налимов и другие науковеды (см., например, [28]).

Традиционный предрассудок состоит в том, что каждый новый результат, полученный исследователем - это кирпич в непрерывно растущее здание науки, который непременно будет проанализирован и использован научным сообществом. Реальная ситуация - совсем иная. Основа профессиональных знаний исследователя и инженера закладывается в период обучения. Затем они пополняются в том узком направлении, в котором работает специалист. Следующий этап - их тиражирование новому поколению. В результате вузовские учебники отстают от современного развития на десятки лет.

Так, учебники по математической статистике, по нашей экспертной оценке, в основном соответствуют 40–60-м годам XX в. А потому тем же годам соответствует большинство вновь публикуемых исследований и тем более - прикладных работ. Одновременно приходится признать, что результаты, которые не вошли в учебники, независимо от их ценности почти все забываются.

Активно продолжается развитие тупиковых направлений. Психологически это понятно. Приведем пример из опыта одного из авторов настоящей монографии. В свое время по заказу Госстандарта А. И. Орлов разработал методы оценки параметров гамма-распределения [29]. Поэтому ему близки и интересны работы по оцениванию параметров по выборкам из распределений, принадлежащих тем или иным параметрическим семействам, понятия функции максимального правдоподобия, эффективности оценок, использование неравенства Рао - Крамера и т.д. К сожалению, известно, что это - тупиковая ветвь, поскольку реальные данные не подчиняются каким-либо параметрическим семействам, надо применять иные статистические методы, о которых речь пойдет ниже. Понятно, что специалистам по параметрической статистике, потратившим многие годы на совершенствование в своей области, психологически трудно согласиться с этим утверждением. В том числе и одному из авторов настоящей монографии.

3.1.4. Точки роста

Отечественная литература по прикладной статистике столь же необозрима, как и мировая. Только в разделе "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" с 1960-х годов опубликовано более 1000 статей. Не будем даже пытаться перечислять коллективы исследователей или основные монографии в этой области (впрочем, см. статью [24]). Отметим только одно издание. По нашему мнению, наилучшей отечественной книгой по прикладной статистике является сборник статистических таблиц Л. Н. Большева и Н. В. Смирнова [30] с подробными комментариями, играющими роль учебника и справочника по классическим вопросам прикладной статистики.

С целью управления развитием статистической науки кратко рассмотрим "точки роста" прикладной статистики [31], т.е. те ее направления, которые представляются перспективными, но пока отодвинуты на задний план традиционными постановками. Двадцать пять лет назад при описании современного этапа развития статистических методов нами были выделены [32] пять актуальных направлений, в которых развивается современная прикладная статистика, т.е. пять "точек роста": непараметрика, робастность, бутстреп, статистика интервальных данных, нечисловая статистика. Обсудим их.

Непараметрическая статистика. В первой трети XX в., одновременно с параметрической статистикой, в работах Спирмена и Кендалла появились первые непараметрические методы, основанные на коэффициентах ранговой корреляции, носящих ныне имена этих статистиков. Но непараметрика, не делающая нереалистических предположений о том, что функции распределения результатов наблюдений принадлежат тем или иным параметрическим семействам распределений, стала заметной частью статистики лишь со второй трети XX века. В 30-е годы появились работы А. Н. Колмогорова и Н. В. Смирнова, предложивших и изучивших статистические критерии, носящие в настоящее время их имена (история этих работ подробно описана в статьях [33, 34]). Эти критерии основаны на использовании так называемого эмпирического процесса - разности между эмпирической и теоретической функциями распределения, умноженной на квадратный корень из объема выборки. В работе А. Н. Колмогорова 1933 г. изучено предельное распределение супремума модуля эмпирического процесса, называемого сейчас статистикой критерия Колмогорова. Затем Н. В. Смирнов исследовал су-

премум и инфимум эмпирического процесса, а также интеграл (по теоретической функции распределения) квадрата эмпирического процесса. Следует отметить, что встречающееся иногда в литературе [35] словосочетание "критерий Колмогорова - Смирнова", как подробно обосновано в [33, 34], некорректно, поскольку эти два статистика никогда не печатались вместе и не изучали один и тот же критерий. Корректно словосочетание "критерий типа Колмогорова - Смирнова", применяемое для обозначения критериев, основанных на использовании супремума функций от одного или нескольких эмпирических процессов или процессов более сложной природы.

После второй мировой войны развитие непараметрической статистики пошло быстрыми темпами [36, 37]. Большую роль сыграли работы американца Вилкоксона и его школы. К настоящему времени с помощью непараметрических методов можно решать практически тот же круг статистических задач, что и с помощью параметрических [35]. Все большую роль играют непараметрические оценки плотности, непараметрические методы регрессии и распознавания образов (дискриминантного анализа). В нашей стране непараметрические методы получили достаточно большую известность после выхода в 1965 г. первого издания упомянутого выше сборника статистических таблиц Л. Н. Большева и Н. В. Смирнова [30], содержащего подробные таблицы для основных непараметрических критериев. Тем не менее параметрические методы всё еще популярнее непараметрических, особенно среди тех прикладников, кто слабо знаком со статистическими методами. Неоднократно публиковались (см. сводки в [4, 38]) экспериментальные данные, свидетельствующие о том, что распределения реально наблюдаемых случайных величин, в частности, ошибок измерения, в подавляющем большинстве случаев отличны от нормальных (гауссовских). Тем не менее теоретики продолжают строить и изучать статистические модели, основанные на гауссовости, а практики - применять подобные методы и модели. Другими словами, как сказано в пословице, "ищут под фонарем, а не там, где потеряли".

Устойчивость статистических процедур (робастность). Если в параметрических постановках на данных накладываются слишком жесткие требования - их функции распределения должны принадлежать определенному параметрическому семейству, то в непараметрических, наоборот, излишне слабые - требуется лишь, чтобы функции распределения были непрерывны. При этом игнорируется априорная информация о том, каков "примерный вид" распределения. Априори можно ожидать, что учет этого "примерного вида" улучшит показатели качества статистических процедур. Развитием этой идеи является теория устойчивости (робастности) статистических процедур, в которой предполагается, что распределение исходных данных мало отличается от некоторого параметрического семейства. С 60-х годов эту теорию разрабатывали П. Хубер [39], Ф. Хампель [40] и многие другие. Из монографий на русском языке, трактующих о робастности и устойчивости статистических процедур, самой ранней и наиболее общей была книга [41], следующей - монография [42]. Современное состояние отражено в [43, 44]. Частными случаями реализации идеи робастности (устойчивости) статистических процедур являются рассматриваемые ниже статистика интервальных данных и нечисловая статистика.

Имеется большое разнообразие моделей робастности в зависимости от того, какие именно отклонения от заданного параметрического семейства допускаются. Сначала наиболее популярной [39, 40] была модель выбросов, в которой исходная выборка "загрязняется" малым числом "выбросов", имеющих принципиально иное распределение. Однако эта модель представляется "тупиковой", поскольку в большинстве случаев большие выбросы либо невозможны из-за ограниченности шкалы прибора, либо от них можно избавиться, применяя статистики, построенные по центральной части вариационного ряда. Кроме того, в подобных моделях обычно считается известной частота за-

сорения, что в сочетании со сказанным выше делает их малопригодными для практического использования. Более перспективной представляется модель, в которой расстояние между распределением каждого элемента выборки и базовым распределением не превосходит заданной малой величины.

Бутстреп (размножение выборок). Третье из упомянутых выше направлений - бутстреп - связано с интенсивным использованием возможностей вычислительной техники, т.е. с применением современных информационно-коммуникационных технологий. Основная идея состоит в том, чтобы теоретическое исследование дополнить или даже заменить вычислительным экспериментом [45, 46]. Вместо описания выборки распределением из параметрического семейства строим большое число "похожих" выборок, т.е. "размножаем" выборку. Затем вместо оценивания характеристик и параметров и проверки гипотез на основе свойств теоретического распределения решаем эти задачи вычислительным методом, рассчитывая интересующие нас статистики по каждой из "похожих" выборок и анализируя полученные при этом распределения. Например, вместо того, чтобы теоретическим путем находить распределение статистики, доверительные интервалы и другие характеристики, моделируют много выборок, похожих на исходную, рассчитывают соответствующие значения интересующей исследователя статистики и изучают их эмпирическое распределение. Квантили этого распределения задают доверительные интервалы, и т.д.

Термин "бутстреп" мгновенно получил известность после статьи Б. Эфрона 1979 г. [47] по этой тематике. Новый подход сразу же стал обсуждаться в массе публикаций, в том числе и научно-популярных [48]. В журнале "Заводская лаборатория. Диагностика материалов" была помещена подборка статей по бутстрепа [49], выпущен сборник статей Б.Эфрона [50]. Основная идея бутстрепа по Б. Эфрону состоит в том, что методом Монте-Карло (статистических испытаний) многократно извлекаются выборки из эмпирического распределения. Эти выборки, естественно, являются вариантами исходной, напоминают ее.

Сама по себе идея "размножения выборок" была известна гораздо раньше. Статья Б.Эфрона [47] называется так: "Бутстреп-методы: новый взгляд на метод складного ножа". Упомянутый "метод складного ножа" (jackknife) предложен М. Кенуем еще в 1949 г., за 30 лет до статьи Б. Эфрона. "Размножение выборок" при этом осуществляется путем исключения одного наблюдения. При этом для выборки объема n получаем n "похожих" на нее выборок объема $(n - 1)$ каждая. Если же исключать по 2 наблюдения, то число "похожих" выборок возрастает до $n(n - 1) / 2$ объема $(n - 2)$ каждая, и т.д.

Преимущества и недостатки бутстрепа как статистического метода обсуждаются в [51]. Там же и в [23] приводится информация о ряде аналогичных методов. Необходимо подчеркнуть, что бутстреп по Эфрону [47 - 50] - лишь один из вариантов методов "размножения выборки" (resampling), и, на наш взгляд, не самый удачный. Метод "складного ножа" представляется более полезным. На его основе можно сформулировать следующую простую практическую рекомендацию.

Вы по выборке делаете какие-либо статистические выводы и хотите узнать также, насколько эти выводы устойчивы. Если у Вас есть другие (контрольные) выборки, описывающие то же явление, то Вы можете применить к ним ту же статистическую процедуру и сравнить результаты. А если таких выборок нет? Тогда Вы можете построить их искусственно. Берете исходную выборку и исключаете один элемент. Получаете похожую выборку. Затем возвращаете этот элемент и исключаете другой. Получаете вторую похожую выборку. Поступив так со всеми элементами исходной выборки, получаете столько выборок, похожих на исходную, каков ее объем. Остается обработать их тем же способом, что и исходную, и изучить устойчивость получаемых выводов - разброс оценок параметров, частоты принятия или отклонения гипотез и т.д.

Можно изменять не выборку, а сами данные. Поскольку всегда имеются погрешности измерения, то реальные данные - это не числа, а интервалы (результат измерения плюс-минус погрешность). Нужна статистическая теория анализа таких данных.

Статистика интервальных данных. Перспективное и быстро развивающееся направление последних десятилетий - математическая статистика интервальных данных. Речь идет о развитии методов математической статистики в ситуации, когда статистические данные - не числа, а интервалы, в частности, порожденные наложением ошибок измерения на значения случайных величин. Полученные результаты отражены, в частности, в дискуссии [52] и в докладах международной конференции ИНТЕРВАЛ-92 [53]. Статистика интервальных данных идейно связана с интервальной математикой, в которой в роли чисел выступают интервалы (см., например, [54]). Это направление математики является дальнейшим развитием известных правил приближенных вычислений, посвященных выражению погрешностей суммы, разности, произведения, частного через погрешности тех чисел, над которыми осуществляются перечисленные операции. Как видно из докладов конференции ИНТЕРВАЛ-92 [53], уже к 1992 г. удалось решить, в частности, ряд задач теории интервальных дифференциальных уравнений, в которых коэффициенты, начальные условия и решения описываются с помощью интервалов.

Одна из ведущих научная школа в области статистики интервальных данных - это школа проф. А. П. Воцинина (1937–2008), активно работающая с конца 70-х годов. Полученные результаты отражены в ряде монографий (см., в частности, [55–57]), статей [52, 58, 59], докладов [53], диссертаций. В частности, изучены проблемы регрессионного анализа, планирования эксперимента, сравнения альтернатив и принятия решений в условиях интервальной неопределенности.

Рассмотрим другое - наше - направление в статистике интервальных данных, которое также представляется перспективным. В нем развиваются асимптотические методы статистического анализа интервальных данных при больших объемах выборок и малых погрешностях измерений. В отличие от классической математической статистики, сначала устремляется к бесконечности объем выборки и только потом - уменьшаются до нуля погрешности. В частности, с помощью такой асимптотики были сформулированы правила выбора метода оценивания параметров гамма-распределения в ГОСТ 11.011-83 [29]. Основные идеи статистики интервальных данных были разработаны нами при подготовке этого стандарта в начале 80-х, однако в научной печати появились лишь в начале 90-х [60, 61]. Общая схема исследования включает расчет нотны (максимально возможного отклонения статистики, вызванного интервальностью исходных данных) и рационального объема выборки (превышение которого не дает существенного повышения точности оценивания). Она применена к оцениванию математического ожидания, дисперсии, коэффициента вариации, параметров гамма-распределения и характеристик аддитивных статистик, при проверке гипотез о параметрах нормального распределения, в т.ч. с помощью критерия Стьюдента, а также гипотезы однородности с помощью критерия Смирнова. Разработаны подходы к рассмотрению интервальных данных в основных постановках регрессионного, дискриминантного и кластерного анализов. В частности, изучено влияние погрешностей измерений и наблюдений на свойства алгоритмов регрессионного анализа, разработаны способы расчета нотн и рациональных объемов выборок, введены и исследованы новые понятия многомерных и асимптотических нотн, доказаны соответствующие предельные теоремы. Начата разработка интервального дискриминантного анализа, в частности, рассмотрено влияние интервальности данных на введенный нами показатель качества классификации. Изучено асимптотическое поведение оценок метода моментов и оценок максимального правдоподобия (а также более общих - оценок минимального

контраста), проведено асимптотическое сравнение этих методов в случае интервальных данных. Найдены общие условия, при которых, в отличие от классической математической статистики, метод моментов дает более точные оценки, чем метод максимального правдоподобия. Основные результаты разработанной нами статистики интервальных данных приведены как развернутые главы в книгах [62–65].

В области асимптотической математической статистики интервальных данных российская наука имеет мировой приоритет. Развертывание работ по рассматриваемой тематике позволит закрепить этот приоритет, получить теоретические результаты, основополагающие в новой области математической статистики и необходимые для обоснованного статистического анализа почти всех типов данных. Со временем во все виды статистического программного обеспечения должны быть включены алгоритмы интервальной статистики, "параллельные" обычно используемым алгоритмам прикладной математической статистики. Это позволит в явном виде учесть наличие погрешностей у результатов наблюдений, сблизить позиции метрологов и статистиков.

Статистика объектов нечисловой природы как часть прикладной статистики. Согласно классификации статистических методов, принятой в [9, 62], прикладная статистика делится на следующие четыре области: статистика (числовых) случайных величин, многомерный статистический анализ, статистика временных рядов и случайных процессов, статистика объектов нечисловой природы. Первые три из этих областей являются классическими. Остановимся на четвертой, только еще входящей в массовое сознание специалистов. Ее именуют также статистикой нечисловых данных или нечисловой статистикой [64].

Исходный объект в математической статистике - это выборка. В вероятностной теории статистики выборка - это совокупность независимых одинаково распределенных случайных элементов. Какова природа этих элементов? В классической математической статистике (той, что обычно преподают студентам) элементы выборки - это числа. В многомерном статистическом анализе - вектора. А в нечисловой статистике элементы выборки - это объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы лежат в пространствах, не имеющих векторной структуры.

Примерами объектов нечисловой природы являются: значения качественных признаков, т.е. результаты кодировки объектов с помощью заданного перечня категорий (градаций); упорядочения (ранжировки) экспертами образцов продукции (при оценке её технического уровня и конкурентоспособности) или заявок на проведение научных работ (при проведении конкурсов на выделение грантов); классификации, т.е. разбиения объектов на группы сходных между собой (кластеры); толерантности, т.е. бинарные отношения, описывающие сходство объектов между собой, например, сходства тематики научных работ, оцениваемого экспертами с целью рационального формирования экспертных советов внутри определенной области науки; результаты парных сравнений или контроля качества продукции по альтернативному признаку ("годен" - "бракован"), т.е. последовательности из 0 и 1; множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга; графы; слова, предложения, тексты; вектора, координаты которых - совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности (форма №1-наука) или заполненная компьютеризированная история болезни, в которой часть признаков носит качественный характер, а часть - количественный; ответы на вопросы экспертной, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возможно, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты; и т.д. Интервальные

данные тоже можно рассматривать как пример объектов нечисловой природы, а именно, как частный случай нечетких множеств.

С начала 70-х годов под влиянием запросов прикладных исследований в технических, медицинских и социально-экономических науках в России активно развивается нечисловая статистика. В создании этой сравнительно новой области прикладной математической статистики приоритет принадлежит российским ученым. Большую роль сыграл основанный в 1973 г. научный семинар "Экспертные оценки и анализ данных". В 60-е годы советское научное сообщество стало интересоваться методами экспертных оценок (об их истории и современном состоянии см. [66, 67]). Началось знакомство с конкретными математизированными теориями, связанными с этими методами. Речь идет о репрезентативной теории измерений, ставшей известной в нашей стране по статье П. Суппеса и Дж. Зинеса [68] и книге И. Пфанцагля [69], о теории нечеткости Л. А. Заде [70], теории парных сравнений, описанной Г. Дэвидом [71]. К этому кругу идей примыкают теория случайных множеств (см., например, книгу Ж. Матерона [72]) и методы многомерного шкалирования (описаны А. Ю. Терехиной [73] и В. Т. Перекрестом [74]). Но наибольшее влияние оказали идеи Дж. Кемени, который аксиоматически ввел расстояние между ранжировками (теперь оно именуется в литературе расстоянием Кемени) и предложил использовать в качестве средней величины решение оптимизационной задачи (теперь - медиана Кемени). Его небольшая книга [75], написанная в соавторстве с Дж. Снеллом, породила большой поток исследований.

В течение 70-х годов на основе запросов теории экспертных оценок (а также социологии, экономики, техники и медицины) развивались конкретные направления нечисловой статистики. Были установлены связи между конкретными видами таких объектов, разработаны для них вероятностные модели [41, 64, 76].

Следующий этап - выделение нечисловой статистики в качестве самостоятельного направления в прикладной статистике, ядром которого являются методы статистического анализа данных произвольной природы. Программа развития этого нового научного направления впервые была сформулирована в 1979 г. в статье [77]. Реализация этой программы была осуществлена в 80-е годы. Для работ этого периода характерна сосредоточенность на внутренних проблемах нечисловой статистики. Ссылки на конкретные монографии, сборники, статьи и иные публикации нескольких десятков авторов приведены в обзорах [78–80]. Отметим лишь сборник [81], специально посвященный нечисловой статистике, и диссертацию [82], относящуюся к непараметрической теории парных сравнений (теории люсианов).

К 90-м годам статистика объектов нечисловой природы с теоретической точки зрения была достаточно хорошо развита, основные идеи, подходы и методы были разработаны и изучены математически, в частности, доказано достаточно много теорем. Однако она оставалась недостаточно апробированной на практике. Это было связано как с ее сравнительной молодостью, так и с общеизвестными особенностями организации науки в 80-е годы, когда отсутствовали достаточные стимулы для того, чтобы теоретики занялись широким внедрением своих результатов. И в 90-е годы наступило время от математико-статистических исследований перейти к применению полученных результатов на практике.

Следует отметить, что в нечисловой статистике, как и в других областях прикладной математической статистики и прикладной математики вообще, одна и та же математическая схема может с успехом применяться и в технических исследованиях, и в медицине, и в социологии, и для анализа экспертных оценок, а потому ее лучше всего формулировать и изучать в наиболее общем виде, для объектов произвольной природы.

3.1.5. Основные идеи нечисловой статистики

В чем её принципиальная новизна? Для классической математической статистики характерна операция сложения. При расчете выборочных характеристик распределения (выборочное среднее арифметическое, выборочная дисперсия и др.), в регрессионном анализе и других областях этой научной дисциплины постоянно используются суммы. Математический аппарат - законы больших чисел, Центральная предельная теорема и другие теоремы - нацелены на изучение сумм. В нечисловой же статистике нельзя использовать операцию сложения, поскольку элементы выборки лежат в пространствах, где нет операции сложения. Методы обработки нечисловых данных основаны на принципиально ином математическом аппарате - на применении различных расстояний в пространствах объектов нечисловой природы.

Кратко рассмотрим несколько идей, развиваемых в статистике объектов нечисловой природы для данных, лежащих в пространствах произвольного вида. Решаются классические задачи описания данных, оценивания, проверки гипотез - но для неклассических данных, а потому неклассическими методами.

Первой обсудим проблему определения средних величин. В рамках репрезентативной теории измерений удастся указать вид средних величин, соответствующих тем или иным шкалам измерения [41]. В классической математической статистике средние величины вводят с помощью операций сложения (выборочное среднее арифметическое, математическое ожидание) или упорядочения (выборочная и теоретическая медианы). В пространствах произвольной природы средние значения нельзя определить с помощью операций сложения или упорядочения. Теоретические и эмпирические средние приходится вводить как решения экстремальных задач. Для теоретического среднего это - задача минимизации математического ожидания (в классическом смысле) расстояния от случайного элемента со значениями в рассматриваемом пространстве до фиксированной точки этого пространства (минимизируется указанная функция от этой точки). Для эмпирического среднего математическое ожидание берется по эмпирическому распределению, т.е. берется сумма расстояний от некоторой точки до элементов выборки и затем минимизируется по этой точке. При этом как эмпирическое, так и теоретическое средние как решения экстремальных задач могут быть не единственным элементом пространства, а состоять из множества таких элементов, которое может оказаться и пустым. Тем не менее удалось сформулировать и доказать законы больших чисел для средних величин, определенных указанным образом, т.е. установить сходимость эмпирических средних к теоретическим [64].

Оказалось, что методы доказательства законов больших чисел допускают существенно более широкую область применения, чем та, для которой они были разработаны. А именно, удалось изучить асимптотику решений экстремальных статистических задач, к которым, как известно, сводится большинство постановок прикладной статистики. В частности, кроме законов больших чисел установлена и состоятельность оценок минимального контраста, в том числе оценок максимального правдоподобия и робастных оценок. К настоящему времени подобные оценки изучены также и в статистике интервальных данных.

В статистике в пространствах произвольной природы большую роль играют непараметрические оценки плотности, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализов. В нечисловой статистике предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в частности, доказана их состоятельность, изучена скорость сходимости и установлен примечательный факт совпадения наилучшей скорости схо-

димости в произвольном случае с той, которая имеет быть в классической теории для числовых случайных величин.

Дискриминантный, кластерный, регрессионный анализы в пространствах произвольной природы основаны либо на параметрической теории - и тогда применяется подход, связанный с асимптотикой решения экстремальных статистических задач - либо на непараметрической теории - и тогда используются алгоритмы на основе непараметрических оценок плотности.

Для проверки гипотез могут быть использованы статистики интегрального типа, в частности, типа омега-квадрат. Любопытно, что предельная теория таких статистик, построенная первоначально в классической постановке [83], приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида [84], поскольку при этом удалось провести рассуждения, опираясь на базовые математические соотношения, а не на те частные (с общей точки зрения), что были связаны с конечномерным пространством.

Представляют интерес результаты, связанные с конкретными областями статистики объектов нечисловой природы, в частности, со статистикой нечетких множеств [85], со случайными множествами [41] (следует отметить, что теория нечетких множеств в определенном смысле сводится к теории случайных множеств [41, 86]), с непараметрической теорией парных сравнений [82], с аксиоматическим введением метрик в конкретных пространствах объектов нечисловой природы [81, 87].

Для анализа нечисловых, в частности, экспертных данных весьма важны методы классификации. С другой стороны, наиболее естественно ставить и решать задачи классификации, основанные на использовании расстояний или показателей различия, в рамках статистики объектов нечисловой природы. Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа). Современное состояние дискриминантного и кластерного анализа с точки зрения нечисловой статистики отражено работах в [88–90].

Статистические методы анализа нечисловых данных особенно хорошо приспособлены для применения в экономике, социологии и экспертных оценках, поскольку в этих областях от 50% до 90% данных являются нечисловыми [91].

3.1.6. Итоги анализа структуры прикладной статистики

Выше рассмотрены пять "точек роста" прикладной статистики. Разумеется, они не исчерпывают все многообразие фронта научных исследований в этой области. В частности, решены отнюдь не все проблемы, поставленные в конце 70-х годов в т.н. "цах-кадзорской тетради" [92]. Кроме того, мы почти не затрагивали разнообразные применения статистических методов в конкретных прикладных областях. Много интересных проблем есть в планировании экспериментов, особенно кинетических, при анализе проблем надежности, в статистических методах управления качеством продукции, в вопросах экологии и безопасности и др.

В течение последних более чем 60 лет в России наблюдается огромный разрыв между государственной статистикой и научным сообществом специалистов по статистическим методам (подробнее об этом см. [24]). В учебнике по истории статистики [16], подготовленном ориентированными на обслуживание государственной статистики лицами, даже не упоминаются имена членов-корреспондентов АН СССР Н.В. Смирнова и Л.Н. Большева! Поэтому нет ничего удивительного в том, что тенденции развития современной прикладной математической статистики столь же мало обсуждаются отечественными авторами, как и ее история.

3.2. Теоретические инструменты статистических методов

Набор широко применяемых исследователями теоретических инструментов прикладной математической статистики и статистических методов в целом достаточно ограничен. В настоящем разделе собраны основные математические инструменты (теоремы, методы), постоянно используемые при обосновании новых результатов в области статистических методов. Эти инструменты отнюдь не всегда легко найти в литературе по теории вероятностей и математической статистике. Например, такие рассматриваемые далее теоремы и методы, как многомерная центральная предельная теорема, теоремы о наследовании сходимости и метод линеаризации, даже не включены в энциклопедию «Вероятность и математическая статистика» [1] – наиболее полный, по мнению составителей энциклопедии, свод знаний по заявленной тематике. Последний факт наглядно демонстрирует разрыв между математической дисциплиной «теория вероятностей и математическая статистика» и потребностями прикладной статистики и других статистических методов.

3.2.1. Законы больших чисел

Законы больших чисел позволяют описать поведение сумм случайных величин. Примером является следующий результат, доказанный русским математиком П. Л. Чебышёвым (1821–1894) в 1867 г. Пусть сначала вероятностное пространство состоит из конечного числа элементов.

Теорема Чебышёва. Пусть случайные величины X_1, X_2, \dots, X_k попарно независимы и существует число C такое, что $D(X_i) < C$ при всех $i = 1, 2, \dots, k$. Тогда для любого положительного ε выполнено неравенство

$$P\left\{\left|\frac{X_1 + X_2 + \dots + X_k}{k} - \frac{M(X_1) + M(X_2) + \dots + M(X_k)}{k}\right| \geq \varepsilon\right\} \leq \frac{C}{k\varepsilon^2}. \quad (1)$$

Частным случаем теоремы Чебышева является теорема Бернулли – первый в истории вариант закона больших чисел. Известный математики Якоб Бернулли (1654–1705), живший в городе Базель в Швейцарии, в самом конце XVII века доказал это утверждение в рамках математической модели (опубликовано доказательство было лишь после его смерти, в 1713 году). Современная формулировка теоремы Бернулли такова.

Теорема Бернулли. Пусть m – число наступлений события A в k независимых (попарно) испытаниях, и p есть вероятность наступления события A в каждом из испытаний. Тогда при любом $\varepsilon > 0$ справедливо неравенство

$$P\left\{\left|\frac{m}{k} - p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{k\varepsilon^2}. \quad (2)$$

Ясно, что при росте k выражения в правых частях формул (1) и (2) стремятся к 0. Таким образом, среднее арифметическое попарно независимых случайных величин сближается со средним арифметическим их математических ожиданий.

Выше шла речь лишь о пространствах элементарных событий из конечного числа элементов. Однако приведенные теоремы верны и в общем случае, для произвольных пространств элементарных событий. Однако в список условий закона больших чисел необходимо добавить требование существования дисперсий. Легко видеть, что если существуют дисперсии, то существуют и математические ожидания. Закон больших чисел в форме Чебышёва приобретает следующий вид.

Теорема Чебышёва [2, с.147]. Если $X_1, X_2, \dots, X_k, \dots$ - последовательность попарно независимых случайных величин, имеющих конечные дисперсии, ограниченные одной и той же постоянной,

$$D(X_1) \leq C, D(X_2) \leq C, \dots D(X_i) \leq C, \dots$$

то, каково бы ни было постоянное $\varepsilon > 0$,

$$\lim_{k \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{j=1}^k X_j - \frac{1}{n} \sum_{j=1}^k MX_j \right| < \varepsilon \right\} = 1. \quad (3)$$

С точки зрения прикладных статистических исследований ограниченность дисперсий вполне естественна. Она вытекает, например, из ограниченности диапазона изменения практически всех величин, используемых при реальных расчетах.

В 1923 г. А. Я. Хинчин показал, что если случайные величины не только независимы, но и одинаково распределены, то существование у них математического ожидания является необходимым и достаточным условием для применимости закона больших чисел [2, с.150]. Найдены и более экзотические варианты закона больших чисел. Например, такой.

Теорема [2, с.150–151]. Для того чтобы для последовательности $X_1, X_2, \dots, X_k, \dots$ (как угодно зависимых) случайных величин при любом положительном ε выполнялось соотношение (3), необходимо и достаточно, чтобы при $n \rightarrow \infty$

$$M \left(\frac{\left(\sum_{j=1}^k (X_j - MX_j) \right)^2}{n^2 + \left(\sum_{j=1}^k (X_j - MX_j) \right)^2} \right) \rightarrow 0.$$

Законы больших чисел для случайных величин служат основой для аналогичных утверждений для случайных элементов в пространствах более сложной природы, в частности, в пространствах произвольной природы [3, 4]. Однако здесь мы ограничимся классическими формулировками, служащими основой для современных статистических методов.

Смысл классических законов больших чисел состоит в том, что выборочное среднее арифметическое независимых одинаково распределенных случайных величин приближается (сходится) к математическому ожиданию этих величин. Другими словами, *выборочные средние сходятся к теоретическому среднему*.

Это утверждение справедливо и для других видов средних. Например, выборочная медиана сходится к теоретической медиане. Это утверждение – тоже закон больших чисел, но не классический.

Существенным продвижением в теории вероятностей во второй половине XX в. явилось введение средних величин в пространствах произвольной природы и получение для них законов больших чисел, т.е. утверждений, состоящих в том, что эмпирические (т.е. выборочные) средние сходятся к теоретическим средним [3, 4].

3.2.2. Центральные предельные теоремы

Простейший вариант Центральной предельной теоремы (ЦПТ) теории вероятностей таков.

Центральная предельная теорема (для одинаково распределенных слагаемых). Пусть $X_1, X_2, \dots, X_n, \dots$ – независимые одинаково распределенные случайные величины с математическими ожиданиями $M(X_i) = m$ и дисперсиями $D(X_i) = \sigma^2, i = 1, 2, \dots, n, \dots$ Тогда для любого действительного числа x существует предел

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} < x\right) = \Phi(x),$$

где $\Phi(x)$ – функция стандартного нормального распределения.

Эту теорему иногда называют теоремой Линдеберга – Леви [5, с.122].

В ряде прикладных задач не выполнено условие одинаковой распределенности. В таких случаях центральная предельная теорема обычно остается справедливой, однако на последовательность случайных величин приходится накладывать те или иные условия. Суть этих условий состоит в том, что ни одно слагаемое не должно быть доминирующим, вклад каждого слагаемого в среднее арифметическое должен быть пренебрежимо мал по сравнению с итоговой суммой. Наиболее часто используется теорема Ляпунова.

Теорема Ляпунова - Центральная предельная теорема (для разнораспределенных слагаемых). Пусть $X_1, X_2, \dots, X_n, \dots$ – независимые случайные величины с математическими ожиданиями $M(X_i) = m_i$ и дисперсиями $D(X_i) = \sigma_i^2 \neq 0, i = 1, 2, \dots, n, \dots$. Пусть при некотором $\delta > 0$ у всех рассматриваемых случайных величин существуют центральные моменты порядка $2+\delta$ и безгранично убывает «дробь Ляпунова»:

$$\lim_{k \rightarrow \infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^n M |X_k - m_k|^{2+\delta} = 0,$$

где

$$B_k^2 = \sum_{i=1}^k \sigma_i^2 = D\left(\sum_{i=1}^k X_i\right).$$

Тогда для любого действительного числа x существует предел

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + X_2 + \dots + X_n - m_1 - m_2 - \dots - m_n}{B_n} < x\right) = \Phi(x), \quad (4)$$

где $\Phi(x)$ – функция стандартного нормального распределения.

В случае одинаково распределенных случайных слагаемых

$$m_1 = m_2 = \dots = m_n = m, \quad B_n = D(X_1 + X_2 + \dots + X_n) = \sigma\sqrt{n},$$

и теорема Ляпунова переходит в теорему Линдеберга – Леви.

История получения центральных предельных теорем для числовых случайных величин растянулась на два века – от первых работ Муавра в 30-х гг. XVIII в. для необходимых и достаточных условий, полученных Линдебергом и Феллером в 30-х гг. XX в.

Теорема Линдеберга - Феллера. Пусть $X_1, X_2, \dots, X_n, \dots$ – независимые случайные величины с математическими ожиданиями $M(X_i) = m_i$ и дисперсиями $D(X_i) = \sigma_i^2 \neq 0, i = 1, 2, \dots, n, \dots$. Предельное соотношение (4), т.е. Центральная предельная теорема, выполнено тогда и только тогда, когда при любом $\tau > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-m_k| > \tau B_n} (x - m_k)^2 dF_k(x) = 0,$$

где $F_k(x)$ обозначает функцию распределения случайной величины X_k .

Доказательства перечисленных в настоящем разделе центральных предельных теорем для случайных величин можно найти в классическом курсе теории вероятностей [2].

Для обоснования многих статистических методов большое значение имеет многомерная центральная предельная теорема. В ней речь идет не о сумме случайных величин, а о сумме случайных векторов.

Необходимое и достаточное условие многомерной сходимости [5, с.124]. Пусть F_n обозначает совместную функцию распределения k -мерного случайного вектора $(X_n^{(1)}, \dots, X_n^{(k)})$, $n = 1, 2, \dots$, и $F_{\lambda n}$ – функция распределения линейной комбинации $\lambda_1 X_n^{(1)} + \dots + \lambda_k X_n^{(k)}$. Необходимое и достаточное условие для сходимости F_n к некоторой k -мерной функции распределения F состоит в том, что $F_{\lambda n}$ имеет предел для любого вектора λ .

Приведенная теорема ценна тем, что с ее помощью сходимость распределений случайных векторов сводится к сходимости распределений линейных комбинаций их координат, т.е. к сходимости обычных (числовых) случайных величин, рассмотренных ранее. Однако она не дает возможности непосредственно указать предельное распределение. Это можно сделать с помощью следующей теоремы.

Теорема о многомерной сходимости [5]. Пусть F_n и $F_{\lambda n}$ – те же, что в предыдущей теореме. Пусть F – совместная функция распределения k -мерного случайного вектора (X_1, \dots, X_k) . Если функция распределения $F_{\lambda n}$ сходится при росте объема выборки к функции распределения F_λ для любого вектора λ , где F_λ – функция распределения линейной комбинации $\lambda_1 X_1 + \dots + \lambda_k X_k$, то F_n сходится к F .

Здесь сходимость F_n к F означает, что для любого k -мерного вектора (x_1, \dots, x_k) такого, что функция распределения F непрерывна в (x_1, \dots, x_k) , числовая последовательность $F_n(x_1, \dots, x_k)$ сходится при росте n к числу $F(x_1, \dots, x_k)$. Другими словами, сходимость функций распределения понимается точно так же, как при обсуждении предельных теорем для случайных величин выше. Приведем многомерный аналог этих теорем.

Многомерная центральная предельная теорема [5]. Рассмотрим независимые одинаково распределенные k -мерные случайные вектора

$$U_n' = (U_{1n}, \dots, U_{kn}), \quad n = 1, 2, \dots,$$

где штрих обозначает операцию транспонирования вектора. Предположим, что случайные вектора U_n имеют моменты первого и второго порядка, т.е.

$$M(U_n) = \mu, \quad D(U_n) = \Sigma,$$

где μ – вектор математических ожиданий координат случайного вектора, Σ – его ковариационная матрица. Введем последовательность средних арифметических случайных векторов:

$$\bar{U}_n = (\bar{U}_{1n}, \dots, \bar{U}_{kn}), \quad n = 1, 2, \dots, \quad \bar{U}_{in} = \frac{1}{n} \sum_{j=1}^n U_{ij}.$$

Тогда случайный вектор $\sqrt{n}(\bar{U}_n - \mu)$ имеет асимптотическое k -мерное нормальное распределение $N_k(0, \Sigma)$, т.е. он асимптотически распределен так же, как k -мерная нормальная величина с нулевым математическим ожиданием, ковариационной Σ и плотностью

$$N_k(u | 0, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} u' \Sigma^{-1} u\right\}.$$

Здесь $|\Sigma|$ – определитель матрицы Σ . Другими словами, распределение случайного вектора $\sqrt{n}(\bar{U}_n - \mu)$ сходится к k -мерному нормальному распределению с нулевым математическим ожиданием и ковариационной матрицей Σ .

Напомним, что многомерным нормальным распределением с математическим ожиданием μ и ковариационной матрицей Σ называется распределение, имеющее плотность

$$N_k(u | \mu, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}[(u - \mu)' \Sigma^{-1}(u - \mu)]\right\}.$$

Многомерная центральная предельная теорема показывает, что распределения сумм независимых одинаково распределенных случайных векторов при большом числе слагаемых хорошо приближаются с помощью нормальных распределений, имеющих такие же первые два момента (вектор математических ожиданий координат случайного вектора и его корреляционную матрицу), как и исходные вектора. От одинаковости распределенности можно отказаться, но это потребует некоторого усложнения символики. В целом из теоремы о многомерной сходимости вытекает, что многомерный случай ничем принципиально не отличается от одномерного.

Пример. Пусть X_1, \dots, X_n, \dots – независимые одинаково распределенные случайные величины. Рассмотрим k -мерные независимые одинаково распределенные случайные вектора

$$U_n = (X_n, X_n^2, X_n^3, \dots, X_n^k), \quad n = 1, 2, \dots$$

Их математическое ожидание – вектор теоретических начальных моментов, а ковариационная матрица составлена из соответствующих центральных моментов. Тогда \bar{U}_n – вектор выборочных центральных моментов. Многомерная центральная предельная теорема утверждает, что \bar{U}_n имеет асимптотически нормальное распределение. Как вытекает из теорем о наследовании сходимости и о линеаризации (см. ниже), из распределения \bar{U}_n можно вывести распределения различных функций от выборочных начальных моментов. А поскольку центральные моменты выражаются через начальные моменты, то аналогичное утверждение верно и для них.

3.2.3. Теоремы о наследовании сходимости

Суть проблемы наследования сходимости. Пусть распределения случайных величин X_n при $n \rightarrow \infty$ стремятся к распределению случайной величины X . При каких функциях f можно утверждать, что распределения случайных величин $f(X_n)$ сходятся к распределению $f(X)$, т.е. наследуется сходимость?

Хорошо известно, что для непрерывных функций f сходимость наследуется [5]. Однако в статистических методах используются различные обобщения этого утверждения. Необходимость обобщений связана с тремя обстоятельствами.

1) Статистические данные могут моделироваться не только случайными величинами, но и случайными векторами, случайными множествами, случайными элементами произвольной природы (т.е. функциями на вероятностном пространстве со значениями в произвольном множестве) [6, 7].

2) Переход к пределу должен рассматриваться не только для случая безграничного возрастания объема выборки, но и в более общих случаях. Например, если в постановке статистической задачи участвуют несколько выборок объемов $n(1), n(2), \dots, n(k)$, то вполне обычным является предположение о безграничном росте всех этих объемов (что можно описать и как $\min\{n(1), n(2), \dots, n(k)\} \rightarrow \infty$).

3) Функция f не обязательно является непрерывной. Она может иметь разрывы. Кроме того, она может зависеть от параметров, по которым происходит переход к пределу. Например, может зависеть от объемов выборок. Например, в [8, гл.5] понадобилось рассмотреть функцию $f = f(n(1), n(2), \dots, n(k))$.

Расстояние Прохорова и сходимость по направленному множеству. Введем необходимые для дальнейшего изложения понятия.

Расстояние (метрика) Прохорова. Пусть C – некоторое пространство, A – его подмножество, d – метрика в C . Назовем ε -окрестностью множества A в метрике d следующее множество:

$$S(A, \varepsilon) = \{x \in C: d(A, x) < \varepsilon\}.$$

Таким образом, ε -окрестность множества A – это совокупность всех точек пространства C , отстоящих от A не более чем на положительное число ε . При этом расстояние от точки x до множества A – это точная нижняя грань расстояний от x до точек множества A , т.е.

$$d(A, x) = \inf\{d(x, y): y \in A\}.$$

Пусть P_1 и P_2 – две вероятностные меры на C (т.е. распределения двух случайных элементов со значениями в C). Пусть D_{12} – множество чисел $\varepsilon > 0$ таких, что

$$P_1(A) \leq P_2(S(A, \varepsilon)) + \varepsilon$$

для любого замкнутого подмножества A пространства C . Пусть D_{21} – множество чисел $\varepsilon > 0$ таких, что

$$P_2(A) \leq P_1(S(A, \varepsilon)) + \varepsilon$$

для любого замкнутого подмножества A пространства C . Расстояние Прохорова $L(P_1, P_2)$ между вероятностными мерами (его можно рассматривать и как расстояние между случайными элементами с распределениями P_1 и P_2 соответственно) вводится формулой

$$L(P_1, P_2) = \max(\inf D_{12}, \inf D_{21}).$$

С помощью метрики Прохорова формализуется понятие сходимости распределений случайных элементов в произвольном пространстве.

Расстояние $L(P_1, P_2)$ введено академиком АН СССР РАН Юрием Васильевичем Прохоровым (1929–2013) в середине XX в. [9] и широко используется в современной теории вероятностей.

Сходимость по направленному множеству [10, с.95-96]. Бинарное отношение \geq (упорядочение), заданное на множестве B , называется направлением на нем, если B не пусто и

- (а) если m, n и p – такие элементы множества B , что $m \geq n$ и $n \geq p$, то $m \geq p$;
- (б) $m \geq m$ для любого m из B ;
- (в) если m и n принадлежат B , то найдется элемент p из B такой, что $p \geq m$ и $p \geq n$.

Направленное множество – это пара (B, \geq) , где \geq – направление на множестве B . Направленностью (или «последовательностью по направленному множеству») называется пара (f, \geq) , где f – функция, \geq – направление на ее области определения. Пусть $f: B \rightarrow Y$, где Y – топологическое пространство. Направленность (f, \geq) сходится в топологическом пространстве Y к точке y_0 , если для любой окрестности U точки y_0 найдется p из B такое, что $f(q) \in U$ при любом $q \geq p$. В таком случае говорят также о сходимости по направленному множеству.

Пусть $B = \{(n(1), n(2), \dots, n(k))\}$ – совокупность векторов, каждый из которых составлен из объемов k выборок. Пусть

$$(n(1), n(2), \dots, n(k)) \geq (n_1(1), n_1(2), \dots, n_1(k))$$

тогда и только тогда, когда $n(i) \geq n_1(i)$ при всех $i = 1, 2, \dots, k$. Тогда (B, \geq) – направленное множество, сходимость по которому эквивалентна сходимости при $\min\{n(1), n(2), \dots, n(k)\} \rightarrow \infty$.

Чтобы охватить различные частные случаи, целесообразно предельные теоремы формулировать в терминах сходимости по направленному множеству. Будем писать $B = \{\alpha\}$. Пусть запись $\alpha \rightarrow \infty$ обозначает переход к пределу по направленному множеству.

Формулировка проблемы наследования сходимости. Пусть случайные элементы X_α со значениями в пространстве C сходятся при $\alpha \rightarrow \infty$ к случайному элементу X , где через $\alpha \rightarrow \infty$ обозначен переход к пределу по направленному множеству. Сходимость

случайных элементов означает, что $L(X_\alpha, X) \rightarrow 0$ при $\alpha \rightarrow \infty$, где L – метрика Прохорова в пространстве C .

Пусть $f_\alpha: C \rightarrow Y$ – некоторые функции. Какие условия надо на них наложить, чтобы из $L(X_\alpha, X) \rightarrow 0$ вытекало, что $L_1(f_\alpha(X_\alpha), f_\alpha(X)) \rightarrow 0$ при $\alpha \rightarrow \infty$, где L_1 – метрика Прохорова в пространстве Y ? Другими словами, какие условия на функции $f_\alpha: C \rightarrow Y$ гарантируют наследование сходимости?

В работах [11, 12] найдены необходимые и достаточные условия на функции $f_\alpha: C \rightarrow Y$, гарантирующие наследование сходимости. Описанию этих условий посвящена оставшаяся часть подраздела.

Приведем для полноты изложения строгие формулировки математических предположений.

Математические предположения. Пусть C и Y – полные сепарабельные метрические пространства [10]. Пусть выполнены обычные предположения измеримости: X_α и X – случайные элементы C , $f_\alpha(X_\alpha)$ и $f_\alpha(X)$ – случайные элементы в Y , рассматриваемые ниже подмножества пространств C и Y лежат в соответствующих σ -алгебрах измеримых подмножеств, и т.д.

Понадобятся некоторые *определения*. Разбиение $T_n = \{C_{1n}, C_{2n}, \dots, C_{nn}\}$ пространства C – это такой набор подмножеств C_j , $j = 1, 2, \dots, n$, этого пространства, что пересечение любых двух из них пусто, а объединение совпадает с C . Диаметром $diam(A)$ подмножества A множества C называется точная верхняя грань расстояний между элементами A , т.е.

$$diam(A) = \sup \{d(x, y), x \in A, y \in A\},$$

где $d(x, y)$ – метрика в пространстве C . Обозначим ∂A границу множества A , т.е. совокупность точек x таких, что любая их окрестность $U(x)$ имеет непустое пересечение как с A , так и с $C \setminus A$. Колебанием $\delta(f, B)$ функции f на множестве B называется $\delta(f, B) = \sup \{|f(x) - f(y)|, x \in B, y \in B\}$.

Достаточное условие для наследования сходимости. Пусть $L(X_\alpha, X) \rightarrow 0$ при $\alpha \rightarrow \infty$. Пусть существует последовательность T_n разбиений пространства C такая, что $P(X \in \partial A) = 0$ для любого A из T_n и, основное условие, для любого $\varepsilon > 0$

$$m_\varepsilon(\alpha, n) = \sum P(X \in A) \rightarrow 0 \quad (5)$$

при $n \rightarrow \infty$ и $\alpha \rightarrow \infty$, где сумма берется по всем тем A из T_n , для которых колебание функции f_α на A больше ε , т.е. $\delta(f_\alpha, A) > \varepsilon$. Тогда $L_1(f_\alpha(X_\alpha), f_\alpha(X)) \rightarrow 0$ при $\alpha \rightarrow \infty$.

Необходимое условие для наследования сходимости. Пусть Y – конечномерное линейное пространство, $Y = R^k$. Пусть случайные элементы $f_\alpha(X)$ асимптотически ограничены по вероятности при $\alpha \rightarrow \infty$, т.е. для любого $\varepsilon > 0$ существуют число $S(\varepsilon)$ и элемент направленного множества $\alpha(\varepsilon)$ такие, что $P(\|f_\alpha(X)\| > S(\varepsilon)) < \varepsilon$ при $\alpha \geq \alpha(\varepsilon)$, где $\|f_\alpha(X)\|$ – норма (длина) вектора $f_\alpha(X)$. Пусть существует последовательность T_n разбиений пространства C такая, что

$$\lim_{n \rightarrow \infty} \max \{diam(C_{jn}), C_{jn} \in T_n\} = 0,$$

т.е. последовательность T_n является безгранично измельчающейся. Самое существенное – пусть условие (5) не выполнено для последовательности T_n . Тогда существует последовательность случайных элементов X_α такая, что $L(X_\alpha, X) \rightarrow 0$ при $\alpha \rightarrow \infty$, но $L_1(f_\alpha(X_\alpha), f_\alpha(X))$ не сходится к 0 при $\alpha \rightarrow \infty$.

Несколько огрубляя, можно сказать, что *условие (5) является необходимым и достаточным для наследования сходимости.*

Пример 1. Пусть C и Y – конечномерные линейные пространства, функции f_α не зависят от α , т.е. $f_\alpha \equiv f$, причем функция f ограничена. Тогда условие (5) эквивалентно требованию интегрируемости по Риману – Стильтьесу функции f по мере $G(A) = P(X \in A)$. В частности, условие (5) выполнено для непрерывной функции f .

В конечномерных пространствах C вместо сходимости $L(X_\alpha, X) \rightarrow 0$ при $\alpha \rightarrow \infty$ можно говорить о слабой сходимости функций распределения случайных векторов X_α к функции распределения случайного вектора X . Речь идет о «сходимости по распределению», т.е. о сходимости во всех точках непрерывности функции распределения случайного вектора X . В этом случае разбиения могут состоять из многомерных параллелепипедов [12, гл.2].

Пример 2. Полученные выше результаты дают обоснование для рассуждений типа следующего (ср., например, утверждения в [8, гл.5] выше). Пусть по двум независимым выборкам объемов m и n соответственно построены статистики X_m и Y_n . Пусть известно, что распределения этих статистик сходятся при безграничном росте объемов выборок к стандартным нормальным распределениям с математическим ожиданием 0 и дисперсией 1. Пусть $a(m, n)$ и $b(m, n)$ – некоторые коэффициенты. Тогда согласно результатам примера 1 распределение случайной величины $Z(m, n) = a(m, n)X_m + b(m, n)Y_n$ сближается с распределением нормально распределенной случайной величины с математическим ожиданием 0 и дисперсией $a^2(m, n) + b^2(m, n)$. Если же $a^2(m, n) + b^2(m, n) = 1$, например,

$$a(m, n) = \sqrt{\frac{m}{m+n}}, \quad b(m, n) = \sqrt{\frac{n}{m+n}},$$

то распределение $Z(m, n)$ сходится при безграничном росте объемов выборок к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1.

3.2.4. Метод линеаризации

При разработке статистических методов часто возникает следующая задача (см., например, [5, с.338]). Имеется последовательность k -мерных случайных векторов $X_n = (X_{1n}, X_{2n}, \dots, X_{kn})$, $n = 1, 2, \dots$, такая, что $X_n \rightarrow a = (a_1, a_2, \dots, a_k)$ при $n \rightarrow \infty$, и последовательность функций $f_n: R^k \rightarrow R^1$. Требуется найти распределение случайной величины $f_n(X_n)$.

Основная идея – рассмотреть главный линейный член функции f_n в окрестности точки a . Из математического анализа известно, что

$$f_n(X_n) - f_n(a) = \sum_{j=1}^k \frac{\partial f_n(a)}{\partial x_j} (X_{jn} - a_j) + O_n(\|X_n - a\|^2),$$

где остаточный член является бесконечно малой величиной более высокого порядка малости, чем линейный член. Таким образом, произвольная функция может быть заменена на линейную функцию от координат случайного вектора. Эта замена проводится с точностью до бесконечно малых более высокого порядка. Конечно, должны быть выполнены некоторые математические условия регулярности. Например, функции f_n должны быть дважды непрерывно дифференцируемы в окрестности точки a .

Если вектор X_n является асимптотически нормальным с математическим ожиданием a и ковариационной матрицей Σ/n , где $\Sigma = \|\sigma_{ij}\|$, причем $\sigma_{ij} = nM(X_i - a_i)(X_j - a_j)$, то линейная функция от его координат также асимптотически нормальна. Следовательно, при очевидных условиях регулярности $f_n(X_n)$ – асимптотически нормальная случайная величина с математическим ожиданием $f_n(a)$ и дисперсией

$$\frac{1}{n} \sum_{i=1}^k \sum_{j=1}^k \frac{\partial f_n(a)}{\partial x_i} \frac{\partial f_n(a)}{\partial x_j} \sigma_{ij}.$$

Для практического использования асимптотической нормальности $f_n(X_n)$ остается заменить неизвестные моменты a и Σ на их оценки. Например, если X_n – это среднее

арифметическое независимых одинаково распределенных случайных векторов, то a можно заменить на X_n , а Σ - на выборочную ковариационную матрицу.

Пример. Пусть Y_1, Y_2, \dots, Y_n - независимые одинаково распределенные случайные величины с математическим ожиданием a и дисперсией σ^2 . В качестве X_n (при $k = 1$) рассмотрим выборочное среднее арифметическое

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}.$$

Как известно, в силу закона больших чисел $\bar{Y} \rightarrow a = M(Y)$. Следовательно, для получения распределений функций от выборочного среднего арифметического можно использовать метод линеаризации. В качестве примера рассмотрим $f_n(y) = f(y) = y^2$. Тогда

$$(\bar{Y})^2 - a^2 = \frac{df(a)}{dy}(\bar{Y} - a) + O((\bar{Y} - a)^2) = 2a(\bar{Y} - a) + O((\bar{Y} - a)^2).$$

Из этого соотношения следует, что с точностью до бесконечно малых более высокого порядка

$$(\bar{Y})^2 = a^2 + 2a(\bar{Y} - a).$$

Поскольку в соответствии с Центральной Предельной Теоремой выборочное среднее арифметическое является асимптотически нормальной случайной величиной с математическим ожиданием a и дисперсией σ^2/n , то квадрат этой статистики является асимптотически нормальной случайной величиной с математическим ожиданием a^2 и дисперсией $4a^2\sigma^2/n$. Для практического использования может оказаться полезной замена параметров (асимптотического нормального распределения) на их оценки, а именно, математического ожидания - на $(\bar{Y})^2$, а дисперсии - на $4(\bar{Y})^2 s^2/n$, где s^2 - выборочная дисперсия.

Большое внимание (целая глава!) уделено методу линеаризации в классическом учебнике Е. С. Вентцель [13].

3.2.5. Принцип инвариантности

Пусть Y_1, Y_2, \dots, Y_n - независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$. Многие используемые в статистических методах функции от результатов наблюдений выражаются через эмпирическую функцию распределения $F_n(x)$. К ним относятся, в частности, статистики Колмогорова, Смирнова, омега-квадрат [14]. Отметим, что и другие статистики выражаются через эмпирическую функцию распределения, например:

$$\bar{Y} = \int_{-\infty}^{+\infty} x dF_n(x).$$

Полезным является преобразование Н.В.Смирнова $t = F(x)$. Тогда независимые случайные величины $Z_j = F(Y_j), j = 1, 2, \dots, n$, имеют равномерное распределение на отрезке $[0; 1]$. Рассмотрим построенную по ним эмпирическую функцию распределения $F_n(t), 0 \leq t \leq 1$. *Эмпирическим процессом* называется случайный процесс

$$\xi_n(t) = \sqrt{n}(F_n(t) - t).$$

Рассмотрим критерии проверки согласия функции распределения выборки с фиксированной функцией распределения $F(x)$. Статистика критерия Колмогорова записывается в виде

$$K_n = \sup_{0 \leq t \leq 1} |\xi_n(t)|,$$

статистика критерия Смирнова – это

$$S_n = \sup_{0 \leq t \leq 1} \xi_n(t),$$

а статистика критерия омега-квадрат (Крамера - Мизеса - Смирнова) имеет вид

$$\omega_n^2 = \int_0^1 \xi_n^2(t) dt.$$

Случайный процесс $\xi_n(t)$ имеет нулевое математическое ожидание и ковариационную функцию $M\xi_n(s)\xi_n(t) = \min(s, t) - st$. Рассмотрим гауссовский случайный процесс $\xi(t)$ с такими же математическим ожиданием и ковариационной функцией. Он называется броуновским мостом. (Напомним, что гауссовским процесс именуется потому, что вектор $(\xi(t_1), \xi(t_2), \dots, \xi(t_k))$ имеет многомерное нормальное распределение при любых наборах моментов времени t_1, t_2, \dots, t_k .)

Пусть f – функционал, определенный на множестве возможных траекторий случайных процессов. *Принцип инвариантности* [1] состоит в том, что последовательность распределений случайных величин $f(\xi_n)$ сходится при $n \rightarrow \infty$ к распределению случайной величины $f(\xi)$. Сходимость по распределению обозначим символом \Rightarrow . Тогда принцип инвариантности кратко записывается так: $f(\xi_n) \Rightarrow f(\xi)$. В частности, согласно принципу инвариантности статистика Колмогорова и статистика омега-квадрат сходятся по распределению к распределениям соответствующих функционалов от случайного процесса ξ :

$$K_n = \sup_{0 \leq t \leq 1} |\xi_n(t)| \Rightarrow \sup_{0 \leq t \leq 1} |\xi(t)|, \quad \omega_n^2 = \int_0^1 \xi_n^2(t) dt \Rightarrow \int_0^1 \xi^2(t) dt.$$

Таким образом, от проблем прикладной статистики сделан переход к теории случайных процессов. Методами этой теории найдены распределения случайных величин

$$\sup_{0 \leq t \leq 1} |\xi(t)|, \quad \int_0^1 \xi^2(t) dt.$$

Принцип инвариантности – инструмент получения предельных распределений функций от результатов наблюдений, используемых в прикладной статистике.

Обоснование принципу инвариантности может быть дано на основе теории сходимости вероятностных мер в функциональных пространствах [9, 15]. Более простой подход, позволяющий к тому же получать необходимые и достаточные условия в предельной теории статистик интегрального типа (принцип инвариантности к ним нельзя применить), рассмотрен в [16].

Почему «принцип инвариантности» так назван? Обратим внимание, что предельные распределения рассматриваемых статистик не зависят от их функции распределения $F(x)$. Другими словами, предельное распределение инвариантно относительно выбора $F(x)$.

В более широком смысле термин «принцип инвариантности» применяют тогда, когда предельное распределение не зависит от тех или иных характеристик исходных распределений [1]. В этом смысле наиболее известный «принцип инвариантности» – это Центральная предельная теорема, поскольку предельное стандартное нормальное распределение – одно и то же для всех возможных распределений независимых одинаково распределенных слагаемых (лишь бы слагаемые имели конечные математическое ожидание и дисперсию).

3.3. Распределения реальных статистических данных не являются нормальными

В учебных курсах по теории вероятностей и математической статистике рассматривают различные параметрические семейства распределений числовых случайных величин. А именно, изучают семейства нормальных распределений, логарифмически нормальных, экспоненциальных, гамма-распределений, распределений Вейбулла - Гнеденко и др. Все они зависят от одного, двух или трех параметров. Поэтому для полного описания распределения достаточно знать или оценить одно, два или три числа. Очень удобно. Поэтому широко развита параметрическая теория математической статистики, в которой предполагается, что распределения результатов наблюдений принадлежат тем или иным параметрическим семействам. Эта традиция идет от Карла Пирсона, который в начале XX в. предложил использовать четырехпараметрическое семейство распределений [1]. Перечисленные выше семейства распределений - это подмножества четырехпараметрического семейства Пирсона.

К сожалению, параметрические семейства существуют лишь в головах авторов учебников по теории вероятностей и математической статистике. В реальной жизни их нет. Поэтому современная прикладная статистика [2 - 4] и эконометрика [5] используют в основном непараметрические методы [6, 7], в которых распределения результатов наблюдений могут иметь произвольный вид.

Сначала на примере нормального распределения достаточно подробно обсудим невозможность практического использования параметрических семейств для описания распределений конкретных статистических данных. Затем разберем параметрические методы отбраковки резко выделяющихся наблюдений и продемонстрируем невозможность практического использования ряда методов параметрической статистики, покажем ошибочность выводов, к которым они приводят.

3.3.1. Часто ли распределение результатов наблюдений является нормальным?

В эконометрических и экономико-математических моделях, применяемых, в частности, при изучении и оптимизации процессов маркетинга и менеджмента, управления предприятием и регионом, точности и стабильности технологических процессов, в задачах надежности, обеспечения безопасности, в том числе экологической, функционирования технических устройств и объектов, разработки организационных схем часто применяют понятия и результаты теории вероятностей и математической статистики. При этом зачастую используют те или иные параметрические семейства распределений вероятностей. Как уже отмечалось, наиболее популярно нормальное распределение. Используют также логарифмически нормальное распределение, экспоненциальное распределение, гамма-распределение, распределение Вейбулла-Гнеденко и т.д.

Очевидно, всегда необходимо проверять соответствие моделей реальности. Возникают два вопроса. Отличаются ли реальные распределения от используемых в модели? Насколько это отличие влияет на выводы?

Ниже на примере нормального распределения и основанных на нем методов отбраковки резко отличающихся наблюдений (выбросов) показано, что реальные распределения практически всегда отличаются от включенных в классические параметрические семейства, а имеющиеся отклонения от заданных семейств делают неверными выводы, в рассматриваемом случае, об отбраковке, основанные на использовании этих семейств.

Есть ли основания априори предполагать нормальность результатов измерений?

Иногда утверждают, что в случае, когда погрешность измерения (или иная случайная величина) определяется в результате совокупного действия многих малых факторов, то в силу Центральной Предельной Теоремы (ЦПТ) теории вероятностей эта величина хорошо приближается (по распределению) нормальной случайной величиной. Это утверждение, вообще говоря, неверно.

Точнее, такое утверждение справедливо, если малые факторы действуют аддитивно и независимо друг от друга. Если же они действуют мультипликативно (и независимо друг от друга), то в силу той же ЦПТ аппроксимировать распределение рассматриваемой величины надо логарифмически нормальным распределением. В прикладных задачах обосновать аддитивность, а не мультипликативность действия малых факторов обычно не удается.

Если же зависимость имеет общий характер, не приводится к аддитивному или мультипликативному виду, а также нет оснований принимать модели, дающие экспоненциальное, Вейбулла - Гнеденко, гамма или иные распределения, то о распределении итоговой случайной величины практически ничего не известно, кроме внутриматематических свойств типа регулярности.

При обработке конкретных данных иногда считают, что погрешности измерений имеют нормальное распределение. На предположении нормальности построены классические модели регрессионного, дисперсионного, факторного анализов, метрологические модели, которые еще продолжают встречаться как в отечественной нормативно-технической документации, так и в международных стандартах. На то же предположение опираются модели расчетов максимально достигаемых уровней тех или иных характеристик, применяемые при проектировании систем обеспечения безопасности функционирования экономических структур, технических устройств и объектов. Однако теоретических оснований для такого предположения нет. Необходимо экспериментально изучать распределения погрешностей.

3.3.2. Результаты экспериментов метрологов

Что же показывают результаты экспериментов? В классической монографии В. В. Налимова 1960 г. [8], посвященной применению математической статистики при анализе вещества, рассматриваемой проблеме посвящен специальный раздел "Отклонения от нормального распределения в аналитической работе" (гл. IV, параграф 4, с.122-134). Разбирается распространенное утверждение (со ссылкой на ЦПТ), что "истинное" распределение погрешностей - нормальное, а отклонения от нормальности - результат смешивания (разных генеральных совокупностей, например, серий измерений, проведенных в различных условиях). Вместе с тем приведены следующие экспериментальные данные: "В работе Клэнси [9] было изучено 250 распределений для различных аналитических методов, включающих в общей сложности 50 000 отдельных определений, и показано, что с практической точки зрения только в 10 - 15% случаев имеет место нормальное распределение" [8, с.122 - 123].

Развернутые исследования распределений погрешностей измерений проведены проф. П. В. Новицким (Ленинград) и его научной школой. Сводка, данная в монографии [10], позволяет утверждать, что в большинстве случаев распределение погрешностей измерений отличается от нормального. Так, в Машинно-электротехническом институте (г. Варна в Болгарии) было исследовано распределение погрешностей градуировки шкал аналоговых электроизмерительных приборов. Изучались приборы, изготовленные в Чехословакии, СССР и Болгарии. Согласно [10], закон распределения погрешностей оказался одним и тем же. Он имеет плотность

$$f(x) = 0,534 \exp(1 - |x|^7).$$

Были проанализированы данные о параметрах 219 фактических распределений погрешностей, исследованных разными авторами, при измерении как электрических, так и не электрических величин самыми разнообразными (электрическими) приборами. В результате этого исследования оказалось, что 111 распределений, т.е. примерно 50%, принадлежат классу распределений с плотностью

$$f(x; \alpha, b, \sigma) = \frac{\alpha}{2\lambda\sigma\Gamma(1/\alpha)} \exp\left(-\left|\frac{x-b}{\lambda\sigma}\right|^\alpha\right),$$

где α - параметр степени; b - параметр сдвига; σ - параметр масштаба; $\Gamma(\beta)$ - гамма-функция от аргумента β ;

$$\lambda = \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$$

(см. [10, с. 56]); 63 распределения, т.е. 30%, имеют плотности с плоской вершиной и пологими длинными спадами и не могут быть описаны как нормальные или, например, экспоненциальные. Оставшиеся 45 распределений оказались двухмодальными.

В другой книге известного метролога проф. П. В. Новицкого [11] приведены результаты исследования законов распределения различного рода погрешностей измерения. Он изучил распределения погрешностей электромеханических приборов на кер-нах, электронных приборов для измерения температур и усилий, цифровых приборов с ручным уравниванием. Объем выборок экспериментальных данных для каждого экземпляра составлял 100–400 отсчетов. Оказалось, что 46 из 47 распределений значи-мо отличались от нормального. Исследована форма распределения погрешностей у 25 экземпляров цифровых вольтметров Щ-1411 в 10 точках диапазона. Результаты анало-гичны. Дальнейшие сведения содержатся в монографии [10].

В лаборатории прикладной математики Тартуского государственного университе-та проанализировано 2500 выборок из архива реальных статистических данных [12]. В 92% гипотезу нормальности пришлось отвергнуть.

Приведенные описания экспериментальных данных показывают, что погрешно-сти измерений в большинстве случаев имеют распределения, отличные от нормальных. Это означает, в частности, что большинство применений критерия Стьюдента, класси-ческого регрессионного анализа и других статистических методов, основанных на нор-мальной теории, строго говоря, не является обоснованным, поскольку неверна лежащая в их основе аксиома нормальности распределений соответствующих случайных вели-чин.

Очевидно, для оправдания или обоснованного изменения существующей практи-ки анализа статистических данных требуется изучить свойства процедур анализа дан-ных при "незаконном" применении. Изучение процедур отбраковки показало, что они крайне неустойчивы к отклонениям от нормальности, а потому применять их для обра-ботки реальных данных нецелесообразно (см. ниже); поэтому нельзя утверждать, что произвольно взятая процедура устойчива к отклонениям от нормальности.

Иногда предлагают перед применением, например, критерия Стьюдента однород-ности двух выборок проверять нормальность. Хотя для этого имеется много критериев, но проверка нормальности - более сложная и трудоемкая статистическая процедура, чем проверка однородности (как с помощью статистик типа Стьюдента, так и с помо-щью непараметрических критериев). Для достаточно надежного установления нор-мальности требуется весьма большое число наблюдений. Так, чтобы гарантировать, что функция распределения результатов наблюдений отличается от некоторой нормальной не более, чем на 0,01 (при любом значении аргумента), требуется порядка 2500 наблю-дений [3]. В большинстве экономических, технических, медико-биологических и дру-гих прикладных исследований число наблюдений существенно меньше. Особенно это

справедливо для данных, используемых при изучении проблем, связанных с обеспечением безопасности функционирования экономических структур и технических объектов.

3.3.3. Скорость сходимости в Центральной предельной теореме

Иногда пытаются использовать ЦПТ для приближения распределения погрешности к нормальному, включая в технологическую схему измерительного прибора специальные сумматоры. Оценим полезность этой меры.

Пусть Z_1, Z_2, \dots, Z_k - независимые одинаково распределенные случайные величины с функцией распределения $H = H(x)$ такие, что $M(Z_1) = 0$, $D(Z_1) = 1$, $M |Z_1|^3 = \rho < +\infty$. Рассмотрим

$$w = \frac{Z_1 + Z_2 + \dots + Z_k}{\sqrt{k}}.$$

Показателем обеспечиваемой сумматором близости к нормальности является

$$C = \sup_H \sup_x |P(w < x) - \Phi(x)|.$$

Тогда

$$0,3989 \frac{\rho}{\sqrt{k}} \leq C \leq 0,7975 \frac{\rho}{\sqrt{k}}.$$

Правое неравенство в последнем соотношении вытекает из оценок константы в неравенстве Берри - Эссеена, полученном в книге [13, с.172], а левое - из примера в монографии [14, с.140-141]. Для нормального закона $\rho = 1,6$, для равномерного $\rho = 1,3$, для двухточечного $\rho = 1$ (это - нижняя граница для ρ). Следовательно, для обеспечения расстояния (в метрике Колмогорова) до нормального распределения не более 0,01 для "неудачных" распределений необходимо не менее k_0 слагаемых, где

$$0,4\sqrt{k_0} < 0,01, \quad k_0 > 1600.$$

В обычно используемых сумматорах слагаемых значительно меньше. Сужая класс возможных распределений H , можно получить, как показано в монографии [15], более быструю сходимость, но теория здесь еще не смыкается с практикой. Кроме того, не ясно, обеспечивает ли близость распределения к нормальному (в определенной метрике) также и близость распределения статистики, построенной по случайным величинам с этим распределением, к распределению статистики, соответствующей нормальным результатам наблюдений. Видимо, для каждой конкретной статистики необходимы специальные теоретические исследования, Именно к такому выводу приходит автор монографии [15]. В задачах отбраковки выбросов ответ: "Не обеспечивает" (см. ниже).

Отметим, что результат любого реального измерения записывается с помощью конечного числа десятичных знаков, обычно небольшого (2-5), так что любые реальные данные целесообразно моделировать лишь с помощью дискретных случайных величин, принимающих конечное число значений. Нормальное распределение - лишь аппроксимация реального распределения. Так, например, данные конкретного исследования, приведенные в работе [16], принимают значения от 1,0 до 2,2, т.е. всего 13 возможных значений. Из принципа Дирихле следует, что в какой-то точке построенная по данным работы [16] функция распределения отличается от ближайшей функции нормального распределения не менее чем на 1/26, т.е. на 0,04. Кроме того, очевидно, что для нормального распределения случайной величины вероятность попасть в дискретное множество десятичных чисел с заданным числом знаков после запятой равна 0.

Из сказанного выше следует, что результаты измерений и вообще статистические данные имеют свойства, приводящие к тому, что моделировать их следует случайными величинами с распределениями, более или менее отличными от нормальных. В большинстве случаев распределения существенно отличаются от нормальных, в других нормальные распределения могут, видимо, рассматриваться как некоторая аппроксимация, но никогда нет полного совпадения. Отсюда вытекает как необходимость изучения свойств классических статистических процедур в неклассических вероятностных моделях (подобно тому, как это сделано в [17] для критерия Стьюдента), так и необходимость разработки устойчивых (учитывающих наличие отклонений от нормальности) и непараметрических, в том числе свободных от распределения процедур, их широкого внедрения в практику статистической обработки данных.

Опущенные здесь рассуждения для других параметрических семейств приводят к аналогичным выводам. Итог можно сформулировать так. Распределения реальных данных практически никогда не входят в какое-либо конкретное параметрическое семейство. Реальные распределения всегда отличаются от тех, что включены в параметрические семейства. Отличия могут быть большие или маленькие, но они всегда есть. Попробуем понять, насколько важны эти различия для проведения статистического анализа данных.

3.3.4. Неустойчивость параметрических методов отбраковки резко выделяющихся результатов наблюдений

При обработке реальных статистических данных, полученных в процессе наблюдений, измерений, расчетов, иногда один или несколько результатов наблюдений резко выделяются, т.е. далеко отстоят от основной массы данных. Такие резко выделяющиеся результаты наблюдений часто считают содержащими грубые погрешности, соответственно называют промахами или выбросами. В рассматриваемых случаях возникает естественная мысль о том, что подобные наблюдения не относятся к изучаемой совокупности, поскольку содержат грубую погрешность, а получены в результате ошибки, промаха. В метрологии об этом явлении говорят так: "Грубые погрешности и промахи возникают из-за ошибок или неправильных действий оператора (его психофизиологического состояния, неверного отсчета, ошибок в записях или вычислениях, неправильного включения приборов и т.п.), а также при кратковременных резких изменений проведения измерений (вибрации, поступления холодного воздуха, толчка прибора оператором и т.п.). Если грубые погрешности и промахи обнаруживают в процессе измерений, то результаты, содержащие их, отбрасывают. Однако чаще всего их выявляют только при окончательной обработке результатов измерений с помощью специальных критериев оценки грубых погрешностей" [18, с.46-47].

Есть два подхода к обработке данных, которые могут быть искажены грубыми погрешностями и промахами:

1) отбраковка резко выделяющихся результатов наблюдений, т.е. обнаружение наблюдений, искаженных грубыми погрешностями и промахами, и исключение их из дальнейшей статистической обработки;

2) применение устойчивых (робастных) методов обработки данных, на результаты работы которых мало влияет наличие небольшого числа грубо искаженных наблюдений (см. [19 - 22] и др.).

В настоящем подразделе обсуждаются методы отбраковки.

Наиболее изучена ситуация, когда результаты наблюдений - числа x_1, x_2, \dots, x_n , резко выделяется один результат наблюдения, для определенности, максимальный x_{\max} .

Простейшая вероятностно-статистическая модель такова [23]. При нулевой гипотезе H_0 результаты наблюдения x_1, x_2, \dots, x_n рассматриваются как реализация независимых одинаково распределенных случайных величин числа X_1, X_2, \dots, X_n с функцией распределения $F(x)$. При альтернативной гипотезе H_1 случайные величины X_1, X_2, \dots, X_n также независимы, X_1, X_2, \dots, X_{n-1} имеют распределение $F(x)$, а X_n - распределение $G(x)$, оно "существенно сдвинуто вправо" относительно $F(x)$, например, $G(x)=F(x - A)$, где A достаточно велико. Если альтернативная гипотеза справедлива, то при $A \rightarrow \infty$ вероятность равенства

$$X_n = \max(X_1, X_2, \dots, X_n)$$

стремится к 1, поэтому естественно применять решающее правило следующего вида:

$$\begin{aligned} &\text{если } x_{\max} > d, \text{ то принять } H_1, \\ &\text{если } x_{\max} \leq d, \text{ то принять } H_0, \end{aligned} \quad (1)$$

где d - параметр решающего правила, значение которого следует определять из вероятностно-статистических соображений.

При справедливости нулевой гипотезы

$$P\{\max_{1 \leq i \leq n} X_i \leq d\} = \{F(d)\}^n.$$

Статистический критерий проверки гипотезы H_0 , основанный на решающем правиле вида (1), имеет уровень значимости α , если

$$P\{\max_{1 \leq i \leq n} X_i > d\} = 1 - \{F(d)\}^n = \alpha,$$

т.е.

$$F(d) = \sqrt[n]{1 - \alpha}. \quad (2)$$

Из соотношения (2) определяют граничное значение $d = d(\alpha, n)$ в решающем правиле (1).

При больших n и малых α

$$F(d) = \sqrt[n]{1 - \alpha} = 1 - \frac{\alpha}{n} + O\left(\frac{\alpha^2}{n^2}\right), \quad (3)$$

поэтому в качестве хорошего приближения к $d(\alpha, n)$ рассматривают $(1 - \alpha/n)$ - квантиль распределения $F(x)$.

Пусть правило отбраковки задано в соответствии с выражениями (1) и (2) с некоторой функцией распределения F , однако выборка берется из функции распределения G , мало отличающейся от F в смысле расстояния Колмогорова

$$\rho(F, G) = \sup_x |F(x) - G(x)| \leq \delta. \quad (4)$$

С помощью соотношения (3) получаем, что величина $\gamma = G(d)$ для d из уравнения (2) находится между $\gamma_1 = \max(0, 1 - \frac{\alpha}{n} - \delta)$ и $\gamma_2 = \min(1 - \frac{\alpha}{n} + \delta, 1)$. Уровень значимости критерия, построенного для F , при применении к наблюдениям из G есть $1 - \gamma^n$ и может принимать любые значения в отрезке $[1 - \gamma_2^n; 1 - \gamma_1^n]$. В частности, при $\delta = 0,01$, $\alpha = 0,05$, $n = 5$ возможные значения уровня значимости заполняют отрезок $[0; 0,1]$, т.е. уровень значимости может быть в 2 раза выше номинального, а если n возрастает до 30, то максимальный уровень значимости есть 0,297, т.е. почти в 6 раз выше номинального. При дальнейшем росте n верхняя граница для уровня значимости, как нетрудно видеть, приближается к 1.

Рассмотрим и другой вопрос - насколько правило отбраковки с уровнем значимости α для G может отличаться от такового для F при справедливости неравенства (4). С использованием соотношения (3) заключаем, что из

$$G(d) = 1 - \frac{\alpha}{n} \quad (5)$$

следует, что $\gamma_1 \leq F(d) \leq \gamma_2$, где γ_1 и γ_2 выписаны выше. Решение уравнения (5) может принимать любое значение в отрезке $[F^{-1}(\gamma_1); F^{-1}(\gamma_2)]$. В частности, при $\alpha = 0,05$ и $n = 5$ для стандартного нормального распределения F имеем $d(\alpha, n) = 2,319$, при $\delta = 0,01$ решение уравнения (5) может принимать любое значение в отрезке $[2,054; +\infty]$, при $\delta = 0,005$ - любое значение в отрезке $[2,170; 2,576]$.

При использовании любого другого расстояния между функциями распределения выводы о неустойчивости правил отбраковки также справедливы. Отметим, что проведенные рассуждения выполнены в рамках "общей схемы устойчивости" (см. об устойчивости статистических процедур [19 - 22] и др.).

Рассмотренные примеры показывают, что при конкретном значении $\delta = 0,01$ в неравенстве (4) весьма неустойчивы как уровни значимости при фиксированном правиле отбраковки, так и параметр d правила отбраковки при фиксированном уровне значимости. Обсудим, насколько реалистично определение функции распределения с точностью $\delta \leq 0,01$.

Есть два подхода к определению функции распределения результатов наблюдений: эвристический подбор с последующей проверкой с помощью критериев согласия и вывод из некоторой вероятностной модели.

Пусть с помощью критерия согласия Колмогорова проверяется гипотеза о том, что выборка взята из распределения F . Пусть функции распределения F и G удовлетворяют соотношению (4). Пусть на самом деле выборка взята из распределения G , а не F . При каких δ не удастся различить F и G ? Для определенности, при каких δ гипотеза согласия с F будет приниматься не менее чем в 50% случаев?

Критерий согласия Колмогорова основан на статистике

$$\lambda_n = \sqrt{n} \rho(F_n, H), \quad (6)$$

где расстояние ρ между функциями распределения определено выше в формуле (4); H - та функция распределения, согласие с которой проверяется, а F_n - эмпирическая функция распределения (т.е. $F_n(x)$ равно доле наблюдений, меньших x , в выборке объема n). Как показал А.Н. Колмогоров в 1933 г., функция распределения случайной величины λ_n при росте объема выборки n сходится к некоторой функции распределения $K(x)$, которую ныне называют функцией Колмогорова [3, 23]. При этом $K(1,36) = 0,95$ и $K(0,83) = 0,50$.

Поскольку выборка взята из распределения G , то с вероятностью 0,50

$$\rho(F_n, G) < 0,83/\sqrt{n} \quad (7)$$

(при больших n). Тогда для рассматриваемой выборки с учетом неравенства (4) и неравенства треугольника для расстояния Колмогорова и симметричности этого расстояния имеем

$$\rho(F_n, F) \leq \rho(F_n, G) + \rho(G, F) = \rho(F_n, G) + \rho(F, G) < 0,83/\sqrt{n} + \delta.$$

Если

$$0,83/\sqrt{n} + \delta \leq 1,36/\sqrt{n},$$

т.е.

$$\delta\sqrt{n} \leq 0,53, \quad (8)$$

то, согласно формуле (6), гипотеза согласия принимается по крайней мере с той же вероятностью, с которой выполнено неравенство (7), т.е. с вероятностью не менее 0,50. Для $\delta = 0,01$ это условие выполняется при $n \leq 2809$. Таким образом, для определения

функции распределения с точностью $\delta \leq 0,01$ с помощью критерия согласия Колмогорова необходимо несколько тысяч наблюдений, что для большинства прикладных задач нереально.

При втором из названных выше подходов к определению функции распределения ее конкретный вид выводится из некоторой системы аксиом, в частности, из некоторой модели порождения соответствующей случайной величины. Например, из модели суммирования вытекает нормальное распределение, а из мультипликативной модели перемножения - логарифмически нормальное распределение. Как правило, при выводе используется предельный переход. Так, из Центральной Предельной Теоремы теории вероятностей вытекает, что сумма независимых случайных величин может быть приближена нормальным распределением. Однако более детальный анализ, в частности, с помощью неравенства Берри - Эссеена (см. выше) показывает, что для гарантированного достижения точности $\delta \leq 0,01$ необходимо более полутора тысяч слагаемых. Такого количества слагаемых реально, конечно, указать почти никогда нельзя. Это означает, что при решении практических задач теория дает возможность лишь сформулировать гипотезу о виде функции распределения, а проверять ее надо с помощью анализа реальной выборки объема, как показано выше, не менее нескольких тысяч. Таким образом, в большинстве реальных ситуаций определить функцию распределения с точностью $\delta \leq 0,01$ невозможно.

Итак, показано, что правила отбраковки, основанные на использовании конкретной функции распределения, являются крайне неустойчивыми к отклонениям от нее распределения элементов выборки, а гарантировать отсутствие подобных отклонений невозможно. Поэтому *отбраковка по классическим правилам математической статистики не является научно обоснованной*, особенно при больших объемах выборок. Указанные правила целесообразно применять лишь для выявления "подозрительных" наблюдений, вопрос об отбраковке которых должен решаться из соображений соответствующей предметной области, а не из формально-математических соображений [24].

Выше для простоты изложения рассмотрен лишь случай полностью известного распределения F , для которого изучено правило отбраковки, заданное формулами (1) и (2). Аналогичные выводы о крайней неустойчивости правил отбраковки справедливы, если "истинное распределение" принадлежит какому-либо параметрическому семейству, например, нормальному, Вейбулла - Гнеденко, гамма.

Параметрическим методам отбраковки, основанным на моделях тех или иных параметрических семейств распределений, посвящены тысячи книг и статей. Приходится признать, что они имеют в основном внутриматематический интерес. При обработке реальных данных следует применять устойчивые методы (см. [19 - 22] и др.), в частности, непараметрические [25, 26].

3.4. Выборочные исследования

Термин «выборочные исследования» применяют, когда невозможно изучить все единицы представляющей интерес совокупности. Приходится знакомиться с частью совокупности - с выборкой, а затем с помощью вероятностно - статистических методов и моделей переносить выводы с выборки на всю совокупность. Выборочные исследования - способ получения статистических данных и важный раздел эконометрики и прикладной статистики [5]. Методы выборочных исследований используются при решении различных задач экономики и управления (менеджмента), в маркетинге и социологии.

3.4.1. Организация выборочных исследований

В качестве примера рассмотрим выборочные исследования предпочтений потребителей, которые часто проводят специалисты по маркетингу (изучению рынка).

Оценивание функции спроса. Функция спроса часто встречается в учебниках по экономической теории, но при этом обычно не рассказывается, как она получена. Между тем оценить ее по эмпирическим данным не так уж трудно. Например, можно выяснять ожидаемый спрос с помощью следующего простого приема - спрашиваем потенциальных потребителей: «Какую максимальную цену Вы заплатили бы за такой-то товар?» Пусть для определенности выборка состояла из 20 опрошенных. Они назвали следующие максимально допустимые для них цены:

40, 25, 30, 50, 35, 20, 50, 32, 15, 40,
20, 40, 45, 30, 50, 25, 35, 20, 35, 40.

Сначала названные опрошенными величины упорядочим в порядке возрастания. Результаты представлены в табл.1. В первом столбце - номера различных численных значений (в порядке возрастания), названных потребителями. Во втором столбце приведены сами значения цены, названные ими. В третьем столбце указано, сколько раз названо то или иное значение.

Таблица 1. – Эмпирическая оценка функции спроса и ее использование

№ п/п (i)	Цена p_i	Повторы N_i	Спрос $D(p_i)$	Прибыль $(p_i - 10)D(p_i)$	Прибыль $(p_i - 15)D(p_i)$	Прибыль $(p_i - 25)D(p_i)$
1	15	1	20	100	0	-
2	20	3	19	190	95	-
3	25	2	16	240	160	0
4	30	2	14	280	210	70
5	32	1	12	264	204	84
6	35	3	11	275	220	110
7	40	4	8	240	200	120
8	45	1	4	140	120	80
9	50	3	3	120	105	75

Таким образом, 20 потребителей назвали 9 конкретных значений цены (максимально допустимых, или приемлемых для них значений), каждое из значений, как видно из третьего столбца, названо от 1 до 4 раз. Теперь легко построить выборочную функцию спроса в зависимости от цены. Она будет представлена в четвертом столбце, который заполним снизу вверх. Спрос как функция от цены p обозначен $D(p)$ (от *demand* (англ.) – спрос). Если мы будем предлагать товар по цене свыше 50 руб., то его не купит никто из опрошенных. При цене 50 руб. появляются 3 покупателя. Записываем 3 в четвертый столбец в девятую строку. А если цену понизить до 45? Тогда товар купят четверо – тот единственный, для кого максимально возможная цена - 45, и те трое, кто был согласен на более высокую цену – 50 руб. Таким образом, легко заполнить столбец 4, действуя по правилу: значение в клетке четвертого столбца равно сумме значений в находящейся слева клетке третьего столбца и в лежащей снизу клетке четвертого столбца. Например, за 30 руб. купят товар 14 человек, а за 20 руб. - 19.

Зависимость спроса от цены - это зависимость четвертого столбца от второго. Табл.1 дает нам девять точек такой зависимости. Зависимость можно представить на рисунке, в координатах «спрос – цена». Если абсцисса - это спрос, а ордината - цена, то девять точек на кривой спроса, перечисленные в порядке возрастания абсциссы, имеют вид:

(3; 50), (4; 45), (8; 40), (11; 35), (12; 32),

(14; 30), (16; 25), (19; 20), (20; 15).

Эти девять точек можно использовать для построения кривой спроса каким-либо графическим (сделайте чертеж!) или расчетным способом, например, методом наименьших квадратов. Кривая спроса, как и следует ожидать согласно учебникам экономической теории, убывает, имея направления от левого верхнего угла чертежа к правому. Однако заметны отклонения от гладкого вида функции, связанные, в частности, с естественным пристрастием потребителей к круглым числам. Заметьте, все опрошенные, кроме одного, назвали числа, кратные 5 руб.

Расчет оптимальной цены. Данные табл.1 могут быть использованы для выбора цены продавцом-монополистом. Или организацией, действующей на рынке монополистической конкуренции. Пусть расходы на изготовление или оптовую покупку единицы товара равны 10 руб. По какой цене ее продавать на том рынке, функцию спроса для которого мы только что нашли? Для ответа на этот вопрос вычислим суммарную прибыль, т.е. произведение прибыли на одной единице товара ($p - 10$) на число проданных (точнее, запрошенных) экземпляров $D(p)$. Результаты приведены в пятом столбце табл.1. Видно, что максимальная прибыль, равная 280 руб., достигается при цене 30 руб. за единицу товара. При этом из 20 потенциальных покупателей окажутся в состоянии заплатить за книгу 14, т.е. 70% .

Если же удельные издержки производства, приходящиеся на одну единицу товара (или оптовая цена), повысятся до 15 руб., то данные столбца 6 табл.1 показывают, что максимальная прибыль, равная 220 руб. (она, разумеется, меньше, чем в предыдущем случае), достигается при более высокой цене - 35 руб. Эта цена доступна 11 потенциальным покупателям, т.е. 55% от всех возможных покупателей. При дальнейшем повышении издержек, скажем, до 25 руб., как вытекает из данных столбца 7 табл.1, максимальная прибыль, равная 120 руб., достигается при цене 40 руб. за единицу товара, что доступно 8 лицам, т.е. 40% покупателей. Отметьте, что при повышении оптовой цены на 10 руб. оказалось выгодным увеличить розничную лишь на 5, поскольку более резкое повышение привело бы к такому сокращению спроса, которое перекрыло бы эффект от повышения удельной прибыли (т.е. прибыли, приходящейся на одну проданную единицу товара).

Замечание. При более строгом подходе к использованию терминов надо вместо «прибыли» говорить о «маржинальной прибыли», а вместо «удельных издержек» – о «переменных издержках» (на одну единицу продукции), поскольку постоянные издержки не учитываем. Кроме того, спрос целесообразно выражать не в числе потребителей, а в процентах от общего числа потенциальных потребителей. Мы не сочли необходимым придерживаться подобных уточнений, поскольку цель настоящей главы – в демонстрации возможности использования в маркетинговых исследований подходов, основанных на организационно-экономическом моделировании.

Представляет интерес анализ оптимального объема выпуска при различных значениях удельных издержек (табл.2).

В табл.2 звездочками указаны максимальные значения прибыли при том или ином значении издержек, не включенном в табл.1. Для легкости обозрения результаты об оптимальных объемах выпуска и соответствующих ценах из табл.1 и табл.2 приведены в табл.3.

Как видно из табл.3, с ростом издержек оптимальный выпуск падает, а цена растет. При этом изменение издержек на 5 единиц может вызывать, а может и не вызывать повышения цены. В этом проявляется микроструктура функции спроса – небольшое повышение цены может привести к тому, что значительные группы покупателей откажутся от покупок, и прибыль упадет.

Таблица 2. – Прибыль при различных значениях издержек

№ (i)	Цена p_i	Спрос $D(p_i)$	Прибыль $(p_i - 5)D(p_i)$	Прибыль $(p_i - 20)D(p_i)$	Прибыль $(p_i - 30)D(p_i)$	Прибыль $(p_i - 35)D(p_i)$	Прибыль $(p_i - 40)D(p_i)$
1	15	20	200	-	-	-	-
2	20	19	285	0	-	-	-
3	25	16	320	80	-	-	-
4	30	14	350 *	140	0	-	-
5	32	12	324	144	24	-	-
6	35	11	330	165 *	55	0	-
7	40	8	280	160	80 *	40	0
8	45	4	160	100	60	40	20
9	50	3	135	90	60	45 *	30 *

Таблица 3. – Зависимость оптимального выпуска и цены от издержек

Издержки	5	10	15	20	25	30	35	40
Оптимальный выпуск	14	14	11	11	8	8	3	3
Цена	30	30	35	35	40	40	50	50

Этот эффект напоминает известное в экономической теории разделение налогового бремени между производителем и потребителем. Неверно говорить, что производитель перекладывает издержки или, конкретно, налоги, на потребителя, повышая цену на их величину, поскольку при этом сокращается спрос (и выпуск), а потому и прибыль производителя.

Дальнейшее ясно - если оптовая цена будет повышаться, то и дающая максимальную прибыль розничная цена также будет повышаться, и все меньшая доля покупателей сможет приобрести товар. Крайняя точка - оптовая цена, равная 45 руб. Тогда только трое (15%) купят товар за 50 руб., а прибыль продавца составит только 15 руб. Наглядно видно, что повышение издержек производства приводит к ориентации производителя на наиболее богатые слои населения. Но и повышение цен (до оптимального для монополиста-производителя уровня) не приводит к повышению прибыли, напротив, она снижается, и при этом большинство потенциальных потребителей не в состоянии купить товар.

Отметим, что рыночные структуры не в состоянии обеспечить всех желающих – это просто не выгодно. Так, из 20 опрошенных лишь 14, т.е. 70%, могут рассчитывать на покупку, даже при минимальных издержках и ценах. Если общество желает чем-либо обеспечить всех граждан, оно должно раздавать это благо бесплатно, как это делается, например, с учебниками в школах.

Описанный здесь метод оценивания спроса был разработан в Институте высоких статистических технологий и эконометрики (Москва) в 1993 г.

Для изучения предпочтений потребителей часто используют более изощренные методы. Рассмотрим некоторые из них.

Маркетинговые опросы потребителей. Потенциального покупателя интересует не только цена, но и качество товара, красота упаковки (например, для подарочных наборов конфет) и многое другое. Хочешь узнать, чего желает потребитель - спроси его. Эта простая мысль объясняет популярность маркетинговых опросов.

Бесспорно, что основная цель производственной и торговой деятельности - удовлетворение потребностей людей. Как получить представление об этих потребностях? Очевидно, необходимо опросить потребителей. В американском учебнике по рекламному делу [6] подробно рассматриваются различные методы опроса потребителей и обработки результатов с помощью методов эконометрики. Расскажем о результатах опро-

са потребителей растворимого кофе. Исследование проведено Институтом высоких статистических технологий и эконометрики по заказу АОЗТ «Д-2» в апреле 1994 г. в Москве.

Сбор данных. Один из важнейших разделов прикладной статистики – сбор данных. Обсудим постановку задачи в случае опроса потребителей растворимого кофе. Заказчика интересуют предпочтения как продавцов кофе (розничных и мелкооптовых), так и непосредственно потребителей. В результате совместного обсуждения было признано целесообразным использовать для опроса и тех, и других одну и ту же анкету из 14 основных и 4 социально-демографических вопросов с добавлением двух вопросов специально для продавцов. Анкета была разработана совместно представителями заказчика и исполнителя и утверждена заказчиком. В табл.4 приведен несколько сокращенный вариант этой анкеты.

Таблица 4 – Анкета для потребителей растворимого кофе (в сокращении)

Дорогой потребитель растворимого кофе,

Институт высоких статистических технологий и эконометрики просит Вас ответить на несколько простых вопросов о том, какой кофе Вы любите. Ваши ответы позволят составить объективное представление о вкусах российских любителей кофе и будут способствовать повышению качества этого товара на российском рынке.

1. Часто ли Вы пьете растворимый кофе: иногда, каждый день 1 чашку, 2-3 чашки, больше, чем 3 чашки.

(Здесь и далее подчеркните нужное.)

2. Что Вы цените в кофе: вкус, аромат, крепость, цвет, отсутствие вредных для здоровья веществ, что-либо еще (сообщите нам, что именно)

3. Как часто покупаете кофе: по мере надобности или по возможности?

4. Какую марку растворимого кофе Вы обычно покупаете? _____

5. Какой объем упаковки Вы предпочитаете: в пакетиках, маленькая банка, средняя банка, большая банка, обязательно стеклянная банка, все равно.

6. Где покупаете растворимый кофе: в ларьках, в продуктовых магазинах, в специализированных отделах и магазинах, все равно, где купить, где-либо еще (опишите, пожалуйста) _____.

7. Были ли случаи, когда купленный Вами кофе оказывался низкого качества? Да, нет.

8. Согласны ли Вы, что за высокое и гарантированное качество продукта можно и заплатить несколько дороже? Да, нет.

9. На сколько дороже Вы готовы платить за экологически безопасный кофе?

10. Считаете ли Вы нужным, чтобы вредные для здоровья вещества, в частности, ионы тяжелых металлов, не проникали из материала упаковки в растворимый кофе? Да, нет.

Мы планируем сравнить потребительские предпочтения различных категорий жителей нашей страны. Поэтому просим ответить еще на несколько вопросов.

11. Пол: женский, мужской.

12. Возраст: до 20, 20-30, 30-50, более 50.

13. Род занятий: учащийся, работающий, пенсионер, инженер, врач, преподаватель, служащий, менеджер, предприниматель, научный работник, рабочий, др. (пожалуйста, расшифруйте).

14. Вся Ваша семья любит растворимый кофе или же Вы - единственный любитель этого восхитительного напитка современного человека? Вся семья, я один (одна).

Спасибо за Ваше содействие работе по повышению качества продуктов на российском рынке!

Выбор метода опроса. Широко применяются процедуры опроса, когда респонденты (так социологи и маркетологи называют тех, от кого получают информацию, т.е. опрашиваемых) самостоятельно заполняют анкеты (розданные им или полученные по почте), а также личные и телефонные интервью. Из этих процедур нами было выбрано личное интервью по следующим причинам.

Возврат почтовых анкет сравнительно невелик (в данном случае можно было ожидать не более 5-10%), оттянут по времени и искажает структуру совокупности потребителей (наиболее динамичные люди вряд ли найдут время для ответа на подобную анкету).

Самостоятельное заполнение анкеты, как показали специально проведенные эксперименты, не позволяет получить полные ответы на поставленные вопросы. Респондент утомляется или отвлекается, отказывается отвечать на часть вопросов, иногда не понимает их или отвечает не по существу. Некоторые категории респондентов, например, продавцы в киосках, отказываются заполнять анкеты, но готовы устно ответить на вопросы.

Телефонный опрос искажает совокупность потребителей, поскольку наиболее активных индивидуумов трудно застать дома и уговорить ответить на вопросы анкеты. Репрезентативность нарушается также и потому, что на один номер телефона может приходиться различное количество продавцов и потребителей растворимого кофе, а некоторые из них не имеют телефонов вообще. Анкета достаточно длинна, и разговор по домашнему и тем более служебному телефону респондента может быть прекращен досрочно по его инициативе. Иногородних продавцов и потребителей растворимого кофе, приехавших в Москву, по телефону опросить практически невозможно.

Метод личного интервью лишен перечисленных недостатков. Соответствующим образом подготовленный интервьюер, получив согласие на интервью, удерживает внимание собеседника на анкете, добивается получения ответов на все её вопросы, контролируя при этом соответствие ответов реальной позиции респондента. Ясно, что успех интервьюирования зависит от личных качеств и подготовки интервьюера. Однако расходы на получение одной анкеты при использовании этого метода больше, чем для других рассмотренных методов.

Формулировки вопросов. В маркетинговых и социологических опросах используют три типа вопросов - закрытые, открытые и полузакрытые, они же полуоткрытые. При ответе на закрытые вопросы респондент может выбирать лишь из сформулированных составителями анкеты вариантов ответа. В качестве ответа на открытые вопросы респондента просят изложить свое мнение в свободной форме. Полузакрытые, они же полуоткрытые вопросы занимают промежуточное положение - кроме перечисленных в анкете вариантов, респондент может добавить свои соображения.

В социологических публикациях, посвященных выборочным исследованиям, продолжается дискуссия по поводу «мягких» и «жестких» форм сбора данных. Т.е. фактически о том, какого типа вопросы более целесообразно использовать - открытые или закрытые (см., например, статью известного социолога В. А. Ядова [7]).

Преимущество открытых вопросов состоит в том, что респондент может свободно высказать свое мнение так, как сочтет нужным. Их недостаток - в сложности сопоставления мнений различных респондентов. Для такого сопоставления и получения свод-

ных характеристик организаторы опроса вынуждены сами шифровать ответы на открытые вопросы, применяя разработанную ими схему шифровки.

Преимущество закрытых вопросов в том и состоит, что такую шифровку проводит сам респондент. Однако при этом организаторы опроса уподобляются древнегреческому мифическому персонажу Прокрусту. Как известно, Прокруст приглашал путников заночевать у него. Укладывал их на кровать. Если путник был маленького роста, он вытягивал его ноги так, чтобы они и голова доставали до концов кровати. Если же путник оказывался высоким и ноги его торчали - он обрубал их так, чтобы достигнуть стандарта: «рост» путника должен равняться длине кровати. Так и организаторы опроса, применяя закрытые вопросы, заставляют респондента «вытягивать» или «обрубать» свое мнение, чтобы выразить его с помощью приведенных в формулировке вопроса возможных ответов.

Ясно, что для обработки данных по группам и сравнения групп между собой нужны формализованные данные, и фактически речь может идти лишь о том, кто - респондент или маркетолог (социолог, психолог и др.) - будет шифровать ответы. В проекте «Потребители растворимого кофе» практически для всех вопросов варианты ответов можно перечислить заранее, т.е. можно широко использовать закрытые вопросы. В отличие от опросов с вопросами типа: «Одобряете ли Вы идущие в России реформы?», в которых естественно просить респондента расшифровать, что он понимает под «реформами» (открытый вопрос). Поэтому в используемой в описываемом проекте анкете использовались в основном закрытые и полужакрытые вопросы. Как показали результаты обработки, этот подход оказался правильным - лишь в небольшом числе анкет оказались вписаны свои варианты ответов. Вместе с тем демонстрировалось уважение к мнению респондента, не выдвигалось требование обязательного выбора из заданного множества ответов - респондент мог добавить свое, но редко пользовался этой возможностью (не более чем в 5% случаев).

В последнем вопросе анкеты респонденту предлагалось стать постоянным участником опросов о качестве товаров народного потребления. Ряд респондентов откликнулся на это предложение, в результате стало возможным развертывание постоянной сети «экспертов по качеству», подобной аналогичным в США и других странах.

3.4.2. Модели случайных выборок

Статистические методы выборочных исследований основаны на вероятностных моделях, описывающих получение ответов опрашиваемых на вопросы анкет. В случае ответов типа «да» - «нет» наиболее распространенными являются две вероятностные модели—биномиальная и гипергеометрическая.

В биномиальной модели предполагается, что ответы n опрашиваемых можно рассматривать как совокупность n независимых одинаково распределенных случайных величин X_1, X_2, \dots, X_n , где $X_i = 1$, если i -ый респондент сказал «да», и $X_i = 0$, если его ответ - «нет». Тогда число X ответов «да» в выборке равно

$$X = X_1 + X_2 + \dots + X_n. \quad (1)$$

Из формулы (1) и Центральной предельной теоремы теории вероятностей (см. раздел 3.2 выше) вытекает, что при увеличении объема выборки n распределение X сближается с нормальным распределением. Известно, что распределение X имеет вид

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad (2)$$

где C_n^k - число сочетаний из n элементов по k , а p - доля ответов «да» в генеральной совокупности, т.е. $p = P(X_i = 1)$. Формула (2) задает биномиальное распределение, часто используемое при вероятностном моделировании реальных явлений и процессов.

Гипергеометрическое распределение соответствует иной схеме - случайному отбору респондентов в выборку. Пусть среди N лиц, составляющих генеральную совокупность, имеется D лиц, чье мнение - «да». Случайность отбора респондентов в выборку означает, что каждое лицо имеет одинаковые шансы быть отобранным. Мало того, ни одна пара потенциальных респондентов не должна иметь при отборе в выборку преимущества перед любой другой парой. То же самое — для троек, четверок и т.д. Это условие выполнено тогда и только тогда, когда каждое из C_N^n сочетаний по n лиц из N имеет одинаковые шансы быть отобранным в качестве выборки. Вероятность того, что будет отобрано заранее заданное сочетание, равна, очевидно, $1/C_N^n$.

Пусть Y — число сказавших «да» лиц в случайной выборке, организованной таким образом. Известно, что тогда $P(Y = k)$ – гипергеометрическое распределение, т.е.

$$P(Y = k) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}. \quad (3)$$

Отбор случайной выборки согласно описанным правилам организуют при проведении различных лотерей. Например, отбирают 6 номеров из 49. Тогда генеральная совокупность состоит из 49 единиц (номеров), а выборка - из 6. В этом случае отбирают номера, а не респондентов, но вероятностная модель - та же. Удобно говорить, что генеральная совокупность и выборка состоят из единиц. В одном случае единицы - это люди (лица, потенциальные респонденты), в другом - номера. В статистических метода управления качеством рассматриваются единицы продукции - детали или изделия.

Замечательный математический факт состоит в том, что биномиальная модель (2) и гипергеометрическая модель (3) *весьма близки* (с практической точки зрения совпадают), когда объем генеральной совокупности (партии) по крайней мере в 10 раз превышает объем выборки. Другими словами, можно принять, что

$$P(X = k) = P(Y = k), \quad (4)$$

если объем выборки мал по сравнению с объемом партии. При этом в качестве p в левой части формулы (4) берут D/N .

Близость результатов, получаемых с помощью биномиальной и гипергеометрической моделей, весьма важна не только с практической, но и с методологической точки зрения. Дело в том, что эти модели исходят из принципиально различных методологических предпосылок. В биномиальной модели случайность *присуща каждому респонденту*. Он с какой-то вероятностью отвечает «да», а с какой-то - «нет» (сумма этих вероятностей, очевидно, равна 1). В то же время в гипергеометрической модели ответ респондента полностью определен, а случайность проявляется лишь *в отборе*, вносится социологом или маркетологом при составлении выборки.

В науках о человеке противоречие между рассматриваемыми моделями выборки четко выражено. В среде специалистов, изучающих человека (маркетологов, социологов, психологов, политологов и др.) давно идет дискуссия о роли случайности в поведении человека. А именно, о том, есть ли случайность в поведении отдельно взятого человека или же случайность проявляется лишь в отборе выборки из генеральной совокупности.

Биномиальная модель предполагает, что поведение человека, в частности, выбор им определенного варианта при ответе на вопрос, определяется с участием случайных причин. Например, человек может случайно сказать «да», случайно — «нет». Некоторые философы отрицают случайность, присущую поведению человека согласно биномиальной модели. Они верят в причинность и считают поведение конкретного человека практически полностью детерминированным (его взглядами, психофизиологическими особенностями, прежним опытом и др.). Поэтому они принимают гипергеометрическую модель и считают, что случайность отличия ответов в выборке от ответов во всей

генеральной совокупности определяется всецело случайностью, вносимой при отборе единиц наблюдения в выборку.

Сформулированные выше математические результаты (соотношение (4)) показывают, что позиция в этой давней дискуссии практически не влияет на алгоритмы обработки данных. Следовательно, во многих случаях нет необходимости принимать чью-либо сторону в этом споре, поскольку обе модели дают близкие численные результаты.

Отличия проявляются лишь при обсуждении вопроса о том, какую выборку считать *представительной*. В терминах контроля качества продукции - является ли такой выборка, составленная из 20 изделий, лежащих сверху в первом вскрытом ящике? В биномиальной модели вполне допустим ответ «да», в гипергеометрической - только «нет».

Биномиальная модель легче для теоретического изучения, поэтому будем её рассматривать в дальнейшем. Однако при реальном опросе лучше формировать выборку, исходя из гипергеометрической модели. Это делают, выбирая респондентов из списка избирателей (для включения в выборку) с помощью датчиков псевдослучайных чисел на ЭВМ или с помощью таблиц псевдослучайных чисел. Алгоритмы формирования выборки встраивают во все современные программные продукты, предназначенные для поддержки проведения маркетинговых или социологических опросов, организации статистического контроля качества и др.

Обоснование объема выборки и проведение опроса. Вернемся к анализу результатов опроса потребителей растворимого кофе, о котором шла речь в предыдущем разделе. Как уже говорилось, модели выборочных исследований часто опираются на предположение о том, что реальную выборку можно описывать как «случайную выборку из конечной совокупности». Типа той, когда из списков избирателей с помощью датчика случайных чисел отбирается необходимое число номеров для формирования жюри присяжных заседателей. В рассматриваемом исследовании нельзя обеспечить формирование подобной выборки - не существует реестра потребителей растворимого кофе. Однако в этом и нет необходимости. Поскольку гипергеометрическое распределение хорошо приближается биномиальным, если объем выборки по крайней мере в 10 раз меньше объема всей совокупности (в рассматриваемом случае это так), то правомерно использование биномиальной модели, согласно которой мнение респондента (ответы на все вопросы анкеты) рассматривается как случайный вектор, а все такие вектора независимы между собой. Другими словами, можно использовать модель простой случайной выборки.

3.4.3. Доверительное оценивание доли

Зачем проводятся выборочные исследования? Чтобы получить необходимую информацию о генеральной совокупности. Для этого необходимо перенести выводы с выборки на генеральную совокупность. Как и с какой точностью можно это сделать?

Рассмотрим эту проблему для простейшего случая одного вопроса с двумя возможными ответами - «да» и «нет».

Напомним, что биномиальная модель выборки как раз и применяется для описания ответов на закрытые вопросы, имеющие две подсказки, например, «да» и «нет». Конечно, пары подсказок могут быть иными. Например, «согласен» и «не согласен». Или при опросе потребителей кондитерских товаров первая подсказка может иметь такой вид: «Больше люблю «Марс», чем «Сникерс»». А вторая тогда такова: «Больше люблю «Сникерс», чем «Марс»».

Пусть объем выборки равен n . Тогда ответы опрашиваемых можно представить как X_1, X_2, \dots, X_n , где $X_i = 1$, если i -й респондент выбрал первую подсказку, и $X_i = 0$, если

i -й респондент выбрал вторую подсказку, $i=1,2,\dots,n$. В вероятностной модели предполагается, что случайные величины X_1, X_2, \dots, X_n независимы и одинаково распределены. Поскольку эти случайные величины принимают два значения, то ситуация описывается одним параметром p - долей выбирающих первую подсказку во всей генеральной совокупности. Тогда

$$P(X_i = 1) = p, P(X_i = 0) = 1 - p, i=1,2,\dots,n.$$

Пусть $m = X_1 + X_2 + \dots + X_n$. Оценкой вероятности p является частота $p^* = m/n$. При этом математическое ожидание $M(p^*)$ и дисперсия $D(p^*)$ имеют вид

$$M(p^*) = p, D(p^*) = p(1-p)/n.$$

По Закону Больших Чисел (ЗБЧ) теории вероятностей (в данном случае - по теореме Бернулли) частота p^* сходится (т.е. безгранично приближается) к вероятности p при росте объема выборки (см. разд. 3.2 выше). Это означает, что оценивание проводится тем точнее, чем больше объем выборки. Точность оценивания можно указать. Займемся этим.

По теореме Муавра-Лапласа теории вероятностей

$$\lim_{n \rightarrow \infty} P \left\{ \frac{m - np}{\sqrt{np(1-p)}} \leq x \right\} = \Phi(x),$$

где $\Phi(x)$ - функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy,$$

где $\pi = 3,1415925\dots$ - отношение длины окружности к ее диаметру, $e = 2,718281828\dots$ - основание натуральных логарифмов. График плотности стандартного нормального распределения

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

был очень точно изображен на германской денежной банкноте в 10 немецких марок (до введения евро). Банкнота была посвящена великому немецкому математику Карлу Гауссу (1777–1855), среди основных работ которого есть относящиеся к нормальному распределению. Эта подробность демонстрирует, что в Германии (и тем более в англосаксонских странах) гораздо шире распространено знакомство с основами теории вероятностей и математической статистики, чем в нашей стране.

В настоящее время нет необходимости вычислять функцию стандартного нормального распределения и ее плотность по приведенным выше формулам, поскольку давно составлены подробные таблицы (см., например, [1]), а распространенные программные продукты содержат алгоритмы нахождения этих функций.

С помощью теоремы Муавра-Лапласа могут быть построены доверительные интервалы для неизвестной статистике вероятности. Сначала заметим, что из этой теоремы непосредственно следует, что

$$\lim_{n \rightarrow \infty} P \left\{ -x \leq \frac{m - np}{\sqrt{np(1-p)}} \leq x \right\} = \Phi(x) - \Phi(-x).$$

Поскольку функция стандартного нормального распределения симметрична относительно 0, т.е. $\Phi(x) + \Phi(-x) = 1$, то справедливо полезное равенство $\Phi(x) - \Phi(-x) = 2\Phi(x) - 1$.

Зададим характеристику надежности переноса выводов с выборки на генеральную совокупность - доверительную вероятность γ , близкую к 1. Пусть функция $U(\gamma)$ удовлетворяет условию

$$\Phi(U(\gamma)) - \Phi(-U(\gamma)) = \gamma,$$

т.е.

$$U(\gamma) = \Phi^{-1}\left(\frac{1+\gamma}{2}\right).$$

Из последнего предельного соотношения следует, что

$$\lim_{n \rightarrow \infty} P \left\{ p^* - U(\gamma) \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq p^* + U(\gamma) \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right\} = \gamma.$$

К сожалению, это соотношение нельзя непосредственно использовать для доверительного оценивания, поскольку верхняя и нижняя границы зависят от неизвестной вероятности. Однако с помощью метода наследования сходимости (см. раздел 3.2 выше или [3, п.2.4]) можно доказать, что

$$\lim_{n \rightarrow \infty} P \left\{ p^* - U(\gamma) \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}} \leq p \leq p^* + U(\gamma) \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}} \right\} = \gamma.$$

Следовательно, нижняя доверительная граница имеет вид

$$p_{\text{нижн}} = p^* - U(\gamma) \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}},$$

в то время как верхняя доверительная граница такова:

$$p_{\text{верх}} = p^* + U(\gamma) \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}}.$$

Наиболее распространенным (в прикладных исследованиях) значением доверительной вероятности является $\gamma = 0,95$. Иногда употребляют термин «95% доверительный интервал». Тогда $U(\gamma) = 1,96$.

Пример 1. Пусть $n = 500$, $m = 200$. Тогда $p^* = 0,40$. Найдем доверительный интервал для $\gamma = 0,95$:

$$p_{\text{нижн}} = 0,40 - 1,96 \frac{\sqrt{0,4 \times 0,6}}{\sqrt{500}} = 0,40 - 0,043 = 0,357, \quad p_{\text{верх}} = 0,40 + 0,043 = 0,443.$$

Таким образом, хотя в достаточно большой выборке 40% респондентов говорят «да», можно утверждать лишь, что во всей генеральной совокупности таких от 35,7% до 44,3% - крайние значения отличаются на 8,6%.

Замечание. С достаточной для практики точностью можно заменить 1,96 на 2.

Величина

$$U(\gamma) \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}}$$

называется ошибкой выборки. Обычно, как в примере 1, используют значение доверительной вероятности $\gamma = 0,95$ и множитель $U(\gamma) = 1,96$.

Удобные для использования в практической работе специалиста по выборочным исследованиям, маркетолога и социолога таблицы точности оценивания разработаны во ВЦИОМ (Всероссийском центре по изучению общественного мнения). Приведем здесь несколько модифицированный вариант одной из них (табл.5).

Таблица 5. – Допустимая величина ошибки выборки (в процентах)

Объем группы n	1000	750	600	400	200	100
Доля p^*						
Около 10% или 90%	2	3	3	4	5	7
Около 20% или 80%	3	4	4	5	7	9
Около 30% или 70%	4	4	4	6	9	10
Около 40% или 60%	4	4	5	6	8	11
Около 50%	4	4	5	6	8	11

В условиях рассмотренного выше примера надо взять вторую снизу строку. Объем выборки 500 нет в таблице, но есть объемы 400 и 600, которым соответствуют ошибки в 6% и 5% соответственно. Следовательно, в условиях примера целесообразно оценить ошибку как $((5+6)/2)\% = 5,5\%$. Эта величина несколько больше, чем рассчитанная выше (4,3%). С чем связано это различие? Дело в том, что таблица ВЦИОМ связана не с доверительной вероятностью $\gamma = 0,95$, а с доверительной вероятностью $\gamma = 0,99$, которой соответствует множитель $U(\gamma) = 2,58$. Расчет ошибки по приведенным выше формулам дает 5,65%, что практически совпадает со значением, найденным по табл.5.

Необходимый объем выборки. В биномиальной модели выборки оценивание характеристик происходит тем точнее, чем объем выборки больше. Часто спрашивают: «Какой объем выборки нужен?» Разработан ряд методов определения необходимого объема выборки. Они основаны на разных подходах. Либо на задании необходимой точности оценивания параметров. Либо на явной формулировке альтернативных гипотез, между которыми необходимо сделать выбор. Либо на учете погрешностей измерений (методы статистики интервальных данных). Ни один из этих подходов нельзя применить в рассматриваемом случае.

Минимальный из обычно используемых объемов выборки n в маркетинговых или социологических исследованиях - 100, максимальный - до 5000 (обычно в исследованиях, охватывающих ряд регионов страны, т.е. фактически разбивающихся на ряд отдельных исследований - как в ряде исследований ВЦИОМ). По данным Института социологии Российской академии наук [2], среднее число анкет в социологическом исследовании не превышает 700. Поскольку стоимость исследования растет, по крайней мере, как линейная функция объема выборки, а точность повышается как квадратный корень из этого объема, то верхняя граница объема выборки определяется обычно из экономических соображений. Объемы пилотных исследований (т.е. проводящихся впервые, предварительно или как первые в сериях подобных) обычно ниже, чем объемы исследований по обкатанной программе.

Нижняя граница определяется тем, что в минимальной по численности анализируемой подгруппе должно быть несколько десятков человек (не менее 30), поскольку по ответам попавших в эту подгруппу необходимо сделать обоснованные заключения, например, о предпочтениях соответствующей подгруппы в совокупности всех потребителей растворимого кофе. Учитывая деление опрашиваемых на продавцов и покупателей, на мужчин и женщин, на четыре градации по возрасту и восемь - по роду занятий, наличие 5 - 6 подсказок во многих вопросах, приходим к выводу о том, что в рассматриваемом проекте объем выборки должен быть не менее 400 - 500. Вместе с тем существенное превышение этого объема было признано нецелесообразным, поскольку исследование являлось пилотным.

Поэтому в проекте «Потребители растворимого кофе» объем выборки был выбран равным 500. Анализ полученных результатов позволяет утверждать, что в соответствии с целями исследования выборка следует считать репрезентативной.

3.4.4. Два прикладных выборочных исследования

Продолжим обсуждение выборочного исследования потребителей растворимого кофе.

Организация опроса. Интервьюерами работали молодые люди – студенты первого курса экономико-математического факультета Московского государственного института электроники и математики (технического университета) и лица № 1140, проходившие обучение по экономике, всего 40 человек, имеющих специальную подготовку по изучению рынка и проведению маркетинговых опросов потребителей и продавцов (в объеме 8 часов). Опрос продавцов проводился на рынках г. Москвы, действующих в Лужниках, у Киевского вокзала и в других местах. Опрос покупателей проводился на рынках, в магазинах, на улицах около киосков и ларьков, а также в домашней и служебной обстановке.

Большое внимание уделялось качеству заполнения анкет. Интервьюеры были разбиты на шесть бригад, бригадиры персонально отвечали за качество заполнения анкет. Второй уровень контроля осуществляла специально созданная «группа организации опроса», третий происходил при вводе информации в базу данных. Каждая анкета заверена подписями интервьюера и бригадира, на ней указано место и время интервьюирования. Поэтому необходимо признать высокую достоверность собранных анкет.

Обработка данных. В соответствии с целью исследования основной метод первичной обработки данных - построение частотных таблиц для ответов на отдельные вопросы. Кроме того, проводилось сравнение различных групп потребителей и продавцов, выделенных по социально-демографическим данным, с помощью критериев проверки однородности выборок (см. ниже). При более углубленном анализе применялись различные методы статистики объектов нечисловой природы (более 90% маркетинговых и социологических данных имеют нечисловую природу [4]). Использовались средства графического представления данных.

Итоги опроса. Итак, по заданию одной из торговых фирм были изучены предпочтения покупателей и мелкооптовых продавцов растворимого кофе. Совместно с представителями заказчика был составлен опросный лист (анкета типа социологической) из 16 основных вопросов и 4 дополнительных, посвященных социально-демографической информации. Опрос проводился в форме интервью с 500 покупателями и продавцами кофе. Места опроса - рынки, лотки, киоски, продуктовые и специализированные магазины. Другими словами, были охвачены все виды мест продаж кофе. Интервью проводили более 40 специально подготовленных (примерно по 8-часовой программе) студентов и лицестов, разбитых на 7 бригад. После тщательной проверки бригадирами и группой обработки информация была введена в специально созданную базу данных. Затем проводилась разнообразная статистическая обработка, строились таблицы и диаграммы, проверялись статистические гипотезы и т.д. Заключительный этап - осмысление и интерпретация данных, подготовка итогового отчета и предложений для заказчиков.

Технология организации и проведения маркетинговых опросов лишь незначительно отличается от технологии социологических опросов, многократно описанной в литературе. Так, мы предпочли использовать полуоткрытые вопросы, в которых для опрашиваемого дан перечень подсказок, а при желании он может высказать свое мнение в свободной форме. Не уложившихся в подсказки оказалось около 5 %, их мнения были внесены в базу данных и анализировались дополнительно. Для повышения надежности опроса о наиболее важных с точки зрения маркетинга моментах спрашивалось в нескольких вопросах. Были вопросы - ловушки, с помощью которых контролировалась «осмысленность» заполнения анкеты. Например, в вопросе: «Что Вы цените в

кофе: вкус, аромат, крепость, наличие пенки...» ловушкой является включение «крепости» - ясно, что крепость зависит не от кофе самого по себе, а от его количества в чашке. В ловушку никто из 500 не попался - никто не отметил «крепость». Этот факт свидетельствует о надежности выводов проведенного опроса. Мы считали нецелесообразным задавать вопрос об уровне доходов (поскольку в большинстве случаев отвечают «средний», что невозможно связать с определенной величиной). Вместо такого вопроса мы спрашивали: «Как часто Вы покупаете кофе: по мере надобности или по возможности?». Поскольку кофе не является дефицитным товаром, первый ответ свидетельствовал о наличии достаточных денежных средств, второй - об их ограниченности (потребитель не всегда имел возможность позволить себе купить банку растворимого кофе).

Стоимость подобных исследований - 5–10 долларов США на одного обследованного. При этом трудоемкость (и стоимость) начальной стадии - подготовки анкеты и интервьюеров, пробный опрос и др. - 30% от стоимости исследования. Стоимость непосредственно опроса - тоже 30%, ввод информации в компьютер и проведение расчетов, построение таблиц и графиков - 20%, интерпретация результатов, подготовка итогового отчета и предложений для заказчиков - 20%. Таким образом, стоимость собственно опроса в два с лишним раза меньше стоимости остальных стадий исследования. И в выполнении работы участвуют различные специалисты. На первой стадии – в основном нужны высококвалифицированные аналитики. На второй – многочисленные интервьюеры, в роли которых могут выступать студенты и школьники, прошедшие конкретный курс обучения в 8–10 часов. На третьей – работа с компьютером (надо уметь строить и обчислять электронные таблицы или базы данных, использовать статистические пакеты, составлять и печатать таблицы и диаграммы и т.п.). На четвертой – опять в основном нужны высококвалифицированные аналитики.

Приведем некоторые из полученных результатов.

а) В отличие от западных потребителей, отечественные не отдавали предпочтения стеклянным банкам по сравнению с жестяными. Поскольку жестяные банки дешевле стеклянных, то можно было порекомендовать (в 1994 г., когда проходил опрос) с целью снижения расходов закупку кофе в жестяных банках.

б) Отечественные потребители готовы платить на 10–20% больше за экологически безопасный кофе более высокого качества, имеющий сертификат Минздрава и символ экологической безопасности на упаковке.

в) Средний объем потребления растворимого кофе одной семьей - 850 г в месяц.

г) Потребители растворимого кофе могут быть разделены на классы (в другой терминологии - кластеры). Есть «продвинутые» потребители, обращающие большое внимание на качество и экологическую безопасность, марку и страну производства, терпимо относящиеся к изменению цены. Эти «тонкие ценители» - в основном женщины от 30 до 50 лет, служащие, менеджеры, научные работники, преподаватели, врачи (т.е. лица с высшим образованием), пьющие кофе как дома, так и на работе, причем «кофейный ритуал» зачастую входит в процедуру деловых переговоров или совещаний. Противоположный по потребительскому поведению класс состоит из мужчин двух крайних возрастных групп - школьников и пенсионеров. Для них важна только цена, что очевидным образом объясняется недостатком денег.

Результаты были использованы заказчиком в рекламной кампании. В частности, в ней в соответствии с итогами опроса обращалось внимание на сертификат Минздрава и на экологическую безопасность упаковки.

Оценивание функции спроса и моделирование рынка. Выпускник программы «Топ-менеджер» Академии народного хозяйства при Правительстве Российской Федерации А. А. Пивень в 2003 г. оценил функцию спроса на продукцию своего предприятия. Расчет и установление оптимальной цены на изделие с точки зрения максимизации

ции прибыли был произведен по описанному выше методу. В табл.6 приведена функция ожидаемого спроса в зависимости от цены. Как подсчитал А.А. Пивень, уровень издержек на производство 1 изделия составляет 42824,7 руб. (1350 у.е.). Для удобства все расчеты будем производить в условных единицах.

Таблица 6. – Функция ожидаемого спроса в зависимости от цены

№ п/п	Цена, у.е.	Объем продаж в год, шт.	Издержки на объем производства	Выручка, у.е.	Прибыль, у.е.
1	1 400	1 600	2 160 000	2 240 000	80 000
2	1 500	1 500	2 025 000	2 250 000	225 000
3	1 600	1 200	1 620 000	1 920 000	300 000
4	1 700	1 000	1 350 000	1 700 000	350 000
5	1 800	720	972 000	1 246 000	324 000
6	1 900	500	675 000	950 000	275 000
7	2 000	320	432 000	640 000	208 000
8	2 100	170	229 500	357 000	127 500
9	2 200	110	148 500	242 000	93 500

Как видно из приведенных расчетов, оптимальная цена на подъемник должна находиться в диапазоне 1600–1700 у.е.

На основе многомерной регрессионной зависимости методом наименьших квадратов была построена математическая модель рынка. Она довольно точно отражает реальное положение дел. При исходной цене 1650 у.е. продажи ориентировочно должны составить 1010 шт. На рис.1 приведена кривая спроса.

Эти расчеты были сделаны при допущении, что издержки не меняются в течение длительного промежутка времени. Однако, в реальных условиях постоянный рост стоимости энергоресурсов и непрекращающаяся инфляция издержек (рост затрат на сырье, материалы, комплектующие изделия, рабочую силу) приводит к увеличению издержек. Поэтому А. А. Пивень проанализировал оптимальный объем выпуска при их различных значениях. Данные его расчетов приведены в табл.7. Поскольку инфляция в нашей стране заметно искажает стоимостные характеристики, используем для их описания условные единицы (у.е.).

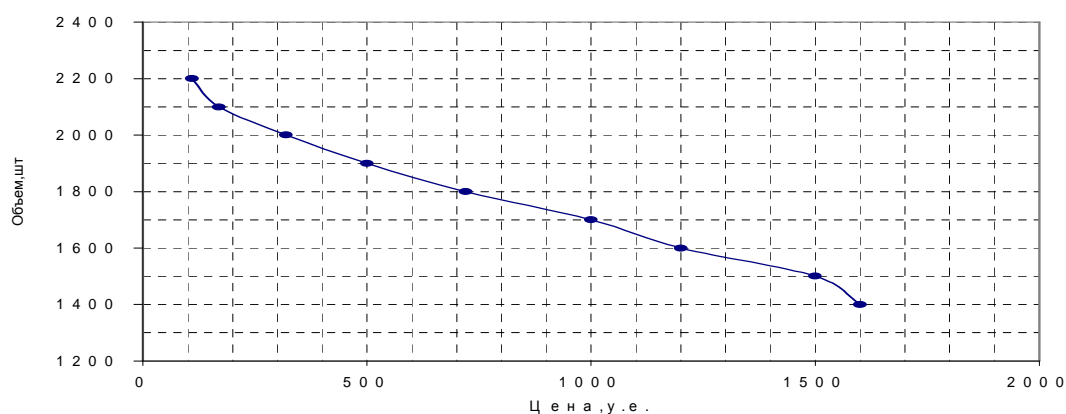


Рисунок 1 – Кривая спроса на изделие.

Для удобства восприятия рассмотренные результаты относительно оптимальных объемов производства при соответствующих ценах приведены в табл.8.

Таблица 7. – Прибыль в зависимости от цены и издержек

№ п/п	Цена, у.е.	Объем продаж, шт.	Прибыль (тыс. у.е.) при издержках на единицу продукции, у.е.							
			1350	1400	1450	1500	1550	1600	1650	1700
1	1 400	1600	80	0	-	-	-	-	-	-
2	1 500	1500	225	150	75	0	-	-	-	-
3	1 600	1200	300	240	180	120	60	0	-	-
4	1 700	1000	350	300	250	200	150	100	50	0
5	1 800	720	324	288	252	216	180	144	108	72
6	1 900	500	275	250	225	200	175	150	125	100
7	2 000	320	208	192	176	160	144	128	112	96
8	2 100	170	127,5	119	110,5	102	93,5	85	76,5	68
9	2 200	110	93,5	88	82,5	77	71,5	66	60,5	55

Таблица 8. – Оптимальные выпуск и цена в зависимости от издержек

Издержки	1350	1400	1450	1500	1550	1600	1650	1700
Оптимальный выпуск	1000	1000	720	720	720	500	500	500
Цена	1700	1700	1800	1800	1800	1900	1900	1900

Как видно из табл.8, увеличение издержек ведет к снижению оптимального выпуска при росте цены. Хотя изменение издержек на 50 у.е. может не сразу привести к изменению цены. Необоснованная цена может “переключить” большую группу потребителей на другое, аналогичное изделие, имеющее сходный по уровню набор технических характеристик, но более низкую рыночную цену.

По данным функции спроса (табл.7) проведем расчет эластичности спроса по цене. Под ценовой эластичностью спроса понимается степень реагирования рыночного спроса на изменение цен. В классическом понимании эластичность спроса по цене показывает, насколько изменится объем спроса при изменении цены на 1%. Спрос квалифицируется как эластичный, если понижение цены вызывает такой рост оборота, при котором увеличение объема продаж с лихвой компенсирует более низкие цены. Если же понижение цены, приводя к некоторому увеличению объема продаж, тем не менее, не ведет к увеличению оборота или даже уменьшает его, то такой спрос называется неэластичным. Коэффициент ценовой эластичности спроса определяется по формуле:

$$K_{цэс} = \frac{(Q_1 - Q_2) / (Q_1 + Q_2)}{(P_1 - P_2) / (P_1 + P_2)},$$

где Q_1, Q_2 – значения объема продаж; P_1, P_2 – значения цены изделия.

В рассматриваемом случае $K_{цэс}$ будет различен на протяжении всей функции спроса (рис.1). Однако, произведем расчет на той части кривой (в том диапазоне), где присутствует расчетная цена подъемника, а именно: $Q_1=1200$ шт.; $Q_2=720$ шт.; $P_1=1600$ у.е.; $P_2=1800$ у.е. В этом случае

$$K_{цэс} = \frac{(1200 - 720) / (1200 + 720)}{(1600 - 1800) / (1600 + 1800)} = -4,25.$$

Коэффициент $K_{цэс}$ имеет отрицательный знак и абсолютную величину, значительно превышающую 1. Это говорит о сильной обратной зависимости объемов продаж от цены. Спрос на подъемник весьма эластичен. Валовая выручка увеличивается при снижении цены и уменьшается при ее повышении. Компании необходимо быть готовой к тому, что покупатели очень чутко реагируют на всякое повышение цены на изделие значительным снижением объемов закупок. Как отмечает А. А. Пивень, снижение эластичности спроса на изделие возможно только при общем росте благосостояния насе-

ления страны и в частности, значительного роста доходной части бюджетов промышленных предприятий.

3.4.5. Проверка однородности двух биномиальных выборок

Проверка однородности – одна из базовых проблем, решаемых статистическими методами. Она часто обсуждается в литературе, а методы проверки однородности применяются при решении многих практических задач. Например, как сравнить две группы - мужчин и женщин, молодых и пожилых, и т.п.? В маркетинге это важно для сегментации рынка. Если две группы не отличаются по ответам, значит, их можно объединить в один сегмент и проводить по отношению к ним одну и ту же маркетинговую политику, в частности, осуществлять одни и те же рекламные воздействия. Если же две группы различаются, то и относиться к ним надо по-разному. Это - представители двух разных сегментов рынка, требующих разного подхода при борьбе за их завоевание.

Обсуждаемая далее постановка задачи в статистических терминах такова. Рассматривается вопрос с двумя возможными ответами, например, «да» и «нет». В первой группе из n_1 опрошенных m_1 человек сказали «да», а во второй группе из n_2 опрошенных m_2 сказали «да». В вероятностной модели предполагается, что m_1 и m_2 - биномиальные случайные величины $B(n_1, p_1)$ и $B(n_2, p_2)$ соответственно. Запись $B(n, p)$ означает, что случайная величина m имеет биномиальное распределение с параметрами n - объем выборки и p - вероятность определенного ответа (скажем, ответа «да»). Такая случайная величина может быть представлена в виде суммы $m = X_1 + X_2 + \dots + X_n$, где случайные величины X_1, X_2, \dots, X_n независимы, одинаково распределены, принимают два значения 1 и 0, причем $P(X_i = 1) = p, P(X_i = 0) = 1 - p, i = 1, 2, \dots, n$.

Однородность двух групп означает, что соответствующие им вероятности равны, неоднородность - что эти вероятности отличаются. В терминах прикладной математической статистики задача ставится так: необходимо проверить гипотезу однородности

$$H_0: p_1 = p_2$$

при альтернативной гипотезе о наличии эффекта

$$H_1: p_1 \neq p_2$$

(Иногда представляют интерес односторонние альтернативные гипотезы $H_1': p_1 > p_2$ и $H_1'': p_1 < p_2$.)

Оценкой вероятности p_1 является частота $p_1^* = m_1/n_1$, а оценкой вероятности p_2 является частота $p_2^* = m_2/n_2$. Даже при совпадении вероятностей p_1 и p_2 частоты, как правило, различаются. Как говорят, «по чисто случайным причинам». Рассмотрим случайную величину $p_1^* - p_2^*$. Тогда

$$M(p_1^* - p_2^*) = p_1 - p_2, D(p_1^* - p_2^*) = p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2.$$

Из теоремы Муавра - Лапласа и теоремы о наследовании сходимости (см. разд. 3.2) следует, что

$$\lim_{n_1 \rightarrow \infty, n_2 \rightarrow \infty} P \left\{ \frac{p_1^* - p_2^* - M(p_1^* - p_2^*)}{\sqrt{D(p_1^* - p_2^*)}} \leq x \right\} = \Phi(x),$$

где $\Phi(x)$ - функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1. Для практического применения этого соотношения следует заменить неизвестную статистику дисперсию разности частот на оценку этой дисперсии:

$$D^*(p_1^* - p_2^*) = p_1^*(1 - p_1^*)/n_1 + p_2^*(1 - p_2^*)/n_2.$$

(Могут использоваться и другие оценки рассматриваемой дисперсии, например, при справедливости нулевой гипотезы - по объединенной выборке.) С помощью указанной

выше математической техники можно показать, что при такой замене предельное распределение не меняется:

$$\lim_{n_1 \rightarrow \infty, n_2 \rightarrow \infty} P \left\{ \frac{p_1^* - p_2^* - M(p_1^* - p_2^*)}{\sqrt{D^*(p_1^* - p_2^*)}} \leq x \right\} = \Phi(x).$$

При справедливости гипотезы однородности (т.е. при отсутствии эффекта) имеем $M(p_1^* - p_2^*) = 0$. Поэтому правило принятия решения при проверке однородности двух выборок выглядит так:

1. Вычислить статистику

$$Q = \frac{p_1^* - p_2^*}{\sqrt{\frac{p_1^*(1-p_1^*)}{n_1} + \frac{p_2^*(1-p_2^*)}{n_2}}}.$$

2. Сравнить значение модуля статистика $|Q|$ с граничным значением K . Если $|Q| \leq K$, то принять гипотезу однородности H_0 . Если же $|Q| > K$, то заявить об отсутствии однородности и принять альтернативную гипотезу H_1 .

Граничное значение K определяется выбором уровня значимости статистического критерия проверки однородности. Из приведенных выше предельных соотношений следует, что при справедливости гипотезы однородности H_0 для уровня значимости $\alpha = P(|Q| > K)$ имеем (при $n_1 \rightarrow \infty, n_2 \rightarrow \infty$)

$$\alpha \rightarrow 1 - \Phi(K) + \Phi(-K) = 2 - 2\Phi(K).$$

Следовательно, граничное значение в зависимости от уровня значимости целесообразно выбирать из условия

$$K = K(\alpha) = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right).$$

Здесь $\Phi^{-1}(\cdot)$ - функция, обратная к функции стандартного нормального распределения. В социально-экономических исследованиях наиболее распространен 5% уровень значимости, т.е. $\alpha = 0,05$. Для него $K = 1,96$.

Пример 2. Пусть в первой группе из 500 опрошенных мужчин ответили «да» 200, а во второй группе из 700 опрошенных женщин сказали «да» 350. Есть ли разница по доле отвечающих «да» между генеральными совокупностями, представленными этими двумя группами?

Для установления взаимопонимания с маркетологом уберем из формулировки примера относящийся к теории статистики термин «генеральная совокупность». Получим следующую постановку.

Пусть из 500 опрошенных мужчин ответили «да, я люблю пепси-колу» 200, а из 700 опрошенных женщин 350 сказали «да, я люблю пепси-колу». Есть ли разница между мужчинами и женщинами по доле отвечающих «да» на вопрос о любви к пепси-коле?

В рассматриваемом примере нужные для расчетов величины таковы: $n_1 = 500, p_1^* = 200/500 = 0,4; n_2 = 700, p_2^* = 350/700 = 0,5$. Вычислим статистику

$$\begin{aligned} Q &= \frac{0,4 - 0,5}{\sqrt{\frac{0,4 \cdot 0,6}{500} + \frac{0,5 \cdot 0,5}{700}}} = \frac{-0,1}{\sqrt{\frac{0,24}{500} + \frac{0,25}{700}}} = \frac{-0,1}{\sqrt{0,00048 + 0,0003571}} \\ &= \frac{-0,1}{\sqrt{0,0008371}} = \frac{-0,1}{0,029} = -3,45. \end{aligned}$$

Поскольку $|Q| = 3,45 > 1,96$, то необходимо отклонить нулевую гипотезу и принять альтернативную. Таким образом, мужчины и женщины отличаются по рассматриваемому признаку - любви к пепси-коле.

Необходимо отметить, что результат проверки гипотезы однородности зависит не только от частот, но и от объемов выборок. Предположим, что частоты (доли) зафиксированы, а объемы выборок растут. Тогда числитель статистики Q не меняется, а знаменатель уменьшается, значит, вся дробь возрастает. Поскольку знаменатель стремится к 0, то дробь возрастает до бесконечности и рано или поздно превзойдет любую границу. Есть только одно исключение - когда в числителе стоит 0. Следовательно, при строгом подходе к формулировкам вывод статистика должен выглядеть так: «различие обнаружено» или «различие не обнаружено». Во втором случае различие, возможно, было бы обнаружено при увеличении объемов выборок.

Как и для доверительного оценивания вероятности, во ВЦИОМ разработаны две полезные таблицы, позволяющие оценить вызванные чисто случайными причинами допустимые расхождения между частотами в группах. Эти таблицы рассчитаны при выполнении нулевой гипотезы однородности и соответствуют ситуациям, когда частоты близки к 50% (табл.9) или к 20% (табл.10). Если наблюдаемые частоты - от 30% до 70%, то рекомендуется пользоваться первой из этих таблиц, если от 10% до 30% или от 70% до 90% - то второй. Если наблюдаемые частоты меньше 10% или больше 90%, то теорема Муавра-Лапласа и основанные на ней асимптотические формулы дают не очень хорошие приближения, целесообразно применять иные, более продвинутые математические средства, в частности, приближения с помощью распределения Пуассона.

В условиях разобранного выше примера табл.9 дает допустимое расхождение 7%. Действительно, объем первой группы 500 отсутствует в таблице, но строки, соответствующие объемам 400 и 600, совпадают для первых двух столбцов слева. Эти столбцы соответствуют объемам второй группы 750 и 600, между которыми расположен объем 700, данный в примере. Он ближе к 750, поэтому берем величину расхождения, стоящую на пересечении первого столбца и второй (и третьей) строк, т.е. 7%. Поскольку реальное расхождение (10%) больше, чем 7%, то делаем вывод о наличии значимого различия между группами. Естественно, этот вывод совпадает с полученным ранее расчетным путем.

Таблица 9. – Допустимые расхождения (в %) между частотами в двух группах, когда наблюдаются частоты от 30% до 70%

Объемы групп	750	600	400	200	100
750	6	7	7	10	12
600	7	8	8	11	13
400	7	8	10	11	14
200	10	11	11	13	16
100	12	13	14	16	18

Таблица 10. – Допустимые расхождения (в %) между частотами в двух группах, когда наблюдаются частоты от 10% до 30% или от 70% до 90%

Объемы групп	750	600	400	200	100
750	5	5	6	8	10
600	5	6	7	8	10
400	6	7	8	9	11
200	8	8	9	10	12
100	10	10	11	12	14

Как и в случае табл.5, значения в таблицах 9 и 10 несколько больше, чем рассчитанные по приведенным выше формулам. Дело в том, что таблицы ВЦИОМ связаны не с уровнем значимости $\alpha = 0,05$, а с уровнем значимости $\alpha = 0,01$, которому соответствует граничное значение 2,58.

Допустимое расхождение $\Delta = \Delta(\alpha)$ между частотами нетрудно получить расчетным путем. Для этого достаточно воспользоваться формулой для статистики Q и определить, при каком максимальном расхождении частот все еще делается вывод о том, что верна гипотеза однородности. Следовательно, допустимое расхождение $\Delta = \Delta(\alpha)$ находится из уравнения

$$K(\alpha) = \frac{\Delta(\alpha)}{\sqrt{\frac{p_1^*(1-p_1^*)}{n_1} + \frac{p_2^*(1-p_2^*)}{n_2}}}$$

Таким образом,

$$\Delta(\alpha) = K(\alpha) \sqrt{\frac{p_1^*(1-p_1^*)}{n_1} + \frac{p_2^*(1-p_2^*)}{n_2}}$$

Для данных примера 2 $\Delta = \Delta(\alpha) = 1,96 \times 0,029 = 0,057$, или 5,7%, для уровня значимости 0,05.

Для других уровней значимости надо использовать другие коэффициенты $K(\alpha)$. Так, $K(0,01) = 2,58$ для уровня значимости 1% и $K(0,10) = 1,64$ для уровня значимости 10%. Для данных примера $\Delta = \Delta(\alpha) = 2,58 \times 0,029 = 0,7482 \approx 0,075$, или 7,5%, для уровня значимости 0,01. Если округлить до ближайшего целого числа процентов, то получим 7%, как при использовании таблицы 9 выше.

Анализ таблиц 9 и 10 показывает, что для обнаружения эффекта (констатации различия генеральных совокупностей) частоты должны отличаться не менее чем на 6%. А при некоторых объемах выборок - более чем на 10%, например, при объемах выборок 100 и 100 - на 19%. Если же частоты отличаются на 5% или менее, можно сразу сказать, что статистический анализ приведет к выводу о том, что различие не обнаружено (для выборок объемов не более 750).

В связи со сказанным возникает вопрос: каково типовое отличие частот в двух выборках из одной и той же совокупности? Разность частот в этом случае имеет нулевое математическое ожидание и дисперсию

$$p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right) = \frac{p(1-p)(n_1+n_2)}{n_1 n_2}$$

Величина $p(1-p)$ достигает максимума при $p = 1/2$, и этот максимум равен 1/4. Если $p = 1/2$, а объемы двух выборок совпадают и равны 500, то дисперсия разности частот равна

$$\sigma^2 = \frac{0,25 \times 1000}{500 \times 500} = \frac{250}{250 \times 1000} = \frac{1}{1000}$$

Следовательно, среднее квадратическое отклонение σ равно 0,032, или 3,2%. Поскольку для стандартной нормальной случайной величины в 50% случаев ее значение не превосходит по модулю 0,67 (а в 50% случаев - больше 0,67), то типовой разброс равен $0,67 \sigma$, а в рассматриваемом случае - 2,1%.

Приведенные соображения дают возможность построить метод контроля правильности (корректности) проведения повторных опросов. Если частоты излишне устойчивы, значения при повторных опросах слишком близки - это подозрительно! Возможно, нарушены правила проведения опросов, выборки не являются случайными, ответы фальсифицированы, и т.д.

3.5. Проверка однородности

В конце предыдущего раздела шла речь о проверке однородности двух биномиальных выборок. В настоящем разделе продолжим обсуждать проблему однородности - рассмотрим систему эконометрических моделей и методов, предназначенных для проверки однородности двух независимых выборок. Подобные системы разработаны в прикладной статистике [21] для решения многих иных задач статистического анализа данных, однако объем настоящей монографии не позволяет провести подробный разбор всех таких задач и систем. Настоящий раздел следует рассматривать как пример разработки системы эконометрических моделей и методов, предназначенной для решения определенной задачи.

3.5.1. Система моделей проверки однородности двух независимых выборок

В прикладных исследованиях часто возникает необходимость выяснить, различаются ли генеральные совокупности, из которых взяты две независимые выборки. Например, надо выяснить, зависят ли от способа упаковки потребительские качества подшипников, измеренные через год после хранения. Или: влияет ли система оплаты на производительность труда.

В математико-статистических терминах постановка задачи такова: имеются две выборки x_1, x_2, \dots, x_m и y_1, y_2, \dots, y_n , требуется проверить их однородность. Напомним, что выборка моделируется как совокупность независимых одинаково распределенных случайных величин. Термин «однородность» уточняется ниже.

Противоположным понятием является «различие» (или «наличие эффекта»). Можно переформулировать задачу: требуется проверить, есть ли различие между выборками. Если различия нет, то для дальнейшего изучения две рассматриваемые выборки часто объединяют в одну.

Например, в маркетинге важно выделить сегменты потребительского рынка. Если установлена однородность двух выборок мнений потребителей, то возможно объединение сегментов, из которых эти выборки взяты, в один. В дальнейшем это позволит осуществлять по отношению к ним одинаковую маркетинговую политику (проводить одни и те же рекламные мероприятия и т.п.). Если же установлено различие, то поведение потребителей в двух сегментах различно, объединять эти сегменты нельзя, и могут понадобиться различные маркетинговые стратегии, своя для каждого из этих сегментов.

Вероятностная модель порождения данных. Для обоснованного выбора и применения организационно-экономических (эконометрических, статистических) методов необходимо прежде всего построить и обосновать вероятностную модель порождения данных. При проверке однородности двух выборок общепринята модель, в которой x_1, x_2, \dots, x_m рассматриваются как результаты m независимых наблюдений некоторой случайной величины X с функцией распределения $F(x)$, неизвестной статистике, а y_1, y_2, \dots, y_n - как результаты n независимых наблюдений, вообще говоря, другой случайной величины Y с функцией распределения $G(x)$, также неизвестной статистике. Предполагается также, что наблюдения в одной выборке не зависят от наблюдений в другой, поэтому выборки и называют независимыми.

Возможность применения модели в конкретной реальной ситуации требует обоснования. Независимость и одинаковая распределенность результатов наблюдений, входящих в выборку, могут быть установлены или исходя из методики проведения конкретных наблюдений, или путем проверки статистических гипотез независимости и

одинаковой распределенности с помощью соответствующих критериев проверки статистических гипотез [2].

Если проведено $(m+n)$ измерений объемов продаж в $(m+n)$ торговых точках, то описанную выше модель, как правило, можно применять. Если же, например, x_i и y_i - объемы продаж одного и того же товара до и после определенного рекламного воздействия, то рассматриваемую модель применять нельзя, поскольку очевидно, что эти объемы продаж определяются не только и не столько рекламным воздействием, сколько особенностями конкретной торговой точки (ее расположением, продолжительностью работы, репутацией и т.д.). В последнем случае используют модель связанных выборок. В ней обычно строят новую выборку $z_i = x_i - y_i$ и используют статистические методы анализа одной выборки, а не двух. Методы проверки однородности для связанных выборок рассматриваются в [21].

При дальнейшем изложении принимаем описанную выше вероятностную модель двух выборок.

Классификация моделей по типам данных. В предыдущем разделе рассматривались результаты измерений по альтернативным признакам. Каждое из чисел x_i и y_i принимало одно из двух значений, для определенности, 0 или 1. Если респондент дает ответ «да» на вопрос анкеты, то $x_i = 1$, если его ответ – «нет», то $x_i = 0$. Такие признаки называют также дихотомическими или бинарными. Распределение элементов первой выборки x_1, x_2, \dots, x_m описывается одним числом $P(x_i = 1) = p_1$. Распределение элементов второй выборки y_1, y_2, \dots, y_n также описывается одним числом $P(y_i = 1) = p_2$. Проверка однородности двух независимых выборок состоит в проверке статистической гипотезы $H_0: p_1 = p_2$ при альтернативной гипотезе о наличии эффекта $H_1: p_1 \neq p_2$. Метод проверки гипотезы H_0 разобран в конце предыдущего раздела. Есть и другие методы проверки этой гипотезы, основанные на использовании иных статистик [16].

Обобщением альтернативного признака является такой, значением которого является элемент некоторого конечного множества. Например, респондент выбирает не из двух ответов («да» или «нет»), а из трех («да», «нет», «может быть»). Пусть множество значений состоит из k элементов (их часто называют градациями признака). Занумеруем их натуральными числами $j = 1, 2, \dots, k$. Для простоты записи будем считать, что элемент выборки – это *номер* значения (градации), которое принимает признак. Тогда распределение случайных элементов двух выборок со значениями в одном и том же конечном множестве описывается вероятностями

$$P(x_i = j) = p_j(1), P(y_i = j) = p_j(2), j = 1, 2, \dots, k.$$

Таким образом, в отличие от альтернативного признака каждое распределение задается не одним числом, а k числами, неотрицательными и в сумме составляющими 1, так что «свободных параметров» всего $(k - 1)$.

Подобные k -значные признаки обычно возникают при измерениях по качественным шкалам (наименований и порядковой). Однако иногда они возникают в результате группировки значений количественных (числовых) признаков. Будем называть их «признаки с конечным числом градаций». Проверка однородности для таких признаков – это проверка сложной гипотезы

$$H_0: p_j(1) = p_j(2), j = 1, 2, \dots, k.$$

Альтернативную гипотезу наиболее общего вида можно записать так:

$$H_1: \sum_{j=1}^k (p_j(1) - p_j(2))^2 > 0.$$

Третий тип рассматриваемых здесь данных – количественные признаки, значения которых – действительные числа, а функции распределения непрерывны.

Уточнения понятия однородности для количественных данных. Понятие «однородность», т. е. «отсутствие различия», может быть формализовано в терминах вероятностной модели различными способами.

Наивысшая степень однородности (абсолютная однородность) достигается, если обе выборки взяты из одной и той же генеральной совокупности, т. е. справедлива нулевая гипотеза

$$H_0: F(x)=G(x) \text{ при всех } x.$$

Отсутствие абсолютной однородности означает, что верна альтернативная гипотеза, согласно которой

$$H_1: F(x_0) \neq G(x_0)$$

хотя бы при одном значении аргумента x_0 . Если гипотеза H_0 принята, то выборки можно объединить в одну, если нет - то нельзя.

В некоторых случаях целесообразно проверять не совпадение функций распределения, а лишь совпадение некоторых характеристик случайных величин X и Y - математических ожиданий, медиан, дисперсий, коэффициентов вариации и др. (однородность тех или иных характеристик). Например, однородность математических ожиданий означает, что справедлива гипотеза

$$H'_0: M(X)=M(Y),$$

где $M(X)$ и $M(Y)$ - математические ожидания случайных величин X и Y , результаты наблюдений над которыми составляют первую и вторую выборки соответственно. Доказательство различия между выборками в рассматриваемом случае - это доказательство справедливости альтернативной гипотезы

$$H'_1: M(X) \neq M(Y).$$

Если гипотеза H_0 верна, то и гипотеза H'_0 верна, но из справедливости H'_0 , вообще говоря, не следует справедливость H_0 . Математические ожидания могут совпадать для различающихся между собой функций распределения. В частности, если в результате обработки выборочных данных принята гипотеза H'_0 , то отсюда *не следует*, что две выборки можно объединить в одну. Однако в ряде ситуаций целесообразна проверка именно гипотезы H'_0 . Например, пусть функция спроса на определенный товар или услугу оценивается путем опроса потребителей (первая выборка) или с помощью данных о продажах (вторая выборка). Тогда маркетологу важно проверить гипотезу об отсутствии систематических расхождений результатов этих двух методов, т.е. гипотезу о равенстве математических ожиданий. Другой пример - из производственного менеджмента. Пусть изучается эффективность управления бригадами рабочих на предприятии с помощью двух организационных схем, результаты наблюдения - объем производства продукции или услуг на одного члена бригады (производительность), а показатель эффективности организационной схемы - средний (по предприятию) объем производства на одного рабочего. Тогда для сравнения эффективности препаратов достаточно проверить гипотезу H'_0 .

Иногда нужно проверить однородность дисперсий. Например, различаются ли два способа измерения по величине случайной ошибки - т.е. по дисперсии случайных погрешностей.

Рассмотрим проверку однородности для признаков с конечным числом градаций, а затем - для количественных признаков.

3.5.2. Проверка согласия и однородности для признаков с конечным числом градаций

Проведем в качестве примера обработку данных, относящихся к известной всем читателям тематике. Дональд А. Уиндзор подсчитал, сколько ученых родилось под ка-

ждым из знаков Зодиака (см. журнал «Химия и жизнь», 1976, №4, с.112–113). Им были взяты две научные специальности – таксономия (т.е. теория классификация биологических организмов) и молекулярная биология. Результаты приведены в табл.1.

Видно, что под знаком Овна родилось гораздо больше молекулярных биологов, чем под любым другим – почти в 1,5 раза больше, чем приходится в среднем на один знак Зодиака. Таксономисты чаще рождались под знаком Рака, а реже всего – под знаком Скорпиона. Для этой специальности среднее число рождений больше числа рождений под знаком Скорпиона почти в 1,6 раза, а отношение максимального числа в столбце таксономистов к минимальному равно $38/18 \approx 2,1$ (!).

Разве все эти факты не доказывают, что специальность ученого и знак Зодиака, под которым он родился, связаны между собой, что молекулярные биологи, скажем, не случайно чаще всего рождаются под знаком Овна?

Таблица 1 – Специальности ученых и знаки Зодиака их дней рождений

Номер (i)	Знак Зодиака	Количество таксономистов ($m_i(1)$)	Количество молекулярных биологов ($m_{2i}(2)$)
1	Овен	28	58
2	Телец	30	32
3	Близнецы	31	39
4	Рак	38	41
5	Лев	32	32
6	Дева	31	42
7	Весы	25	41
8	Скорпион	18	41
9	Стрелец	27	40
10	Козерог	25	33
11	Водолей	26	35
12	Рыба	31	36
Всего (n_i)		342	470
В среднем на знак Зодиака ($n_i/12$)		28,5	39,17

Нет, не доказывают. Почему превышение над средним уровнем в 1,5 раза считать большим, а, к примеру, в 1,1 раза – малым? Может быть, и то, и другое вызывается число случайными причинами?

Как теория организационно-экономического моделирования рекомендует поступать? Прежде всего необходимо сформулировать гипотезу, которую будем проверять. Или несколько гипотез. А для этого построим вероятностно-статистическую модель, в терминах которой сформулируем гипотезу. Модель нужна, чтобы дальнейшие расчеты опирались на теорию математической статистики.

Принимаем, что для каждой из двух генеральных совокупностей ученых (таксономистов и молекулярных биологов) существуют 12 вероятностей событий, состоящих в рождении ученого под определенным знаком Зодиака. Обозначим $p_1(1)$ вероятность того, что таксономист родился под знаком Овна, $p_2(1)$ – что он родился под знаком Тельца, и так далее до $p_{12}(1)$ – вероятности рождения под знаком Рыбы (знаки Зодиака перенумерованы в табл.2.1). Кроме того, считаем, что $n_1 = 342$ изученных Дональдом А. Уиндзором таксономиста выбраны из всей совокупности ученых этой специальности таким способом, который никак не связан с днями и месяцами их рождений – ведь иначе мы не можем распространить выводы, полученные по выборке, на всю совокупность. Короче, рассматриваемая выборка является представительной.

Итак, вероятностно-статистическая модель такова. Считаем, что в столбце таксономистов табл.1 записаны результаты $n_1 = 342$ опытов, проведенных независимо друг

от друга, в каждом из которых осуществляется одно из 12 событий – с вероятностью $p_1(1)$ качественный признак принимает значение 1 (интерпретируется как «родился под знаком Овна»), с вероятностью $p_2(1)$ признак принимает значение 2 (т.е. «родился под знаком Тельца»), и так далее до значения 12 («родился под знаком Рыбы»), которое этот признак принимает с вероятностью $p_{12}(1)$. Модель для описания результатов $n_2 = 470$ опытов, приведенных в столбце молекулярных биологов, отличается только другими обозначениями вероятностей, а именно, $p_1(2), p_2(2), \dots, p_{12}(2)$ для рождений под знаками Овна, Тельца, ..., Рыбы соответственно.

Самая простая гипотеза, которая приходит в голову, такова: шансы родиться под каждым знаком Зодиака одинаковы. Поскольку всего знаков 12, то имеется 1 шанс из 12 родиться под знаком Овна, 1 шанс из 12 – под знаком Тельца, и т.д. Значит, все вероятности равны между собой и потому равны $1/12$. Речь идет о нулевых гипотезах

$$H_0(1): p_1(1) = p_2(1) = \dots = p_{12}(1) = \frac{1}{12},$$

$$H_0(2): p_1(2) = p_2(2) = \dots = p_{12}(2) = \frac{1}{12}.$$

При взгляде на табл.1 возникает, скажем, гипотеза, что для таксономистов $p_4(1)$ больше $p_8(1)$ (вероятность родиться под знаком Рака больше, чем вероятность родиться под знаком Скорпиона). Если эта гипотеза справедлива, то знаки Зодиака не являются равноправными, и реальный мир устроен более сложно, чем в случае равных вероятностей. По принципу экономии мышления (известен также как «бритва Оккама»¹) необходимо сначала проверить, не соответствуют ли всё-таки данные табл.1 гипотезе равноправности шансов, и только в случае обнаружения противоречия переходить к более сложным гипотезам.

Итак, мы пришли к следующим задачам проверки статистических гипотез: не противоречат ли данные табл.1 гипотезам $H_0(1)$ и $H_0(2)$?

Критерий χ^2 (хи-квадрат) для проверки согласия с фиксированным распределением. В математической статистике со времен великого английского статистика Карла Пирсона (1857–1936) рассматривают задачу проверки согласия эмпирического распределения с теоретическим. А именно, пусть в результате опыта осуществляется один и только один из k исходов. Пусть эти исходы занумерованы натуральными числами от 1 до k , и p_1, p_2, \dots, p_k – вероятности этих исходов. Пусть проведены n опытов, в результате которых m_1 раз осуществился первый исход, m_2 раз – второй, ..., m_k раз – k -ый исход. По этим статистическим данным требуется проверить статистическую гипотезу о том, что вероятности исходов опыта совпадают с заданными:

$$H_0: p_j = p_{j0}, j = 1, 2, \dots, k.$$

Альтернативная гипотеза, которую обычно рассматривают, состоит в том, что нарушается хотя бы одно из этих равенств. Ее можно записать так:

$$H_1: \sum_{j=1}^k (p_j - p_{j0})^2 > 0.$$

Для проверки этой гипотезы согласия эмпирического распределения с теоретическим естественно исходить из того, что при ее справедливости случайная величина m_j

¹ «Бритва (лезвие) Оккама» — методологический принцип, получивший название по имени английского монаха-францисканца, философа-номиналиста Вильяма Оккама (Ockham, Ockam, Oссam; ок. 1285–1349). В упрощенном виде он гласит: «Не следует множить сущее без необходимости» (либо «Не следует привлекать новые сущности без самой крайней на то необходимости»). Этот принцип формирует базис методологического редукционизма. Его называют также *принципом бережливости*, или *законом экономии*.

имеет биномиальное распределение с вероятностью p_{j0} и числом опытов n , а потому ее математическое ожидание равно np_{j0} . Следовательно, отклонение эмпирического распределения от теоретического описывается величинами $m_j - np_{j0}$. С некоторой естественной точки зрения [13] наилучший способ проверки согласия основан на введенной Карлом Пирсоном статистике критерия хи-квадрат, рассчитываемой по формуле

$$\chi^2 = \sum_{j=1}^k \frac{(m_j - np_{j0})^2}{np_{j0}}.$$

При росте объема выборки n распределение только что введенной статистики критерия хи-квадрат стремится к известному в теории вероятностей распределению хи-квадрат с $(k - 1)$ степенью свободы, т.е. к распределению суммы $(k - 1)$ независимых случайных величин, каждая из которых – квадрат стандартной нормальной случайной величины (с математическим ожиданием 0 и дисперсией 1). Исходя из этого математического утверждения, для проверки гипотезы согласия по уровню значимости α находят квантиль $\chi^2(\alpha, k-1)$ порядка $(1 - \alpha)$ распределения хи-квадрат с $(k - 1)$ степенью свободы (с помощью распространенных таблиц, имеющих, в частности, в сборнике [2], или с помощью различных программных продуктов). Правило принятия решений при проверке гипотезы согласия таково: если рассчитанное по эмпирическим данным значение статистики хи-квадрат таково, что

$$\chi^2 \leq \chi^2(\alpha, k - 1),$$

то гипотезу согласия принимают на заданном уровне значимости (и констатируют, что эмпирические данные не противоречат гипотезе $H_0 : p_j = p_{j0}, j = 1, 2, \dots, k$); если же

$$\chi^2 > \chi^2(\alpha, k - 1),$$

то гипотезу согласия отклоняют (и принимают альтернативную гипотезу).

Замечание 1. Чтобы можно было опираться на предельное соотношение, требуется, чтобы величины np_{j0} не были малыми. Для этого достаточно, чтобы $np_{j0} \geq 5$ при всех $j = 1, 2, \dots, k$.

Замечание 2. В математической статистике еще несколько критериев проверки гипотез называются «критериями хи-квадрат», например, критерий для проверки однородности распределений значений конечнозначных признаков (см. ниже), критерий для проверки независимости признаков на основе таблицы сопряженности. Причина проста – распределения статистик этих критериев сходятся к распределению хи-квадрат.

Замечание 3. Если соответствующее опыту распределение мало отличается от теоретического, то при сравнительно небольшом объеме выборки скорее всего будет принята гипотеза согласия. При увеличении же объема выборки может быть обнаружено отличие распределения от теоретического. Поэтому в случае $\chi^2 \leq \chi^2(\alpha, k - 1)$ точнее формулировать вывод так: эмпирические данные не позволяют отклонить гипотезу согласия, в то время как в случае $\chi^2 > \chi^2(\alpha, k - 1)$ констатируем, что отклонение обнаружено.

Замечание 4. Если значение статистики χ^2 мало, то данные, возможно, фальсифицированы. Это утверждение основано на то, что при справедливости гипотезы согласия распределение статистики близко к хи-квадрат распределению, а потому осуществление маловероятных событий – слишком большого или слишком малого значения статистики χ^2 – практически невозможно.

Вернемся к данным табл.1. Соответственно числу знаков Зодиака $k = 12$. Проверим гипотезу равномерности, т.е. частный случай гипотезы согласия с $p_{j0} = 1/12, j = 1, 2, \dots, 12$. Из последней строки табл.1 следует, что приведенное в замечании 1 условие выполнено.

Проведя расчеты по приведенной выше формуле, получаем, что для таксономистов $\chi^2 = 9,36$, а для молекулярных биологов $\chi^2 = 13,85$. По таблицам [2] (и любым другим) для числа степеней свободы $k - 1 = 12 - 1 = 11$ квантиль распределения хи-квадрат порядка 0,9 есть 17,3, а квантиль порядка 0,95 равен 19,7. Это значит, что гипотеза согласия принимается (точнее, не отклоняется) при уровнях значимости) 0,1, а также 0,05 и всех иных, используемых на практике. По тем же таблицам

$$P(\chi^2 \leq 9,36) = 0,41, \quad P(\chi^2 \leq 13,85) = 0,76,$$

так что значения статистики попадают в среднюю часть распределения – они не слишком велики и не слишком малы.

Итак, данные табл.1 *идеально согласуются с равномерным распределением* моментов рождения по знакам Зодиака как для таксономистов, так и для молекулярных биологов. Отклонения от равномерности, отмеченные в начале подраздела, объясняются чисто случайными причинами.

Критерий χ^2 (хи-квадрат) проверки однородности распределений признаков с конечным числом градаций. Может быть, хотя гипотезу равноправности знаков Зодиака нельзя отвергнуть ни для одной группы ученых, но зато эти группы сильно различаются *между собой*, отклоняясь о среднего, как сказать, в разные стороны? Речь идет, очевидно, о проверке однородности распределений двух признаков с конечным числом градаций. Опишем соответствующую вероятностно-статистическую модель.

Пусть в результате опыта осуществляется один и только один из k исходов. Пусть эти исходы занумерованы натуральными числами от 1 до k . Пусть $p_1(1), p_2(1), \dots, p_k(1)$ – вероятности этих исходов для одной генеральной совокупности, а $p_1(2), p_2(2), \dots, p_k(2)$ – вероятности этих же исходов для другой генеральной совокупности. Другими словами, рассмотрим два признака (две случайные величины) X и Y , возможные значения которых – рассматриваемые k исходов. Распределения этих признаков таковы:

$$P(X = j) = p_j(1), \quad P(Y = j) = p_j(2), \quad j = 1, 2, \dots, k.$$

Пусть для признака X проведены $n(1)$ независимых опытов, в результате которых $m_1(1)$ раз осуществился первый исход, $m_2(1)$ раз - второй, ..., $m_k(1)$ раз – k -ый исход. Другими словами, проведено $n(1)$ независимых испытаний, в результате которых получено $n(1)$ независимых значений случайной величины X , причем эта случайная величина $m_1(1)$ раз приняла значение 1, $m_2(1)$ раз – значение 2, ..., $m_k(1)$ раз – значение k . Аналогичным образом получены статистические данные для случайной величины Y – проведено $n(2)$ независимых испытаний, в которых эта случайная величина $m_1(2)$ раз приняла значение 1, $m_2(2)$ раз – значение 2, ..., $m_k(2)$ раз – значение k . Причем испытания для X проведены независимо от испытаний для Y .

По этим статистическим данным, т.е. по двум независимым выборкам значений двух конечнозначных случайных величин, требуется проверить статистическую гипотезу о том, что распределения этих случайных величин совпадают. Другими словами, проверить гипотезу однородности распределений, т.е. сложную гипотезу

$$H_0: p_j(1) = p_j(2), \quad j = 1, 2, \dots, k.$$

В качестве альтернативной обычно рассматривают гипотезу о том, что хотя бы одно из этих k равенств не выполнено. Эту гипотезу наиболее общего вида можно записать так:

$$H_1: \sum_{j=1}^k (p_j(1) - p_j(2))^2 > 0.$$

Статистика критерия хи-квадрат имеет вид [22, с.275]:

$$\chi^2 = \chi^2(n(1), n(2)) = n(1)n(2) \sum_{j=1}^k \frac{1}{m_j(1) + m_j(2)} \left(\frac{m_j(1)}{n(1)} - \frac{m_j(2)}{n(2)} \right)^2.$$

Известно, что при росте объемов выборок $n(1)$ и $n(2)$ распределение статистики $\chi^2(n(1), n(2))$ стремится к распределению хи-квадрат с $(k - 1)$ степенями свободы. Поэтому при больших $n(1)$ и $n(2)$ правило принятия решений при проверке гипотезы однородности таково: если рассчитанное по эмпирическим данным значение статистики хи-квадрат таково, что

$$\chi^2(n(1), n(2)) \leq \chi^2(\alpha, k - 1),$$

то гипотезу однородности принимают на заданном уровне значимости; если же

$$\chi^2(n(1), n(2)) > \chi^2(\alpha, k - 1),$$

то гипотезу однородности отклоняют (и принимают альтернативную гипотезу). Здесь, как и ранее, $\chi^2(\alpha, k - 1)$ - это квантиль порядка $(1 - \alpha)$ распределения хи-квадрат с $(k-1)$ степенью свободы.

Замечание. Более точно, в случае $\chi^2(n(1), n(2)) > \chi^2(\alpha, k - 1)$ можно констатировать, что обнаружено различие распределений (как говорят, доказано наличие эффекта (на данном уровне значимости), в том время как в случае $\chi^2(n(1), n(2)) \leq \chi^2(\alpha, k - 1)$ можно сказать лишь, что эффект не обнаружен (нет оснований отвергнуть предположение о совпадении распределений).

Расчет по данным табл.1 дает, что

$$\chi^2 = 14,22 < \chi^2(0,1; 11), \quad P(\chi^2 \leq 14,22) = 0,78.$$

Таким образом, на любом из практически используемых уровней значимости констатируем однородность распределений. Наблюдаемое значение статистики (т.е. 14,22) попадает в среднюю часть распределения – при справедливости нулевой гипотезы значение статистики в 78% случаев меньше наблюдаемого и в 22% случаев больше наблюдаемого. Итак, не обнаружено никакой связи между специальностью ученого и знаком Зодиака, под которым он рожден.

3.5.3. Проверка однородности характеристик для количественных признаков

Перейдем к следующему элементу системы моделей, описанной в начале настоящего раздела – к моделям, нацеленным на проверку равенства характеристик двух распределений, из которых взяты две независимые выборки. Исходим из общепринятой базовой модели, в которой элементы первой выборки x_1, x_2, \dots, x_m рассматриваются как результаты m независимых наблюдений некоторой числовой случайной величины X с функцией распределения $F(x)$, неизвестной статистику, а элементы второй выборки y_1, y_2, \dots, y_n - как результаты n независимых наблюдений, вообще говоря, другой случайной величины Y с функцией распределения $G(x)$, также неизвестной статистику. Предполагается также, что наблюдения в одной выборке не зависят от наблюдений в другой, поэтому выборки и называют независимыми.

Замечание. Обратите внимание, что объемы выборок обозначены здесь не так, как в предыдущем разделе. Это сделано специально, а именно, для того, чтобы у читателя не создавался ложный стереотип, мешающий воспринимать многообразие литературных источников с соответствующим многообразием обозначений.

Традиционный метод проверки однородности (критерий Стьюдента). Для дальнейшего критического разбора опишем традиционный статистический метод проверки однородности. Он широко использовался в течение всего XX в. Хотя к настоя-

шему времени этот метод устарел (см. ниже), но продолжает встречаться в учебной литературе, и потому и применяется для анализа конкретных данных.

При использовании традиционного метода проверки однородности вычисляют выборочные средние арифметические в каждой выборке

$$\bar{x} = \frac{1}{m} \sum_{1 \leq i \leq m} x_i, \quad y = \frac{1}{n} \sum_{1 \leq i \leq n} y_i,$$

затем выборочные дисперсии

$$s_x^2 = \frac{1}{m-1} \sum_{1 \leq i \leq m} (x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n-1} \sum_{1 \leq i \leq n} (y_i - \bar{y})^2$$

и статистику Стьюдента t , на основе которой принимают решение,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(m-1)s_x^2 + (n-1)s_y^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}. \quad (1)$$

По заданному уровню значимости α и числу степеней свободы $(m+n-2)$ из таблиц распределения Стьюдента (см., например, [2]) находят критическое значение $t_{кр}$. Если $|t| > t_{кр}$, то гипотезу однородности (отсутствия различия) отклоняют, если же $|t| \leq t_{кр}$, то принимают. (При односторонних альтернативных гипотезах вместо условия $|t| > t_{кр}$ проверяют, что $t > t_{кр}$; эту постановку рассматривать не будем, так как в ней нет принципиальных отличий от обсуждаемой здесь.)

В литературе зачастую описывается только приведенный выше алгоритм. Этого недостаточно для квалифицированного анализа статистических данных. Рассмотрим условия применимости традиционного метода проверки однородности, основанного на использовании статистики t Стьюдента, а также обсудим современные методы проверки однородности двух выборок.

Классические условия применимости критерия Стьюдента. Согласно математико-статистической теории должны быть выполнены два классических условия применимости критерия Стьюдента, основанного на использовании статистики t , заданной формулой (1):

а) результаты наблюдений имеют нормальные распределения:

$$F(x) = N(x; m_1, \sigma_1^2), \quad G(x) = N(x; m_2, \sigma_2^2)$$

с математическими ожиданиями m_1 и m_2 и дисперсиями σ_1^2 и σ_2^2 в первой и во второй выборках соответственно;

б) дисперсии результатов наблюдений в первой и второй выборках совпадают:

$$D(X) = \sigma_1^2 = D(Y) = \sigma_2^2.$$

Если условия а) и б) выполнены, то нормальные распределения $F(x)$ и $G(x)$ отличаются только математическими ожиданиями, а поэтому *обе* гипотезы H_0 и H'_0 (см. подраздел 3.5.1) сводятся к гипотезе

$$H''_0 : m_1 = m_2,$$

а *обе* альтернативные гипотезы H_1 и H'_1 сводятся к гипотезе

$$H''_1 : m_1 \neq m_2.$$

Если условия а) и б) выполнены, то статистика t при справедливости H''_0 имеет распределение Стьюдента с $(m+n-2)$ степенями свободы. Только в этом случае описанный выше традиционный метод обоснован безупречно. Если хотя бы одно из условий а) и б) не выполнено, то нет никаких оснований считать, что статистика t имеет распределение Стьюдента, поэтому применение традиционного метода, строго говоря, не обосновано. Обсудим возможность проверки этих условий и последствия их нарушений.

Имеют ли результаты наблюдений нормальное распределение? Как подробно показано в литературе (см. раздел 3.3 выше, а также, например, [21, гл.5.1], [20, гл.4.1]), априори нет оснований предполагать нормальность распределения результатов экономических, технико-экономических, технических, медицинских и иных наблюдений. Следовательно, нормальность надо проверять. Разработано много статистических критериев для проверки нормальности распределения результатов наблюдений [2]. Однако проверка нормальности - более сложная и трудоемкая статистическая процедура, чем проверка однородности (как с помощью статистики t Стьюдента, так и с использованием непараметрических критериев, рассматриваемых ниже).

Для достаточно надежного установления нормальности требуется весьма большое число наблюдений. В указанных выше литературных источниках показано, что для того, чтобы гарантировать, что функция распределения результатов наблюдений отличается от некоторой нормальной не более чем на 0,01 (при любом значении аргумента), требуется порядка 2500 наблюдений. В большинстве технических, экономических, медицинских и иных исследований число наблюдений существенно меньше.

Как отмечалось в литературе, есть и еще одна общая причина отклонений от нормальности: любой результат наблюдения записывается конечным (обычно 2–5) количеством цифр, а с математической точки зрения вероятность такого события равна 0. Точнее, для случайной величины с непрерывной плотностью распределения вероятность попадания в счетное множество рациональных чисел равна 0. Следовательно, при статистической обработке данных в организационно-экономических исследованиях распределение результатов наблюдений практически всегда более или менее отличается от нормального распределения.

Последствия нарушения условия нормальности. Если условие а) не выполнено, то распределение статистики t не является распределением Стьюдента. Однако можно показать, используя Центральную предельную теорему теории вероятностей и теоремы о наследовании сходимости [21, гл.4], что при справедливости H'_0 и условия б) распределение статистики t при росте объемов выборок приближается к стандартному нормальному распределению $\Phi(x) = N(x; 0, 1)$. К этому же распределению приближается распределение Стьюдента при возрастании числа степеней свободы. Другими словами, несмотря на нарушение условия нормальности традиционный метод (критерий Стьюдента) можно использовать (при определенных условиях!) для проверки гипотезы H'_0 при больших объемах выборок. При этом вместо таблиц распределения Стьюдента достаточно пользоваться таблицами стандартного нормального распределения $\Phi(x)$.

Сформулированное в предыдущем абзаце утверждение справедливо для любых функций распределения $F(x)$ и $G(x)$ таких, что $M(X) = M(Y)$, $D(X) = D(Y)$ и выполнены некоторые внутриматематические условия, обычно считающиеся справедливыми в реальных задачах. Если же $M(X) \neq M(Y)$, то нетрудно вычислить, что при больших объемах выборок

$$P(t \leq x) \approx \Phi(x - a_{mn}), \quad (2)$$

где

$$a_{mn} = \frac{\sqrt{mn}[M(X) - M(Y)]}{\sqrt{mD(X) + nD(Y)}}. \quad (3)$$

Формулы (2) - (3) позволяют приближенно вычислить мощность t -критерия (точность возрастает при увеличении объемов выборок m и n).

О проверке условия равенства дисперсий. Иногда условие б) вытекает из методики получения результатов наблюдений, например, когда с помощью одного и того же прибора или методики m раз измеряют характеристику первого объекта и n раз - второ-

го, а параметры распределения погрешностей измерения при этом не меняются. Однако ясно, что в постановках большинства исследовательских и практических задач нет оснований априори предполагать равенство дисперсий.

Целесообразно ли проверять равенство дисперсий статистическими методами, например, как это иногда предлагают, с помощью F -критерия Фишера? Этот критерий основан на нормальности распределений результатов наблюдений. А от нормальности неизбежны отклонения (см. выше). Причем хорошо известно, что в отличие от t -критерия распределение F -критерия Фишера сильно меняется при малых отклонениях от нормальности [3]. Кроме того, F -критерий отвергает гипотезу $D(X) = D(Y)$ лишь при большом различии выборочных дисперсий. Так, для данных [2] о двух группах результатов химических анализов отношение выборочных дисперсий равно 1,95, т.е. существенно отличается от 1. Тем не менее, гипотеза о равенстве теоретических дисперсий принимается при применении F -критерия на 1%-м уровне значимости. Следовательно, при проверке однородности применение F -критерия для предварительной проверки равенства дисперсий с целью обоснования возможности использования критерия Стьюдента нецелесообразно.

Итак, в большинстве технических, экономических, медицинских и иных задач условие б) нельзя считать выполненным, а проверять его перед проверкой однородности нецелесообразно.

Последствия нарушения условия равенства дисперсий. Если объемы выборок m и n велики, то можно показать, что распределение статистики t описывается с помощью только математических ожиданий $M(X)$ и $M(Y)$, дисперсий $D(X)$, $D(Y)$ и отношения объемов выборок, а именно:

$$P(t \leq x) \approx \Phi(b_{mn}x - a_{mn}), \quad (4)$$

где a_{mn} определено формулой (3),

$$b_{mn}^2 = \frac{\lambda D(X) + D(Y)}{D(X) + \lambda D(Y)}, \quad \lambda = \frac{m}{n}. \quad (5)$$

Если $b_{mn} \neq 1$, то распределение статистики t отличается от распределения, заданного формулой (2), полученной в предположении равенства дисперсий. Когда $b_{mn} = 1$? В двух случаях – при $m = n$ и при $D(X) = D(Y)$. Таким образом, при больших и равных объемах выборок требовать выполнения условия б) нет необходимости. Кроме того, ясно, что если объемы выборок мало различаются, то b_{mn} близко к 1. Так, для данных [2] о двух группах результатов химических анализов имеем $b_{mn}^* = 0,987$, где b_{mn}^* – оценка b_{mn} , полученная заменой в формуле (5) теоретических дисперсий на их выборочные оценки.

Область применимости традиционного метода проверки однородности с помощью критерия Стьюдента. Подведем итоги рассмотрения t -критерия. Он позволяет проверять гипотезу H'_0 о равенстве математических ожиданий, но не гипотезу H_0 о том, что обе выборки взяты из одной и той же генеральной совокупности. Классические условия применимости критерия Стьюдента в подавляющем большинстве технических, экономических, медицинских и иных задач не выполнены. Тем не менее, при больших и примерно равных объемах выборок его можно применять. При конечных объемах выборок традиционный метод носит неустранимо приближенный характер.

Критерий Крамера-Уэлча равенства математических ожиданий. Вместо критерия Стьюдента целесообразно для проверки H'_0 использовать критерий Крамера-Уэлча [12, с.492], основанный на статистике

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{ns_x^2 + ms_y^2}}. \quad (6)$$

Критерий Крамера-Уэлча имеет прозрачный смысл – разность выборочных средних арифметических для двух выборок делится на естественную оценку среднего квадратического отклонения этой разности. Естественность указанной оценки состоит в том, что неизвестные статистику дисперсии заменены их выборочными оценками. Из многомерной центральной предельной теоремы и из теорем о наследовании сходимости [21, гл.4] вытекает, что при росте объемов выборок распределение статистики T Крамера-Уэлча сходится к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1. Итак, при справедливости H'_0 и больших объемах выборок распределение статистики T приближается с помощью стандартного нормального распределения $\Phi(x)$, из таблиц которого и следует брать критические значения.

При $m = n$, как следует из формул (1) и (6), $t = T$. При $m \neq n$ этого равенства нет. В частности, при s_x^2 в формуле (1) стоит множитель $(m - 1)$, в формуле (6) - множитель n .

Если $M(X) \neq M(Y)$, то при больших объемах выборок

$$P(T \leq X) \approx \Phi(x - c_{mn}), \quad (7)$$

где

$$c_{mn} = \frac{\sqrt{mn}[M(X) - M(Y)]}{\sqrt{nD(X) + mD(Y)}}. \quad (8)$$

При $m = n$ или $D(X) = D(Y)$, согласно формулам (3) и (8), $a_{mn} = c_{mn}$, в остальных случаях равенства нет.

Из асимптотической нормальности статистики T , формул (7) и (8) следует, что правило принятия решения для критерия Крамера-Уэлча выглядит так:

- если $|T| \leq \Phi^{-1}(1 - \alpha/2)$, то гипотеза однородности (равенства) математических ожиданий принимается на уровне значимости α ,

- если же $|T| > \Phi^{-1}(1 - \alpha/2)$, то гипотеза однородности (равенства) математических ожиданий отклоняется на уровне значимости α .

В прикладной статистике наиболее часто применяется уровень значимости $\alpha = 0,05$. Тогда значение модуля статистики T Крамера-Уэлча надо сравнивать с граничным значением $\Phi^{-1}(1 - \alpha/2) = 1,96$.

Из сказанного выше следует, что применение критерия Крамера-Уэлча при анализе организационно-экономических данных более обосновано, чем применение критерия Стьюдента. Дополнительное преимущество критерия Крамера-Уэлча по сравнению с критерием Стьюдента - не требуется равенства дисперсий $D(X) = D(Y)$. Распределение статистики T не является распределением Стьюдента, однако и распределение статистики t , как показано выше, не является таковым в реальных ситуациях.

Распределение статистики T при объемах выборок $m = n = 6, 8, 10, 12$ и различных функциях распределений выборок $F(x)$ и $G(x)$ изучено нами совместно с Ю.Э. Камнем и Я.Э. Камнем методом статистических испытаний (Монте-Карло). Рассмотрены различные варианты функций распределения $F(x)$ и $G(x)$. Результаты (частично опубликованы в статье [8]) показывают, что даже при таких небольших объемах выборок точность аппроксимации предельным стандартным нормальным распределением вполне удовлетворительна. Поэтому представляется целесообразным во всех тех случаях, когда в соответствии с устаревшими литературными источниками рекомендуют применять критерий Стьюдента, заменить его на критерий Крамера-Уэлча. Конечно, такая замена потребует переделки ряда нормативно-технических и методических документов, исправления учебников и учебных пособий для вузов.

Пример 1. Пусть объем первой выборки $m = 120$, $\bar{x} = 13,7$, $s_x = 5,3$. Для второй выборки $n = 541$, $\bar{y} = 14,1$, $s_y = 8,4$. Вычислим величину статистики Крамера-Уэлча

$$T = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{ns_x^2 + ms_y^2}} = \frac{\sqrt{120 \times 541}(13,7 - 14,1)}{\sqrt{541 \times 5,3^2 + 120 \times 8,4^2}} = \frac{\sqrt{64920}(-0,4)}{\sqrt{541 \times 28,09 + 120 \times 70,56}} =$$

$$= \frac{254,79 \times (-0,4)}{\sqrt{15196,69 + 8467,2}} = \frac{-101,916}{\sqrt{23663,89}} = \frac{-101,916}{153,83} = -0,66.$$

Поскольку полученное значение по абсолютной величине меньше 1,96, то гипотеза однородности математических ожиданий принимается на уровне значимости 0,05.

Непараметрические методы проверки однородности. В большинстве управленческих, технических, экономических, медицинских и иных задач представляет интерес не проверка равенства математических ожиданий или иных характеристик распределения, а обнаружение различия генеральных совокупностей, из которых извлечены выборки, т.е. проверка гипотезы H_0 . Методы проверки гипотезы H_0 позволяют обнаружить не только изменение математического ожидания, но и любые иные изменения функции распределения результатов наблюдений при переходе от одной выборки к другой (увеличение разброса, появление асимметрии и т. д.). Как установлено выше, методы, основанные на использовании статистик t Стьюдента и T Крамера-Уэлча, не позволяют проверять гипотезу H_0 . Априорное предположение о принадлежности функций распределения $F(x)$ и $G(x)$ к какому-либо определенному параметрическому семейству (например, семействам нормальных, логарифмически нормальных, распределений Вейбулла - Гнеденко, гамма-распределений и др.), как также показано выше, обычно нельзя достаточно надежно обосновать. Поэтому для проверки H_0 следует использовать методы, пригодные при любом виде $F(x)$ и $G(x)$, т.е. непараметрические методы. (Напомним, что термин «непараметрический метод» означает, что при использовании этого метода нет необходимости предполагать, что функции распределения результатов наблюдений принадлежат какому-либо определенному параметрическому семейству.)

Для проверки гипотезы H_0 разработано много непараметрических методов - критерии Смирнова, типа омега-квадрат (Лемана - Розенблатта), Вилкоксона (Манна-Уитни), Ван-дер-Вардена, Сэвиджа, хи-квадрат и др. [5, 21, 24]. Распределения статистик всех этих критериев при справедливости H_0 не зависят от конкретного вида совпадающих функций распределения $F(x) \equiv G(x)$. Следовательно, таблицами точных и предельных (при больших объемах выборок) распределений статистик этих критериев и их процентных точек [21, 24] можно пользоваться при любых непрерывных функциях распределения результатов наблюдений.

Какой из непараметрических критериев применять? Как известно [3], для выбора одного из нескольких критериев необходимо сравнить их мощности, определяемые видом альтернативных гипотез. Сравнению мощностей критериев посвящена обширная литература.

Хорошо изучены свойства критериев при альтернативной гипотезе сдвига

$$H_{1c} : G(x) = F(x-d), d \neq 0.$$

Критерии Вилкоксона, Ван-дер-Вардена и ряд других ориентированы для применения именно в этой ситуации. Если m раз измеряют характеристику одного объекта и n раз - другого, а функция распределения погрешностей измерения произвольна, но не меняется при переходе от объекта к объекту (это более жесткое требование, чем условие равенства дисперсий), то рассмотрение гипотезы H_{1c} оправдано. Однако в большинстве организационно-экономических, технических, медицинских и иных исследо-

ваний нет оснований считать, что функции распределения, соответствующие выборкам, различаются только сдвигом.

3.5.4. Двухвыборочный критерий Вилкоксона

Рассмотрим подробнее часто используемый непараметрический критерий Вилкоксона. В частности, покажем (и это - основной теоретический результат настоящего подраздела), что двухвыборочный критерий Вилкоксона (в литературе его называют также критерием Манна-Уитни) предназначен для проверки гипотезы

$$H_0: P(X < Y) = 1/2,$$

где X - случайная величина, распределенная как элементы первой выборки, а Y - случайная величина, распределенная как элементы второй выборки. Это - непараметрическая гипотеза. Но из нее не следует, что функции распределения двух выборок совпадают. Обратное, конечно, верно: если X и Y одинаково распределены, то $P(X < Y) = 1/2$.

В описанной выше вероятностной модели двух независимых выборок без ограничения общности можно считать, что объем первой из них не превосходит объема второй, $m \leq n$, в противном случае выборки можно поменять местами. Обычно предполагается, что функции $F(x)$ и $G(x)$ непрерывны и строго возрастают. Из непрерывности этих функций следует, что с вероятностью 1 все $m + n$ результатов наблюдений различны. При рассмотрении реальных статистических данных иногда наблюдаются совпадения результатов наблюдений, но сам факт их наличия - свидетельство нарушений предпосылок только что описанной базовой математической модели.

Расчет значения статистики критерия Вилкоксона и правило принятия решений. Статистика S двухвыборочного критерия Вилкоксона определяется следующим образом. Все элементы объединенной выборки $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ упорядочиваются в порядке возрастания. Элементы первой выборки X_1, X_2, \dots, X_m занимают в общем вариационном ряду места с номерами R_1, R_2, \dots, R_m , другими словами, имеют ранги R_1, R_2, \dots, R_m (напомним, что ранг - это номер в упорядоченном ряду). Тогда статистика Вилкоксона - это сумма рангов элементов первой выборки

$$S = R_1 + R_2 + \dots + R_m.$$

Статистика U Манна-Уитни определяется как число пар (X_i, Y_j) таких, что $X_i < Y_j$, среди всех mn пар, в которых первый элемент - из первой выборки, а второй - из второй. Как известно [5, с.160],

$$U = mn + m(m+1)/2 - S.$$

Поскольку S и U линейно связаны, то обычно говорят не о двух критериях - Вилкоксона и Манна-Уитни, а об одном - критерии Вилкоксона (Манна-Уитни).

Критерий Вилкоксона - один из самых известных инструментов непараметрической статистики (наряду с критериями на основе статистик типа Колмогорова-Смирнова, омега-квадрат и коэффициентами ранговой корреляции). Свойствам этого критерия и таблицам его критических значений уделяется место во многих монографиях по математической и прикладной статистике (см., например, [2, 5, 21, 24]).

Однако в литературе имеются и неточные утверждения относительно возможностей критерия Вилкоксона. Так, отдельные авторы полагают, что с его помощью можно обнаружить любое различие между функциями распределения $F(x)$ и $G(x)$. По мнению других, этот критерий нацелен на проверку равенства медиан распределений, соответствующих выборкам. И то, и другое неверно. Это будет ясно из дальнейшего изложения.

Введем некоторые обозначения. Пусть $F^{-1}(t)$ - функция, обратная к функции распределения $F(x)$. Она определена на отрезке $[0; 1]$. Положим $L(t) = G(F^{-1}(t))$. Поскольку $F(x)$ непрерывна и строго возрастает, то $F^{-1}(t)$ и $L(t)$ обладают теми же свойствами.

Важную роль в дальнейшем изложении будет играть величина $a = P(X < Y)$. Как нетрудно показать,

$$a = P(X < Y) = \int_0^1 t dL(t).$$

Введем также параметры

$$b^2 = \int_0^1 L^2(t) dt - (1-a)^2, \quad g^2 = \int_0^1 t^2 dL(t) - a^2.$$

Тогда математические ожидания и дисперсии статистик Вилкоксона и Манна-Уитни согласно [5, с.160] выражаются через введенные величины:

$$\begin{aligned} M(U) &= mna, \quad M(S) = mn + m(m+1)/2 - M(U) = mn(1-a) + m(m+1)/2, \\ D(S) &= D(U) = mn[(n-1)b^2 + (m-1)g^2 + a(1-a)]. \end{aligned} \quad (9)$$

Когда объемы обеих выборок безгранично растут, распределения статистик Вилкоксона и Манна-Уитни являются асимптотически нормальными (см., например, [5, гл. 5 и 6]) с параметрами, задаваемыми формулами (9).

Если выборки полностью однородны, т.е. их функции распределения совпадают, другими словами, справедлива гипотеза

$$H_0: F(x) = G(x) \text{ при всех } x, \quad (10)$$

то $L(t) = t$ для t из отрезка $[0, 1]$, $L(t) = 0$ для всех отрицательных t и $L(t) = 1$ для $t > 1$, соответственно $a = 1/2$. Подставляя в формулы (9), получаем, что

$$M(S) = m(m+n+1)/2, \quad D(S) = mn(m+n+1)/12. \quad (11)$$

Следовательно, распределение нормированной и центрированной статистики Вилкоксона

$$T = (S - m(m+n+1)/2) (mn(m+n+1)/12)^{-1/2} \quad (12)$$

при росте объемов выборок приближается к стандартному нормальному распределению (с математическим ожиданием 0 и дисперсией 1).

Из асимптотической нормальности статистики T следует, что при больших объемах выборок правило принятия решения для критерия Вилкоксона выглядит так:

- если $|T| \leq \Phi^{-1}(1 - \alpha/2)$, то гипотеза (2.10) однородности (тождества) функций распределений принимается на уровне значимости α ,

- если же $|T| > \Phi^{-1}(1 - \alpha/2)$, то гипотеза (2.10) однородности (тождества) функций распределений отклоняется на уровне значимости α .

В прикладной статистике наиболее часто применяется уровень значимости $\alpha = 0,05$. Тогда значение модуля статистики T Вилкоксона надо сравнивать с граничным значением $\Phi^{-1}(1 - \alpha/2) = 1,96$.

Пример 1. Пусть даны две выборки. Первая содержит $m = 12$ элементов 17; 22; 3; 5; 15; 2; 0; 7; 13; 97; 66; 14. Вторая содержит $n = 14$ элементов 47; 30; 2; 15; 1; 21; 25; 7; 44; 29; 33; 11; 6; 15. Проведем проверку однородности функций распределения двух выборок с помощью критерия Вилкоксона.

Первым шагом является построение общего вариационного ряда для элементов двух выборок (табл.2). Общий вариационный ряд – в средней строке. Ниже для каждого его элемента указано, из какой выборки он взят. Построение верхней строки «Ранги» описано далее.

Хотя с точки зрения теории математической статистики вероятность совпадения двух элементов выборок равна 0, в реальных выборках статистических данных совпадения встречаются. Так, в рассматриваемых выборках, как видно из табл.2, два раза повторяется величина 2, два раза - величина 7 и три раза - величина 15. В таких случаях говорят о наличии «связанных рангов», а соответствующим совпадающим величинам

приписывают среднее арифметическое тех рангов, которые они занимают (т.е. среднее арифметическое номеров тех мест, которые они занимают в общем вариационном ряду). Так, величины 2 (из первой выборки) и 2 (из второй выборки) занимают в объединенном вариационном ряду места 3 и 4, поэтому им приписывается ранг $(3+4)/2=3,5$. Величины 7 и 7 занимают в объединенном вариационном ряду места 8 и 9, поэтому им приписывается ранг $(8+9)/2=8,5$. Величины 15, 15 и 15 занимают в объединенном вариационном ряду места 13, 14 и 15, поэтому им приписывается ранг $(13+14+15)/3=14$.

Таблица 2. – Общий вариационный ряд для элементов двух выборок

Ранги	1	2	3,5	3,5	5	6	7	8,5	8,5	10	11	12	14
Элементы выборки	0	1	2	2	3	5	6	7	7	11	13	14	15
Номера выборок	1	2	1	2	1	1	2	1	2	2	1	1	1
Ранги	14	14	16	17	18	19	20	21	22	23	24	25	26
Элементы выборки	15	15	17	21	22	25	29	30	33	44	47	66	97
Номера выборок	2	2	1	2	1	2	2	2	2	2	2	1	1

Таким образом, после построения объединенного вариационного ряда выделяют группы «связанных рангов» и проводят описанные выше расчеты. В итоге получают строку «Ранги».

Следующий шаг - подсчет значения статистики Вилкоксона, т.е. суммы рангов элементов первой выборки

$$S = R_1 + R_2 + \dots + R_m = 1+3,5+5+6+8,5+11+12+14+16+18+25+26=146.$$

Подсчитаем также сумму рангов элементов второй выборки

$$S_1 = 2+3,5+7+8,5+10+14+14+17+19+20+21+22+23+24= 205.$$

Величина S_1 может быть использована для контроля вычислений. Дело в том, что суммы рангов элементов первой выборки S и второй выборки S_1 вместе составляют сумму рангов объединенной выборки, т.е. сумму всех натуральных чисел от 1 до $m+n$. Следовательно,

$$S + S_1 = (m+n)(m+n+1)/2 = (12+14)(12+14+1)/2 = 351.$$

В соответствии с ранее проведенными расчетами $S+S_1 = 146 + 205 = 351$. Необходимое условие правильности расчетов выполнено. Ясно, что справедливость этого условия не гарантирует правильности расчетов.

Перейдем к расчету статистики T . Согласно формуле (11)

$$M(S) = 12(12+14+1)/2 = 162, D(S) = 12 \cdot 14(12+14+1)/12 = 378.$$

Следовательно,

$$T = (S - 162) / (378)^{1/2} = (146-162) / 19,44 = - 0,82.$$

Поскольку $|T| \leq 1,96$, то гипотеза однородности принимается на уровне значимости 0,05.

Что будет, если поменять выборки местами, вторую назвать первой? Тогда вместо S надо рассматривать S_1 . Имеем

$$M(S_1) = 14(12+14+1)/2 = 189, D(S) = D(S_1) = 378,$$

$$T_1 = (S_1 - 189) / (378)^{1/2} = (205-189) / 19,44 = 0,82.$$

Таким образом, значения статистики критерия отличаются только знаком (можно показать, что это утверждение верно всегда). Поскольку в правиле принятия решения используется только абсолютная величина статистики, то принимаемое решение не зависит от того, какую выборку считаем первой, а какую второй. Для уменьшения объема таблиц критических значений принято считать первой выборку меньшего объема.

Мощность критерия Вилкоксона. Продолжим обсуждение критерия Вилкоксона. Правила принятия решений и таблицы критических значений для критерия Вилкоксона строятся в предположении справедливости гипотезы полной однородности, описываемой формулой (10). А что будет, если эта гипотеза неверна? Другими словами, какова мощность критерия Вилкоксона?

Пусть объемы выборок достаточно велики, так что можно пользоваться асимптотической нормальностью статистики Вилкоксона. Тогда в соответствии с формулами (9) статистика T будет асимптотически нормальна с параметрами

$$M(T) = (12mn)^{1/2} (1/2 - a) (m+n+1)^{-1/2}, \\ D(T) = 12 [(n-1)b^2 + (m-1)g^2 + a(1-a)] (m+n+1)^{-1}. \quad (13)$$

Из формул (13) видно большое значение гипотезы

$$H_{01}: a = P(X < Y) = 1/2. \quad (14)$$

Если эта гипотеза неверна, то, поскольку $m \leq n$, справедлива оценка

$$|M(T)| \geq (12mn(2n+1)^{-1})^{1/2} |1/2 - a|,$$

а потому $|M(T)|$ безгранично растет при росте объемов выборок. В то же время, поскольку

$$b^2 \leq \int_0^1 L^2(t)dt \leq 1, \quad g^2 \leq \int_0^1 t^2 dL(t) \leq 1, \quad \alpha(1-\alpha) \leq 1/4,$$

то

$$D(T) \leq 12 [(n-1) + (m-1) + 1/4] (m+n+1)^{-1} \leq 12. \quad (15)$$

Следовательно, вероятность отклонения гипотезы H_{01} , когда она неверна, т.е. мощность критерия Вилкоксона как критерия проверки гипотезы (14), стремится к 1 при возрастании объемов выборок, т.е. критерий Вилкоксона является состоятельным для этой гипотезы при альтернативе

$$AH_{01}: a = P(X < Y) \neq 1/2. \quad (16).$$

Если же гипотеза (14) верна, то статистика T асимптотически нормальна с математическим ожиданием 0 и дисперсией, определяемой формулой

$$D(T) = 12[(n-1)b^2 + (m-1)g^2 + 1/4](m+n+1)^{-1}. \quad (17)$$

Гипотеза (14) является сложной, дисперсия (17), как показывают приводимые ниже примеры, в зависимости от значений b^2 и g^2 может быть как больше 1, так и меньше 1, но согласно неравенству (15) никогда не превосходит 12.

Критерий Вилкоксона не позволяет проверять абсолютную однородность. Приведем пример двух функций распределения $F(x)$ и $G(x)$ таких, что гипотеза (14) выполнена, а гипотеза (10) - нет. Поскольку

$$a = P(X < Y) = \int_{-\infty}^{+\infty} F(x)dG(x), \quad 1 - a = P(Y < X) = \int_{-\infty}^{+\infty} G(x)dF(x) \quad (18)$$

и $a = 1/2$ в случае справедливости гипотезы (10), то для выполнения условия (14) необходимо и достаточно, чтобы

$$\int_{-\infty}^{+\infty} (F(x) - G(x))dF(x) = 0, \quad (19),$$

а потому естественно в качестве $F(x)$ рассмотреть функцию равномерного распределения на интервале $(-1; 1)$. Тогда формула (19) переходит в условие

$$\int_{-\infty}^{+\infty} (F(x) - G(x))dF(x) = -\frac{1}{2} \int_{-1}^{+1} \left(G(x) - \frac{(x+1)}{2} \right) dx = 0.$$

Это условие выполняется, если функция $(G(x) - (x+1)/2)$ является нечетной.

Пример 2. Пусть функции распределения $F(x)$ и $G(x)$ сосредоточены на интервале $(-1; 1)$, на котором

$$F(x) = (x + 1)/2, G(x) = (x + 1 + 1/\pi \sin \pi x)/2.$$

Тогда

$$x = F^{-1}(t) = 2t - 1, L(t) = G(F^{-1}(t)) = (2t + 1/\pi \sin \pi(2t - 1))/2 = t + 1/2 \pi \sin \pi(2t - 1).$$

Условие (19) выполнено, поскольку функция $(G(x) - (x + 1)/2)$ является нечетной. Следовательно, $a = 1/2$. Начнем с вычисления

$$g^2 = \int_0^1 t^2 dL(t) - 1/4 = \int_0^1 t^2 d\left(t + \frac{1}{2\pi} \sin \pi(2t - 1)\right) - \frac{1}{4}.$$

Поскольку

$$d\left(t + \frac{1}{2\pi} \sin \pi(2t - 1)\right) = (1 + \cos \pi(2t - 1))dt,$$

то

$$g^2 = \int_0^1 t^2 (1 + \cos \pi(2t - 1))dt - \frac{1}{4} = \frac{1}{12} + \int_0^1 t^2 \cos \pi(2t - 1)dt.$$

С помощью замены переменных $t = (x + 1) / 2$ получаем, что

$$\int_0^1 t^2 \cos \pi(2t - 1)dt = \frac{1}{8} \left(\int_{-1}^1 x^2 \cos \pi x dx + 2 \int_{-1}^1 x \cos \pi x dx + \int_{-1}^1 \cos \pi x dx \right).$$

В правой части последнего равенства стоят табличные интегралы (см., например, справочник [23, с.71]). Проведя соответствующие вычисления, получаем, что в правой части стоит $1/8 \times (-4/\pi^2) = -1/(2\pi^2)$. Следовательно,

$$g^2 = 1/12 - 1/(2\pi^2) = 0,032672733...$$

Перейдем к вычислению b^2 . Поскольку

$$b^2 = \int_0^1 L^2(t)dt - \frac{1}{4} = \int_0^1 \left(t + \frac{1}{2} \pi \sin \pi(2t - 1) \right)^2 dt - \frac{1}{4},$$

то

$$b^2 = \frac{1}{12} + \frac{1}{\pi} \int_0^1 (t \sin \pi(2t - 1))dt + \left(\frac{\pi}{2} \right)^2 \int_0^1 \sin^2 \pi(2t - 1)dt.$$

С помощью замены переменных $t = (x+1)/2$ переходим к табличным интегралам (см., например, справочник [23, с.65]):

$$b^2 = \frac{1}{12} + \frac{1}{4\pi} \int_{-1}^1 x \sin \pi x dx + \frac{1}{4\pi} \int_{-1}^1 \sin \pi x dx + \frac{1}{8\pi^2} \int_{-1}^1 \sin^2 \pi x dx.$$

Проведя необходимые вычисления, получим, что

$$b^2 = \frac{1}{12} + \frac{1}{4\pi} \left(-\frac{2}{\pi} \right) + 0 + \frac{1}{8\pi^2} = \frac{1}{12} - \frac{3}{8\pi^2} = 0,045337893...$$

Следовательно, для рассматриваемых функций распределения нормированная и центрированная статистика Вилкоксона (см. формулу (12)) асимптотически нормальна с математическим ожиданием 0 и дисперсией (см. формулу (17))

$$D(T) = (0,544 n + 0,392 m + 2,064) (m+n+1)^{-1}.$$

Как легко видеть, дисперсия всегда меньше 1. Это значит, что в рассматриваемом случае гипотеза полной однородности (10) при проверке с помощью критерия Вилкоксона будет приниматься чаще, чем если она на самом деле верна.

Сказанное означает, что критерий Вилкоксона нельзя считать критерием для проверки гипотезы (10) при альтернативе общего вида. Он не всегда позволяет проверить однородность - не при всех альтернативах. Точно так же критерии типа хи-квадрат нельзя считать критериями проверки гипотез согласия и однородности в случае непре-

рывных распределений - они позволяют обнаружить не все различия, поскольку некоторые из них «скрадывает» группировка.

Критерий Вилкоксона не позволяет проверять равенство медиан. Обсудим теперь, действительно ли критерий Вилкоксона нацелен на проверку равенства медиан распределений, соответствующих выборкам.

Пример 3. Построим семейство пар функций распределения $F(x)$ и $G(x)$ таких, что их медианы различны, но для $F(x)$ и $G(x)$ выполнена гипотеза (14). Пусть распределения сосредоточены на интервале $(0; 1)$, и на нем $G(x) = x$, а $F(x)$ имеет кусочно-линейный график с вершинами в точках $(0; 0)$, $(\lambda, 1/2)$, $(\delta, 3/4)$, $(1; 1)$. Следовательно,

$$\begin{aligned} F(x) &= 0 \text{ при } x < 0; \\ F(x) &= x/(2\lambda) \text{ на } [0; \lambda); \\ F(x) &= 1/2 + (x - \lambda)/(4\delta - 4\lambda) \text{ на } [\lambda; \delta); \\ F(x) &= 3/4 + (x - \delta)/(4 - 4\delta) \text{ на } [\delta; 1]; \\ F(x) &= 1 \text{ при } x > 1. \end{aligned}$$

Очевидно, что медиана $F(x)$ равна λ , а медиана $G(x)$ равна $1/2$.

Согласно соотношению (17) для выполнения гипотезы (14) достаточно определить δ как функцию λ , т.е. $\delta = \delta(\lambda)$, из условия

$$\int_0^1 F(x) dx = \frac{1}{2}.$$

Вычисления дают

$$\delta = \delta(\lambda) = 3(1 - \lambda)/2.$$

Учитывая, что δ лежит между λ и 1, не совпадая ни с тем, ни с другим, получаем ограничения на λ , а именно, $1/3 < \lambda < 3/5$. Итак, построено искомое семейство пар функций распределения.

Пример 4. Пусть, как и в примере 3, распределения сосредоточены на интервале $(0; 1)$, и на нем $F(x)=x$. А $G(x)$ - функция распределения, сосредоточенная в двух точках - β и 1. Т.е. $G(x) = 0$ при x , не превосходящем β ; $G(x) = h$ на $(\beta; 1]$; $G(x) = 1$ при $x > 1$. С такой функцией $G(x)$ легко проводить расчеты. Однако она не удовлетворяет принятым выше условиям непрерывности и строгого возрастания. Вместе с тем легко видеть, что она является предельной (сходимостью в каждой точке отрезка $[0; 1]$) для последовательности функций распределения, удовлетворяющих этим условиям. А распределение статистики Вилкоксона для пары функций распределения примера 4 является предельным для последовательности соответствующих распределений статистики Вилкоксона, полученных в рассматриваемых условиях непрерывности и строгого возрастания.

Условие $P(X < Y) = 1/2$ выполнено, если $h = (1 - \beta)^{-1}/2$ (при β из отрезка $[0; 1/2]$). Поскольку $h > 1/2$ при положительном β , то очевидно, что медиана $G(x)$ равна β , в то время как медиана $F(x)$ равна $1/2$. Значит, при $\beta = 1/2$ медианы совпадают, при всех иных положительных β - различны. При $\beta = 0$ медианой $G(x)$ является любая точка из отрезка $[0; 1]$.

Легко подсчитать, что в условиях примера 4 параметры предельного распределения имеют вид

$$b^2 = \beta(1 - \beta)^{-1}/4, \quad g^2 = (1 - 2\beta)/4.$$

Следовательно, распределение нормированной и центрированной статистики Вилкоксона будет асимптотически нормальным с математическим ожиданием 0 и дисперсией

$$D(T) = 3 [(n-1) \beta(1 - \beta)^{-1} + (m-1)(1 - 2\beta) + 1] (m+n+1)^{-1}.$$

Проанализируем величину $D(T)$ в зависимости от параметра β и объемов выборок m и n . При достаточно больших m и n

$$D(T) = 3w\beta(1-\beta)^{-1} + 3(1-w)(1-2\beta)$$

с точностью до величин порядка $(m+n)^{-1}$, где $w = n/(m+n)$. Значит, $D(T)$ - линейная функция от w , а потому достигает экстремальных значений на границах интервала изменения w , т.е. при $w = 0$ и $w = 1$. Легко видеть, что при $\beta(1-\beta)^{-1} < 1-2\beta$ минимум равен $3\beta(1-\beta)^{-1}$ (при $w = 1$), а максимум равен $3(1-2\beta)$ (при $w = 0$). В случае $\beta(1-\beta)^{-1} > 1-2\beta$ максимум равен $3\beta(1-\beta)^{-1}$ (при $w = 1$), а минимум равен $3(1-2\beta)$ (при $w = 0$). Если же $\beta(1-\beta)^{-1} = 1-2\beta$ (это равенство справедливо при $\beta = \beta_0 = 1 - 2^{-1/2} = 0,293$), то $D(T) = 3(2^{1/2}-1) = 1,2426\dots$ при всех w из отрезка $[0; 1]$.

Первый из описанных выше случаев имеет быть при $\beta < \beta_0$. При этом минимум $D(T)$ возрастает от 0 (при $\beta=0, w=1$ - предельный случай) до $3(2^{1/2}-1)$ (при $\beta=\beta_0, w$ - любым), а максимум уменьшается от 3 (при $\beta=0, w=0$ - предельный случай) до $3(2^{1/2}-1)$ (при $\beta=\beta_0, w$ - любым). Второй случай относится к β из интервала $(\beta_0; 1/2]$. При этом минимум убывает от приведенного выше значения для $\beta=\beta_0$ до 0 (при $\beta=1/2, w=0$ - предельный случай), а максимум возрастает от того же значения при $\beta=\beta_0$ до 3 (при $\beta=1/2, w=0$).

Таким образом, $D(T)$ может принимать все значения из интервала $(0; 3)$ в зависимости от значений β и w . Если $D(T) < 1$, то при применении критерия Вилкоксона к выборкам с рассматриваемыми функциями распределения гипотеза однородности (10) будет приниматься чаще (при соответствующих значениях β и w - с вероятностью, сколь угодно близкой к 1), чем если бы она самом деле была верна. Если $1 < D(T) < 3$, то гипотеза (10) также принимается достаточно часто. Так, если уровень значимости критерия Вилкоксона равен 0,05, то (асимптотическая) критическая область этого критерия, как показано выше, имеет вид $\{T: |T| \geq 1,96\}$. Если - самый плохой случай - $D(T) = 3$, то гипотеза (10) принимается с вероятностью 0,7422.

Гипотеза сдвига. При проверке гипотезы однородности мы рассмотрели различные виды нулевых и альтернативных гипотез - гипотезу (10) и ее отрицание в качестве альтернативы, гипотезу (14) и ее отрицание, гипотезы о равенстве или различии медиан. В теоретических работах по математической статистике часто рассматривают гипотезу сдвига, в которой альтернативой гипотезе (10) является гипотеза

$$H_1: F(x) = G(x+r) \quad (20)$$

при всех x и некотором сдвиге r , отличным от 0. Если верна альтернативная гипотеза H_1 , то вероятность $P(X < Y)$ отлична от 1/2, а потому при альтернативе (20) критерий Вилкоксона является состоятельным.

В некоторых прикладных постановках гипотеза (20) представляется естественной. Например, если одним и тем же прибором проводятся две серии измерений двух значений некоторой величины (физической, химической и т.п.). При этом функция распределения $G(x)$ описывает погрешности измерения одного значения, а $G(x+r)$ - другого. Вопреки распространенному заблуждению, хорошо известно, что распределение погрешностей измерений, как правило, не является нормальным (см. об этом раздел 3.3, а также [21, гл.5.1], [20, гл.4.1]). Однако при анализе подавляющего большинства конкретных статистических данных, как правило, нет никаких оснований считать, что отсутствие однородности всегда выражается столь однозначным образом, как следует из формулы (20). Поэтому для проверки однородности необходимо использовать статистические критерии, состоятельные против любого отклонения от гипотезы однородности (10), а не только против альтернативы сдвига.

Почему же математики так любят гипотезу сдвига (20)? Да потому, что она дает возможность доказывать глубокие математические результаты, например, об асимптотической оптимальности критериев. К сожалению, с точки зрения организационно-экономического моделирования это напоминает поиск ключей под фонарем, где светло, а не в кустах, где они потеряны.

Отметим еще одно обстоятельство. Часто говорят (в соответствии с классическим подходом математической статистики), что нельзя проверять нулевые гипотезы без рассмотрения альтернативных. Однако при анализе данных, полученных в ходе организационно-экономических, технических, медицинских или иных исследований, зачастую полностью ясна формулировка той гипотезы, которую желательно проверить (например, гипотезы абсолютной (иногда говорят, полной) однородности - см. формулу (10)), в то время как формулировка альтернативной гипотезы не очевидна. То ли это гипотеза о неверности равенства (10) хотя бы для одного значения x , то ли это альтернатива (16), то ли - альтернатива сдвига (20), и т. д. В таких случаях целесообразно «обернуть» постановку задачи - исходя из статистического критерия найти альтернативы, относительно которых он состоятелен. Именно это и проделано в настоящем подразделе для критерия Вилкоксона.

Подведем итоги рассмотрения критерия Вилкоксона.

1. Критерий Вилкоксона (Манна-Уитни) является одним из самых распространенных непараметрических ранговых критериев, используемых для проверки однородности двух выборок. Его значение не меняется при любом монотонном преобразовании шкалы измерения (т.е. он пригоден для статистического анализа данных, измеренных в порядковой шкале).

2. Распределение статистики критерия Вилкоксона определяется функциями распределения $F(x)$ и $G(x)$ и объемами m и n двух выборок. При больших объемах выборок распределение статистики Вилкоксона является асимптотически нормальным с параметрами, выписанными выше (см. формулы (9), (11) и (13)).

3. При альтернативной гипотезе, когда функции распределения выборок $F(x)$ и $G(x)$ не совпадают, распределение статистики Вилкоксона зависит от величины вероятности $a = P(X < Y)$. Если a отличается от $1/2$, то мощность критерия Вилкоксона стремится к 1, и он отличает нулевую гипотезу $F \equiv G$ от альтернативной. Если же $a = 1/2$, то это не всегда имеет место. В примере 2 приведены две *различные* функции распределения выборок $F(x)$ и $G(x)$ такие, что гипотеза однородности $F \equiv G$ при проверке с помощью критерия Вилкоксона будет приниматься *чаще*, чем если бы она на самом деле была верна.

4. Следовательно, в случае общей альтернативы критерий Вилкоксона не является состоятельным, т.е. не всегда позволяет обнаружить различие функций распределения. Однако это не лишает его практической ценности, точно так же, как несостоятельность критериев типа хи-квадрат при проверке согласия, независимости или однородности не мешает отклонять нулевую гипотезу во многих практически важных случаях. Однако принятие нулевой гипотезы с помощью критерия Вилкоксона может означать не совпадение F и G , а всего лишь выполнение равенства $a = 1/2$.

5. Иногда утверждают, что с помощью критерия Вилкоксона можно проверять равенство медиан функций распределения F и G . Это не так. В примерах 3 и 4 указаны функции распределения F и G с $a = 1/2$, но с различными медианами. Во многих случаях это различие нельзя обнаружить с помощью критерия Вилкоксона, как это показано при численном анализе асимптотической дисперсии в примере 4.

6. Указанные выше недостатки критерия Вилкоксона исчезают для специального вида альтернативы - т.н. «альтернативы сдвига» $H_1: F(x) = G(x + r)$. В этом частном случае при справедливости альтернативной гипотезы мощность стремится к 1, разли-

чие медиан также всегда обнаруживается. Однако альтернатива сдвига не всегда естественна. Ее целесообразно принять, если одним и тем же прибором проводятся две серии измерений двух значений некоторой величины (физической, химической и т.п.). При этом функция распределения $G(x)$ описывает результаты измерений (с погрешностями) одного значения, а $F(x) = G(x+r)$ - другого. Другими словами, меняется лишь измеряемое значение, а собственно распределение погрешностей - одно и то же, присущее используемому средству измерения (и обычно описанное в его техническом паспорте). Однако в большинстве прикладных статистических исследований нет никаких оснований считать, что при альтернативе функция распределения второй выборки лишь сдвигается, но не меняется каким-либо иным образом.

7. При всех своих недостатках критерий Вилкоксона прост в применении и часто позволяет обнаруживать различие групп (поскольку оно часто сводится к отличию $a = P(X < Y)$ от $1/2$). Приведенные здесь критические замечания не следует понимать как призыв к полному отказу от использования критерия Вилкоксона. Однако для проверки гипотезы однородности в случае альтернативы общего вида можно порекомендовать состоятельные критерии, в частности, рассматриваемые в следующем разделе критерии Смирнова и типа омега-квадрат (Лемана-Розенблатта).

8. В литературе по прикладным статистическим методам соседствуют два стиля изложения. Один из них исходит из формулировок нулевой и альтернативных гипотез (или описания набора гипотез, из которого надо выбрать наиболее адекватную), для проверки которых строятся те или иные критерии. При другом стиле изложения упор делается на алгоритмическое описание критериев для проверки тех или иных гипотез, а об альтернативах даже не упоминается.

Например, в литературе по математической статистике часто говорится, что для проверки нормальности используются критерии асимметрии и эксцесса (они описаны, например, в лучшем справочнике 1960–1980-х годов [2, табл. 4.7]). Однако эти критерии позволяют проверять некоторые соотношения между моментами распределения, но отнюдь не являются состоятельными критериями нормальности (не все отклонения от нормальности обнаруживают). Впрочем, для прикладной статистики эти критерии большого практического значения не имеют, поскольку заранее известно, что распределения конкретных технических, экономических, медицинских и иных статистических данных скорее всего отличны от нормальных.

Так что недостатки критерия Вилкоксона не является исключением, мощность ряда иных популярных в математической статистике критериев заслуживает тщательного изучения, при этом заранее можно сказать, что зачастую они не позволяют проверять те гипотезы, с которыми традиционно связаны. При применении подобных критериев к анализу реальных данных необходимо тщательно взвешивать их достоинства и недостатки.

В организационно-экономических исследованиях начинать следует с построения вероятностно-статистической модели, формулировки в ее терминах проверяемых гипотез. Лишь на основе подобной модели можно изучить свойства тех или иных методов и алгоритмов обработки данных. За статистическим критерием всегда стоит вероятностно-статистическая модель порождения данных, определяющая его свойства.

3.5.5. Состоятельные критерии проверки однородности независимых выборок

В соответствии с методологией организационно-экономического моделирования естественно потребовать, чтобы рекомендуемый для массового использования в управленческих, экономических, технических, медицинских и иных исследованиях критерий однородности был состоятельным. Напомним: это значит, что для любых отличных

друг от друга функций распределения $F(x)$ и $G(x)$ (другими словами, при справедливости альтернативной гипотезы H_1) вероятность отклонения гипотезы H_0 должна стремиться к 1 при увеличении объемов выборок m и n . Из перечисленных выше (в конце раздела 3.5.3) критериев однородности состоятельными являются только критерии Смирнова и типа омега-квадрат.

Проведенное в Институте высоких статистических технологий и эконометрики исследование мощности (методом статистических испытаний) первых четырех из перечисленных выше критериев (при различных вариантах функций распределения $F(x)$ и $G(x)$) подтвердило преимущество критериев Смирнова и омега-квадрат и при малых объемах выборок 6–12. Рассмотрим эти критерии подробнее.

Критерий Смирнова однородности двух независимых выборок. Он был предложен членом-корреспондентом АН СССР Н. В. Смирновым в 1939 г. (см. справочник [2]). Единственное ограничение - функции распределения $F(x)$ и $G(x)$ должны быть непрерывными. Напомним, что согласно Л. Н. Большеву и Н. В. Смирнову [2] значение эмпирической функции распределения в точке x равно доле результатов наблюдений в выборке, меньших x . Критерий Смирнова основан на использовании эмпирических функций распределения $F_m(x)$ и $G_n(x)$, построенных по первой и второй выборкам соответственно. Значение статистики Смирнова

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|$$

сравнивают с соответствующим критическим значением (см., например, [2]) и по результатам сравнения принимают или отклоняют гипотезу H_0 о совпадении (однородности) функций распределения. Практически значение статистики $D_{m,n}$ рекомендуется согласно монографии [2] вычислять по формулам

$$D_{m,n}^+ = \max_{1 \leq r \leq n} \left[\frac{r}{n} - F_m(y_r) \right] = \max_{1 \leq s \leq m} \left[G_n(x'_s) - \frac{s-1}{m} \right], \quad (21)$$

$$D_{m,n}^- = \max_{1 \leq r \leq n} \left[F_m(y_r) - \frac{r-1}{n} \right] = \max_{1 \leq s \leq m} \left[\frac{s}{m} - G_n(x'_s) \right], \quad (22)$$

$$D_{m,n} = \max(D_{m,n}^+, D_{m,n}^-), \quad (23)$$

где $x'_1 < x'_2 < \dots < x'_m$ - элементы первой выборки x_1, x_2, \dots, x_m , переставленные в порядке возрастания, а $y'_1 < y'_2 < \dots < y'_n$ - элементы второй выборки y_1, y_2, \dots, y_n , также переставленные в порядке возрастания. Поскольку функции распределения $F(x)$ и $G(x)$ предполагаются непрерывными, то вероятность совпадения каких-либо выборочных значений равна 0.

Пример 1. Рассчитаем значение статистики двухвыборочного критерия Смирнова для тех же выборок, для которых в предыдущем разделе было рассчитано значение статистики критерия Вилкоксона. Первая из них содержит $m = 12$ элементов. Переставим их в порядке возрастания $0 < 2 < 3 < 17 < 5 < 7 < 13 < 14 < 15 < 22 < 66 < 97$. Вторая содержит $n = 14$ элементов. Также переставим их в порядке возрастания $1 < 2 < 6 < 7 < 11 < 15 < 15 < 21 < 25 < 29 < 30 < 33 < 44 < 47$. Точнее, в порядке неубывания, поскольку два элемента совпадают. С точки зрения теории вероятность совпадения двух элементов равна 0, но из-за неизбежных округлений эта вероятность положительна. Поскольку совпадений мало (как внутри одной выборки, так и для элементов разных выборок), то использование теории, основанной на нулевой вероятности совпадения элементов выборок, является допустимым.

Расчет значений статистик представлен в табл.3 ($D_{m,n}^+$) и табл.4 ($D_{m,n}^-$).

Беря максимум по столбцу (6) табл.3, получаем, что $D_{m,n}^+ = 0,167$. Таков же максимум и по столбцу (9), как и должно быть в соответствии с приведенным выше равенст-

вом. Максимум по столбцу (6) табл.4 равен 0,262, в то время как максимум по столбцу (9) той же таблицы есть 0,312. Это различие вызвано тем, что некоторые выборочные значения совпадают, а потому равенство (22), справедливое при отсутствии совпадений, не выполняется. В таких случаях рекомендуют брать максимальное из полученных двумя способами значений, т.е. следует положить $D_{m,n}^- = \max(0,262; 0,312) = 0,312$. По формуле (23) двухвыборочная статистика Смирнова $D_{m,n} = \max(0,167; 0,312) = 0,312$.

Таблица 3. – Расчет значения статистики $D_{m,n}^+$

№ п/п	Элементы выборки x	Номера выборки	$F_m(x)$	r/n	$r/n - F_m(x)$	$G_n(x)$	$\frac{s-1}{m}$	$G_n(x) - (s-1)/m$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	0	1	0			0	0	0
2	1	2	0,083	0,071	-0,012	0		
3	2	1	0,083			0,071	0,083	-0,012
4	2	2	0,083	0,143	0,06	0,071		
5	3	1	0,167			0,143	0,167	-0,024
6	5	1	0,25			0,143	0,25	-0,107
7	6	2	0,333	0,214	-0,119	0,143		
8	7	1	0,333			0,214	0,333	-0,119
9	7	2	0,333	0,286	-0,047	0,214		
10	11	2	0,417	0,357	-0,06	0,286		
11	13	1	0,417			0,357	0,417	-0,06
12	14	1	0,5			0,357	0,5	
13	15	1	0,583			0,357	0,583	-0,226
14	15	2	0,583	0,429	-0,154	0,357		
15	15	2	0,583	0,5	-0,083	0,357		
16	17	1	0,667			0,5	0,667	-0,167
17	21	2	0,75	0,571	-0,179	0,5		
18	22	1	0,75			0,571	0,75	-0,179
19	25	2	0,833	0,643	-0,19	0,571		
20	29	2	0,833	0,714	-0,119	0,643		
21	30	2	0,833	0,786	-0,047	0,714		
22	33	2	0,833	0,857	0,024	0,786		
23	44	2	0,833	0,929	0,096	0,857		
24	47	2	0,833	1,0	0,167	0,929		
25	66	1	0,833			1,0	0,883	0,167
26	97	1	0,917			1,0	0,883	0,167

В табл.6.5а справочника [2] приведены критические значения для двухвыборочной статистики Смирнова, соответствующие обычно используемым уровням значимости (см. табл.5). Поскольку полученное по статистическим данным значение меньше критического значения для уровня значимости $\alpha = 0,1$, а потому и для всех остальных рассматриваемых уровней значимости, то нет оснований отклонять гипотезу однородности. Как и при использовании критерия Вилкоксона, эффект не обнаружен, нулевую гипотезу абсолютной однородности принимаем.

Разработаны алгоритмы и программы для ЭВМ, позволяющие рассчитывать точные распределения, процентные точки и достигаемый уровень значимости для двухвыборочной статистики Смирнова $D_{m,n}$, рассчитаны подробные таблицы (см., например, методику [14], содержащую описание алгоритмов, тексты программ и подробные таблицы).

Таблица 4. – Расчет значения статистики $D_{m,n}^-$

№ п/п	Элементы выборки x	Номера выборки	$F_m(x)$	$(r-1)/n$	$F_m(x) - (r-1)/n$	$G_n(x)$	s/m	$s/m - G_n(x)$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	0	1	0			0	0,083	0,083
2	1	2	0,083	0	0,083	0		
3	2	1	0,083			0,071	0,167	0,096
4	2	2	0,083	0,071	0,012	0,071		
5	3	1	0,167			0,143	0,25	0,107
6	5	1	0,25			0,143	0,333	0,19
7	6	2	0,333	0,143	0,19	0,143		
8	7	1	0,333			0,214	0,417	0,203
9	7	2	0,333	0,214	0,119	0,214		
10	11	2	0,417	0,286	0,131	0,286		
11	13	1	0,417			0,357	0,5	0,143
12	14	1	0,5			0,357	0,583	
13	15	1	0,583			0,357	0,667	0,31
14	15	2	0,583	0,357	0,226	0,357		
15	15	2	0,583	0,429	0,154	0,357		
16	17	1	0,667			0,5	0,75	0,25
17	21	2	0,75	0,5	0,25	0,5		
18	22	1	0,75			0,571	0,883	0,312
19	25	2	0,833	0,571	0,262	0,571		
20	29	2	0,833	0,643	0,19	0,643		
21	30	2	0,833	0,714	0,119	0,714		
22	33	2	0,833	0,786	0,047	0,786		
23	44	2	0,833	0,857	-0,024	0,857		
24	47	2	0,833	0,929	-0,096	0,929		
25	66	1	0,833			1,0	0,917	-0,083
26	97	1	0,917			1,0	0,917	-0,083

Таблица 5. – Критические значения и истинные уровни значимости для двухвыборочной статистики Смирнова ($m = 12, n = 14$)

Номинальный уровень значимости	10%	5%	2%	1%
Критическое значение (дробь)	39/84	43/84	47/84	52/84
Критическое значение (десятичное число)	0,464	0,512	0,559	0,619
Истинный уровень значимости	8,7	4,4	2,0	0,8

Однако у критерия Смирнова есть и недостатки. Его распределение сосредоточено в сравнительно небольшом числе точек. Ясно, что принимаемые этой статистикой значения пропорциональны величине $1/L$, где L – наименьшее общее кратное объемов выборок m и n . Поэтому функция распределения растет большими скачками. Для рассматриваемого примера L – наименьшее общее кратное 12 и 14, т.е. 84. Следовательно, принимаемые статистикой Смирнова входят в арифметическую прогрессию с шагом $1/84 = 0,012$. Именно поэтому критические значения в сборнике [2] приведены в виде дроби с знаменателем $L = 84$.

Кроме того, не удастся выдержать заданный уровень значимости. Реальный (другими словами, истинный) уровень значимости может значительно, даже в несколько раз отличаться от номинального (подробному обсуждению неклассического феномена существенного отличия реального уровня значимости от номинального посвящена работа [8] и раздел 3.5.6 ниже).

При больших объемах выборок можно воспользоваться доказанной Н. В. Смирновым в 1939 г. теоремой: в случае совпадения непрерывных функций распределения элементов двух независимых выборок

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} P \left\{ \sqrt{\frac{mn}{m+n}} D_{m,n} < y \right\} = K(y),$$

где $K(y)$ – функция распределения Колмогорова, заданная формулой

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k \exp\{-2k^2 y^2\}.$$

Поскольку согласно [2] квантиль порядка 0,9 функции распределения Колмогорова равна 1,224, то критическое значение двухвыборочной статистики Смирнова $D_{m,n}$, соответствующее уровню значимости 10%, при больших объемах выборок имеет вид

$$1,224 \sqrt{\frac{m+n}{mn}}.$$

При $m=12$, $n=14$ эта формула дает 0,4815, в то время как точное значение равно 0,464 (см. табл.5 выше). Видим, что приближение удовлетворительное, т.е. рассматриваемые объемы выборок (более 10 элементов) можно считать большими. Для построения правил принятия решений на основе значений двухвыборочной статистики Смирнова, соответствующих другим уровням значимости, можно воспользоваться небольшой табл.6 квантилей функции распределения Колмогорова, взятой из справочника [2].

Таблица 6. – Квантили функции распределения Колмогорова

Величина a	0,8	0,9	0,95	0,98	0,99
Квантиль порядка a	1,07275	1,22385	1,35810	1,51743	1,62762

Критерий типа омега-квадрат (Лемана-Розенблатта). Статистика критерия типа омега-квадрат для проверки однородности двух независимых выборок имеет вид:

$$A = \frac{mn}{m+n} \int_{-\infty}^{\infty} (F_m(x) - G_n(x))^2 dH_{m+n}(x),$$

где $H_{m+n}(x)$ – эмпирическая функция распределения, построенная по объединенной выборке. Легко видеть, что

$$H_{m+n}(x) = \frac{m}{m+n} F_m(x) + \frac{n}{m+n} G_n(x).$$

Статистика A типа омега-квадрат была предложена Э. Леманом в 1951 г., изучена М. Розенблаттом в 1952 г., а затем и другими исследователями. Она зависит лишь от рангов элементов двух выборок в объединенной выборке. Пусть x_1, x_2, \dots, x_m – первая выборка, $x'_1 < x'_2 < \dots < x'_m$ – соответствующий вариационный ряд, y_1, y_2, \dots, y_n – вторая выборка, $y'_1 < y'_2 < \dots < y'_n$ – вариационный ряд, соответствующий второй выборке. Поскольку функции распределения независимых выборок непрерывны, то с вероятностью 1 все выборочные значения различны, совпадения отсутствуют. Статистика A представляется в виде (см., например, [2]):

$$A = \frac{1}{mn(m+n)} \left[m \sum_{i=1}^m (r_i - i)^2 + n \sum_{j=1}^n (s_j - j)^2 \right] - \frac{4mn-1}{6(m+n)},$$

где r_i – ранг x'_i и s_j – ранг y'_j в общем вариационном ряду, построенном по объединенной выборке.

Правила принятия решений при проверке однородности двух выборок на основе статистик Смирнова и типа омега-квадрат, т.е. таблицы критических значений в зависимости от уровней значимости и объемов значимости приведены, например, в таблицах [2]. При достаточно больших объемах выборок правило принятия решения формулируется просто: если наблюдаемое значение статистики меньше соответствующего

квантиля предельного распределения, гипотеза однородности принимается, в противном случае отклоняется.

Расчет значения статистики A типа омега-квадрат (статистики Лемана-Розенблатта) по тем же данным, по которым были найдены значения статистик критериев Вилкоксона и Смирнова, представлен в табл.7. Суммируя значения в столбце (6), получаем, что

$$\sum_{i=1}^{12} (r_i - i)^2 = 598.$$

Аналогично получаем с помощью столбца (9), что

$$\sum_{j=1}^{14} (s_j - j)^2 = 880.$$

Следовательно,

$$\begin{aligned} A &= \frac{1}{12 \times 14 \times 26} [12 \times 598 + 14 \times 880] - \frac{4 \times 12 \times 14 - 1}{6 \times 26} = \\ &= \frac{1}{4368} [7176 + 12320] - \frac{671}{156} = 4,4634 - 4,3013 = 0,1621 \end{aligned}$$

Таблица 7 – Расчет значения статистики A Лемана-Розенблатта

№ п/п	Элементы выборки x	Номера выборки	i	$r_i - i$	$(r_i - i)^2$	j	$s_j - j$	$(s_j - j)^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	0	1	1	0	0			
2	1	2				1	1	1
3	2	1	2	1	1			
4	2	2				2	2	4
5	3	1	3	2	4			
6	5	1	4	2	4			
7	6	2				3	4	16
8	7	1	5	3	9			
9	7	2				4	5	25
10	11	2				5	5	25
11	13	1	6	5	25			
12	14	1	7	5	25			
13	15	1	8	5	25			
14	15	2				6	8	64
15	15	2				7	8	64
16	17	1	9	7	49			
17	21	2				8	9	81
18	22	1	10	8	64			
19	25	2				9	10	100
20	29	2				10	100	100
21	30	2				11	10	100
22	33	2				12	10	100
23	44	2				13	10	100
24	47	2				14	10	100
25	66	1	11	14	196			
26	97	1	12	14	196			

Известно [21], что

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} P\{A < x\} = a_1(x)$$

(в обозначениях [2]), где $a_1(x)$ – предельная функция распределения классической статистики омега-квадрат (Крамера - Мизеса - Смирнова), используемой для проверки согласия эмпирического распределения с заданным теоретическим.

Квантили функции распределения $a_1(x)$ приведены в табл.8. Известно [2, 21], что в случае статистики Лемана-Розенблатта предельным распределением можно пользоваться и для выборок умеренного объема (5 и 7, 6 и 7, 7 и 7,8 и 8 и т.д.). Поскольку наблюдаемое значение $A = 0,1621$ меньше любого критического значения в табл.8, то гипотезу однородности двух рассматриваемых выборок следует принять.

Таблица 8. – Квантили предельной функции распределения статистики омега-квадрат (Крамера - Мизеса - Смирнова)

Величина a	0,8	0,9	0,95	0,98	0,99
Квантиль порядка a	0,245	0,347	0,461	0,620	0,743

Рекомендации по выбору критерия однородности. Для критерия типа омега-квадрат (Лемана-Розенблатта) нет выраженного эффекта различия между номинальными и реальными уровнями значимости. Поэтому мы *рекомендуем для проверки однородности функций распределения (гипотеза H_0) применять статистику A типа омега-квадрат. Если методическое, табличное или программное обеспечение для статистики Лемана - Розенблатта отсутствует, рекомендуем использовать критерий Смирнова. Для проверки однородности математических ожиданий (гипотеза H'_0) целесообразно применять критерий Крамера-Уэлча.* По нашему мнению, статистики Стьюдента, Вилкоксона и др. допустимо использовать лишь в отдельных частных случаях, рассмотренных выше.

Кратко сформулируем некоторые соображения о внедрении современных методов прикладной статистики в практику технических, экономических, медицинских и иных исследований. Даже из проведенного выше разбора лишь одной из типичных статистических задач организационно-экономического моделирования - задачи проверки однородности двух независимых выборок - можно сделать вывод о целесообразности широкого развертывания работ по критическому анализу сложившейся практики статистической обработки данных и по внедрению накопленного арсенала современных методов прикладной статистики. По нашему мнению, широкого внедрения заслуживают, в частности, методы многомерного статистического анализа, планирования эксперимента, статистики объектов нечисловой природы. Очевидно, рассматриваемые работы должны быть плановыми, организационно оформленными, проводиться мощными самостоятельными организациями и подразделениями. Целесообразно создание службы статистических консультаций в системе научно-исследовательских учреждений и вузов технического, экономического, медицинского профиля, а также в рамках корпораций и промышленных предприятий. Этот инновационный проект подробно разработан в специальной литературе [18, 19].

3.5.6. Реальные и номинальные уровни значимости в задачах проверки статистических гипотез

Во многих монографиях, справочниках и таблицах (например, [1, 6, 7]) при проверке статистических гипотез критические значения статистик указаны для априорно фиксированных (номинальных в терминологии [8]) уровней значимости α_H . В качестве таковых обычно используются значения из тройки чисел 0,01, 0,05, 0,1, к которым иногда добавляют еще несколько: 0,001, 0,005, 0,02 и др.

Однако ясно, что для дискретных статистик (т.е. статистик с дискретными функциями распределения), к которым, в частности, относятся все непараметрические статистические критерии [2, 24], реальные уровни значимости α_p могут и не совпадать с номинальными. Под α_p понимается максимально возможный уровень значимости дискретной статистики, не превосходящий заданный номинальный α_H (т.е. при переходе к следующему по величине возможному значению дискретной статистики соответствующий уровень значимости оказывается больше заданного номинального). Поэтому в лучших таблицах [2, 24] для ограниченных объемов выборок (2 - 100) табулируются точные распределения дискретных статистик. Для каждой конкретной статистики реальный уровень значимости α_p - функция от объемов выборок $n = (n_1, \dots, n_t)$, т.е. $\alpha_p = \alpha_p(n)$. (Здесь t - число выборок, по которым рассчитывается значение статистики; рассматриваем в основном случай двух выборок, т.е. $t = 2$.)

В одних таблицах приведены α_p [2, 24], в других - нет [1, 6, 7]. Возникает естественный вопрос: с чем это связано? Либо в работах [1, 6, 7] нарушена культура табулирования, либо реальные α_p и номинальные α_H уровни значимости практически совпадают для всех n . Продемонстрируем, что по крайней мере для некоторых статистик выполнено первое из этих двух утверждений.

В качестве примера рассмотрим критерий серий (Вольфовица) проверки однородности двух независимых выборок. Статистика этого критерия V - это число серий, т.е. частей общего вариационного ряда двух выборок, каждая из которых состоит из элементов одной выборки. При справедливости нулевой гипотезы о тождестве функций распределения, соответствующих двум независимым выборкам объемов n_1 и n_2 , известно точное распределение [2, табл.6.7]

$$P(V = r | n_1, n_2) = \begin{cases} \frac{2C_{n_1-1}^{k-1} C_{n_2-1}^{k-1}}{C_{n_1+n_2}^{n_1}}, & \text{если } r = 2k, \\ \frac{C_{n_1-1}^k C_{n_2-1}^{k-1} + C_{n_1-1}^{k-1} C_{n_2-1}^k}{C_{n_1+n_2}^{n_1}}, & \text{если } r = 2k+1, \end{cases}$$

где $r = 2, 3, \dots, 2n_1$ при $n_1 = n_2$ и $r = 2, 3, \dots, 2n_1 + 1$ при $n_1 < n_2$ (без ограничения общности можно принять, что объем первой выборки не превосходит объема второй выборки, т.е. $n_1 \leq n_2$).

Несложный расчет для номинального уровня значимости $\alpha_H = 0,05$ показывает, что

при $n_1 = n_2 = 6$ реальный уровень значимости $\alpha_p = 0,0260$;

при $n_1 = n_2 = 8$ реальный уровень значимости $\alpha_p = 0,0178$;

при $n_1 = n_2 = 10$ реальный уровень значимости $\alpha_p = 0,0370$;

при $n_1 = n_2 = 12$ реальный уровень значимости $\alpha_p = 0,0190$.

Таким образом, для рассматриваемых объемов выборок реальный уровень значимости в 2-3 раза меньше, чем номинальный. Это, очевидно, необходимо учитывать при интерпретации результатов анализа реальных статистических данных.

Соотношение реальных (истинных) и номинальных уровней значимости было изучено нами [8] на примере непараметрических критериев проверки однородности двух независимых выборок. В табл.9, построенной в [8] по данным [2, 4, 24], для ряда непараметрических критериев проверки однородности двух независимых выборок приведены реальные уровни значимости $\alpha_p(n)$ для номинального уровня значимости $\alpha_H = 0,05$ и объемов выборок $n_1 = n_2 = 6, 8, 10, 12$. Проанализированы пять критериев.

1. Двухвыборочный критерий Вилкоксона, являющийся линейной функцией от критерия Манна-Уитни и подробно рассмотренный выше в разд. 3.5.4. Напомним, что статистика Вилкоксона S - это сумма рангов элементов первой выборки

$$S = R_1 + R_2 + \dots + R_{n_1}$$

в общем вариационном ряду, построенном по объединенной выборке, включающей в себя все элементы обеих выборок (без ограничения общности можно принять, что объем первой выборки не превосходит объема второй выборки, т.е. $n_1 \leq n_2$).

2. Критерий Ван-дер-Вардена [2, 4], представляющий собой дальнейшее развитие (модификацию) критерия Вилкоксона и предназначенный для анализа выборок, распределение которых близко к нормальному. Статистика X критерия Ван-дер-Вардена имеет вид

$$X = \Phi^{-1} \left\{ \frac{R_1}{n_1 + n_2 + 1} \right\} + \Phi^{-1} \left\{ \frac{R_2}{n_1 + n_2 + 1} \right\} + \dots + \Phi^{-1} \left\{ \frac{R_{n_1}}{n_1 + n_2 + 1} \right\},$$

где $\Phi^{-1}(p)$ есть квантиль порядка p стандартного нормального распределения $\Phi(x)$ с математическим ожиданием 0 и дисперсией 1, т.е. $\Phi^{-1}(p)$ - обратная функция к $\Phi(x)$.

3. Двухвыборочный двухсторонний критерий Смирнова однородности двух независимых выборок, рассмотренный в разд. 3.5.5. Он основан на использовании разности эмпирических функций распределения $F_{n_1}(x)$ и $G_{n_2}(x)$ построенных по первой и второй выборкам соответственно. Термин «двухсторонний» означает, что берется супремум модуля этой разности. Статистика двухвыборочного двухстороннего критерия Смирнова

$$D = D(n_1, n_2) = \sup_x |F_{n_1}(x) - G_{n_2}(x)|$$

в случае равенства объемов выборок $n_1 = n_2$ принимает значения, кратные $1/n_1$, поскольку только такие значения принимают эмпирические функции распределения $F_{n_1}(x)$ и $G_{n_2}(x)$, а потому рассматриваемая статистика имеет (n_1+1) возможных значений.

4. Критерий знаков Z используют в случае равенства объемов выборок $n_1 = n_2$. Статистика этого критерия равна числу положительных разностей $X_k - Y_k$ элементов двух выборок с одинаковыми номерами. При справедливости нулевой гипотезы статистика Z имеет биномиальное распределение $B(1/2; n_1)$, а потому имеет (n_1+1) возможных значений.

5. Критерий серий (Вольфовица) V , о котором шла речь выше в начале настоящего подраздела. Число его возможных значений не превосходит $2n_1$.

Таблица 9. – Реальные уровни значимости $\alpha_p(n)$ для $\alpha_H = 0,05$

Наименование и обозначение критерия	Объемы выборок $n_1 = n_2$				Примечания и ссылки
	6	8	10	12	
Вилкоксона S	0,0320	0,0400	0,0480	0,0420	[24, с.280-281], [4, с.418]
Ван-дер-Вардена X	0,0498	0,0498	0,0500	0,0500	Рассчитано по методике [4, с.249-250]
Смирнова D	0,0044	0,0372	0,0246	0,0158	[2, с.412], [24, с.406-427]
Знаков Z	0,0312	0,0078	0,0214	0,0386	[24, с.273-274]
Вольфовица (серий) V	0,0260	0,0178	0,0370	0,0190	Рассчитано по методике [2, с.91-92]

Анализ содержания табл.9 подтверждает предположение о существенности отличия реальных уровней значимости $\alpha_p(n)$ от номинальных уровней значимости α_H .

Предположим теперь, что, несмотря на установленные отличия, мы используем при проверке гипотезы однородности таблицы [1, 6, 7], в которых указаны $\alpha_H > \alpha_p$, а не α_p . Это приводит к снижению мощности критерия по сравнению с соответствующим рандомизированным критерием, обеспечивающим равенство α_p и α_H .

Разъяснение. Поясним, что такое рандомизированный критерий. Пусть Y – статистика некоторого статистического критерия, принимающая дискретные значения, числа a и b , где $a < b$ – два соседних значения этой статистики, такие, что $P(Y > b) < \alpha_H$ и $P(Y > a) > \alpha_H$ (вероятности взяты в предположении справедливости нулевой гипотезы). Если критическое значение критерия равно b , т.е. нулевая гипотеза принимается при $Y \leq b$, то $\alpha_p = P(Y > b) < \alpha_H$. Если же критическое значение равно следующему возможному (при движении в сторону уменьшения) значению a , т.е. нулевая гипотеза принимается при $Y \leq a$, то $\alpha_p = P(Y > a) > \alpha_H$. Рандомизированный критерий получим, если при $Y = b$ в некоторой доле p случаев будем принимать нулевую гипотезу, а в остальных случаях – альтернативную. Поскольку

$$P(Y = b) = P(Y > a) - P(Y > b),$$

то (реальный) уровень значимости рандомизированного критерия равен

$$(1 - p) P(Y = b) + P(Y > b) = (1 - p) P(Y > a) + p P(Y > b).$$

Ясно, что при соответствующем выборе параметра рандомизации p уровень значимости рандомизированного критерия совпадет с заданным номинальным уровнем α_H .

Для малых объемов выборок (2–20 элементов) понижение мощности из-за того, что $\alpha_H > \alpha_p$, может быть существенным. Для иллюстрации этого в табл.10 приведены результаты моделирования наиболее употребительных (согласно [2]) критериев проверки однородности двух независимых выборок.

Таблица 10. – Мощности статистических критериев при $\alpha_H = 0,05$

Номер эксперимента	Объем выборок $n_1 = n_2$	Параметры				Мощность M статистического критерия				
		m_1	m_2	σ_1^2	σ_2^2	S	V	X	D	t
1	6	0	1	1	1	0,318	0,006	0,298	0,238	0,396
2	8	0	1	1	1	0,452	0,104	0,426	0,068	0,484
3	10	0	1	1	1	0,520	0,180	0,534	0,116	0,598
4	12	0	1	1	1	0,632	0,076	0,618	0,462	0,682
5	6	0	2	1	1	0,828	0,308	0,808	0,716	0,904
6	8	0	2	1	1	0,958	0,510	0,954	0,458	0,976
7	10	0	2	1	1	0,984	0,704	0,990	0,632	0,988
8	12	0	2	1	1	0,996	0,568	0,996	0,978	0,998

Моделируются выборки одинакового объема из нормальных законов распределения с математическими ожиданиями m_1 и m_2 и дисперсиями σ_1^2 и σ_2^2 . Номинальный уровень значимости, определяющий конкретные критические значения для критериев, принят равным $\alpha_H = 0,05$. Мощность критерия определяется моделированием $N = 5000$ пар выборок. При использовании $N = 5000$ моделируемых пар выборок среднее квадратическое отклонение оценок мощности $\sigma_M \leq 0,0223$ (при $M \geq 0,95$ имеем $\sigma_M \leq 0,01$).

Изучены критерии Вилкоксона S , Вольфовица V , Ван-дер-Вардена X , Смирнова D . Критерий Стьюдента t (см. например, [2]), как равномерно наиболее мощный в классе

нормальных законов распределения, приведен для сравнительной оценки мощностей рассматриваемых непараметрических критериев. (Моделирование и расчеты, приведенные в настоящем подразделе, выполнены Ю.Э. Камнем и Я.Э. Камнем [8].)

Замечание. Приведенные в табл.10 значения мощностей критериев интересны нам с точки зрения обсуждения их зависимости от различия реальных и номинальных уровней значимости. При этом необходимо подчеркнуть, что эти значения зависят от предположений, принятых при моделировании. Так, критерии Вилкоксона и Ван-дер-Вардена «настроены» на использование в случае распределений, близких к нормальному семейству. При проверке гипотезы о совпадении функций распределения двух независимых выборок из логистического распределения с альтернативой сдвига критерий Вилкоксона является асимптотически оптимальным. А в случае выборок из нормального распределения аналогичным свойством обладает критерий Ван-дер-Вардена, причем известно, что семейства нормальных и логистических распределений весьма близки – расстояние Колмогорова между ними не превышает 0,01 (см. по вопросам асимптотической оптимальности непараметрических критериев [10, 11, 15]). Поэтому нет ничего удивительного в том, что мощности критериев Вилкоксона и Ван-дер-Вардена близки к оптимуму в случае нормального распределения – к мощности критерия Стьюдента. При этом мощности критериев Смирнова и особенно критерия Вольфовица заметно меньше. Однако для выборок из других распределений (например, распределений Вейбулла-Гнеденко или гамма-распределений) ситуация иная – критерий Смирнова, как показывает компьютерное моделирование, оказывается более мощным, чем критерии Вилкоксона и Ван-дер-Вардена. Более того, критерий Смирнова – состоятельный, т.е. позволяет отклонить любую конкретную альтернативу (при соответствующих объемах выборок), а критерии Вилкоксона и Ван-дер-Вардена не являются состоятельными, некоторых альтернатив они «не чувствуют» (см. подраздел 3.5.4). Поэтому вполне обоснованной является рекомендация о широком использовании состоятельных критериев Смирнова и типа омега-квадрат (Лемана-Розенблатта), данная в подразделе 3.5.5. Что же касается критерия серий (Вольфовица), то из-за его отрицательных свойств (выраженной дискретности, низкой мощности) он в настоящее время выходит из употребления при анализе реальных данных, несмотря на прозрачность определения.

Рассмотрения настоящего раздела позволяют сделать следующие выводы [8].

1. При создании методик и таблиц, а также программных продуктов необходимо соблюдать определенную культуру табулирования. В качестве положительных примеров можно указать работы [2, 24].

2. При малых объемах выборок использовать номинальные уровни значимости α_H вместо реальных уровней значимости α_p для дискретных статистик недопустимо.

3. При конечных объемах выборок выбор того или иного критерия с дискретной статистикой должен сопровождаться исследованием влияния варьирования уровня значимости на качественную интерпретацию результатов проверки гипотез. В частности, выбор одного из двух конкурирующих непараметрических критериев K_1 и K_2 прежде всего должен зависеть от априорного выбора исследователем реального уровня значимости α_{p_1} или α_{p_2} , соответствующего первому критерию K_1 или второму K_2 , в качестве номинального уровня значимости α_H .

Последний вывод демонстрирует сложность сравнения критериев с дискретными статистиками между собой, поскольку точки скачков распределений их статистик не совпадают. Следовательно, в отличие от критериев с непрерывными статистиками нельзя выбрать единый фиксированный уровень значимости и сравнить свойства критериев при этом уровне значимости.

В заключение отметим, что для любого критерия проверки статистических гипотез реальный уровень значимости приближается к номинальному при безграничном

возрастании объемов выборок, т.е. $\alpha_p(n) \rightarrow \alpha_H$ при $\min(n_1, \dots, n_l) \rightarrow \infty$. Поэтому для прикладных исследований значительный интерес представляет определение верхней оценки скорости сходимости $\alpha_p(n)$ к α_H . Соответствующие теоретические результаты для критериев проверки однородности двух независимых или связанных выборок можно получить, основываясь на оценках скорости сходимости в принципе инвариантности [21, гл.4]. Некоторые оценки приведены в [17, гл.2]. Скорость сходимости также может быть оценена методом статистических испытаний (Монте-Карло). Пример подобного исследования подробно рассмотрен в [9] в ходе обсуждения проблем вероятностно-статистического моделирования помех, создаваемых электровозами.

В настоящем разделе затронута лишь небольшая часть непараметрических методов анализа числовых статистических данных. В частности, обратим внимание на непараметрические оценки плотности, которые используются для описания данных, проверки однородности, в задачах восстановления зависимостей и других областях эконометрики. Непараметрические оценки плотности рассмотрены в [21, раздел 5.6].

3.6. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей

Метод наименьших квадратов восстановления зависимости – один из наиболее распространенных статистических методов анализа данных. В настоящем разделе рассмотрена непараметрическая постановка: восстанавливаемая зависимость – сумма линейной функции и периодической составляющей произвольного вида (с известным периодом), распределение случайных погрешностей (остатков, невязок) произвольно.

3.6.1. Задача восстановления линейной зависимости

Начнем с простейшего случая – задачи восстановления линейной зависимости. Пусть t – независимая переменная, а x – зависимая. Рассмотрим задачу восстановления зависимости $x = x(t)$ на основе набора n пар чисел (t_k, x_k) , $k = 1, 2, \dots, n$, где t_k – значения независимой переменной, а x_k – соответствующие им значения зависимой переменной.

Восстанавливать зависимость можно на основе различных моделей. Обычно применяют модели временных рядов, включающие три составляющие: трендовую (T), периодическую (S) и случайную (E). Рассматривают, как в [1] и аналогичных изданиях, аддитивную модель $T + S + E$ и мультипликативную модель $T \times S \times E$.

Простейшая аддитивная модель имеет вид

$$x_k = a(t_k - \bar{t}) + d + e_k = a(t_k - \bar{t}) + d + f(t_k) + E_k, \quad k = 1, 2, \dots, n. \quad (1)$$

Здесь трендовая составляющая – линейная функция $a(t_k - \bar{t}) + d$ (такая запись тренда предпочтительнее для облегчения выкладок); периодическая составляющая $f(t)$ обычно описывает сезонность, т.е. период известен (в зависимости от моделируемой ситуации он равен году, неделе, суткам и т.п.); случайная составляющая представлена слагаемыми E_k , которые являются реализациями независимых одинаково распределенных случайных величин с нулевым математическим ожиданием и дисперсией σ^2 , неизвестной статистику. В рассматриваемой модели $e_k = f(t_k) + E_k$, $k = 1, 2, \dots, n$, т.е. отклонения от линейного тренда e_k не являются одинаково распределенными. Однако их распределения отличаются лишь сдвигами (на значения детерминированной периодической составляющей).

Соответствующая модели (1) мультипликативная модель имеет вид

$$y_k = [Bt_k^a] \times f_1(t_k) \times [1 + \varepsilon_k], \quad k = 1, 2, \dots, n. \quad (2)$$

В модели (2) сомножители имеют описанный выше смысл. При логарифмировании модель (2) переходит в аналог модели (1), следовательно, достаточно рассматривать модель (1).

Иногда принимают предположение о нормальности распределения погрешностей. Однако давно известно, что распределения реальных данных, как правило, отличаются от нормальных [2]. Поэтому далее рассматриваем *непараметрическую модель*, не предполагая, что распределение погрешностей входит в то или иное параметрическое семейство. Отказ от задания распределения погрешностей в параметрическом виде – одно из оснований для того, чтобы именовать рассматриваемые модель и метод непараметрическими. Второе основание – отказ от выбора периодической составляющей из какого-либо параметрического семейства функций.

Практическая значимость модели (1) очевидна. Однако расчетные методы, описанные в [1] и аналогичных изданиях, являются эвристическими. Цель настоящей статьи – *построить непараметрическую вероятностно-статистическую теорию прогноза временного ряда на базе линейного тренда с учетом аддитивной периодической составляющей*.

Метод наименьших квадратов был разработан К. Гауссом в 1794 г. [2]. Согласно нему этому методу для расчета наилучшей функции, приближающей линейным образом зависимость x от t в модели (1), следует рассмотреть функцию двух переменных

$$f(a, d) = \sum_{k=1}^n (x_k - a(t_k - \bar{t}) - d)^2.$$

Оценки метода наименьших квадратов (кратко: оценки МНК) – это такие значения a^* и d^* , при которых функция $f(a, d)$ достигает минимума по всем значениям аргументов. Как известно (см., например, [2]), оценки МНК имеют вид

$$a^* = \frac{\sum_{k=1}^n x_k (t_k - \bar{t})}{\sum_{k=1}^n (t_k - \bar{t})^2}, \quad d^* = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k. \quad (3)$$

Следуя эвристическому подходу [1], изучим асимптотическое поведение оценок МНК a^* и d^* , заданных формулами (3), установим их асимптотическую нормальность в предположениях модели (19), а затем состоятельно оценим периодическую составляющую $f(t)$ и построим интервальный прогноз для $x(t)$.

3.6.2. Асимптотические распределения оценок параметров

Из формулы (3) следует, что

$$d^* = \frac{a}{n} \sum_{k=1}^n (t_k - \bar{t}) + d + \frac{1}{n} \sum_{k=1}^n e_k = d + \frac{1}{n} \sum_{k=1}^n e_k = d + \frac{1}{n} \sum_{k=1}^n f(t_k) + \frac{1}{n} \sum_{k=1}^n E_k. \quad (4)$$

Согласно Центральной предельной теореме (для выполнения ее условий необходимо предположить, например, что погрешности e_k , $k = 1, 2, \dots, n$, финитны или имеют конечный третий абсолютный момент; однако заострять внимание на этих внутриматематических «условиях регулярности» здесь нет необходимости) оценка d^* имеет асимптотически нормальное распределение с математическим ожиданием $d + \frac{1}{n} \sum_{k=1}^n f(t_k)$ и

дисперсией σ^2/n , ее оценка приводится ниже. Из формул (3) и (4) вытекает, что

$$x_k - \bar{x} = a(t_k - \bar{t}) + d + e_k - d - \frac{1}{n} \sum_{k=1}^n e_k = a(t_k - \bar{t}) + e_k - \frac{1}{n} \sum_{k=1}^n e_k,$$

$$(x_k - \bar{x})(t_k - \bar{t}) = a(t_k - \bar{t})^2 + e_k(t_k - \bar{t}) - \frac{(t_k - \bar{t})}{n} \sum_{k=1}^n e_k.$$

Последнее слагаемое во втором соотношении при суммировании по k обращается в 0, поэтому

$$a^* = a + \sum_{k=1}^n c_k e_k = a + \sum_{k=1}^n c_k f(t_k) + \sum_{k=1}^n c_k E_k, \quad c_k = \frac{(t_k - \bar{t})}{\sum_{k=1}^n (t_k - \bar{t})^2}. \quad (5)$$

Формулы (5) показывают, что оценка a^* является асимптотически нормальной с математическим ожиданием $a + \sum_{k=1}^n c_k f(t_k)$ и дисперсией

$$D(a^*) = \sum_{k=1}^n c_k^2 D(E_k) = \frac{\sigma^2}{\sum_{k=1}^n (t_k - \bar{t})^2}.$$

Отметим, что многомерная нормальность имеет быть, когда каждое слагаемое в формуле (5) мало сравнительно со всей суммой, т.е.

$$\lim_{n \rightarrow \infty} \max |t_k - \bar{t}| / \left\{ \sum_{k=1}^n (t_k - \bar{t})^2 \right\}^{1/2} = 0. \quad (6)$$

Условие (6) выполнено, например, если t_k образуют (полную, т.е. без пропусков) арифметическую прогрессию, число членов которой безгранично растет.

Итак, дисперсии оценок МНК параметров a^* и d^* линейного тренда – те же, что и при отсутствии сезонных искажений (см., например, [2]). А вот их математические ожидания зависят от периодической составляющей. Однако в случае

$$\sum_{i=1}^n f(t_i) = 0, \quad \sum_{i=1}^n (t_i - \bar{t}) f(t_i) = 0 \quad (7)$$

оценки a^* и d^* являются несмещенными.

Условия (7) являются необходимыми и достаточными для несмещенности и состоятельности оценок МНК коэффициентов линейной зависимости. Проверка условий (7) рассмотрена в конце статьи.

Несмещенность (в предположениях (7) и асимптотическая нормальность оценок метода наименьших квадратов позволяют легко указывать для них асимптотические доверительные границы и проверять статистические гипотезы, например, о равенстве определенным значениям, прежде всего 0.

3.6.3. Асимптотическое распределение трендовой составляющей

Из формул (4) и (5) следует, что при справедливости соотношений (7)

$$M\{a^*(t - \bar{t}) + d^*\} = M(a^*)(t - \bar{t}) + M(d^*) = a(t - \bar{t}) + d,$$

т.е. оценка $y^*(t) = a^*(t - \bar{t}) + d^*$ трендовой составляющей $y(t) = a(t - \bar{t}) + d$ рассматриваемой зависимости является несмещенной. Поэтому

$$D(y^*(t)) = D(a^*)(t - \bar{t})^2 + 2M\{(a^* - a)(d^* - d)(t - \bar{t})\} + D(d^*).$$

При этом, поскольку погрешности E_k независимы в совокупности и $M(E_k) = 0$, то

$$M\{(a^* - a)(d^* - d)(t - \bar{t})\} = \frac{1}{n} \sum_{k=1}^n c_k (t - \bar{t}) M(E_k^2) = \frac{1}{n} (t - \bar{t}) \sigma^2 \sum_{k=1}^n c_k = 0.$$

Таким образом,

$$D(y^*(t)) = \sigma^2 \left\{ \frac{1}{n} + \frac{(t - \bar{t})^2}{\sum_{k=1}^n (t_k - \bar{t})^2} \right\}. \quad (8)$$

Итак, оценка $y^*(t)$ является несмещенной и асимптотически нормальной. Для ее практического использования (построения доверительных интервалов, проверки статистических гипотез) необходимо состоятельно уметь оценивать остаточную дисперсию $M(E_k^2) = \sigma^2$.

В частности, не представляет труда выписывание нижней и верхней границ для трендовой составляющей прогностической функции:

$$y_{\text{нижн}}(t) = a^*(t - \bar{t}) + d^* - \delta(t), \quad y_{\text{верх}}(t) = a^*(t - \bar{t}) + d^* + \delta(t),$$

где полуширина доверительного интервала $\delta(t)$ имеет вид

$$\delta(t) = U(\gamma) \sqrt{D^*(y^*(t))} = U(\gamma) \sigma^* \sqrt{\frac{1}{n} + \frac{(t - \bar{t})^2}{\sum_{k=1}^n (t_k - \bar{t})^2}}. \quad (9)$$

Здесь γ - доверительная вероятность, $U(\gamma)$ - квантиль нормального распределения порядка $\frac{1+\gamma}{2}$, т.е. $U(\gamma) = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$, где $\Phi(x)$ - функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1. При $\gamma = 0,95$ (наиболее применяемое значение) имеем $U(\gamma) = 1,96$. В формуле (9) $D^*(y^*(t))$ - состоятельная оценка дисперсии $y^*(t)$. В соответствии с (8) она является произведением состоятельной оценки σ^* среднего квадратического отклонения σ случайных погрешностей E_k на известную исследователю детерминированную функцию от t .

3.6.4. Математическое ожидание остаточной суммы квадратов

В точках t_k , $k = 1, 2, \dots, n$, имеются исходные значения зависимой переменной x_k и восстановленные значения $y^*(t_k)$. Рассмотрим остаточную сумму квадратов

$$SS = \sum_{k=1}^n (y^*(t_k) - x_k)^2 = \sum_{k=1}^n \left\{ (a^* - a)(t_k - \bar{t}) + (d^* - d) - f(t_k) - E_k \right\}^2.$$

При отсутствии периодической составляющей используют [2] состоятельные оценки σ^* среднего квадратического отклонения σ случайных погрешностей, построенные на основе остаточной суммы квадратов $\sigma^* = \sqrt{\frac{SS}{n}}$ или $\sigma^* = \sqrt{\frac{SS}{n-2}}$. Однако при наличии периодической составляющей так делать нельзя. Приходится использовать «обходный путь».

В соответствии с формулами (4) и (5) при справедливости условий (7)

$$\begin{aligned} SS &= \sum_{k=1}^n \left\{ (t_k - \bar{t}) \sum_{j=1}^n c_j E_j + \frac{1}{n} \sum_{j=1}^n E_j - f(t_k) - E_k \right\}^2 = \\ &= \sum_{k=1}^n \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j - f(t_k) - E_k \right\}^2 = \sum_{k=1}^n SS_k. \end{aligned}$$

Найдем математическое ожидание каждого из слагаемых:

$$M(SS_k) = M \left\{ \sum_{j=1}^n \left[c_j(t_k - \bar{t}) + \frac{1}{n} E_j - f(t_k) - E_k \right] \right\}^2 = M \left\{ \sum_{j=1}^n \left[c_j(t_k - \bar{t}) + \frac{1}{n} E_j \right] \right\}^2 - 2M \left\{ \sum_{j=1}^n \left[c_j(t_k - \bar{t}) + \frac{1}{n} E_j \right] \right\} (f(t_k) + E_k) + M(f(t_k) - E_k)^2.$$

Поскольку E_k независимы, одинаково распределены и имеют нулевое математическое ожидание, то

$$M \left\{ \sum_{j=1}^n \left[c_j(t_k - \bar{t}) + \frac{1}{n} E_j \right] \right\}^2 = \sum_{j=1}^n \left\{ c_j(t_k - \bar{t}) + \frac{1}{n} \right\}^2 \sigma^2.$$

Далее,

$$-2M \left\{ \sum_{j=1}^n \left[c_j(t_k - \bar{t}) + \frac{1}{n} E_j \right] \right\} (f(t_k) + E_k) = -2 \left\{ c_k(t_k - \bar{t}) + \frac{1}{n} \right\} \sigma^2.$$

Наконец,

$$M(f(t_k) - E_k)^2 = f^2(t_k) + \sigma^2.$$

На основе трех последних равенств можно показать, что при выполнении условия асимптотической нормальности (б)

$$\lim_{n \rightarrow \infty} M(SS_k) = f^2(t_k) + \sigma^2.$$

Следовательно,

$$M \left(\frac{SS}{n} \right) = \sigma^2 + \frac{1}{n} \sum_{k=1}^n f^2(t_k). \quad (10)$$

В правой части (10) первое слагаемое соответствует вкладу случайной составляющей, второе – вкладу периодической составляющей.

В некоторых случаях второе слагаемое в правой части (10) может быть известно из предыдущего опыта или же оценено экспертами, однако в большинстве ситуаций целесообразно исходить из оценки периодической составляющей.

3.6.5. Оценивание периодической составляющей

В литературе рассматривают как параметрические, так и непараметрические подходы. Популярный метод исходит из того, что достаточно гладкую функцию можно разложить в ряд Фурье и получить хорошее приближение с помощью небольшого числа гармоник. В простейшем случае – одна гармоника. Так, динамику индекса инфляции можно попытаться изучать с помощью модели

$$x_k = a(t_k - \bar{t}) + d + f(t_k) + E_k = a(t_k - \bar{t}) + d + g \cos(2\pi k) + E_k, \quad k = 1, 2, \dots, n$$

(время t измеряется в годах). Тогда неизвестные параметры a , b , g оцениваются методом наименьших квадратов.

Однако обычно нет оснований предполагать, что периодическая составляющая входит в то или иное параметрическое семейство функций. Приходится строить непараметрические оценки. Опишем одну из возможных постановок.

Пусть в согласии с предположениями (7) рассматривается целое число периодов, т.е. $n = tq$, где n – объем наблюдений, t – количество периодов, q – число наблюдений в одном периоде. Предполагается, что первые q моментов наблюдения при сдвиге на длину периода дают следующие q моментов времени, при сдвиге на две длины периода дают третий набор из q моментов наблюдения, и т.д. Тогда в соответствии с определением периодической составляющей справедливы равенства

$$f(t_s) = f(t_{q+s}) = f(t_{2q+s}) = \dots = f(t_{(m-1)q+s}), \quad s = 1, 2, \dots, q. \quad (11)$$

Если наблюдения проводятся ежемесячно в течение m лет, то число наблюдений в одном периоде $q = 12$, общий объем наблюдений $n = 12m$, далее s – номер месяца в году, $s = 1, 2, \dots, 12$. Пусть g_s - общее значение в (11). Для оценки периодической составляющей требуется оценить g_1, g_2, \dots, g_q .

Естественный подход состоит в том, чтобы усреднить m значений $x_k - y^*(t_k)$, соответствующих моментам времени, отстоящим друг от друга на целое число периодов. Другими словами, усреднить «очищенные» от трендовой составляющей исходные данные, соответствующие одноименным месяцам различных лет. Речь идет об оценках

$$g_s^* = \frac{1}{m} \sum_{j=1}^m (x_{s+(j-1)q} - y^*(t_{s+(j-1)q})), \quad s = 1, 2, \dots, q. \quad (12)$$

Оценка периодической составляющей распространяется на весь интервал наблюдений очевидным образом:

$$f^*(t_s) = f^*(t_{q+s}) = f^*(t_{2q+s}) = \dots = f^*(t_{(m-1)q+s}) = g_s^*, \quad s = 1, 2, \dots, q. \quad (13)$$

Сложив восстановленные значения трендовой и периодической составляющей, получим оценку зависимости, «очищенную» от случайной составляющей

$$x^*(t) = y^*(t) + f^*(t) = a^*(t - \bar{t}) + d^* + f^*(t). \quad (14)$$

Здесь оценки a^* и d^* находят по формулам (3), а оценки $f^*(t)$ – по формулам (12) – (13).

С помощью формулы (14) можно строить точечный прогноз, используя ее вне интервала наблюдений. Для этого достаточно распространить сезонную составляющую $f^*(t)$ вплоть до рассматриваемого момента времени по правилу (13) и суммировать ее с прогнозом трендовой составляющей $y^*(t)$. Интерполяция и экстраполяция на моменты времени t , не входящие в исходное множество $\{t_k, k = 1, 2, \dots, n\}$ и множества, полученные из него сдвигами на целое число периодов, может быть осуществлена путем линейной интерполяции ближайших значений или иным методом сглаживания.

Обсудим свойства оценок (12) – (14).

При безграничном росте объема данных и справедливости условий (6) и (7) оценки a^* и d^* параметров трендовой составляющей являются состоятельными и несмещенными, а потому, как можно показать, в рассматриваемых в настоящей статье условиях суммы (12) оценивают периодическую составляющую состоятельно (при $m \rightarrow \infty$) и несмещенно. Как следствие,

$$\frac{1}{n} \sum_{k=1}^n [f^*(t_k)]^2 - \frac{1}{n} \sum_{k=1}^n f^2(t_k) \rightarrow 0 \quad (15)$$

по вероятности при $n \rightarrow \infty$. В соответствии с (10) последнее соотношение дает возможность оценить σ^2 , а затем построить интервальный прогноз для трендовой составляющей согласно (9).

Отметим, что в рассматриваемой ситуации, как правило, n растет, увеличиваясь на величины, кратные q – числу наблюдений в одном периоде. Как следствие, уменьшаемое в (15) – константа, зависимости от n нет. Эти особенности связаны с тем, что выполнение условий (7) предполагает рассмотрение целого числа периодов.

Рассмотрим оценки (12) подробнее. Как вытекает из (4.1.19), (11) и (12),

$$g_s^* = f(t_s) - (a^* - a) \frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t}) - (d^* - d) + \frac{1}{m} \sum_{j=1}^m E_{s+(j-1)q}, \quad s = 1, 2, \dots, q.$$

С учетом (4), (5) и (7) получаем, что

$$g_s^* = f(t_s) - \left(\sum_{k=1}^n c_k E_k \right) \left(\frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t}) \right) - \frac{1}{n} \sum_{k=1}^n E_k + \frac{1}{m} \sum_{j=1}^m E_{s+(j-1)q}, \quad s = 1, 2, \dots, q.$$

Таким образом,

$$g_s^* = f(t_s) + \sum_{k=1}^n h_{ks} E_k, \quad s = 1, 2, \dots, q \quad (16)$$

где $h_{ks} = -c_k r_s - \frac{1}{n} + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $h_{ks} = -c_k r_s - \frac{1}{n}$ при всех остальных значениях индекса суммирования k , здесь $r_s = \frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t})$.

Соотношение (16) означает, что рассматриваемые оценки есть суммы независимых случайных величин, а потому с помощью Центральной предельной теоремы можно построить доверительные интервалы для рассматриваемых значений периодической составляющей (в предположении справедливости условий (6)).

3.6.6. Интервальный прогноз

Точечный прогноз строят по формуле (11) на основе $x^*(t)$ - оценки зависимости, «очищенной» от случайной составляющей, но включающей трендовый и периодический компоненты. Если выполнены условия (7), то

$$Mx^*(t) = x(t) = a(t - \bar{t}) + d + f(t),$$

т.е. оценка $x^*(t)$ является несмещенной.

При справедливости условий (7) с учетом (4), (5) и (16) получаем, что для момента времени t , входящего в исходное множество $\{t_k, k = 1, 2, \dots, n\}$ или в множества, полученные из него сдвигами на целое число периодов,

$$x^*(t) - x(t) = (t - \bar{t}) \sum_{k=1}^n c_k E_k + \frac{1}{n} \sum_{k=1}^n E_k + \sum_{k=1}^n h_{ks} E_k. \quad (17)$$

В (17) при определении значений коэффициентов h_{ks} в качестве s следует взять номер наименьшего из исходных моментов времени $\{t_k, k = 1, 2, \dots, n\}$, отстоящих от рассматриваемого момента t на целое число периодов. С помощью (16) заключаем, что

$$x^*(t) - x(t) = \sum_{k=1}^n w_{ks} E_k,$$

где $w_{ks} = c_k(t - \bar{t} - r_s) + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $w_{ks} = c_k(t - \bar{t} - r_s)$ при всех остальных значениях индекса суммирования k , здесь r_s - то же, что и в формуле (16).

В правой части формулы (17) стоит сумма независимых случайных величин, поэтому оценка $x^*(t)$ является асимптотически нормальной (при справедливости условий (6)) с математическим ожиданием $x(t)$ и дисперсией

$$D(x(t)) = \sum_{k=1}^n w_{ks}^2 D(E_k) = \sigma^2 \sum_{k=1}^n w_{ks}^2. \quad (18)$$

Следовательно, нижняя $x_{нижн}(t)$ и верхняя $x_{верх}(t)$ доверительные границы для прогностической функции (с учетом как трендовой, так и периодической составляющих) имеют вид:

$$x_{нижн}(t) = a^*(t - \bar{t}) + d^* + f^*(t) - \Delta(t), \quad x_{верх}(t) = a^*(t - \bar{t}) + d^* + f^*(t) + \Delta(t),$$

где

$$\Delta(t) = U(\gamma) \sqrt{D^*(x^*(t))} = U(\gamma) \sigma^* \sqrt{\sum_{k=1}^n w_{ks}^2}. \quad (19)$$

Здесь γ - доверительная вероятность, $U(\gamma)$ - квантиль нормального распределения порядка $\frac{1+\gamma}{2}$. В формуле (19) $D^*(x^*(t))$ - состоятельная оценка дисперсии точечного прогноза $x^*(t)$. В соответствии с (18) она является произведением состоятельной оценки σ^* среднего квадратического отклонения σ случайных погрешностей E_k на известную статистику детерминированную функцию от t . Величину σ^* рассчитывают согласно (10) и (15).

3.6.7. Пример применения непараметрического метода наименьших квадратов в модели с периодической составляющей

Обработаем фактические данные ОАО «Магнитогорский металлургический комбинат» о закупочных ценах на лом черных металлов [3]. Как показано в [3], может быть использована модель (1) линейного тренда с периодической составляющей. Для облегчения расчетов оставим из каждого квартала данные только по одному месяцу. Введем условные моменты времени, а именно, будем измерять время в кварталах, начиная с первого квартала 2003 г. Исходные данные для демонстрации примера применения непараметрического метода наименьших квадратов в модели с периодической составляющей - пары чисел (t_k, x_k) , $k = 1, 2, \dots, 12$, - представлены в табл.1 в столбцах (3) и (4) соответственно.

По формулам (3) найдем оценки параметров a^* и d^* , что позволяет построить оценку трендовой составляющей

$$y^*(t) = a^*(t - \bar{t}) + d^* = 212,26(t - 6,5) + 3967,17 = 212,26t + 2587,48.$$

Численные значения трендовой составляющей приведены в столбце (5) табл.1.

Рассчитав отклонения исходных значений закупочных цен от оценок трендовой составляющей (столбец (6) табл.1), возведя их в квадрат и сложив, получаем остаточную сумму квадратов $SS = 4\,539\,214$ и $SS/n = SS/12 = 378\,267,843$.

Таблица 1 – Построение модели прогнозирования цен на лом марки 3А

№ п/п	Периоды времени	Условные моменты времени	Закупочные цены, руб./т	Оценка тренда	Отклонения от оценки тренда	Восстановленные значения	Кажущиеся невязки
k		t_k	x_k	$y^*(t_k)$	$x_k - y^*(t_k)$	x_k^*	$x_k - x_k^*$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	январь.03	1	2 750	2 800	- 50	2 424	326
2	апрель.03	2	3 800	3 012	788	3 545	255
3	июль.03	3	2 900	3 224	- 324	2 655	245
4	октябрь.03	4	3 100	3 437	- 337	3 848	- 748
5	январь.04	5	2 761	3 649	- 888	3 273	- 512
6	апрель.04	6	4 602	3 861	741	4394	208
7	июль.04	7	3 540	4 073	- 533	3504	36
8	октябрь.04	8	5 268	4 286	982	4 697	571
9	январь.05	9	4 307	4 498	- 191	4 122	185
10	апрель.05	10	4 779	4 710	69	5 243	- 464
11	июль.05	11	4 071	4 922	- 851	4 353	- 280
12	октябрь.05	12	5 723	5 135	588	5546	177

Сгруппировав отклонения исходных значений закупочных цен от оценок трендовой составляющей по месяцам (табл.2), наглядно убеждаемся в наличии периодической составляющей. Взяв среднее арифметическое отклонений от тренда за конкретный ме-

сяц, рассчитываем оценку $f^*(t_s)$ периодической составляющей (в соответствии с формулой (12)). Результаты приведены в табл.2.

Рассчитав по формуле (13) оценки периодической составляющей на весь интервал времени и сложив их с оценками трендовой составляющей, получаем в соответствии с формулой (14) оценку зависимости, «очищенную» от случайной составляющей, т.е. восстановленные значения (столбец (7) табл.1). Кажущиеся невязки, т.е. отклонения исходных значений закупочных цен от восстановленных значений, приведены в столбце (8) табл.1. Сравнивая столбцы (6) и (8), убеждаемся в целесообразности введения в модель периодической составляющей. В 9 случаях из 12 абсолютные величины отклонений уменьшились, в остальных трех, хотя и возросли, но лишь до среднего уровня среди остальных.

Таблица 2 – Оценивание периодической составляющей

Номер квартала s	Месяц	Отклонения от тренда			Оценка $g_s^* = f^*(t_s)$ периодической составляющей
		В 2003 г.	В 2004 г.	В 2005 г.	
1	Январь	-50	- 888	-191	- 376
2	Апрель	788	741	69	533
3	Июль	- 324	- 533	- 851	- 569
4	Октябрь	- 337	982	588	411

Возведя в квадрат оценки периодической составляющей (табл.2), сложив эти квадраты, умножив на число лет и поделив на n , получаем, что $\frac{1}{n} \sum_{k=1}^n (f^*(t_k))^2 = 229\,537$. В соответствии с формулой (10) оценкой дисперсии случайной составляющей является

$$(\sigma^*)^2 = \frac{SS}{n} - \frac{1}{n} \sum_{k=1}^n (f^*(t_k))^2 = 378\,267,83 - 229\,537 = 148\,731,$$

а оценкой среднего квадратического отклонения

$$\sigma^* = \sqrt{148731} = 385,7.$$

В соответствии с формулами (4) и (5) оценим дисперсии оценок параметров:

$$D^*(a^*) = \sum_{k=1}^n c_k^2 D^*(E_k) = \frac{(\sigma^*)^2}{\sum_{k=1}^n (t_k - \bar{t})^2} = \frac{148731}{143} = 1040,$$

$$D^*(d^*) = \frac{(\sigma^*)^2}{n} = \frac{148731}{12} = 12394.$$

Средние квадратические отклонения a^* и d^* оцениваются как 32,25 и 111,33 соответственно, а доверительные интервалы для доверительной вероятности 0,95 таковы: $[a_{\min}; a_{\max}] = [149,05; 275,47]$, $[d_{\min}; d_{\max}] = [3748,96; 4185,38]$.

Первое из условий (7) выполнено в силу построения оценок периодической составляющей по целому числу периодов. Действительно, согласно данным табл.2 сумма оценок периодической составляющей для 12 точек наблюдений равна (-3), незначительное отклонение от 0 вызвано ошибками округления.

В соответствии с формулой (5) смещение оценки a^* оценивается как

$$\sum_{k=1}^n c_k f^*(t_k) = \frac{\sum_{k=1}^n (t_k - \bar{t}) f^*(t_k)}{\sum_{k=1}^n (t_k - \bar{t})^2} = \frac{5568}{143} = 38,94.$$

Таким образом, смещение имеет тот же порядок, что и среднее квадратичное отклонение оценки a^* , и заведомо меньше, чем полуширина доверительного интервала. Дальнейшее сравнение может быть проведено на основе оценки дисперсии смещения – случайной величины

$$Z = \frac{\sum_{k=1}^n (t_k - \bar{t}) f^*(t_k)}{\sum_{k=1}^n (t_k - \bar{t})^2}.$$

Алгоритм вычисления дисперсии Z аналогичен таковому для периодической составляющей и интервального прогноза (см. (16) и (18) соответственно), но более сложен, поэтому не включен в статью. Таким образом, можно считать, что предположения (7) модели (1) выполнены для данных табл.1.

Перейдем к оценке дисперсий значений периодической составляющей. Как следует из равенства (16),

$$D(g_s^*) = \sigma^2 \sum_{k=1}^n h_{ks}^2, \quad s = 1, 2, \dots, q,$$

где $h_{ks} = -c_k r_s - \frac{1}{n} + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $h_{ks} = -c_k r_s - \frac{1}{n}$ при иных

значениях индекса суммирования k , здесь $r_s = \frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t})$.

Начнем со значения $s = 1$ (периодическая составляющая для января). Тогда $r_1 = \frac{1}{3}((1-6,5) + (5-6,5) + (9-6,5)) = -1,5$. Понадобятся значения

$$c_k = \frac{t_k - \bar{t}}{\sum_{k=1}^n (t_k - \bar{t})^2} = \frac{t_k - 6,5}{143} = \frac{k - 6,5}{143}.$$

Расчет удобно проводить с помощью таблицы (табл.3).

Таблица 3 – Расчет дисперсии периодической составляющей

k	$t_k - \bar{t}$	$c_k r_1$	$-1/n$	$+1/m$	h_{k1}	h_{k1}^2
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	- 5,5	0,0577	- 0,0833	0,3333	0,3077	0,09468
2	- 4,5	0,0472	- 0,0833	-	- 0,0361	0,00130
3	- 3,5	0,0367	- 0,0833	-	- 0,0466	0,00217
4	- 2,5	0,0262	- 0,0833	-	- 0,0571	0,00326
5	- 1,5	0,0157	- 0,0833	0,3333	0,2657	0,07060
6	- 0,5	0,0052	- 0,0833	-	- 0,0781	0,00610
7	0,5	- 0,0052	- 0,0833	-	- 0,0885	0,00783
8	1,5	- 0,0157	- 0,0833	-	- 0,0990	0,00980
9	2,5	- 0,0262	- 0,0833	0,3333	0,2238	0,05009
10	3,5	- 0,0367	- 0,0833	-	0,1200	0,01440
11	4,5	- 0,0472	- 0,0833	-	0,1305	0,01703
12	5,5	- 0,0577	- 0,0833	-	0,1410	0,01988

В табл. 3 столбец (3) получен из столбца (2) умножением на $\frac{r_1}{143} = \frac{-1,5}{143} = -0,01049$, каждый элемент столбца (6) равен сумма элементов столбцов (3), (4) и (5), стоящих в той же строке, а в столбце (7) стоят квадраты соседних элементов из столбца (6). Цель построения табл.3 – расчет суммы элементов столбца (7). Эта сумма равна 0,28275. Следовательно,

$$\sqrt{D^*(g_1^*)} = \sigma^* \sqrt{\sum_{k=1}^n h_{k1}^2} = 385,7 \times \sqrt{0,28275} = 204,8.$$

Доверительный интервал для значения периодической составляющей в январе ($-376 - 1,96 \times 204,8$; $-376 + 1,96 \times 204,8$) захватывает 0 (при доверительной вероятности 0,95), отличие значения периодической составляющей от 0 не значимо (на уровне значимости 0,05).

Аналогичный расчет для значения $s = 2$ (периодическая составляющая для апреля) дает

$$\sum_{k=1}^n h_{k2}^2 = 0,25524, \quad \sqrt{D^*(g_2^*)} = \sigma^* \sqrt{\sum_{k=1}^n h_{k2}^2} = 385,7 \times \sqrt{0,25524} = 194,86.$$

Доверительный интервал для значения периодической составляющей в апреле ($533 - 1,96 \times 194,86$; $533 + 1,96 \times 194,86$) = ($533 - 381,93$; $533 + 381,93$) не захватывает 0 (при доверительной вероятности 0,95), отличие значения периодической составляющей от 0 значимо (на уровне значимости 0,05).

Приступим к завершающему этапу анализа данных табл.1 – построению интервального прогноза. Необходимо рассчитать величины $w_{ks} = c_k(t - \bar{t} - r_s) + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $w_{ks} = c_k(t - \bar{t} - r_s)$ при всех остальных значениях индекса суммирования k , где r_s – то же, что и в формуле (16), поскольку точечный прогноз $x^*(t)$ является несмещенным, асимптотически нормальным, а его дисперсия оценивается согласно (18) так:

$$D^*(x^*(t)) = (\sigma^*)^2 \sum_{k=1}^n w_{ks}^2.$$

Начнем с прогноза на январь 2006 г. (по данным за 2003 - 2005 гг.). Тогда $t = 13$, $s = 1$, $r_1 = -1,5$, $w_{k1} = 8c_k + \frac{1}{3}$, если $k \in \{1 + 4(j-1), j = 1, 2, 3\}$, и $w_{k1} = 8c_k$ при всех остальных значениях индекса суммирования. При этом $8c_k = 8 \frac{k-6,5}{143} = \frac{8k-52}{143}$. Расчет удобно проводить с помощью таблицы (табл.4).

Сумма значений, стоящих в последнем столбце табл.4, равна 0,61299. Согласно формуле (19)

$$\Delta(13) = U(0,95) \sqrt{D^*(x^*(13))} = 1,96 \times 385,7 \times \sqrt{0,61299} = 591,88.$$

Согласно (14) точечный прогноз прогностической функции таков:

$$x^*(13) = a^*(13 - \bar{t}) + d^* + f^*(13) = 212,26 \times 13 + 2587,48 + (-376) = 4971.$$

Нижняя и верхняя доверительные границы для прогностической функции (с учетом как трендовой, так и периодической составляющих) имеют вид:

$$x_{\text{нижн}}(13) = 4971 - 592 = 4379, \quad x_{\text{верх}}(13) = 4971 + 592 = 5563.$$

Реальное значение согласно [3] равно 4336. Оно практически совпадает с нижней доверительной границей прогностической функции $x_{\text{нижн}}(13)$.

Таблица 4 – Расчет дисперсии прогностической функции

k	$\frac{8k - 52}{143}$	$1/m$	w_{k1}	w_{k1}^2
1	- 0,3077	0,3333	0,0256	0,00066
2	- 0,2517	-	- 0,2517	0,06336
3	- 0,1958	-	- 0,1958	0,03834
4	- 0,1399	-	- 0,1399	0,01957
5	- 0,0839	0,3333	0,2494	0,06220
6	- 0,0280	-	- 0,0280	0,00078
7	0,0280	-	0,0280	0,00078
8	0,0839	-	0,0839	0,00700
9	0,1399	0,3333	0,4732	0,22392
10	0,1958	-	0,1958	0,03834
11	0,2517	-	0,2517	0,06336
12	0,3077	-	0,3077	0,09468

Аналогичные расчеты для апреля 2006 г. ($t = 14$, $s = 2$, $r_2 = -0,5$) дают $\Delta(14) = 1,96 \times 385,7 \times \sqrt{0,72480} = 643,60$. Точечный прогноз равен $x^*(14) = 6092$, а нижняя и верхняя доверительные границы таковы: $x_{нижн}(14) = 5448$, $x_{верх}(14) = 6736$. Реальное значение [3] – 5430. Оно практически совпадает с нижней доверительной границей прогностической функции $x_{нижн}(14)$.

3.6.8. Интервальный прогноз индивидуальных значений

Формула (19) позволяет строить интервальный прогноз для прогностической функции, т.е. для математического ожидания временного ряда. Наблюдаемое значение отличается от него на величину невязки. Распределение невязки можно оценить по значениям кажущихся невязок (см. столбец (8) в табл.1). Напомним, что это распределение не является нормальным, не описывается элементом какого-либо параметрического семейства. Интервальный прогноз индивидуального значения построить, скорректировав интервальный прогноз для прогностической функции с помощью выборочных квантилей кажущихся невязок.

Для рассмотренного выше примера вариационный ряд $n = 12$ кажущихся невязок таков: -748, - 512, - 464, - 280, 36, 177, 185, 208, 245, 255, 326, 571. Нижний дециль оценим как второй член вариационного ряда (-512), верхний – как предпоследний (одинадцатый) член вариационного ряда 326. Для расчета нижней доверительной границы индивидуального значения надо взять нижнюю доверительную границу прогностической функции и отнять 512. Для расчета верхней доверительной границы индивидуального значения надо взять верхнюю доверительную границу прогностической функции и прибавить 326.

Итак, для данных табл.1 индивидуальные значения лежат «глубоко внутри» доверительных интервалов. Прогнозы полностью оправдались.

3.6.9. О проверке условий (7)

Рассмотрим три вопроса. Верны ли условия (7) в моделях, соответствующих реальным ситуациям? Как проверять справедливость условий по результатам наблюдений? Каковы свойства оценок, если эти условия оказываются невыполненными?

В условиях (7) важную роль играет система точек наблюдения t_k , $k = 1, 2, \dots, n$. Более тщательно рассмотрим ранее принятую модель с целым числом периодов, для ко-

торой справедливо соотношение (11). При этом объем наблюдений $n = tq$, где t – количество периодов, q – число наблюдений в одном периоде. Предполагается, что первые q моментов наблюдения при сдвиге на длину периода дают следующие q моментов времени, при сдвиге на две длины периода дают третий набор из q моментов наблюдения, и т.д. Для значений периодической составляющей выше построены точечные оценки и доверительные интервалы (в предположении, что количество периодов t безгранично растет), в чем и состоит оценивание периодической составляющей. (Для гладкой функции $f(t)$ при безграничном росте числа наблюдений q в одном периоде можно получить сходимость оценок периодической составляющей не только в q точках, но и на всем периоде. При этом от оценок в q точках придется перейти к оценкам на всем периоде, например, кусочно-линейным, соединив соседние точки графика отрезками прямых.)

Описанная модель справедлива, когда, например, в течение некоторого числа лет имеются поквартальные или помесечные данные бухгалтерского учета. При изучении посещений сайта или торгового заведения – почасовые данные за целое число недель. Если в ряду наблюдений есть пропуски (временной ряд не является полным) – предположения модели не выполняются. Если система точек наблюдения не образует арифметическую прогрессию,

В рассматриваемой модели естественно принять, что

$$\sum_{k=1}^q f(t_k) = 0, \quad (20)$$

суммарное отклонение значений восстанавливаемой функции от линейного тренда за один период является нулевым. Тогда первое из условий (7) выполнено:

$$\sum_{i=1}^n f(t_i) = m \sum_{k=1}^q f(t_k) = 0 .$$

В реальных ситуациях система точек наблюдения может включать в себя, кроме целого числа периодов, еще несколько начальных точек следующего периода. Можно априори принять первое условие (7), для этого изменив – при необходимости – величину свободного члена d в модели тренда (та же логика рассуждений, что и при принятии условий $M(e_k) = 0$ – в модели без периодической составляющей – и $M(E_k) = 0$ в общем случае). Однако возникает противоречие между первым условием (7) и условием (20). Условие первое условие (7) автоматически обеспечивается методом наименьших квадратов, а условие (20) соответствует логике моделирования. Однако поскольку рассматриваем асимптотическую теорию при безграничном росте числа периодов, указанное различие исчезает при $m \rightarrow \infty$. Таким образом, первое из условий (7) вытекает из свойств рассматриваемой модели и потому вообще не требует проверки по экспериментальным данным, в отличие от второго условия (7), которое выполнено не всегда.

Добавим к модели с целым числом периодов два предположения - симметричности множества $\{t_k, k = 1, 2, \dots, n\}$ относительно \bar{t} и четности периодической составляющей $f(t)$ относительно той же точки. Эти предположения выполнены, если, например, график $f(t)$ симметричен относительно середины года. Тогда второе условие (7) выполнено. Ясно, что обычно нет оснований априори считать, что реальные данные описываются такой моделью.

3.6.10. Проверка второго условия (7) по экспериментальным данным

Естественно использовать статистику

$$Y = \sum_{j=1}^n (t_j - \bar{t}) f^*(t_j),$$

где $f^*(t_j)$ - ранее построенная оценка периодической составляющей $f(t)$. Оценка $f^*(t_j)$ является несмещенной, а потому

$$M(Y) = \sum_{j=1}^n (t_j - \bar{t}) f(t_j).$$

При справедливости (6) распределение Y является асимптотически нормальным (при безграничном росте количества периодов m). Для проверки второго условия (7), т.е. для проверки нулевой гипотезы $H_0: M(Y) = 0$ при альтернативной гипотезе о неравенстве математического ожидания 0 достаточно оценить дисперсию Y .

В соответствии с (11) формулу (16) можно записать для любого $j = 1, 2, \dots, n$, если под $k = k(j)$ понимать $k(j) = j - aq$ при максимально возможном a , при котором $k(j)$ остается положительным, т.е. $k(j)$ - это остаток от деления j на q , если этот остаток ненулевой, и $k(j) = q$ при нулевом остатке. Таким образом,

$$f^*(t_j) = f(t_j) + \sum_{i=1}^n h_{ik(j)} E_i, \quad j = 1, 2, \dots, n, \quad (21)$$

где h_{ik} - те же, что и в формуле (16). В соответствии с определением Y из (21) следует, что

$$Y = \sum_{j=1}^n (t_j - \bar{t}) f^*(t_j) = \sum_{j=1}^n (t_j - \bar{t}) f(t_j) + \sum_{j=1}^n \left[(t_j - \bar{t}) \sum_{i=1}^n h_{ik(j)} E_i \right]. \quad (22)$$

Изменим порядок суммирования во втором слагаемом в (22):

$$\sum_{j=1}^n \left[(t_j - \bar{t}) \sum_{i=1}^n h_{ik(j)} E_i \right] = \sum_{i=1}^n \left[\sum_{j=1}^n (t_j - \bar{t}) h_{ik(j)} \right] E_i.$$

Следовательно, поскольку E_i - независимые одинаково распределенные случайные величины с математическим ожиданием 0 и дисперсией σ^2 , то

$$D(Y) = \sum_{i=1}^n \left[\sum_{j=1}^n (t_j - \bar{t}) h_{ik(j)} \right]^2 \sigma^2. \quad (23)$$

Величину σ^2 оцениваем по формулам (10) и (15), величины $h_{ik(j)}$ описаны после формулы (16). Подставив оценку σ^2 в (23), получаем оценку $D^*(Y)$ дисперсии Y .

В соответствии с асимптотической нормальностью Y правило принятия решений при проверке гипотезы $H_0: M(Y) = 0$ таково: если

$$\left| \frac{Y}{\sqrt{D^*(Y)}} \right| \leq C(\alpha) = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right), \quad (24),$$

где $C(\alpha)$ - критическое значение, соответствующее уровню значимости α , то нулевая гипотеза принимается (второе условие (7) выполнено), если же неравенство (24) не выполнено, то принимается альтернативная гипотеза (второе условие (7) не выполнено).

3.6.11. Асимптотическая несмещенность оценки параметра a

Приведем пример, когда второе условие (7) не выполнено. Измерять время будем в месяцах. Пусть данные берутся на середину квартала. Тогда последовательность моментов времени такова: 2, 5, 8, 11, 14, 17, 20, 23, ... Задан период - год. Периодическая составляющая задается четырьмя числами: $g_1 = -1$, $g_2 = -2$, $g_3 = -3$, $g_4 = 6$. Для таких дан-

ных выполнено равенство (20), т.е. $\sum_{k=1}^q f(t_k) = -1 - 2 - 3 + 6 = 0$. Следовательно, выполнено первое условие (7). Используя это условие, можно упростить второе условие (7):

$$\sum_{i=1}^n (t_i - \bar{t}) f(t_i) = \sum_{i=1}^n t_i f(t_i) - \bar{t} \sum_{i=1}^n f(t_i) = \sum_{i=1}^n t_i f(t_i) = 0.$$

Для простоты расчетов ограничимся двумя годами. Тогда

$$\sum_{i=1}^n t_i f(t_i) = 2 \times (-1) + 5 \times (-2) + 8 \times (-3) + 11 \times 6 + 14 \times (-1) + 17 \times (-2) + 20 \times (-3) + 23 \times 6 = (-2) + (-10) + (-24) + 66 + (-14) + (-34) + (-60) + 138 = 60$$

Второе условие (7) не выполнено. Оно не будет выполнено и для любого иного числа лет. Действительно, если x – начало года (для первого года $x = 0$, для второго $x = 12$, и т.д.), то вклад этого года в рассматриваемую сумму будет равен

$$(x+2) \times (-1) + (x+5) \times (-2) + (x+8) \times (-3) + (x+11) \times 6 = 2 \times (-1) + 5 \times (-2) + 8 \times (-3) + 11 \times 6 = 30$$

Причина нарушения второго условия (7) ясна – периодическая составляющая не симметрична в течение года. Такое поведение периодической составляющей естественно для сельскохозяйственных предприятий. Противоположную ситуацию демонстрирует периодическая составляющая для временного ряда цен на лом черных металлов (по данным Магнитогорского металлургического комбината), проанализированного выше.

Смещение оценки параметра a равно

$$M(a^*) - a = \frac{\sum_{i=1}^n f(t_i)(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2} = \frac{\sum_{i=1}^n f(t_i)t_i}{\sum_{i=1}^n (t_i - \bar{t})^2}. \quad (25)$$

В рассматриваемом примере числитель за m лет равен $30m$. А знаменатель, очевидно, имеет порядок m^3 . Смещение имеет порядок m^{-2} , т.е. быстро убывает с ростом числа периодов. Оценка a^* параметра a является асимптотически несмещенной.

Нетрудно показать, что для модели с целым числом периодов всегда имеет асимптотическая несмещенность оценки a^* параметра a . Если второе условие (7) выполнено – эта оценка является несмещенной, если не выполнено – смещенной, но смещение стремится к 0 при росте числа периодов. Таким образом, выполнение второго условия (7) не является необходимым для применения рассматриваемых методов. Тем не менее проверка второго условия (7) по экспериментальным данным является полезным для решения о том, можно ли пользоваться асимптотической несмещенностью оценки при имеющемся объеме данных.

3.6.12. Обсуждение полученных в разделе результатов

Подведем итоги. По сравнению с эвристическими алгоритмами, разобранными в [1] и других литературных источниках, разработанная в настоящем разделе теория позволила:

- 1) дать общее обоснование этим алгоритмам в рамках асимптотических методов математической статистики и указать условия их применимости (формула (6));
- 2) выявить принципиально важные условия (7), необходимые и достаточные для несмещенности и состоятельности рассматриваемых оценок;

3) построить доверительные интервалы для зависимости (прогностической функции), трендовой и периодической составляющих, индивидуальных значений временного ряда.

Обсуждение отдельных сторон рассматриваемой проблемы проведено в работах [2, 4, 5].

В рамках математической статистики удастся провести анализ не всех распространенных эвристических алгоритмов. Так, довольно часто рекомендуют вначале провести сглаживание («выравнивание») временного ряда, например, методом скользящих средних [1, с.137]. При этом периодическая (сезонная) составляющая меняется (также сглаживается), а погрешности (отклонения от суммы трендовой и периодической составляющих) становятся зависимыми случайными величинами, что делает невозможным применение описанных в настоящем разделе методов.

Теория устойчивости [6] отвергает идею поиска оптимального метода, поскольку зачастую оказывается, что для любого выбранного для рассмотрения метода анализа данных можно подобрать такое понимание оптимальности, что именно этот метод является оптимальным. Например, метод наименьших квадратов в определенном смысле оптимален, если погрешности имеют нормальное распределение, в то время как метод наименьших модулей оптимален, если погрешности имеют распределение Лапласа. В задаче проверки однородности двух независимых выборок установлено [7], что для любого из обычно используемых критериев однородности существует такое распределение на множестве альтернативных гипотез, что рассматриваемый критерий является оптимальным (в том смысле, который определен в [7]).

Работа выполнена в рамках новой парадигмы прикладной (математической) статистики [8, 9]. Изучена *непараметрическую модель*, не предполагающая, что распределение погрешностей (ошибок, невязок) входит в то или иное параметрическое семейство. Второе основание для того, чтобы именовать рассматриваемые модель и метод непараметрическими – оценивание периодической составляющей произвольного вида, т.е. отказ от выбора периодической составляющей из какого-либо параметрического семейства функций.

Полученные в статье [10] научные результаты, касающиеся средних величин и законов больших чисел в пространствах произвольной природы, могут быть применены для анализа данных в различных научных и прикладных областях. В отличие от них результаты настоящей работы нацелены прежде всего на анализ динамических рядов экономических показателей (временных рядов), необходимость которого часто возникает при организационно-экономическом моделировании с целью решения задач управления хозяйственными единицами [11]. Именно потребности экономики и управления ставят во главу угла модели с одной независимой переменной – временем. Длина периода задается существом рассматриваемой прикладной задачи (для оценки длины периода по статистическим данным нужен другой математический аппарат, разработанный в [12]). Рассмотренные в настоящем разделе постановки можно относить к эконометрике [2], т.е. статистическим методам в экономике. Полученные результаты могут быть применены для прогнозирования и построения экономико-математических моделей, в частности, в рамках неформальной информационной экономики будущего [13].

ГЛАВА 4. ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ КОНТРОЛЛИНГА, ИННОВАЦИЙ И МЕНЕДЖМЕНТА В РАКЕТНО- КОСМИЧЕСКОЙ ПРОМЫШЛЕННОСТИ

4.1. Организационно-экономическое обеспечение ракетно-космической промышленности

В течение последние годы в рамках сотрудничества МГТУ им. Н.Э. Баумана с Роскосмосом (прежде всего с Центральным научно-исследовательским институтом машиностроения) нами был проведен ряд исследований по проблемам разработки организационно-экономического обеспечения решения задач управления в ракетно-космической промышленности (РКП), прежде всего в области управления проектами разработки ракетно-космической техники. Настоящий раздел имеет целью подведение предварительных итогов этого цикла исследований.

Очевидно, результаты космической деятельности (КД) используются как для обеспечения обороноспособности государств, так и для решения социально-экономических задач развития страны. Исходя из современной геополитической обстановки, первую из этих задач следует считать основной. Следовательно, в РКП *нельзя следовать догмам "рыночной экономики"*, упирать на конкурентоспособность предприятий и видов деятельности и т.п.

Поскольку основное финансирование КД в России ведется в соответствии с утвержденными государственными органами целевыми программами из государственного бюджета, то среди показателей финансово-хозяйственной деятельности предприятий основное внимание следует уделять не максимизации прибыли, а уменьшению издержек.

Проекты разработки ракетно-космической техники являются весьма ресурсоемкими. Поэтому необходима оценка *реализуемости* проектов в области КД, в первую очередь, по научно-технической обоснованности и удовлетворению социально-экономических потребностей, а также ресурсному обеспечению. Важен анализ всех видов ресурсов - материальных, производственных, кадровых, временных, а не только финансовых.

Из сказанного видна необходимость разработки организационно-экономического обеспечения решения задач управления КД, соответствующего современным внешнеполитическим условиям и достижениям науки. В качестве базовой организационно-экономической теории предлагаем использовать солидарную информационную экономику, менеджмент высоких технологий, контроллинг, развиваемые в МГТУ им. Н.Э. Баумана *на основе новой парадигмы математических методов экономики*, прежде всего эконометрики, теории принятия решений, организационно-экономического моделирования.

Проекты по созданию изделий ракетно-космической техники (РКТ) обладают рядом особенностей по сравнению с проектами в других высокотехнологичных отраслях. Во-первых, в них велика инновационная составляющая, обусловленная необходимостью решения новых научно-технических задач. Как следствие, велики инновационные *риски*. Во-вторых, проекты по созданию РКТ требуют для своей реализации значительных объемов ресурсного обеспечения (трудовых, временных, материальных и производственных ресурсов), значительных инвестиций. Поэтому такие проекты мы с точки зрения конструирования процедур управления рассматриваем как *инновационно-инвестиционные*.

Современная теория управления проектами – основа организационно-экономического обеспечения решения задач управления КД (ОЭО РЗУ КД). Управле-

ние инновационно-инвестиционными проектами в области КД, в частности, оценка эффективности таких проектов, должны исходить *из всей совокупности* социальных, технологических, экологических, экономических, политических факторов. ОЭО РЗУ КД используется на основе системы *контроллинга*, т.е. системы информационно-аналитической поддержки процесса принятия управленческих решений на предприятиях РКП и в отрасли в целом.

При управлении проектами по созданию изделий РКТ необходимо учитывать *риски* их реализации. При оценке реализуемости проектов по созданию РКТ необходимо проводить анализ и оценку рисков, а также применять современные статистические и экспертные методы *прогнозирования* динамики технико-экономических показателей проектов, в том числе на основе *ситуационных* центров.

ОЭО РЗУ КД должно быть основано на современных научных достижениях, в частности, на современной парадигме в области математических методов исследования, включая статистические и экспертные методы принятия решений в условиях неопределенности и риска. Контроллинг в этой области – это разработка процедур *управления соответствием* используемых и вновь создаваемых (внедряемых) организационно-экономических методов поставленным задачам.

Как показывает практика, для ОЭО РЗУ КД достаточно часто должны разрабатываться *новые* организационно-экономические и экономико-математические модели и методы в рамках соответствующих НИР. Для эффективной плановой разработки ОЭО РЗУ КД необходимо создание базы знаний в рассматриваемой области и адекватное наполнение ее современными знаниями *с учетом данных Российского индекса научного цитирования*.

В связи с длительностью реализации проектов разработки ракетно-космической техники отметим необходимость учета инфляции при планировании и оценке финансово-хозяйственной деятельности предприятий, организаций, отрасли в целом.

Сформулированные выше положения соответствуют докладу [1], прочитанному одним из авторов настоящей монографии на XL Академических чтениях по космонавтике (январь 2016 г.). Раскроем содержание этих положений на основе опубликованных ранее научных работ.

4.1.1. Методологические основы разработки организационно-экономического обеспечения решения задач управления космической деятельностью

Методология - это учение об организации деятельности [2]. Методологические основы определяют успех выбранного исследователем варианта решения рассматриваемых задач. Сводка разработанных нами подходов к разработке организационно-экономического обеспечения инновационной деятельности дана в статье [3]. Можно также указать соответствующие разделы монографий [4–6] и статьи [7, 8].

Базовой публикацией в рассматриваемой области, в которой как единое целое рассмотрены основные подходы к разработке организационно-экономического обеспечения решения задач управления в аэрокосмической отрасли, является статья [9]. В ней показано, что для решения стратегических и оперативных задач управления в аэрокосмической отрасли используются разнообразные организационно-экономические модели и методы. Их совокупность назовем организационно-экономическим обеспечением. В статье [9] обсуждаются современные подходы к разработке адекватного организационно-экономического обеспечения в аэрокосмической отрасли. В частности, рассмотрены проблемы оценки эффективности и управления инновационно-инвестиционными проектами по созданию авиационной и ракетно-космической техники.

Дальнейшему развитию методологического анализа проблем разработки организационно-экономического обеспечения решения задач управления космической деятельностью посвящена статья [10]. В ней обсуждаются основания для разработки организационно-экономического обеспечения (ОЭО) в ракетно-космической отрасли (РКО). Рассмотрены проблемы оценки эффективности инновационно-инвестиционных проектов, ОЭО управления проектами по созданию ракетно-космической техники. На основе анализа состояния и перспектив развития выработаны предложения по ОЭО инновационной деятельности в РКО.

Предлагаемые нами подходы были обсуждены и одобрены на XLIX научных чтений памяти К.Э. Циолковского [11], XXXIX и XL академических чтений по космонавтике, посвященных памяти академика С.П. Королева [1, 12].

Перейдем к обсуждению отдельных составляющих предлагаемых нами подходов.

4.1.2. Моделирование рисков при выполнении проектов разработки ракетно-космической техники

Наши разработки организационно-экономического обеспечения в ракетно-космической отрасли начались с вопросов моделирования рисков при выполнении проектов.

Многообразие рисков рассмотрено нами в [13, 14]. Во многих областях - в экономике, управлении качеством, ракетно-космической промышленности, медицине, экологии, при обеспечении безопасности полетов и др. - задачи анализа, оценки и управления рисками имеют много общего. Поэтому мы считаем нужным развивать общую теорию риска, подходы и методы которой позволяют единообразно решать задачи риск-менеджмента в конкретных предметных областях. На основе анализа научных публикаций и отраслевых нормативных документов приходится констатировать, что частные теории риска имеют тенденцию замыкаться внутри себя, создавать свои внутренние стандарты и системы нормативных документов. Отдельно – для банковской деятельности, отдельно – для безопасности полетов, отдельно – для промышленных аварий, и т.д. С целью построения общей теории риска анализируем употребление термина «риск» в различных областях, рассматриваем многообразие видов рисков, даем основные определения в области анализа, оценки и управления риском. Обсуждаем планетарные риски (на уровне Земли в целом), глобальные риски (на уровне одного или нескольких государств), финансовые риски, коммерческие риски (риски на уровне непосредственного окружения компании), производственные (внутренние, операционные) риски, относящиеся к деятельности отдельного предприятия (организации), личные риски. Инструментарий общей теории риска позволяет единообразно решать основные проблемы анализа, оценки и управления риском для всех предметных областей.

В обзорной статье [15] обоснована концепция контроллинга рисков на основе общей теории риска. С точки зрения управления рисками в ракетно-космической промышленности рассмотрено современное состояние риск-менеджмента в нашей стране. Рассказано о работах по контроллингу рисков, выполненных в Лаборатории экономико-математических методов в контроллинге МГТУ им. Н.Э. Баумана.

С целью оценки рисков в ракетно-космической промышленности впервые разработана в общем виде аддитивно-мультипликативная модель оценки рисков (вероятностей рисковых событий). В ней в двухуровневой схеме на нижнем уровне оценки рисков объединяются аддитивно, на верхнем – мультипликативно. Аддитивно-мультипликативная модель применена для оценки рисков (1) выполнения инновационных проектов в вузах (с участием внешних партнеров), (2) выпуска новых инновационных изделий, (3) проектов создания ракетно-космической техники [16].

Особенности оценки рисков в РКП были проанализированы в [17]. Аддитивно-мультипликативная модель применена для оценки рисков проектов создания ракетно-космической техники. Выделено 44 частных риска на нижнем уровне и 8 – на верхнем, соответственно этапам выполнения проекта. Рассмотрены примеры расчетов. Полученные результаты обсуждены на международных конференциях [18, 19].

4.1.3. Оценка реализуемости проектов создания изделий ракетно-космической техники

Один из основных вопросов управления проектами в РКП можно сформулировать так: можно ли реализовать конкретный проект в заданные сроки и с заданными затратами при имеющихся ограничениях на различные виды ресурсов - научно-технические, материальные, производственные, кадровые?

Для обеспечения возможности получения обоснованного ответа на этот вопрос в [20, 21] разработана общая теоретическая модель оценки реализуемости инновационно-инвестиционного проекта. Для конкретизации общей модели выделены типовые этапы разработки проектов в ракетно-космической отрасли. Организационно-экономические подходы к оценке реализуемости проектов по созданию ракетно-космической техники представлены в виде алгоритмов. Они учитывают специфику ракетно-космической отрасли, в силу которой подобные проекты имеют как инновационную, так и инвестиционную составляющие.

В [22] конкретизированы требования к этапам разработки проектов создания изделий ракетно-космической техники. Предложен алгоритм оценки реализуемости таких проектов с учетом их инновационной и инвестиционной составляющих.

Итоги исследований по разработке методов оценки и управления реализуемостью проектов по созданию ракетно-космической техники подведены в кандидатской диссертации В.А. Волкова [23].

4.1.4. О контроллинге научной деятельности в РКП

Под контроллингом [24, 25] понимаем информационно-аналитическую поддержку процесса принятия решений в организации (на предприятии, в территориальном органе власти и т.п.). Выделяя те или иные стороны информационно-аналитической поддержки процесса принятия решений, получаем различные виды контроллинга - контроллинг инноваций [3, 26], контроллинг качества [27] и др.

В [28, 29] нами введено понятие «контроллинг организационно-экономических методов». Даны определения терминов в цепочке «задача – модель – метод – условия применимости». Описана базовая организационно-экономическая модель промышленного предприятия, в рамках которой рассмотрены проблемы разработки современных организационно-экономических методов. Продемонстрирована актуальность разработки теории и методологии организационно-экономического моделирования. В качестве примеров рассмотрено применение статистических методов на различных этапах жизненного цикла продукции, проблемы внутренних рисков на промышленном предприятии и учет инфляции при анализе хозяйственной деятельности организации.

В РКП весьма значима научная составляющая. Для успешного решения задач, стоящих перед РКП, необходимо эффективное управление наукой, т.е. научными работниками, их сообществами (секторами, лабораториями, отделами, отделениями и другими подразделениями) и научно-исследовательскими институтами в целом.

Наука как объект управления проанализирована в [30]. В этой статье наука рассмотрена как отрасль народного хозяйства. Обсуждается взаимоотношение областей

человеческой деятельности, прикладной науки и фундаментальной науки. В качестве примера рассмотрено развитие теории принятия решений и экспертных оценок в ходе выполнения прикладных научных работ в авиации и ракетно-космической промышленности. Подчеркнуто, что основное в науке – новизна результатов. Обсуждается проблема оценки эффективности научной деятельности. Рассмотрены преимущества и недостатки оценок на основе библиометрических баз данных и индексов цитирования, показана основная роль экспертных технологий. Рассмотрена роль глобализации и патриотизма в развитии науки. Показано принципиальное отличие получения знания и продвижения научного результата. Обоснована необходимость проведения развернутых исследований в области науковедения и разработки на их основе научно обоснованных рекомендаций по управлению наукой.

Для создания таких рекомендаций необходимо прежде всего разделить прикладную науку и фундаментальную науку. К первой из них относятся научные исследования, проводимые для нужд конкретных заказчиков. Применительно к тематике настоящей статьи - для нужд ракетно-космической промышленности. Фундаментальная наука не имеет конкретного заказчика, она нацелена на получение нового знания. Эффективность прикладных научных исследований оценивается по результатам решения поставленных задач. Например, осуществление первого космического полета с человеком на борту - это величайший триумф прикладной космической науки, и для обоснования этого вывода не требуется публикации статей и защит диссертаций. В фундаментальной науке нет столь явных свидетельств успеха, для оценки эффективности исследований используют наукометрические и экспертные показатели, основанные на мнениях деятелей фундаментальной науки.

Может ли наука успешно развиваться изолированно от внешнего мира, в рамках отдельно взятой страны? Полагаем, что да [31]. Для прикладной науки лучшим аргументом является успешно реализованный космический проект. Фундаментальная наука также стимулируется прежде всего внутренними причинами. К сожалению, в области науковедения распространены неверные утверждения, на основе которых, к сожалению, принимаются решения, вредные для отечественной науки и нашей страны в целом.

Например, в [32] утверждается: «мировая наука – единый живой организм, а изоляционизм есть полная или частичная закупорка кровеносных сосудов, соединяющих «нашу» часть организма с другими. Хорошо известно, к чему такая закупорка приводит: к гангрене и отмиранию. Если при разработке методов оценки ученых мы будем это учитывать, то у российской науки остается шанс выжить и воспрянуть духом». Из подобных неверных утверждений вытекают вредные для страны решения, например, о нацеливании научных работников на публикацию статей в зарубежных научных журналах. Поскольку такие решения подкреплены финансовыми стимулами, научное сообщество начинает двигаться во вредном для страны направлении (это наше заключение подробно обосновано в [30, 31]).

На приведенное утверждение отвечает С.Н. Гринченко: «Чтобы считаться «организмом», мировая наука, прежде всего, недостаточно автономна. В рамках такой аналогии она скорее «орган» единой системы Человечества, которую как раз следует уподобить «организму». Другими такими «органами» являются «мировое производство», «мировое образование», «совокупность мировых языков» и т.п. Итак, «мировая наука» **не является** «организмом» [33].

В работе К.С. Хруцкого [34] подвергается острой критике тезис «провинциализма», который в отдельных статьях дискуссии на страницах сборника [35, 36] приводится как «диагноз» (объяснение) текущей «болезни» российского научного организма. Он обосновывает формы поддержки развития научной деятельности в российской провин-

ции; а также доказываемся, что ценные наукометрические предложения (прозвучавшие в ходе дискуссии по проблемам управления наукой [35, 36]) заработают именно в выдвигаемых автономных научных сферах.

Из многих актуальных проблем науковедения в [37] рассмотрены методы оценки эффективности и качества работы ученого, научной деятельности подразделения, организации, журнала. Показатели эффективности научной деятельности используются как важная составная часть при оценке вузов, инновационного потенциала предприятий и т.п. Для оценки эффективности научной деятельности естественно использовать хорошо зарекомендовавшие себя в других предметных областях интеллектуальные инструменты. К таким инструментам относятся, в частности, система сбалансированных показателей, основанная на ключевых показателях эффективности (отсюда и название статьи [37]), а также контроллинг, прежде всего контроллинг научной деятельности. Подробно разработаны и широко применяются два инструмента оценки эффективности научной деятельности - наукометрические показатели и экспертные оценки. Их критическому анализу и посвящена статья [37]. Различные варианты манипулирования значениями наукометрических показателей в РФ, по нашей оценке, пока еще применяются сравнительно редко. Возможно, это связано со сравнительно небольшим сроком их использования при управления наукой. Поскольку такой показатель, как число цитирований работ исследователя, позволяет объективно оценить его вклад в науку, то применение этого наукометрического показателя для управления наукой оправдано. В то же время число публикаций и особенно индекс Хирша не позволяют объективно оценить эффективность научной деятельности, особенно с учетом свойств реальных библиометрических баз данных. Экспертные процедуры имеют ряд недостатков. В статье [37] обсуждается реальная эффективность экспертных процедур в таких областях их применения, как присвоение ученых степеней и выборы в государственные академии наук (прежде всего в РАН). Основные принципы экспертизы в рассматриваемых областях остаются неизменными в течение последних 70 лет. На основе анализа практики приходится констатировать недостаточную эффективность экспертных оценок в указанных областях. Обоснование сказанному приведено в статье [37].

С целью создания эффективной системы управления научной деятельностью нами выделена новая область контроллинга – контроллинг научной деятельности. В статье [38] рассмотрены основные проблемы развития этой области, прежде всего проблема выбора ключевых показателей эффективности. Установлено, что стимулированная административными мерами погоня за числом опубликованных статей в научных журналах мешает развитию науки. Методологические ошибки – упор на индексы цитирования, импакт-факторы и т.п. – приводят к неправильным управленческим решениям. Как показывает опыт Великобритании, в управлении наукой необходимо применять экспертизы. Кратко обсуждаются некоторые недостатки сложившейся системе научных специальностей. Предлагается развернуть научные исследования по науковедению и контроллингу научной деятельности. Обсуждаются проблемы контроллинга в научно-исследовательских организациях прикладного профиля.

Следующий шаг - применение контроллинга для рационализации организации деятельности научно-исследовательских институтов [39]. Совершенствование организационных структур позволяет повысить эффективность работы предприятий. Контроллинг персонала на предприятиях типа "Научно-исследовательский институт" является инструментом поддержки принятия кадровых решений, способствует выполнению поставленных стратегических целей и тактических задач. В статье [39] описаны основные виды организационных структур, их свойства, социометрическое исследование как инструмент менеджера, этапы внедрения модели кадрового контроллинга в систему управления персоналом для предприятий типа "Научно-исследовательский институт".

Контроллинг персонала заключается в регламентации процессов управления персоналом, определении контрольных показателей, мониторинге выполнения поставленных целей, учете затрат на осуществление улучшения системы управления и т.д. Он направлен на определение качества, оптимальности и эффективности специфических механизмов, технологий и методов реализации функций управления персоналом. Объективно объем реализации функций управления персоналом зависит от наличия определенного количества материальных, трудовых, финансовых и других ресурсов, от целей предприятия на различных этапах жизненного цикла, а также от численности и квалификации персонала. Качество реализации функций управления персоналом зависит от понимания значимости управления персоналом в деятельности предприятия ее руководством, а также уровня квалификации среднего руководящего звена. Контроллинг функций управления персоналом позволяет сформировать информационную базу для принятия эффективных управленческих решений, с помощью которых можно оптимизировать систему управления персоналом в сложившихся условиях рыночной среды, что является необходимой основой для успешного развития предприятий, работающих в области наукоемкой продукции и услуг.

4.1.5. Разработка в интересах РКП новых математических методов

Решение конкретных задач по созданию организационно-экономического обеспечения РКП потребовало разработки новых математических методов.

Речь идет прежде всего о теории принятия решений. Были проанализированы принятия решений для научно-технического и экономического развития. Конкретные факты, приведенные в статье [40], демонстрируют большое значение в современном мире стратегического менеджмента, методов управления инновациями и инвестициями и роль теории принятия решений в этих экономических дисциплинах. Проведен ретроспективный анализ развития исследований по ядерной физике. Для развития фундаментальной и прикладной науки во второй половине XX в. весьма большое значение имели два события: решение президента США Рузвельта о развертывании атомного проекта (принятое в ответ на письмо Эйнштейна) и совпадение по времени момента завершения разработки и момента окончания Второй мировой войны. Ядерная бомбардировка Хиросимы и Нагасаки определила развитие ситуации в научно-технической сфере на всю вторую половину XX в. Впервые за всю мировую историю руководители ведущих стран наглядно убедились в том, что фундаментальные научные исследования способны принести большую прикладную пользу (с точки зрения руководителей стран). А именно, дать принципиально новое сверхмощное оружие. Следствием явилась широкая организационная и финансовая поддержка фундаментальных и вытекающих из них прикладных научных исследований. Проанализировано влияние фундаментальной и прикладной науки на развитие и эффективное использование новой техники и технического прогресса. В [40] рассмотрено также развитие математических методов исследования и информационных технологий, в частности, миф об "искусственном интеллекте".

При разработке управленческих решений с целью совместного учета и соизмерения различных факторов, частичного снятия неопределенности широко используются рейтинги. В теории принятия решений практически в том же смысле используются термины "обобщенный показатель" или "интегральный показатель". Статья [41] посвящена математической теории рейтингов - инструментов изучения социально-экономических систем. Рассмотрены, прежде всего, линейные рейтинги - линейные функции от единичных (частных) показателей (факторов, критериев), построенные с помощью коэффициентов важности (весомости, значимости). Обсуждаются причины,

влияющие на величины рейтингов. На величину линейного рейтинга влияют три группы причин: способы измерения единичных показателей; выбор набора показателей; значения коэффициентов важности. Подробнее рассмотрены бинарные рейтинги, когда рейтинговая оценка принимает два значения. Для сравнения рейтингов предлагаем использовать новый показатель качества диагностики - прогностическую силу [42]. Существенно, что во многих управленческих ситуациях значительные различия между объектами выявляются при использовании любого рейтинга. Согласно фундаментальным результатам теории устойчивости [43] одни и те же исходные данные целесообразно обрабатывать несколькими способами. Совпадающие выводы, полученные при применении нескольких методов, скорее всего, отражают свойства реальности. Различия – следствие субъективного выбора метода. При использовании результатов сравнения объектов по нескольким показателям (критериям, рейтингам), в том числе в динамике, полезным является выделение множества Парето. В [41] обсуждаются примеры применения теории принятия решений, экспертных оценок и рейтингов при разработке сложных технических систем.

Большое значение имеют методы прогнозирования и принятия решений в условиях неопределенности и риска. В [44] выделены основные источники неопределенностей в различных производственных и экономических ситуациях, на основе полученных результатов рассмотрены роль и задачи прогнозирования при управлении промышленными предприятиями, в частности, в ракетно-космической промышленности. Обсуждаются основные методы организационно-экономического прогнозирования – статистические, экспертные, комбинированные, в том числе форсайт (который оценен в целом отрицательно). Даны предложения по совершенствованию механизмов прогнозирования и планирования для практического использования при создании космических комплексов.

Развитие статистических и экспертных методов прогнозирования для нужд РКП рассмотрено в докладах [45, 46]. Пример нового научного результата в области статистических методов представлен в статье [47], в которой рассмотрена непараметрическая задача восстановления зависимости, описываемой суммой линейного тренда и периодической функции с известным периодом. В [47] получены асимптотические распределения оценок параметров и трендовой составляющей, а также разработаны методы оценивания периодической компоненты и построения интервального прогноза. В рамках модели точек наблюдения, естественной для приложений, обоснованы условия применимости, в частности, установлена асимптотическая несмещенность оценки коэффициента линейного члена. О новых методах сбора и анализа экспертных оценок рассказано в [48, 49].

4.1.6. Разработка в интересах РКП новых экономических методов

По нашему мнению, в РКП базовой экономической теорией должна быть не рыночная экономика, а разрабатываемая нами солидарная информационная экономика (первоначальное название - неформальная информационная экономика будущего). Ее основные идеи проанализированы в [50]. Обосновано использование солидарной информационной экономики (СИЭ) как базовой современной организационно-экономической теории взамен «economics». Стержнем исследований в области СИЭ является прогнозирование развития будущего общества и его экономики, разработка необходимых для будущего организационно-экономических методов и моделей, предназначенных для повышения эффективности процессов управления. Экономика - это наука о том, как производить, а не о том, как делить прибыль. Основное ядро современной экономической теории - это экономика предприятия. В качестве экономической

составляющей государственной идеологии России мы предлагаем использовать солидарную информационную экономику. Согласно солидарной информационной экономике современные информационные технологии и теория принятия решений позволяют на основе «открытого сетевого общества» построить информационно-коммуникационную систему, предназначенную для выявления потребностей людей и организации производства с целью их удовлетворения. Для реализации этой возможности необходима лишь воля руководства хозяйственной единицей, нацеленная на преобразование системы управления этой единицей. В частности, как уже и происходит во всех развитых странах, российское государство должно стать основным действующим лицом в экономике.

Дальнейшее развитие СИЭ дано в [51]. Эта теория опирается на взгляды Аристотеля, с краткого рассмотрения экономических взглядов Аристотеля в соотнесении с основными идеями СИЭ и начинается статья [51]. Затем обосновываем отмирание семьи, частной собственности и государства. Обсуждаем эволюцию денег - от золотых монет к долговым распискам и условным средствам обращения. Констатируем, что, как показал П. Друкер, рыночная экономика осталась в XIX веке, а основное течение (мейнстрим) в современной экономической науке – обоснование несостоятельности рыночной экономики и необходимости перехода к плановой системе управления хозяйством. Рассматриваем влияние информационно-коммуникационных технологий на хозяйственную деятельность. Разрабатываем подходы к организации принятия решений в СИЭ. На основе современных достижений теории принятия решений (прежде всего экспертных процедур) и информационно-коммуникационных технологий в РКП, РФ и на Земле в целом удастся избавиться от хрематистики. Все будут понимать термин "экономика" по Аристотелю.

В интересах РКП кроме методологических и теоретических результатов на основе СИЭ был разработан ряд конкретных экономических методов.

К управлению инновационно-инвестиционными проектами относится статья [52]. Оценка погрешностей характеристик финансовых потоков инвестиционных проектов необходима для принятия адекватных управленческих решений, в частности, в ракетно-космической промышленности. Организационно-экономические подходы к оценке реализуемости инновационно-инвестиционных проектов создания изделий ракетно-космической техники предполагают интенсивное использование числовых характеристик финансовых потоков многолетних проектов рассматриваемого типа. В организационно-экономическом обеспечении решения задач управления в аэрокосмической отрасли предусмотрена необходимость получения оценок погрешностей характеристик финансовых потоков. Такие оценки - неотъемлемая часть организационно-экономического обеспечения инновационной деятельности в ракетно-космической отрасли. Их можно сравнить с интервальными прогнозами, т.е. доверительным оцениванием прогнозных значений. Половина длины доверительного интервала - это и есть оценка погрешности прогнозирования. В статье [52] разработан новый метод оценки погрешностей основных характеристик инвестиционных проектов. Основное внимание уделено чистой текущей стоимости NPV . Метод оценки погрешностей основан на результатах статистики интервальных данных [53], являющейся неотъемлемой частью системной нечеткой интервальной математики [54]. Построена асимптотическая теория, соответствующая малым отклонениям коэффициентов дисконтирования. Погрешность NPV найдена как асимптотическая нотна. С точностью до бесконечно малых более высокого порядка погрешность NPV является линейной функцией от максимально возможной погрешности коэффициентов дисконтирования.

В начале настоящего раздела отмечена необходимость учета инфляции при принятии управленческих решений в РКП. Статья [55] посвящена работам нашего научно-

го коллектива по сбору и анализу независимо собранной информации о ценах, т.е. по изучению реальной инфляции. Подход к измерению роста цен основан на выборе и фиксации инструмента экономиста и управленца - потребительской корзины, не меняющейся со временем. На основе физиологических норм потребления Института питания РАМН нами составлена минимальная потребительская корзина, т.е. указан годовой объем потребления по основным продовольственным товарам, необходимый для поддержания нормальной жизнедеятельности человеческого организма. В 1993 - 2016 гг. нами проводился независимый сбор цен. Получены стоимости потребительской корзины и индексы инфляции. Проведено сравнение с данными официальной статистики. Наша работа направлена на ликвидацию монополии Росстата при расчетах индексов инфляции, величин прожиточного минимума и реальных располагаемых денежных доходов населения. Использование одной и той же потребительской корзины обеспечивает возможность сопоставления результатов расчетов за различные временные периоды. Этим наши работы выгодно отличаются от подхода официальной статистики. Дан более подробный анализ инфляции в XXI веке. Кратко рассмотрено использование индексов инфляции при анализе проблем домохозяйств, организаций и предприятий, страны в целом.

Подводя итоги, можно констатировать, что в настоящем разделе описано решение ряда актуальных для РКП научных задач. При их решении понадобилось разработать ряд новых математических и экономических методов, которые могут быть с успехом применены в различных отраслях народного хозяйства и областях научных исследований. Но в первую очередь они были применены в интересах РКП. Таким образом, для решения прикладных задач понадобилось получить фундаментальные научные результаты, которые затем были применены в исходной прикладной области (в РКП). Итак, на примере наших работ продемонстрирована диалектическая связь прикладных и фундаментальных научных исследований.

4.2. Основные идеи солидарной информационной экономики - базовой организационно-экономической теории

Качество государственного управления, успешность осуществления государственных политик в разных сферах жизнеустройства страны во многом, если не в основном, определяются методами государственного управления народным хозяйством [1]. Практика экономической жизни, т.е. управления народным хозяйством в целом и его составными частями (регионами, отдельными организациями и др.), опирается на экономическую теорию. Как подробно показано в курсе лекций С.Г. Кара-Мурзы [2, 3], в нашей стране экономическая теория, как и обществоведение в целом, находится в кризисном состоянии и не может быть надежной основой для принятия конкретных управленческих решений. К сожалению, мы в МГТУ им. Н.Э. Баумана долго использовали внедренную с Запада «economics» в качестве базовой теории, на основе которой выполнялись конкретные разработки. Осознав неадекватность «economics», летом 2007 г. мы на основе идей научной школы МГТУ им. Н.Э. Баумана в области экономики и организации производства приступили к разработке новой базовой организационно-экономической теории как замены «economics».

В настоящее время мы рассматриваем внедрение «economics» в начале 1990-х годов как один из эффективных способов [4] борьбы с российской государственностью.

Разразившийся вскоре мировой экономический кризис выявил необходимость немедленной разработки новых организационно-экономических механизмов управле-

ния экономическими системами. Организация производства должна быть основана на адекватной экономической теории. Подтвердилось, что этой теорией не может быть рыночная экономика (economics). Нужна другая теоретическая основа. По нашему мнению, следует исходить из солидарной информационной теории - новой базовой организационно-экономической теории, разрабатываемой в Научно-образовательном комплексе «Контроллинг и управленческие инновации» Московского государственного технического университета им. Н.Э. Баумана. Солидарная информационная экономика построена на основе экономики предприятия (инженерной экономики), теории управления (современного менеджмента, теории активных систем и принятия решений), современных информационных технологий.

Перспективные организационно-экономические механизмы управления [5] производственно-хозяйственной деятельностью предприятий, интегрированных производственно-корпоративных структур, регионов и страны в целом предлагаем конструировать на основе солидарной информационной экономики, разрабатываемой как методологическая основа конкретных исследований в области организационно-экономического моделирования, экономики и управления народным хозяйством, прежде всего в области промышленности, менеджмента, инноваций.

По нашим наблюдениям, основное течение (мейнстрим) в современной экономической науке – обоснование несостоятельности рыночной экономики и необходимости перехода к плановой системе управления хозяйством. Характерны слова директора Центрального экономико-математического института РАН академика РАН В.Л. Макарова: "Я сторонник проектной экономики развития, которая должна прийти на смену рыночным отношениям, причем все проекты должны быть не только коммерческими, но и социально ориентированными. И наша задача – создавать модели развития общества, которые учитывали бы материальные аспекты и духовные ценности россиян" [6]. Дискуссии идут о выборе наиболее адекватного варианта плановой системы. Например, внутри предприятия или корпорации план должен быть жестким и одновременно позволяющим адекватно реагировать на нештатные ситуации, в рамках региона или страны в целом - индикативным.

К мейнстриму плановой экономики относится разрабатываемая нами с 2007 г. солидарная информационная экономика. Ранее мы использовали термин «неформальная информационная экономика будущего». На 26 сентября 2016 г. основной Интернет-ресурс [7] по солидарной информационной экономике просмотрен более 112,4 тыс. раз, издана 44 публикации [8] (статьи и тезисы докладов).

Общепризнано, что управленческие решения необходимо принимать на основе всей совокупности социальных, технологических, экономических, экологических, политических факторов. Это касается прежде всего стратегических решений. Для практически работающих руководителей, как авторы настоящей монографии, очевидно, что экономическая теория – часть менеджмента как науки об управлении людьми [9]. Согласно солидарной информационной экономике современные информационные технологии и теория принятия решений [10, 11] (включая экспертные технологии [12, 13]) позволяют построить информационно-коммуникационную систему, предназначенную для выявления потребностей людей и организации производства с целью их удовлетворения. Для реализации этой возможности необходима лишь воля руководства хозяйственной единицей, нацеленная на преобразование системы управления этой единицей. В частности, как уже и происходит во всех развитых странах, российское государство должно стать основным действующим лицом в экономике.

Эффективному решению современных проблем государственной политики и управления мешают широко распространенные неадекватные представления о рациональном ведении хозяйства. По оценке ведущего американского исследователя П. Дру-

кера, 1873 г. – «конец эры либерализма – конец целого столетия, на протяжении которого политическим кредо была политика невмешательства в экономику» [14]. Но и сейчас, 150 лет спустя, архаичное представление о рациональности рыночных отношений, о «невидимой руке рынка» широко распространено в России и мешает инновационной модернизации систем управления.

По мнению Джозефа Стиглица, лауреата Нобелевской премии по экономике, «экономисты виноваты в кризисе, но есть шанс исправить дело». Говоря о моделях, на которых базируется «экономическая теория», Стиглиц констатирует, что они «провалились полностью, и решения, на них основанные, оказались неверными» [15]. Очевидно, что Стиглиц признал неадекватность буржуазной экономической теории Запада. Следствие – она должна быть заменена на новую теорию, соответствующую реалиям XXI в. Очевидно, эта новая теория – солидарная информационная экономика - должна использоваться в науке, управлении и преподавании.

Соображения, приведенные в настоящем разделе монографии, достаточно просты. Аналогичные мысли высказывали многие авторы. Однако лица, принимающие решения, зачастую исходят из давно устаревших и методологически неверных представлений, ведущих к неверным управленческим решениям. Поэтому изложение основных идей солидарной информационной экономики представляется полезным.

4.2.1. Исходные идеи солидарной информационной экономики

Поясним термины, составляющие название рассматриваемой организационно-экономической теории. Термин «солидарная» (как и первоначальный не вполне точный термин "неформальная") подчеркивает необходимость и возможность совместной деятельности независимых экономических агентов, отсутствие формальной иерархической схемы, добровольность совместной деятельности, принятие удовлетворяющих всех решений в результате переговоров и компромиссов, преобладание синергетической самоорганизации и роевых структур, отсутствие формального принуждения. Вслед за П.А. Кропоткиным [16] мы рассматриваем взаимную помощь как двигатель прогресса.

Термин «информационная» отражает все возрастающую роль информационно-коммуникационных технологий, в том числе сетевых, развитие которых позволяет предсказать в области управления хозяйством и обществом в целом революционный «переход количества в качество», в частности, переход от представительных органов к прямому действию, от государства как аппарата насилия к общественному договору.

Термин «экономика» означает, что рассматривается производственная, организационно-экономическая сторона деятельности общества. Под экономикой в соответствии с Аристотелем [17] понимается управление хозяйством.

В первоначальном варианте названия рассматриваемой организационно-экономической теории был термин «будущее». Он подчеркивал ориентацию исследований на прогнозирование и конструирование будущего развития хозяйственных систем, без привязки к сложившимся традициям хозяйствования. Таким образом, рассматривалась первая задача нормативного прогнозирования – формулировка цели развития (целеполагание). Мы отказались от использования термина «будущее» потому, что «будущее уже наступило». Наша задача - выработка рекомендаций для руководителей хозяйственных структур, государственных и муниципальных органов, предназначенных для применения в ближайшие годы и в среднесрочной перспективе, а не в неопределенном будущем.

Обсудим исходные идеи солидарной информационной экономики.

Повторим, общепризнанно, что при принятии управленческих решений необходимо учитывать не только экономические, но и технологические, социальные, экологи-

ческие, политические факторы [9]. Экономика в целом – служанка общества, выполняет его требования [18]. Цели общества первичны, экономические механизмы вторичны, предназначены для реализации потребностей общества [19]. Несмотря на очевидность сказанного, иногда пытаются экономические факторы считать основными, например, ставить во главу угла прибыль (обычно не уточняя, какой именно тип прибыли из многих [20] имеется в виду).

Цели общества определяются потребностями общества. Исходим из того, что эти потребности могут быть сформулированы и согласованы обществом. Из множества индивидуальных и групповых предложений в результате обсуждений и компромиссов могут быть выявлены потребности общества в целом. Для малой группы (семьи, клана) выявление общих потребностей миллиарды раз осуществлялось в истории человечества. По мере роста масштаба человеческой общности применялись различные варианты агрегирования потребностей – сходка общины, новгородское вече [21], власть самодержца, представительная демократия, использование государственных органов. Решение общегосударственных задач должно сочетаться с обеспечением прав и свобод отдельных граждан и групп. Однако до недавнего времени не было видно путей решения основной проблемы – учета и согласования мнений всех заинтересованных лиц из-за большого их числа. Развитие информационных технологий позволяет использовать необходимые вычислительные ресурсы. Теория и практика разработки и принятия управленческих решений, в частности, методы экспертных оценок, дают возможность разработать эффективные справедливые процедуры выявления общественных потребностей.

Стихийное развитие идет именно в этом направлении, особенно если рассматривать его в масштабе столетий. Значит, нужны работы (к ним относится и наша), нацеленные на прогнозирование развития методов принятия решений в больших системах. О масштабе необходимых управленческих инноваций говорит то, что в результате разработки методов выявления потребностей общества будет фактически устранено государство как посредник между физическими лицами и обществом – произойдет отмирание государства с заменой его на непосредственное народоправие.

При управлении хозяйственной системой (предприятием, корпорацией, государством) самое сложное – целеполагание. Какие потребности удовлетворять, другими словами, как сформулировать цель, поставить задачу оптимизации? Фантасты (как футурологи, исследователи будущего, разработчики моделей развития человечества) предложили различные варианты решений. Например, И.А. Ефремов описал будущее общественное устройство, подобное структуре человеческого мозга: постоянно действующий форум со своими исследовательскими и координационно-ассоциативными центрами [22]. Спроектировать подобную систему разработки и принятия управленческих решений – задача современной теории принятия решений, интенсивно использующей информационные технологии.

Если цель поставлена, то для ее достижения можно и нужно разработать оптимальный план (в натуральных единицах измерения) и отследить его выполнение. Методологически это более простая задача, чем выявление потребностей. Но и здесь до недавнего времени не хватало вычислительных мощностей. К настоящему времени составление оптимального плана производства и распределения продукции и услуг в масштабах Земли в целом – вполне решаемая задача, как показывает опыт крупных транснациональных компаний.

4.2.2. Предшественники и единомышленники

Весьма актуальными для разработки солидарной информационной экономики являются взгляды Аристотеля. Его размышления являются поразительно современными.

В книгах по истории экономической мысли обычно пишут так: «Первым, кто подверг анализу экономические явления и попытался выявить закономерности развития общества стал древнегреческий мыслитель Аристотель (384-322 г. до н.э. - согласно традиционной хронологии - А.О.). Поэтому его можно с полным правом назвать первым экономистом в истории науки» [23].

Выделим положения, на которых базируется солидарная информационная экономика.

По Аристотелю «экономика» - наука о разумном ведении хозяйства. В современных терминах речь идет об организации производства и экономике предприятия, государственном и муниципальном управлении, поскольку Аристотель обсуждает управление хозяйствующими субъектами разного уровня – предприятие, город (полис), регион (сатрапия), государство [17].

Аристотель ввел специальный термин «хрематистика», под которой он понимал деятельность, направленную на извлечение прибыли, на накопление богатства, в отличие от экономики – как деятельности, направленной на удовлетворение потребностей людей, т.е. на производство и приобретение благ для дома и государства. Хрематистику как форму организации хозяйства Аристотель считал противоестественной. Его особое негодование вызывал процент, который он расценивал как самую противоестественную форму дохода.

В настоящее время хрематистика – это прежде всего деятельность банков, фондовых бирж и других структур, действующих по схеме «деньги – деньги», без производства товаров и услуг. Произошла подмена понятий – хрематистику некоторые лица стали называть экономикой.

Вслед за Аристотелем мы полагаем, что деятельность, направленная на извлечение прибыли, является противоестественной, вредной для общества. Экономическая деятельность всегда направлена на удовлетворение потребностей людей, в отличие от хрематистики.

Анализируя природу денег, Аристотель настаивал на том, что деньги являются результатом соглашения между людьми и «в нашей власти сделать их неупотребительными»... Вслед за Аристотелем мы полагаем, что ведение хозяйства вполне возможно без денег. Наш опыт работы в качестве руководителя (менеджера) в Группе авиакомпаний «Волга-Днепр» показал, что в управленческой деятельности решение финансовых вопросов составляет лишь небольшую часть.

Однако на период перехода от современности к будущему хозяйству использование денег целесообразно. Согласно Аристотелю, если деньги относятся к «экономике» - то это символ стоимости, обусловленный законом или обычаем, а если к «хрематистике» - то они выступают как реальный представитель неистинного богатства».

Рассмотрим дальнейшее развитие представлений об экономике и хрематистике.

Френсис Бэкон предупреждал: «Противоестественно, чтобы деньги рождали деньги» [24, с.445]. Отметим: Ф. Бэкон против хрематистики, против процента.

Робинзон Крузо (согласно роману Дефо), попав на необитаемый остров, успешно организовал свое хозяйство. Для этого он разрабатывал методы организации производства, вел управленческий учет, прогнозировал и планировал. Робинзон управлял своим предприятием без использования денег.

Жюль Верн в романе «Таинственный остров» рассказал о том, как успешно работали инженер С. Смит и его товарищи. Они вышли на современный им научно-технический уровень, не обращаясь к хрематистике, поскольку само понятие прибыли было их коммуне чуждо.

Генри Форд в книге «Моя жизнь. Мои достижения» [25] писал: «... Задача предприятия – производить для потребления, а не для наживы или спекуляции... Работу на

общую пользу ставь выше выгоды...». Выделим мысль Форда: цель производства – удовлетворение потребностей, а не получение прибыли.

Неслучайно один из наиболее продуктивных и известных современных экономистов проф. В. Н. Лившиц присоединяется [26] к чеканным формулировкам Джона Мейнарда Кейнса: «Алчность – грех, давать деньги в рост преступно, а любовь к деньгам отвратительна» [27, с. 67] (см. также рецензию [28] на замечательную по своей глубине и охвату книгу [26]).

Эффективные механизмы принятия и реализации плановых решений должны опираться на современные информационные технологии. База разработки таких технологий – кибернетика (вспомним работы Н. Винера, А. И. Берга, Н. Н. Моисеева, многих других). В послевоенные годы в нашей стране, как и во всем мире, разрабатывались различные типы автоматизированных систем управления. Наиболее грандиозный проект - в начале 1960-х гг. В. М. Глушков предложил правительству СССР создать Общегосударственную автоматизированную систему управления экономикой страны (ОГАС), для чего, по его оценкам, требовалось как минимум 15-20 лет и 20 млрд. тогдашних рублей, однако выигрыш стоил того: ОГАС давала реальный шанс построить самую эффективную экономику в мире. В. М. Глушков писал [29]:

«Отныне только «безмашинных» усилий для управления мало. Первый информационный барьер или порог человечество смогло преодолеть потому, что изобрело товарно-денежные отношения и ступенчатую структуру управления. Электронно-вычислительная техника – вот современное изобретение, которое позволит перешагнуть через второй порог. Происходит исторический поворот по знаменитой спирали развития. Когда появится государственная автоматизированная система управления, мы будем легко охватывать единым взглядом всю экономику. На новом историческом этапе, с новой техникой, на новом возросшем уровне мы как бы «проплываем» над той точкой диалектической спирали, ниже которой, отделенный от нас тысячелетиями, остался лежать период, когда свое натуральное хозяйство человек без труда обозревал невооруженным глазом».

Своеобразная ОГАС, хотя и куда более скромная по масштабам, была введена на практике в другом уголке Земли – в Чили, во время президентства Сальвадора Альенде. Один из основоположников кибернетики Стаффорд Бир разработал автоматизированную систему управления национализированными предприятиями Чили. Проект получил название «Киберсин» [30]. Он представлял собой автоматизированную систему сбора и обработки информации, которая состояла из четырех основных компонент: «Кибернет» - система связи (на основе телексов), «Киберстрайд» - компьютерные программы, «Чико» - математическая модель чилийской экономики - и ситуационная комната, из которой велось управление (зал с экранами, на которых отображалось в виде графиков и схем состояние экономики Чили). Можно было управлять производством всей страны в реальном времени (каждым конкретным предприятием), сразу же видеть результаты принятых решений и при необходимости вносить поправки. Кроме того, в каждом населенном пункте создавались «опросные пункты», где производился автоматизированный опрос населения по поводу принимаемых мер. Эти центры были включены в систему «Киберсин», и правительство быстро узнавало реакцию населения на очередное нововведение. Разработки Бира дают прообразы (прототипы) для следующего этапа развития информационных систем управления предприятиями и их объединениями – интегрированными производственно-корпоративными структурами, а также регионами – муниципальными образованиями, субъектами федерации, Россией, международными объединениями, Землей в целом.

Многочисленные работы посвящены конкретным проблемам развития информационных технологий управления. Вопросам информационного обеспечения солидар-

ной информационной экономики посвящены разработки коллектива молодых исследователей, выпустивших специальный номер журнала «Великое кольцо» [31]. В названии журнала подчеркивается связь с идеями И. А. Ефремова. В частности, один из основных авторов этого журнала И. Герасимов считает, что информационные технологии управления уже на сегодняшнем этапе позволяют разработать и внедрять интегрированные информационно-управляющие системы, предназначенные для координации людей, ресурсов, потребностей, предложений; объединения в рабочие группы по реализации экономических проектов; установления прямых связей между производителями и потребителями; оптимальной координации инициатив и проектов в масштабе всего общества. К конкретным функциям в рамках экономического блока можно отнести: учет и распределение ресурсов, обмен экономическим опытом и технологиями, выявление потребностей населения в товарах и услугах, формирование коллективов новых хозяйствующих субъектов, аккумуляция и распределение инвестиций, координация действий трудовых коллективов, публичная оценка потребителями работы хозяйствующих субъектов, ведение публичного диалога между потребителями и производителями товаров и услуг, публикация жалоб добросовестных хозяйствующих субъектов на деструктивные действия официальных и неофициальных паразитических структур, выработка схем оптимального ресурсообмена и планов экономического развития [32].

Новым по сравнению с временами В. М. Глушкова и С. Бира является широкое распространение сетевых технологий, позволяющее аппаратно реализовать право граждан на участие в принятии касающихся их решений. Не менее важным, чем участие в экономической жизни, является наделение всех граждан возможностями по непосредственному участию в политическом управлении обществом. Это, в частности, формирование общественных советов и рабочих групп по коллективному изучению и решению тех или иных социальных проблем, оценка качества работы должностных лиц, разработка, обсуждение и оценка нормативных документов, выдвижение собственных предложений и доведение их до управленческих структур, организация публичного диалога между административными органами и населением, обсуждение кандидатов на выборные должности, публикация сведений о характеризующих личность граждан совершенных ими социально значимых действиях, мобилизация населения на акции прямого гражданского волеизъявления... Качественно новым уровнем по сравнению с разрозненными сайтами фирм и административных учреждений должны стать интегрированные порталы целых секторов экономики и территорий. Эти порталы должны обладать выраженной обратной связью, формироваться по единым стандартам, подчиняться принципу «открытой архитектуры», быть соединенными между собой каналами регулярного обмена данными и находиться под контролем формируемых населением общественных советов различных уровней.

Приведем пример институализации неформальных акций с помощью информационных технологий. В США по инициативе ученых Массачусетского технологического института - разумеется, вопреки воле официальных властей - была разработана интерактивная сетевая база данных, позволяющая гражданам свободно добавлять и получать информацию, касающуюся должностных лиц и политиков. Задуманная как «асимметричный ответ» на создание системы тотальной слежки правительственных органов за простыми людьми, она получила название GIA - Government Information Awareness, или «Информационная осведомленность о правительстве». Таким образом, передовой мировой опыт также указывает верное направление развития.

Открытый процесс создания реальных организационных модулей системы, привлечения участников, прокладывания горизонтальных связей, осуществления операций с ее помощью, по мнению И. Герасимова [32], можно - как синтез англоязычных терминов Open Source, Open Architecture, Commons-based Peer Production, Peer-to-Peer и

Open Society («Открытое общество») – назвать Open P2P Society. По-русски этот термин будет звучать как «Открытое сетевое общество», где топология связей будет не иерархической, а «от каждого к каждому», или «Peer to Peer».

Одним из основных направлений развития современной экономической и управленческой мысли являются информационные технологии управления, причем не только на уровне предприятия, но и на макроуровне. Их математической основой может являться теория управления организационными системами, развиваемая неформальным научным коллективом вокруг Института проблем управления РАН [33]. Выделяют такие ее разделы, как теория активных систем, теория принятия решений, в том числе на основе экспертных технологий. Все они применяются в солидарной информационной экономике.

4.2.3. Механизмы организации хозяйственной деятельности согласно солидарной информационной экономике

Оставшиеся с прошлых веков привычные принципы управления как экономикой, так и государством входят во все большее противоречие с возможностями в сфере технологий управления, предоставляемые прогрессом информационных технологий. Падает роль представительной демократии, при которой подавляющее большинство граждан делегируют полномочия по принятию решений специалистам-депутатам. Возрастает роль неформальной, «роевой» деятельности, как противовеса иерархическим структурам. На основе современных информационных технологий и методов разработки и принятия управленческих решений возникла возможность и необходимость реализации идеи В. М. Глушкова и Ст. Бира.

В перспективе путем предварительного обсуждения и планирования можно будет снять проблему нерационального производства товаров и услуг. Управление организацией предполагает необходимость плана. Делегирование полномочий не отменяет централизацию принятия решений. Отметим, что в рамках плановой системы можно смоделировать любые рыночные отношения, а потому плановое хозяйство заведомо не менее эффективно, чем рыночное. Основная критика плановой системы в масштабах государства базировалась на невозможности произвести необходимое количество вычислений, в результате плановые решения запаздывали и не могли охватить необходимую номенклатуру товаров и услуг. Как показали шотландские экономисты В. Пол Кокшотт и Аллин Ф. Коттрелл [34], современные информационные технологии снимают эти проблемы. Отметим, что с плановым подходом несовместимо перевыполнение плана, планирование от достигнутого.

Возможность самореализации личности, быстрота реализации творческих решений, полностью реализуется в продвинутом варианте планового хозяйства, построенного на основе современной теории принятия решений. Возможно полностью удовлетворить индивидуальные пожелания потребителей.

В реальных ситуациях процессы управления экономическими системами реализуются с достаточно высоким уровнем неопределенности. Велика роль нечисловой информации как на «входе», так и на «выходе» процесса принятия управленческого решения. Неопределенность и нечисловая природа управленческой информации должны быть учтены путем проведения анализа устойчивости экономико-математических методов и моделей, например, на основе разработанной нами общей схемы устойчивости [35–37]. Для обоснованного практического применения математические модели процессов управления экономическими системами и основанных на них экономико-математических методов должна быть изучена их устойчивость по отношению к допустимым отклонениям исходных данных и предпосылок моделей.

Механизмы организации производства, хозяйственной деятельности, экономической жизни могут быть разными. Использование таких изобретений человечества, как денежное обращение и конкуренция (а тем более фондовые рынки и опционы), совсем не обязательно. Робинзон и община «Таинственного острова» обходились без них – небольшим общинам в хозяйственной деятельности деньги не нужны. Развитие информационных технологий сняло ограничение на размер хозяйства. Современный шотландский экономист П. Кокшотт [38] демонстрирует теоретическую возможность организации производства с целью удовлетворения потребностей в масштабах страны или человечества в целом. Особо отметим выполненный им тщательный анализ объемов вычислений, необходимых для этого. Как показал П. Кокшотт, мощностей стандартных современных компьютеров вполне достаточно для расчетов управленческих решений на базе экономико-математических моделей. Таким образом, применять инструменты управления хозяйственной деятельностью, основанные на деньгах, теоретически не обязательно.

Как писал один из основоположников научного менеджмента А. Файоль [39], «управлять – значит прогнозировать и планировать, организовывать, руководить командой, координировать и контролировать» (ни слова про деньги, конкуренцию и коммерческую тайну – внутри предприятия они не обязательны, более того, неуместны). Метафорой предприятия является семья, поскольку исходной формой является семейное хозяйство. Разве нужны деньги, разве извлекают прибыль внутри одной семьи?

На уровне предприятия или корпорации никто не отрицает необходимость планирования и контроля. Взгляды А. Файоля вполне созвучны современной концепции контроллинга [40]. Однако на уровне государства отдельные лица все еще считают эффективным «рынок», т.е. анархию производства, при которой решения отдельных субъектов экономической жизни не требуют согласования между собой.

Денежное обращение должно уйти в прошлое. Ему, как и «рыночным отношениям», нет места в солидарной информационной экономике будущего. Прежде всего это касается производных финансовых инструментов (деривативов), позволяющих участникам финансового рынка действовать по схеме «деньги – деньги», исключив из классической схемы «деньги – товар – деньги» центральный элемент. Для предотвращения экономических кризисов должен быть ликвидирован финансовый капитал, не обеспеченный материальными и интеллектуальными ценностями. В частности, это касается таких деривативов, как валютный своп, кредитный дефолтный своп, опцион, процентный своп, свопцион, соглашение о будущей процентной ставке, форвард, фьючерс, варрант, конвертируемые облигации, депозитарная расписка, кредитные производные и др. Следующий шаг – прекращение обращения (выпуска, котировок, продаж) акций, ликвидация ценных бумаг, в том числе государственных, закрытие валютных и товарных бирж. На переходный период могут быть сохранены деньги как средство обмена и средство соизмерения ценностей. По нашему мнению, они должны быть обеспечены не золотом или энергией, а рабочим временем (одна денежная единица соответствует одному часу неквалифицированного труда).

Необходимость ликвидации частной (но не личной) собственности и процента также вполне очевидна, поскольку управление организациями осуществляется согласно солидарной информационной экономике не личностью, а коллективно. Проект ликвидации денежного обращения и частной собственности в прошлом выдвигался неоднократно, но рассматривался как утопия. Однако современные информационные технологии и теория принятия решений позволяют претворить этот проект в жизнь.

Солидарное принятие решений – важнейшая черта будущих информационных систем принятия решений, разработанных на основе рассматриваемой теории. В-первых, свобода передачи информации – от каждого к каждому, без административных

фильтров. Во-вторых, участие всех заинтересованных лиц и организаций в выработке и реализации решений (принцип прямой демократии). Конечно, для этого необходимы адекватные процедуры принятия решений. Разработать их должны специалисты по теории принятия решений. Необходима техническая база, позволяющая реализовать «Открытое сетевое общество». Наконец, необходима политическая воля для ликвидации административных препятствий на пути внедрения достижений солидарной информационной экономики – базовой организационно-экономической теории, разрабатываемой в рамках научной школы МГТУ им. Н. Э. Баумана по экономике и организации производства.

Приведем список наших основных публикаций по солидарной информационной экономике (неформальной информационной экономике будущего) [41–56].

4.3. Организационно-экономические подходы к оценке реализуемости инновационно- инвестиционных проектов

4.3.1. Основные понятия

Начнем с краткого обсуждения терминов, используемых в настоящем разделе.

Под *проектом* (от *proiectus* (лат.) — брошенный вперед, выступающий, выдающийся вперед) обычно понимают замысел, идею, образ, оформленные в виде описания, обоснования, расчетов, чертежей, которые раскрывают сущность замысла и обосновывают возможность его практической реализации. Иногда под проектом понимают научно-технический документ, содержащий описание замысла, его обоснование, план реализации (осуществления).

Под *управлением проектами* обычно понимают область деятельности, в ходе которой определяются и достигаются цели проекта при балансировании между объемами работ, ресурсами (материальными, производственными, кадровыми, временными, финансовыми и др.), качеством и рисками. Ключевыми факторами успеха проектного управления считают наличие заранее определенного плана, минимизации рисков и отклонений от плана, эффективного управления изменениями. Можно сказать, что управление проектами – это частный вид теории и практики управления, имеющий целью осуществление проекта (замысла).

Управление проектами – развитая сфера научно-технической и организационно-экономической деятельности. В нашей стране наиболее серьезные результаты получены в рамках научной школы Института проблем управления РАН [1, 2, 3]. О практической важности этой сферы свидетельствует наличие государственных стандартов по управлению проектами, в несколько другой терминологии – по проектному менеджменту [4, 5, 6].

Пусть проект подготовлен. Удастся ли его осуществить? Какие условия для этого необходимы? Совокупность этих вопросов имеется в виду, когда говорят о «*реализуемости проекта*». Можно сказать, что главное при оценке реализуемости проекта (возможности его осуществления) – наличие объективных предпосылок для выполнения проекта. Такими предпосылками могут быть технологические возможности организации (предприятия, корпорации, отрасли, государства в целом), реализующей проект, экономические и социальные показатели проекта, при работе на рынок – рыночные предпосылки (наличие потенциального спроса, приемлемый уровень конкуренции).

Целесообразно выделить различные стороны реализуемости проекта – техническую (включая физическую, технологическую, производственную), экономическую, организационную, политическую (включая социальную и экологическую) ...

Техническая реализуемость проекта определяется, прежде всего, возможностью составления (сетевое) графика выполнения работ, необходимых для осуществления проекта, с учетом того, что эти работы взаимосвязаны в пространстве и времени, должны быть сбалансированы с имеющимися ресурсами (как по объектам потребления ресурсов, так и с интенсивностью их потребления по этапам реализации проекта). Отметим, что выполнение проекта должно быть обеспечено ресурсами не только в целом, но и в каждый момент времени. Физическая реализуемость проекта обеспечивается соответствием замысла проекта современным научным представлениям. Технологическая – наличием технологий, необходимых для реализации работ, составляющих проект. Производственная – реализованностью необходимых технологий на производствах, имеющих в распоряжении исполнителей проекта. В целом техническая реализуемость проекта обеспечивается наличием (при необходимости – созданием) производственной структуры, имеющей потенциальную технологическую возможность для выполнения планового комплекса работ требуемого качества и в установленные сроки при оговоренном уровне затрат.

Экономическая реализуемость проекта определяется возможностью обеспечения финансирования всех этапов его выполнения (за счет различных источников финансирования – государственного бюджета, средств инвесторов, собственных или заемных средств). При принятии решения о целесообразности реализации проекта проводят предварительную оценку экономической эффективности проекта по различным критериям.

Организационная реализуемость проекта обеспечивается эффективной структурой управления, наличием у руководителей, принимающих решения, необходимой компетентности и требуемого уровня полномочий для эффективной координации совместной деятельности всех участников выполнения проекта.

Политическая сторона реализуемости проекта (включая социальную и экологическую) определяется интересами государства и общества, которые могут как содействовать, так и противодействовать осуществлению проекта.

Выделяют инновационные и инвестиционные проекты.

Термин «инновация» происходит от лат. «*innovato*», что означает обновление или улучшение. В самом общем плане под инновацией можно понимать особую культурную ценность (материальную или нематериальную), которая в данное время и в данном месте воспринимается людьми как новая. Только в начале XX века стали изучаться закономерности технических нововведений.

Принято считать, что понятие «нововведение» является русским вариантом английского слова *innovation*. Буквальный перевод с английского означает «введение новаций» или в нашем понимании этого слова «введение новшеств». Под новшеством понимается новый порядок, новый обычай, новый метод, изобретение, новое явление. Русское словосочетание «нововведение» (в буквальном смысле «введение нового») означает процесс внедрения и использования новшества.

В мировой экономической литературе «инновация» интерпретируется как превращение потенциального научно-технического прогресса (НТП) в реальный, воплощающийся в новых продуктах и технологиях. Проблематика нововведений в нашей стране на протяжении многих лет разрабатывалась в рамках экономических исследований НТП.

В литературе насчитывается множество определений понятия «инновация». Будем понимать под инновацией любое внедренное нововведение или изменение [7]. Естественно выделять научно-технические и управленческие инновации. Первые основаны на новых научных и технических решениях, а вторые - на новых организационных решениях.

Согласно сказанному выше инновационный проект – это проект по разработке и внедрению новшеств. Управление инновациями (другими словами, инновационный менеджмент) – научная и практическая область, посвященная управлению инновационными проектами. Заметная часть этой области относится к инновационным проектам, замысел которых принадлежит одному лицу или группе специалистов. Такие проекты на первых этапах выполняются малыми предприятиями, поскольку на этих этапах не требуют значительных капиталовложений, хотя сопряжены со значительным риском прекращения разработки по тем или иным причинам. Для подобных инновационных проектов характерным является переход прав собственности на них от одних хозяйствующих субъектов к другим, в том числе путем Интернет-аукционов [8].

Инвестиционные проекты, как следует из названия, выделяются среди всех проектов необходимостью значительных капиталовложений (термин «инвестиции» в переводе на русский язык означает «капиталовложения»). Инвестирование – один из наиболее важных аспектов деятельности любой развивающейся организации. Причины, обуславливающие необходимость инвестиций, в целом можно подразделить на три вида: обновление имеющейся материально-технической базы; наращивание объемов производственной деятельности; освоение новых видов деятельности. Любой инвестиционный проект может быть охарактеризован с различных сторон: финансовой, технологической, организационной, временной, экологической, социальной и др. Каждая из них по-своему важна, однако финансовые аспекты инвестиционной деятельности во многих случаях имеют решающее значение.

Управление инвестициями (другими словами, инвестиционный менеджмент) ориентирован на управление капиталоемкими проектами, осуществление которых не связано со значительными рисками. Поэтому анализу, оценке и управлению рисками уделяется ограниченное внимание, причем в основном рискам, связанным с экономическими отношениями [9].

Большое значение имеют инновационно-инвестиционные проекты – инновационные проекты, реализация которых требует значительных инвестиций. Многие нововведения требуют значительных финансовых затрат, вложений капитала научно-исследовательские и опытно-конструкторские разработки, в подбор и обучение сотрудников, в новые здания, сооружения, станки, оборудование, запасы сырья и материалов, используемых в производстве. Следует финансировать научно-техническую деятельность, проведение исследований и проектирования изделий и технологических процессов, оплачивать работы сотрудников, на начальном этапе не приносящих прибыли, рекламную кампанию и др.

Проекты по созданию ракетно-космической техники (РКТ) обладают рядом специфических особенностей по сравнению с проектами в других отраслях промышленности. Во-первых, в них велика инновационная составляющая, обусловленная необходимостью решения вновь возникших научно-технических задач. Как следствие, большую роль играют инновационные риски. Во-вторых, проекты по созданию РКТ требуют для своей реализации значительного ресурсного обеспечения и времени.

Итак, основные проекты, выполняемые в ракетно-космической отрасли – это проекты, реализация которых требует значительного ресурсного обеспечения. Как правило, их следует считать инновационно-инвестиционными проектами.

При анализе и управлении инновационно-инвестиционными проектами в ракетно-космической отрасли необходимо учитывать как многообразие рисков, с которыми связана их реализация, так и необходимость значительных капиталовложений. Следовательно, методы оценки реализуемости инновационно-инвестиционных проектов в ракетно-космической отрасли должны быть построены на основе современных научных достижений, в частности, в области математических методов экономики, включая ста-

тистические и экспертные методы принятия решений в условиях неопределенности и риска [10].

4.3.2. Подходы к оценке реализуемости инновационно-инвестиционных проектов, в том числе по созданию РКТ

Общим подходам к оценке реализуемости инновационно-инвестиционных проектов посвящена работа [11], согласно которой целью оценки реализуемости научно-технического проекта является определение степени сбалансированности работ и мероприятий, предусмотренных проектом, с прогнозируемыми возможностями по их выполнению (в предположении выполнения имеющихся в конкретной ситуации ограничений). По мнению автора работы [11], реализуемость проекта должна рассматриваться, по крайней мере, в трех аспектах: научно-техническом, временном и ресурсном. *Научно-технический аспект* предусматривает оценку возможностей достижения заданных характеристик продукции (технических, технологических, эксплуатационных и др.), которая будет получена в результате реализации проекта, с учетом имеющегося научно-технического задела и возможностей опытно-экспериментальной базы. *Временной аспект* предполагает оценку возможности выполнения включенных в проект работ и мероприятий в установленные сроки. *Ресурсный аспект* касается обеспечения работ и мероприятий проекта всеми необходимыми видами ресурсов: финансовыми, материальными, трудовыми, производственными.

Инструментальные методы оценки реализуемости наукоемкого инновационно-инвестиционного проекта разработаны в статье [12]. В этой работе предложен методологический подход к разработке инструментальных методов, предназначенных для многофакторной оценки реализуемости отдельного наукоемкого проекта, направленного на создание новой высокотехнологичной продукции или услуг на имеющихся или создаваемых (реконструируемых) мощностях предприятий.

Из всех частных аспектов реализуемости инновационно-инвестиционных проектов наиболее часто обсуждается финансовый. Методологии оценки финансовой значимости и реализуемости инновационных проектов создания интеллектуальной продукции посвящена работа [13]. В ней рассмотрены методологические проблемы концептуального построения инструментальных методов и средств, позволяющих оценить финансово-экономический эффект и практическую реализуемость проектов создания инновационной интеллектуальной продукции. Финансовая устойчивость наукоемкого предприятия в качестве фактора оценки реализуемости инновационного проекта рассмотрена в статье [14], посвященной разработке и анализу качественно-количественных методов оценки финансовой устойчивости наукоемких предприятий, функционирующих в рыночной нестабильной среде, позволяющих снизить сопутствующие производству риски и повысить степень реализуемости инновационных проектов. Финансовая устойчивость – одна из составляющих организационно-экономической устойчивости. Основные результаты научной школы МГТУ им. Н. Э. Баумана в области организационно-экономической устойчивости отражены в монографиях [15, 16].

Концепция оценки и управления риском при реализации инновационных проектов создания интеллектуальной продукции разработана в статье [17]. Обсуждаются концептуальные основы оценки и управления уровнем риска при обосновании и реализации программ и планов создания и производства инновационной интеллектуальной

продукции. Приведены структура, состав и основное содержание постановок задач и методов, направленных на обеспечение компенсации риска.

Институциональный метод повышения реализуемости наукоемких инновационно-инвестиционных проектов предложен в статье [18], посвященной риск-менеджменту подобных проектов. В организационной структуре предприятия по производству наукоемкой и высокотехнологичной продукции предлагается создать подразделение управления риском, показываются управляющие и информационные связи между составляющими его структурными элементами, приводятся рекомендации по принципам их работы.

При анализе и управлении проектами в ракетно-космической отрасли необходимо учитывать риски их реализации в условиях значительных капиталовложений. Так, в работе [19] впервые разработана в общем виде аддитивно-мультипликативная модель оценки рисков. В двухуровневой схеме на нижнем уровне оценки рисков объединяются аддитивно, на верхнем уровне – мультипликативно. Аддитивно-мультипликативная модель применена для оценки рисков проектов создания РКТ. Выделено 44 частных риска на нижнем уровне и 8 – на верхнем, соответственно этапам выполнения проекта. Получены численные оценки рисков и вероятностей осуществления инновационно-инвестиционных проектов (на условных примерах).

При реализации реализуемости проектов по созданию РКТ необходимо проводить анализ, оценку и управление рисками, а также применять статистические и экспертные методы прогнозирования динамики развития технико-экономических показателей проектов. Следовательно, организационно-экономические подходы к оценке реализуемости проектов по созданию РКТ должны быть построены на основе современных научных достижений, в частности, в области математических методов экономики [10], включая статистические [20] и экспертные [21] методы принятия решений в условиях неопределенности и риска [22, 23].

4.3.3. Общая модель оценки реализуемости проекта

Нами разработана общая теоретическая модель оценки реализуемости инновационно-инвестиционного проекта (далее – проект), которая содержит методические подходы по оценке реализуемости. Алгоритмы расчетов реализуемости конкретных проектов разрабатываются с учетом специфики самих проектов и области их внедрения.

Принимаем, что процесс реализации проекта состоит из m этапов.

Для реализации проекта необходимы ресурсы k типов (т.е. заданы исходные данные, включающие информацию о трудовых, временных, материальных, производственных, финансовых и др. ресурсах, необходимых для реализации проекта). Цены на ресурсы являются сопоставимыми, т.е. приведены к фиксированному моменту времени (например, моменту проведения расчетов), что позволяет исключить влияние инфляции на принятие решений.

На i -ом этапе, $i = 1, 2, \dots, m$, для реализации проекта необходимы ресурсы в объемах, записанных в виде k -мерного вектора

$$X(i) = (x(i, 1), x(i, 2), \dots, x(i, k)).$$

Здесь $x(i, j)$ – объем ресурсов типа j , $j = 1, 2, \dots, k$, необходимых для выполнения i -го этапа проекта. Если ресурс типа j не используется для выполнения i -го этапа проек-

та, то $x(i, j) = 0$, если используется, то $x(i, j) > 0$. Вектор $X(i)$ размерности k описывает набор ресурсов, необходимых для выполнения i -го этапа проекта.

Объективно существуют ограничения на объемы ресурсов, имеющихся в распоряжении руководства хозяйствующего субъекта (предприятия, корпорации, отрасли), реализующей проект. Пусть $a(i, j)$ – максимальный объем ресурсов типа j , $j = 1, 2, \dots, k$, который может быть выделен для выполнения i -го этапа проекта. Вектор

$$A(i) = (a(i, 1), a(i, 2), \dots, a(i, k))$$

размерности k описывает набор ресурсов, имеющихся в распоряжении руководства хозяйствующего субъекта (далее – организации) и могущих быть использованными для выполнения i -го этапа проекта.

Проект может быть реализован тогда и только тогда, когда

$$x(i, j) \leq a(i, j)$$

при всех $i = 1, 2, \dots, m, j = 1, 2, \dots, k$, или в векторной форме

$$X(i) \leq A(i),$$

при всех $i = 1, 2, \dots, m$. Если хоть одно из указанных неравенств не будет выполнено, проект реализовать не удастся.

Если выяснено, что проект реализовать не удастся, то могут быть предприняты меры для исправления ситуации. Одно направление разработки таких мер – изменение плана выполнения проекта, приводящее к изменению объемов необходимых ресурсов $X(i)$, $i = 1, 2, \dots, m$, например, продление сроков выполнения работ, частичная замена натуральных испытаний компьютерным моделированием и т.п. Другое направление разработки мер для исправления ситуации – изменение наборов ресурсов $A(i)$, $i = 1, 2, \dots, m$, имеющихся в распоряжении руководства организации и могущих быть использованными для выполнения проекта, например, привлечение новых сотрудников, покупка нового оборудования, взятие кредита в банке. Третье направление – корректировка объемов работ проекта путем управления требованиями. Как указано в работе [24], в процессе реализации комплексных проектов по созданию перспективных образцов РКТ порой возникают существенные отклонения фактических сроков и бюджетов от плановых. Основная причина заключается в недостаточном внимании к этапу формирования, согласования и управления требованиями в процессе реализации проекта. Сделан вывод о том, что для повышения эффективности реализации проектов целесообразно использовать концепцию и инструментарий управления требованиями (requirements engineering).

Затем для скорректированного проекта необходимо проверить выполнение указанных выше неравенств. Если все неравенства выполнены, то проверка закончена, проект реализуем. Если хотя бы одно неравенство не выполнено, то необходимо принять решение либо о целесообразности следующей итерации, либо о признании невозможности реализации проекта и прекращении его разработки.

Если имеется несколько проектов, то возникает возможность управлять их совместным выполнением. Для этого следует поставить и решить оптимизационные задачи о выборе подмножества проектов для реализации, о планировании выполнения проектов, распределении ресурсов и т.п., в математическом плане достаточно близкие к задачам оптимальной организации производства [25]. При этом проект рассматривается как заказ, который надо выполнить.

Приведенный выше подход позволяет провести также анализ, оценку и управление рисками при выполнении проекта в ракетно-космической отрасли. Риски разбиваем

на две группы. Одни, назовем их рисками проекта, изменяют объемы необходимых ресурсов $X(i)$, $i = 1, 2, \dots, m$ (как правило, увеличивают, но иногда, при нахождении и внедрении новых научно-технических решений, могут и уменьшить). Другие, назовем их рисками предприятия РКП, приводят к изменению наборов ресурсов $A(i)$, $i = 1, 2, \dots, m$, имеющихся в распоряжении предприятия РКП и используемых для выполнения проекта.

Алгоритм проверки реализуемости проекта для общей теоретической модели представлен на рис.1.

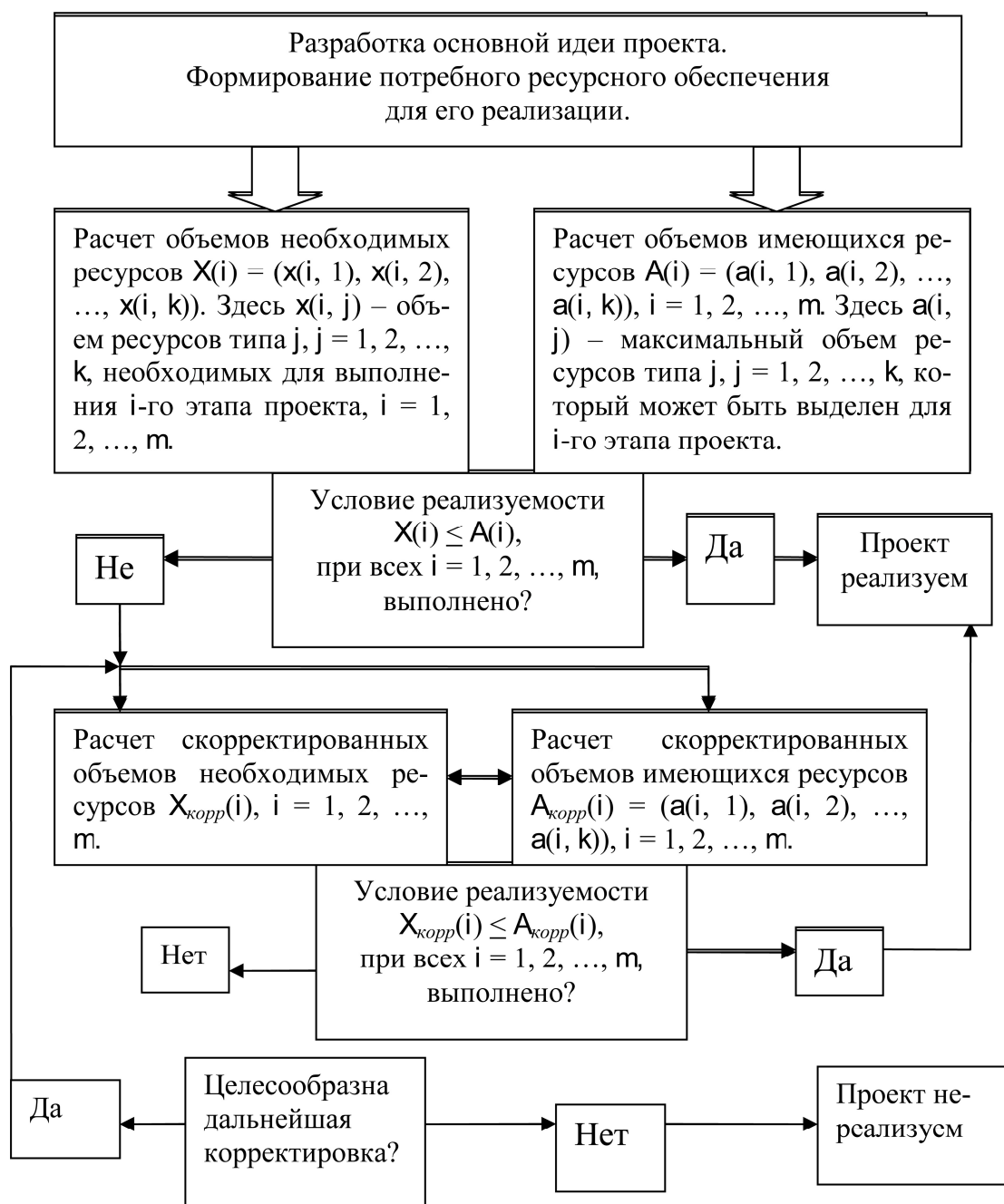


Рисунок 1 – Алгоритм проверки реализуемости проекта в целом.

Анализ и управление рисками проводится на основе математических моделей. В настоящее время наиболее широко используются модели на основе теории вероятностей и математической статистики.

При вероятностно-статистическом моделировании вектора $X(i)$ и $A(i)$, $i = 1, 2, \dots, m$, рассматриваются как случайные вектора, распределения которых оцениваются статистическими (на основе прошлой информации) и экспертными методами, а также на основе комбинированных моделей. В результате может быть получена вероятность успешной реализации проекта

$$P = P(X(i) \leq A(i), \text{ при всех } i = 1, 2, \dots, m),$$

т.е. вероятность одновременного выполнения всех неравенств, указанных в скобках.

Возможность успешной реализации проекта может быть найдена с помощью других методов описания неопределенности, в частности, теории нечеткости, интервальной математики и их синтеза – системной нечеткой интервальной математики [26, 27].

Для прогнозирования потребного ресурсного обеспечения и возможных рисков (возможностей реализации опасных событий, величин ущербов, в вероятностно-статистических моделях – вероятностей таких событий и математических ожиданий случайных ущербов) могут быть использованы современные методы прогнозирования (см. [22, 23] и др.).

4.3.4. Подходы к оценке реализуемости проектов по созданию РКТ

Для конкретизации общей модели оценки реализуемости выделим типовые этапы разработки проектов в ракетно-космической отрасли, которые будем применять для моделирования реализуемости проекта:

1. Разработка аванпроекта (технического предложения).
2. Разработка эскизного проекта и рабочей документации.
3. Изготовление опытных изделий комплекса. Наземная отработка.
4. Лётные испытания.
5. Подготовка документации на изделия серийного производства.
6. Серийное производство.
7. Прием в эксплуатацию и эксплуатация РКТ.

Деление на этапы может быть проведено и несколько иначе. Так, этап «Разработка эскизного проекта и рабочей документации» может быть разделен на два: «Эскизный проект» и «Разработка рабочей документации на опытные образцы и макеты». В этапе «Изготовление опытных изделий комплекса. Наземная отработка» могут быть выделены стадии изготовления макетов и опытных изделий комплекса, автономных испытания, комплексных испытаний, межведомственных испытаний и корректировки рабочей документации. На этапе «Серийное производство» можно выделить стадии подготовки и освоения серийного производства, изготовления, испытания изделий, корректировки документации на изделия серийного производства. На этапе «Прием в эксплуатацию и эксплуатация РКТ» выделяют ввод в эксплуатацию, собственно эксплуатацию и завершение жизненного цикла комплекса (его изделий) на стадии «Утилизация».

Для описания предлагаемых нами подходов к оценке реализуемости проектов нет необходимости в рамках настоящей статьи сосредотачивать внимание на выделении конкретных этапов создания РКТ. Отметим, что может быть также применено деление на этапы, использованное в [19].

Первый этап «Разработка аванпроекта (технического предложения)» принципиально отличается от последующих. До этого этапа должны быть проведены, как правило, научно-исследовательские (в том числе системные, проектно-поисковые) работы по

обоснованию целесообразности создания изделия (ракетного комплекса и т.д.), формированию его технического облика, правовой охране создаваемых результатов интеллектуальной деятельности и т.д.



Рисунок 2 – Алгоритм проверки реализуемости проекта по созданию РКТ (этап 1 «Разработка аванпроекта»).

Разработка аванпроекта (технического проекта), как правило, проводится на конкурсных началах. Организация – победитель конкурса назначается единственным исполнителем опытно-конструкторских работ на весь срок проекта.

Первый этап, в частности, принципиально отличается от последующих тем, что на этапе разработки концепции укрупненно оценивается потребное ресурсное обеспечение, которое отражается в техническом задании на проект.

На следующих этапах конкретизируются и реализуются решения, принятые на первом этапе. Поэтому для этапа 1 характерно целеполагание (рис.2), для этапов 2 - 7 – планирование с возможной корректировкой и учетом рисков (рис.3).

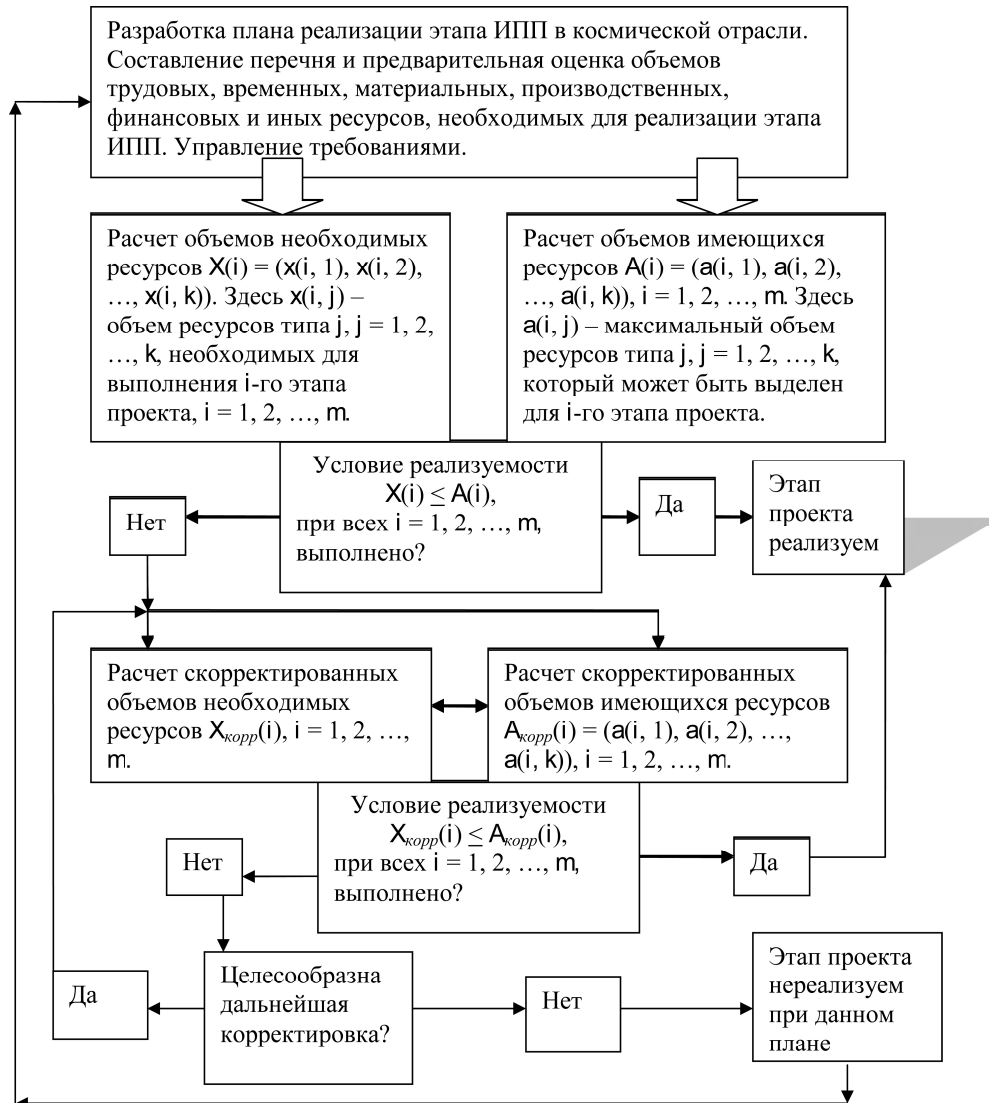


Рисунок 3 – Алгоритм проверки реализуемости проекта по созданию РКТ (этапы 2 – 7 реализации проекта).

На рис. 2 и рис. 3, в отличие от рис.1, не предусмотрено заключение «Проект не реализуем». Вместо этого для обеспечения стоящих перед проектом задач в необходимых случаях предусмотрен возврат к началу алгоритма и поиск приемлемого решения на основе управления требованиями.

4.3.5. Промежуточные итоги

Разработанные в настоящем подразделе организационно-экономические подходы к оценке реализуемости проектов по созданию РКТ представлены в виде алгоритмов на

рис. 1–3. Они учитывают специфику ракетно-космической отрасли, в силу которой подобные проекты имеют как инновационную, так и инвестиционную составляющие.

Решение о реализации таких проектов принимается на основе всей совокупности СТЭП-факторов (социальных, технологических, экологических, экономических, политических) [2, 7]. Для реализации проектов по созданию РКТ необходимо ресурсное обеспечение, а также соответствующая система организационного управления. Подчеркнем, что для успешного выполнения проекта недостаточно только наличия финансовых ресурсов.

Важная особенность рассматриваемых проектов – значительные риски [19] различной природы, что делает необходимым регулярное применение методов управления требованиями [24].

4.4. Управление требованиями и оценка реализуемости проектов создания изделий ракетно-космической техники

4.4.1. Системный подход к оценке реализуемости космического проекта

Под оценкой реализуемости космического проекта (КП) понимается установление расчетно-аналитическим и (или) экспертным методом степени соответствия потребностей проектных работ во всех видах ресурсов (интеллектуальных, производственных, трудовых, финансовых, материальных, временных и др.) с возможностями удовлетворения этих потребностей в ходе выполнения проекта. При оценке реализуемости и принятии проектных решений дополнительно необходимо учитывать влияние научно-технических рисков разработки и производства наукоемкой продукции со сложной структурой жизненного цикла.

Для оценки реализуемости КП необходимо провести укрупненно его декомпозицию на составляющие с целью дальнейшего анализа реализуемости подпроектов по разработке, изготовлению и эксплуатации этих составляющих КП.

Под проектом (согласно ГОСТ Р 54869-2011 «Проектный менеджмент. Требования к управлению проектом») понимается комплекс взаимосвязанных мероприятий, направленный на создание уникального продукта или услуги в условиях временных и ресурсных ограничений.

Под космическим проектом (КП) понимается комплекс взаимосвязанных мероприятий, направленных на создание изделий ракетно-космической техники (РКТ) (космических аппаратов (КА), ракетносителей (РН), разгонных блоков (РБ) и т.д.) в условиях временных и ресурсных ограничений.

В настоящем подразделе декомпозицию проводим укрупненно по составным элементам ракетного космического комплекса (РКК). Предлагаемая декомпозиция представлена на рис.1.

Выделим крупные сегменты РКК – космического аппарата (КА), ракетносителя (РН), наземной космической инфраструктуры (НКИ).

На следующем уровне декомпозиции выделяем элементы РКК (например, целевая аппаратура, конструкция и др.).

Далее - элементы подсистемы (система энергоснабжения, система управления и др.).

В элементах подсистемы выделяем на верхнем уровне классификационной иерархии оборудование и приборы (например, батареи, солнечные панели и др.). На нижнем уровне декомпозиции рассматриваем детали оборудования и приборов.

Общий порядок проведения оценки реализуемости космического проекта предусматривает:

- формирование варианта реализации проекта;
- расчет показателей затрат ресурсов, необходимых для выполнения проекта по всему жизненному циклу его работ и мероприятий;
- количественное определение ресурсных ограничений;
- расчет технико-экономических показателей работ и мероприятий проекта;
- расчет показателей реализуемости проекта;
- анализ показателей и коррекция (в случае необходимости) исходных данных для последующих циклов расчетов.

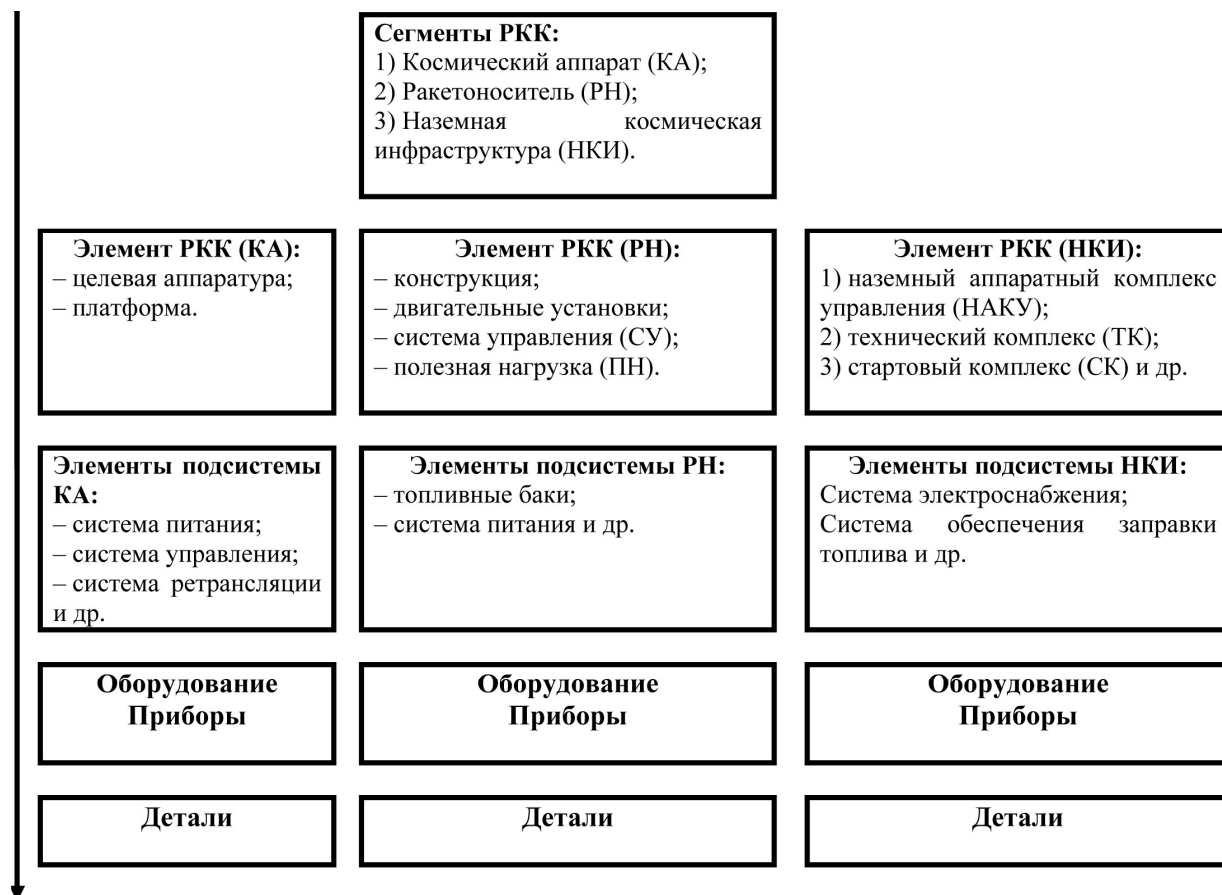


Рисунок 1 – Декомпозиция РКК

Оценка реализуемости КП является обязательным этапом процесса его формирования. Она осуществляется и по мере выполнения проекта, если в структуре или составе работ или мероприятий, а также в составе участников или структуре их основных фондов происходят серьезные для целей проекта изменения.

Таким образом, процедуру оценки реализуемости проекта в целом и отдельных его составных элементов необходимо производить практически на всех этапах формирования и выполнения проекта.

В соответствии с подходом, задача оценки реализуемости КП может быть подввергнута декомпозиции на отдельные задачи оценки реализуемости его составных элементов и потребует отдельного анализа. Например, если изготовление деталей может быть поручено сторонним организациям, в том числе – в отдельных случаях – зарубежным, то по мере продвижения к верхним уровням декомпозиции участие сторонних организаций становится недопустимым. Вместе с тем на основе фрактального подхода системного анализа (методологии самоподобия при моделировании) ясно, что все от-

дельные задачи реализуемости составляющих КП могут быть рассмотрены в рамках единого подхода, развиваемого далее.

4.4.2. Этапы разработки космического проекта

Для упрощения восприятия разрабатываемого подхода под космическим проектом будем понимать процесс создания КА. Тогда для оценки реализуемости создания КА необходимо проанализировать этапы его разработки.

Начальный этап – формирование цели КП (КА) (целеполагание). Разрабатываются и утверждаются назначение и требования к КП (КА). В соответствии с методологией стратегического планирования [1] будем говорить о целях и задачах КП (КА). При разработке концепции КП (КА) необходимо рассмотреть последовательность:

Цели – задачи – мероприятия – показатели реализации.

В результате получаем концепцию создания КА и требования к КА. Для этого предлагается использовать новое для нашей страны направление в теории и практике управления созданием сложных технических систем – управления требованиями (Requirements Engineering).

В процессе реализации КП (создания КА) на основе создания изделий РКТ возникают существенные отклонения фактических сроков разработки и выделенных бюджетных ассигнований от плановых. Основная причина заключается в недостаточном внимании к этапу формирования, согласования и управления требованиями в процессе реализации проекта. В работе [2] сделан вывод о том, что для повышения эффективности реализации проектов целесообразно использовать концепцию и инструментарий управления требованиями (requirements engineering (RE)).

Дословный перевод с английского термина «Requirements Engineering» – выработка требований к проектируемой системе; разработка технических условий и т.п. На практике в RE включают более широкий спектр функций:

- выявление и документирование требований;
- моделирование и анализ требований;
- проверка и согласование требований;
- управление изменениями к требованиям по этапам реализации проекта и т.д.

К началу следующего этапа имеем концепцию КП (КА) (например, выбраны орбита и полезный груз), на основе которой разработаны требования к КП (КА) до уровня элементов КА. Декомпозицию концепции создания КП (КА) целесообразно разделить на три компонента – КА, РН и НКИ.

Итогом этапа является разработка первого варианта бюджета КП (КА) на основе технико-экономического обоснования (ТЭО) затрат на его создание.

Далее проводится оценка концепции создания КП (КА), в частности, путем оценки соответствия затрат ресурсов прогнозируемым результатам. В результате оценки концепции создания КП (КА) может быть либо продолжено, либо возвращено на стадию разработки концепции или даже на начальную стадию формирования цели и решаемых задач КП (КА) (рис. 2).

В случае положительной оценки концепции определяется состав КП (КА), осуществляется проработка КП (КА) до уровня элементов подсистемы и состава НКИ. Укрупненно оценивается бюджет создания КП (КА).

Проводится оценка системы, в результате которой разработка КП (КА) может быть либо продолжена, либо возвращена на стадию разработки концепции или даже на начальную стадию формирования цели и задач создания КП (КА) (рис. 2).

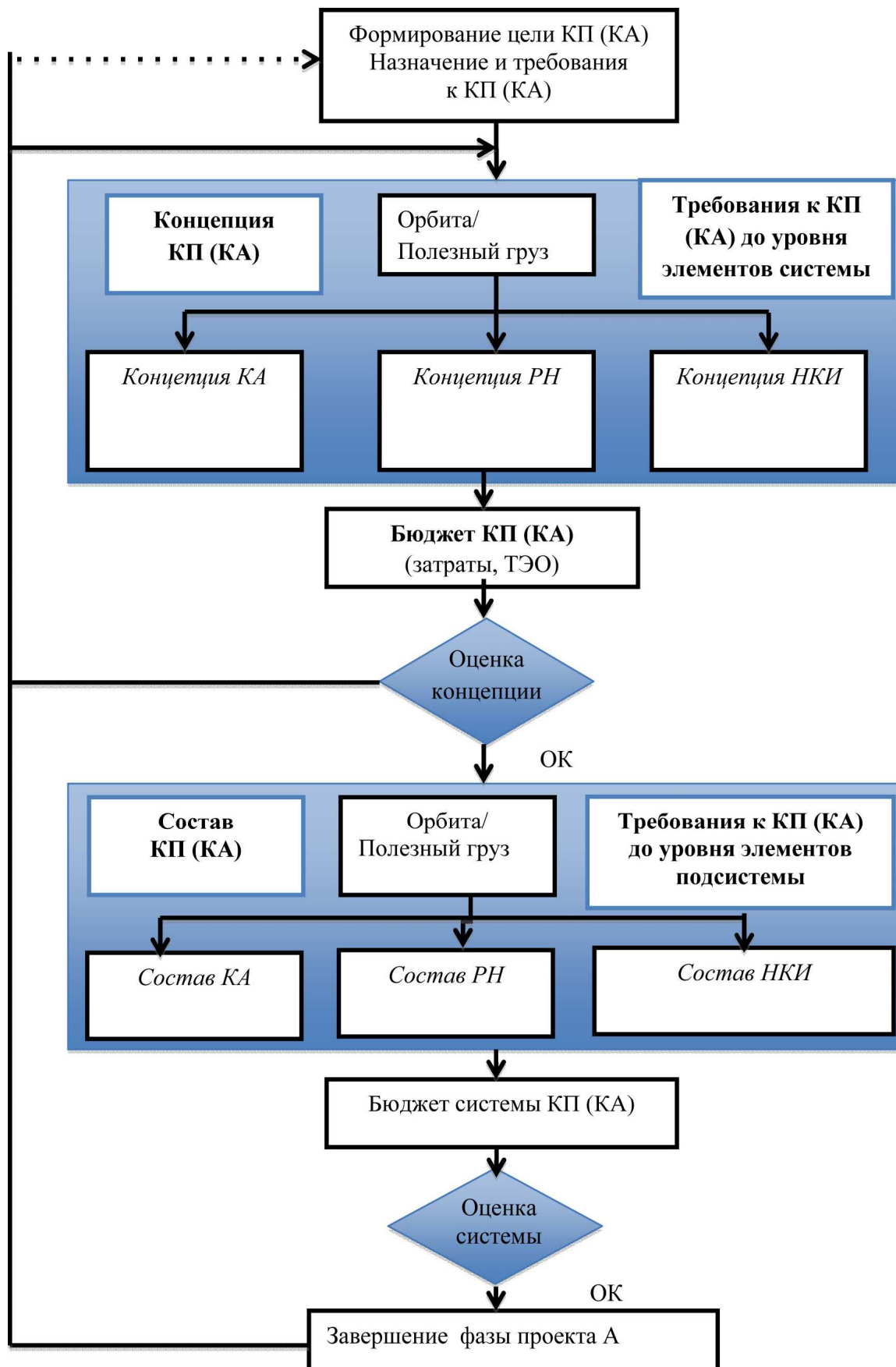


Рисунок 2 – Блок-схема работ по подготовке КП (КА) (первая фаза А)

Методические рекомендации по оценке реализуемости инновационно-инвестиционных проектов создания изделий ракетно-космической техники представим в виде алгоритма.

4.4.3. Алгоритм управления космическим проектом

Алгоритм представим в виде последовательности шагов 1 – 13.

Алгоритм управления требованиями к КП до этапа изготовления макетов и опытных изделий приведен на рис. 3 (часть 1) и рис. 4 (часть 2).

Шаг 1. Выполнение КП начинается с формирования целей и задач КП, а вслед за этим – формирование требований к КА.

Шаг 2. После формирования цели, задач и требований к КП предлагается оценить бюджет ресурсов (финансовых, материальных, производственных, кадровых, временных), необходимых для реализации КП. В частности, оценить капитальные и эксплуатационные затраты КП, себестоимость доставки полезного груза на орбиту. Применяются различные методы прогнозирования бюджета, проводится оценка и сопоставление себестоимости доставки груза конкурентами. Необходимо отметить [6] сложность сопоставления экономических величин, выраженных в денежной форме, из-за:

1) заметной инфляции всех денежных единиц, т.е. падения их покупательной способности, причем с разной скоростью в зависимости от вида денежной единицы;

2) отличия (в разы) официальных курсов валют и курсов по паритету покупательной способности (ППС).

Шаг 3. Проверяем, устраивает ли бюджет КП заказчика? Если да, переходим к Шагу 4. Если нет, возвращаемся к шагу 1 (к формированию требований к КП).

Шаг 4 - оценка рисков и реализуемости КП (в целом). Применяются методы оценки рисков концепции [7] и методы оценки реализуемости [8 - 9]. Если оценка рисков и/или реализуемости отрицательна – переходим к шагу 1 (к формированию требований к КП). Если оценка рисков и реализуемости положительна – переходим к следующему уровню декомпозиции – к системам КП.

Шаг 5 – оценка реализуемости системы. Поскольку цель и задачи системы определены на более высоком уровне декомпозиции (при разработке концепции КП), то начинаем с управления требованиями, касающимися рассматриваемой системы. Уточняем оценку бюджета системы, используя методы прогнозирования системы и проводя сравнения с возможностями конкурентов.

Шаг 6. Проверяем, укладывается ли система в бюджет реализации миссии. Если да, укладывается, переходим к оценке рисков и реализуемости системы. Если нет, то возвращаемся к шагу 5 - управлению требованиями, касающимися рассматриваемой системы.

Шаг 7 - оценка рисков и реализуемости системы. Применяются методы оценки рисков системы [7] и методы оценки реализуемости системы [8 – 9]. Если оценка рисков системы и/или реализуемости системы отрицательна, то возвращаемся к шагу 5 – управлению требованиями, касающимися рассматриваемой системы. Если оценка рисков системы и реализуемости системы положительна – переходим к следующему уровню декомпозиции – к подсистемам КП.

Шаг 8 – оценка реализуемости подсистемы. Поскольку цель и задачи подсистемы определены на более высоком уровне декомпозиции (при разработке системы КП), то

начинаем с управления требованиями, касающимися рассматриваемой подсистемы, а затем переходим к оценке бюджета подсистемы. Используем различные методы оценки бюджета подсистемы, целевых издержек (Target Costing) подсистемы [10], сравнение с конкурентами.

Шаг 9. Проверяем, укладывается ли подсистема в бюджет системы. Если нет – переходим к функционально-стоимостному анализу (ФСА) подсистемы (см., например, [11]), а затем - к оценке бюджета подсистемы. После повторной оценки бюджета подсистемы проверяем, увеличился ли суммарный бюджет системы после проведения ФСА. Если увеличился, то снова переходим к ФСА подсистемы, а затем - к оценке бюджета подсистемы. Если нет, то переходим к следующему шагу.

Шаг 10 - оценка рисков и реализуемости подсистемы. Применяются методы оценки рисков подсистемы и методы оценки реализуемости подсистемы. Если оценка рисков подсистемы и/или реализуемости подсистемы отрицательна, то возвращаемся к шагу 8 - управлению требованиями, касающимися рассматриваемой подсистемы. Если оценка рисков подсистемы и реализуемости подсистемы положительна – переходим к следующему уровню декомпозиции – к компонентам КА.

Шаг 11 – оценка реализуемости компонента. Поскольку цель и миссия компонента определены на более высоком уровне декомпозиции (при разработке подсистемы КА), то начинаем с управления требованиями, касающимися рассматриваемого компонента, а затем переходим к оценке бюджета компонента. Используем различные методы оценки бюджета компонента, Target Costing компонента, сравнение с конкурентами.

Шаг 12. Проверяем, укладывается ли компонент в бюджет подсистемы. Если нет – переходим к функционально-стоимостному анализу компонента, а затем - к оценке бюджета компонента. После повторной оценки бюджета компонента проверяем, увеличился ли суммарный бюджет подсистемы после проведения ФСА. Если увеличился, то снова переходим к ФСА компонента, а затем - к оценке бюджета компонента. Если нет, то переходим к следующему шагу.

Шаг 13 - оценка рисков и реализуемости компонента. Применяются методы оценки рисков компонента и методы оценки реализуемости компонента. Если оценка рисков компонента и/или реализуемости компонента отрицательна, то возвращаемся к шагу 11 - управлению требованиями, касающимися рассматриваемого компонента. Если оценка рисков компонента и реализуемости компонента положительна – переходим к изготовлению макетов и опытных изделий.

Аналогично предыдущему алгоритм может быть доведен до нижнего уровня декомпозиции – до уровня деталей. Этого не делаем, поскольку осуществление работ на этом уровне хорошо изучено и относится к стандартным задачам организации производства.

Итак, базовая формулировка методических рекомендаций по оценке реализуемости инновационно-инвестиционных проектов на этапе планирования в ракетно-космической отрасли представлена в виде алгоритма на рис. 3 и 4 и описана выше в виде шагов 1–13.

Отметим, что использование термина «инновационно-инвестиционные проекты» применительно к проектам по созданию изделий ракетно-космической техники подробно обосновано в статьях [8–9]. В них велика инновационная составляющая, обусловленная необходимостью решения вновь возникших научно-технических задач. Как следствие, большую роль играют инновационные риски.

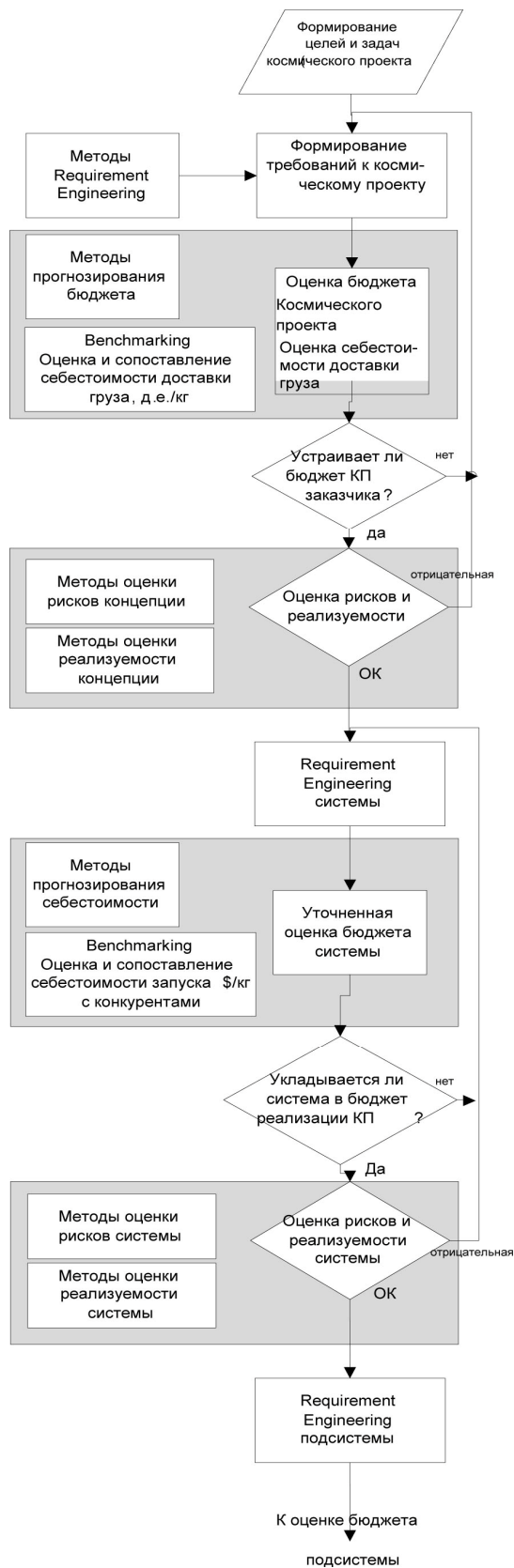


Рисунок 3 – Алгоритм управления КП до этапа изготовления макетов и опытных изделий (часть 1)

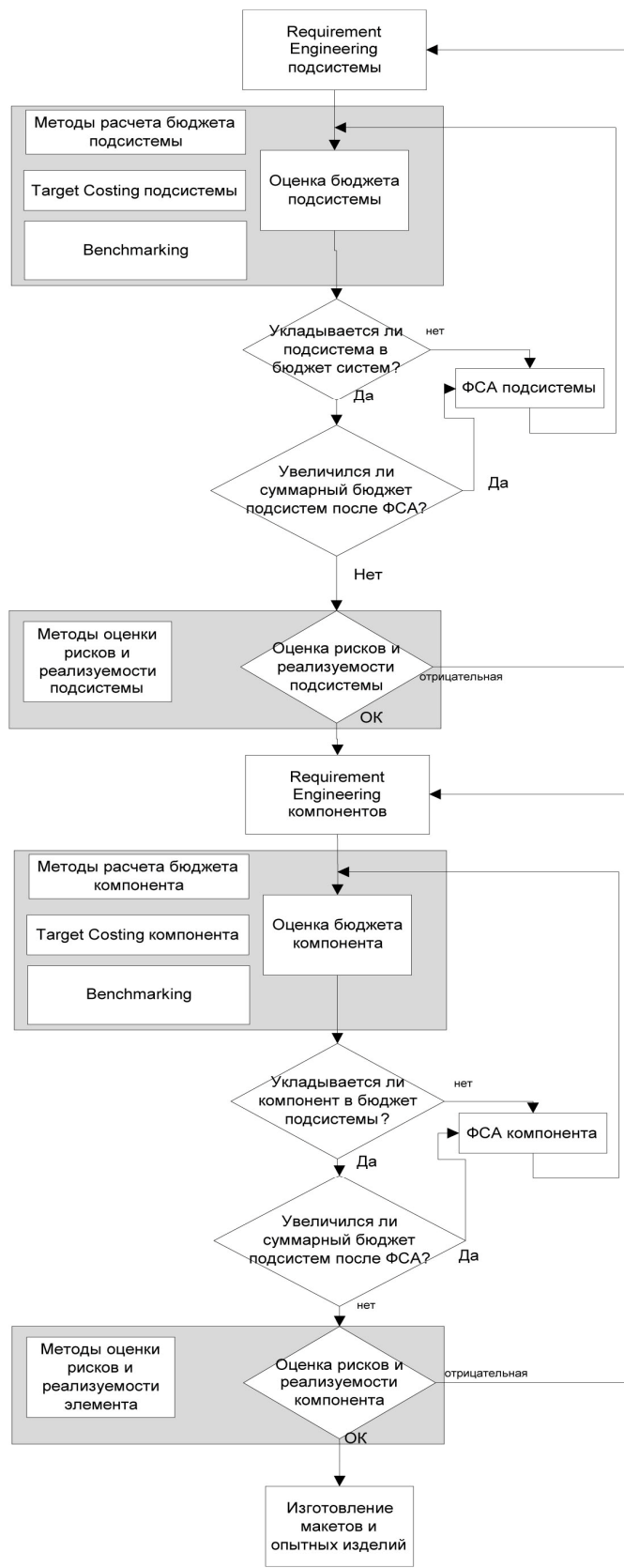


Рисунок 4 – Алгоритм управления КП до этапа изготовления макетов и опытных изделий (часть 2)

При этом проекты по созданию РКТ требуют для своей реализации значительного ресурсного обеспечения и времени, являются инвестиционными.

4.5. Аддитивно-мультипликативная модель оценки рисков при создании ракетно-космической техники

4.5.1. Об основных понятиях общей теории риска

Предложено большое число различных определений основных понятий теории риска (см. обзоры [1–3]). Не вдаваясь в их обсуждение, выделим два базовых основных понятия – «рисковое событие» и «ущерб от осуществления рискованного события», необходимые для дальнейшего изложения.

Под «рисковым событием» будем понимать возможное нежелательное событие, приводящее к отрицательным последствиям. Оно связано с определенным видом риска, может произойти или не произойти. «Ущерб от осуществления рискованного события» может выражаться как в натуральных показателях (число погибших, получивших непоправимый ущерб здоровью и т.п.), так и в финансовых (оценка потерь в денежных единицах). Используют также понятие «ущерб от риска», который определен всегда, он равен 0, если «рисковое событие» не состоялось, положителен в противном случае. Тогда «ущерб от осуществления рискованного события» - это «ущерб от риска» при условии, что «рисковое событие» состоялось.

Выделяют три этапа исследования риска – анализ, оценку и управление [4]. Анализ проводится средствами предметной области. Для оценки и управления может использоваться различный математический аппарат, основанный на одной из трех базовых теорий - теории вероятностей и математической статистике, теории нечетких множеств, интервальной математике и статистике (о второй и третьей из них см., например, [5]).

Наиболее простой и наиболее распространенный подход к исследованию риска основан на теории вероятностей и математической статистике в ситуации, когда ущерб от осуществлении рискованного события выражен в денежных единицах. Тогда «рисковое событие» моделируется случайным событием (в смысле теории вероятностей) с некоторой вероятностью p , а «ущерб от осуществления рискованного события» – случайной величиной X (в смысле теории вероятностей).

Простейшая единая характеристика риска – это средний ущерб $pM(X)$, т.е. математическое ожидание ущерба от риска, подробнее, произведение вероятности p рискованного события на математическое ожидание $M(X)$ ущерба от осуществления рискованного события. Под управлением риском часто понимают выбор управляющих воздействий с целью минимизации среднего ущерба.

Известен ряд более сложных постановок, в том числе многокритериальных, в которых, например, математическое ожидание $M(X)$ заменяют на медиану или, скажем, квантиль порядка 0,999999; минимизируют два критерия – математическое ожидание и дисперсию (естественно, в двухкритериальной задаче один из критериев переводят в ограничение или применяют иной подход, обеспечивающий получение решения), и т.п. [1, 4].

Из сказанного ясна необходимость разработки методов оценки вероятности p рискованного события. Оценка этой вероятности может быть как шагом к оценке среднего ущерба $pM(X)$, так и представлять самостоятельную ценность для разработки практических рекомендаций по управлению риском. По сравнению с оценкой вероятности p рискованного события оценка математического ожидания $M(X)$ ущерба от осуществления

рискового события требует дополнительных исследований, зачастую достаточно трудоемких [6].

Для оценки вероятности p рисков события при создании ракетно-космической техники предлагаем использовать аддитивно-мультипликативную модель оценки рисков, рассматриваемую ниже.

4.5.2. Аддитивно-мультипликативная модель оценки рисков (общий случай)

Для оценки и управления рисками можно разрабатывать математические модели различной степени общности и сложности. Целесообразно выделить класс моделей, достаточно общих для применений в различных предметных областях, но при этом достаточно простых и приспособленных для практических применений и расчетов. По нашему мнению, рассмотренная ниже аддитивно-мультипликативная модель оценки рисков, а именно, оценки вероятности p рисков события, относится к этому классу.

Аддитивно-мультипликативная модель оценки рисков основана на двухуровневой иерархической схеме декомпозиции риска (рис.1). При этом на нижнем уровне агрегированные оценки групповых рисков строятся аддитивно (поскольку вероятности конкретных видов нежелательных событий – частные риски нижнего уровня - малы), а на верхнем уровне итоговая оценка риска рассчитывается по групповым рискам по мультипликативной схеме.

В общем случае аддитивно-мультипликативная модель оценки риска исходит из следующих предпосылок.

1. Цель разработки модели – оценка риска R наступления нежелательного события. Для расчета этого риска применяем вероятностную модель, согласно которой наступление нежелательного события является случайным событием – подмножеством множества всех возможных элементарных событий. Риск (нежелательное событие) будем обозначать R , его числовую вероятностную оценку Q . Пусть Q – вероятность наступления нежелательного события R , тогда $P = 1 - Q$ есть вероятность того, что нежелательного события удастся избежать. Для простоты описания пусть Q – вероятность неудачи, тогда $P = 1 - Q$ есть вероятность успеха, например, вероятность успешного выполнения инновационно-инвестиционного проекта по созданию изделия ракетно-космической техники (или его определенного этапа). В дальнейшем изложении используется двойственность Q и P (с прикладной точки зрения важна оценка риска Q , в то время как модель описывается с помощью вероятностей P).

2. Примем, что для успеха (осуществления случайного события B) необходимо одновременное выполнение m независимых условий (должны одновременно осуществиться случайные события B_1, B_2, \dots, B_m). Предполагаем, что **случайные события B_1, B_2, \dots, B_m независимы в совокупности** (в терминах теории вероятностей [7]). Тогда вероятность успеха, т.е. вероятность P осуществления случайного события B , равна произведению вероятностей P_1, P_2, \dots, P_m осуществления случайных событий B_1, B_2, \dots, B_m , т.е. $P = P_1 P_2 \dots P_m$. Следовательно, оценка Q риска R , т.е. вероятность наступления нежелательного события, равна $Q = 1 - P = 1 - P_1 P_2 \dots P_m$. Оценка суммарного риска Q всегда больше оценки частного риска $Q_i = 1 - P_i$, поскольку итоговая вероятность P всегда меньше частной вероятности успеха P_i .

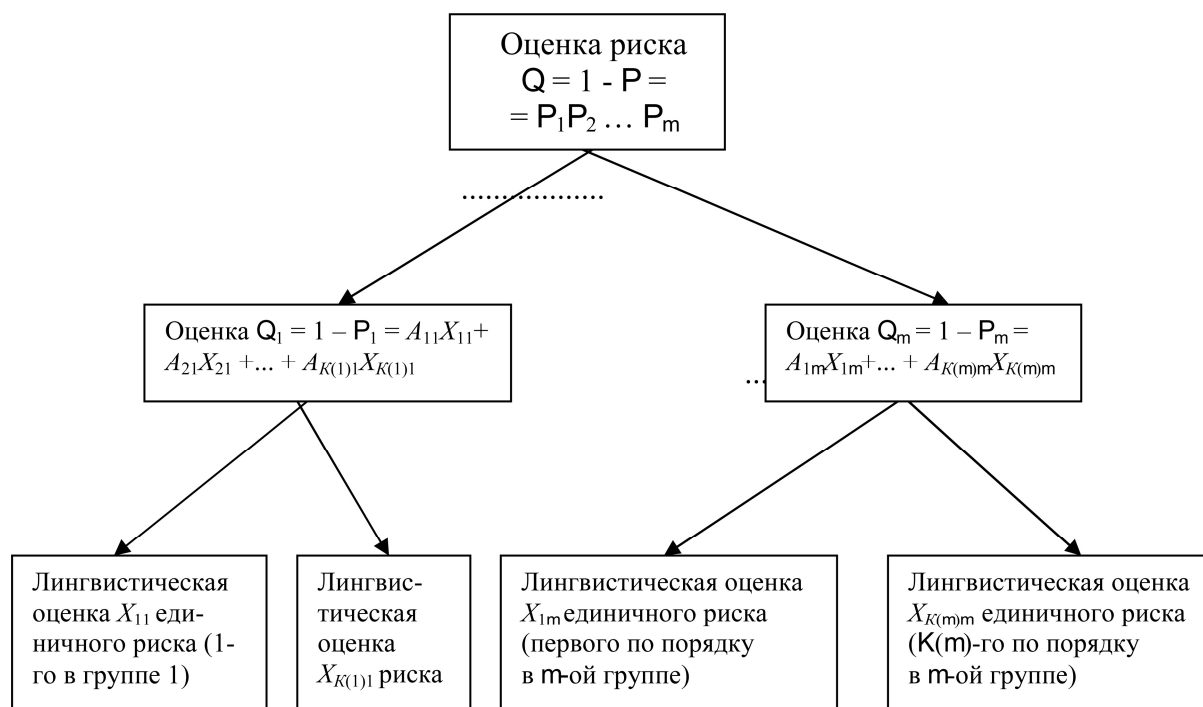


Рисунок 1 – Декомпозиция рисков с помощью дерева событий.

3. Принимаем, что для осуществления i -го условия должны одновременно осуществиться случайные события $B_{i1}, B_{i2}, \dots, B_{ik(i)}$, имеющие вероятности $P_{i1}, P_{i2}, \dots, P_{ik(i)}$ соответственно. Здесь $k(i)$ – число событий второго (нижнего) уровня декомпозиции (см. иерархическую схему на рис.1), соответствующих i -му событию на первом (верхнем) уровне декомпозиции. Оценки частных рисков второго порядка R_i равны $Q_{ij} = 1 - P_{ij}, j = 1, 2, \dots, k(i)$. При моделировании предполагаем, что оценки частных рисков Q_{ij} малы, а частные вероятности успеха P_{ij} достаточно близка к 1.

Как выразить вероятность события B_i первого уровня через вероятности событий $B_{i1}, B_{i2}, \dots, B_{ik(i)}$ второго уровня? Рассмотрим два варианта:

(А) события $B_{i1}, B_{i2}, \dots, B_{ik(i)}$ второго уровня независимы в совокупности (и дополнительные к ним, соответствующие реализациям частных рисков, также независимы);

(Б) нежелательные события (т.е. соответствующие частным рискам) несовместны.

В случае (А) независимости:

$$P_i = P_{i1}P_{i2} \dots P_{ik(i)} = (1 - Q_{i1})(1 - Q_{i2}) \dots (1 - Q_{ik(i)}). \quad (1)$$

В случае (Б) несовместности (принимаем, что риски реализуются редко, поэтому возможностями одновременного осуществления двух или нескольких нежелательных событий можно пренебречь):

$$P_i = 1 - Q_i = 1 - Q_{i1} - Q_{i2} - \dots - Q_{ik(i)}. \quad (2)$$

Формула (2) означает, что оценка Q_i частного риска R_i есть сумма оценок Q_{ij} частных рисков второго порядка R_{ij} , т.е. $Q_i = Q_{i1} + Q_{i2} + \dots + Q_{ik(i)}$. Поскольку оценки Q_{ij} частные риски второго порядка R_{ij} малы, то, раскрывая скобки в правой части формулы (1), получаем, что с точностью до бесконечно малых второго порядка $(1 - Q_{i1})(1 - Q_{i2}) \dots (1 - Q_{ik(i)}) = 1 - Q_{i1} - Q_{i2} - \dots - Q_{ik(i)}$. Таким образом, два принципиально разных подхода (А) и (Б) дают одно и то же численное значение (с точностью до бесконечно малых более высокого порядка), что повышает обоснованность использования формул (1) и (2).

4. Каждый из частных рисков (факторов риска) второго порядка R_{ij} имеет два показателя – *выраженность* (показывает частоту встречаемости) и *весомость* (насколько

влияет на риск более высокого уровня). Эти показатели можно оценивать на основе различных моделей.

Рассмотрим оценку выраженности (другими словами, распространенности, возможности, вероятности). Если есть возможность – ее целесообразно проводить по статистическим данным (как частоту реализации нежелательного события). Можно использовать экспертные оценки. При этом естественно давать оценки рисков с помощью лингвистических переменных. Например, члены экспертной комиссии оценивают риск R_{ij} с помощью градаций лингвистической переменной X_{ij} , выбирая ее значения из списка:

- 0 - практически невозможное событие (с вероятностью не более 0,01),
- 1 - крайне маловероятное событие (с вероятностью от 0,01 до 0,05),
- 2 - маловероятное событие (вероятность от 0,05 до 0,10),
- 3 - событие с вероятностью, которой нельзя пренебречь (от 0,10 до 0,20),
- 4 - достаточно вероятное событие (вероятность от 0,20 до 0,30),
- 5 - событие с заметной вероятностью (более 0,30).

Этот список может меняться в соответствии с конкретной задачей оценки и управления риском. В частности, могут быть изменены: количество градаций; способ **оцифровки** градаций (например, вместо ряда 0 - 1 - 2 - 3 - 4 - 5 может использоваться ряд 0 - 0,2 - 0,4 - 0,6 - 0,8 - 1); граничные значения для вероятностей (например, если нежелательные события являются редкими, но соответствующий им ущерб велик, то вероятность практически невозможного события должна быть не более 10^{-5} , вместо «не более 0,01», как выше, и т.п.). Естественно принять, что значения X_{ij} , используемые для оцифровки градаций, неотрицательны.

Кратко скажем о других возможных моделях. Лингвистические переменные естественно моделировать с помощью **теории нечеткости** (см., например, [1, 5]). Тогда X_{ij} – нечеткие числа. Можно использовать «треугольные» нечеткие числа, у которых функция принадлежности описывается тремя числовыми параметрами a, b, c ($a < b < c$) и имеет треугольный вид – функция принадлежности равна 0 левее a и правее c , в точке b равна 1, на интервалах (a, b) и (b, c) линейна. Арифметические операции над такими числами описываются проще, чем для функций принадлежности общего вида. Для реализации аддитивно-мультипликативной модели необходимо тем или иным способом определить арифметические операции над нечеткими числами. Другое обобщение – моделирование лингвистических переменных с помощью **интервальных чисел** (см., например, [1, 5]). Тогда X_{ij} – интервал (a, b) (или $[a, b]$, $(a, b]$, $[a, b)$), т.е. описывается двумя числовыми параметрами a и b . В примере, описанном выше, «крайне маловероятное событие» описывается интервалом $(0,01; 0,05]$.

Сбор и анализ экспертных оценок должны быть описаны в соответствующей методике в соответствии с общими положениями монографии [8]. В частности, согласно теории измерений итоговую оценку целесообразно рассчитывать как медиану индивидуальных оценок (при четном числе членов экспертной комиссии - как правую медиану).

5. В оценке Q_{ij} риска R_{ij} можно учесть весомость (другими словами, важность, значимость, тяжесть) этого вида риска:

$$Q_{ij} = A_{ij}X_{ij}, \quad (3)$$

где A_{ij} – показатель весомости (важности), например, оценка экономических потерь, вызванных данным видом риска, X_{ij} – показатель выраженности (распространенности). **Эта формула обобщает известный способ оценки риска как произведения среднего ущерба (математического ожидания ущерба) на вероятность нежелательного события.**

5. В соответствии с формулами (2) и (3) имеем

$$\begin{aligned}
 P_i &= 1 - Q_i = 1 - Q_{i1} - Q_{i2} - \dots - Q_{ik(i)} = \\
 &= 1 - A_{i1}X_{i1} - A_{i2}X_{i2} - \dots - A_{ik(i)}X_{ik(i)}, \quad i = 1, 2, \dots, m, \quad (4)
 \end{aligned}$$

где $X_{i1}, X_{i2}, \dots, X_{ik(i)}$ – оценки факторов риска второго порядка, используемые при вычислении оценки частного риска типа i , положительные числа $A_{i1}, A_{i2}, \dots, A_{ik(i)}$ – коэффициенты весомости (важности) этих факторов.

Значения факторов $X_{i1}, X_{i2}, \dots, X_{ik(i)}$ оценивают эксперты для каждого конкретного инновационного проекта, в то время как значения коэффициентов весомости $A_{i1}, A_{i2}, \dots, A_{ik(i)}$ задаются одними и теми же для всех проектов – по результатам специально организованного экспертного опроса.

6. Вероятность P_i должна быть неотрицательна при всех возможных значениях $X_{i1}, X_{i2}, \dots, X_{ik(i)}$. Если все оценки факторов риска (частных рисков) принимают свои максимальные значения, то и риск R_i должен принять свое максимальное значение, равное 1. Следовательно, коэффициенты весомости (важности) должны удовлетворять условию $A_{i1} \max X_{i1} + A_{i2} \max X_{i2} + \dots + A_{ik(i)} \max X_{ik(i)} = 1$. В рассмотренном выше варианте оцифровки максимальные значения X_{ij} равны 5. Следовательно, сумма $A_{i1}, A_{i2}, \dots, A_{ik(i)}$ должна равняться $1/5 = 0,2$.

Все составляющие аддитивно-мультипликативной модели (АМ-модели) описаны. Обсудим некоторые дополнительные вопросы. АМ-модель позволяет рассчитывать не только риск реализации проекта в целом, но и вероятности частичной его реализации. Например, если проект предусматривает выполнение четырех этапов: НИР, ОКР, экспертизы, создание (и испытание) опытного образца, то представляют интерес вероятности успешного выполнения (1) НИР, (2) НИР и ОКР, (3) НИР, ОКР и экспертизы, т.е. первого этапа; первого и второго этапа; трех начальных этапов. В случае сетевого графика может быть полезен расчет вероятности успешного прохождения отдельных ветвей. АМ-модель позволяет описывать риски реализации инновационно-инвестиционного проекта, состоящего из нескольких этапов. Простейший вариант – верхний уровень иерархической схемы рис.1 соответствует этапам (m – число этапов). Более сложный вариант АМ-модели – рис.1 соответствует одному этапу, риски выполнения этапов независимы между собой, т.е. добавляется еще один уровень иерархии, на котором агрегирование вероятностей происходит по мультипликативной схеме. Если какой-либо из рисков первого или второго порядка оказывается недопустимо велик (больше заданного порога), то естественно признать риск проекта в целом недопустимо большим и прекратить дальнейшие оценки и расчеты. Порог задают эксперты. Риск проекта может быть выражен в вербальной форме, т.е. численное значение может быть переведено в значение лингвистической переменной (например, по схеме п.4 выше). АМ-модель позволяет ввести коррективы в выполнение проекта. Если тот или иной риск недопустимо велик, проект не может быть успешно реализован с высокой вероятностью, то можно выявить критические факторы риска и разработать управленческие решения, позволяющие добиться успешной реализации проекта. Например, повторить этап, с теми же или иными параметрами. Определить, как надо изменить значения критических факторов риска, чтобы добиться реализации проекта с вероятностью, не меньшей заданной, а затем оценить объемы ресурсов, необходимые для изменения значений выявленных критических факторов риска. Таким образом, АМ-модель может быть развита в различных направлениях.

Аддитивно-мультипликативная модель оценки рисков может применяться для решения различных прикладных задач. Так, в [1] на с.359-370 она использована для моделирования рисков выполнения инновационных проектов в вузах (с участием внешнего партнера), в этой модели $m = 4$. В [9] – для оценки рисков выпуска нового инновационного изделия, снова $m = 4$. В [10] – для оценки рисков проектов создания

ракетно-космической техники, в этой модели $m = 8$. Рассмотрим подробнее три перечисленных примера применения аддитивно-мультипликативной модели оценки рисков.

4.5.3. Аддитивно-мультипликативная модель оценки рисков выполнения инновационных проектов в вузах (с участием внешнего партнера)

Обычно под инновационным проектом в вузе понимают проект, который опирается на ранее проведенные научно-технические разработки, приведшие к перспективным для практического использования результатам. Поскольку вузы, как правило, не занимаются сами производственной и коммерческой деятельностью, то предполагается, что коммерческая реализация будет осуществляться внешним партнером (или партнерами).

Структура и выраженность рисков реализации инновационных проектов в вузах несколько отличаются от таковых для инновационных проектов вообще и тем более от рисков разнообразных инвестиционных проектов. На первое место выходят риски невыполнения работы в соответствии с техническим заданием и невозврата (полного или частичного) средств. Риски могут быть, в частности, связаны с различными трудностями (см. приведенную выше схему).

Возможные итоги выполнения инновационной работы можно описать следующим образом:

а) работа и финансовые обязательства всех партнеров выполнены в полном объеме;

б) научно-исследовательская часть работы выполнена полностью, но по каким-либо причинам внешний партнер свои обязательства, в том числе финансовые, выполнил не в полном объеме;

в) научно-исследовательская часть работы выполнена полностью, но коммерческая часть проекта сорвана (внешним партнером), финансовые обязательства не выполнены;

г) научно-исследовательская часть работы не выполнена полностью, но получены существенные научные результаты; для окончания работы требуется некоторое время;

д) научно-исследовательская часть работы не выполнена, но получены некоторые интересные научные результаты; однако планируемый вначале научный результат не будет достигнут в обозримое время;

е) выполнение в вузе инновационной работы сорвано полностью.

Также при любом из вышеперечисленных исходов существует вероятность осуществления макроэкономического риска, которое может еще более ухудшить результат выполнения инновационного процесса.

Таким образом, только в двух случаях из шести оценка однозначна: итог а) - это полный успех, а итог е) - это полный провал. В остальных случаях - итоги б), в), г), д) - получены некоторые научные результаты, а в случае итога б) - также и некоторые коммерческие результаты. При этом в случае итогов а), б), в) научно-исследовательский коллектив выполнил все, что от него требовалось, хотя «полный успех» имеет место только в одном из этих трех случаев - в зависимости от результатов работы внешнего партнера.

Для оценки рисков реализации инновационных проектов в вузах применим описанную выше аддитивно-мультипликативную модель. Будем исходить из двухступенчатой схемы: сначала работает научно-исследовательский коллектив, затем он передает свои разработки внешнему партнеру, и тот начинает коммерческий этап. Считаем, что научно-исследовательский коллектив и внешний партнер работают независимо друг от

друга (в теоретико-вероятностном смысле). Вероятность того, что научно-исследовательский коллектив полностью выполнит свою работу, зависит от двух групп факторов, определяемых ситуациями соответственно внутри коллектива исполнителей и внутри вуза. Будем считать, что эти группы факторов также независимы между собой. Четвертый фактор риска - макроэкономический, т.е. ситуация в народном хозяйстве (степень выраженности неплатежей, инфляции, нерациональной налоговой политики и т.д.). Таким образом, выделяются четыре основные группы факторов риска, связанные с коллективом исполнителей, с вузом, с внешним партнером, с общей экономической обстановкой соответственно. Принимаем, что все четыре фактора независимы между собой (в модели - в теоретико-вероятностном смысле). В соответствии со сказанным выше основная формула аддитивно-мультипликативной модели оценки рисков реализации инновационных моделей в вузах и соответствующих вероятностей имеет вид: $P = P_1 P_2 P_3 P_4$, где P - вероятность «полного успеха», т.е. итога а) согласно приведенной выше классификации, при этом риск того, что инновационный проект не будет осуществлен полностью, оценивается вероятностью «отсутствия полного успеха», т.е. величиной $(1 - P)$; P_1 - вероятность того, что ситуация внутри коллектива исполнителей не помешает выполнению инновационного проекта (следовательно, риск коллектива оценивается величиной $1 - P_1$); P_2 - вероятность того, что ситуация внутри вуза не помешает выполнению инновационного проекта ($1 - P_2$ - риск вуза); P_3 - вероятность того, что внешний партнер полностью выполнит свою работу, после того, как научно-исследовательский коллектив полностью выполнит свою часть работы ($1 - P_3$ - риск партнера); P_4 - вероятность того, что ситуация в народном хозяйстве не помешает выполнению инновационного проекта (здесь $1 - P_4$ - макроэкономический риск, т.е. риск ситуации в стране). Структуризации вероятностей P_1-P_4 посвящены соответствующие подразделы ниже. Небольшое изменение обозначений по сравнению с предыдущим подразделом 4.5.2 (один нижний индекс вместо двух) не мешает пониманию.

Риск коллектива. Введем следующие переменные:

X_1 - на выполнении инновационного проекта скажется недооценка сложности научно-технической задачи (включая возможный выбор принципиально неверного направления работ),

X_2 - на выполнении работы скажется нехватка времени (из-за неправильного планирования процесса выполнения инновационного проекта, в то время как основное направление работ выбрано правильно),

X_3 - на выполнении работы скажутся возникшие в ходе ее выполнения проблемы, связанные с научным руководителем темы, в частности, с его длительным отсутствием или сменой (из-за длительной командировки, болезни, смерти, ухода на пенсию, перехода на другую работу и т.д.),

X_4 - на выполнении работы скажутся возникшие в ходе ее выполнения проблемы, связанные с иными непосредственными участниками работы (кроме руководителя).

Заметим, что в двух последних позициях (факторы X_3 и X_4) причинами невыполнения работы могут быть и недостаточная квалификация руководителя работы либо иных членов научно-исследовательского коллектива.

Экспертный опрос дал следующие значения весовых коэффициентов: $A_1 = 0,02$, $A_2 = 0,08$, $A_3 = 0,07$, $A_4 = 0,03$.

Пример 1. Если итоговая оценка экспертов такова: $X_1 = 3$; $X_2 = 2$; $X_3 = 4$; $X_4 = 1$, то $P_1 = 1 - A_1 X_1 - A_2 X_2 - A_3 X_3 - A_4 X_4 = 1 - 0,02 \times 3 - 0,08 \times 2 - 0,07 \times 4 - 0,03 \times 1 = 1 - 0,06 - 0,16 - 0,28 - 0,03 = 1 - 0,53 = 0,47$.

Таким образом, в данном конкретном случае эксперты достаточно скептически относятся к возможности выполнения работы в срок, причем основная причина скепси-

са - в возможном отъезде научного руководителя (риск оценивается как 0,28), вторая заметная причина - возможный недостаток времени (риск оценивается как 0,16).

Риск вуза. Для оценивания P_2 введем следующие переменные:

Y_1 - на возможности выполнения инновационного проекта скажутся организационные изменения в вузе, предпринятые руководством вуза,

Y_2 - на возможности выполнения инновационного проекта скажутся внутривузовские экономические проблемы (например, работы будут на какое-то время приостановлены из-за решения руководства вуза (несостоятельном с правовой точки зрения) о направлении средств, выделенных на финансирование инновационного проекта, на оплату труда преподавателей),

Y_3 - на возможности выполнения инновационного проекта скажется отсутствие в вузе соответствующей материальной базы (оборудования, материалов, вычислительной техники, площадей и т.д.).

Предварительный экспертный опрос дал следующие значения коэффициентов: $A_1 = 0,10$; $A_2 = 0,08$; $A_3 = 0,02$.

Пример 2. Если итоговые (групповые) оценки экспертов таковы: $Y_1 = 1$; $Y_2 = 4$; $Y_3 = 0$, то $P_2 = 1 - A_1 Y_1 - A_2 Y_2 - A_3 Y_3 = 1 - 0,10 \times 1 - 0,08 \times 4 - 0,02 \times 0 = 1 - 0,01 - 0,32 - 0 = 0,67$.

По мнению экспертов, для данного проекта и вуза наибольшее отрицательное влияние могут оказать внутривузовские экономические проблемы (вклад в общий риск оценен как 0,32).

Риск партнера. Для оценивания риска P_3 , связанного с деятельностью внешнего партнера, введем следующие переменные:

Z_1 - на возможности выполнения инновационного проекта скажутся финансовые проблемы внешнего партнера, связанные с недостатками в работе его сотрудников,

Z_2 - на выполнение проекта повлияют финансовые проблемы внешнего партнера, связанные с деятельностью конкретных государственных органов и частных фирм (например, неплатежи, административные решения),

Z_3 - работу над проектом сорвет изменение поведения возможных потребителей, например, из-за изменения моды или из-за решений соответствующих вышестоящих органов (министерств (ведомств) или регионального руководства), связанных, в частности, с выдачей лицензий, закрытием информации или с таким выбором технической политики, который делает ненужным (для большинства возможных потребителей) результатов инновационного проекта,

Z_4 - на возможности выполнения инновационного проекта отрицательно скажутся организационные преобразования у внешнего партнера, в частности, смена руководства.

Пилотный экспертный опрос дал следующие значения весовых коэффициентов: $A_1 = 0,03$, $A_2 = 0,06$, $A_3 = 0,06$, $A_4 = 0,05$.

Пример 3. Если итоговые (групповые) оценки экспертов таковы: $Z_1 = 3$; $Z_2 = 5$; $Z_3 = 1$; $Z_4 = 4$, то $P_3 = 1 - A_1 Z_1 - A_2 Z_2 - A_3 Z_3 - A_4 Z_4 = 1 - 0,03 \times 3 - 0,06 \times 5 - 0,06 \times 1 - 0,05 \times 4 = 1 - 0,09 - 0,30 - 0,06 - 0,20 = 1 - 0,65 = 0,35$.

Таким образом, эксперты достаточно скептически относятся к возможности успешного выполнения внешним партнером своих обязательств по договору, связанному с коммерческой реализацией разработок, выполненных по инновационному проекту. Основные «подводные камни», по их мнению, это действия конкретных государственных органов (вклад в общий риск оценен как 0,30), и нежелательные организационные преобразования (кадровые изменения) у внешнего партнера (вклад в риск равен 0,20).

Макроэкономический риск. Под макроэкономическим риском понимаем риск, определяемый внешними по отношению к системе «вуз - внешний партнер» фактора-

ми, прежде всего теми, которые являются общими для всего народного хозяйства. Для оценивания P_4 введем переменные:

W_1 - на возможности выполнения инновационного проекта скажется отсутствие или сокращение номинального финансирования (неплатежи со стороны бюджета),

W_2 - на возможности выполнения инновационного проекта скажется резкое сокращение реального финансирования (в сопоставимых ценах) из-за инфляции,

W_3 - на возможности выполнения инновационного проекта скажется изменение статуса и/или задач вуза или его внешнего партнера (в частности, из-за ликвидации или реорганизации вуза) по решению вышестоящих органов (министерства (ведомства) или регионального руководства),

W_4 - на возможности выполнения инновационного проекта скажутся относящиеся к инновационному проекту решения соответствующих вышестоящих органов (министерств (ведомств) или регионального руководства), связанные, например, с закрытием информации или с таким выбором технической политики, который делает ненужным или нецелесообразным выполнение инновационного проекта.

Пилотный экспертный опрос дал следующие значения весовых коэффициентов: $A_1 = 0,10$, $A_2 = 0,05$, $A_3 = 0,03$, $A_4 = 0,02$.

Пример 4. Если итоговые (групповые) оценки экспертов таковы: $W_1 = 3$; $W_2 = 4$; $W_3 = 1$; $W_4 = 2$, то $P_4 = 1 - A_1 W_1 - A_2 W_2 - A_3 W_3 - A_4 W_4 = 1 - 0,10 \times 3 - 0,05 \times 4 - 0,03 \times 1 - 0,02 \times 2 = 1 - 0,30 - 0,20 - 0,03 - 0,04 = 1 - 0,57 = 0,43$.

Таким образом, эксперты считают, что общая экономическая ситуация в стране может негативно сказаться на возможности выполнения рассмотренного ими инновационного проекта. Причем наиболее опасаются неплатежей со стороны государства (отсутствия или сокращения перечисления средств для выполнения проекта) и в несколько меньшей мере - уменьшения реального финансирования из-за инфляции (что, возможно, отвлечет членов научно-исследовательского коллектива на побочные заработки).

Итоговые оценки. Сведем вместе полученные результаты. Вероятность успешного выполнения инновационного проекта оценивается по формуле:

$$P = P_1 P_2 P_3 P_4,$$

где

$$P_1 = 1 - 0,02 X_1 - 0,08 X_2 - 0,07 X_3 - 0,03 X_4,$$

$$P_2 = 1 - 0,10 Y_1 - 0,08 Y_2 - 0,02 Y_3,$$

$$P_3 = 1 - 0,03 Z_1 - 0,06 Z_2 - 0,06 Z_3 - 0,05 Z_4,$$

$$P_4 = 1 - 0,10 W_1 - 0,05 W_2 - 0,03 W_3 - 0,02 W_4.$$

Для данных, приведенных в примерах 1–4, вероятность того, что научно-исследовательский коллектив в вузе полностью выполнит свою работу, равна: $P_1 P_2 = 0,47 \times 0,67 = 0,3149$, а вероятность успешного осуществления проекта в целом значительно меньше: $P = P_1 P_2 P_3 P_4 = 0,47 \times 0,67 \times 0,35 \times 0,43 = 0,0473924$. Таким образом, имеется лишь примерно 1 шанс из 20, что рассматриваемый инновационный проект будет успешно завершен (в намеченные сроки и с запланированным экономическим эффектом).

В табл.1 (см. ниже) приведены результаты расчета вероятностей, связанных с реализацией четырех типовых инновационных проектов. Видно, какое влияние оказывает изменение того или иного фактора на общую величину вероятности выполнения проекта. Выполнение первого проекта практически в одинаковой степени зависит от всех четырех факторов. Низкая вероятность выполнения второго проекта связана с относительно высокими показателями всех четырех видов риска. Вероятность выполнения третьего проекта – наименьшая, что связано с высоким риском внутри коллектива исполнителей и внутри вуза. У четвертого проекта наибольший риск связан с политиче-

ской и экономической обстановкой в стране. Вероятность выполнения пятого проекта относительно невысокая, но она выше, чем у второго, третьего и четвертого проектов.

Таблица 1 – Оценки вероятности реализации инновационных проектов в вузе

	Проект 1	Проект 2	Проект 3	Проект 4	Проект 5
1. Риск для коллектива исполнителей					
A_n	X_{n1}	X_{n2}	X_{n3}	X_{n4}	X_{n5}
0,02	0	2	4	2	1
0,08	0	3	5	2	2
0,07	1	2	4	2	2
0,03	1	2	2	3	0
$P_1 =$	0,9	0,52	0,18	0,57	0,68
2. Риск внутри вуза					
A_n	Y_{n1}	Y_{n2}	Y_{n3}	Y_{n4}	Y_{n5}
0,1	0	3	4	1	1
0,08	1	2	5	1	2
0,02	1	3	4	0	2
$P_2 =$	0,92	0,48	0,12	0,82	0,70
3. Риск партнера					
A_n	Z_{n1}	Z_{n2}	Z_{n3}	Z_{n4}	Z_{n5}
0,03	0	2	3	1	2
0,06	1	2	2	1	0
0,06	1	3	2	1	1
0,05	0	1	1	1	1
$P_3 =$	0,880	0,590	0,620	0,800	0,830
4. Макроэкономический риск					
A_n	W_{n1}	W_{n2}	W_{n3}	W_{n4}	W_{n5}
0,1	0	3	2	5	2
0,05	1	2	2	4	2
0,03	1	1	1	5	1
0,02	0	2	0	5	1
$P_4 =$	0,92	0,53	0,67	0,05	0,65
Вероятность выполнения данного проекта					
$P =$	0,670	0,078	0,009	0,019	0,26
Вероятность выполнения работ без учета риска партнера					
$P_1 P_2 P_4$	0,76	0,13	0,01	0,02	0,3
Вероятность выполнения работ без учета риска страны					
$P_1 P_2 P_3$	0,73	0,15	0,01	0,37	0,4
Вероятность выполнения работ без учета риска вуза					
$P_1 P_3 P_4$	0,73	0,16	0,07	0,02	0,37
Вероятность выполнения работ в вузе					
$P_1 P_2$	0,83	0,16	0,07	0,02	0,37

Выбор инновационных проектов для финансирования целесообразно проводить на основе описанной выше процедуры вероятностно-статистической (с учетом мнений экспертов) оценки их рисков реализации, разработанной нами при выполнении НИР «Разработка методологии оценки рисков реализации инновационных проектов высшей школы» [11]. Работа проводилась по заданию Отделения инновационных проектов и программ РИНКЦЭ Миннауки (1996г.). Основные результаты опубликованы в статье [12] и включены в [1]. Они могут быть модифицированы в соответствии со спецификой конкретного вуза.

Позже аналогичная процедура была разработана для оценки рисков при выпуске инновационного изделия. Группы рисков были, естественно, иными.

4.5.4. Оценки рисков при выпуске нового инновационного изделия

Выделим согласно [13] производственные, коммерческие, финансовые и глобальные риски. Принимаем, что все четыре группы независимы между собой (в модели - в теоретико-вероятностном смысле). Будем использовать описанную выше аддитивно-мультипликативную модель.

Для оценки P_1 - вероятности того, что производственные риски не окажут влияние на выполнение проекта, введем следующие частные факторы риска: X_{11} - недооценка сложности производства, и, как следствие, высокая доля брака, X_{21} - принципиальные ошибки при проектировании, из-за которых не удастся наладить серийный выпуск продукции, X_{31} - риски аварий на производстве, X_{41} - риски, связанные с возможным отсутствием (болезнь, увольнение) специалистов, без которых невозможно наладить производство, а также проблемы, возникшие в ходе выполнения работы, связанные с иными непосредственными участниками работы.

Для оценивания P_2 - вероятности того, что коммерческие риски не окажут влияние на выполнение проекта, введем частные факторы риска (кратко - риски): X_{12} - риски, связанные с деятельностью поставщиков (сроки, качество поставки и т.д.), X_{22} - риски, связанные с потребителями (товар не привлекателен (плохой маркетинг), высокая цена и т.д.), X_{32} - риски связанные с деятельностью конкурентов (выпуск конкурентами аналогичных товаров, сговор и т.д.), X_{42} - риски связанные с деятельностью органов государственной и муниципальной власти.

Для оценивания P_3 - вероятности того, что финансовые риски не окажут влияние на выполнение проекта, введем частные факторы риска: X_{13} - риски, связанные с изменением законодательства, X_{23} - риски изменения курсов валют, курсов акций, X_{33} - риски, связанные с ростом цен (инфляцией).

Для оценки глобальных рисков введем: X_{14} - государственные и международные риски (политические, военные, терроризм), X_{24} - природные риски (наводнения, землетрясения и т.д.).

В табл.2 приведены примеры - экспертные оценки весовых коэффициентов A_{in} и частных факторов риска для пяти проектов: X_{in} - для проекта 1 и соответственно Y_{in} - для проекта 2, V_{in} - для проекта 3, W_{in} - для проекта 4, Z_{in} - для проекта 5. При принятии решения о выборе инновационного изделия для выпуска изделия целесообразно использовать табл.2, в которой приведены оценки вероятностей ненаступления неблагоприятных событий для сравниваемых изделий (за время освоения). Вероятность «полного успеха» оценивается как $P = P_1P_2P_3P_4$. Для окончательного принятия решения о выборе проекта для реализации необходимо составить матрицу вероятностей благоприятного исхода и возможной прибыли, сопутствующей производству каждого изделия [9].

4.5.5. Аддитивно-мультипликативная модель оценки рисков при разработке ракетно-космической техники (РКТ)

Для демонстрации предлагаемого подхода примем на основе опроса экспертов, что разработка РКТ состоит из следующих восьми этапов:

1. Концепция.
2. Разработка технического проекта (аванпроекта и эскизного проекта).
3. Разработка рабочей (конструкторской и технологической) документации.
4. Изготовление макета и опытных изделий (опытного образца).

5. Наземная отработка (испытания).
6. Корректировка документации.
7. Летные испытания и доработка документации для производства.
8. Запуск.

Таблица 2 – Оценки вероятности реализации инновационных проектов

Коэффициенты весомости A_{in} и вероятности	Проект 1 X_{in}	Проект 2 Y_{in}	Проект 3 V_{in}	Проект 4 W_{in}	Проект 5 Z_{in}
1. Производственные риски					
0,08	1	2	0	2	1
0,07	0	1	0	1	1
0,02	0	0	0	0	0
0,03	1	0	0	0	1
$P_1 =$	0,89	0,77	1	0,77	0,82
2. Коммерческие риски					
0,05	0	1	1	1	1
0,07	1	2	5	1	2
0,02	0	1	1	1	0
0,06	1	1	1	1	1
$P_2 =$	0,87	0,73	0,52	0,80	0,75
3. Финансовые риски					
0,06	0	0	0	0	0
0,07	1	1	1	1	1
0,07	0	0	0	0	0
$P_3 =$	0,93	0,93	0,93	0,93	0,93
4. Глобальные риски					
0,11	1	1	1	1	1
0,09	0	0	0	0	0
$P_4 =$	0,89	0,89	0,89	0,89	0,89
Вероятность «полного успеха» изделия на рынке					
$P =$	0,64	0,47	0,43	0,51	0,51

На каждом этапе имеются те или иные риски. Будем оценивать риски и рассчитывать вероятности P_1, P_2, \dots, P_8 успешного выполнения перечисленных выше этапов, используя их сокращенные (условные) названия. Примем, что риски на разных этапах порождены независимыми случайными причинами, поскольку они возникают в непересекающиеся интервалы времени. Тогда вероятность P успешного выполнения разработки равна произведению вероятностей успешного выполнения этапов (мультипликативная составляющая модели): $P = P_1 P_2 \dots P_8$. Следовательно, оценка Q суммарного риска R , т.е. вероятность невыполнения разработки в заданный срок, равна $Q = 1 - P = 1 - P_1 P_2 \dots P_8$.

Будем применять аддитивно-мультипликативную модель с $m = 8$. На каждом этапе выделим частные риски R_{ij} второго порядка, всего 44 вида рисков (табл.3).

Оценки рисков по этапам и суммарного риска будем давать с помощью описанной выше в п.2 аддитивно-мультипликативной модели, используя лингвистические оценки X_{ij} рисков R_{ij} с шестью градациями, которым придаются числовые значения 0, 1, 2, 3, 4, 5. Соответствие между значениями лингвистических переменных и интервалами вероятностей приведено выше в подразделе 4.5.2.

Для оценки рисков по этапам в соответствии с формулой (4) необходимо использовать коэффициенты весомости (важности) $A_{i1}, A_{i2}, \dots, A_{ik(i)}$ факторов риска второго порядка. Эти коэффициенты были получены по результатам специально организованного экспертного опроса. По каждому из восьми этапов экспертам предлагалось распределить 100 баллов по видам рисков в соответствии с их влиянием на риск по этапу.

Таблица 3 – Риски по этапам и экспертные оценки весовых коэффициентов

№ п/п	Виды частных рисков по этапам	Оценки экспертов
	1. Концепция	
1	R_{11} – риск неверного технического решения (глобального, т.е. для изделия в целом)	10
2	R_{21} – риск персонала (риск срыва работ из-за организационных проблем в НИИ)	10
3	R_{31} – внешний риск (срыв работ по причинам, лежащим вне НИИ)	80
	2. Разработка технического проекта	
4	R_{12} – риск неверных технических решений (локальных, т.е. для отдельных блоков)	5
5	R_{22} – риск недостаточно высокого качества подготовки проекта	20
6	R_{32} – риск недостатка ресурсов (временных, кадровых, материальных, финансовых)	10
7	R_{42} – организационный риск (из-за организационных проблем внутри предприятия)	5
8	R_{52} – внешний риск	60
	3. Разработка рабочей документации	
9	R_{13} – риск ошибок при реализации технических решений	5
10	R_{23} – риск недостаточно высокого качества подготовки документации	10
11	R_{33} – риск недостатка ресурсов (кадровых, компьютерных, временных и др.)	35
12	R_{43} – риски, связанные с невыполнением обязательств смежниками и подрядчиками	25
13	R_{53} – организационный риск (риск срыва работ из-за плохой их организации)	10
14	R_{63} – внешний риск (кроме причин, указанных при описании рисков R_{43})	15
	4. Изготовление опытного образца	
15	R_{14} – риск ошибок при изготовлении деталей и блоков	5
16	R_{24} – риск ошибок при сборке	10
17	R_{34} – риск недостатка ресурсов (станочного парка, кадровых, компьютерных, временных и др. ресурсов)	35
18	R_{44} – риски, связанные с невыполнением обязательств смежниками и подрядчиками	25
19	R_{54} – организационный риск (риск срыва работ из-за плохой их организации)	5
20	R_{64} – риск, вызванный действиями поставщиков сырья, комплектующих, материалов (низкое качество, нарушение сроков)	15
21	R_{74} – внешний риск (по другим причинам)	5
	5. Наземная отработка (испытания)	
22	R_{15} – риск ошибок из-за нарушения работы испытательного оборудования	10
23	R_{25} – риск ошибок испытателей	10
24	R_{35} – риск недостатка ресурсов	60
25	R_{45} – организационный риск (риск срыва работ из-за плохой их организации)	5
26	R_{55} – внешний риск	15
	6. Корректировка документации	
27	R_{16} – риск ошибок при корректировке документации	5
28	R_{26} – риск недостаточно высокого качества подготовки документации	15
29	R_{36} – риск недостатка ресурсов (кадровых, компьютерных, временных и др.)	50
30	R_{46} – организационный риск (риск срыва работ из-за плохой их организации)	10
31	R_{56} – внешний риск	20
	7. Летные испытания и доработка документации	
32	R_{17} – риск ошибок из-за нарушения работы испытательного оборудования	5
33	R_{27} – риск ошибок испытателей (человеческий фактор)	10
34	R_{37} – риск недостатка ресурсов	25
35	R_{47} – организационный риск (риск срыва работ из-за плохой их организации)	10
36	R_{57} – «риск машины»	10
37	R_{67} – «риск среды»	10
38	R_{77} – внешний риск (по другим причинам)	30
	8. Запуск	
39	R_{18} – риск дефектности изделия;	5
40	R_{28} – риск дефектности оборудования (на стартовом столе и в ЦУПе);	10
41	R_{38} – группа факторов риска «Человек»	30
42	R_{48} – группа факторов риска «Машина»	10
43	R_{58} – группа факторов риска «Среда»	15
44	R_{68} – внешний риск (по другим причинам)	30

Ответы экспертов оказались достаточно хорошо согласованными, и в табл.3 в столбце «Оценки экспертов» приведены итоговые значения - медианы ответов экспертов.

В соответствии с п.2 сумма коэффициентов весомости (важности) $A_{i1}, A_{i2}, \dots, A_{ik(i)}$ должна равняться 0,2. Поэтому эти коэффициенты были получены путем деления соответствующих итоговых экспертных оценок на 500. Например оценка 5 переходит в 0,01, оценка 15 – в 0,03, и т.д. Численные значения коэффициентов весомости приведены в табл.2 в столбце «Веса A_{ij} ».

Другой экспертный опрос дал возможность получить лингвистические оценки рисков X_{ij} для двух проектов (табл.4).

Таблица 4 – Поэтапные и суммарные оценки рисков при разработке РКТ

№ п/п	Риски R_{ij}	Веса A_{ij}	Проект 1		Проект 2	
			Оценки рисков X_{ij}	Вклад $A_{ij}X_{ij}$	Оценки рисков X_{ij}	Вклад $A_{ij}X_{ij}$
1. Концепция (i = 1)						
1	R_{11}	0,02	3	0,06	2	0,04
2	R_{21}	0,02	1	0,02	1	0,02
3	R_{31}	0,16	1	0,16	0	0
Оценка Q_1 риска невыполнения в срок этапа 1				0,24		0,06
Вероятность успешного выполнения этапа P_1				0,76		0,94
2. Разработка технического проекта (i = 2)						
4	R_{12}	0,01	2	0,02	3	0,03
5	R_{22}	0,04	0	0	1	0,04
6	R_{32}	0,02	1	0,02	0	0
7	R_{42}	0,01	0	0	0	0
8	R_{52}	0,12	1	0,12	0	0
Оценка Q_2 риска невыполнения в срок этапа 2				0,16		0,07
Вероятность успешного выполнения этапа P_2				0,84		0,93
3. Разработка рабочей документации (i = 3)						
9	R_{13}	0,01	3	0,03	1	0,01
10	R_{23}	0,02	1	0,02	1	0,02
11	R_{33}	0,07	1	0,07	0	0
12	R_{43}	0,05	1	0,05	0	0
13	R_{53}	0,02	0	0	2	0,04
14	R_{63}	0,03	0	0	0	0
Оценка Q_3 риска невыполнения в срок этапа 3				0,17		0,07
Вероятность успешного выполнения этапа P_3				0,83		0,93
4. Изготовление опытного образца (i = 4)						
15	R_{14}	0,01	4	0,04	1	0,01
16	R_{24}	0,02	2	0,04	0	0
17	R_{34}	0,07	1	0,07	0	0
18	R_{44}	0,05	1	0,05		
19	R_{54}	0,01	0	0	2	0,01
20	R_{64}	0,03	2	0,06	1	0,03
21	R_{74}	0,01	0	0	1	0,01
Оценка Q_4 риска невыполнения в срок этапа 4				0,26		0,06
Вероятность успешного выполнения				0,74		0,94

ния этапа P_4						
5. Наземная отработка (испытания) ($i = 5$)						
22	R_{15}	0,02	1	0,02	1	0,02
23	R_{25}	0,02	2	0,04	1	0,02
24	R_{35}	0,12	2	0,24	0	0
25	R_{45}	0,01	0	0	1	0,01
26	R_{55}	0,03	1	0,03	0	0
Оценка Q_5 риска невыполнения в срок этапа 5				0,33		0,05
Вероятность успешного выполнения этапа P_5				0,67		0,95
6. Корректировка документации ($i = 6$)						
27	R_{16}	0,01	0	0	1	0,01
28	R_{26}	0,03	2	0,06	0	0
29	R_{36}	0,10	1	0,10	0	0
30	R_{46}	0,02	1	0,02	1	0,02
31	R_{56}	0,04	1	0,04	0	0
Оценка Q_6 риска невыполнения в срок этапа 6				0,22		0,03
Вероятность успешного выполнения этапа P_6				0,78		0,97
7. Летные испытания и доработка документации ($i = 7$)						
32	R_{17}	0,01	1	0,01	0	0
33	R_{27}	0,02	3	0,06	1	0,02
34	R_{37}	0,05	2	0,10	0	0
35	R_{47}	0,02	0	0	0	0
36	R_{57}	0,02	2	0,04	1	0,02
37	R_{67}	0,02	2	0,04	1	0,02
38	R_{77}	0,06	2	0,12	0	0
Оценка Q_7 риска невыполнения в срок этапа 7				0,37		0,06
Вероятность успешного выполнения этапа P_7				0,63		0,94
8. Запуск ($i = 8$)						
39	R_{18}	0,01	1	0,01	2	0,02
40	R_{28}	0,02	0	0	0	0
41	R_{38}	0,06	2	0,12	0	0
42	R_{48}	0,02	1	0,02	0	0
43	R_{58}	0,03	2	0,06	1	0,03
44	R_{68}	0,06	1	0,06	0	0
Оценка Q_8 риска невыполнения в срок этапа 8				0,27		0,05
Вероятность успешного выполнения этапа P_8				0,73		0,95
Вероятность $P = P_1 P_2 \dots P_8$ успешного выполнения проекта				0,094		0,629
Оценка $Q = 1 - P$ риска R , т.е. вероятность невыполнения проекта в срок				0,906		0,371

Согласно табл.4 вероятность выполнения в срок проекта 1 равна 0,094, т.е. это событие маловероятно (чуть менее 1 шанса из 10). Для этого проекта вероятности успешного выполнения этапов не малы - меняются от 0,63 до 0,84 (т.е. оценки рисков составляют от 0,16 до 0,37). Накапливаясь от этапа к этапу, оценка риска возрастает до явно недопустимого значения 0,906.

Для проекта 2 вероятности успешного выполнения этапов близки к 1 – меняются от 0,93 до 0,97 (т.е. оценки рисков составляют от 0,03 до 0,07). Тем не менее, накопи-

ваясь от этапа к этапу, оценка риска возрастает до заметного значения 0,371, соответственно вероятность успешного выполнения проекта 2 (т.е. в срок) заметно меньше 1 и составляет 0,629 (несколько менее 2 шансов из 3).

4.5.6. О дальнейшем развитии аддитивно-мультипликативной модели

С помощью АМ-модели, построенной на начальных этапах выполнения проекта (например, на этапе 1 «Концепция»), можно выделить факторы риска, вносящие наибольший вклад в суммарный риск. Управленческие воздействия, направленные на снижение соответствующих рисков, могут оказаться наиболее полезными для снижения суммарного риска. Так, согласно данным табл.4, для проекта 1 из рисков по этапам наибольшим является риск (с оценкой 0,37), соответствующий этапу 7 «Летные испытания и доработка документации», а для проекта 2 – риск (с оценкой 0,07), соответствующий этапу 3 «Разработка рабочей документации». Именно на эти этапы целесообразно обратить наибольшее внимание при управлении проектами. Согласно данным табл.4, отдельные риски второго порядка вносят в риски по этапам заметно больший вклад, чем другие виды рисков на тех же этапах. Так, для проекта 2 в риск на этапе 3 «Разработка рабочей документации» наибольший вклад (0,04 из 0,07, т.е. больше половины) вносит риск R_{53} – организационный риск (риск срыва работ из-за плохой их организации). Для проекта 1 в риск на этапе 7 «Летные испытания и доработка документации» наибольший вклад вносят риски R_{37} (риск недостатка ресурсов) и R_{77} (внешний риск) – 0,10 и 0,12 соответственно из оценки риска на этапе 0,37. На других этапах также можно выделить риски второго порядка, которые вносят в риски по этапам основной вклад. Так, для проекта 1 в риск на этапе 5 «Наземная отработка (испытания)» (оценка риска на этапе 0,33) наибольший вклад вносит риск R_{35} – риск недостатка ресурсов с оценкой 0,24 (более 70% от риска на этапе). Таким образом, можно выделить риски, управленческие воздействия на которые могут в наибольшей степени снизить общий риск.

Предположим, что известна стоимость реализации тех или иных управленческих решений. На основе АМ-модели можно рассчитать эффективность набора управленческих решений, т.е. снижение риска в результате применения этого набора. Согласно АМ-модели риск оценивается числом от 0 до 1. Более продвинутое исследование может дать возможность финансовой оценки рисков (в частности, потерь от реализации рискованных ситуаций и затрат на ликвидацию последствий). Решая задачу оптимизации, при заданном бюджете на снижение рисков находим наилучший набор управленческих решений, а также решаем двойственную задачу – определяем величину финансовых вложений, необходимых для снижения риска на заданную величину и в дальнейшем поддержания приемлемого уровня риска. Такой подход аналогичен примененному при разработке автоматизированной системы прогнозирования и предотвращения авиационных происшествий [14].

Для реализации описанного направления развития АМ-модели необходимо внести в модель информацию о последствиях невыполнения очередного этапа в ходе разработки РКТ. Можно оценивать накопленный риск на основе накопленной вероятности выполнения первых m этапов: $P = P_1 P_2 \dots P_k$, где k последовательно принимает значения 1, 2, 3, ..., m . Если вероятность P достаточно близка к 1, нет необходимости дополнять АМ-модель. Такая необходимость возникает, если вероятность P становится ниже определенного порога (заданного экспертно).

Что произойдет, если очередной этап не будет выполнен? Согласно предлагаемой здесь *простейшей модели* этап будет повторен и выполнен полностью. Тогда разработка РКТ идет по одной из следующих траекторий:

- 1) 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8
- 2) 1 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8
- 3) 1 – 2 – 2 – 3 – 4 – 5 – 6 – 7 – 8
- 4) 1 – 1 – 2 – 2 – 3 – 4 – 5 – 6 – 7 – 8
- 5) 1 – 2 – 3 – 3 – 4 – 5 – 6 – 7 – 8
- 6) 1 – 1 – 2 – 3 – 3 – 4 – 5 – 6 – 7 – 8
- 7) 1 – 2 – 2 – 3 – 3 – 4 – 5 – 6 – 7 – 8
- 8) 1 – 1 – 2 – 2 – 3 – 3 – 4 – 5 – 6 – 7 – 8
- 9) 1 – 2 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 10) 1 – 1 – 2 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 11) 1 – 2 – 2 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 12) 1 – 1 – 2 – 2 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 13) 1 – 2 – 3 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 14) 1 – 1 – 2 – 3 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 15) 1 – 2 – 2 – 3 – 3 – 4 – 4 – 5 – 6 – 7 – 8
- 16) 1 – 1 – 2 – 2 – 3 – 3 – 4 – 4 – 5 – 6 – 7 – 8

И т.д. (всего $2^8 = 256$ возможных траекторий). Вероятность каждой из этих траекторий может быть вычислена на основе данных табл.4. Поскольку известны оценки финансовых, временных и иных ресурсов для выполнения отдельных этапов, можно рассчитать необходимые ресурсы для каждой из траекторий, а затем построить функции распределения необходимых ресурсов и их характеристики (математическое ожидание, медиану, дисперсию, квантили...). Отметим, что согласно простейшей модели общий расход ресурсов увеличивается не более чем в 2 раза.

В связи с дальнейшим развитием АМ-модели отметим, что в подразделе 4.5.2 предложены модели на основе теории нечеткости и интервальных чисел. Одним из направлений развития является дальнейшая разработка процедур сбора и анализа экспертных оценок. Необходимо изучить устойчивость выводов по отношению к допустимым отклонениям исходных данных и предпосылок АМ-модели [15, 16].

4.6. Многообразие методов прогнозирования

С целью разработки адекватных методов прогнозирования выделим основные источники неопределенностей в различных производственных и экономических ситуациях. Рассмотрим здесь роль и задачи прогнозирования при управлении промышленными предприятиями ракетно-космической промышленности. Обсудим основные методы организационно-экономического прогнозирования, развивая положения статьи [1].

Основные функции менеджмента были выделены около 100 лет назад одним из основоположников научного менеджмента французом Анри Файолем: «Управлять – значит прогнозировать и планировать, организовывать, руководить командой, координировать и контролировать» [2]. Вслед за А. Файолем констатируем, что первая из основных функций менеджмента – прогнозирование и планирование.

Прогнозирование – это взгляд в будущее, оценка возможных путей развития, последствий тех или иных решений. Планирование же – это разработка последовательности действий, позволяющей достигнуть желаемого. В работе руководителя (управленца, менеджера) они тесно связаны. Результаты прогнозирования необходимы для планирования.

Практика показывает, что прогнозировать достаточно сложно. Иногда прогноз основывается на хорошо изученных закономерностях и осуществляется «наверняка». Например, при прогнозировании астрономических явлений. Однако в социально-экономической области обычно не удается дать однозначный полностью обоснованный прогноз. Причины – неопределённости в различных аспектах производственной и экономической ситуации.

4.6.1. Источники неопределенностей

Часть неопределенностей связана с недостаточностью знаний о природных явлениях и процессах, в частности:

- неопределенности, порожденные недостаточными знаниями о природе (например, неизвестен точный объем полезных ископаемых в конкретном месторождении, а потому невозможно точно предсказать развитие добывающей промышленности и объем налоговых поступлений от ее предприятий);

- неопределенности самих природных явлений (погода, влияющая на возможность запуска космического аппарата, на урожайность, на затраты на отопление, на загрузку транспортных путей).

Многие возможные неопределенности связаны с социально-экономическим окружением предприятия ракетно-космической промышленности или иной отрасли народного хозяйства, соответствующая служба которого занимается прогнозированием:

- неопределенности, относящиеся к деятельности (в текущий момент) участников экономической жизни (прежде всего партнеров рассматриваемого предприятия), в частности, с их деловой активностью, финансовым положением, соблюдением обязательств;

- неопределенности, обусловленные социальными и административными факторами в конкретных регионах, в которых данное предприятие имеет деловые интересы. Речь идет о взаимоотношении предприятия с органами местной и региональной власти, как официальными, так и криминальными;

- неопределенности, связанные с будущими действиями поставщиков в связи с меняющимися заданиями вышестоящих органов управления и предпочтениями рынка;

- неопределенности, порожденные конкурентным окружением (в том числе структур вероятного противника), от действий которого зависит многое в судьбе конкретного предприятия. Здесь имеют место возможности государственного или промышленного шпионажа, проникновения конкурентов в коммерческие тайны, и иное воздействие на внутренние дела предприятия, вплоть до диверсий.

Большое значение имеют и неопределенности на уровне страны, в частности:

- неопределенность будущей рыночной ситуации в стране, в том числе отсутствие достоверной информации о будущих предпочтениях государственного заказчика и других потребителей;

- неопределенности, связанные с колебаниями цен (динамикой инфляции), нормы (банковского) процента, валютных курсов, курсов акций и других макроэкономических показателей;

- неопределенности, порожденные нестабильностью законодательства и текущей экономической политики (т.е. с деятельностью руководства страны, министерств и ведомств), связанные с политической ситуацией, действиями партий, профсоюзов, экологических и других организаций в масштабе страны.

Часто приходится учитывать и внешнеэкономические неопределенности, связанные с ситуацией в зарубежных странах и международных организациях, с которыми предприятие поддерживает деловые отношения.

Большое значение имеют неопределенности, связанные с производством:

- дефектность продукции. Известно, что при массовом производстве, как правило, невозможно обеспечить выпуск продукции без дефектов;

- неопределенности, относящиеся к проектируемым продукции или технологическим процессам. Они могут быть связаны с ошибками разработчиков или физической невозможностью осуществления того или иного процесса;

- неопределенности, связанные с осуществлением действующих технологических процессов. Возможны аварии различной степени тяжести, от незначительных нарушений технологических процессов до катастроф с человеческими жертвами. Как следствие нарушения технологических процессов возникают экологические неопределенности, связанные с аварийными сбросами в реки технологических жидкостей, выбросами в атмосферу газов и др.

Среди неопределенностей на предприятии есть и социальные (риски персонала), связанные с различными конфликтами – между службами (отделами, цехами), между руководителями высшего звена (топ-менеджерами), между профсоюзами и администрацией (по поводу заработной платы, условий труда и др.).

Неопределенности порождают соответствующие риски. Поэтому прогнозирование необходимо для адекватного принятия решений в условиях неопределенности и риска.

4.6.2. Что и зачем прогнозировать?

В современных условиях хозяйственной независимости основной массы промышленных предприятий для многих из них стал весьма актуальным вопрос о прогнозировании. При составлении плана производства важны не только возможности предприятия, но и спрос на выпускаемую продукцию. Сейчас, когда предприятия вынуждены работать по «рыночным законам», руководители (менеджеры) хотят знать перспективы развития своего предприятия, взглянуть в будущее, чтобы оценить возможные пути развития, предугадать последствия тех или иных решений.

Эффективная деятельность промышленных предприятий в современных условиях в значительной степени зависит от прогнозирования, т.е. насколько достоверно менеджеры предвидят дальнюю и ближнюю перспективу своего развития. Прогнозирование – частный вид моделирования как основы познания и управления [3].

Роль прогнозирования в управлении предприятием очевидна. Первично соответствующим службам предприятия ракетно-космической отрасли необходимо прогнозировать:

- поведение государственных органов управления,
- поведение потребителей,
- поведение поставщиков,
- поведение конкурентов,
- научно-технический прогресс.

Вторичными прогнозируемыми показателями, определяющими успешное существование промышленного предприятия в долгосрочной перспективе, являются:

- величина прибыли (в различных определениях этого термина – от основной деятельности, балансовой (налогооблагаемой), чистой, реинвестируемой и др.),
- объем реализации продукции;
- рентабельность, в том числе активов различных видов;
- фондоотдача,
- показатели ликвидности,
- производительность труда и т.д.

Наличие неопределенностей у этих факторов значительно усложняют процесс управления промышленным предприятием. Это связано с тем, что в связи с развитием предприятия ракетно-космической промышленности возникают новые цели и конкретные задачи (в терминах стратегического менеджмента [2]), налаживаются новые хозяйственные связи, формируются новые механизмы управления. Обеспечение методической и инструментальной базы для поддержки основных функций менеджмента лежит на контроллинге. Это новая концепция управления, порожденная практикой современного менеджмента [4]. Можно говорить о том, что одним из инструментов руководителя (менеджера), обеспечивающий успех предприятию, является контроллинг. Это относится и к управлению деятельностью научно-исследовательских организаций, в том числе в ракетно-космической отрасли [5].

Контроллеру (специалисту службы контроллинга) и сотрудничающему с ним специалисту по анализу данных (эконометрику) нужна разнообразная экономическая и управленческая информация, а также удобные инструменты ее анализа. Следовательно, необходима информационная поддержка контроллинга. Без современных компьютерных инструментов анализа и управления, основанных на продвинутых эконометрических и экономико-математических методах и моделях, невозможно эффективно принимать управленческие решения. Недаром специалисты по контроллингу большое внимание уделяют проблемам создания, развития и применения компьютерных систем поддержки принятия решений. Высокие статистические технологии и эконометрика – неотъемлемые части любой современной системы поддержки принятия экономических и управленческих решений [6]. Итак, эконометрика – важная составляющая инструментария контроллера, воплощенного в компьютерной системе поддержки принятия решений [3].

4.6.3. Методы организационно-экономического прогнозирования

Разработаны различные методы прогнозирования. Их теоретической основой являются математические дисциплины: теория вероятностей, математическая статистика, дискретная математика, дифференциально-разностные уравнения, математическое моделирование, исследование операций, а также экономика предприятия (инженерная экономика), экономическая теория, экономическая статистика, менеджмент (в том числе менеджмент высоких технологий [7]), социология, политология и другие социально-экономические науки. Выбор конкретного метода является одной из наиболее важных задач прогнозирования. При этом можно указать три основные группы причин, влияющих на выбор метода прогнозирования.

Первая группа состоит в увеличении числа классов (групп, видов, вариантов) методов прогнозирования, которое и в перспективе будет возрастать в связи с разнообразием практических задач прогнозирования. Отметим, что и в настоящее время число конкретных методов прогнозирования, строго говоря, бесконечно много. Например, при использовании линейной прогностической функции:

$$x(t) = at + b, \quad (1)$$

неизвестные параметры a и b могут быть найдены путем решения оптимизационной задачи:

$$\sum_{i=1}^n |x_i - at_i - b|^c \rightarrow \min, c > 0, \quad (2)$$

где t – независимая переменная, x – зависимая переменная, (t_i, x_i) , $i = 1, 2, \dots, n$, – исходные данные, c – параметр, определяющий метод оценивания неизвестных параметров. При этом, если $c = 2$, то имеем метод наименьших квадратов (наиболее популярный в рассматриваемом классе методов), если $c = 1$ – метод наименьших модулей, если же

параметр $c \rightarrow +\infty$ – метод Чебышёва (минимизация максимального отклонения). Поскольку параметр c – действительное число, то всего в рассматриваемом классе столько же методов, сколько действительных чисел. На языке математики – множество методов рассматриваемого класса имеет мощность континуума.

Вторая группа причин заключается в том, что в современных социально-экономических условиях существенно возрастает сложность как самих решаемых задач, так и объектов прогнозирования (в частности, из-за создания корпоративных групп, холдингов, объединений и других сложных организационно-производственных структур).

Третья группа причин связана с возрастанием динамичности (подвижности) социально-экономической среды, ускорением темпов инновационных процессов в ракетно-космической промышленности в соответствии с решениями руководства страны.

Поэтому на выбор конкретного метода (или группы методов) прогнозирования влияют:

- существо проблемы, подлежащей решению;
- динамические характеристики объекта прогнозирования;
- вид и характер информационного обеспечения;
- выбранный период упреждения прогноза (и его соотношение с продолжительностью цикла разработки товара или услуги);
- требования к результатам прогнозирования (точности, надежности и достоверности) [8].

По нашему мнению, среди методов прогнозирования базисными являются две группы – статистические и экспертные. Они используются как непосредственно, так и в составе более сложных комбинированных систем прогнозирования. Дадим краткое представление об этих группах методов прогнозирования (статистических, экспертных, комбинированных).

4.6.4. Статистические методы прогнозирования

Кратко охарактеризуем основные методы прогнозирования, основанные на вероятностно-статистических моделях.

4.6.4.1. ПРОГНОЗ ВРЕМЕННЫХ РЯДОВ

Под временным рядом понимают значения экономической (или иной) величины (или величин), выраженной как функция времени. При этом время является дискретным, в противном случае говорят о случайных процессах, а не о временных рядах. Обычно в поведении временного ряда выявляют две основные компоненты – тренд и отклонения от тренда. Под трендом понимают детерминированную зависимость от времени, которую выявляют тем или иным способом сглаживания (например, экспоненциального сглаживания) либо расчетным путем, в частности, с помощью метода наименьших квадратов. Другими словами, тренд – это очищенная от случайностей основная тенденция временного ряда. В вероятностно-статистических методах под трендом понимают математическое ожидание временного ряда (как функцию времени).

Временной ряд обычно колеблется вокруг некоторой достаточно простой функции от времени, причем отклонения от нее иногда обнаруживают правильность. Часто это связано с естественной или назначенной периодичностью, например, сезонной или недельной, месячной или квартальной. Иногда наличие периодичности и тем более ее причины не ясны, и задача статистического анализа данных – выяснить, действительно ли имеется периодичность [9].

Временной ряд может быть как одномерным, так и многомерным (число зависимых переменных больше одной). Для анализа временного ряда наиболее часто используется метод наименьших квадратов. Методы наименьших модулей, метод Чебышёва (минимакса) и иные применяются реже. Применяются также эвристические приемы: метод скользящих средних, метод экспоненциального сглаживания и др.

4.6.4.2. РЕГРЕССИОННЫЙ АНАЛИЗ

Регрессионный анализ предназначен для восстановления зависимости определенной величины (или нескольких величин) от других величин. Он используется для различных видов прогнозирования (объема инвестиций, уровня затрат, финансовых результатов, объемов продаж и т.п.). Многомерная регрессия, в том числе с использованием непараметрических оценок плотности распределения – основной на настоящий момент математико-статистический аппарат прогнозирования.

4.6.4.3. АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Адаптивные методы прогнозирования позволяют оперативно корректировать прогнозы при появлении новых точек. Выделяют адаптивные методы оценивания параметров моделей и адаптивные методы непараметрического оценивания. Используют и более специальные модели, такие, как модель авторегрессии, модель Бокса-Дженкинса, системы эконометрических уравнений [10].

4.6.4.4. КОМПЬЮТЕРНЫЕ СТАТИСТИЧЕСКИЕ ТЕХНОЛОГИИ

Для установления возможности применения асимптотических результатов при малых объемах выборок полезны компьютерные статистические технологии. Они позволяют также строить различные имитационные модели.

4.6.4.5. МЕТОДЫ СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ

При разработке алгоритмов прогнозирования широко используют методы статистических испытаний. Этот термин применяется для обозначения компьютерных технологий, в которых в эконометрическую модель искусственно вводится большое число случайных элементов. Обычно моделируется последовательность независимых одинаково распределенных случайных величин, на основе которых рассчитываются функции от них, например, последовательность, построенная на основе последовательности накапливающихся (кумулятивных) сумм.

Необходимость в методе статистических испытаний возникает потому, что чисто теоретические методы лишь в исключительных случаях дают точное решение. Это происходит, когда исходные случайные величины имеют вполне определенные функции распределения, например, нормальные (как правило, на практике подобные предположения не выполняются [11]), либо когда объемы выборок очень велики (с практической точки зрения бесконечны).

Второе название методов статистических испытаний – методы Монте-Карло. В методах статистических испытаний можно выделить две составляющие. База – датчики псевдослучайных чисел. Результатом работы таких датчиков являются последовательности чисел, которые обладают некоторыми свойствами последовательностей случайных величин. Надстройкой служат различные алгоритмы, использующие последовательности псевдослучайных чисел [11].

Модель в общем смысле (обобщенная модель) есть создаваемый с целью получения и (или) хранения информации специфический объект (в форме мысленного образа, описания знаковыми средствами либо материальной системы), отражающий свойства, характеристики и связи объекта-оригинала произвольной природы, существенные для

задачи, решаемой субъектом [12]. Для теории принятия решений наиболее полезны модели, которые выражаются словами или формулами, алгоритмами и иными математическими средствами. При более тщательном анализе словесных моделей, как правило, недостаточно. Необходимо применение достаточно сложных математических моделей. Так, при принятии решений в менеджменте производственных систем, в том числе в ракетно-космической отрасли, используются:

- модели технологических процессов (прежде всего модели контроля и управления);
- модели обеспечения качества продукции (в частности, модели оценки и контроля надежности);
- модели массового обслуживания;
- модели управления запасами (модели логистики);
- имитационные и эконометрические модели деятельности предприятия в целом, и др.

4.6.4.6. МЕТОДЫ РАЗМНОЖЕНИЯ ВЫБОРОК

Методы размножения выборок – методы, используемые при изучении свойств статистических процедур. Рассмотрим основную идею группы методов «размножения выборок», наиболее известным представителем которых является бутстреп – дальнейшее развитие «метода складного ножа». Сам термин «бутстреп» (bootstrap) буквально означает «вытягивание себя за шнурки от ботинок».

Идея, которую предложил в 1949 году М. Кенуй («метод складного ножа») состоит в том, чтобы из одной выборки сделать много, исключая по одному наблюдению (и возвращая ранее исключенные). Б. Эфрон разработал новый способ размножения выборок, существенно использующий датчики псевдослучайных чисел. Он предложил строить новые выборки, моделируя выборки из эмпирического распределения [11].

Есть много способов развития идеи размножения выборок [13]. Можно по исходной выборке построить эмпирическую функцию распределения, а затем каким-либо образом от кусочно-постоянной функции перейти к непрерывной функции распределения. Другой вариант – перейти к непрерывному распределению, построив непараметрическую оценку плотности. После этого рекомендуется брать размноженные выборки из этого непрерывного распределения (являющегося состоятельной оценкой исходного), непрерывность защитит от совпадений элементов в этих выборках. Следующий вариант построения размноженных выборок – более прямой. Исходные данные не могут быть определены совершенно точно и однозначно. Поэтому предлагается к исходным данным добавлять малые независимые одинаково распределенные погрешности. При таком подходе одновременно соединяем вместе идеи устойчивости и бутстрепа.

В новых научно-практических областях со сложными алгоритмами, свойства которых недостаточно ясны, бутстреп представляет собой ценный инструмент для изучения ситуации.

Не всегда статистические методы используются в чистом виде. Часто их включают в виде важных элементов в комплексные методики, предусматривающие сочетание статистических методов с другими, например, экспертными оценками.

Статистические методы прогнозирования продолжают бурно развиваться. Так, недавно разработан непараметрический метод наименьших квадратов для восстановления зависимости с периодической составляющей [14].

Статистические методы основаны на использовании объективной информации. В прогнозировании весьма полезными являются также экспертные методы сбора и анализа субъективной информации, полученной от специалистов.

4.6.5. Экспертные методы прогнозирования

Для принятия решения, например, об экономических, социальных, экологических и иных проектах, в том числе требующих крупных инвестиций, в случае чрезвычайной сложности системы (объекта прогнозирования), его новизны, недостаточной полноты информации и невозможности полной математической формализации процесса, необходимо обратиться к методам экспертных оценок.

Методы экспертных оценок – это методы организации работы со специалистами-экспертами и обработки мнений экспертов [15].

Эти мнения обычно выражены частично в количественной, частично в качественной формах. Экспертные исследования проводят с целью подготовки информации для принятия решений лицом, принимающим решения (ЛПР). Для проведения работы по методу экспертных оценок создают Рабочую группу (сокращенно РГ), которая и организует по поручению ЛПР деятельность экспертов, объединенных (формально или по существу) в экспертную комиссию (ЭК).

Экспертные оценки бывают:

- индивидуальные – это оценки одного специалиста.
- коллективные – это сведенные вместе оценки многих специалистов.

Существует масса методов получения экспертных оценок. В одних с каждым экспертом работают отдельно, он даже не знает, кто еще является экспертом, а потому высказывает свое мнение независимо от авторитетов. В других – экспертов собирают вместе для подготовки материалов для ЛПР, при этом эксперты обсуждают проблему друг с другом, учатся друг у друга, и неверные мнения отбрасываются. В одних методах число экспертов фиксировано и таково, чтобы статистические методы проверки согласованности мнений и затем их усреднения позволяли принимать обоснованные решения. В других – число экспертов растет в процессе проведения экспертизы. Существует множество методов обработки ответов экспертов, они компьютеризированы и весьма насыщены математикой. На языке математики – множество экспертных методов имеет мощность континуума. Рассмотрим несколько конкретных процедур экспертных оценок.

4.6.5.1. МЕТОД ДЕЛЬФИ

Название данному методу дано по ассоциации с Дельфийским храмом, куда согласно древнегреческой легенде было принято обращаться для получения поддержки при принятии решений. Он был расположен у выхода ядовитых вулканических газов. Жрицы храма, надышавшись отравы, начинали пророчествовать, произнося непонятные слова. Специальные «переводчики» – жрецы храма – толковали эти слова и отвечали на вопросы паломников, пришедших со своими проблемами.

В США в 1960-х годах методом Дельфи называли экспертную процедуру прогнозирования научно-технического развития. В первом туре эксперты называли вероятные даты тех или иных будущих свершений. Во втором туре каждый эксперт знакомился с прогнозами всех остальных. Если его прогноз сильно отличался от прогнозов основной массы, его просили пояснить свою позицию, и иногда он изменял свои оценки, приближаясь к средним значениям. Процедуру повторяли, пока средние значения не переставали меняться. Эти средние значения и выдавались заказчику как групповое мнение.

С помощью метода Дельфи момент высадки человека на Луну был предсказан с точностью до месяца. Этот успех привел к широкому развертыванию как прикладных работ с использованием экспертных оценок, так и теоретических исследований по разработке новых экспертных процедур и изучению их свойств.

4.6.5.2. МЕТОД СЦЕНАРИЕВ

Метод сценариев применяется, прежде всего, для экспертного прогнозирования. Рассмотрим его основные идеи. Технологическое, экологическое или социально-экономическое прогнозирование, как и любое прогнозирование вообще, может быть успешным лишь при некоторой стабильности условий. Однако решения органов власти, отдельных лиц, иные события меняют условия, и события развиваются по-иному, чем ранее предполагалось. Целесообразно выделить сценарии (пути) развития ситуации. Например, при разработке методологического, программного и информационного обеспечения анализа риска химико-технологических проектов признано необходимым составление детального каталога сценариев аварий, связанных с утечками токсических химических веществ. Каждый из таких сценариев описывает аварию своего типа, со своим индивидуальным происхождением, развитием, последствиями, возможностями предупреждения.

Метод сценариев – это метод декомпозиции задачи прогнозирования, предусматривающий выделение набора отдельных вариантов развития событий (сценариев), в совокупности охватывающих все возможные варианты развития. При этом каждый отдельный сценарий должен допускать возможность достаточно точного прогнозирования, а общее число сценариев должно быть обозримо.

Возможность подобной декомпозиции не очевидна и обосновывается в каждом конкретном случае.

При применении метода сценариев необходимо осуществить два этапа исследования:

- построение исчерпывающего, но обозримого набора сценариев;
- прогнозирование в рамках каждого конкретного сценария с целью получения ответов на интересующие исследователя вопросы.

Каждый из этих этапов лишь частично формализуем. Существенная часть рассуждений проводится на качественном уровне, как это принято в общественно-экономических и гуманитарных науках. Одна из причин заключается в том, что стремление к излишней формализации и математизации приводит к искусственному внесению определенности там, где ее нет по существу, либо к использованию громоздкого математического аппарата. Так, рассуждения на словесном уровне считаются доказательными в большинстве ситуаций, в то время как попытка уточнить смысл используемых слов с помощью, например, теории нечетких множеств, приводит к весьма громоздким математическим моделям. Набор сценариев должен быть обозрим. Приходится исключать различные маловероятные события. Само по себе создание набора сценариев – предмет экспертного исследования. Кроме того, эксперты могут оценить вероятности реализации того или иного сценария. Прогнозирование в рамках каждого конкретного сценария с целью получения ответов на интересующие исследователя вопросы также осуществляется в соответствии с описанной выше методологией прогнозирования. При стабильных условиях могут быть применены статистические методы прогнозирования временных рядов. Однако этому предшествует анализ с помощью экспертов, причем зачастую прогнозирование на словесном уровне является достаточным (для получения интересующих исследователя и ЛПР выводов) и не требующим количественного уточнения. Примером является работа [16], посвященная сценариям социально-экономического развития России.

Другой вариант метода сценариев часто применяют при составлении бизнес-планов [2]. Финансовый поток инвестиционного проекта рассматривают как вероятный. Оптимистический сценарий соответствует тому, что поступления увеличиваются на определенный процент, например, на 10%, а платежи – уменьшают на 10%. В пес-

симистическом сценарии, наоборот, поступления уменьшаются на определенный процент, например, на 10%, а платежи – увеличиваются на 10%. Затем рассчитываются характеристики инвестиционного проекта, соответствующие трем сценариям, и сопоставляются между собой. Моделям и методам анализа и синтеза сценариев развития социально-экономических систем посвящен фундаментальный двухтомник [17, 18].

4.6.5.3. МОЗГОВОЙ ШТУРМ

«Мозговой штурм» организуется как дискуссия на собрании экспертов, на выступления которых наложено одно, но очень существенное ограничение – нельзя критиковать предложения других. Можно их развивать, можно высказывать свои идеи, но нельзя критиковать. В ходе заседания эксперты, «заражаясь» друг от друга, высказывают все более экстравагантные соображения. Часа через два записанное на диктофон или видеокамеру заседание заканчивается, и начинается второй этап мозгового штурма – анализ высказанных идей. Обычно из 100 идей 30 заслуживают дальнейшей проработки, из них 5 - 6 дают возможность сформулировать прикладные проекты, а 2 - 3 оказываются перспективными для реализации, и одна из них в итоге приносит полезный эффект – увеличение прибыли, повышение экологической безопасности и т.п. При этом интерпретация идей – творческий процесс.

4.6.5.4. АНАЛИЗ СИТУАЦИИ

Для разработки управленческих решений с помощью экспертов используют метод «анализа ситуации». Полезно выделить этапы анализа ситуации, подготовки и принятия решения, анализа их последствий [19]:

- 1) Уяснить ситуацию.
- 2) Установить наличие проблемы, подлежащей решению.
- 3) Сформировать возможные решения.
- 4) Описать последствия решений.
- 5) Выбрать решение.
- 6) Обобщить накопленный опыт принятия решений.

Целесообразно уточнить содержание каждого из перечисленных этапов. Например, для уяснения ситуации целесообразно ответить на пять вопросов:

- 1) КТО должен или обязан (или хочет) принять решение?
- 2) ГДЕ (в каком месте, в каком окружении, в какой среде, при каких обстоятельствах) предстоит принимать решение?
- 3) КОГДА (до какого срока, или насколько часто, с какой периодичностью) необходимо принимать решение?
- 4) КАК (каким образом, в какой форме, каким документом) должно быть выражено решение?
- 5) ЧТО обуславливает решение? Зачем оно нужно? В чем его цель? Какой замысел лежит в его основе? Для чего оно служит? Зачем его надо принимать?

После того, как ситуация обдумана, необходимо рассмотреть варианты управленческих решений. Разрабатываются новые математические методы оценки эффективности управленческих решений [20].

Экспертные методы прогнозирования продолжают бурно развиваться. Так, недавно в связи с разработкой автоматизированной системы прогнозирования и предотвращения авиационных происшествий предложены и применены новые экспертные технологии для оценивания вероятностей редких событий [21].

Статистические и экспертные методы зачастую являются блоками в более развитой прогностической системе.

4.6.6. Комбинированные методы

Особое место в классификации методов прогнозирования занимают комбинированные методы, которые предполагают одновременное использование различных методов прогнозирования. Использование комбинированных методов особенно актуально для сложных социально-экономических систем, когда при разработке прогноза показателей каждого элемента системы могут быть использованы различные сочетания методов прогнозирования. Разновидностью комбинированных методов можно считать эконометрическое моделирование, в которых экономико-математическое моделирование сочетается со статистическими методами оценивания параметров и проверки гипотез [22].

Ключевые слова при рассмотрении проблем создания и использования современных комбинированных методов прогнозирования и – на его основе – планирования таковы: системные исследования, экспертные оценки, прогнозирование и планирование, контроллинг. Большое значение имеет использование современных информационно-коммуникационных технологий [23].

Как отмечено в статье [24]: «... при поиске механизмов принятия эффективных корпоративных решений (например, по вопросам инвестирования, менеджмента, инноваций и др.) напрашивается особое внимание уделить использованию информационных систем поддержки решений (СПР, Decision Support Systems). Они предназначены для интеграции возможностей современных средств сбора, аналитической обработки и визуального представления информации, а также поддержки групповой деятельности экспертов. Большой интерес проявляется сегодня к ситуационным комнатам, которые позволяют быстро «погрузить» участников процесса принятия решений в рассматриваемую проблему, «научить их говорить» на одном языке, помочь разобраться в проблеме, правильно сформулировать запросы к внешним источникам информации и совместно подготовить хорошее (не обязательно наилучшее) решение».

Среди комбинированных методов известна технология Форсайта [25]. В переводе с английского Foresight – предвидение, в то время как Prediction – предсказание, Forecasting – прогнозирование, Prognostication - прогнозирование, предсказание, предзнаменование. Как сказано на официальном сайте Форсайт-центра Высшей школы экономики: «Форсайт представляет собой систему методов экспертной оценки стратегических направлений социально-экономического и инновационного развития, выявления технологических прорывов, способных оказать воздействие на экономику и общество в средне- и долгосрочной перспективе» (<http://foresight.hse.ru/whatforesight/>). Метод весьма трудоемок, поскольку требует привлечения тысяч экспертов, а не десятков, как в основной массе экспертных технологий [15]: «... в японских долгосрочных прогнозах научно-технологического развития, проводимых каждые пять лет, участвует более 2-х тысяч экспертов, которые представляют все важнейшие направления развития науки, технологий и техники, а в последнем корейском проекте участвовали более 10 тысяч экспертов...» (<http://foresight.hse.ru/whatforesight/>).

Форсайт ориентирован на совмещение функций прогнозирования и планирования: «... Форсайт исходит из того, что наступление «желательного» варианта будущего во многом зависит от действий, предпринимаемых сегодня, поэтому выбор вариантов сопровождается разработкой мер, обеспечивающих оптимальную траекторию инновационного развития ... Важным результатом является развитие неформальных взаимосвязей между их участниками, создание единого представления о ситуации... Третье главное отличие Форсайта от традиционных прогнозов – нацеленность на разработку практических мер по приближению выбранных стратегических ориентиров...» (<http://foresight.hse.ru/whatforesight/>).

Сказанное показывает, что Форсайт плохо сочетается с менеджментом высоких технологий [7], в частности, не является перспективным для ракетно-космической промышленности из-за ее высокой наукоемкости, с одной стороны, и централизованной системой принятия решений, в другой. Его сфера применения, как и подчеркнуто выше, совсем другая, в частности, связанная с развитием социально-экономических систем, например, регионов, когда важно налаживание "неформальных взаимосвязей" с целью предотвращения конфликтов.

На основе проведенного выше анализа многообразия методов прогнозирования можно сформулировать ряд практических предложений, ориентированных на ракетно-космическую промышленность. Отметим, что они могут быть использованы и в других высокотехнологичных наукоемких отраслях промышленности.

4.6.7. Предложения по совершенствованию механизмов прогнозирования и планирования для практического использования при создании космических комплексов

4.6.7.1. Действующую систему планирования целесообразно совершенствовать в соответствии с положениями Федерального закона «О стратегическом планировании в Российской Федерации» N 172-ФЗ [26] и рекомендациями теории планирования в стратегическом менеджменте [2], теории управления организационными системами (<http://mtas.ru>), крупномасштабными системами, проектами, в том числе инновационными и инвестиционными, контроллингом, короче, в соответствии с достижениями современной организационно-экономической науки. Конкретные рекомендации по совершенствованию системы планирования могут быть выработаны в результате выполнения соответствующей научно-исследовательской работы с учетом специфики отрасли.

4.6.7.2. Целесообразно создать функциональную структуру прогнозирования, учитывая при этом рекомендации по созданию такой структуры, полученные в диссертации [22]. Функциональная структура прогнозирования может действовать в рамках службы контроллинга [5].

4.6.7.3. Поскольку в настоящее время отсутствует банк знаний (энциклопедия) по методам прогнозирования, то целесообразно выполнить комплекс научно-исследовательских работ по систематизации (аналитическому обзору) методов прогнозирования с целью создания указанного банка знаний.

4.6.7.4. На основе банка знаний по методам прогнозирования целесообразно разработать ориентированный на практическое применение при создании космических комплексов программный продукт, предназначенный для реализации отобранных в соответствии с п.4.6.7.3 методов прогнозирования.

4.6.7.5. Из статистических методов прогнозирования целесообразно использовать непараметрические (на основе рекомендаций монографии [11]).

4.6.7.6. Для нужд прогнозирования предлагаем использовать весь арсенал современных экспертных технологий [15].

4.6.7.7. Если сформулировать коротко, то предлагаем разработать

А) систему современных методов прогнозирования для практического использования при создании космических комплексов и

Б) систему современных методов планирования, в том числе стратегического, с адекватным организационным и программным обеспечением, предназначенную для практического использования при создании космических комплексов.

4.7. Оценка инфляции по независимой информации

Работами по сбору и анализу независимо собранной информации о ценах, т.е. по изучению реальной инфляции, мы занялись в середине 1990-х гг. по заказу Министерства обороны РФ. Заказчика интересовали размеры финансирования НИР в реальных (сопоставимых) ценах. Был создан творческий коллектив (под руководством одного из авторов настоящей монографии) из преподавателей и студентов Московского государственного института электроники и математики (технического университета), который и проводил эту работу. Наиболее активные члены коллектива указаны ниже в числе соавторов публикаций. В международной газете «Наука и технология в России» членами нашего коллектива помещен ряд статей об инфляции.

Первая научная публикация по индексу инфляции формально имела статус учебного материала для студентов – методических рекомендаций по курсу «Основы экономики» [1]. По существу же это была небольшая научная монография. Еще интереснее, что подготовлена она была вместе со студентами, поступившими в вуз в 1993 г. – т.е. к моменту выхода книги они перешли на второй курс. Авторы этой и дальнейших работ Балашов В.В., Куроптев О.В., Канакова Е.М., Рафальская А.С., Иванова И.Г., Точенная Н.С. – наиболее активные из студентов, работавших в нашем творческом коллективе. Отметим также, что в этой работе были сформулированы фундаментальные «теорема умножения» и «теорема сложения» для индекса инфляции. Авторы настоящей монографии будут благодарны за информацию об аналогичных формулировках в других учебных изданиях. Обычно обсуждение идет на словесном уровне.

Вслед за основополагающей публикацией [1] последовала серия статей и докладов [2–9]. Развернутые публикации по результатам проведенных исследований выпущены в конце 1990-х гг. [10–11]. Работы на близкую тему – о дифференциации по доходам – появились на рубеже тысячелетий [12–14].

Полученные при изучении инфляции научные результаты включены в ряд учебных курсов. Они послужили основой для главы 7 учебника «Эконометрика» [15] и аналогичных разделов других учебников А.И. Орлова (см. перечень в [16]), выпущенных уже в XXI веке.

Работы по сбору и анализу независимой информации о ценах можно отнести к экспериментальным исследованиям, как следствие, они достаточно трудоемки. Исходя из интересов обеспечения экономической науки экспериментальным материалом, их целесообразно продолжать и углублять. В настоящее время такие работы выполняются студентами как лабораторные. В частности, на основе собранной студентами факультета «Инженерный бизнес и менеджмент» МГТУ им. Н.Э. Баумана информации о реальных ценах весны 2004 г. проанализировано распределение индекса инфляции по различным точкам сбора данных в Москве и Московской области [17].

Вопросы применения индексов инфляции при анализе финансово-хозяйственной деятельности предприятий и организаций рассмотрены в кандидатской диссертации В.С. Муравьевой [18]. Такие вопросы стали практически обязательными для дипломников кафедры ИБМ-2 "Экономика и организация производства" МГТУ им. Н. Э. Баумана.

Исследования продолжаются. За последние годы опубликованы статьи и тезисы докладов [19–21]. Заново написана глава по инфляции для 4-го издания учебника "Эконометрика" [22]. Целесообразна дальнейшая глубокая теоретическая и прикладная проработка рассматриваемой тематики. Наш коллектив наработал большой задел.

Индекс инфляции применяют для приведения к сопоставимому виду экономических величин, измеренных в различные моменты времени. Так, номинальный размер пенсий, доходность пенсионных фондов (в накопительной системе) сравнивают с рос-

том цен. Если доходность фонда больше роста цен, пенсионные накопления в реальном исчислении (в сопоставимых ценах) растут, если меньше – их покупательная способность уменьшается. Поэтому при анализе данных о пенсионных накоплениях важен измеритель роста цен – индекс инфляции. Его используют для приведения значений экономических показателей к сопоставимому виду (к ценам на фиксированный момент времени). Нами установлено, что **индексы инфляции Росстата на десятки и сотни процентов преуменьшают реальные индексы инфляции, полученные при независимых измерениях.**

С 1993 г. мы используем один и тот же инструмент исследования - *минимальную потребительскую корзину физиологически необходимых продовольственных товаров*, разработанную в 1993 г. на основе исходных данных Института питания Российской академии медицинских наук (РАМН). Она описана в табл.1. Нами рассчитаны индексы инфляции (индексы потребительских цен), приведенные в табл.2. Установлено, что официальная статистика занижала реальную инфляцию в 1,5 – 2,0 раза (табл.3). Обсуждаем данные о реальных значениях средней заработной платы, используемых в Пенсионном фонде (табл.4). Далее проведен более подробный анализ на материалах последних лет, собранных под нашим руководством.

4.7.1. Потребительская корзина - инструмент измерения роста цен

Как известно [22], подход к измерению роста цен основан на выборе и фиксации инструмента экономиста и управленца - *потребительской корзины* ($Q_1(t), Q_2(t), \dots, Q_n(t)$), *не меняющейся со временем*, т.е. ($Q_1(t), Q_2(t), \dots, Q_n(t) \equiv (Q_1, Q_2, \dots, Q_n)$). Здесь Q_i – заданный при формировании потребительской корзины объем потребления i -го товара или услуги из n включенных в корзину. Пусть $p_i(t), i=1,2,\dots,n$, - цена i -го товара (услуги) в момент t . Стоимость $S(t)$ потребительской корзины в момент времени t такова:

$$S(t) = \sum_{1 \leq i \leq n} p_i(t) Q_i$$

Для оценки роста цен необходимо сравнить стоимости $S(t_1)$ и $S(t_2)$ потребительской корзины (Q_1, Q_2, \dots, Q_n) в старых $p_i(t_1), i=1,2,\dots,n$, и новых $p_i(t_2), i=1,2,\dots,n$, ценах.

Определение. *Индексом инфляции называется отношение стоимостей потребительской корзины в конечный t_2 и начальный t_1 моменты времени*

$$I(t_1, t_2) = \frac{S(t_2)}{S(t_1)} = \frac{\sum_{1 \leq i \leq n} p_i(t_2) Q_i}{\sum_{1 \leq i \leq n} p_i(t_1) Q_i}$$

Здесь индексируемые величины – цены, а весами служат объемы потребления, зафиксированные в принятой исследователем потребительской корзине. С математической точки зрения индекс инфляции – это функция двух переменных, а именно, двух моментов времени – начального, или базового, момента t_1 и конечного, или текущего, момента t_2 . Когда говорят об инфляции за определенный промежуток времени, то t_1 – начало этого промежутка (года, месяца), а t_2 – его конец. Обычно $t_1 < t_2$, хотя с математической точки зрения это не обязательно.

На протяжении более двадцати лет мы (Институт высоких статистических технологий и эконометрики (ИВСТЭ) МГТУ им. Н. Э. Баумана - исследовательский коллектив под руководством проф. А.И. Орлова) используем одну и ту же (в отличие от Росстата, который постоянно меняет корзины по конъюнктурным соображениям) *минимальную потребительскую корзину физиологически необходимых продовольственных товаров*, разработанную в 1993 г. ИВСТЭ на основе исходных данных Института

питания Российской академии медицинских наук (РАМН). Почти совпадающая потребительская корзина использовалась в первой половине 1990-х гг. Министерством труда Российской Федерации для оценки прожиточного минимума. Рассматриваемую минимальную потребительскую корзину, впервые опубликованную в [1], обозначим сокращенно «корзина ИВСТЭ».

В корзине ИВСТЭ продукты питания разделены на 11 групп: (1) хлеб и хлебопродукты; (2) картофель; (3) овощи; (4) фрукты и ягоды; (5) сахар; (6) мясопродукты; (7) рыба и рыбопродукты; (8) молоко и молочные продукты; (9) яйца; (10) масло растительное и маргарин; (11) прочие. Общая стоимость «прочих» видов продуктов - 6% от стоимости первых 10 групп продуктов данной потребительской корзины.

На основе физиологических норм потребления Института питания РАМН в ИВСТЭ составлена минимальная потребительская корзина, т.е. указан годовой объем потребления по основным продовольственным товарам, необходимый для поддержания нормальной жизнедеятельности человеческого организма (табл.1). При разработке корзины исходили из трех принципов:

1. Суммарное содержание белков, жиров, углеводов и калорий должно быть не менее нормативов, определяющих согласно науке о питании (как части медицины) возможность продолжения существования человеческого организма без физиологического вырождения.

2. На основе включенных в корзину продуктов может быть разработано меню трехразового питания на год.

3. Стоимость корзины должна быть минимальна.

Первый и третий принципы позволяют сформулировать задачу оптимизации (линейного программирования). Ее решение таково (в расчете на день): 812 г черного хлеба, 705 г картофеля, 180 г молока и 10 г сыра. Хотя этот набор продуктов обеспечивает необходимое количество белков, жиров, углеводов и калорий, ежедневно питаться таким образом невозможно. Второй принцип обеспечивает человека полноценным трехразовым питанием. Но стоимость корзины возрастает примерно на четверть.

Потребительская корзина, представленная в табл.1, не описывает реальное потребление большинства граждан. Например, типовой москвич покупает значительно больше колбасы, сала, копченостей, чем включено в корзину, и в несколько раз меньше муки. Корзина табл.1 предназначена прежде всего для измерения инфляции (роста потребительских цен). Однако еще одно ее использование – оценка минимально допустимых расходов на продовольственные товары, обеспечивающих нормальную жизнедеятельность человеческого организма. Таковы расходы в ряде закрытых учреждений – больницах, тюрьмах, домах престарелых и др.

Для расчета индекса инфляции достаточно заполнить последний столбец значениями реальных цен, затем рассчитать расходы на приобретение товаров в базовый и текущий моменты времени, сложить эти расходы для получения стоимостей потребительских корзин и вычислить индекс инфляции. Процедура расчетов подробно показана в [1, 15, 22]. В частности, при изучении динамики цен следует фиксировать цены в одних и тех же торговых заведениях, чтобы избежать влияния разброса цен в пространстве. В качестве базового момента времени взят март 1991 г., поскольку рост цен в России начался со 2 апреля 1991 г., когда постановлением Кабинета министров СССР были в 2–3 раза подняты цены на основные потребительские товары. На базовый момент времени 14.03.1991 суммарная стоимость минимальной потребительской корзины продуктов питания в расчете на год составила 325 руб. 30 коп., а в расчете на месяц 27 руб. 11 коп.

Таблица 1 – Номенклатура, годовые нормы потребления и цены в базовый момент времени для потребительской корзины ИВСТЭ

Наименования продуктов питания	Годовые нормы, кг	Цены на 14.03.91, руб.	Текущие цены
<i>1. Хлеб и хлебопродукты</i>			
1.1 Мука пшеничная	18,5	0,46	
1.2 Рис	3,5	0,88	
1.3 Другие крупы (геркулес)	4,9	0,62	
1.4 Хлеб пшеничный	59,8	0,50	
1.5 Хлеб ржаной	65,3	0,20	
1.6 Макароны изделия	4,9	0,70	
<i>2. Картофель</i>	124,2	0,10	
<i>3. Овощи</i>			
3.1 Капуста	30,4	0,20	
3.2 Огурцы и помидоры	2,8	0,85	
3.3 Столовые корнеплоды	40,6	0,20	
3.4 Прочие (лук и др.)	27,9	0,50	
<i>4. Фрукты и ягоды</i>			
4.1 Яблоки свежие	15,1	1,50	
4.2 Яблоки сушеные	1,0	3,00	
<i>5. Сахар и кондитерские изделия</i>			
5.1 Сахар	19,0	0,90	
5.2 Конфеты	0,8	4,50	
5.3 Печенье и торты	1,2	1,40	
<i>6. Мясо и мясопродукты</i>			
6.1 Говядина	4,4	2,00	
6.2 Баранина	0,8	1,80	
6.3 Свинина	1,4	2,00	
6.4 Субпродукты (печень)	0,5	1,40	
6.5 Птица	16,1	2,40	
6.6 Сало	0,7	2,40	
6.7 Копчености	0,7	3,70	
<i>7. Рыба и рыбопродукты</i>			
7.1 Свежая (минтай)	10,9	0,37	
7.2 Сельди	0,8	1,40	
<i>8. Молоко и молочные продукты</i>			
8.1 Молоко, кефир, л	110,0	0,32	
8.2 Сметана, сливки	1,6	1,70	
8.3 Масло животное	2,5	3,60	
8.4 Творог	9,8	1,00	
8.5 Сыр и брынза	2,3	3,60	
<i>9. Яйца, шт.</i>	152,0	0,09	
<i>10. Масло растительное, маргарин</i>			
10.1 Масло растительное, л	3,8	1,80	
10.2 Маргарин	6,3	1,20	
11. Прочие (6% от стоимости товаров групп 1–10)			

4.7.2. Результаты измерения роста цен

В 1993–2016 гг. нами проводился независимый сбор цен. Получены стоимости потребительской корзины и индексы инфляции (табл.2). Обратите внимание, что при переходе от 37-й строки к 38-й учтено изменение масштаба цен в 1000 раз в результате деноминации, приведенной 1 января 1998 г.

Таблица 2 – Стоимости потребительской корзины ИВСТЭ
и индексы инфляции в 1993 - 2015 гг.

№	Дата снятия цен, <i>t</i>	Стоимость потребительской корзины <i>S(t)</i> (руб.)	Индекс инфляции <i>I(18.03.91; t)</i>
1	31.03.91	26,60	1,00
2	14.08.93	17 691,00	665,08
3	15.11.93	28 050,00	1054,51
4	14.03.94	40 883,00	1536,95
5	14.04.94	44 441,00	1670,71
6	28.04.94	47 778,00	1796,17
7	26.05.94	52 600,00	1977,44
8	08.09.94	58 614,00	2203,53
9	06.10.94	55 358,00	2081,13
10	10.11.94	72 867,00	2739,36
11	01.12.94	78 955,00	2968,23
12	29.12.94	97 897,00	3680,34
13	02.02.95	129 165,00	4855,83
14	02.03.95	151 375,00	5690,79
15	30.03.95	160 817,00	6045,75
16	27.04.95	159 780,00	6006,77
17	01.06.95	167 590,00	6300,38
18	29.06.95	170 721,00	6418,08
19	27.07.95	175 499,00	6597,71
20	31.08.95	173 676,00	6529,17
21	28.09.95	217 542,00	8178,27
22	26.10.95	243 479,00	9153,35
23	30.11.95	222 417,00	8361,54
24	28.12.95	265 716,00	9989,32
25	01.02.96	287 472,55	10 807,24
26	05.03.96	297 958,00	11 201,43
27	05.04.96	304 033,44	11 429,83
28	08.05.96	305 809,55	11 496,60
29	05.06.96	302 381,69	11 367,73
30	03.07.96	306 065,21	11 506,21
31	03.08.96	308 963,42	11 615,17
32	07.09.96	288 835,07	10 858,46
33	01.10.96	278 235,35	10 459,98
34	05.11.96	287 094,77	10 793,04
35	04.12.96	298 024,76	11 203,94
36	03.01.97	314 287,16	11 815,31
37	04.02.97	334 738,24	12 584,14
38	04.01.98	345,72	12,997
39	03.01.99	630,07	20,395
40	05.01.00	737,80	27,737
41	03.01.01	886,84	33,340
42	03.01.02	1051,79	39,541
43	03.01.03	1210,62	45,512
44	03.01.04	1355,91	50,974
45	14.05.04	1369,10	51,470
46	11.01.05	1463,98	55,037
47	10.01.06	1525,62	57,354
48	26.11.06	1571,26	59,070
49	10.01.07	1580,89	59,432
50	02.07.07	1644,38	61,819
51	03.01.08	1891,04	71,092
52	03.07.08	2286,54	85,960

53	10.01.09	2458,03	92,407
54	02.07.09	2682,08	100,83
55	11.01.10	2833,04	106,48
56	01.07.10	2954,86	111,07
57	11.01.11	3082,35	115,85
58	01.07.11	3267,29	122,80
59	11.01.12	3496,00	131,40
60	01.07.12	3740,72	141,91
61	11.01.13	4227,01	160,36
62	01.07.13	4507,77	170,79
63	11.01.14	4776,53	181,20
64	01.07.14	5015,35	190,26
65	11.01.15	5349,77	202,94
66	01.07.15	5681,46	215,52
67	11.01.16	6039,82	229,12
68	01.07.16	6291,41	238,28

Примечание. Стоимость потребительской корзины приводится без включения группы «прочие» (6% от стоимости основной части корзины).

4.7.3. Сравнение с данными официальной статистики

Сравним наши данные (табл.2) и данные Росстата. Сопоставление столбцов (4) и (5) в табл. 3 показывает, что официальная статистика занижала реальную инфляцию в 1,5 – 2,0 раза.

Таблица 3 – Индексы инфляции в РФ (по данным Росстата) в сравнении с независимыми данными ИВСТЭ (табл.2)

Год	Индекс инфляции	Индекс инфляции в %	Накопленная инфляция с января 1992	Накопленная инфляция с марта 1991	Данные ИВСТЭ к столбцу (4)
(1)		(2)	(3)	(4)	(5)
1991				2,6	-
1992	26,1	25100	26,1	67,86	-
1993	9,4	840	245,34	637,88	1235,42
1994	3,15	215	772,82	2009,33	3680,34
1995	2,31	131	1785,22	4641,56	9989,32
1996	1,218	21,8	2174,39	5653,42	11815,31
1997	1,11	11,0	2413,58	6274,30	12997
1998	1,844	84,4	4,451	11,572	20,395
1999	1,365	36,5	6,075	15,795	27,737
2000	1,202	20,2	7,303	18,986	33,340
2001	1,186	18,6	8,661	22,517	39,541
2002	1,151	15,1	9,968	25,917	45,512
2003	1,12	12,0	11,164	29,028	50,974
2004	1,117	11,7	12,471	32,424	55037
2005	1,109	10,9	13,830	35,958	57,354
2006	1,09	9,0	15,075	39,194	59,342
2007	1,119	11,9	17,830	46,358	85,960

Примечание. Накопленные индексы инфляции с 1998 г. даются с учетом деноминации.

Наша работа направлена на ликвидацию монополии Росстата при расчетах индексов инфляции, величин прожиточного минимума и реальных располагаемых денежных доходов населения. Как уже сказано, в 1990-е годы Генеральный штаб Министерства обороны Российской Федерации поручил нашему коллективу расчет индексов ин-

фляции по независимо собранной информации. Заказчика интересовали размеры финансирования оборонных НИР в реальных (сопоставимых) ценах. Полученные нами численные значения индексов инфляции оказались примерно в 2 раза выше, чем данные официальной статистики. На рубеже тысячелетий различия в оценке темпов роста цен сгладились, в частности, из-за того, что официальные статистические органы стали использовать потребительскую корзину, приведенную в табл.1. Однако в 2000-е годы Росстат разорвал связь между понятиями «индекс инфляции» и «индекс потребительских цен». В его публикациях стали называть индексом инфляции результаты расчетов с использованием весьма обширной потребительской корзины, не имеющей отношения к реальной жизни подавляющего большинства граждан России. Индексы потребительских цен, полученные по корзинам типа приведенной в табл.1, публикуются сравнительно редко и превышают примерно в 2 раза данные об индексах инфляции. Описанный лингвистический прием позволяет успокоить население, но мешает принимать адекватные управленческие решения.

Отслеживание динамики потребительских цен по независимо собранной информации с использованием описанной выше методики нетрудно. Трудозатраты для единичного сбора данных и расчета индекса инфляции – около 2 часов. Работы такого рода ведутся в ряде регионов. Так, Управление персонала Магнитогорского металлургического комбината ведет мониторинг прожиточного минимума сотрудников комбината. Волгоградская областная организация Общества защиты интересов вкладчиков банков и владельцев ценных бумаг отслеживает фактическую стоимость советского рубля, рассчитывая индекс инфляции. По нашему мнению, от единичных примеров целесообразно перейти к повсеместному мониторингу, органы управления территориальными и муниципальными образованиями должны получать достоверные данные о динамике цен и реальных располагаемых денежных доходах населения, рассчитанные по независимо собранной информации.

Основные полученные результаты на 2009 г. приведены в учебнике [22]. Исследования продолжаются (см., например, доклад [21] 2012 г.), в том числе мониторинг цен (см. статьи [23] 2013 г. и [24] 2014 г.).

Очевидно, значение индекса инфляции зависит от используемой потребительской корзины. Нами были сопоставлены результаты расчетов по пяти продовольственным корзинам [15, 22]. Оказалось, что результаты совпадают с точностью до 10%. Более заметный разброс – в 1,5 – 2 раза – определяется местом сбора цен в одном регионе. Динамика роста цен сравнительно мало отличается в разных регионах.

Таблица 4 – Среднемесячная заработная плата в РФ

№ п/п	Дата	Среднемесячная заработная плата в РФ (по данным Пенсионного фонда РФ, ноябрь 2004), руб.	Индекс инфляции $I(31.3.91;t)$, табл.2	Среднемесячная заработная плата в РФ, в % от уровня 1990
1	1990	303 (за 1990 г.)	1,00	100
2	Август 1993	65400	665,08	32,45
3	Декабрь 1994	354200	3680,34	32,08
4	Декабрь 1995	735500	9989,32	24,30
5	Декабрь 1996	1017000	11815,31	28,32
6	Декабрь 1997	760000	12997,00	19,30
7	Декабрь 1998	760,0	23,395	10,72
8	Декабрь 1999	1086,0	32,004	10,97
9	Декабрь 2000	1584,0	35,684	14,80
10	Декабрь 2001	1671,0	43,321	12,73

Для решения задач управления предприятиями и организациями целесообразно использовать потребительские корзины, составленные из товаров и услуг различных типов. В области управления персоналом – продовольственные корзины типа приведенной в табл.1. В других областях – составленные из закупаемых предприятием товаров и услуг производственного назначения.

При обсуждении пенсионных систем представляет интерес реальная величина заработной платы. Среднемесячная заработная плата в РФ (номинальная и в процентах от уровня 1990 г.) представлена в табл.4. Она составлена по данным Пенсионного фонда РФ, использующего сведения о средней заработной плате при расчета величин пенсий. Обратим внимание, что средняя заработная плата по данным Пенсионного фонда РФ отличается от таковой по данным Росстата.

4.7.4. Более подробный анализ: инфляция в XXI веке

Использование одной и той же потребительской корзины обеспечивает возможность сопоставления результатов расчетов за различные временные периоды. Этим работы ИВСТЭ выгодно отличаются от подхода официальной статистики. Как известно, Госкомстат РФ (ныне – Росстат) в 1993–2013 гг. из конъюнктурных соображений неоднократно менял состав потребительской корзины и объемы потребления входящих в нее товаров. Однако в начале XXI в. потребительская корзина официальной статистики мало отличалась от нашей. Здравый смысл восторжествовал - статистическое ведомство решило исходить из тех же разработок диетологов РАМН, на которые мы опирались еще в 1993 г.

Студенты факультета «Инженерный бизнес и менеджмент» МГТУ им. Н.Э. Баумана в соответствии программой изучения дисциплин «Организационно-экономическое моделирование» и «Эконометрика» два раза в месяц собирали данные о ценах и рассчитывали индексы инфляции. С целью сопоставимости результатов студент все шесть раз собирал данные на продукты одних и тех же конкретных наименований (марок, сортов) и в одних и тех же торговых организациях. По данным за пять моментов времени методом наименьших квадратов строился точечный и непараметрический интервальный прогноз на шестой момент времени, который затем сопоставлялся с реальностью.

Собранные данные позволили изучить разброс значений индекса инфляции в зависимости от конкретных мест сбора данных. Рассмотрим значения индекса инфляции $I(t_1, t_2)$ на текущий момент $t_2 = 14$ мая 2004 г., соответствующие базовому моменту $t_1 = 14$ марта 1991 г. (табл.5 - по Москве, табл.6 - по Московской области).

Статистические характеристики для двух выборок индексов инфляции, приведенных в табл.5 и 6, содержатся в табл.7. Они показывают, что индекс инфляции - это не число, а типичная нечисловая экономическая величина (см. главу 7 учебника [15]). Индекс инфляции в Москве можно описать интервалом [39,1; 62,43], а в Московской области - интервалом [44,02; 57,32].

Таблица 5 – Индексы инфляции в Москве в мае 2004 г.

39,10	40,50	40,56	40,70	41,56	41,73	44,03	47,18
47,18	47,30	48,40	49,27	51,45	52,67	52,70	53,04
54,60	55,00	55,01	55,33	55,62	56,40	57,15	57,29
57,65	57,72	57,80	58,26	58,40	59,59	62,43	

Таблица 6 – Индексы инфляции в Московской области в мае 2004 г.

44,02	48,11	50,40	51,02	51,08
54,12	54,12	55,65	57,32	

Таблица 7 – Статистический анализ данных об инфляции (май 2004 г.)

Статистические характеристики	Москва	Подмосковье
Минимум	39,1	44,02
Максимум	62,43	57,32
Объем выборки	31	9
Выборочное среднее арифметическое	51,47	51,76
Среднее квадратическое отклонение	6,80	4,08

Нечисловой характер индекса инфляции позволяет, тем не менее, сделать полезные для практического применения выводы:

1) В мае 2004 г. индекс инфляции близок к 50 (+20%), т.е. 50 руб. мая 2004 г. по покупательной способности соответствуют 1 руб. марта 1991 г.

2) Индексы инфляции в Москве и Московской области близки.

В мае 2004 г. в Москве стоимость минимальной продовольственной корзины оценивается как (27 руб. 11 коп.) \times 51,47 = 1400 руб., а прожиточный минимум - как 2800 руб. в месяц.

Индекс инфляции - это эконометрический инструмент, позволяющий доказательно обсуждать и решать те или иные экономические проблемы. Например, проблему соотношения зарплаты и прожиточного минимума. Средства массовой информации часто рассматривают эту тематику. К сожалению, не всегда обсуждение является доказательным, а выводы - обоснованными. Так, в статье [25] утверждается, что мы в конце 2003 г. «живем, как в 1985 году». Это не так.

Сравним уровни жизни в 1985 г. и в 2003 г. Поскольку цены на основные продовольственные товары до марта 1991 г. не росли, можно признать, что индекс инфляции с 1985 г. по конец 2003 г. совпадает с таковым с марта 1991 г. по конец 2003 г., т.е. согласно табл.2 с достаточной для расчетов точностью равен 50. В статье в «Комсомольской правде» приведены значения средней зарплаты по стране - 199 руб. в 1985 г. и 5722 руб. в конце 2003 г. Номинальная зарплата выросла в 29 раз, а цены - в 50 раз. Значит, реальная зарплата сократилась в 1,7 раза. В 1985 г. средняя зарплата почти в 4 раза превосходила прожиточный минимум, а в 2003 г. - лишь в 2 с небольшим раза. Следовательно, в 2004 г. среднестатистический гражданин РФ живет гораздо хуже, чем в 1985 г. Основную причину назвал Президент РФ В. В. Путин в Послании 2004 г. Федеральному Собранию РФ: валовой внутренний продукт (в сопоставимых ценах) в 2003 г. меньше, чем в 1989 г. (динамика макроэкономических показателей России анализируется нами в [26–28]). Большое значение имеет резко возросшая дифференциация доходов. Измеряющий ее децильный коэффициент увеличился за эти годы с 3 до 15; в развитых странах его значение - около 7.

Таблица 8 – Индексы инфляции в Москве в ноябре 2006 г.

46,14	46,44	48,26	49,21	49,27	49,71
50,29	51,05	51,07	52,52	53,64	53,75
53,83	54,36	54,68	55,07	57,16	57,83
57,83	58,7	59,11	59,12	60,41	60,41
60,53	60,57	63,81	65,9	68,01	72,15
72,15	72,15	72,15	72,23	73,3	83,61

Таблица 9 – Индексы инфляции в Московской области в ноябре 2006 г.

39,84	49,15	52,58	58,65
63,51	66,09	68,09	69,18

Крупное исследование было проведено через три с половиной года. Студенты собрали данные о ценах и рассчитали индексы инфляции в Москве и Московской обл. за период с $t_1 = 14$ марта 1991 г. до $t_2 = 26$ ноября 2006 г. (табл. 8 и 9).

В Москве индексы инфляции были рассчитаны по ценам в таких торговых организациях, как гипермаркет «Ашан», супермаркет «SPAR», гипермаркет «Метро», супермаркет «Перекресток», другие магазины, рынки. Проверка на однородность двух выборок – индексов инфляции в гипермаркете «Ашан» и индексов инфляции в «других магазинах», не входящих в сети – с помощью критерия Крамера-Уэлча [29] показала, что выборки однородны, а, следовательно, их можно объединить в одну.

Слушатели программы «Топ-менеджер» (Мастер делового администрирования / MBA) Академии народного хозяйства при Правительстве РФ по нашему предложению собрали данные о ценах и рассчитали индексы инфляции в ряде регионов РФ 2006 г. (табл.10).

Таблица 10 – Индексы инфляции по регионам России

№	Город, регион	Дата	Индекс
1	Владимир	22.02.07	44,5
		22.03.07	46,8
2	Иркутск	09.01.07	42,38
		09.02.07	42,97
3	Красноярск (1)	25.11.06	59,50
		30.01.07	61,77
4	Красноярск (2)	25.11.06	64,60
		08.02.07	66,86
5	Калужская обл., г. Малоярославец	20.12.06	46,86
		10.02.07	48,51
6	Нижний Новгород	10.11.06	43,16
		21.05.07	47,0
7	Новосибирск	01.03.07	51,79
		01.05.07	53,74
8	Петропавловск-Камчатский	10.11.06	28,94
		25.01.07	32,96
9	Ростов-на-Дону (1)	01.02.05	51,99
		01.02.07	67,67
10	Ростов-на-Дону (2)	01.02.07	39,66
		01.03.07	46,63
11	Ростов-на-Дону (3)	01.02.07	43,83
12	Татарстан, г. Бавлы	10.11.06	33,72
		25.01.07	36,05
13	Томск	25.12.06	49,86
		01.02.07	51,03
14	Тюменская обл., п. Боровский	Янв. 07	38,37
		Март 07	41,27
15	Череповец	01.11.06	48,1
		20.01.07	54,3

Примечание. Несколько исследований, проведенных в одном городе, указаны под разными порядковыми номерами. Следует иметь в виду, что стоимости потребительской корзины ИВСТЭ в марте 1991 г. для разных регионов различаются, иногда существенно.

Статистические характеристики для выборок индексов инфляции, приведенных в табл. 8 – 10, содержатся в табл. 11. Они показывают, что индекс инфляции имеет заметный разброс, это не число, а типичная нечисловая экономическая величина (см. [30] и [15, гл. 7 и 9]). В соответствии с приведенными данными индекс инфляции в Москве

можно описать интервалом [46,14; 83,61], в Московской области - интервалом [39,84; 69,18]. Статистическая обработка данных, приведенных в табл. 10, является не вполне обоснованной, поскольку регионы, в которых проводились исследования, не представляют собой представительную (репрезентативную) выборку из генеральной совокупности регионов России (см. главу 1 в [15]). Кроме того, различаются даты снятия информации о ценах. Поэтому для включения в посвященный РФ столбец отображена лишь часть данных. Тем не менее табл. 10 дает некоторое представление о динамике цен в регионах России. В частности, подтверждается высказанное ранее утверждение о том, что официальные статистические органы систематически занижают индексы инфляции: приведенное в табл.3 значение 39,194 Росстата меньше 24 из 29 индексов инфляции, замеренных слушателями Академии народного хозяйства.

Таблица 11 – Результаты статистической обработки данных об индексах инфляции $I(1990, 11.2006)$ в ноябре 2006 г.

Статистические характеристики	Москва	Подмосковье	РФ
Минимум	46,14	39,84	42,38
Максимум	83,61	69,18	64,6
Объем выборки	36	8	6
Выборочное среднее арифметическое	59,07	58,39	53,40
Среднее квадратическое отклонение	9,74	9,04	7,67

Судя по собранным О.Ю. Проскуриной данным, структура стоимости потребительской корзины (т.е. соотношение стоимостей групп продуктов) в среднем по Москве сравнительно мало изменилась с марта 1991 г. по ноябрь 2006 г. [31].

Нечисловой характер индекса инфляции позволяет, тем не менее, сделать полезные для практического применения выводы:

1) в ноябре 2006 г. индекс инфляции равен приблизительно 60, т.е. 60 руб. ноября 2006 г. по своей покупательной способности примерно соответствуют 1 руб. марта 1991 г.;

2) индексы инфляции в Москве и Московской области практически совпадают и достаточно близки к индексам инфляции по большинству других регионов РФ;

3) в ноябре 2006 г. в Москве стоимость минимальной продовольственной корзины оценивается как $(27,11 \text{ руб.}) \times 59,07 = 1601,4 \text{ руб.}$, а прожиточный минимум - как 3202,8 руб. в месяц (в соответствии с методом Оршански [15, 22] с коэффициентом Энгеля $C = 2,0$).

Отметим для сравнения, что по методике и данным Росстата (www.gks.ru), стоимость минимального набора продуктов питания в среднем по России в конце ноября 2006 г. составила 1443,6 рубля в расчете на месяц.

В 2007–2008гг. наблюдаем всплеск роста цен (табл.12–13). Табл. 12 получена слушателями программы «Топ-менеджер» (МБА) Бизнес-школы АНХ при Правительстве РФ, табл. 13 (Москва) – слушателями Бизнес-школы МВА МИРБИС, т.е. действующими менеджерами высшего звена организаций и предприятий различных регионов РФ и Москвы.

Анализ приведенных в табл.12–13 результатов измерений роста цен приводит к ряду интересных и практически полезных выводов. В частности, в Екатеринбурге цены за полтора месяца выросли на 67%, в Красноярске за два с половиной – на 32%, в Малоярославце за последний год – на 53%. Данные Росстата - 11,9% за 2007 год. Средний результат по табл.12 и 13 – 13,8 % за последние месяц-два. Средний рост цен с 1990 г. – в 81,69 раз, т.е. на один рубль можно было купить в 1990-м году столько же, сколько на 81 рубль 69 копеек в январе 2008 года. А год назад индекс инфляции был заметно

меньше - 59,07 (табл.11, Москва). Рост на 38,3%. В три с лишним раза больше, чем по данным Росстата.

Таблица 12 – Индексы инфляции в РФ на конец 2007 г. – начало 2008 г.

№	Регион	Дата t_1	$I(90,t_1)$	Дата t_2	$S(t_2)$, месяц	$I(90,t_2)$	$I(t_1,t_2)$
1	Якутск	01.11.07	76,067	01.01.08	2074,67	84,519	11,1%
2	Хабаровск	30.11.07	66,87	30.12.07	1759,21	68,79	2,9%
3	Петропавловск-Камчатский	09.12.07	93,70	10.01.08	2576,69	102,31	9%
4	Малоярославец	12.06	55,09	01.08	2153,49	84,21	53%
5	Красноярск	08.12.07	82,81	12.01.08	2486,29	97,58	18%
6	Тюмень	05.12.07	75,36	10.01.08	2297,28	82,94	10%
7	Красноярск	26.10.07	47,77	10.01.08	1594,42	63,17	32,2%
8	Нижевартовск	01.11.07	54,45	01.01.08	1949,92	62,513	14,8%
9	Екатеринбург	01.12.07	50,28	10.01.08	2142,82	84,10	67%
10	Рязань	01.11.07	63,28	01.01.08	1785,92	69,84	10%
11	Москва	26.11.07	88,80	26.12.07	2280,69	88,84	0,05%
12	Новосибирск	01.12.07	84,09	13.01.08	2167,30	84,75	0,78%
13	Самара	28.11.07	67,5	08.01.08	1762,17	68,9	2%

Таблица 13 – Индексы инфляции в Москве на 2007 - 2008 гг.

№	Дата t_1	$I(90, t_1)$	Дата t_2	$S(t_2)$,мес.	$I(90, t_2)$	$I(t_1, t_2)$
1	28.12.07	115,08	30.01.08	2992,42	117,01	1,68%
2	28.12.07	73,61	29.01.08	1944,66	76,04	3,30%
3	27.12.07	76,60	31.01.08	2034,95	79,57	3,88%
4	27.12.07	65,53	31.01.08	-	68,37	4,33%
5	27.12.07	79,25	31.01.08	2074,60	81,12	2,36%
6	28.12.07	115,492	29.01.08	3109,94	121,01	5,30%
7	01.01.08	75,38	01.02.08	2035,45	79,59	5,58%
8	16.01.08	92,73	31.01.08	2472,84	96,70	4,28%
9	01.01.08	98,96	01.02.08	2695,68	105,41	6,52%
10	10.12.07	66,99	10.01.08	1735,22	67,85	1,29%

Отметим, что расхождение результатов расчетов по независимо собранной под нашим руководством информации и данных официальной статистики частично объясняется тем, что в последние годы Росстат, как уже отмечалось, в очередной раз сменил потребительскую корзину. Это делает еще более неясной связь сообщаемых им численных значений инфляции с динамикой реальных экономических процессов (на эту неясность обращали внимание участники дискуссии, проведенной в ходе журналистского расследования [32]). Как следствие, констатируем, что каждое физическое и юридическое лицо может самостоятельно измерять рост цен с помощью методики, подробно описанной в настоящей статье. Таким путем целесообразно бороться с монополией Росстата на результаты измерения инфляции.

Отметим еще одну особенность последних лет. Коэффициент Энгеля $C = 2,0$ при оценке прожиточного минимума по методу Оршански [15, 22] как $CS(t)$, где $S(t)$ – стоимость потребительской корзины, был получен нами на основе бюджетного исследования середины 90-х. Им можно пользоваться лишь при условии постоянства структуры расходов. Однако в последние годы резко растет доля расходов на оплату жилищно-коммунальных услуг. Это означает, что доля расходов на продовольствие у всех семей и особенно у бедных семей заметно снижается. Следовательно, коэффициент Энгеля $C = 2,0$ должен быть повышен, по экспертной оценке, до 3,0 (в 2009 г.).

Ряд полезных материалов по проблемам роста цен, уровня жизни и использования данных Росстата содержится в [33–35].

4.7.5. Использование индексов инфляции при решении задач экономики и управления

Индексы инфляции целесообразно использовать при анализе проблем домохозяйств, организаций и предприятий, страны в целом. Достаточно подробное обсуждение этой тематики дано в [15, 18, 21, 22]. Здесь лишь приведем лишь краткие замечания.

В Указе Президента Российской Федерации от 21 августа 2012 года № 1199 «Об оценке эффективности деятельности органов исполнительной власти субъектов Российской Федерации» вместо множества критериев эффективности в управленческий обиход введены 11 интегральных показателей деятельности органов исполнительной власти субъектов Российской Федерации, в том числе такой показатель, как "реальные располагаемые денежные доходы населения". Для его расчета, очевидно, необходимо использовать индексы инфляции, рассчитанные по потребительским корзинам домохозяйств.

Предприятию при ценообразовании и заключении договоров с потребителями выпускаемой продукции необходимо учитывать рост цен на закупаемые сырье, материалы, полуфабрикаты, электроэнергию, горюче-смазочные материалы и т.п. При управлении персоналом важно отслеживать покупательную способность заработной платы (т.е. величину реальных располагаемых денежных доходов сотрудников предприятия).

Несколько подробнее о макроэкономических показателях. Валовой внутренний продукт (ВВП), валовой национальный продукт (ВНД) и другие характеристики экономического положения страны рассчитываются в текущих ценах. Для перехода к неизменным ценам надо поделить на индекс инфляции (т.е. умножить на дефлятор). В 2 раза занизишь индекс инфляции - в 2 раза завысишь валовой национальный продукт, валовой внутренний продукт, национальный доход и иные макроэкономические характеристики.

По данным Правительства РФ к концу 1998 г. валовой внутренний продукт составил 55,7% от уровня 1990 г. (динамика макроэкономических показателей России анализируется нами в [26–28]). Используя же коэффициент занижения инфляции со стороны Госкомстата РФ, равный 2, получаем более реальную цифру - 25% от уровня 1990 г. Падение в 4 раза! Эта оценка близка к выводам ряда специалистов, независимых от правительства.

Напомним, что номинальный ВВП исчисляется в текущих рыночных ценах. Чтобы определить реальный ВВП, необходимо выразить его в сопоставимых ценах базисного года. Для этого применяется так называемый дефлятор ВВП, т.е. индекс инфляции, который отражает изменение среднего уровня цен самой широкой группы товаров и услуг за определенный период, охватывающей все составляющие ВВП. Расчеты проводят с помощью т.н. «системы национальных счетов» [36].

Нет ничего удивительного в том, что официальный дефлятор ВВП отличается от индекса инфляции Росстата. Так, индекс-дефлятор ВВП за 2006 г. по отношению к ценам 2005 г. составил 15,4%, в то время как индекс инфляции Росстата за 2006 г. равен 9%. Разные корзины – разные результаты. Сопоставление этих численных значений, полученных разными официальными структурами, позволяет предположить, что и индекс инфляции должен составлять примерно 15,4%, т.е. Росстат занизил инфляцию за 2006 г. в $15,4/9,0 = 1,71$ раза. Однако более точные результаты организации, предприятия, государственные структуры и частные лица могут получить при независимом сборе и анализе данных о ценах по методике, описанной в настоящем разделе.

Международные сопоставления часто проводятся на основе паритетов покупательной способности национальных валют, методики расчетов которых основаны на использовании потребительских корзин.

Данные о росте цен с течением времени естественным образом устаревают. Однако описанная в настоящем разделе методика позволяет самостоятельно рассчитать индексы инфляции с целью их использования "в режиме текущего времени". Данная методика представляется весьма актуальной в современных (2014–2016 гг.) условиях ускорения роста цен.

4.8. Прогноз развития информационно-коммуникационных технологий

Прогнозирование научно-технического прогресса необходимо для принятия обоснованных управленческих решений [1]. В настоящем разделе прогнозируем развитие информационно-коммуникационных технологий с целью решения частного, но важного вопроса разработки профессиональных стандартов в ракетно-космической промышленности [2]. Как известно, профессиональный стандарт - это характеристика квалификации, необходимой работнику для осуществления определенного вида профессиональной деятельности. Профессиональные стандарты разрабатываются согласно статье 195.1 Трудового кодекса Российской Федерации, для применения:

- работодателями при формировании кадровой политики и в управлении персоналом, при организации обучения и аттестации работников, разработке должностных инструкций, тарификации работ, присвоении тарифных разрядов работникам и установлении систем оплаты труда с учетом особенностей организации производства, труда и управления;

- образовательными организациями профессионального образования при разработке профессиональных образовательных программ;

- при разработке в установленном порядке федеральных государственных образовательных стандартов профессионального образования [3].

Очевидно, содержание профессиональных стандартов должно отражать результаты прогнозирования научно-технического прогресса в соответствующей области, в данном случае - в ракетно-космической промышленности. Общеизвестно возрастание роли информационно-коммуникационных технологий в производстве и управлении. Поэтому в качестве примера прогнозирования научно-технического прогресса обсудим развитие информационно-коммуникационных технологий.

4.8.1. О прогнозировании: предварительные соображения

Начнем с констатации: прогнозы имеют тенденцию капитально подводить своих авторов. Вспомните, что предсказывали творцы литературы и техники полвека назад о нашем времени? Полеты на планеты Солнечной Системы и к звездам, освоение далеких планет, разумные роботы - и средства фиксации информации в виде микрофильмов, голосовых кассет и киноплёнки. Нередким было описание субсветового звездолета с чем-то вроде арифмометра в качестве устройства расчета курса.

И что мы имеем? Не добрались даже до Луны (пара прыжков и прогулок - если они действительно были [4] - при всем огромном уважении к астронавтам все же нельзя считать даже началом освоения небесного тела), зато в жароупорном и ударостойком брелке для ключей можем носить всю Ленинскую библиотеку, а портативное устройство связи в масштабах планеты без особых проблем может быть практически у каждого землянина. Забыты космос и океанские глубины - зато вполне можно поставить на стол устройство, дающее доступ к весьма большой части духовного богатства человечества

и любоваться живописью и кинематографом другого конца мира в любой момент и не выходя из дома. Практически все прогнозы полувековой давности не сбылись.

Для разработки профессионального стандарта специалиста ракетно-космической промышленности желательно знать, во что превратятся к 2020–2030 гг. персональные компьютеры? Что из себя будет представлять "работа на компьютере"?

Будущее определяется прошлым. Вспомним, что из себя представлял компьютер 33 года назад - в 1983-м году. Это был год разгара "холодной войны" - и стабильной и развитой экономики Союза, год первых стартов шаттлов - и активной работы советских космонавтов на орбитальных станциях. И год компьютеров - больших шкафов с электроникой и черно-белыми, изредка цветными дисплеями, использовавшихся в основном в народном хозяйстве, науке и военном деле. Интернет только начал свою победную поступь: то был год принятия протокола TCP/IP в качестве основного для Всемирной Сети и первой разработки системы DNS, а работа "в Интернете" представляла собой лишь передачу файлов и сообщений между компьютерами и запуск программ на удаленных машинах. Персональные компьютеры были тогда в СССР редкими и дорогими вещами: в основном их приобретали крупные научные центры и производственные учреждения, да еще военные. Мощность же процессоров тех компьютеров вполне сравнима с мощностью нынешних оптических мышей и наручных часов...

Однако обратите внимание - **общие принципы работы** с компьютером как с техническим устройством и через более чем 30 лет остались практически прежние! По-прежнему информация выводится в визуальном виде, причем большей частью в виде текста. По-прежнему ввод информации ведется с помощью набора кнопок с символами - т.е. путем работы рук. И даже общая логика программной части компьютера особенно не изменялась - все так же данные на жестких дисках хранятся в виде файлов, все так же они делятся на программные файлы и файлы данных. Резко же шагнули вперед такие характеристики компьютера, как мощность и удобство интерфейса. И не менее резко снизилась стоимость его компонентов.

Почему так произошло? И как эти тенденции экстраполировать на будущее? Для начала следует попробовать оценить, какие факторы влияют на развитие информационно-коммуникационных (компьютерных) технологий.

4.8.2. Факторы

3.1. Первый и самый главный фактор - это **высочайшая наукоемкость** компонентов компьютера и столь же **малая их ресурсоемкость**. Иными словами, чтобы разработать, скажем, микропроцессор и наладить его производство, требуются ресурсоемкие многолетние усилия - а вот само производство, будучи налаженным, является дешевым. В самом деле: сырье для производства процессора можно добыть в ближайшем песчаном карьере, а металла и нефти на производство жесткого диска потребуется куда как меньше, чем на выпуск руля для автомобиля. Человеческого труда на собственно производство тоже требуется очень мало: линии по производству процессоров, винчестеров, CD-ROM'ов автоматические, и, один раз наладив, можно эксплуатировать их очень долго.

Этот фактор приводит к двум тенденциям: первая - это возможность **постепенного удешевления** "железной" части компьютера по мере наладки производства. А вторая - это необходимость при дальнейшем усовершенствовании компьютерных информационно-коммуникационных технологий **базироваться на предыдущих наработках**. Ибо кардинальное изменение даже не принципов, а технологии производства потребует весьма значительных научных усилий и финансовых затрат.

Наиболее четко выраженным этот фактор является в производстве программного обеспечения и вообще информационных продуктов. Для создания новой копии программы, песни или фильма не нужно даже автоматической линии - достаточно не самого современного компьютера. А вот первоначальное производство программного продукта (пакета программ) требует и времени, и сил, и денег. Вместе с тем в настоящее время никто не пишет программы в машинных кодах - для создания одних программ используются другие программы: специальные пакеты разработки.

3.2. Второй фактор - это **капиталистический** характер мировой экономики. Если сформулировать вкратце и не вполне точно, но в виде, пригодном для обсуждения в данном месте монографии, то "капитализм" - это принцип деятельности, при котором основной целью является достижение **максимальной прибыли** (в отличие, скажем, от "социализма", при котором основная цель - удовлетворение потребностей людей, пусть даже и за счет меньших прибылей). С рассматриваемой точки зрения к "социализму" относится, например, солидарная информационная экономика [5, 6].

Этот фактор в какой-то степени противодействует первому фактору, не допуская снижения цен на "железо" до себестоимости - ибо максимальная прибыль достигается при определенной "оптимальной" цене, ниже которой снижать цену уже невыгодно даже при практически нулевой стоимости производства. Особенно сильно это противодействие проявляется в индустрии программного обеспечения: снабдить всех пользователей компьютеров по всему миру копией какой-либо программы труда не составит, однако такое снабжение снижает прибыли компании-производителя.

3.3. Третий фактор, тоже экономический - это **конкурентность** мировой экономики. Фирмы и предприятия мира борются за наивысшую прибыльность своей работы путем расширения рынка и снижения издержек производства.

Этот фактор приводит к **увеличению разнообразия товаров при сохранении их принципиального устройства**. В самом деле: значительно дешевле не придумывать какой-то совершенно новый (инновационный) товар, при налаживании производства которого вполне возможны весьма большие трудности и возникает необходимость крупных денежных вложений, а усовершенствовать старый, снабдив его новыми дополнительными качествами - например, улучшенным дизайном. К примеру, телевизоры различной стоимости в каждой из фирм-производителей обычно собираются на одной и той же электронике (производятся с помощью одного и того же технологического процесса), и различия между моделями нередко заключаются лишь в том, что у более дешевых моделей отключены уже имеющиеся функции. Эксперимент может провести каждый - с помощью определенных комбинаций клавиш пульта дистанционного управления зачастую можно включить прием телетекста даже на том телевизоре, в котором, согласно инструкции, такая функция не предусмотрена.

Кроме того, обратите внимание, что в последние 20–30 лет не было сделано практически никаких действительно революционных изменений компьютерного мира: шло лишь улучшение и удешевление уже созданных технологий. Те же модемы, CD-ROM'ы, жидкокристаллические мониторы, мыши, Flash-память, жесткие диски и сами процессоры были придуманы еще на заре развития персональных компьютеров, и с тех пор принципиально не изменились, хотя, бесспорно, достигли куда как большего совершенства и мощи, чем были изначально. Это не случайно: пока есть спрос, **безопаснее совершенствовать старое, чем создавать принципиально новое** - такое создание весьма рискованно и затратно.

Вместе с тем удачное принципиальное решение может дать такой экономический эффект, который превысит все ожидания и даст огромное преимущество в конкуренции. Поэтому поиск новых, революционных технологий продолжается, но им занима-

ются либо крупные корпорации, могущие выделить средства на рискованные проекты, либо мелкие фирмы, для которых такой поиск - единственный шанс выйти на рынок.

3.4. Ну и, наконец, фактор, который присутствовал у человека с самого первого дня его появления независимо от политических, экономических и иных перипетий и, будем надеяться, останется с ним всегда. Это - жизненная сила человечества: **стремление человека к творчеству** и преобразованию мира, к созданию нового и к развитию. Сотни тысяч инженеров и изобретателей работают во множестве фирм по всему миру, и многие из них делают это не ради лишь получения денег, а и с целью сделать мир лучше - независимо от той прибыли, которую такая деятельность даст.

3.5. Есть еще пятый фактор, который, впрочем, сильно отличается по своей сути от первых четырех. Это - **наличие мощных каналов связи**, Интернета. Причем именно "наличие" - Интернет, как известно, первоначально создавался не коммерческими, а военными и научными организациями, большей частью государственными, для решения своих специальных задач. Торговая деятельность в Интернете началась лишь через 25 лет после появления этой сети. Имели ли бы мы сейчас Сеть Сетей, если бы не "холодная война" и не ученые - неизвестно; но, во всяком случае, она у нас есть, и сбрасывать со счетов ее наличие нельзя.

4.8.3. Взгляд в целом на развитие информационно-коммуникационных технологий

Однако вместо выделения факторов, влияющих на развитие информационных технологий, можно взглянуть на вопрос и с иной точки зрения - оценить развитие компьютеров как технической системы в целом. Согласно так называемой "Теории решения изобретательских задач" (см., например, [7], а также [8–10]), любая техническая система проходит в своем развитии четыре этапа:

1. Разработка основных принципов работы и компонентов системы.
2. Усовершенствование этих принципов и компонентов.
3. Приобретение системой динамичности и гибкости.
4. Замена системы на иную, выполняющую ту же функцию, но более совершенную.

Эти этапы хорошо прослеживаются практически у всех технических систем. Взять, скажем, полиграфию - изначально был придуман сам принцип: получение текста на бумаге путем нанесения оттиска краской, придуман основной компонент системы - типографский набор. Впоследствии печатный станок Ивана Федорова и Гуттенберга превратился в мощные роталитные машины, действующие по прежнему принципу: набор текста (но уже тоже на специальных наборных машинах) и нанесение оттиска. Потом пришла пора динамизации системы: громоздкие машины и свинцовые литеры были заменены лазерными барабанами и компьютерными комплексами, в связи с чем, издание одной книги из трудоемкого процесса создания набора свелось к направлению файла на печать нажатием пары кнопок. И вот сейчас уже появляются признаки грядущего перехода к новой системе представления текстовой информации - не на бумаге путем нанесения оттиска краски, а на электронных дисплеях.

Путь развития компьютерных систем полностью соответствует указанной схеме: как уже упоминалось, основные узлы современного компьютера были созданы на самой заре его развития и с тех пор прошли долгий путь усовершенствований. Можно сказать, что в настоящее время идет третий этап развития - динамизация системы: продажи мобильных компьютеров (ноутбуков, смартфонов и т.п.) растут и уже заметно превышают продажи традиционных стационарных компьютеров. Когда-нибудь настанет время и четвертого этапа: но вот что из себя он будет представлять - вопрос... Может быть, система передачи информации сразу в сознание?

4.8.4. Тенденции

Учитывая упомянутые факторы, постараемся выявить тенденции развития информационно-коммуникационных компьютерных технологий на ближайшие пару десятилетий.

4.8.4.1. ЖЕЛЕЗО (HARDWARE)

4.8.4.1.1. Первая и самая главная тенденция - это максимальное удешевление производства компьютерных (и сетевых) комплектующих, сочетающееся с ростом их мощности. В настоящее время разработаны и отлажены технологии производства всех компонентов компьютера, а конкуренция на этом рынке достаточно велика, чтобы приводить к стремлению цен к себестоимости, которая, в свою очередь, также медленно снижается. Подтверждение может найти каждый, просто взяв прайс-лист какой-нибудь "компьютерной" фирмы пятилетней давности и сегодняшний: вы увидите, что "современный" процессор пять лет назад можно было купить долларов за 200, а сейчас - менее чем за 100, и это при условии, что нынешний "современный" кристалл раз в 50 мощнее тогдашнего.

Удешевление компьютерных комплектующих приведет к их широчайшему распространению. С ними произойдет то же самое, что и с другими достижениями цивилизации: синтетическая футболка требует для своего производства нефтяные скважины, трубопроводы, нефтеперегонные заводы - а носят ее даже папуасы на островах Тихого океана: все потому, что производство налажено и дешево. Компьютеры или приборы с чипами появятся даже в самых бедных странах. Кто знает, может быть, на среднюю месячную зарплату в России через 20 лет можно будет купить 30 батонов хлеба или десяток компьютеров средней руки - как сейчас произошло с калькуляторами, 30 лет назад стоившими как раз этот самый месячный заработок.

Бесспорно, стремление производителей к максимизации прибыли затормозит этот процесс, но не остановит его - как не остановило до сегодняшнего дня.

4.8.4.1.2. Одним из способов удешевления производства является "**централизация**" - объединение нескольких компонентов в один. Например, 40 лет назад тот компонент компьютера, который мы сейчас называем "микропроцессором", состоял из нескольких микросхем - всевозможные кэши и контроллеры были отдельными чипами. А то, что мы называем "чипом" и считаем основной характеристикой материнской платы, ранее вообще было рассредоточено по плате: всевозможные контроллеры винчестера, дисководы и т.д. можно было по отдельности заменять и перепаявать. Но - оказалось дешевле и проще все компоненты уместить в одном кремниевом кристалле.

Этот процесс наверняка продолжится и в будущем. Технологический директор компании Sun, Грег Папандопулос, даже предсказывал, что через десяток лет в компьютере вообще останется лишь один чип, который будет включать в себя все компоненты компьютера, кроме разве что дисплея, клавиатуры да накопителей (см. об этом в [11]). И он явно недалек от истины: уже сейчас производятся материнские платы, в которых и видеосистема, и звуковая карта, и даже все сетевые контроллеры размещены в одном чипе.

Но и в "централизованном" компьютере обязательно сохранится **модульность**. В самом деле: пользователи всегда будут желать самостоятельно улучшать показатели своего компьютера, приобретая к нему дополнительные компоненты. Собственно, именно модульность устройства IBM PC и привела к тому, что персональный компьютер из дорогой игрушки превратился в массовое устройство: выпускать его компоненты взялись все, кому не лень, и за счет этого достигли такого удешевления производства,

что покупка компьютера стала доступна индивидуальным пользователям и небольшим компаниям. Поэтому накопители информации, устройства ввода и вывода останутся отдельными устройствами, а вот содержимое всего остального переключается в один чип, как это уже произошло со звуковой подсистемой.

4.8.4.1.3. Третья тенденция, в отличие от первых двух, видна не сразу. Это стремление к уменьшению размеров компьютеров - к тому, что принято называть **миниатюризацией**. Вы, наверняка, сразу же возразите, что ваш компьютер несколько не уменьшился в размерах по сравнению с теми, которые продавались десятком лет назад, и будете правы. Однако все больше пользователей приобретают не стационарные компьютеры, а ноутбуки (в Японии, например, ноутбуки составляют половину от всех продаваемых компьютеров), да и, положив руку на сердце, скажите: какой компьютер вы бы приобрели при равной стоимости - стационарный или ноутбук? Что ни говори, а ноутбуки, смартфоны, планшеты и другие аналогичные устройства растут в популярности одновременно со снижением цен на них, ибо они куда как удобнее большинству пользователей, чем серые ящики, занимающие полстола. И вышеупомянутая централизация тому в помощь: один чип всегда будет меньше комплекса из нескольких...

В настоящее время весьма многообещающей разработкой можно назвать "*воздушный дисплей*" - "Heliodisplay": с помощью оригинального технического решения изображение создается прямо в воздухе над проектором, причем это изображение может воспринимать воздействие извне - является еще и сенсорным экраном. На создание и отладку Heliodisplay были потрачены годы труда и миллионы долларов, однако же при массовом налаженном производстве такой дисплей вполне может оказаться дешевле традиционных, как это уже постепенно происходит с жидкокристаллическими дисплеями. Вместе с тем в плане миниатюризации (а значит, кстати, и определенного удешевления в производстве - материалов-то потребуется меньше) Heliodisplay более перспективен, чем традиционные "поверхностные" (пусть даже и гибкие, и ударопрочные), так как размер отображаемого им в воздухе экрана значительно больше его собственных размеров. Станет ли эта технология массовой или останется в ряду курьезов вследствие нежелания производителей менять налаженные технологии - покажет время.

4.8.4.1.4. Вместе с тем **не следует ожидать** внедрения **новых методов работы** - во всяком случае, в ближайшие 20 лет. Разрекламированные фантастами системы общения с компьютером на "естественном" языке, особенно ярко показанные в легендарном фильме Стенли Кубрика, так и останутся фантастикой. Причем даже не потому, что их сложно реализовать - при достаточном быстродействии и объеме памяти можно создать такое программное обеспечение, что компьютер вполне может выполнить тест Тьюринга (то есть разговор с ним будет нельзя отличить от общения с реальным человеком), а потому, что это не нужно пользователям, а, значит, не будет воплощаться и производителями.

В 2014 г. компьютерная программа разработчиков из России и Украины впервые в истории прошла тест Тьюринга, который выявляет искусственный интеллект. Чат-бот Eugene Goostman («Евгений Густман»), имитирующий 13-летнего подростка из Одессы, "обманул" жюри конкурса Turing Test - 2014, которое приняло его за человека. Программу начали разрабатывать в 2001 г. в Санкт-Петербурге Владимир Веселов из России и Евгений Демченко с Украины (Веселов в данный момент живет в США, а Демченко - в России). Английский математик Алан Тьюринг (1912–1954) предложил идею теста, который дает ответ на вопрос, может ли компьютер мыслить. Во время теста человек общается с виртуальным собеседником. По итогам беседы он должен решить, с кем разговаривал - с человеком или машиной. Тест пройден в том случае, если минимум 30% респондентов утверждают, что они общались с человеком. Задача компьютерной программы - заставить собеседника сделать неверный выбор. 7 июня программа

«Евгений Густман» впервые прошла тест Тьюринга за его 65-летнюю историю, сообщается на сайте Университета Ридинга, Великобритания (The University of Reading). В конкурсе Turing Test - 2014, который проводился при поддержке Лондонского королевского общества, Школы системного инжиниринга (University's School of Systems Engineering) и проекта RoboLaw (финансируется ЕС) приняли участие пять суперкомпьютеров. 33% жюри конкурса приняли программу «Евгений Густман» за человека. В 2012 г. программа смогла убедить 29% судей в том, что является человеком. Идея создания персонажа программы «Евгений Густман» заключалась в том, что 13-летний подросток может не знать некоторых вещей, но в то же время способен отвечать на вопросы собеседника, цитирует Владимира Веселова Университет Рединга. Команда потратила много времени, чтобы персонаж получился правдоподобным, отметил Веселов. В 2014 г. был улучшен диалоговый контроллер, который делает реплики программы похожими на человеческие, добавил ученый. По словам Веселова, он планирует и дальше работать над программой, чтобы сделать ее «умнее» [12].

По нашему мнению, компьютер - это средство работы с информацией, а не друг и собеседник (впрочем, есть любители играть в шахматы с компьютером), а с информацией проще работать привычными методами - набором символов с помощью клавиатуры с компьютерной мышью и восприятием текста с дисплея. В самом деле: куда как проще крутить колесико мыши, чем заставлять компьютер читать вслух страницу или чем даже просто отдавать команды вроде "двигай вниз" (ну, а собеседников не так и трудно найти через Интернет). Да и у опытных пользователей скорость набора текста нередко превышает скорость речи. Голосовое управление может стать лишь вспомогательным, предназначенным максимум для надиктовки длинных текстов, но никак не преваляющим: как ни крути, а мышью двигать быстрее, чем языком ворочать. Не говоря уже про проблемы точности распознавания речи и передачи ее в соответствии с правилами правописания.

Так что клавиши, указатель и текст еще долго будут нашими спутниками, хотя и могут стать практически неузнаваемыми в процессе совершенствования: про Heliodisplay уже было упомянуто, а перспективные разработки клавиатуры и мыши представляют собой проецирующиеся на любую ровную поверхность изображения клавиш и перстень с датчиком положения в пространстве, воспринимающий движения руки.

По размерам будущий компьютер может представлять собой прибор размером с карандаш, булавку или пуговицу, поскольку системный блок имеет пренебрежимо малые размеры, клавиатура и дисплей будут виртуальными, передача любых объем информации осуществляется через виртуальный офис в Интернете.

4.8.4.2. Софт (SOFTWARE) и ИНТЕРНЕТ

Тенденции развития мира программного обеспечения - это тема отдельного исследования. Слишком много факторов необходимо учитывать, причем многие из них отличаются от тех, что влияли на развитие компьютерного "железа". Но перечислить наиболее существенные из них целесообразно.

4.8.4.2.1. Усиление средств защиты от копирования

Несмотря на то, что возможности информационных технологий позволяют снабдить каждого пользователя компьютеров необходимыми ему программами фактически бесплатно, вследствие капиталистического характера мировой экономики производители софта не могут этого допустить и, вкупе с совершенствованием программ, снабжают их и системами защиты от свободного копирования.

А системы эти могут быть весьма эффективны. Особенно если они функционируют с использованием Интернет-технологий, проверяя каждую копию программы на каком-либо специально созданном для этого сервере. Заодно такая система сможет и сообщать разработчикам программы о всех замеченных случаях копирования и тем более попыток взлома. Бесспорно, взломать можно все, что угодно - но вот насколько это будет трудоемко...

То, что в настоящее время многие фирмы не используют защиту от копирования, может быть объяснено, например, их стремлением закрепиться на рынке. Не секрет, что успех фирмы Microsoft в России (ставшей практически "народной" компанией, продукты которой известны и используются всеми пользователями компьютера за очень редким исключением) обусловлен именно деятельностью пиратов, снабдивших всех желающих копиями программ Microsoft практически бесплатно. Однако уже в настоящее время в Microsoft'овские разработки встраиваются не такие уж и простые методы предотвращения копирования, а в будущем следует ждать их совершенствования.

Указанная тенденция приводит к двум другим:

4.8.4.2.2. Развитие мира свободнораспространяемого программного обеспечения

Стремление к разуму и свободе присуще немалой части людей, поэтому мир свободного и бесплатного распространения информации останется всегда. И тому подтверждение - история операционной системы Linux, по удобству интерфейса в настоящее время уже догоняющей Microsoft Windows. Лишь тот факт, что большинство копий Windows в мире сделано без оглядки на всякие лицензии, приводит к тому, что Linux'оидов значительно меньше, чем любителей продукции Microsoft.

Если же Microsoft усилит свою "борьбу с пиратством" и начнет делать попытки вытрясти немалые деньги по крайней мере с большинства пользователей Windows, то наверняка найдутся компании или группы энтузиастов, которые дополнят Linux Windows-подобным интерфейсом (собственно, это уже во многом сделано) и полной поддержкой FAT, после чего Linux ожидает триумфальное шествие по компьютерам всего мира. Открытый стандарт Linux и возможность всем желающим дорабатывать эту ОС приведет к тому же, к чему привело открытие стандарта IBM на заре развития персональных компьютеров: за доработку Linux возьмутся множество фирм и частных программистов, что наверняка приведет к появлению качественных дистрибутивов по цене чуть выше стоимости носителя, а то и вообще бесплатных.

4.8.4.2.3. Развитие технологии "аренды программ" через Интернет

Другим последствием совершенствования систем защиты от копирования станет развитие принципа оплаты программ "по действию", а не "за приобретение". Иными словами, производители станут предлагать пользователям покупать не программы, а право работы с ними или выполняемые ими функции: например, плата будет браться за распознавание одной отсканированной страницы или за час набора текста. В самом деле: если пользователь работает с программой распознавания максимум раз в месяц, то "коробочную" версию за пару сотен долларов он вряд ли купит, а вот потратить пару центов в месяц он вполне может себе позволить. Разумеется, такая работа должна быть организована с использованием средств Интернета: основная часть программы вместе с блоком учета оплаты будет располагаться на удаленном сервере, а небольшая клиентская часть - на компьютере пользователя.

Вполне возможно, что и продажи самих компьютеров будут напоминать современное распространение сотовых телефонов: к компьютеру может прилагаться контракт на использование набора программ, и при покупке такого контракта на компьютер будет даваться скидка. Наверняка появятся и "заблокированные" компьютеры (то

есть не допускающие установки других программ, кроме имеющихся, особенно бесплатных), и услуги снятия такой блокировки. Вместе с тем наличие альтернативного мира свободного программного обеспечения не допустит сильного взлета цен на аренду программ.

4.8.4.2.4. Увеличение надежности и ресурсоемкости программ

С ресурсоемкостью - понятно из сказанного выше, а с надежностью как? И почему эти две характеристики перечислены вместе? Потому, что программы будущего наверняка будут создаваться на основе уже проработанных и протестированных блоков кода. Как, например, сделано и сейчас в средствах разработки вроде Visual Basic. Принцип объектно-ориентированного программирования позволяет один раз сделать блок, реализующий ту или иную функцию, отладить его, а потом использовать везде, где нужно, просто вызывая его из программы. Создав набор таких блоков и тщательно их оттестировав, можно свести вероятность технического сбоя в программах из них к минимуму - останется лишь опасность неправильного составления алгоритма самой программы, что, ясное дело, лежит лишь на совести программиста. Однако, так как в блоки включаются как нужные, так и не нужные в данной программе функции, то такой блок для своего выполнения будет требовать и больше ресурсов, нежели код, написанный непосредственно для решения конкретной задачи.

Впрочем, для ответа на вопрос "Будет ли софт будущего содержать больше ошибок и "багов"?" можно и просто посмотреть назад. Думается, никто не будет спорить, что Windows XP, будучи куда как сложнее Windows 95, еще и куда как более стабильна в работе. В самом деле: Windows XP очень редко "падает" от некорректно написанных программ, добавление новых устройств в компьютер обходится без переустановки системы. Бесспорно, свои "баги" есть и у этой ОС, однако их значимость для пользователя многократно ниже, чем ошибки Windows 95/98.

4.8.4.2.5. Сохранение общих принципов интерфейса

Как будет выглядеть программное обеспечение будущего? Вариантов, наверное, будет много. Проще сказать, чего не будет: не будет приспособления интерфейса к реальности. Старый добрый фильм о "виртуальной реальности", где доступ к файлам представлялся как выдвигание ящиков из сейфа, на которое надо затрачивать мышечные усилия, останется сказкой - ведь, в самом деле, куда как проще привыкнуть к пиктограммам и курсору, чем воспроизводить законы физики там, где они, в общем-то, только усложнят жизнь. Не будет "искусственного интеллекта" и общения с компьютером на языке образов: Впрочем, о невозможности создания искусственного интеллекта подробно сказано в статьях [13, 14]. Однозначно останется текстово-рисуночное представление данных, разве что с добавками в виде звуковых и видео-комментариев, хотя использование компьютерной техники как хранилища мультимедиа-содержимого, кинофильмов и музыки, станет едва ли не основным. Как уже говорилось, не будет системы общения с компьютером на "естественном" языке, разве что за исключением устройств для инвалидов.

Вместе с тем, привычные нам представления информации могут достичь наивысшей точки своего развития. Например, в электронной книге, посвященной кинорежиссеру, можно будет найти обсуждаемые фрагменты из его фильмов, вставленные в текст так же, как сейчас вставляются картинки. Текст художественной книги может сопровождаться музыкой, меняющейся по мере его прочтения, а также видеотрекками. Обзор ресурсов Интернета может представлять собой не набор ссылок с их описанием, как сейчас, а текст с включенными в него первыми страницами сайтов, при наведении

курсора раскрывающимися на весь экран. Но это все останется на откуп самим авторам информации: самые же общие тенденции пребудут неизменными.

4.8.4.3. ДРУГИЕ ОТРАСЛИ

Вот где нас могут ожидать действительно революционные преобразования - так это в производственной (станки, датчики) и бытовой технике. Кратко обсудим будущее на примере бытовой техники.

Удешевление производства чипов и систем беспроводной связи приведет к тому, что встроить процессор и беспроводной адаптер можно будет во все бытовые приборы без особого удорожания последних. Это весьма широко расширит возможности создателей домашнего оборудования: как вам, например, выключатели, отслеживающие наличие в комнате человека и автоматически включающие и выключающие освещение? А они уже сейчас выпускаются, хоть и значительно дороже обычных.

Но это пока. Уже сейчас есть фирмы, предлагающие создать "Умный дом" - объединить различные устройства квартиры или дома в единую систему, позволяющую, например, с одного пульта (или с мобильного телефона) открывать и закрывать окна и шторы в комнатах дома. Или даже настроить автоматическое закрытие окон в случае начала дождя. Причем цены на все это удовольствие отнюдь не заоблачные, а компоненты системы обладают изрядной долей автономности: отказ центрального процессора квартиры приведет разве что к невозможности управлять окнами и дверями, сидя на диване перед пультом, а отнюдь не к вынужденному затворничеству жителей.

Будет ли сеть дома иметь подключение к Интернету? Вполне возможно, и по простой прозаической причине - за возможность издалека контролировать события в квартире готовы платить очень и очень многие пользователи (особенно те, у кого есть дети). Однако такая возможность все же появится лишь после разработки особых защищенных программ и протоколов, не допускающих возможности доступа к управлению домом нежелательных лиц. Нет, бесспорно, взлом будет возможен всегда - только вот усилий на него потребуются не меньше, чем на физический взлом двери в квартиру.

Одно время весьма популярная идея внедрения компьютера в одежду реальностью вряд ли станет: в доме среднего американца можно найти до трех тысяч одних только костюмов и рубашек - как вы думаете, зачем в них всех компьютеры? Впрочем, полярники, спасатели, летчики наверняка будут использовать "компьютеризированные" куртки или шлемы. Вполне возможны и "извращения" капиталистической системы: в качестве курьеза через 20 лет будет демонстрироваться пакет с соком, демонстрирующий на дисплее количество оставшегося сока и его температуру вкупе с питательным составом в виде красивого коллажа. А что - вставляла же компания Logitech в упаковку своих оптических мышей ради рекламы и привлечения покупателей мигающий диод, лет 30 назад стоивший десятки долларов - даже несмотря на то, что упаковка после покупки будет выброшена...

Перейдет ли в Интернет система телевидения и радио? Возможно, но лишь частично, и - главное! - лишь для удовлетворения пользователей Сети, а не в качестве "нового носителя информации" вместо радиоволн. Теле- и радиопередачи мы все так же будем ловить на УКВ/АМ/FM, и лишь энтузиасты станут смотреть их на своих компьютерах через Интернет. Ибо отказываться от налаженной системы телевидения (и от огромных экранов), переводить ее на совершенно иной принцип лишь ради некоей "продвинутости" никому не надо (особенно самим производителям электроники).

Разумеется, будут развиваться и компьютеры, используемые в промышленности, науке, медицине. Но это - тема уже отдельного исследования, ибо там несколько иные закономерности. Хотя автоматические дозаторы лекарственных средств, имплантируемые в организм и, например, вводящие в кровь инсулин в зависимости от уровня глю-

козы в ней, пойдут в массовое производство наверняка: слишком высока потребность в таких устройствах.

4.8.5. Прогноз - итоги

Итак, подведем итоги в заключительной части раздела. Вот перенеслись мы на 20 лет вперед и зашли в "компьютерный" магазин. Что мы там увидим? Спрогнозируем на основе всего вышеизложенного [15].

Продавец с улыбкой преподнесет нам небольшой цилиндрок величиной с хорошую гаванскую сигару. Видя наше недоумение, он поставит его на стол и нажмет неприметную кнопку. И - над цилиндрком в воздухе появится удивительно четкое изображение с логотипом операционной системы, которое можно будет увеличивать или уменьшать по своему вкусу. А на столе - окажется изображение клавиатуры, спроецированное лазерным сенсором внутри цилиндрика, причем на ней вполне можно будет набирать текст. К услугам пользователя будет и система распознавания речи, которая, впрочем, не станет главенствующей: традиционный набор текста клавишами после небольшого опыта набора ввода текстов становится куда как быстрее, чем речь. Разумеется, компьютер будет снабжен и парой портов для соединения с более старыми моделями, а в магазине для любителей наверняка будет иметься нечто вроде переходника, с помощью которого можно будет присоединить к современным компьютерам старые жесткие диски и принтеры-сканеры из музея техники.

Впрочем, для полной победы "виртуальных" дисплеев и клавиатур 20 лет все же может быть малым сроком. Поэтому не исключено, что и те, и другие все же останутся на материальных носителях, но носители эти будут представлять собой гибкие и прочные листы, которые нетрудно будет складывать и скатывать.

Компьютер сохранит модульность, но это будут такие модули, которые уже объединить будет никак нельзя. В нашей фантазии, например, "сигара" будет являться столбиком "таблеток" - главной, включающей в себя систему расчетов, блока-проектора экрана и блока-проектора клавиатуры, и - наверняка - из пары-тройки многотерабайтных накопителей на основе энергонезависимой памяти. При желании хранить в компьютере больше информации нетрудно будет добавить в столбик еще "таблеток" - накопителей, а при необходимости расширить его возможности - специальные модули: например, электрокардиограф или звукорежиссерскую систему.

Привычные системы ввода-вывода информации - сканирование и печать - сохранятся: от бумаги человечество, бесспорно, откажется (как отказалось от папируса, пергамента), но явно не через 20 лет. Однако их техническое устройство придет почти что к совершенству: возможно, сканер будет из себя представлять подобие цифрового фотоаппарата, которым достаточно сфотографировать объект, чтобы получить изображение. Наверняка на рынке появятся системы с автоматическим переворачиванием страниц - будет достаточно сунуть в приемник книгу, чтобы она была оцифрована. Скорее всего, к 30-м годам 21 века практически все бумажные носители информации, наработанные человечеством, будут оцифрованы. Принтер же, скорее всего, сохранит свою принципиальную схему (бумагопротяжный механизм + печатающее устройство), однако наверняка появятся и портативные принтеры (наверное, напоминающие ползающие по бумаге спичечные коробки).

Общение между компьютерами и устройствами будет происходить по беспроводным технологиям - не в последнюю очередь из-за относительной дороговизны кабелей. Доступ в Интернет, скорее всего, останется лишь по системе сотовой связи - при налаженном производстве базовых станций развивать сотовую сеть куда как выгоднее, чем

проводную, из-за большой трудо- и ресурсоемкости последней. Разве что междугороднее и международное сообщение останется на основе оптоволоконных кабелей.

Отдав за компьютер небольшую часть нашего месячного заработка, мы принесем его домой и начнем работать. На компьютере окажется установленной операционная система и набор ссылок на программы в Интернете, за использование которых придется платить кредитной карточкой. Если нас это не устроит, то нам никто не мешает посетить ресурсы Интернета, созданные сторонниками "свободного софта", и загрузить оттуда не менее функциональные, но зато крайне дешевые, а то и бесплатные программы, за использование которых ничего и никому не придется отчислять. Заодно такая операционная система закроет от всевозможных контролирующих организаций накопитель нашего компьютера, и при желании мы сможем без опасений обмениваться с друзьями музыкой, фильмами и картинками, притом что пользователи платных ОС смогут делать такое, только отчисляя процент "правообладателям".

Как только наш компьютер будет внесен в квартиру, он включится в общую сеть электронных устройств дома, и с него можно будет управлять этими приборами. Например, посмотреть, что делается перед подъездом или кто стоит у входа в квартиру. Так как компьютер помещается в карман, то его нетрудно носить с собой по квартире.

Кроме своей основной задачи, компьютер сможет принимать программы ТВ (но распространяющиеся традиционным способом - в диапазоне УКВ), а также служить мобильным телефоном. Последние, впрочем, тоже останутся и будут успешно конкурировать с компьютерами, отличаясь от них большей направленностью на развлекательно-мультимедийное содержимое и более удобным для разговора дизайном.

Все рано или поздно приходит к завершению, и дальнейшим этапом развития компьютеров как средства работы с информацией будет уже совершенно иная техническая система. Может быть, она будет представлять собой средство внесения знаний прямо в мозг - без посредников в виде экрана, клавиатуры, бумаги... Но это будет уже явно не через 20, и вряд ли даже через 50 лет...

При разработке профессиональных стандартов в ракетно-космической промышленности необходимо учитывать прогноз развития информационно-коммуникационных технологий. Такой прогноз имеет две стороны - прогноз развития указанных технологий как таковых, позволяющий сформировать представление о компьютерах будущего, и прогноз их использования в производстве и управлении, в хозяйственной деятельности предприятий ракетно-космической отрасли. Первая из указанных сторон рассмотрена в настоящем разделе, второй посвящен следующий раздел монографии. Более специальные вопросы рассмотрены в статьях [16, 17], монографии [18] и других работах. Для получения количественных оценок в ситуациях, рассмотренных в настоящем разделе, целесообразно использовать статистические и экспертные методы прогнозирования [19].

4.9. Профессиональные стандарты, информационно-коммуникационные технологии и управление в ракетно-космической промышленности

Для реализации инновационных стратегий нужны подготовленные кадры [1]. Поэтому вполне естественно большое внимание, уделяемое кадровому обеспечению процессов управления инновационной деятельностью на предприятиях ракетно-космической отрасли [2]. Подготовка кадров и управление персоналом [3, 4] в соответствии с действующим законодательством должны проводиться на основе профессиональных стандартов. Как известно, профессиональный стандарт - это характеристика квалификации, необходимой работнику для осуществления определенного вида про-

фессиональной деятельности. Профессиональные стандарты разрабатываются согласно статье 195.1 Трудового кодекса Российской Федерации.

Работник приобретает квалификацию для многолетней работы. Поэтому бесспорно, что содержание профессиональных стандартов должно отражать результаты прогнозирования научно-технического прогресса в соответствующей области, например, в ракетно-космической промышленности [2–4]. Общеизвестно возрастание роли информационно-коммуникационных технологий в производстве и управлении. Прогнозированию развития информационно-коммуникационных технологий посвящена статья [5]. Как следующий шаг, необходимо прогнозирование тенденций использования информационно-коммуникационных технологий при решении проблем управления в социально-экономической области с целью отражения этих тенденций в профессиональных стандартах. Подходам к решению этой задачи и посвящена настоящий раздел.

4.9.1. Понятие профессионального стандарта

В настоящее время в России осуществляется масштабное обновление национальной системы квалификаций работников и формирование системы профессиональных стандартов (ПС), затрагивающее практически все отрасли российской экономики и социальной сферы. Планируется, что в ближайшем времени профессиональные стандарты заменят привычные для всех квалификационные справочники. В ближайшее время Правительство РФ планирует разработать до 800 профессиональных стандартов, которые станут основой для расчета тарифных ставок и систем оплаты труда работников.

Разработка профессиональных стандартов ведется в соответствии с п.1 известного Майского Указа Президента Российской Федерации (от 7 мая 2012 г. № 597 "О мероприятиях по реализации государственной социальной политики" [6]). Через полгода правила разработки, утверждения и применения профессиональных стандартов были утверждены Постановлением Правительства РФ №23 от 22 января 2013 г. [7].

Правительство РФ поручило координации разработки профессиональных стандартов Министерству труда и социальной защиты РФ. Это министерство разработало основополагающие документы, в том числе

- перечень проектов профессиональных стандартов (ПС), разработка которых предусмотрена за счет средств федерального бюджета, в котором утвержден список, порядок, срок разработки, наименования и прочие требования к подготовке ПС;
- методические рекомендации по разработке ПС;
- макет профессионального стандарта;
- уровни квалификации в целях разработки проектов ПС;
- требования к процедуре разработки и общественного обсуждения проекта ПС и др.

Профессиональный стандарт является ключевым механизмом саморегулирования рынка труда. Он является многофункциональным нормативным документом, устанавливающий в рамках конкретного вида (области) профессиональной деятельности требования:

- к содержанию и качеству труда;
- к условиям осуществления трудовой деятельности;
- к уровню квалификации работника;
- к практическому опыту, профессиональному образованию и обучению, необходимому для соответствия данной квалификации.

Профессиональный стандарт состоит из структурных единиц, каждая из которых относится к определенному квалификационному уровню и содержит описание:

- необходимых знаний и умений;

- уровня ответственности и самостоятельности;
- уровня сложности выполняемой трудовой функции.

Официально понятие профессионального стандарта введено в Трудовой кодекс РФ (ТК РФ) и статью 1 Федерального закона «О техническом регулировании» Федеральным законом N 236-ФЗ от 03.12.2012 г. [8].

В статье 195.1 ТК РФ "Понятия квалификации работника, профессионального стандарта" даны следующие определения, которые будем использовать в настоящей статье:

Квалификация работника — уровень знаний, умений, профессиональных навыков и опыта работы работника.

Профессиональный стандарт — характеристика квалификации, необходимой работнику для осуществления определенного вида профессиональной деятельности.

Практическое применение профессионального стандарта многообразно. *Для работодателя* – в ПС даны формулировки требований к работникам. *Для работника* – ПС позволяет провести оценку соответствия имеющихся у него компетенций требованиям рынка труда и конкретного работодателя. *Для системы образования* – на основе ПС проводится разработка образовательных стандартов и программ, соответствующих требованиям рынка труда. Кроме того, что наличие разработанных профессиональных стандартов позволит сформировать принципиально новый классификатор профессий для конкретного вида экономической деятельности или области профессиональной деятельности, весьма важно, что в процессе разработки ПС у работников и работодателей происходит обновление и уточнение содержания трудовой деятельности.

Констатируем, что ПС позволяют контролировать профессионализм работников, поддерживать и улучшать стандарты качества в определенной области деятельности. Поскольку профессиональный стандарт является тем профессиональным минимумом, которому должны соответствовать все работники и руководители отрасли, он может быть использован в качестве основы для создания корпоративных стандартов, стандартов предприятий и документов, посвященных должностным обязанностям (должностных инструкций). При этом возможно расширение или уточнение функций работников с учетом особенностей организации производства, труда и управления в конкретной ситуации, а также прав и ответственности работников на предприятии или в компании. Профессиональный стандарт является основой для определения профессионального уровня и совершенствования профессиональных компетенций работников и их сертификации. Требования профессионального стандарта используются при составлении образовательных стандартов, программ обучения и учебно-методических материалов, а также при выборе форм и методов обучения в системе профессионального образования и внутрикорпоративного обучения персонала [9].

Итак, профессиональный стандарт - характеристика квалификации, необходимой работнику для осуществления определенного вида профессиональной деятельности. Профессиональные стандарты разрабатываются согласно статье 195.1 Трудового кодекса Российской Федерации, для применения:

- работодателями при формировании кадровой политики и в управлении персоналом, при организации обучения и аттестации работников, разработке должностных инструкций, тарификации работ, присвоении тарифных разрядов работникам и установлении систем оплаты труда с учетом особенностей организации производства, труда и управления;

- образовательными организациями профессионального образования при разработке профессиональных образовательных программ;

- при разработке в установленном порядке федеральных государственных образовательных стандартов профессионального образования [10].

4.9.2. Основные проблемы разработки профессиональных стандартов

Каким должен быть профессиональный стандарт в ракетно-космической отрасли? Основная проблема состоит в том, что хотя стандарт должен быть введен в действие в ближайшее время, его реальное влияние на отрасль начнется через 5–10 лет и будет продолжаться еще по крайней мере 10 лет, т.е. до 2030-х годов. Действительно, на его основе должны быть созданы государственные образовательные стандарты и учебные программы, на что потребуется некоторое время. Подготовка специалистов требует 5–6 лет, бакалавров - 4, магистров - 6. Таким образом, первые выпускники, получившие образование в соответствии с профессиональным стандартом, появятся в 2020-х годах.

Каков период "амортизации" полученных знаний, умений, навыков (кратко - знаний)? Очевидно, разные виды знаний обесцениваются с разной скоростью. Так, по общепринятой экспертной оценке математический анализ (дифференцирование и интегрирование) и основы экономики предприятия будут полезны в течение всей профессиональной деятельности инженера РКП. В то время как компетенции, навыки, умения и знания (далее кратко - знания), относящиеся к быстро развивающимся областям, могут устареть еще за время учебы. В частности, знания, касающихся конкретных технических систем (например, мобильных телефонов) и нормативных документов.

Пример. В 70-х годах считалось естественным, что с ЭВМ работают программисты, которые должны владеть двоичной системой счисления. В настоящее время стало ясно, что готовить надо не столько программистов, сколько квалифицированных (продвинутых) пользователей, но и тем, и другим нет необходимости знакомиться с двоичной системой счисления.

Отсюда следует, что в профессиональных стандартах требования к овладению "долгоиграющими" знаниями могут быть прописаны достаточно подробно, в то время как информация о "краткосрочных" знаниях может быть дана в общем виде. Выявить "срок амортизации" тех или иных знаний можно на основе экспертных оценок. Важно, чтобы этот срок был указан в самом профессиональном стандарте, что позволит своевременно обновлять его содержание (например, не реже чем один раз в пять лет).

Профессиональный стандарт должен исходить из принятой в МГТУ им. Н. Э. Баумана концепции "Образование через науку", т.е. знания, умения, навыки, компетенции, предусмотренные профессиональным стандартом, должны быть основаны на современных научных достижениях. Так, математические методы исследования должны исходить из новой парадигмы этой области знаний [11–22], а статистические методы анализа данных должны соответствовать высоким статистическим технологиям [23–24].

Профессиональный стандарт должен предусматривать непрерывное образование, ежегодное повышение квалификации, в том числе дистанционное. В настоящее время Россия отстает от западных стран, интенсивность непрерывного образования на порядок ниже.

Непрерывное научно-техническое развитие приводит к необходимости внесения в профессиональный стандарт соответствующих изменений. Он должен регулярно, раз в 5 лет, пересматриваться. При необходимости - чаще. По нашей экспертной оценке, утвержденный в ближайшее время стандарт перестанет быть актуальным через 15 лет - к 2030 г. К этому времени его содержание будет полностью заменено.

Содержание профессионального стандарта должно быть тщательно проработано сообществом профессионалов. Для этого целесообразно организовать серию экспертиз. Печальным примером является совокупность государственных стандартов СССР по

статистическим методам управления качеством продукции, основная часть которых содержала грубые ошибки, порожденные некомпетентностью разработчиков и несовершенством экспертных процедур, применяемых в процессе разработки стандартов. Большинство указанных стандартов были отменены в 1987 г. Подробно эта ситуация рассмотрена в учебниках А.И. Орлова (см., например, [25–27]). Нельзя допустить повторения подобной ситуации.

4.9.3. Что нужно учитывать при разработке профессионального стандарта?

Для разработки профессионального стандарта в области ракетно-космической промышленности (РКП) необходимо спрогнозировать характеристики квалификации (уровень знаний, умений, профессиональных навыков и опыта работы), необходимой работнику для осуществления профессиональной деятельности в РКП в 2020–2030 гг.

Для этого необходимо опираться на результаты прогнозирования научно-технического прогресса в отрасли. Для такого прогнозирования необходимо исходить из сценариев социально-экономического развития России, которые определяют ограничения (планы) по финансированию космической деятельности и по программе пусков космических аппаратов. Необходимо учитывать развитие ситуации в отрасли, прежде всего ограничения по производственным возможностям предприятий «Роскосмоса» и по их кадровому составу. Основное - прогноз технического прогресса и организации производства в отрасли (можно ожидать массового перехода научных работников, инженеров и менеджеров к дистанционной работе, всеобщее использование современных компьютеров и информационно-коммуникационных технологий. Кроме того, необходим прогноз системы образования, в частности, динамическое развитие системы специальностей, расширение структур непрерывного образования, диверсификация подготовки, в частности, внедрение широкого спектра лекционных курсов и семинаров по выбору обучающихся.

Реализация постоянно идущего процесса прогнозирования может быть реализована с помощью технологий ситуационных центров. Как следует из сказанного выше, базовым методом прогнозирования является метод сценариев. Обсудим его основные идеи [28].

Метод сценариев - это метод декомпозиции задачи прогнозирования, предусматривающий выделение набора отдельных вариантов развития событий (т.е. сценариев), в совокупности охватывающих все возможные варианты развития. При этом каждый отдельный сценарий должен допускать возможность достаточно точного прогнозирования, а общее число сценариев должно быть обозримо.

Возможность подобной декомпозиции не очевидна. При применении метода сценариев необходимо осуществить два этапа исследования:

- построение исчерпывающего, но обозримого набора сценариев;
- прогнозирование в рамках каждого конкретного сценария с целью получения ответов на интересующие исследователя вопросы.

Каждый из этих этапов лишь частично формализуем. Существенная часть рассуждений проводится на качественном уровне, как это принято в общественно-экономических и гуманитарных науках. Одна из причин заключается в том, что стремление к излишней формализации и математизации приводит к искусственному внесению определенности там, где ее нет по существу, либо к использованию громоздкого математического аппарата. Так, рассуждения на словесном уровне считаются доказательными в большинстве ситуаций, в то время как попытка уточнить смысл используемых слов с помощью, например, теории нечетких множеств приводит к весьма гро-

моздким математическим моделям. Можно выразиться и в терминах теории устойчивости [29–32] – выводы из рассуждений на качественном уровне более устойчивы к малым колебаниям исходных данных и предпосылок модели, чем выводы из рассуждений на базе количественных экономико-математических моделей.

Кроме макроэкономических характеристик необходимо учитывать состояние и динамику отечественного массового сознания, политических, в то числе внешнеполитических реалий, поскольку на рассматриваемом интервале времени экономика зачастую следует за политикой, а не наоборот [33–36].

4.9.4. Влияние развития информационно-коммуникационных технологий на хозяйственную деятельность

Прогнозу развития информационно-коммуникационных технологий посвящена статья [5]. Обсудим изменения в хозяйственной деятельности (в производстве и управлении), являющиеся следствиями этого развития.

Современные информационно-коммуникационные технологии создают принципиально новую ситуацию в организации хозяйства.

Возникла возможность управлять из одного центра работой подразделений организации, разбросанными по всему миру. Например, из офиса управляющей компании в Москве можно управлять технологическим процессом на заводе в Иркутске. Дело не столько в глобальной централизации управления, сколько в возможности передачи в любую точку Земли информации, используемой для принятия управленческих решений. В том числе на домашний компьютер управляющего. Скайп и телеконференции во многом делают ненужными командировки менеджеров. Конечно, пользу личных контактов отрицать не будем, но удаленная работа имеет свои преимущества. Недаром она получила правовое оформление в действующей редакции Трудового Кодекса РФ.

Автор настоящей статьи за последние годы выпустил две книги в Германии, три - в Ростове-на-Дону, два - в Краснодаре. Но никуда не ездил по издательским делам. Уже в настоящее время теоретик может вести научную работу исключительно за домашним компьютером, отсылать статьи и отчеты по электронной почте. Сколько места можно освободить в зданиях НИИ, если признать очевидное - научному работнику теоретического плана гораздо эффективнее работать дома, чем на "рабочем месте" в НИИ!

Требование присутствия на рабочем месте - во многом пережиток прошлого. Преимуществ удаленной работы много, в том числе:

- освобождение времени, которое ранее тратилось на дорогу на работу и обратно;
- уменьшение затрат на содержание рабочего помещения (можно спланировать собрания членов подразделений в одном помещении, например, раз в неделю, вместо использования рабочих площадей для каждого подразделения в отдельности, что дает возможность сократить площадь офиса в разы);

- свободный график сотрудников, возможность планирования личного графика выполнения производственных заданий и личных дел.

Есть и недостатки:

- необходимо повышение требований к планированию работы и процедурам контроля и оценки результатов;

- как следствие, требуется повышение квалификации организаторов производства (недостаточно квалифицированные руководители будут против удаленной работы).

Очевидно, что удаленная работа совместима не со всеми профессиями. Однако ее важность подчеркнута тем, что возможность удаленной работы внесена в Трудовой кодекс РФ.

Развитие информационно-коммуникационных технологий приводит к заметному изменению практики хозяйственной деятельности. Например, магазины постепенно переходят в выставки, т.е. в них уже не выбирают объект для покупки из числа представленных, а на основе информации о возможных вариантах и знакомства с выставочными образцами формируют заказ. Так, получила распространение услуга "Книга по требованию": покупатель выбирает конкретное издание из выставленных, и для него печатается персональный экземпляр. На этом примере видна польза информационно-коммуникационных технологий: нет необходимости печатать тираж и распределять его по книжным магазинам, нет физической возможности для появления залежей нераспроданных остатков, заметно сокращается объем складов... Обобщая, можно сказать, что информационно-коммуникационные технологии позволяют избежать производства товаров, не находящих покупателей, поскольку каждый экземпляр товара производится по заказу конкретного потребителя. Конкурируют не произведенные товары как материальные сущности, а их виртуальные прообразы.

Ожидаем грядущий переход к биометрическому режиму оплаты купленных товаров (по оттиску пальца или по компьютерному образу лица). Такой переход ведет к сокращению и в перспективе - к ликвидации оборота наличных денег. Биометрическая идентификация человека позволит отказаться от документов, удостоверяющих личность, как следствие, от охранных структур, сотрудники которых смогут заняться производительным трудом.

Необходимо критически изучить использовать опыт управления транснациональными корпорациями и обосновать рекомендации по его использованию в грядущей информационной экономике. Такая рекомендация основана на том, что величины активов многих из таких корпораций зачастую превышают национальное богатство не только малых, но и средних стран.

4.9.5. Прогноз подходов к организации принятия управленческих решений

В течение периода прогноза (т.е. до 2030 г.) произойдут заметные изменения в системах принятия управленческих решений, которые необходимо учесть в профессиональном стандарте. Они связаны с грядущим развитием информационно-коммуникационных технологий, соответствующими расширениями управленческих и экономических концепций и теорий, нацеленными на вовлечение в процессы принятия решений большого числа заинтересованных физических и юридических лиц.

Выделим основной этап в процессе подготовки управленческого решения. Если потребности сформулированы, т.е. составлено задание для производственных структур, то дальнейшие действия достаточно понятны. Необходимо составить календарный план выполнения этого задания, оценить реализуемость проекта, сопоставить с другими проектами. Если ресурсов не хватает, то придется вернуться к этапу целеполагания. Все эти действия детально проработаны в менеджменте, имеются соответствующие интеллектуальные инструменты организационно-экономического моделирования [37 - 39].

Основное - процедуры целеполагания. Как мнения отдельных лиц привести "к единому знаменателю"? Для малых групп - семья, студенческая группа, дружеская компания, сотрудники подразделения - опыт самоорганизации всем известен из личного опыта. Для более широких общностей также известны различные способы - самодержавие, представительная демократия, процедуры демократического централизма, прямая демократия (например, сходка крестьян русской деревенской общины). К сообщениям статей [40-45], посвященных солидарной информационной экономике, доба-

вим, что в теории экспертных оценок [46, 47] имеются разнообразные процедуры выработки коллективного мнения, в частности, основанные на использовании сетей экспертов [48] и на «модели Команды Syntegrity» Ст. Бира [49]. Интересен опыт массовых петиций многомиллионных добровольных сообществ "Avaaz.org", "Change.org", "Народная инициатива", "Российская общественная инициатива", "Активный Гражданин" и др., действующих на основе соответствующих информационно-коммуникационных систем (Интернет-ресурсов). С помощью подобных опросов может быть выяснено общественное мнение по поводу выбора того или иного направления, для примера, космической деятельности. Например, может быть выбор между развитием орбитальных космических систем, созданием лунной базы, различными вариантами полетов на Марс, в том числе в один конец.

Для создания адекватных процедур выработки коллективного мнения на уровне государства бесценен опыт новгородского вече [50], особенно с учетом реконструкции истории, полученной в рамках новой хронологии [51]. Согласно этой реконструкции новгородское вече - высший орган власти всего русского государства, а не его части. Вече "является собственной (в соответствии с органической принадлежностью к естественной организации российского цивилизационного строя) – «естественной и справедливой формой народного самоуправления»" [50]. "Примечательный (и парадоксальный) текущий момент состоит в том, что авторы (исследующие феномен Новгородского вече) – в подавляющем значении используют термины и понятия, характеризующие современный Западный Тип общественного устройства (как «республика», «демократические процедуры», «институты гражданского общества», «народное представительство», «государственная дума»" (там же). Такой подход некорректен, поскольку "современная демократическая республика – это продукт принципиально иной цивилизации (по сути, чужеродной к Древнерусскому общественному строю), в первую очередь по отношению к тем историческим социокультурным основаниям, что послужили плодородной почвой как раз для возникновения и успешного развития рассматриваемого Органического общественного образования – Новгородского вече". Как пишут К.С. Хруцкий и А.В. Карпов: "Вече стало естественным проявлением как раз автономного Органицистского (в аспекте методологии изучения – Биокосмологического) Типа цивилизационного устройства и социокультурной активности".

По нашему мнению, организация принятия решений должна определяться Ноократией. "С точки зрения научного Органицизма, Ноократия – это «приоритет человеческого разума» (благодаря активной свободной деятельности лучших его носителей, селекция которых является главной целью), и когда активность лучших умов имеет определяющее значение в организации управления обществом. ... Вече не служило в качестве «института народного представительства», но как раз осуществляло и обеспечивало признание и выдвижение *лучших умов* (в отношении к способности достигать «общего блага») среди членов данного общества (взятого как целое) – для осуществления ими (за счет своих добродетельных – превосходных – природных способностей) – естественного (успешного, Органицистского) управления общественными делами. Таким образом, механизм Вече осуществлял не выборность (посредством демократических процедур народного избрания как такового), но скорее признание естественных (виртуальных, присущих) качеств у некоторых из своих сограждан – их разумности и добродетельности, в достижении всеобщей целесообразности, т.е. *общего блага* – и их выдвижение (и наделение ответственностью) как лучших представителей своего общества к осуществлению присущего для них дела управления обществом и его успешным (благополучным) развитием. ... Конечной задачей выступают не интересы определенной группы или класса общества (большинства или меньшинства), но на первый план выходят цели достижения всеобщего благополучия (как в примере с физиологическим

организмом, где нет ничего «не нужного», но все скоординировано в интересах целостного – текущего и онтогенетического – благополучного развития). Это ‘всеобщее благополучие’ касается как онтогенеза (жизненного развития в целом) отдельного индивида и общества в целом, так и отношений общества с окружающей средой (т.е. экологических вопросов), и отношений между странами и государствами в целом мире (т.е. вопросов глобального развития). ... Существенно, что управление общественными делами, реализуемое через (Органицистскую) форму Вечевой организации – предстает как в высшей степени эффективная форма общественной организации" [50].

Отмеченные черты новгородского вече позволяют рассматривать его как прообраз будущей системы общественного самоуправления, соответствующей развиваемой в МГТУ им. Н. Э. Баумана функционалистско-органической солидарной информационной экономике - экономике без рынка и денег [40–45].

Подчеркнем, что хозяйственная деятельность должна иметь нравственную составляющую [52]. Кратко: экономика для человека, а не человек для экономики.

Развитие функционалистско-органической солидарной информационной экономики продолжается. Основные предшественники - В.М. Глушков и Стаффорд Бир. Напомним основную информацию об их работах.

Эффективные механизмы принятия и реализации плановых решений должны опираться на современные информационные технологии. База разработки таких технологий – кибернетика (вспомним работы Н. Винера, А.И. Берга, Н.Н. Моисеева, многих других). В послевоенные годы в нашей стране, как и во всем мире, разрабатывались различные типы автоматизированных систем управления. Наиболее грандиозный проект - в начале 1960-х гг. В.М. Глушков предложил правительству СССР создать Обще-государственную автоматизированную систему управления экономикой страны (ОГАС), для чего, по его оценкам, требовалось как минимум 15–20 лет и 20 млрд. тогдашних рублей, однако выигрыш стоил того: ОГАС давала реальный шанс построить самую эффективную экономику в мире. В.М. Глушков писал:

«Отныне только «безмашинных» усилий для управления мало. Первый информационный барьер или порог человечество смогло преодолеть потому, что изобрело товарно-денежные отношения и ступенчатую структуру управления. Электронно-вычислительная техника – вот современное изобретение, которое позволит перешагнуть через второй порог. Происходит исторический поворот по знаменитой спирали развития. Когда появится государственная автоматизированная система управления, мы будем легко охватывать единым взглядом всю экономику. На новом историческом этапе, с новой техникой, на новом возросшем уровне мы как бы «проплываем» над той точкой диалектической спирали, ниже которой, отделенный от нас тысячелетиями, остался лежать период, когда свое натуральное хозяйство человек без труда обозревал невооруженным глазом» [53].

Своеобразная ОГАС, хотя и куда более скромная по масштабам, была реализована на практике в другом уголке Земли – в Чили, во время президентства Сальвадора Альенде. Один из основоположников кибернетики Стаффорд Бир разработал автоматизированную систему управления национализированными предприятиями Чили. Проект получил название «Киберсин» [54]. Он представлял собой автоматизированную систему сбора и обработки информации, которая состояла из четырех основных компонент: «Кибернет» - система связи (на основе телексов), «Киберстрайд» - компьютерные программы, «Чико» - математическая модель чилийской экономики - и ситуационная комната, из которой велось управление (зал с экранами, на которых отображалось в виде графиков и схем состояние экономики Чили). Можно было управлять производством всей страны в реальном времени (каждым конкретным предприятием), сразу же видеть результаты принятых решений и при необходимости вносить поправки. Кроме того, в

каждом населенном пункте создавались «опросные пункты», где производился автоматизированный опрос населения по поводу принимаемых мер. Эти центры были включены в систему «Киберсин», и правительство быстро узнавало реакцию населения на очередное нововведение. Разработки Бира дают прообразы (прототипы) для следующего этапа развития информационных систем управления предприятиями и их объединениями – интегрированными производственно-корпоративными структурами, а также регионами – муниципальными образованиями, субъектами федерации, Россией, международными объединениями, Землей в целом.

На основе идей функционалистско-органической (солидарной) информационной экономики и современных достижений теории принятия решений [26, 55, 56] (прежде всего экспертных процедур [46–48]) и информационно-коммуникационных технологий могут быть реализованы идеи В.М. Глушкова и Ст. Бира.

С содержанием настоящего раздела монографии перекликается опубликованный в 2016 г. доклад Всемирного банка «Цифровые дивиденды» [57]. В нем обоснованно утверждается что мы переживаем величайшую информационно-коммуникационную революцию в истории человечества. Более 40 процентов населения планеты имеет доступ к интернету, и каждый день в сеть выходят новые пользователи. Даже среди беднейших 20 процентов домохозяйств мобильный телефон есть почти в каждом 7 из 10. Поразительно, что число беднейших домохозяйств, располагающих мобильным телефоном, выше, чем число имеющих доступ к канализации или чистой питьевой воде. Авторы доклада [57] установили, что воплотить в жизнь преобразовательный потенциал революции в сфере цифровых технологий мешают традиционные проблемы развития. Нынешнее расширение доступа к цифровым технологиям несет многим людям богатство выбора и большие удобства. За счет усиления социальной интеграции, повышения эффективности и внедрения инноваций такой доступ открывает бедным и обездоленным слоям населения возможности, которых они прежде были лишены. В настоящее время по всему миру в популярную поисковую систему Google ежедневно отправляется свыше 4 миллиардов запросов – и, в то же время, 4 миллиарда человек по-прежнему не имеют доступа к интернету. Величайший в истории подъем информационно-коммуникационных технологий не станет поистине революционным до тех пор, пока выгоду от него не ощутят все люди во всех уголках планеты.

Развитие информационно-коммуникационных технологий иногда вызывает опасения. Еще в 50-е годы Н. Винер обсуждал проблемы создания заводов - автоматов, на которых места рабочих займут роботы, что может привести к массовой безработице [58]. За прошедшие 60 лет этого не произошло, в частности, из-за дороговизны труда роботов по сравнению с трудом работников - людей. Необходимо подчеркнуть, что даже при реализации утопии заводов - автоматов разумно организованное общество сумеет направить освободившихся работников в другие сферы деятельности. Весьма высокой (практически неограниченной) может быть занятость в науке и образовании. Появление массовой безработицы в результате развития информационно-коммуникационных технологий - это миф, порожденный отсутствием глубокого анализа возможностей использования человеческих ресурсов.

5.1. АСК-анализ влияния экологических факторов на качество жизни населения региона

Без опоры на науку невозможно становление полноценного экологического сознания. Чтобы повысить обоснованность и вес выводов о влиянии экологии на качество жизни, необходимо количественно оценить силу и направление влияния на него разнородных экологических факторов. Однако, оказывается, что сделать это довольно проблематично по целому ряду причин. Во-первых, это отсутствие или малодоступность необходимых для подобных исследований исходных данных. Те же данные, которые все же удастся найти, охватывают небольшие периоды наблюдений (малый лонгитюд), а их восполнение, в т.ч. путем проведения экспериментов, принципиально невозможно. В результате невозможно требовать от таких данных полных повторностей, что является необходимым условием корректного применения факторного анализа. Во-вторых, экологические факторы описываются разнородными показателями, измеренными в различных типах измерительных шкал (номинальных, порядковых и числовых) и в различных единицах измерения. Математические методы сопоставимой обработки подобных данных, а также реализующий эти методы программный инструментарий, фактически отсутствуют. В-третьих, подобные задачи относятся к задачам большой размерности, т.е. в них идет речь не о 5 или максимум 7 факторах, как в факторном анализе, а о сотнях и тысячах. В четвертых исходные данные зашумлены и требуют устойчивых методов. В-пятых, экологические факторы взаимосвязаны и требуют нелинейных непараметрических подходов. Для решения этих проблем предлагается применить новую инновационную интеллектуальную технологию: автоматизированный системно-когнитивный анализ и его программный инструментарий – систему «Эйдос». Приводится краткий численный пример оценки влияния экологических факторов на продолжительность жизни и причины смерти

«Системы искусственного интеллекта позволяют решать сложнейшие проблемы, которые не возникали, пока этих систем не было»

/Mahaguru I.T./

5.1.1. Формулировка проблемы

Все в принципе согласны с тем, что «хорошая экология – это хорошо, а плохая экология – это плохо». Чтобы дать подобную оценку на качественном уровне нет необходимости в проведении каких-то специальных научных исследований, т.е. это не сложно. Соответственно и влияние подобных малосодержательных по существу выводов на экологическое сознание и на принятие решений руководителями различных уровней не недостаточно.

Чтобы повысить обоснованность подобных выводов и их вес в формировании экологического сознания необходимо количественно оценить силу и направление влияния разнородных экологических факторов, например, таких как уровень загрязнения окружающей среды (воздуха, вод и почв), на какие-то значимые интегральные показате-

тели, непосредственно касающиеся основной массы населения, например на качество жизни населения региона [1-4].

Совершенно четко однозначно сформулируем основополагающую на наш взгляд мысль: *«Экология важна не только сама по себе, но и потому, что она оказывает существенное влияние на то, что важно для нас: на наше физическое и психическое здоровье, качество жизни, рождаемость и смертность, физическая продолжительность жизни, а также продолжительность активной и творческой жизни и работоспособного состояния и многое-многое другое».*

Чтобы исследовать влияние экологических факторов на все эти аспекты жизни недостаточно лишь экологических баз данных о степени загрязнения земли, воды, воздуха, пищи, строительных материалов, одежды, мебели, игрушек и т.п. и т.д. Необходимы также базы данных, отражающие наше физическое и психическое здоровье, качество жизни, рождаемость и смертность, продолжительность жизни, продолжительность активной и творческой жизни и работоспособного состояния. И все эти базы данных необходимо обрабатывать *совместно* в сопоставимой форме по одной методологии, технологии и методике и в одной реализующей их программной системе. Проблема состоит не в том, что такой методологии, технологии, методики программной системы нет, т.к. они есть, а в том, что они совершенно неизвестны специалистам в конкретных областях, для которых они предназначены, в частности совершенно неизвестны экологам.

5.1.2. Традиционные подходы к решению проблемы и их недостатки

Экологи до сих пор надеются на то, что их задачи позволит решить MS Excel и системы «Статистика» и SPSS. Но постепенно их иллюзии рассеиваются, и они начинают понимать, что возможности инструмента решения проблемы должны соответствовать сложности проблемы, и что для этого малопригодны математические методы, разработанные 100 лет назад и более, например факторный анализ.

Экологи с завидным упорством пытаются применять эти методы, однако оказывается, что *корректно сделать это довольно проблематично* по целому ряду вполне объективных, независимых от исследователей, реально имеющих место причин. Мы назовем лишь некоторые из них.

Во-первых, это отсутствие или малодоступность необходимых для подобных исследований исходных данных. Те же данные, которые все же удается найти, охватывают небольшие периоды наблюдений (малый лонгитюд), а их восполнение, в т.ч. путем проведения экспериментов, принципиально невозможно. В результате невозможно требовать от таких данных полных повторностей, что является необходимым условием корректно применения факторного анализа.

Во-вторых, экологические факторы описываются разнородными показателями, измеренными в различных типах измерительных шкал (номинальных, порядковых и числовых) и в различных единицах измерения. Математические методы сопоставимой обработки подобных данных, а также реализующий эти методы программный инструментарий, фактически отсутствуют.

В-третьих, подобные задачи относятся к задачам большой размерности, т.е. в них идет речь не о 5 или максимум 7 факторах, как в факторном анализе, а о сотнях и тысячах. Обычно в руководствах по факторному анализу начинаются с сакраменталь-

ной фразы: «Выберем небольшое число наиболее важных факторов, которые будем исследовать». Но при этом авторы этих руководств благоразумно воздерживаются от рассмотрения методологических, методических и практических подходов к тому, как это сделать, т.к. они просто отсутствуют или малоизвестны, как и необходимый для этого программный инструментарий. На практике обычно все сводится не к исследованию объекта, который надо исследовать, а к исследованию данных, которые фактически есть и удовлетворяют этим жестким требованиям, но мягко говоря, не очень полно отражают исследуемый объект.

В-четвертых, факторный анализ является *неустойчивым* методом в том смысле, что, даже небольшие вариации значений исходных данных приводят к сильному изменению результатов применения метода, т.е. требует, чтобы исходные данные были абсолютно точными. Ясно, что реальные исходные данные сильно зашумлены и не удовлетворяют этому требованию. Да и даже в принципе вряд ли можно всерьез говорить о каких-то гипотетических абсолютно точных данных, т.е. ясно, что это некая абстракция, которой в полной мере практически ничего в действительности не соответствует.

В-пятых, факторный анализ является линейным, параметрическим методом, т.е. требует выполнения нормального распределения и независимости исследуемых факторов. Дело в том, что нормальное распределение выполняется только при действии большого числа случайных и независимых друг от друга аддитивных факторов, а на практике они конечно не случайны и часто взаимозависимы, не аддитивны, зависят от каких-то третьих более фундаментальных факторов.

5.1.3. Предлагаемое решение проблемы с применением АСК-анализа и системы «Эйдос»

Для решения подобных задач предлагается применить новую инновационную интеллектуальную технологию: автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – систему «Эйдос».

АСК-анализ имеет ряд особенностей, которые обусловили его выбор в качестве метода решения проблемы:

1. Имеет *теоретическое обоснование*, основой которого является *семантическая мера целесообразности информации* А.Харкевича.

2. Обеспечивает *корректную сопоставимую количественную* обработку *разнородных* по своей природе взаимосвязанных факторов, измеряемых в *различных единицах* измерения, *высокую точность* и независимость результатов расчетов от единиц измерения исходных данных.

3. Обеспечивает построение *многомерных моделей* объекта моделирования непосредственно на основе *неполных (фрагментированных) и зашумленных (искаженных)* эмпирических данных о нем.

4. Имеет развитую и *доступную программную реализацию* в виде универсальной когнитивной аналитической системы «Эйдос» (открытое программное обеспечение: http://lc.kubagro.ru/aidos/_Aidos-X.htm).

Об АСК-анализе написано довольно много. На момент написания данной статьи это 21 монография и учебное пособие [5-25] и сотни статей в изданиях, входящих в Перечень ВАК РФ [26]. На программный инструментарий АСК-анализа – интеллектуальную систему «Эйдос» и различные ее режимы и подсистемы получено 27 свидетельств

РосПатента [27]. Краткое описание АСК-анализа дано в работе [28], в которой есть и основные необходимые ссылки.

По этим причинам нет необходимости в описании теоретических основ, математической модели, методики численных расчетов (т.е. алгоритмов и структур данных) и программного инструментария АСК-анализа и мы кратко остановимся лишь на двух ключевых моментах: метризации измерительных шкал и нелинейности моделей в АСК-анализе.

5.1.4. Суть метода АСК-анализа – преобразование исходных данных в информацию, а ее в знания путем метризации измерительных шкал

АСК-анализ представляет собой один из современных инновационных методов искусственно интеллекта, который имеет теоретическое обоснование и оснащен широко и успешно апробированным универсальным программным инструментарием, позволяющим решить эти вопросы не только как обычно на теоретическом концептуальном уровне, но и на практике [5, 29].

Модели знаний АСК-анализа основаны на нечеткой декларативной модели представления знаний, предложенной автором в 1979 году и являющейся гибридной моделью, сочетающей в себе преимущества фреймовой, нейросетевой и четкой продукционной моделей и обеспечивающей создание моделей очень больших размерностей до 10 млн. раз превышающих максимальные размерности моделей знаний экспертных систем с четкими продукциями:

- от фреймовой модели модель представления знания системы «Эйдос» отличается существенно упрощенной программной реализацией и более высоким быстродействием без потери функциональности;

- от нейросетевой тем, что обеспечивает хорошо обоснованную теоретически содержательную интерпретацию весовых коэффициентов на рецепторах и обучение методом прямого счета;

- от четкой продукционной модели – нечеткими продукциями, представленными в декларативной форме, что обеспечивает эффективное использование знаний без их многократной генерации для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта.

АСК-анализ является непараметрическим методом, устойчивым к шуму в исходных данных, позволяющий корректно обрабатывать неполные (фрагментированные) исходные данные, описывающие воздействие взаимозависимых факторов на нелинейный объект моделирования.

Суть метода АСК-анализа в том, что он позволяет рассчитать на основе исходных данных какое *количество информации* содержится в значениях факторов, обуславливающих переходы объекта моделирования в различные будущие состояния, причем как в желательные, так и в нежелательные [9]².

² Вопреки тому, как его поняли некоторые авторы

Он состоит в целенаправленном *последовательном повышении степени формализации* исходных данных до уровня, который позволяет ввести исходные данные в компьютерную систему, а затем преобразовать исходные данные в информацию; информацию преобразовать в знания; использовать знания для решения задач прогнозирования, принятия решений и исследования предметной области.

Рассмотрим подробнее вопросы выявления, представления и использования знаний в АСК-анализе и системе «Эйдос».

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ) для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на объект моделирования к каким его изменениям приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количест-

венных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу.

Знания – это информация, полезная для достижения целей³.

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).
2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно **последовательно повышать степень формализации** исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач управления, принятия решений и исследования предметной области.

Соответственно, АСК-анализ имеет следующие этапы:

АСК-анализ имеет следующие этапы [5-25]:

- когнитивно-целевая структуризация предметной области;

³ Основные публикации автора по вопросам выявления, представления и использования знаний:

– <http://www.twirpx.com/file/793311/>

– Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.

– Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012\0197, IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.

- формализация предметной области (формирование классификационных и описательных шкал и градаций и обучающей выборки);
- синтез и верификация статистических и системно-когнитивных моделей;
- решение задач идентификации, прогнозирования, принятия решений и исследования предметной области в наиболее достоверных из созданных моделей.

Единственный неавтоматизированный в системе «Эйдос» этап – это первый, а остальные приведены на рисунках 1 и 2:

О соотношении содержания понятий: «Данные», «Информация» и «Знания»

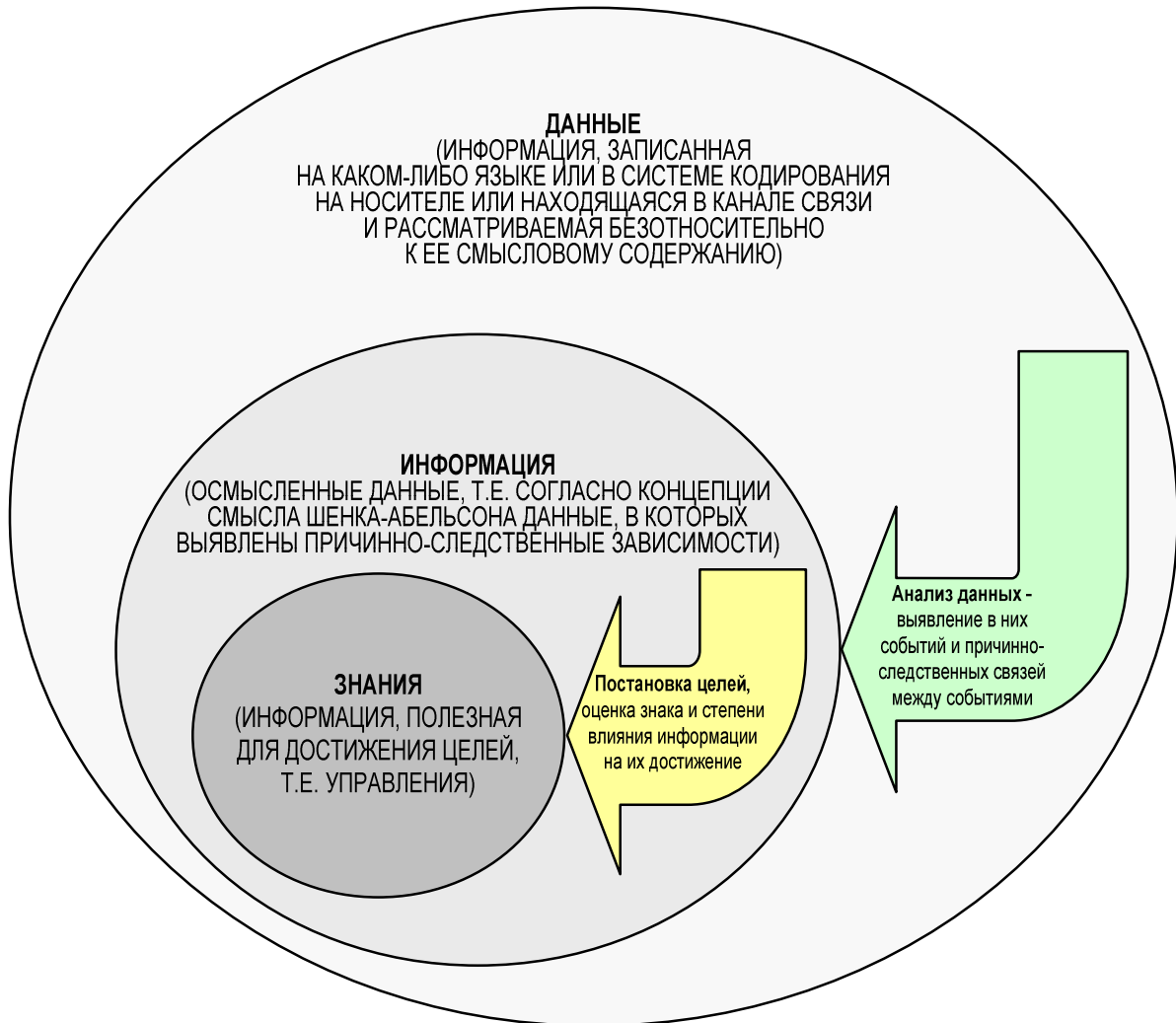


Рисунок 1 – Соотношение содержания понятий: «Данные», «Информация», «Знания»

Последовательность обработки данных, информации и знаний в системе «Эйдос-Х++»

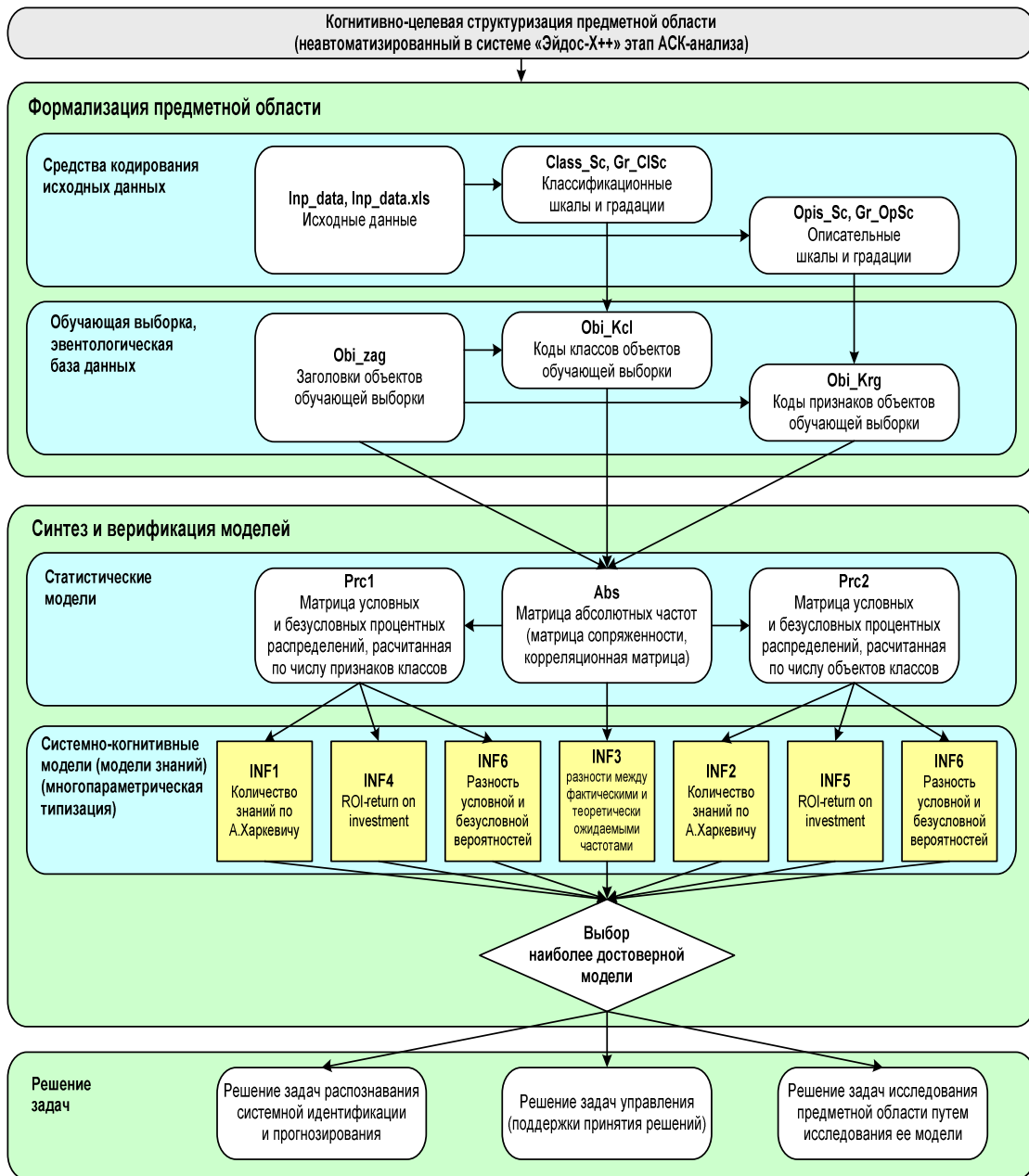


Рисунок 2 – Этапы последовательного повышения степени формализации модели от данных к информации, а от нее к знаниям

После выполнения когнитивной структуризации и формализации предметной области осуществляется синтез статистических моделей и моделей знаний, в которых все шкалы, в которых описаны исходные данные, преобразуются к одному типу: числовому, и к одним единицам измерения: единицам измерения информации, т.е. проводится метризация шкал. В настоящее время в системе «Эйдос» применяется 7 способов метризации шкал [29].

В АСК-анализе факторы формально описываются шкалами, а значения факторов – градациями шкал. Существует три основных группы факторов: физические, социально-экономические и психологические (субъективные) и в каждой из этих групп есть много различных видов факторов, т.е. есть много различных физических факторов, много социально-экономических и много психологических, но в АСК-анализе все они

рассматриваются *с одной единственной точки зрения*: сколько информации содержится в их значениях о переходе объекта, на который они действуют, в определенное состояние, и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: **единицах количества информации**.

Именно по этой причине вполне корректно складывать (в аддитивных интегральных критериях) силу и направление влияния всех действующих на объект значений факторов, независимо от их природы, и определять результат *совместного* влияния на объект *системы* значений факторов.

При этом в общем случае объект является *нелинейным* и факторы внутри него взаимодействуют друг с другом, т.е. для них не выполняется принцип суперпозиции. АСК-анализ позволяет создавать и применять нелинейные модели влияния взаимосвязанных факторов на сложные объекты управления, т.к. является непараметрическим методом [30].

Кратко рассмотрим численный пример применения АСК-анализа и системы «Эйдос» для моделирования влияния экологических факторов на качество жизни населения региона, а именно на один из основных интегральных показателей качества жизни – ее продолжительность.


5.1.5. Краткий численный пример

5.1.5.1. Источники исходных данных

Для создания модели, отражающей влияние экологических факторов на продолжительность жизни в России, использовались данные Федеральной службы государственной статистики⁴ по регионам России, с одной стороны по продолжительности жизни, а с другой стороны – по экологии:


– данные по продолжительности жизни при рождении по регионам России (ожидаемая) за 2013 год взята с сайта: http://www.statdata.ru/spg_reg_rf;

– данные по числу умерших по регионам России с указанием причин на странице: «Социальное положение и уровень жизни населения России»: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138698314188

 Приложение к сборнику (информация по субъектам Российской Федерации), 2014г. (0,3 Мб)

http://www.gks.ru/free_doc/doc_2014/pril_soc-pol2014.rar

– данные по экологическим факторам по регионам России на странице: «Охрана окружающей среды в России»: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1139919459344

 Приложение к сборнику (информация в разрезе субъектов Российской Федерации), 2014г. (0,4 Мб)

http://www.gks.ru/free_doc/doc_2014/Pril-ohrana.rar

Все эти данные из различных источников были собраны в одном Excel-файле, приведенном в таблице 1:

⁴ Огромная благодарность сотрудницам Росстата Новиковой Ирине Владимировне и Шашловой Наталье Викторовне за содействие в поиске необходимых исходных данных

Таблица 1 – Исходные данные по регионам России для модели влияния экологических факторов на продолжительность жизни и причины смерти

Регион Российской Федерации	Продолжительность жизни (Оба пола)	Продолжительность жизни (Мужчины)	Продолжительность жизни (Женщины)	Умерших от инфи. параз. болезней	Умерших от туберкулеза	Умерших от инд. заболеваний	Умерших от ост. сист. кровообращения	Умерших от бол. орг. дыхания	Умерших от бол. орг. пищеварения	Умерших от внешних причин	Умерших от внешних причин	Всего умерших на 100 тыс. чел.	Выборсы в атм. загр. вещ.- всего	Выборсы в атм. загр. вещ. от стационар. источ.	Выборсы в атм. загр. вещ. от передвиж. источ.	Число водн. источ. ников-всего	Водн.источ. чн.не соотв.сан. нормам,%	Водн.источ. чн.не соотв.норм. м.по сан.-хим.сост. %	Водн.источ. чн.не соотв.норм. м.по микробиол. п.%
Алтайский край	68,72	64,11	75,44	42,55	24,22	226,80	660,05	72,79	46,73	140,65	1213,78	430,60	46,73	53,27	40,00	0,00	9,30	0,00	
Анурская область	66,38	60,59	72,59	24,81	21,01	194,08	724,49	60,19	68,17	212,87	1305,62	216,50	57,92	42,08	450,00	7,78	39,10	28,70	
Архангельская область	70,16	64,11	76,27	10,19	5,60	240,85	743,77	44,78	60,82	157,81	1263,81	365,00	67,23	32,77	916,00	31,22	26,10	22,70	
Астраханская область	71,34	65,81	76,72	28,27	21,47	195,92	683,41	41,47	50,93	114,26	1135,74	225,10	57,97	42,03	18,00	0,00	45,80	12,50	
Белгородская область	72,16	66,86	77,32	4,54	1,88	213,47	862,31	49,40	41,16	98,47	1271,23	274,60	42,86	57,14	2991,00	13,07	47,00	30,20	
Брянская область	69,75	63,32	76,32	15,46	10,82	214,32	903,89	62,81	79,80	160,16	1447,25	109,10	33,64	66,36	6090,00	15,48	38,00	37,20	
Владимирская область	69,13	62,78	75,44	12,84	7,27	259,60	980,84	53,47	103,35	142,15	1559,52	133,90	24,20	75,80	1291,00	22,77	52,80	44,60	
Волгоградская область	71,42	66,11	76,51	22,01	13,39	227,29	752,71	66,96	80,24	113,31	1275,92	386,70	44,69	55,31	628,00	9,90	36,60	27,00	
Вологодская область	69,35	63,21	75,63	9,63	5,36	214,59	892,19	46,87	91,23	152,07	1411,93	600,30	83,16	16,84	464,00	20,91	38,20	38,10	
Воронежская область	70,89	64,81	77,03	9,60	5,97	201,84	763,35	45,42	62,80	142,04	1250,81	325,20	23,31	76,69	176,00	0,00	42,20	4,60	
г. Москва	76,37	72,31	80,17	11,94	3,26	204,34	529,61	24,06	39,29	51,44	863,83	995,90	6,63	93,37	73,00	17,81	2,50	5,50	
г. Санкт-Петербург	74,22	69,43	78,38	26,56	6,69	252,43	697,66	39,17	45,71	83,96	1152,18	537,00	13,46	86,54	1,00	0,00	0,00	0,00	
Еврейская автономная область	64,94	58,94	71,68	45,47	34,40	225,62	774,80	49,55	83,95	209,30	1423,09	37,10	65,23	34,77	277,00	19,13	19,30	17,70	
Забайкальский край	67,11	61,47	73,10	24,43	14,09	172,50	566,45	71,38	52,89	214,04	1115,78	244,30	52,03	47,97	1287,00	3,57	15,70	7,20	
Ивановская область	69,84	63,90	75,42	18,07	5,64	222,12	645,38	95,22	113,57	124,28	1223,27	135,90	22,37	77,63	6990,00	30,31	45,80	52,50	
Иркутская область	66,72	60,32	73,28	61,94	32,81	205,65	680,69	72,81	75,41	192,01	1321,31	974,00	70,42	29,58	1922,00	4,89	23,20	6,60	
Кабардино-Балкарская Республика	73,71	69,03	78,08	13,98	9,90	142,54	552,81	24,46	37,03	71,04	851,75	62,70	3,51	96,49	0,00	0,00	0,00	0,00	
Калининградская область	70,51	65,10	75,68	21,07	8,24	207,21	721,31	33,58	62,99	135,15	1189,53	139,20	14,73	85,27	347,00	19,31	28,60	9,60	
Калужская область	70,02	64,43	75,51	15,92	7,96	239,58	885,61	53,73	81,29	147,15	1431,24	118,20	12,94	87,06	4090,00	31,49	17,50	35,40	
Камчатский край	67,98	62,59	74,07	18,11	14,99	183,95	611,49	29,04	57,15	138,66	1053,39	90,50	36,13	63,87	46,00	26,26	0,00	18,60	
Карачаево-Черкесская Республика	73,94	69,21	78,33	5,52	2,97	135,50	567,08	32,07	38,66	81,98	863,78	63,80	33,07	66,93	22,00	0,00	12,50	36,40	
Кемеровская область	67,72	61,50	74,04	50,14	27,50	276,85	647,73	61,17	68,69	192,78	1274,87	1575,70	86,08	13,92	778,00	26,24	22,70	24,80	
Кировская область	70,28	64,31	76,29	9,51	6,24	214,30	888,23	72,09	57,26	175,06	1422,69	216,90	47,63	52,37	205,00	0,00	35,00	25,40	
Костромская область	69,86	64,31	75,29	7,60	3,95	237,62	924,82	48,81	61,58	128,64	1413,24	106,30	47,22	52,78	640,00	9,44	26,60	33,60	
Краснодарский край	72,28	67,16	77,27	15,35	8,68	196,74	675,22	43,75	60,13	98,33	1099,19	732,90	28,02	71,98	156,00	12,20	3,90	12,30	
Красноярский край	69,06	63,99	74,77	32,46	19,16	610,57	610,57	69,52	69,27	164,37	1192,88	2812,20	88,80	11,20	6501,00	36,44	36,00	9,90	
Курганская область	68,27	61,89	74,97	42,20	31,43	270,58	667,78	84,86	62,97	192,64	1352,46	138,10	39,61	60,39	1158,00	7,60	43,70	16,30	
Курганская область	70,14	64,27	76,00	14,03	11,53	228,58	779,65	68,72	67,73	118,58	1288,81	141,70	26,75	73,25	4694,00	20,13	3,00	3,80	
Ленинградская область	70,36	64,73	76,05	31,47	11,44	224,75	813,75	46,37	73,06	168,02	1368,85	418,30	58,50	41,50	717,00	29,57	54,10	38,70	
Липецкая область	70,66	64,56	76,77	6,55	4,13	199,99	740,46	64,60	56,85	150,73	1223,30	475,30	72,94	27,06	692,00	15,75	37,90	25,90	
Магаданская область	67,12	61,84	72,77	15,86	9,25	210,79	544,49	58,15	62,11	181,06	1081,71	57,10	50,79	49,21	0,00	0,00	62,50	20,80	
Московская область	70,78	65,10	76,30	14,13	6,83	230,78	849,76	42,03	64,07	111,98	1325,56	942,60	21,11	78,89	12376,00	12,79	28,00	31,70	
Мурманская область	70,46	65,15	75,26	13,15	5,67	173,26	639,15	29,91	57,37	102,62	1021,12	338,10	79,80	20,20	26,00	0,00	34,60	18,20	
Ненецкий автономный округ	65,76	60,22	75,21	6,99	2,33	174,80	522,06	25,64	46,61	198,10	976,54	77,90	93,32	6,68	37,00	24,32	16,70	9,10	
Нижегородская область	69,42	63,06	75,74	14,97	9,74	230,58	991,33	44,53	75,78	130,72	1497,65	466,60	26,98	73,02	3473,00	9,39	46,80	49,20	
Новгородская область	67,67	60,89	74,79	16,50	7,21	228,79	1076,01	60,08	91,00	169,03	1648,62	115,40	39,17	60,83	1372,00	52,48	52,40	31,80	
Новосибирская область	70,19	64,29	76,13	44,11	20,88	197,74	747,62	49,15	53,23	138,08	1250,81	506,50	38,64	61,36	226,00	9,73	100,00	0,00	
Омская область	69,74	63,86	75,57	25,08	16,31	213,02	722,09	80,65	53,04	153,95	1264,14	376,10	56,79	43,21	622,00	4,02	49,50	16,00	
Оренбургская область	69,90	63,10	74,02	36,08	15,31	230,28	770,46	60,08	69,87	147,39	1329,47	763,40	67,17	32,83	111,00	20,72	31,70	7,40	
Орловская область	70,22	64,39	75,92	5,95	2,20	260,57	1024,18	51,62	77,37	123,17	1545,07	121,90	19,52	80,48	1425,00	10,74	11,50	15,30	
Пензенская область	71,54	66,47	77,53	9,16	5,06	208,05	928,77	52,98	70,72	134,47	1499,20	137,00	20,62	79,38	996,00	6,33	19,40	14,40	
Пермский край	68,75	62,61	74,89	29,83	15,14	199,71	770,61	53,09	78,47	174,29	1329,12	637,10	57,76	42,24	1426,00	12,20	28,20	25,70	
Приморский край	68,19	62,77	73,92	33,56	24,45	219,72	729,00	58,67	79,57	155,95	1300,91	406,60	44,49	55,51	2146,00	27,06	38,00	55,60	
Псковская область	67,82	61,81	74,05	15,93	11,38	252,03	1130,57	63,73	91,04	176,77	1741,46	117,00	23,08	76,92	487,00	13,14	26,30	32,20	
Республика Адыгея	71,80	66,56	76,97	16,84	10,55	220,02	788,48	39,96	56,35	103,05	1235,25	48,60	17,70	82,30	119,00	8,40	42,10	4,50	
Республика Алтай	67,34	61,48	73,44	27,49	18,96	154,51	502,39	74,41	44,08	216,12	1037,96	36,60	25,14	74,86	229,00	0,00	7,20	4,80	
Республика Башкортостан	69,63	63,66	75,84	16,70	9,79	146,92	601,64	64,15	52,98	149,97	1042,36	785,10	57,18	42,82	222,00	10,52	16,50	9,60	
Республика Бурятия	67,67	62,32	73,02	23,95	10,90	176,40	557,47	79,56	63,53	215,87	1127,68	213,90	53,34	46,66	1109,00	53,92	12,30	11,50	
Республика Дагестан	75,63	72,31	78,82	8,26	5,52	73,40	234,12	61,20	21,19	50,90	454,77	226,70	7,06	92,94	670,00	17,01	12,90	12,80	
Республика Ингушетия	78,84	75,07	81,32	9,38	7,82	47,59	184,53	7,82	5,81	25,25	288,19	28,20	2,13	97,87	24,00	37,50	43,90	27,20	
Республика Калмыкия	71,35	65,65	77,25	15,90	12,01	162,85	476,20	30,73	50,52	136,36	884,57	40,10	16,71	83,29	140,00	39,29	30,50	24,20	
Республика Карелия	69,19	63,17	75,05	15,89	9,12	235,35	803,74	44,68	69,38	140,33	1318,48	197,00	60,15	39,85	621,00	30,12	40,30	18,10	
Республика Коми	69,27	63,22	75,39	15,29	10,84	197,98	850,47	47,58	89,35	166,94	1066,45	851,40	90,94	9,06	609,00	28,30	59,00	31,10	
Республика Марий Эл	69,30	62,82	76,13	9,28	6,53	179,26	677,45	105,88	71,21	195,65	1245,26	73,10	36,80	63,20	1058,00	8,98	16,70	13,60	
Республика Мордовия	70,56	64,79	76,39	12,76	4,78	184,09	669,91	62,92	53,84	134,67	1121,96	98,70	36,78	63,22	175,00	8,57	34,00	13,80	
Республика Саха (Якутия)	69,13	63,54	75,00	11,83	6,60	126,78	403,70	30,15	42,82	160,50	782,38	262,50	62,90	37,10	167,00	22,75	23,50	13,40	
Республика Северная Осетия-Алания	73,94	68,48	79,08	12,77	9,79	164,95	689,05	21,13	63,40	61,84	1017,92	67,60	7,40	92,60	0,00	0,00	0,00	0,00	
Республика Татарстан	72,12	66,39	77,73	13,19	6,48	176,28	634,12	48,35	51,98	118,99	1047,39	605,00	49,27	50,73	1861,00	12,68	32,00	14,30	
Республика Тыва	61,79	56,97	67,51	67,82	59,46	110,60	362,88	72,00	55,89	338,88	1073,54	38,60	48,96	51,04	545,00	2,57	3,30	4,70	
Республика Хакасия	68,57	62,66	74,14	18,56	13,87	211													



Рисунок 3 – Экранная форма с требованиями к файлу исходных данных (Help - универсального программного интерфейса системы «Эйдос» с внешними базами данных)

5.1.5.2. КОГНИТИВНО-ЦЕЛЕВАЯ СТРУКТУРИЗАЦИЯ И ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ

Далее в соответствии с этапами АСК-анализа выполняем когнитивную структуризацию и формализацию предметной области.

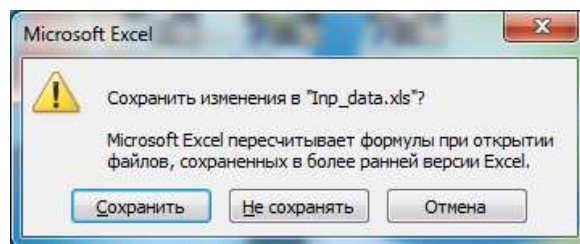
В результате когнитивной структуризации мы решаем, что и на основе чего мы хотим узнать. В данном случае мы хотим выяснить какова сила и направление влияния экологических факторов на продолжительность жизни причины смерти по регионам России.

В таблице 1 соответствующие колонки с классификационными шкалами выделены ярко-желтым и светло-желтым фоном, а с описательными шкалами – светло-зеленым и светло-голубым фоном.

В результате формализация предметной области база исходных данных, представленная в таблице 1, *нормализуется*, т.е. разрабатываются справочники классификационных и описательных шкал и градаций, с использованием которых исходные данные кодируются и создаются база событий (эвентологическая база данных) и обучающая выборка (рисунок 2).

Для автоматической формализации предметной области выполняем в системе «Эйдос» режим 2.3.2.2 с параметрами, указанными на рисунке 4.

После запуска этого режима за заднем фоне за всеми конами появится окно:



на котором нужно выбрать вариант «Сохранить» или «Не сохранять». Это окно появляется потому, что в Excel-файле есть несколько страниц и расчетные колонки.

После этого появляется окно внутреннего калькулятора (рисунок 5), на котором надо задать по 5 градаций и в классификационных, и описательных шкалах и градациях, затем «Пересчитать шкалы и градации» и затем «Выйти на создание модели». Именно 5 градаций на шкалах рекомендуется выбрать в связи с малым объемом выборки, чтобы все интервальные значения были представлены наблюдениями.

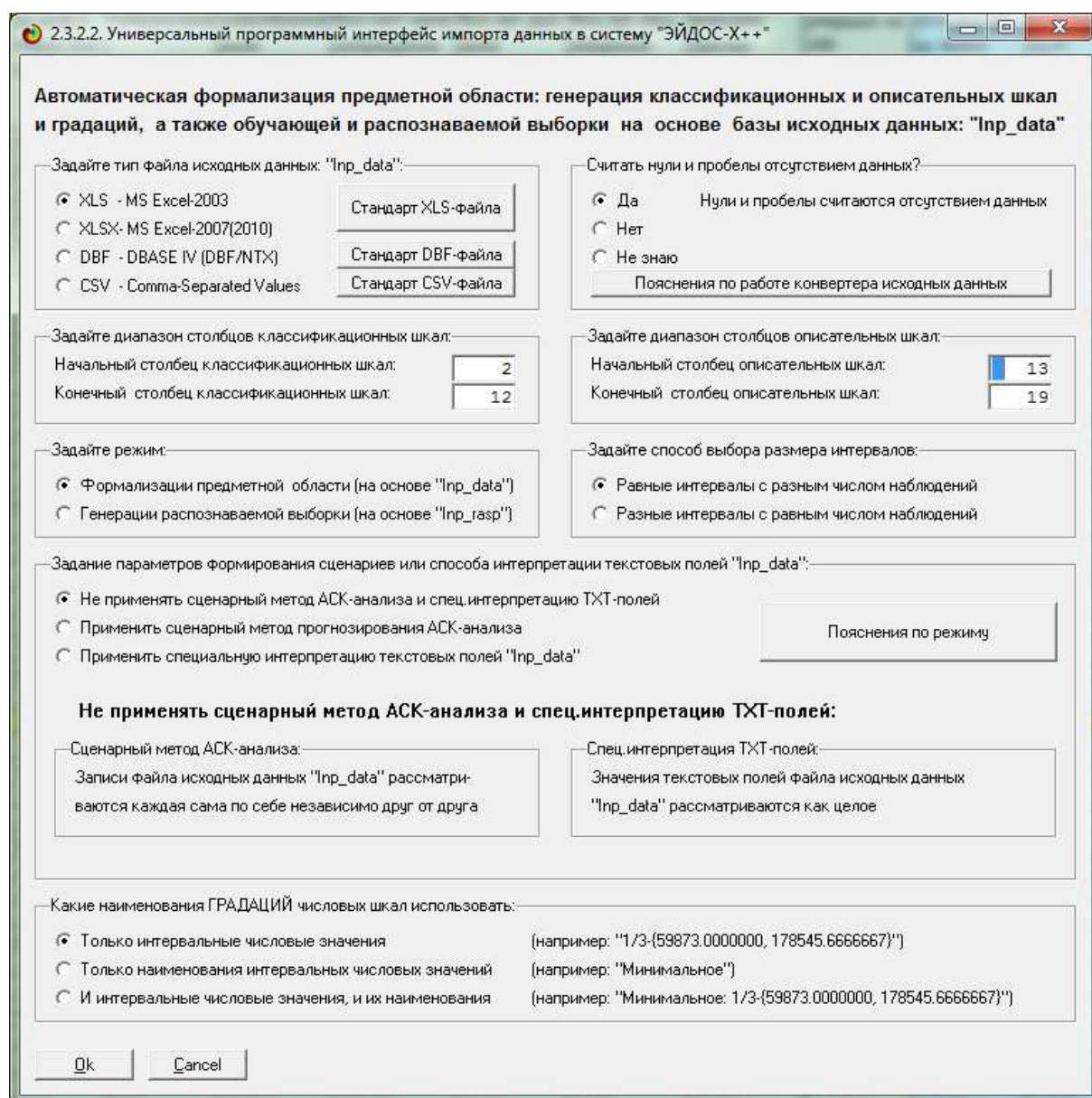


Рисунок 4 – Экранная форма универсального программного интерфейса системы «Эйдос» с внешними базами данных)

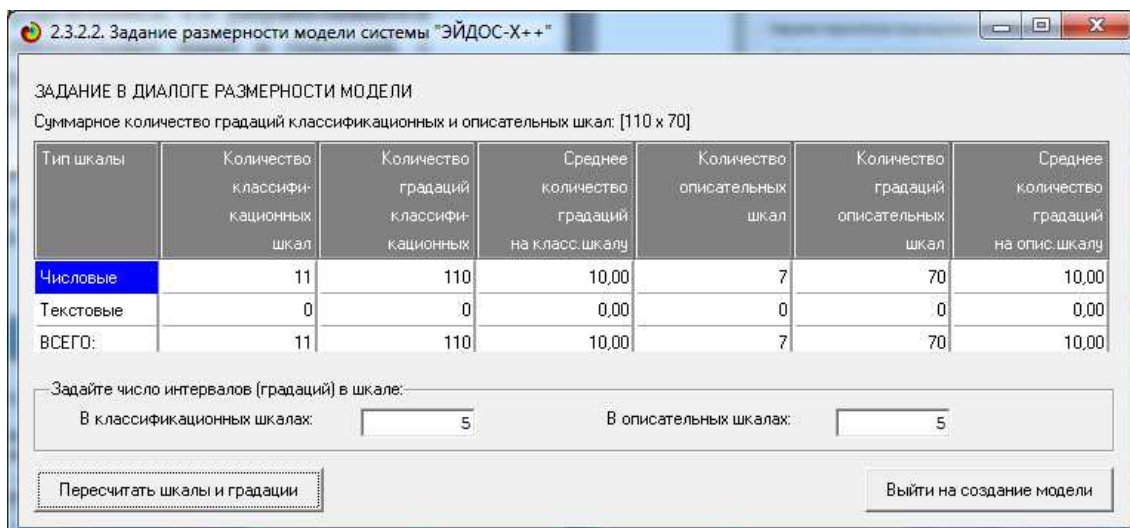


Рисунок 5 – Экранная форма внутреннего калькулятора универсального программного интерфейса системы «Эйдос» с внешними базами данных)

Формализация предметной области происходит быстро (за 2 секунды). При том выполняются этапы, перечисленные на экранной форме отображения стадии исполнения, представленной на рисунке 6:

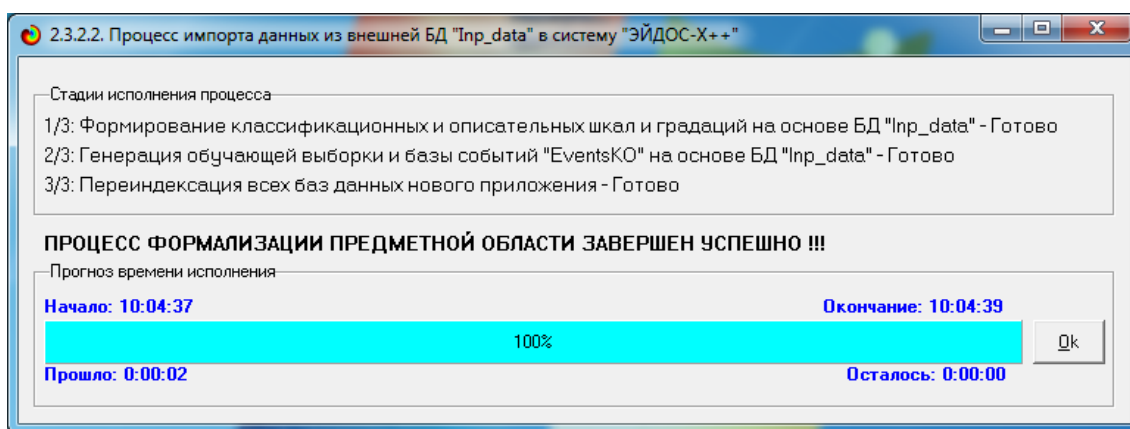


Рисунок 6 – Итоговая экранная форма отображения стадии исполнения этапа формализации предметной области

5.1.5.3. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛИ

Этот этап АСК-анализа выполняется в режиме 3.5 системы «Эйдос» при следующих параметрах (рисунок 7):

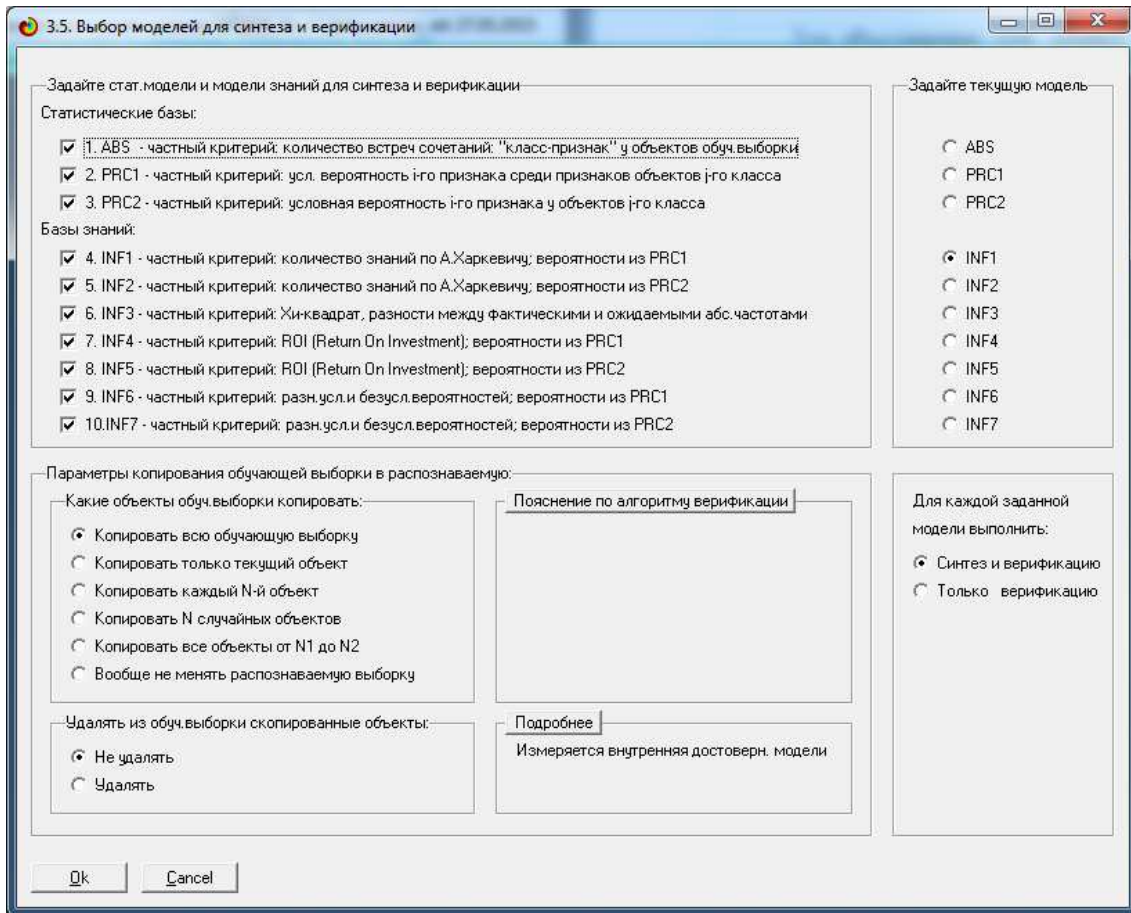


Рисунок 7 – Экранная форма параметров синтеза и верификации модели

На рисунке 8 показана итоговая форма этого режима:

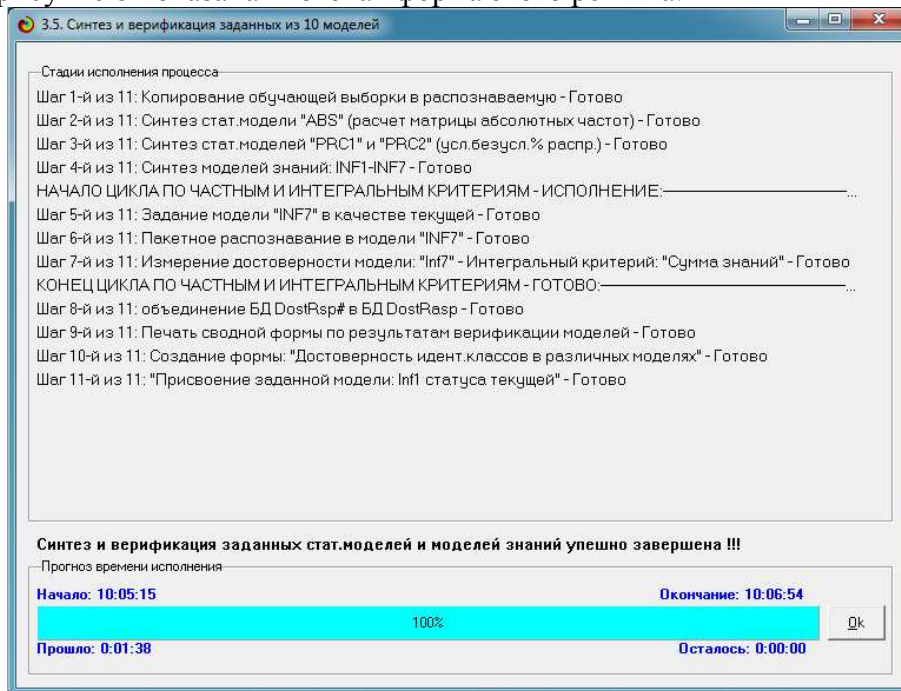


Рисунок 8 – Итоговая экранная форма отображения стадии исполнения этапа синтеза и верификации модели

В результате созданы и верифицированы все модели, приведенные на рисунках 2 и 7. Видно, что этот процесс занял 1 минуту 38 секунд.

Из формы, представленной на рисунке 9, видно, что достоверность моделей знаний достаточно высока для данной предметной области, на основе чего можно говорить об определенной степени зависимости продолжительности жизни и причин смерти от экологической обстановки:

Наименование модели и частного критерия	Интегральный критерий	Вероятность правильной идентификац...	Вероятность правильной не идентиф...	Средняя вероятно... результата	Дата получения результата	Время получения результата
1. ABS - частный критерий: количество встреч сочетаний "клас...	Корреляция абс. частот с обр...	100.000	0.954	50.477	02.06.2015	10:05:31
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма абс. частот по признак...	100.000	0.001	50.001	02.06.2015	10:05:32
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн. частот с о...	100.000	0.954	50.477	02.06.2015	10:05:42
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн. частот по приз...	100.000	0.001	50.001	02.06.2015	10:05:42
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн. частот с о...	100.000	0.954	50.477	02.06.2015	10:05:50
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн. частот по приз...	100.000	0.001	50.001	02.06.2015	10:05:50
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	67.105	72.061	69.583	02.06.2015	10:05:59
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	72.149	64.308	68.229	02.06.2015	10:05:59
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	67.654	72.337	69.995	02.06.2015	10:06:08
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	72.917	63.308	68.112	02.06.2015	10:06:08
6. INF3 - частный критерий: Хинквадрат, разности между фактич...	Семантический резонанс зна...	74.452	63.245	68.848	02.06.2015	10:06:18
6. INF3 - частный критерий: Хинквадрат, разности между фактич...	Сумма знаний	74.452	63.245	68.848	02.06.2015	10:06:18
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	55.263	83.245	69.254	02.06.2015	10:06:26
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	76.974	58.684	67.829	02.06.2015	10:06:26
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	55.154	82.836	68.995	02.06.2015	10:06:35
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	78.947	58.600	68.774	02.06.2015	10:06:35
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	70.395	66.025	68.210	02.06.2015	10:06:44
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	74.452	57.258	65.855	02.06.2015	10:06:44
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	70.943	65.328	68.136	02.06.2015	10:06:53
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	77.632	55.710	66.671	02.06.2015	10:06:53

Рисунок 9 – Экранная форма с оценкой достоверности моделей с разными частными и интегральными критериями на основе предложенной автором метрики, сходной с F-критерием, но не предполагающей нормальность распределения, а лишь интегрально учитывающей верные и ошибочные результаты идентификации и не идентификации

Из экранной формы на рисунке 9 видно также, что модели знаний имеют примерно на 20% более высокую достоверность, чем статистические модели, которые работают по принципу положительного псевдопрогноза (рисунок 10):

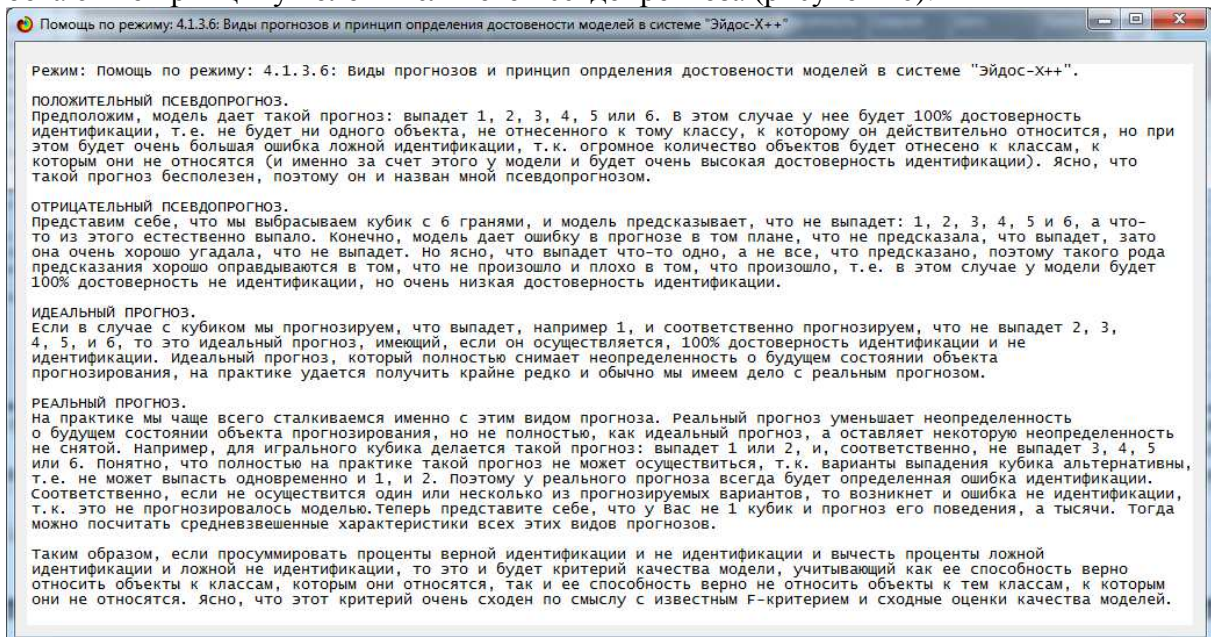


Рисунок 10 – Экранная форма с описанием видов прогнозов с точки зрения оценки их достоверности на основе предложенной автором метрики, сходной с F-критерием

На основе модели Abs (матрица абсолютных частот) принимать решения не целесообразно из-за ее низкой достоверности (рисунок 9) и разного количества примеров по классам (обобщенным категориям).

В модели Prc2 (условные и безусловные процентные распределения) зависимость представленных в модели значений от числа примеров по классам снята, но достоверность у нее такая же низкая, как у Abs. Кроме того для принятия решений на основе той модели необходимо вручную сравнивать значения условных и безусловных вероятностей, что трудоемко и едва ли возможно при больших размерностях моделей.

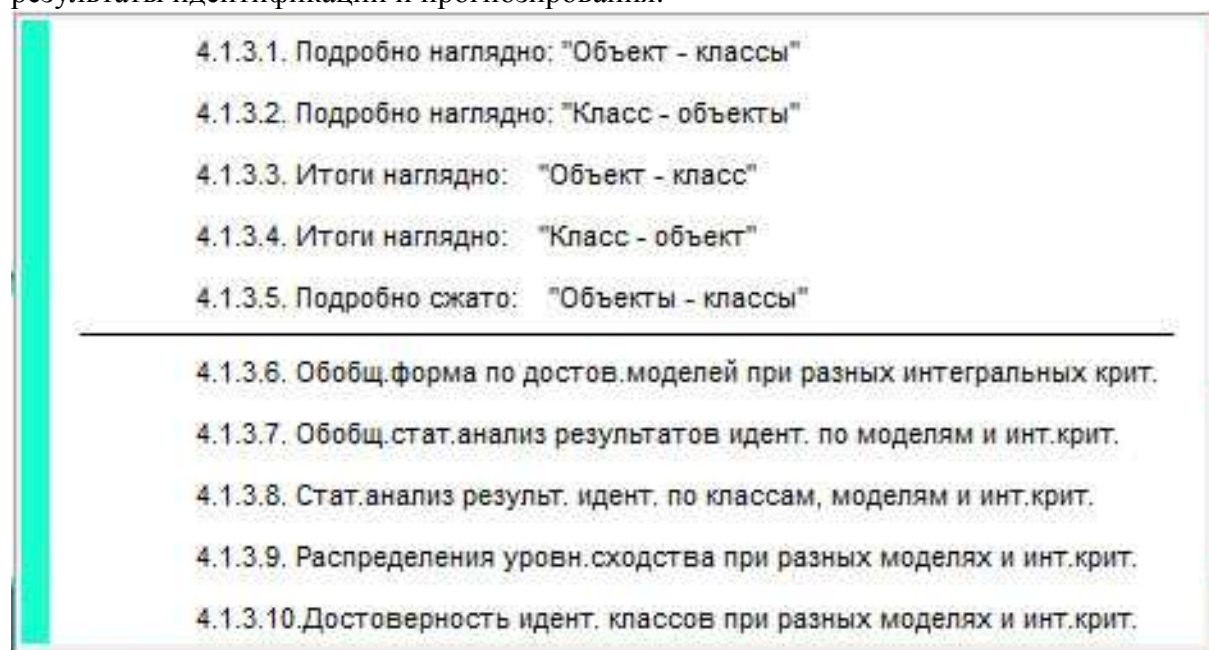
Модель знаний Inf1, основанная на мере А.Харкевича, получена в результате автоматизированного сравнения значения условных и безусловных вероятностей, представленных в модели Prc1, сходной с Prc2, и имеет довольно высокую достоверность⁵.

Поэтому в соответствии с технологией АСК-анализа преобразования данных в информацию, а ее в знания (рисунки 1 и 2) именно модель Inf1 будем использовать для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области, путем исследования ее модели.

5.1.5.4. РЕШЕНИЕ ЗАДАЧ НА ОСНОВЕ СОЗДАННОЙ МОДЕЛИ

Задачи идентификации и прогнозирования отличаются тем, что при идентификации действующие факторы и состояние объекта моделирования относятся к одному моменту времени, а при прогнозировании факторы относятся к настоящему, а состояние объекта, на который они действуют – к будущему. В остальном они практически не отличаются.

В системе «Эйдос» есть много экранных форм и Excel-отчетов, отображающих результаты идентификации и прогнозирования:



из которых на рисунке 14 приведены лишь первые две.

Для их получения в модели Inf1, в соответствии с рисунком 2, сделаем ее текущей (рисунок 12) и проведем в ней идентификацию и прогнозирование (рисунок 13):

⁵ особенно если учесть высокую сложность предметной области, которую мы моделируем

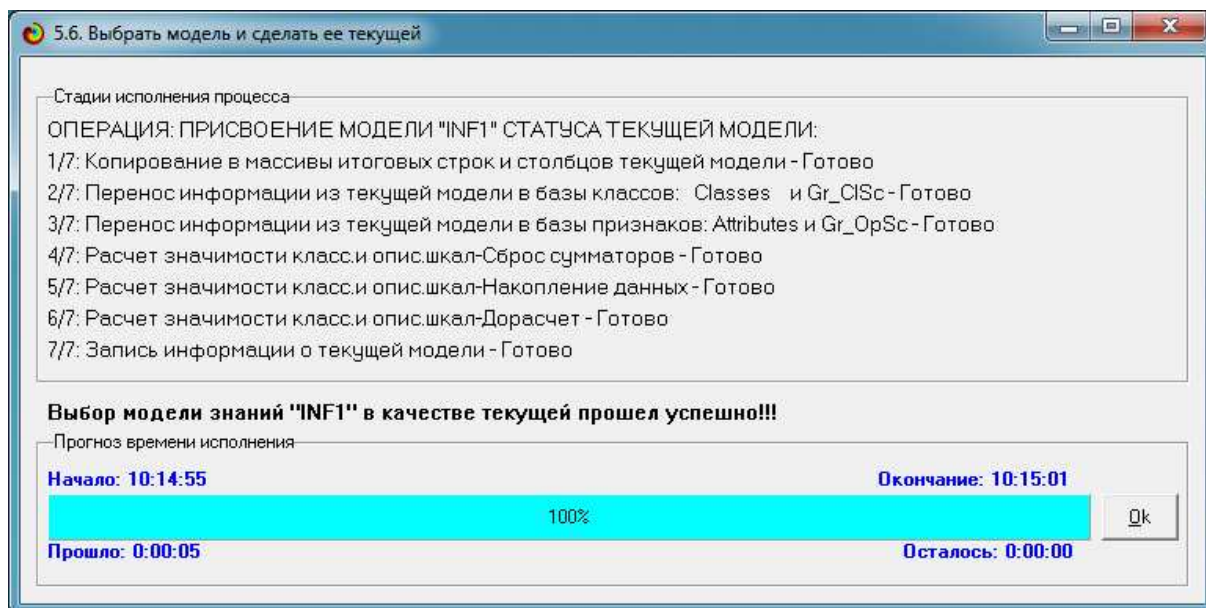


Рисунок 12 – Итоговая экранная форма отображения стадии исполнения придания модели Inf1 статуса текущей модели

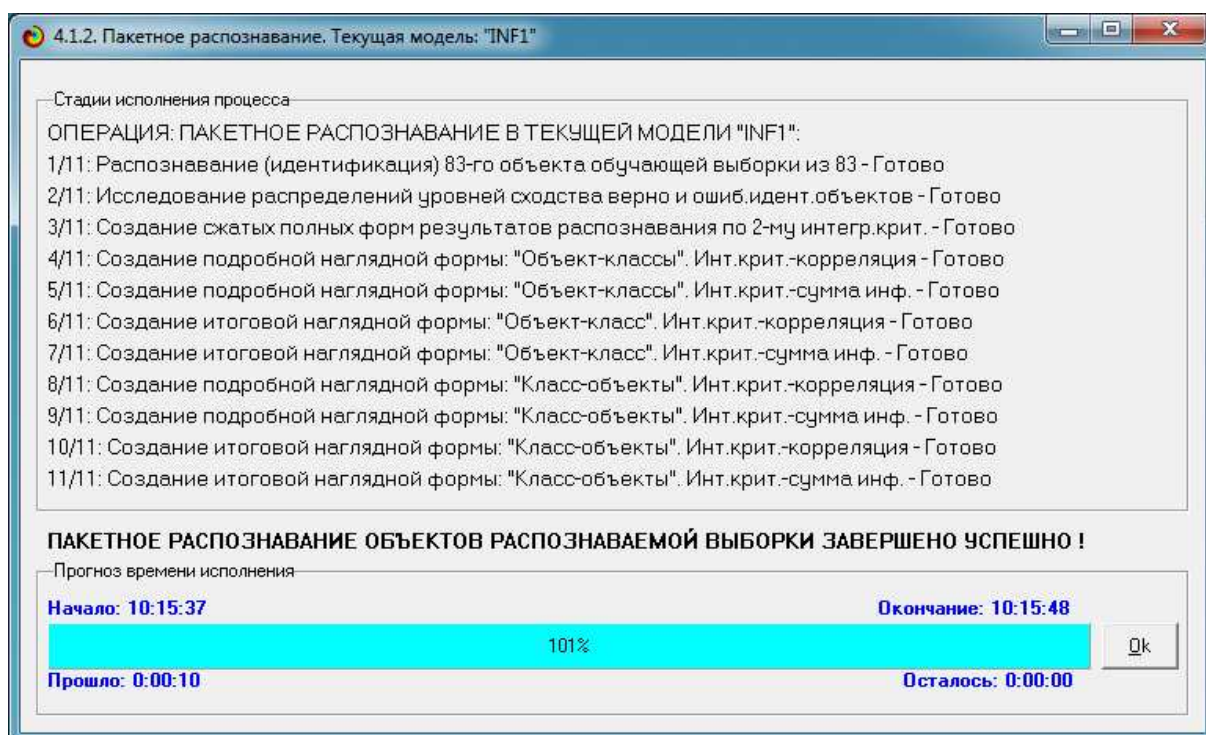


Рисунок 13 – Итоговая экранная форма отображения стадии пакетного распознавания (идентификации, прогнозирования) в модели Inf1

5.1.5.4.1. Идентификация и прогнозирование

Некоторые примеры экранных форм с результатами решения задачи идентификации и прогнозирования приведены на рисунке 14:

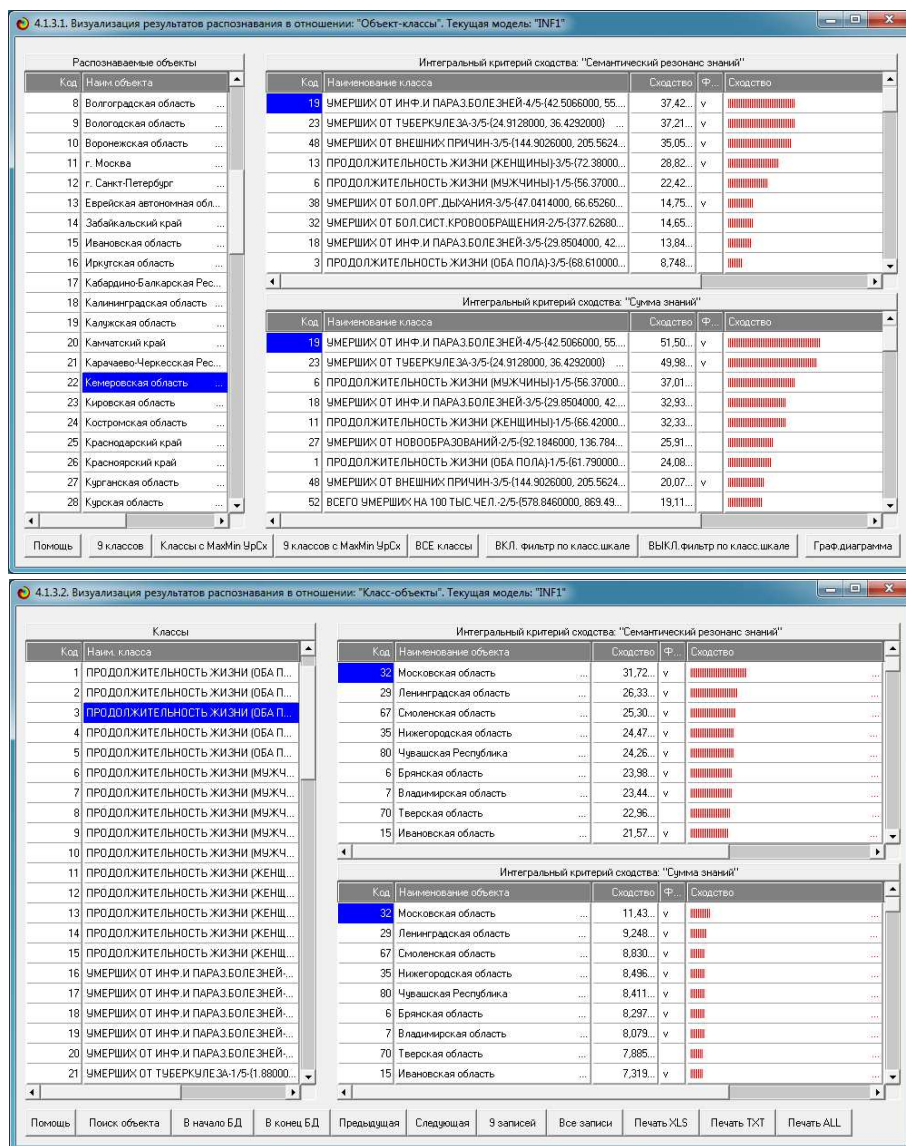


Рисунок 14 – Экранные формы с результатами идентификации и прогнозирования

5.1.5.4.2. Принятие решений

Принятие решения – это задача, *обратная* задаче прогнозирования. Если при прогнозировании мы по значениям действующих факторов определяем будущее состояние объекта управления, то при принятии решений мы наоборот, по будущему состоянию (желательному, целевому, или наоборот нежелательному, а вообще говоря по любому исследуемому) определяем какие значения факторов его обуславливают, т.е. вызывают переход объекта управления в это состояние.

Отметим, что как только мы информацию, содержащуюся в моделях Inf1 – Inf7, начинаем использовать для принятия решений, для достижения целей, для управления, она сразу становится знаниями (см. раздел 4 и рисунки 1 и 2).

В системе «Эйдос» есть много различных режимов и выходных форм, содержащих *знания* для принятия решений. Приведем экранные формы, получающиеся в двух из них (4.2.1 и 4,4,10). Для сопоставимости с результатами, представленными на рисунке 14, выберем класс с кодом 3 для исследования его системы детерминации (рисунки 15 и 16):

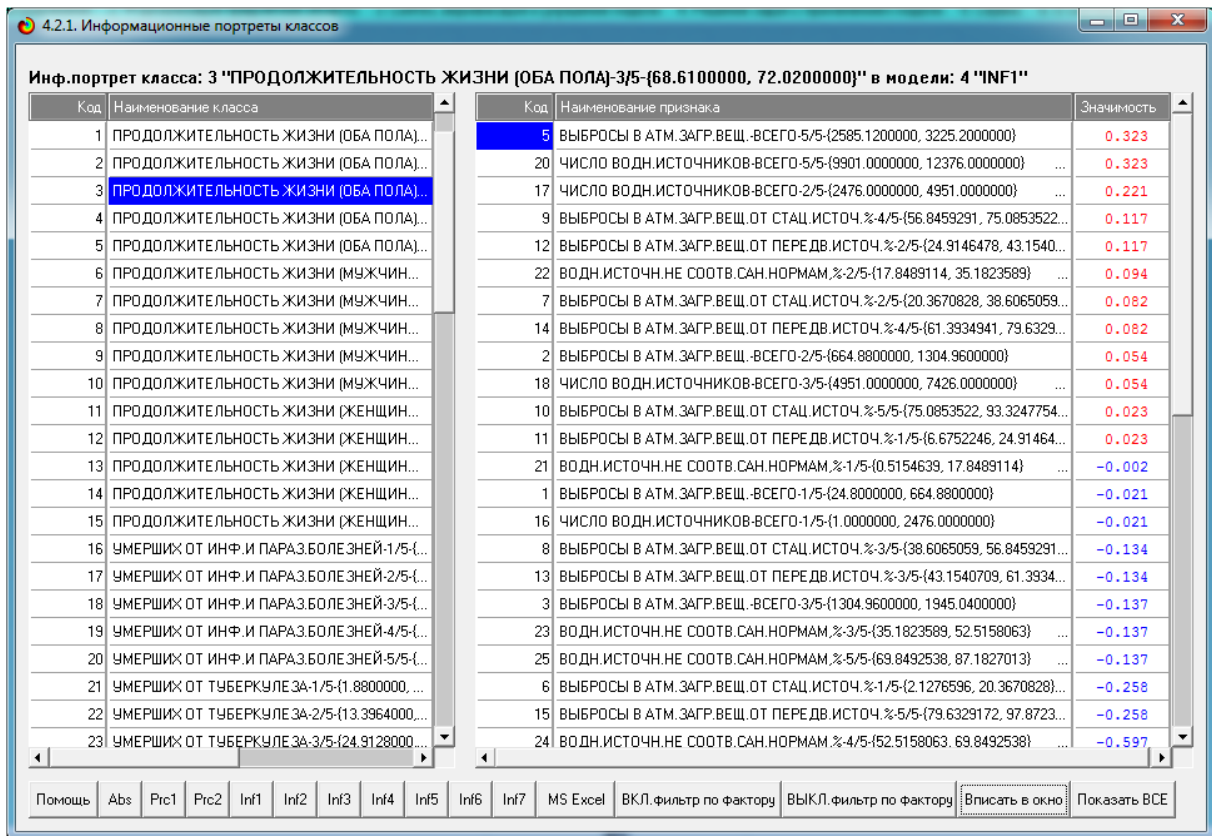
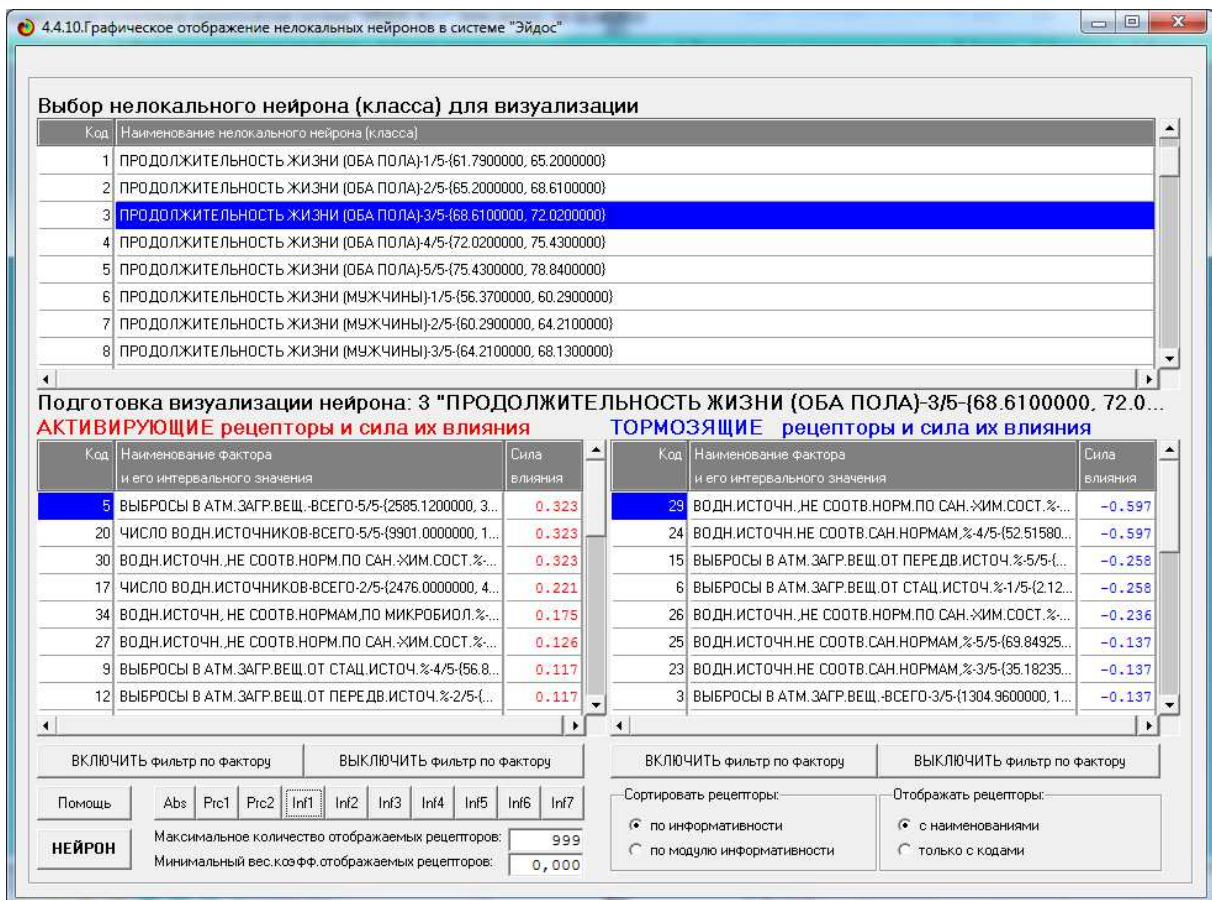


Рисунок 15 – Экранная форма, отображающая систему детерминации класса



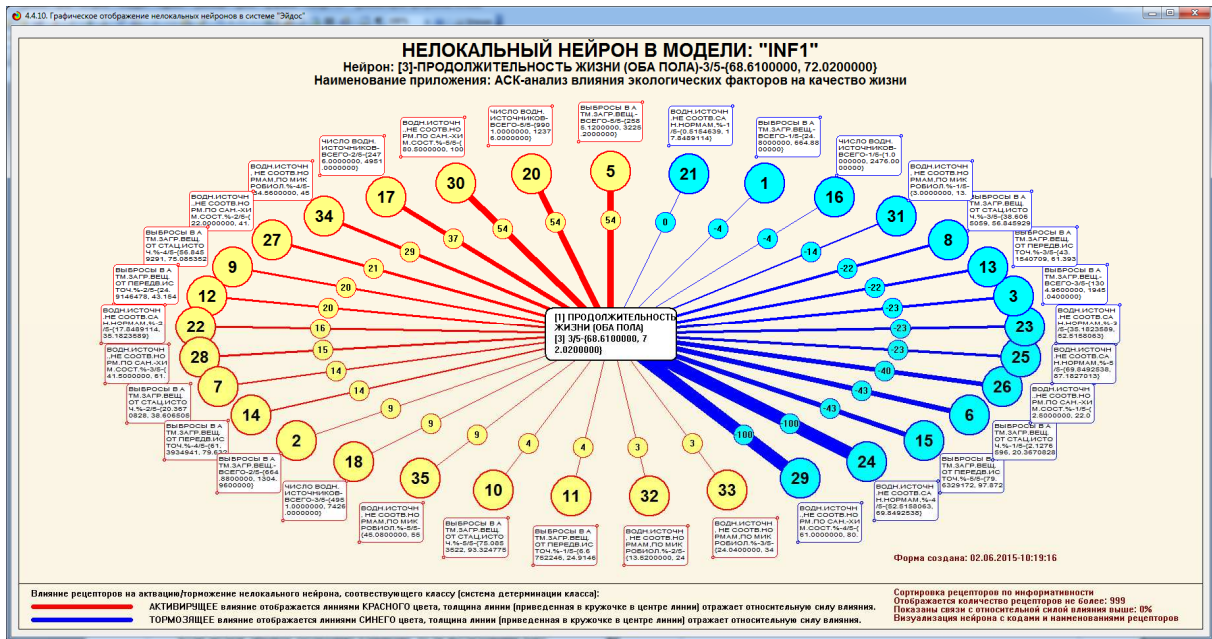


Рисунок 16 – Экранная форма, отображающая систему детерминации класса

На всех формах красным показаны значения факторов, способствующие достижению этого результата, а синим – препятствующие. Сила влияния (значимость) выражена в битах.

5.1.5.4.3. Исследование моделируемого объекта

Если модель объекта достаточно адекватна, то ее исследование корректно считать исследованием самого моделируемого объекта. В нашем случае это именно так.

Прежде всего, приведем Паретто-кривую значимости градаций описательных шкал, т.е. значений экологических факторов (рисунок 17).

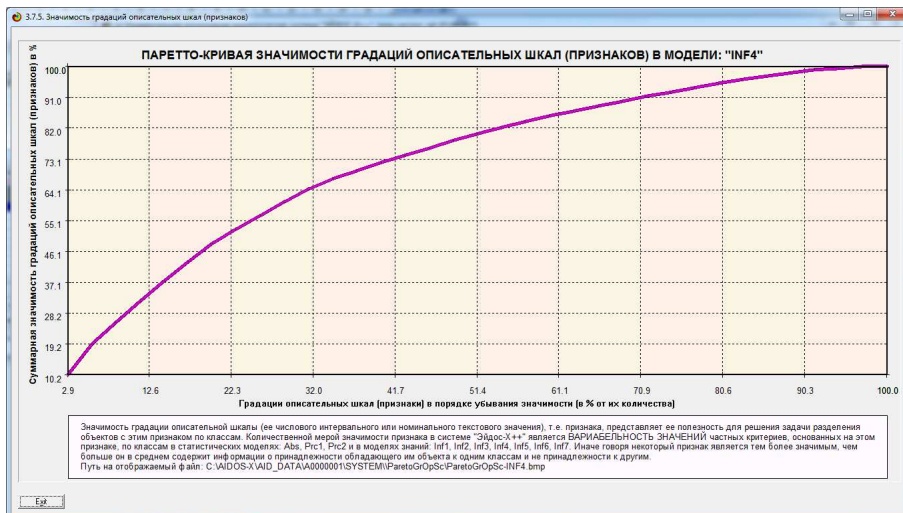


Рисунок 17 – Паретто-кривая значимости значений экологических факторов

И нее хорошо видно, что в модели Inf4 около 50% суммарного влияния всех значений экологических факторов обусловлено всего 15% от их числа, а остальные 85% все вместе играют примерно такую же роль, как этих 15 наиболее сильно влияющих. По-видимому, на этих наиболее сильно влияющих значениях факторов и следует со-

средоточить основное внимание при планировании и проведении экологических мероприятий.

В заключение приведем одну из когнитивных функций (рисунок 18), которая отражает, какое количество информации содержится в значениях аргумента о значениях функции [31].

В когнитивной функции количество информации в аргументе о значении функции отображено цветом. Аппроксимация сделана на основе предложенной автором модификации взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции [32, 33, 34]. Ширина полосы аппроксимации обратно пропорциональна количеству информации.

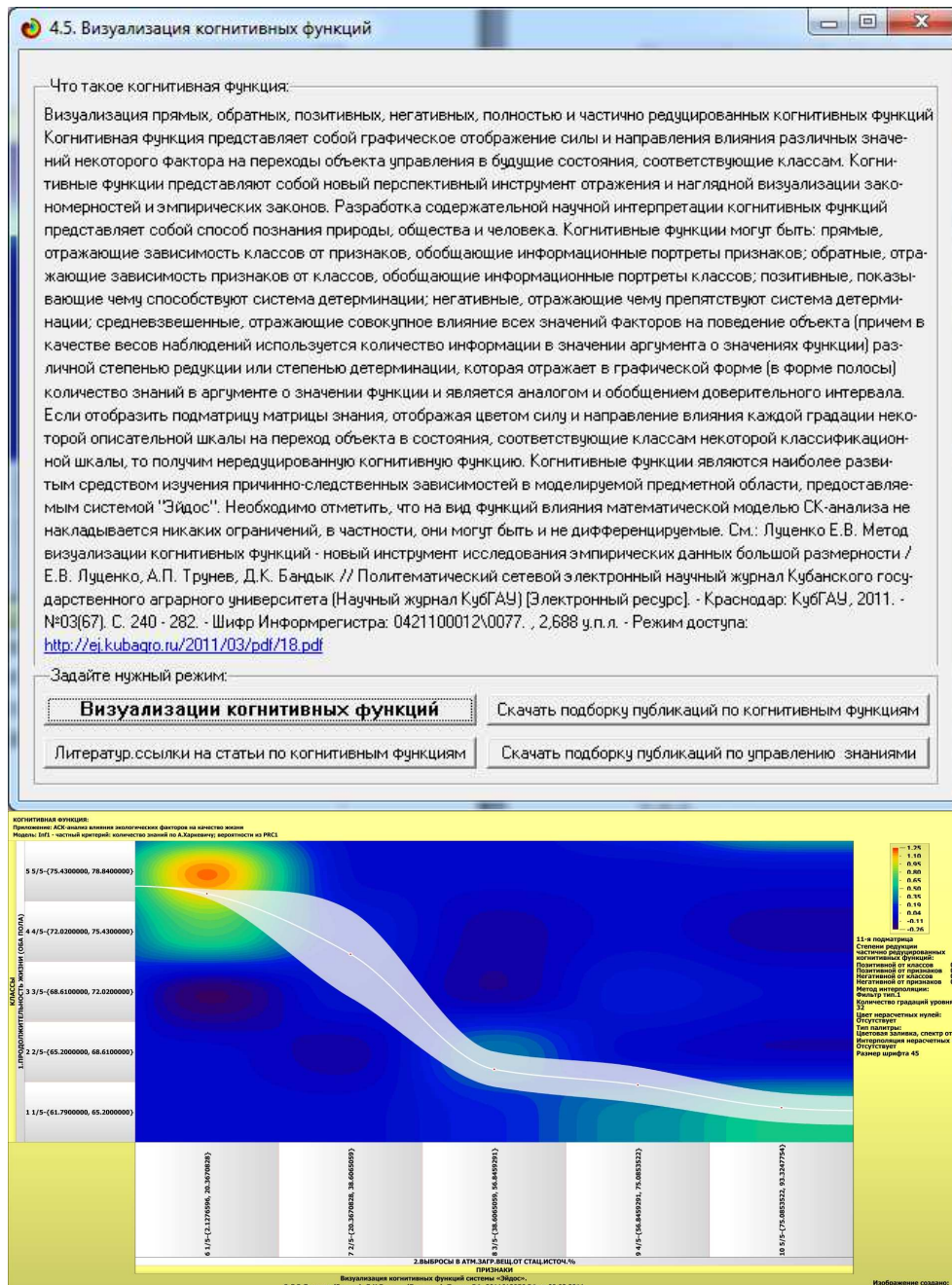


Рисунок 18 – Количество информации о продолжительности жизни (оба пола), в сведениях о доле (от всех выбросов) выбросов загрязняющих веществ в атмосферу от стационарных источников

Видно, что чем больше в общем объеме выбросов *доля* выбросов загрязняющих веществ в атмосферу от стационарных источников, тем меньше продолжительность жизни (по обоим полам), т.е. зависимость между ними практически обратно пропорциональная. *Этот факт можно содержательно объяснить особой опасностью для человека выбросов именно от стационарных источников по сравнению с другими. Отсюда следует научно-обоснованная рекомендация уделить особое внимание уменьшению именно этой категории выбросов. В соответствии с методологией Функционально-стоимостного анализа и «Директ-Костинг» это может быть наиболее эффективными затратами на экологию, от которых есть основания ожидать наиболее ощутимого эффекта (рентабельности) в улучшении здоровья населения.*

Система «Эйдос» обеспечивает автоматизацию SWOT-анализа [35].

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных. Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». В статье на реальном численном примере подробно описывается возможность проведения количественного автоматизированного SWOT-анализа средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» без использования экспертных оценок непосредственно на основе эмпирических данных. Предложено решение прямой и обратной задач SWOT-анализа. PEST-анализ рассматривается как SWOT-анализ, с более детализированной классификацией внешних факторов. Поэтому выводы, полученные в данном разделе на примере SWOT-анализа, можно распространить и на PEST-анализ [35].

Ниже приведены SWOT-матрица и SWOT-диаграмма влияния экологических факторов на достижение высокой продолжительности жизни (рисунок 19).

Когнитивная диаграмма значений экологических факторов отражает их сходство и различие по тем параметрам качества жизни населения региона, которые они обуславливают.

Необходимо отметить, что величина сходства и различия в когнитивных диаграммах получена в результате расчета матриц сходства на основе моделей знаний, а не в результате неформализуемых экспертных оценок на основе интуиции и профессиональной компетенции («на глазок»).

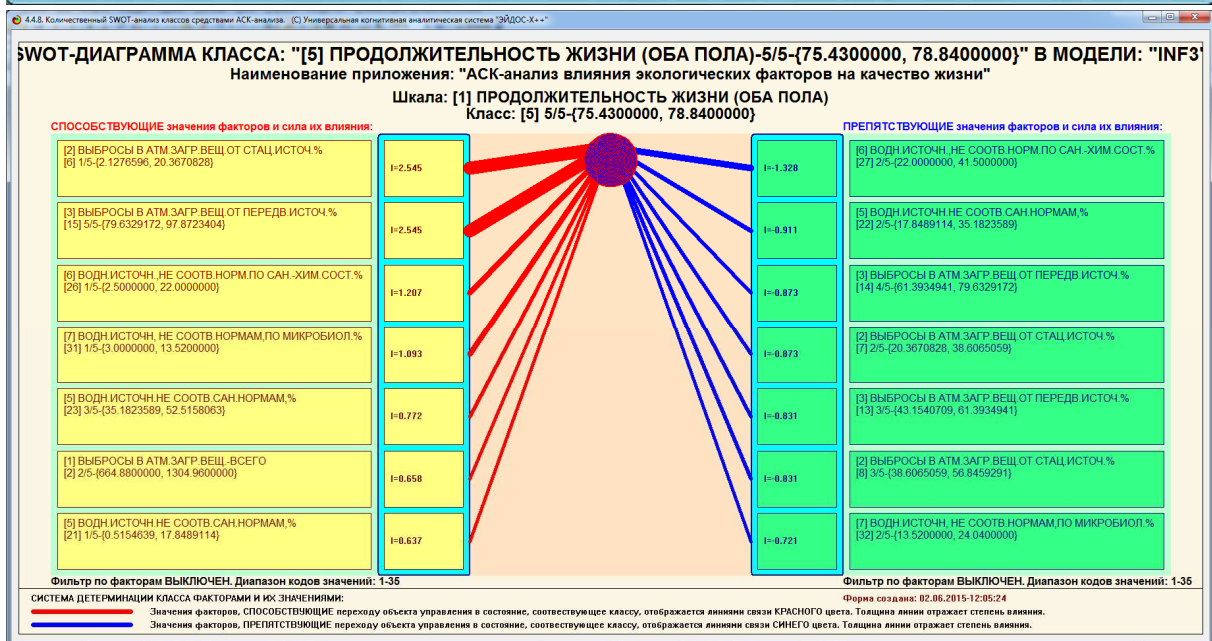
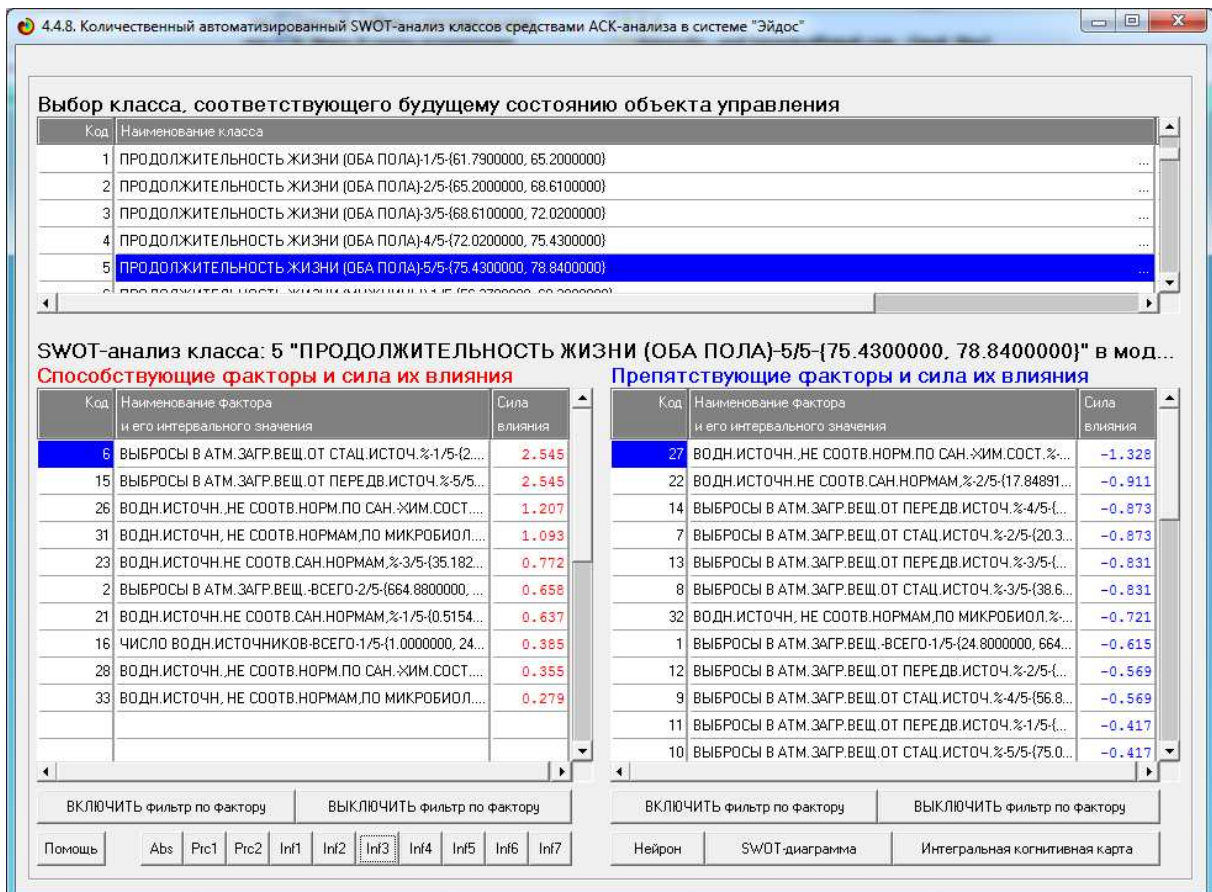


Рисунок 19 – SWOT-матрица и SWOT-диаграмма влияния экологических факторов на достижение высокой продолжительности жизни

На рисунке 20 приведена когнитивная диаграмма классов, а на рисунке 21 – признаков.

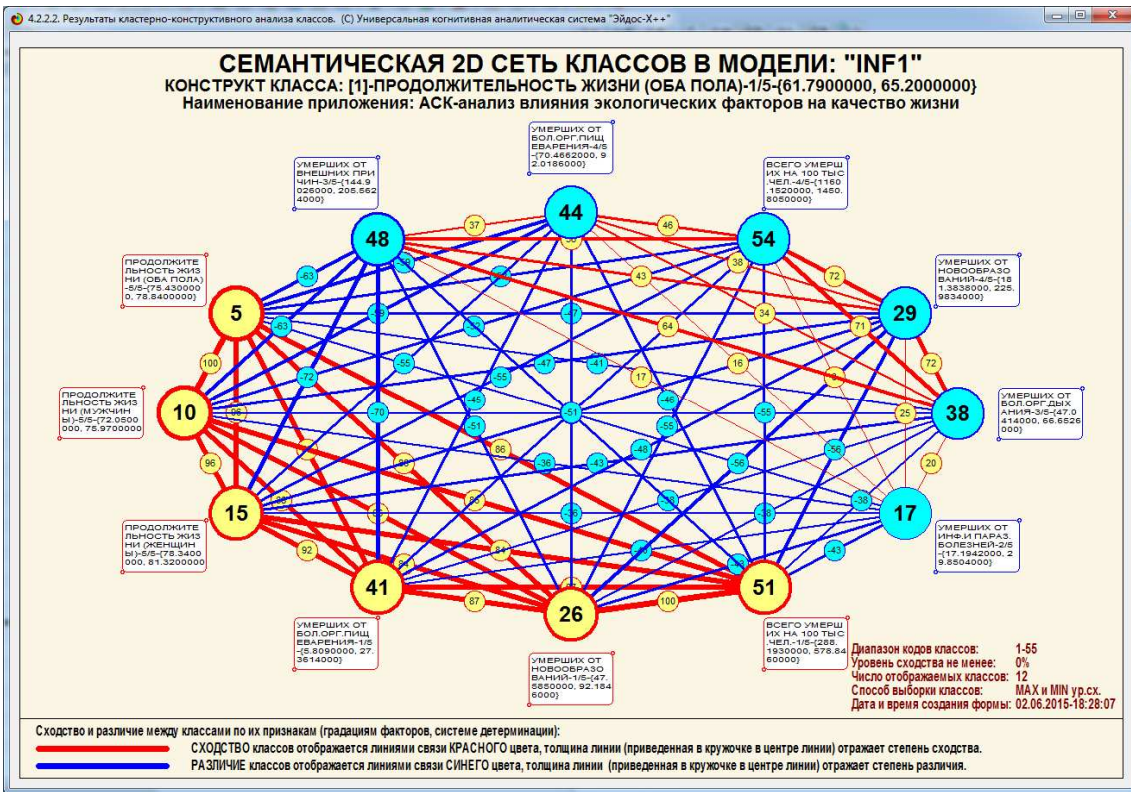


Рисунок 20. Когнитивная диаграмма классов, соответствующих параметрам качества жизни

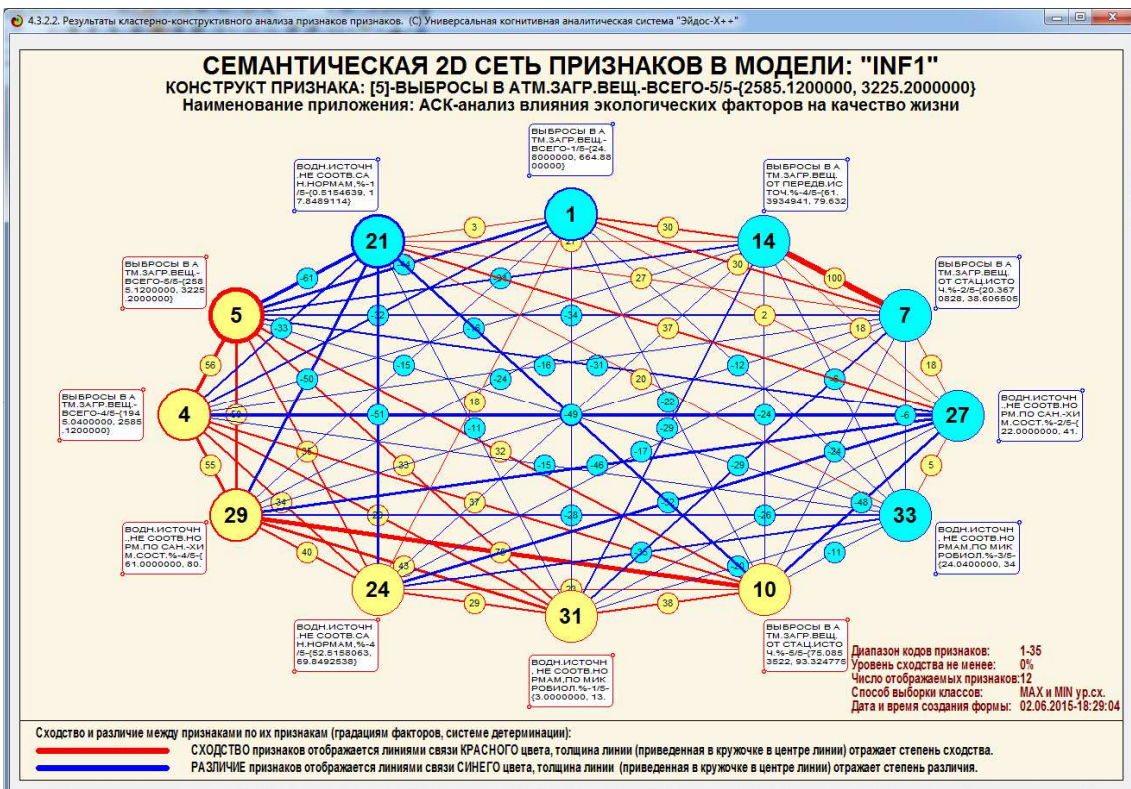


Рисунок 20. Когнитивная диаграмма значений экологических факторов

Когнитивная диаграмма классов отражает их сходство и различие по системе обуславливающих их значений экологических факторов.

5.1.6. Выводы

В статье предложена и продемонстрирована на численном примере возможность исследования влияния экологических факторов на уровень качества жизни региона с применением АСК-анализа и системы «Эйдос».

Отметим, что в статье отражена лишь небольшая доля возможностей исследования моделей, предоставляемых системой «Эйдос».

Материалы данного раздела могут быть использованы при преподавании дисциплин: интеллектуальные системы; инженерия знаний и интеллектуальные системы; интеллектуальные технологии и представление знаний; представление знаний в интеллектуальных системах; основы интеллектуальных систем; введение в нейроматематику и методы нейронных сетей; основы искусственного интеллекта; интеллектуальные технологии в науке и образовании; управление знаниями; автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»; которые автор ведет в настоящее время⁶, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки).

Этим и другим применениям должно способствовать и то, что система «Эйдос» находится в полном открытом бесплатном доступе (с открытыми исходными текстами) на сайте автора по адресу: <http://lc.kubagro.ru/aidos/Aidos-X.htm>.

Данный раздел посвящен краткому описанию возможностей исследования влияния экологических факторов на качество жизни, предоставляемых АСК-анализом и системой «Эйдос». Выше говорилось о том, что совершенно аналогично можно исследовать влияние экологии и на другие самые различные аспекты природы и общества, благо исходные данные достаточные для этого сегодня есть в открытом доступе. Но подробнее рассмотреть собственно экологические результаты в данной работе проблематично из-за ее ограниченного объема и это планируется сделать в последующих публикациях.

5.1.7. Предложения

Предлагается создать региональный центр автоматизированных системно-когнитивных исследований. Область деятельности такого центра может быть весьма широка и позволяет поставить на регулярную основу исследования и разработки по добные тем, которые проводились автором на протяжении многих лет [1-34]⁷.

5.2. Системно-когнитивные модели прогнозирования развития многоотраслевой агропромышленной корпорации

5.2.1. Когнитивно-целевая структуризация и формализация предметной области

В данном разделе в соответствии с методологией Автоматизированного системно-когнитивного анализа (АСК-анализ) рассматривается реализация 1-го и 2-го этапов АСК-анализа: когнитивной структуризации и формализации предметной области. На этапе когнитивной структуризации предметной области, исследователи решают, что рассматривать в качестве объекта моделирования, влияющих на него факторов и результатов их действия. В соответствии с результатами когнитивной структури-

⁶ http://lc.kubagro.ru/My_training_schedule.doc

⁷ <http://ej.kubagro.ru/a/viewaut.asp?id=11>

зации подготавливаются база исходных данных для исследования (обучающая выборка или база прецедентов). На этапе формализации предметной области база исходных данных нормализуются, т.е. разрабатываются классификационные и описательные шкалы и градации, и с их использованием база исходных данных кодируется. В результате формируется база событий (эвентологическая база данных) и обучающая выборка. Этап когнитивной структуризации и подготовки исходных данных не формализован, а этап формализации предметной области полностью автоматизирован и выполняются непосредственно с применением универсальной когнитивной аналитической системы «Эйдос», которая является программным инструментарием АСК-анализа. Этапы когнитивной структуризации и формализации предметной области АСК-анализа являются первыми этапами преобразования данных в информацию, а ее в знания. Последующие этапы: синтез и верификация системно-когнитивной модели, решение задач идентификации, прогнозирования и принятия решений, а также исследования моделируемого объекта путем исследования его модели планируется рассмотреть в будущих работах

Автоматизированный системно-когнитивный анализ (АСК-анализ) предложен в 2002 году проф. Е.В.Луценко [1]. Это инновационный метод искусственного интеллекта, оснащенный общедоступным программным инструментарием, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система «Эйдос» [2]. Скачать и запустить систему «Эйдос» (самую новую на текущий момент версию) или обновление системы до текущей версии всегда можно здесь: <http://lc.kubagro.ru/aidos/Aidos-X.htm>. Это наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с исходными текстами, находящаяся в полном открытом бесплатном доступе (около 50 Мб). Обновление имеет объем около 3 Мб. АСК-анализ – это непараметрический метод, позволяющий исследователю сложные нелинейные объекты управления на основе неполных, зашумленных данных о них большой размерности, измеренных в различных типах шкал и различных единицах измерения [3].

Эти особенности АСК-анализа обусловили его выбор в качестве метода создания системно-когнитивной модели прогнозирования развития многоотраслевой агропромышленной корпорации. Авторы имеют опыт решения подобных задач в АПК [4-19].

АСК-анализ обеспечивает преобразование исходных данных в информацию, а ее в знания [1], и включает следующие этапы:

1. Когнитивная структуризация предметной области.
2. Формализация предметной области.
3. Синтез и верификация системно-когнитивной модели.
4. Решение задач идентификации, прогнозирования и принятия решений.
5. Исследование моделируемого объекта путем исследования его системно-когнитивной модели.

Содержание понятий: «Данные», «Информация» и «Знания» и последовательность преобразования данных в информацию, а ее в знания в АСК-анализе и системе «Эйдос» рассмотрены в ряде работ по АСК-анализу⁸.

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

⁸ Подборка публикаций по этим вопросам: <http://www.twirpx.com/file/793311/>

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ), для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на корпорацию к каким ее изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу [3].

Знания – это информация, полезная для достижения целей.

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).

2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

– вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);

– знания, формализованные в естественном вербальном языке;

– знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);

– знания в форме технологий, организационных, производственных, социально-экономических и политических структур;

– знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно последовательно повышать степень формализации исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

– преобразовать исходные данные в информацию;

– преобразовать информацию в знания;

– использовать знания для решения задач управления, принятия решений и исследования предметной области.

В данной работе в соответствии с методологией Автоматизированного системно-когнитивного анализа (АСК-анализ) рассматривается реализация 1-го и 2-го этапов АСК-анализа: когнитивной структуризации и формализации предметной области. На этапе когнитивной структуризации предметной области исследователи решают, что рассматривать в качестве объекта моделирования, влияющих на него факторов и результатов их действия.

В данном случае в качестве объекта моделирования рассматривается многоотраслевая агропромышленная корпорация («Холдинг»), в качестве факторов, влияющих на ее развитие: Выручка от реализации (Отгрузка), Себестоимость приобретения, Валовая прибыль, Коммерческие расходы, Чистая прибыль, текущие и за 4 последним квартала по предприятиям, входящим в холдинг, а в качестве прогнозируемых показателей – эти показатели по холдингу в целом в двух следующих кварталах.

В соответствии с результатами когнитивной структуризации подготавливаются база исходных данных для исследования (обучающая выборка или база прецедентов) (таблица 1):

Таблица 1 – Исходные данные (фрагмент)

Наименование шкалы	Тип данных	2000г., 1 кв.	2000г., 2 кв.	2000г., 3 кв.	2000г., 4 кв.
В 1-м кв.: ВС.ПО ХОЛД.: Выручка от реализации	N	0,27383403	0,05769639	0,15437844	-0,20388887
В 1-м кв.: ВС.ПО ХОЛД.: Себестоимость приобретения	N	0,28029666	0,05654697	0,14149217	-0,20277262
В 1-м кв.: ВС.ПО ХОЛД.: Валовая прибыль	N	0,34808617	0,01529807	0,25395728	-0,26520805
В 1-м кв.: ВС.ПО ХОЛД.: Коммерческие расходы	N	0,17911279	0,01163287	0,19818931	-0,12430448
В 1-м кв.: ВС.ПО ХОЛД.: Чистая прибыль	N	0,69558329	0,02558452	0,3015067	-0,42765802
В 2-м кв.: ВС.ПО ХОЛД.: Выручка от реализации	N	0,05769639	0,15437844	-0,20388887	0,26426991
В 2-м кв.: ВС.ПО ХОЛД.: Себестоимость приобретения	N	0,05654697	0,14149217	-0,20277262	0,27119177
В 2-м кв.: ВС.ПО ХОЛД.: Валовая прибыль	N	0,01529807	0,25395728	-0,26520805	0,31997283
В 2-м кв.: ВС.ПО ХОЛД.: Коммерческие расходы	N	0,01163287	0,19818931	-0,12430448	0,17289596
В 2-м кв.: ВС.ПО ХОЛД.: Чистая прибыль	N	0,02558452	0,3015067	-0,42765802	0,62070126
001.БАКАЛЕЯ ООО : Выручка от реализации (Отгрузка)	N	59872.904	68605.74	81330.38	90666.871
001.БАКАЛЕЯ ООО : Себестоимость приобретения	N	54663.068	62013.394	73244.71	82504.499
001.БАКАЛЕЯ ООО : Валовая прибыль	N	5162.304	6515.82	7962.462	8074.803
001.БАКАЛЕЯ ООО : Коммерческие расходы:	N	4855.162	5177.562	5708.661	6243.681
001.БАКАЛЕЯ ООО : Чистая прибыль	N	4.937	960.807	1867.996	1432.32
002.РОССИНГРИДГРУПП : Выручка от реализации	N	10875.6	12007.9	12784.1	13784
002.РОССИНГРИДГРУПП : Себестоимость приобретения:	N	9642.754	10829.924	11628.944	12529.951
002.РОССИНГРИДГРУПП : Валовая прибыль	N	1317.49	1289.439	1268.928	1375.051
002.РОССИНГРИДГРУПП : Коммерческие расходы:	N	885.334	934.097	994.073	1055.042
002.РОССИНГРИДГРУПП : Чистая прибыль	N	316.831	231.382	118.053	142.828

003.КОРМИЛИЦА	: Выручка от реализации (Отгрузка)	N	74196.02	87070.587	84697.234	106818.561
003.КОРМИЛИЦА	: Себестоимость приобретения	N	64274.399	78033.995	75411.831	93112.188
003.КОРМИЛИЦА	: Валовая прибыль	N	9489.662	8742.812	8835.316	13144.883
003.КОРМИЛИЦА	: Коммерческие расходы:	N	4584.478	5146.744	5115.603	6287.31
003.КОРМИЛИЦА	: Чистая прибыль	N	3611.378	1945.21	2250.487	4964.347
004.КУБАНЬ АЛКО ООО	: Выручка от реализации (Отгрузка)	N	9907.906	12479.705	17467.6	21390.205
004.КУБАНЬ АЛКО ООО	: Себестоимость приобретения	N	8502.326	10818.892	15205.326	18536.645
004.КУБАНЬ АЛКО ООО	: Валовая прибыль	N	1405.579	1660.813	2256.274	2845.56
004.КУБАНЬ АЛКО ООО	: Коммерческие расходы:	N	964.271	1177.139	1342.223	1542.788
004.КУБАНЬ АЛКО ООО	: Чистая прибыль	N	380.18	406.289	826.097	1210.297
005.КУБТОРГ ЗАО	: Выручка от реализации (Отгрузка)	N	222588.061	261211.761	254091.703	320455.682
005.КУБТОРГ ЗАО	: Себестоимость приобретения	N	176831.595	214111.287	206767.531	254373.733
005.КУБТОРГ ЗАО	: Валовая прибыль	N	17596.183	22349.374	19233.869	28821.405
005.КУБТОРГ ЗАО	: Коммерческие расходы:	N	11946.08	13915.558	13673.261	17032.353
005.КУБТОРГ ЗАО	: Чистая прибыль	N	2349.761	3924.535	714.98	2356.409
006.МОСКВИЧКА ООО	: Выручка от реализации	N	57378.5	75468.3	70557.5	109825.1
006.МОСКВИЧКА ООО	: Себестоимость приобретения:	N	52161.899	68321.089	63439.83	97980.652
006.МОСКВИЧКА ООО	: Валовая прибыль	N	5399.882	7363.734	7406.39	12510.962
006.МОСКВИЧКА ООО	: Коммерческие расходы:	N	3010.511	3939.762	3445.43	4247.655
006.МОСКВИЧКА ООО	: Чистая прибыль	N	2158.976	3174.71	3740.921	7811.79

Исходные данные, приведенные в таблице 1, охватывают период с 2000 по 2009 годы (10-летний лонгитюд), включают 340 строк (классификационных и описательных шкал) и подготовлены в соответствии с требованиями одного из программных интерфейсов системы «Эйдос» с внешними базами данных. Вместо описания этих требований приведем на рисунке 3 Help к этому режиму:

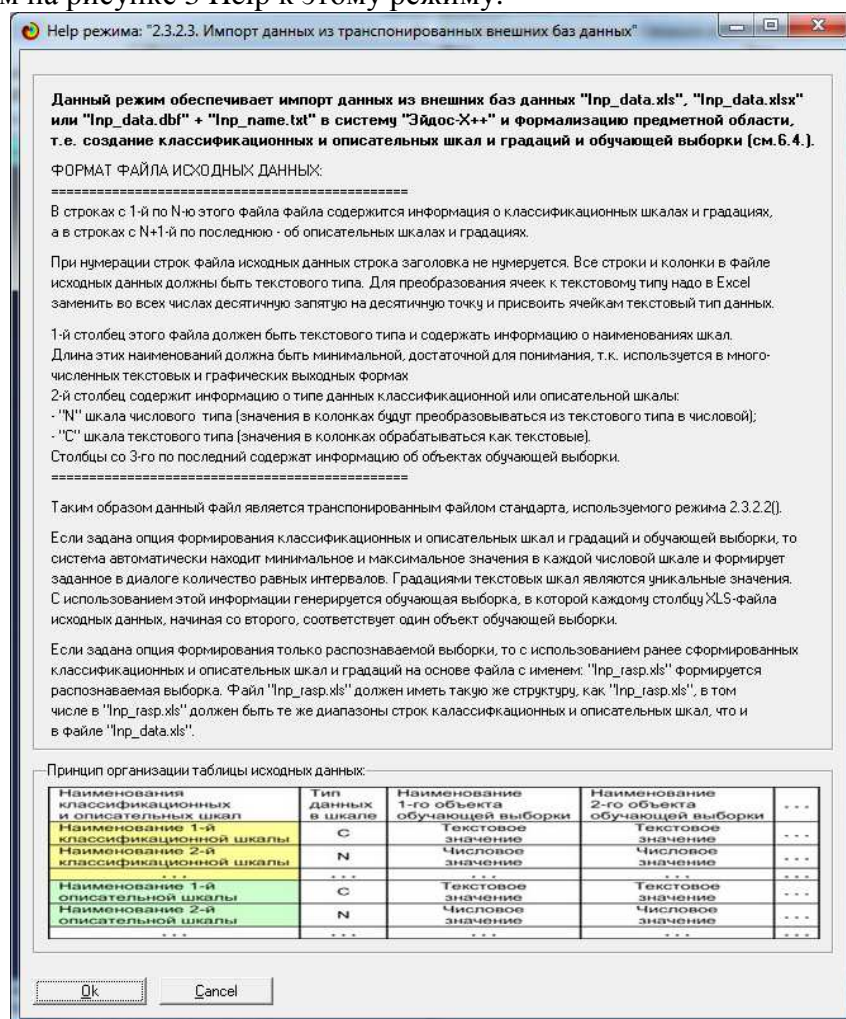


Рисунок 3 – Help программного интерфейса системы «Эйдос», используемого для формализации предметной области на основе исходных данных из таблицы 1

На этапе формализации предметной области база исходных данных нормализуются, т.е. разрабатываются классификационные и описательные шкалы и градации и с их использованием база исходных данных кодируется. В результате формируется база событий (эвентологическая база данных) и обучающая выборка. Этап когнитивной структуризации и подготовки исходных данных не формализован, а этап формализации предметной области полностью автоматизирован и выполняются непосредственно с применением универсальной когнитивной аналитической системы «Эйдос», которая является программным инструментарием АСК-анализа. Этапы когнитивной структуризации и формализации предметной области АСК-анализа являются первыми этапами преобразования данных в информацию, а ее в знания.

На рисунке 4 приведено окно программного интерфейса 2.3.2.3 для задания параметров формализации предметной области:

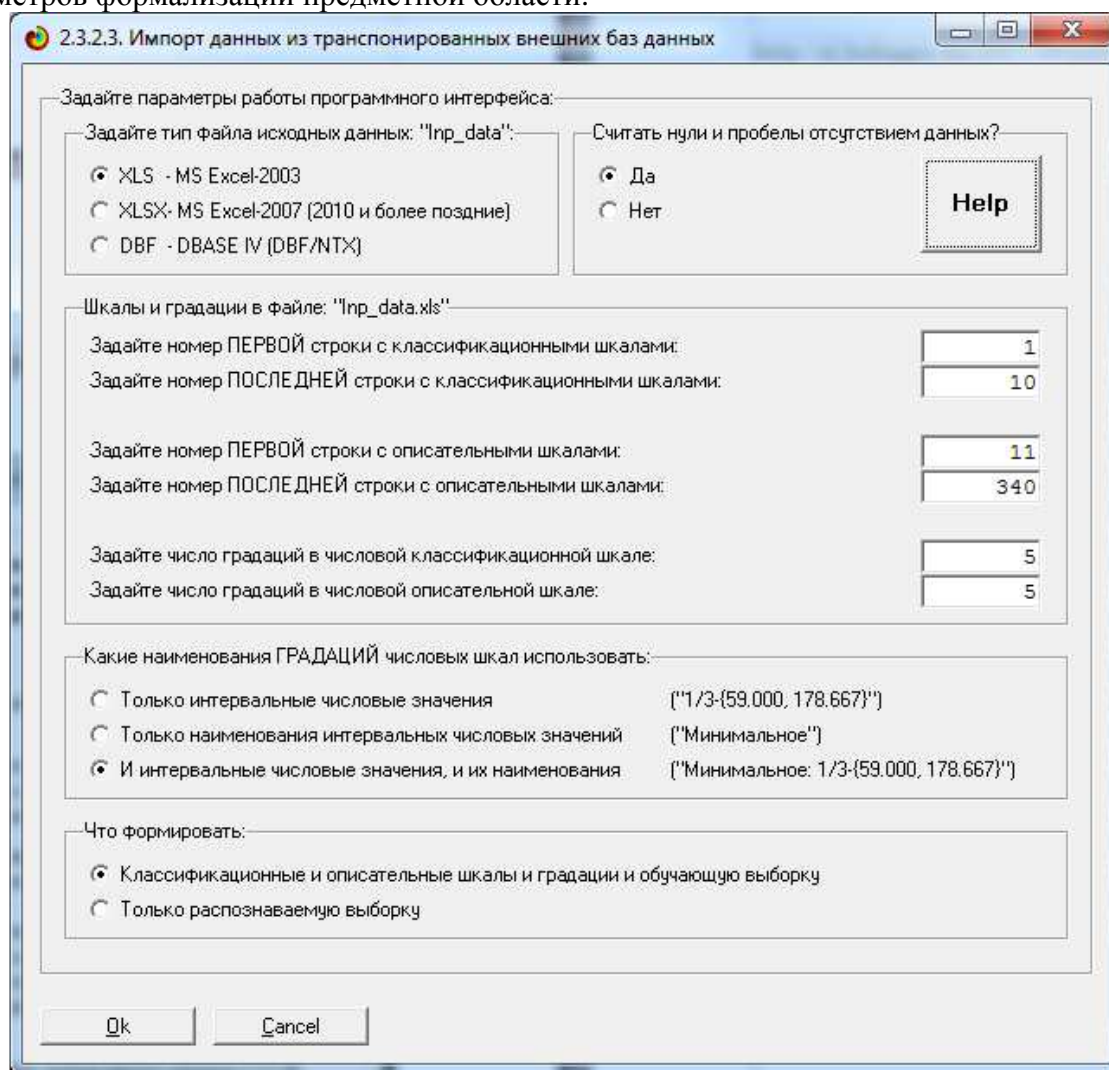


Рисунок 4 – Окно программного интерфейса 2.3.2.3 для задания параметров формализации предметной области

В результате работы данного программного интерфейса созданы классификационные и описательные шкалы и градации и исходные данные, представленные в таблице 1, закодированы с их использованием, в результате чего создана обучающая выборка и база событий (эвентологическая база данных) (рисунки 5, 6, 7):

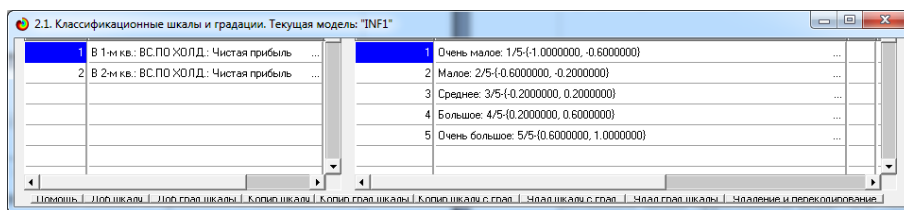


Рисунок 5 – Экранная форма просмотра и корректировки классификационных шкал и градаций

Будущие состояния холдинга формально описываются в модели как градации классификационных шкал, т.е. классы.

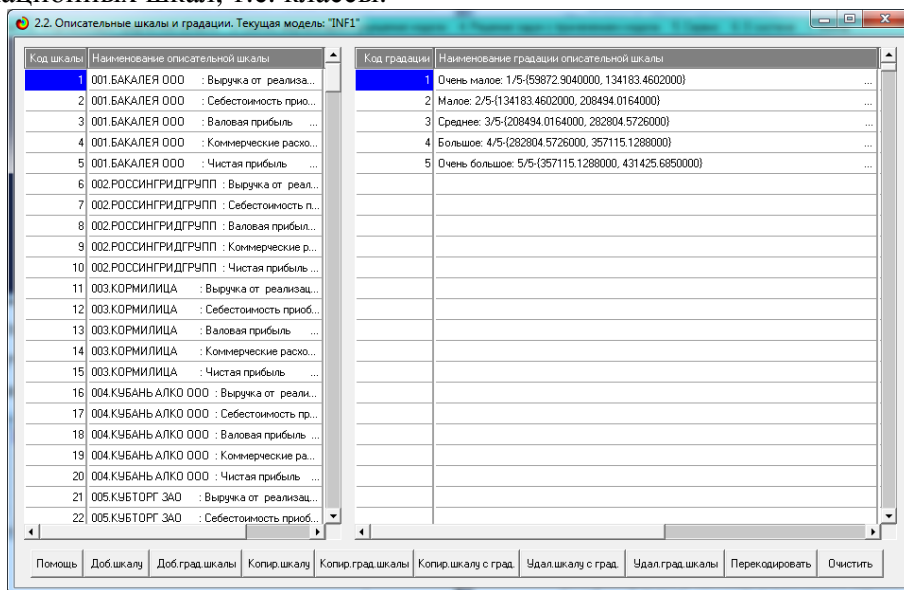


Рисунок 6 – Экранная форма просмотра и корректировки описательных шкал и градаций

Показатели работы предприятий холдинга в прошлый период формально описываются в модели как градации описательных шкал, т.е. факторы и их значения, влияющие на будущие состояния холдинга.

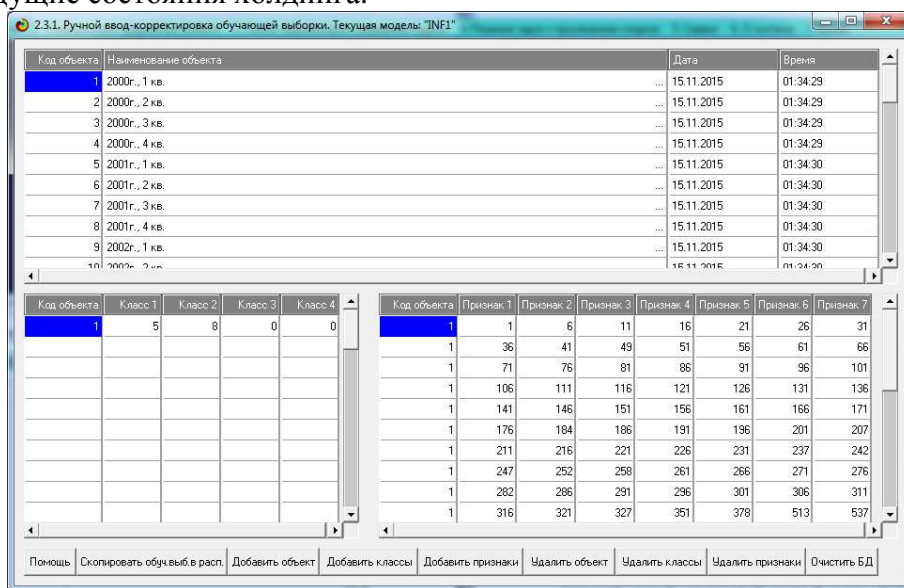


Рисунок 7. Экранная форма просмотра и корректировки обучающей выборки (база событий)

Выводы.

В результате проведенной работы подготовлены все условия для выполнения последующих этапов АСК-анализа: синтез и верификация системно-когнитивной модели, решение задач идентификации, прогнозирования и принятия решений, а также исследования моделируемого объекта путем исследования его модели, которые планируется описать в будущих статьях.

5.2.2. Синтез и верификация модели

В данном разделе в соответствии с методологией Автоматизированного системно-когнитивного анализа (АСК-анализ) рассматривается реализация 3-го АСК-анализа: синтез и верификация модели прогнозирования развития многоотраслевой агропромышленной корпорации. На этом этапе осуществляется синтез и верификация 3 статистических и 7 системно-когнитивных моделей: ABS – матрица абсолютных частот, PRC1 и PRC2 – матрицы условных и безусловных процентных распределений, INF1 и INF2 – частный критерий: количество знаний по А.Харкевичу, INF3 – частный критерий: Хи-квадрат: разности между фактическими и теоретически ожидаемыми абсолютными частотами INF4 и INF5 – частный критерий: ROI - Return On Investment, INF6 и INF7 – частный критерий: разность условной и безусловной вероятностей (коэффициент взаимосвязи). Достоверности созданных моделей оценивались в соответствии с предложенной метрикой, сходной с известным F-критерием, но не предполагающей выполнение нормального распределения, линейности объекта моделирования, независимости и аддитивности действующих на него факторов. Достоверность полученных моделей оказалось достаточно высокой для решения последующих задач идентификации, прогнозирования и принятия решений, а также исследования моделируемого объекта путем исследования его модели, которые планируется рассмотреть в будущих статьях

Автоматизированный системно-когнитивный анализ (АСК-анализ) предложен в 2002 году проф. Е.В.Луценко [1]. Это инновационный метод искусственного интеллекта, оснащенный общедоступным программным инструментарием, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система «Эйдос» [2]. Скачать и запустить систему «Эйдос» (самую новую на текущий момент версию) или обновление системы до текущей версии всегда можно здесь: <http://lc.kubagro.ru/aidos/Aidos-X.htm>. Это наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с исходными текстами, находящаяся в полном открытом бесплатном доступе (около 50 Мб). Обновление имеет объем около 3 Мб. АСК-анализ – это непараметрический метод, позволяющий исследовать сложные нелинейные объекты управления на основе неполных, зашумленных данных о них большой размерности, измеренных в различных типах шкал и различных единицах измерения [3].

Эти особенности АСК-анализа обусловили его выбор в качестве метода создания системно-когнитивной модели прогнозирования развития многоотраслевой агропромышленной корпорации. АСК-анализ включает следующие этапы:

1. Когнитивная структуризация предметной области.
2. Формализация предметной области.
3. Синтез и верификация системно-когнитивной модели.
4. Решение задач идентификации, прогнозирования и принятия решений.
5. Исследование моделируемого объекта путем исследования его системно-когнитивной модели.

В работе [4] описана реализация первых двух этапов АСК-анализа при системно-когнитивные модели прогнозирования развития многоотраслевой агропромышленной корпорации. В данном разделе рассмотрим третий и четвертый этапы. Авторы имеют опыт решения подобных задач в АПК [5-20]. Для синтеза и верификации системно-когнитивной модели запустим с параметрами по умолчанию режим 3.5 системы «Эйдос» (рисунок 1):

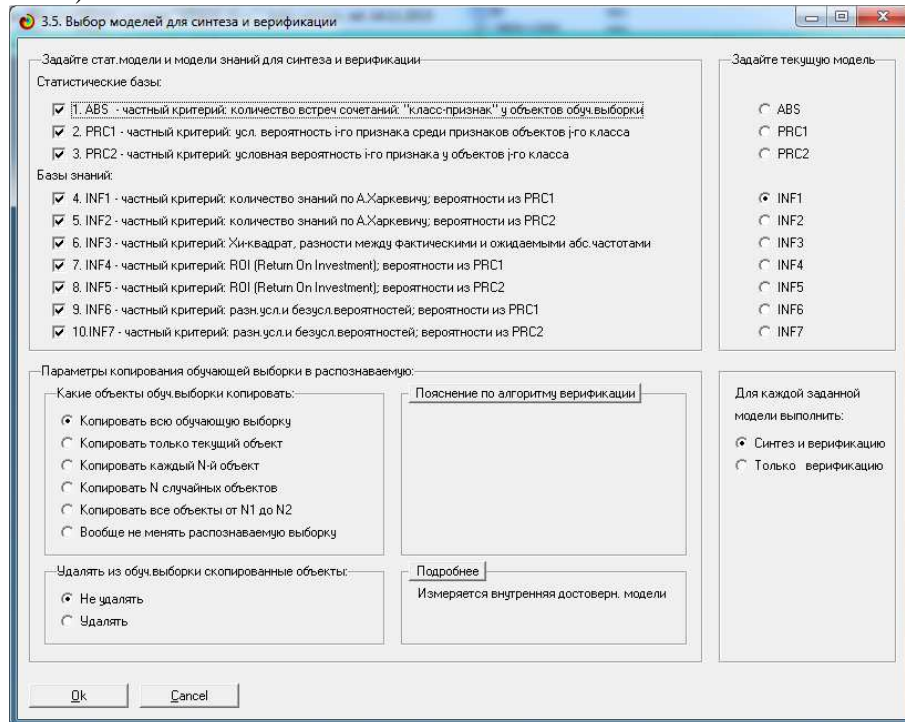


Рисунок 1. Экранная форма задания параметров работы режима синтеза и верификации моделей системы «Эйдос»

На рисунке 2 приведена экранная форма отображения стадии исполнения и прогнозирования времени окончания работы данного режима:

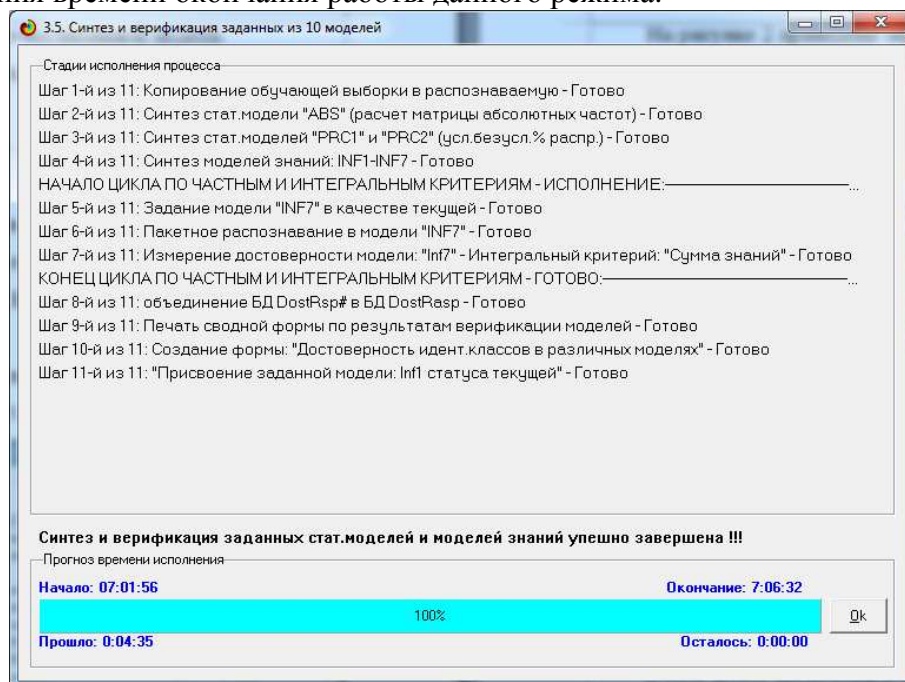


Рисунок 2. Экранная форма задания параметров работы режима синтеза и верификации моделей системы «Эйдос»

В результате работы данного режима перечисленные модели сначала создаются, а потом проверяются на достоверность путем идентификации объектов обучающей выборки и подсчета количества ошибок первого и второго рода (ложной идентификации и неидентификации). На рисунке 3 приведена информация о достоверности созданных моделей в соответствии с предложенной проф. Е.В. Луценко метрикой, сходной с известным F-критерием, но не предполагающей выполнение нормального распределения, линейности объекта моделирования, независимости и аддитивности действующих на него факторов:

Наименование модели и частного критерия	Интегральный критерий	Вероятность правильной идентификации	Вероятность правильной неидентиф.	Средняя вероятность правильного результата	Дата получения результата	Время получения результата
1.ABS - частный критерий: количество встреч сочетаний "клас...	Сумма абс. частот по признак...	100.000	50.000	60.007	15.11.2015	07:05:37
2.PRC1 - частный критерий: усл. вероятность его признака сред...	Корреляция усл.отн. частот с о...	88.750	31.265	60.007	15.11.2015	07:05:42
2.PRC1 - частный критерий: усл. вероятность его признака сред...	Сумма усл.отн. частот по при...	100.000	50.000	60.007	15.11.2015	07:05:42
3.PRC2 - частный критерий: условная вероятность его признака...	Корреляция усл.отн. частот с о...	88.750	31.265	60.007	15.11.2015	07:05:46
3.PRC2 - частный критерий: условная вероятность его признака...	Сумма усл.отн. частот по при...	100.000	50.000	60.007	15.11.2015	07:05:49
4.INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	78.750	75.415	77.082	15.11.2015	07:05:55
4.INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	78.750	61.720	70.235	15.11.2015	07:05:55
5.INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	78.750	77.006	77.878	15.11.2015	07:06:01
5.INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	80.000	68.709	74.355	15.11.2015	07:06:01
6.INF3 - частный критерий: Хинквадрат, разности между фактич...	Семантический резонанс зна...	81.250	69.170	75.210	15.11.2015	07:06:07
6.INF3 - частный критерий: Хинквадрат, разности между фактич...	Сумма знаний	81.250	69.170	75.210	15.11.2015	07:06:07
7.INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	75.000	91.979	83.490	15.11.2015	07:06:13
7.INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	80.000	60.468	70.234	15.11.2015	07:06:13
8.INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	75.000	92.289	83.644	15.11.2015	07:06:19
8.INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	85.000	60.601	72.801	15.11.2015	07:06:19
9.INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	81.250	69.921	75.586	15.11.2015	07:06:25
9.INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	81.250	62.465	71.857	15.11.2015	07:06:25
10.INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	78.750	70.186	74.468	15.11.2015	07:06:31
10.INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	85.000	59.003	72.002	15.11.2015	07:06:31

Рисунок 3. Достоверность созданных моделей

На рисунке 4 приведена экранная форма Help данного режима, поясняющая смысл используемой метрики оценки достоверности:

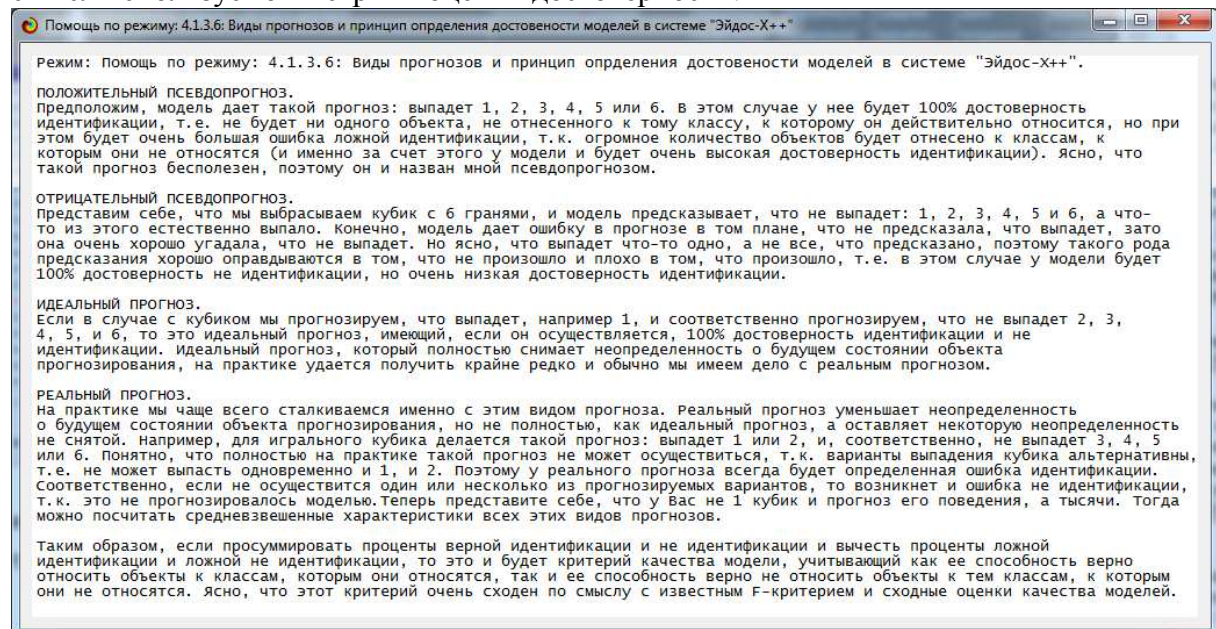


Рисунок 4. Смысл используемой метрики оценки достоверности

Из рисунка 3 видно, что наиболее достоверной оказалась модель INF5. В соответствии со схемой преобразования данных в информацию, а ее в знания в АСК-

анализе, зададим эту модель в качестве текущей (рисунок 7) и выполним пакетное распознавание в наиболее достоверной модели (рисунок 8):

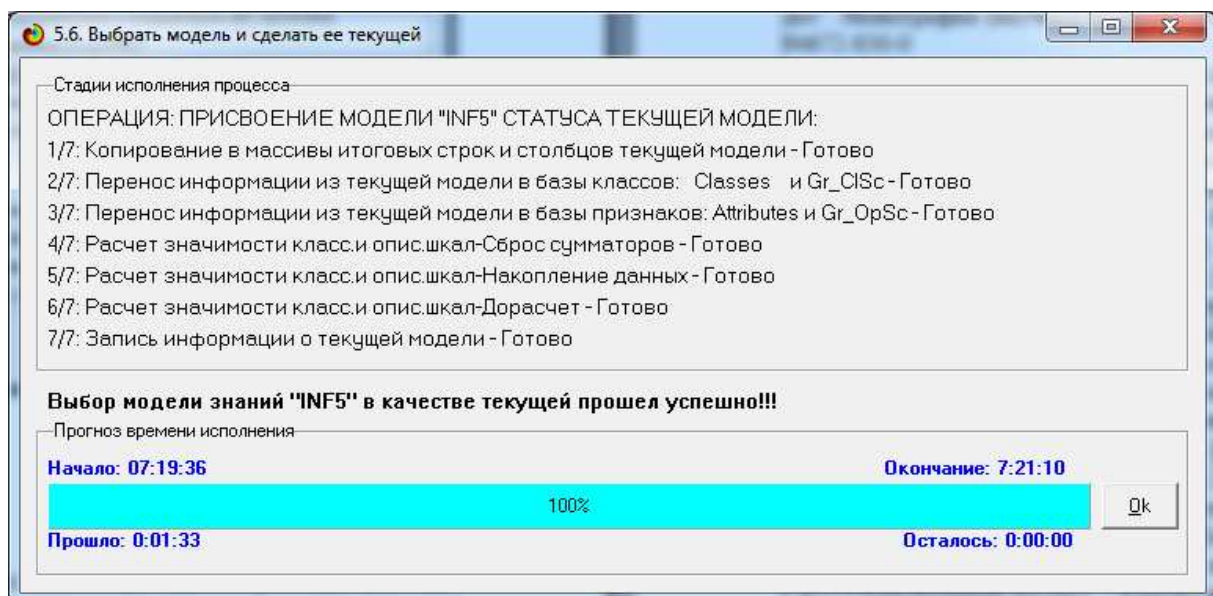
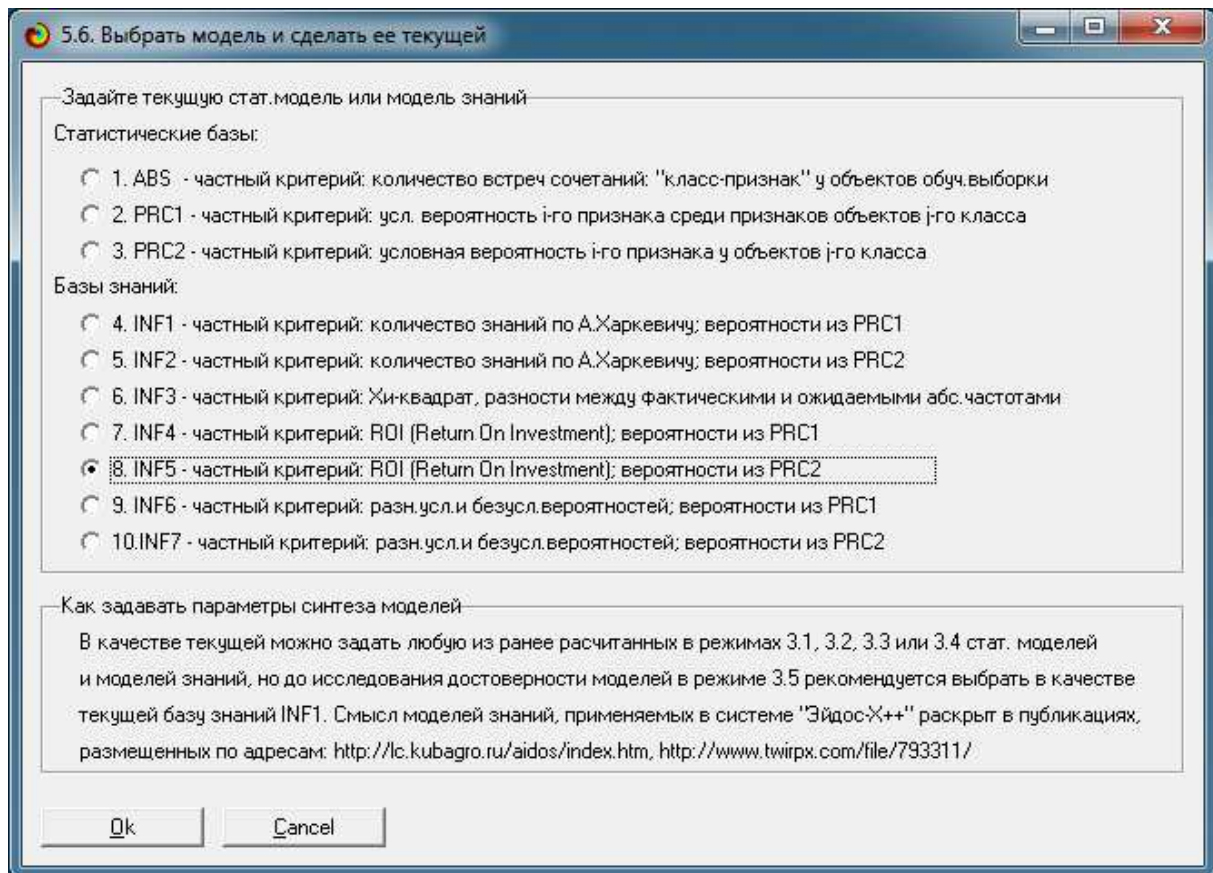


Рисунок 7. Экранные формы режима задания модели в качестве текущей

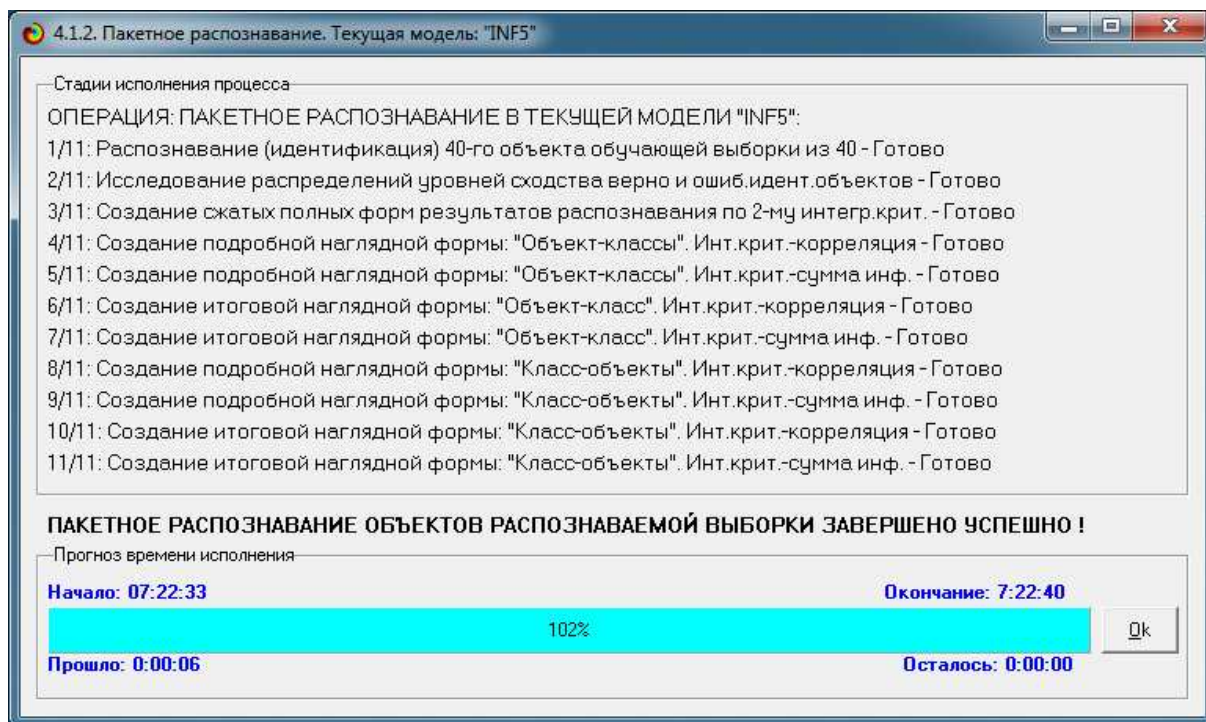


Рисунок 7. Экранная форма режима пакетного распознавания

Выводы.

В результате проведенной работы подготовлены все условия для выполнения последующих этапов АСК-анализа: решение задач идентификации, прогнозирования и принятия решений, а также исследования моделируемого объекта путем исследования его модели, которые планируется описать в будущих статьях.

ГЛАВА 6. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ В ОЦЕНКЕ РЕЗУЛЬТАТОВ НАУЧНОЙ И ПРЕПОДАВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ

6.1. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос», системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине

В разеле предлагается применить автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий систему «Эйдос» как для синтеза, так и для применения адаптивных интеллектуальных измерительных систем с целью измерения не значений параметров объектов, а для системной идентификации состояний сложных многофакторных нелинейных динамических систем. Кратко рассматривается математический метод АСК-анализа, реализованный в его программном инструментарии – универсальной когнитивной аналитической системе «Эйдос-Х++». Математический метод АСК-анализа основан на системной теории информации (СТИ), которая создана в рамках реализации программной идеи обобщения всех понятий математики, в частности - теории информации, базирующихся на теории множеств, путем тотальной замены понятия множества на более общее понятие системы и тщательного отслеживания всех последствий этой замены. Благодаря

математическому методу, положенному в основу АСК-анализа, этот метод является непараметрическим и позволяет сопоставимо обрабатывать десятки и сотни тысяч градаций факторов и будущих состояний объекта управления (классов) при неполных (фрагментированных), зашумленных данных числовой и нечисловой природы измеряемых в различных единицах измерения. Приводится развернутый численный пример применения АСК-анализа и системы «Эйдос-Х++» как для синтеза системно-когнитивной модели, обеспечивающей многопараметрическую типизацию состояний сложных систем, так и для системной идентификации их состояний, а также для принятия решений об управляющем воздействии, так изменяющем состав объекта управления, чтобы его качество (уровень системности) максимально повышалось при минимальных затратах на это. Для численного примера в качестве сложной системы выбран коллектив фирмы, а его компонент – сотрудники и кандидаты (персонал). Однако необходимо отметить, что этот пример следует рассматривать шире, т.к. АСК-анализ и система «Эйдос» разрабатывались и реализовались в очень обобщенной постановке, постановке, не зависящей от предметной области, и с успехом могут быть применены и в других областях

«...законы природы являются лишь высказываниями о пространственно-временных совпадениях...»

Альберт Эйнштейн

6.1.1. Интеллектуальные измерительные системы, как закономерный этап развития информационно-измерительных систем

Очевидно, смысл процесса измерения в том, что в его результате мы получаем определенное количество **информации** о степени выраженности тех или иных свойств у измеряемого объекта или о его состоянии. Информация может рассматриваться с двух точек зрения: с количественной и с качественной, т.е. содержательной, семантической. Парадокс заключается в том, что традиционно внимание обращается только на *содержание* информации, полученной в процессе измерения, тогда как на *количество* этой информации обычно вообще не обращают никакого внимания. Между тем количество информации полученной в результате измерений также очень важно, т.к. непосредственно связано с *точностью* измерений. Точность измерений принципиально ограничено соотношением неопределенностей Гейзенберга, поэтому принципиально ограничено и максимальное количество информации (Фишера), которое можно получить об объекте в процессе измерений⁹.

Например, если в результате измерения температуры с помощью бытового наружного термометра со стандартной шкалой от -50°C до $+50^{\circ}\text{C}$ мы получили содержательную информацию о том, что температура воздуха на улице равна 25°C , то мы получили $I = \log_2 100 \sim 7$ бит информации, если же мы узнали более точное значение температуры $25,4^{\circ}\text{C}$, то это увеличивает количество полученной информации: $I = \log_2 1000 \sim 10$ бит информации.

Проблема заключается в полном отсутствии **универсальных** инструментальных средств для синтеза измерительных систем в различных предметных областях, *которые бы позволяли вычислять какое количество информации содержится в результатах измерения о том, что измеряемая величина примет то или иное значение.*

Вроде бы существуют и учебные пособия по интеллектуальным измерительным системам [1], и государственный стандарт в этой области [2]. Однако, сформулирован-

⁹ См., «Выражение конечного доступного количества информации Фишера»: <http://dic.academic.ru/dic.nsf/ruwiki/44423>

ная проблема ими не решается, т.к. в них даже не ставится задача ее решения. Данный раздел посвящен описанию подхода к решению данной проблемы, основанного на применении автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос» [3, 4].

С точки зрения теории информации измерение представляет собой и *процесс* отображения, и *результат* отображения одной системы в другой. *Процесс* отображения одной системы в другой представляет собой процесс моделирования, *результатом* которого является *модель* измеряемого или отображаемого объекта. В этой связи возникает много сложных методологических проблем [5, 6]. В соответствии с принципом Эшби более простая система может адекватно отображаться в более сложной, тогда как более сложная система в более простой всегда отображается с необратимой потерей информации, т.е. неадекватно. Значит, что для того чтобы измерительная (управляющая) система была адекватной она должна быть сложнее измеряемого объекта (объекта управления).

Поэтому вполне естественно и закономерно, что история измерений – это по сути история эволюция информационно-измерительных систем, в процессе которой они постоянно усложнялись и становились все более и более точными. Измерения развивались по двум основным направлениям: с одной стороны от измерения объективных свойств к социальным и субъективным измерениям, а с другой стороны от измерения степени выраженности отдельных свойств объектов к измерению состояний систем в целом.

Первоначально измерения зародились в естественных науках, но потом постепенно стали проникать, особенно в последнее время, и в гуманитарные науки. Сначала измерялась степень выраженности объективных свойств объектов, принадлежащие им по самой их природе (физика: вес, скорость и т.п.), затем стали измеряться социально-экономические свойства, отражающие отношения людей с помощью вещей (эконометрика: потребительная и меновая стоимость), а после этого пришла очередь и субъективных свойств (педагогические измерительные системы, измеряющие уровень предметной обученности, а также психологические измерительные системы, обеспечивающие измерение степени выраженности психологических свойств личности). При этом сначала измерялась степень выраженности свойств объектов. Затем выяснилось, что не существует совершенно независимых друг от друга свойств, т.е. все объекты являются в той или иной степени нелинейными, а представление об абсолютно линейных объектах являются предельной абстракцией, наподобие математической точки. Но существуют практически линейные объекты, для которых нелинейностью можно вполне обоснованно пренебречь и для которых измерение степени выраженности их свойств вполне корректно. *Однако для систем с высоким уровнем системности и, соответственно, ярко выраженной нелинейностью более правильно говорить не об измерении свойств, а об идентификации состояний систем по их свойствам, т.е. о системной идентификации [7]¹⁰*. Системы с ярко выраженными нелинейными свойствами широко распространены. Это квантовые природные и технические системы, качественно изменяющие свое состояние в процессе штатной эксплуатации, глобальные природные системы [8], биологические и экологические системы, а также так называемые организационные системы, т.е. все системы с участием людей: социально-экономические системы, коллективы и отдельные люди – личности.

Например, в медицине, подход к лечению, основанный на измерении свойств пациента и приведении значений этих свойств к норме, называется симптоматическим лечением, т.е. лечением клинических признаков, а не человека, как целостной системы.

¹⁰ Автор впервые узнал о существовании чрезвычайно удачного термина: «Системная идентификация» именно из этой работы Шитикова В.К., Розенберга Г.С., Зинченко Т.Д.

При таком лечении патологические значения клинических признаков приводятся к нормальным значениям, но причины заболевания не устраняются.

Измерение всегда осуществляется во *взаимодействии* измеряемого объекта и измерительной системы, поэтому и процесс, и результат отображения является *взаимным*, т.е. измерительная система в процессе оказывает влияние на измеряемый объект и изменяет его и сама также изменяется в процессе измерения, поэтому она измеряет не состояние измеряемого объекта «самого по себе», каким оно было до измерения, а то состояние, которое возникло и стабилизировалось у него в результате возникновения равновесия в процессе измерения. Казалось бы, подобные методологические рассуждения играют роль лишь в квантово-механической теории измерений, но это далеко не так. При измерении состояний сложных нелинейных физических, социально-экономических, биологических и психологических систем мы часто наблюдаем существенное, часто необратимое влияние измерительной системы на измеряемую систему, что совершенно недопустимо ни с какой точки зрения.

Например, мы прекрасно знаем, сколь болезненными для пациента и небезопасными для него по своим последствиям могут быть медицинские анализы, в частности гистологические пробы на рак, которые резко активируют развитие рака. Многие психологические и педагогические измерительные материалы (тесты) с *ложными*, неполными и вообще неверными вариантами ответов, типа ЕГЭ, *необратимо дезориентируют тестируемых* в самом процессе тестирования и поэтому дают закономерное снижение качества результатов при повторном тестировании, т.е. имеют недопустимо низкую ретестовую надежность. Но главное даже не в этом, а в том, что само измерение с помощью подобных грубых измерительных систем *необратимо повреждает* измеряемый объект, в результате чего он существенно изменяет свои свойства и по сути становится иным, чем до измерения, а вот это уже недопустимо. В этом случае сам измерительный инструмент, недопустимо сильно влияющий на измеряемый объект, следует признать непригодным для измерений (за исключением случая проведения краш-тестов). Представьте себе, чтобы Вы сказали о термометре для измерения температуры воды, если бы он в процессе измерения эту воду заморозил бы или вскипятил. Но когда подобными измерительными инструментами и методами проводятся педагогические измерения уровня предметной обученности у миллионов школьников нашей страны, то это почему-то считается вполне приемлемым.

В естественных науках прогресс во многом определяется совершенствованием технологий *измерений* и накопления фактов, а также развитием методов *извлечения знаний из фактов*.

Само понятие измерения претерпело значительную эволюцию [8].

Раньше под измерением понимали сам факт обнаружения и идентификации объекта или обнаружения (идентификации) у изучаемого объекта какого-либо свойства, что выражалось *качественной*, номинальной или текстовой величиной.

Затем возникло представление о степени выраженности различных свойств объектов и возможности между ними отношений «больше», «меньше».

Позже возникло понятие об отношениях *эквивалентности* между степенью выраженности свойства измеряемого объекта с каким-либо *эталоном*, который стал рассматриваться как единица измерения. Примером может быть измерение веса продуктов на весах с помощью гирь. Развитие этих представлений привело к формированию понятий об измерительных шкалах различных типов [9] и о единицах измерения и числовых измерениях и тогда под измерением стали понимать установление *количественного* значения некоторого свойства объекта. При этом сначала использовались измерительные шкалы с условным нулем, а затем и с абсолютным нулем.

Дальнейшее развитие науки привело к пониманию, что измерение любой количественной величины всегда осуществляется с некоторой принципиально неустранимой *погрешностью*. Абсолютно точное значение измерения недостижимо по ряду причин. Прежде всего, для абсолютно точной записи любой величины потребовалось бы бесконечное количество знаков, а значит информационный носитель бесконечной емкости и бесконечное время для записи этой информации, а также бесконечная скорость передачи информации по каналу связи и такая же скорость записи на носитель. С другой стороны сам процесс измерения всегда занимает некоторое конечное время и за это время возможно получить лишь ограниченный объем информации Фишера об измеряемом объекте, а сама изменяемая величина, вообще говоря, может и измениться за это время. Кроме того, в любом реальном процессе измерения измерительная система *взаимодействует* с исследуемым объектом, т.е. не только получает информацию о его состоянии, но и влияет на него, т.е. изменяет его состояние в процессе измерения. Иначе говоря, существует принципиально неустранимое влияние наблюдателя на наблюдаемую им реальность. Следовательно, от самого наблюдателя в определенной степени, конечно, зависит, что он наблюдает и что он *в принципе* может наблюдать. В этой связи возникает много сложных методологических проблем [5, 6]. Поэтому результаты количественных измерений стали записывать с указанием погрешностей, а затем и *доверительного интервала*, в который с определенной вероятностью (обычно 0.95) попадает истинное значение измеряемой величины. Таким образом, понимание, что измерение всегда осуществляется с некоторой погрешностью, привело к переходу к доверительным интервалам и *интервальным оценкам* в измерениях.

Следующий этап развития теории измерений связан с пониманием того, что измеряемая величина каким-то образом, в общем случае *неравномерно*, распределена внутри доверительного интервала. В результате под измерением стали понимать установление *статистических характеристик вероятностных распределений числовых величин* и это ознаменовало следующий этап развития понятия «Измерение». Чем выше кривизна кривой частот внутри интервала, тем чаще должны быть расположены точки измерений, чтобы отразить эту форму (теорема Котельникова об отсчетах). Поэтому в системно-когнитивном анализе и его программном инструментарии – интеллектуальной системе «Эйдос» [3, 4] возникла и была реализована идея использования интервальных шкал с *адаптивным размером интервала*, при котором его размер изменяется таким образом, чтобы внутри разных интервалов было примерно одинаковое количество наблюдений. В качестве примеров научных исследований, использующих такой уровень понимания сущности измерений, являются работа [8].

Но когда стали анализировать эти распределения, то оказалось, что для того, чтобы сделать обоснованные выводы о характере влияния исследуемых факторов на систему необходимо сравнивать ее поведение под действием этих факторов с какой-то *базой сравнения*, например с поведением той же системы в условиях отсутствия действия этих факторов и при прочих равных условиях, т.е. сравнивать с *контрольной группой*. Так появился метод контрольных групп, без которого немыслимо научное измерение влияния факторов.

Однако в связи с природой самого объекта исследования реально на практике исследователи чаще всего не имеют возможности изучить влияние на объект всех возможных сочетаний значений факторов и выделить контрольную группу. Поэтому приходится сравнивать поведение объектов в различных группах с его поведением в среднем по всей выборке, для чего были предложены «метод среднего и отклонений от среднего» и «метод переменных контрольных групп» [8].

Другой проблемой является *выделение полезного сигнала из шума*, т.к. в общем случае измеряемая величина является суммой «истинного» значения и шума, и обеспе-

чение *сопоставимости* изучения влияния факторов различной природы, как качественных, так количественных, измеряемых в различных типах измерительных шкал [9, 10] и в различных единицах измерения. Все эти проблемы решены в новом методе исследования: системно-когнитивном анализе («АСК-анализ») и его программном инструментарии – универсальной когнитивной аналитической системе «Эйдос» [3, 4].

Соответственно эволюции понятия «Измерение» эволюционировали и измерительные системы, технологии и методики их применения. Для установления фактов, осознанно или нет, но всегда использовался некоторый *инструмент* и способ или методика его применения. Исторически первыми такими инструментами стали сами органы восприятия человека, прежде всего зрение, а способом – наблюдение. Создание новых инструментов всегда приводило к революции в науке или даже возникновению новых наук. Достаточно вспомнить, как изменились биология и медицина после изобретения микроскопа Антони Ван Левенгуком, и как изменилась астрономия после изобретения телескопа Галилео Галилеем. Последовавшее затем изобретение радиотелескопа привело к возникновению радиоастрономии, а рентгеновского телескопа, соответственно – рентгеновской астрономии и т.д. и т.д. Как микроскоп или телескоп многократно увеличивают возможности естественного зрения, *если оно есть*, так и системы искусственного интеллекта многократно увеличивают возможности интеллекта естественного, *если он есть*. Микроскоп или телескоп не заменяют зрения, а лишь усиливают возможности естественного зрения или компенсируют его недостатки. Аналогично и системы искусственного интеллекта не заменяют естественного интеллект, а лишь усиливают возможности естественного интеллекта или компенсируют его недостатки.

Адаптивная интеллектуальная измерительная система (АИИС) также представляет собой новый инструмент исследования, своего рода специфический микроскоп и телескоп одновременно, появление которого в руках исследователей может многократно увеличить возможности человека к наблюдению и осмыслению их результатов, а значит и привести к возникновению новых направлений науки [8, 11, 12, 13]. В качестве источника фактов для АИИС выступают базы знаний, отражающие свойства систем и характеристику их состояний. В качестве эффективной системы выделения сверхслабого полезного сигнала из многократно превосходящего его шума в АИИС применяются методы и технологии искусственного интеллекта, в частности предлагается применить для этого АСК-анализ и систему «Эйдос».

В общем случае синтез измерительной системы представляет собой процесс обучения с учителем системы распознавания образов на основе примеров и создание базы знаний, отражающей причинно-следственные зависимости между показаниями датчиков и результатами измерения.

*Синтез измерительной системы предполагает наличие **двух** или более параллельных и независимых друг от друга источников и соответствующих им параллельных каналов передачи информации об измеряемых объектах: источник априорной информации о **сущностных** значениях свойств или измеряемых состояний объектов и источник **косвенной** информации, получаемой непосредственно в процессе измерения объектов с помощью датчиков¹¹. После выявления **причинно-следственных взаимосвязей** между этими двумя потоками информации измерительная система на основе знания этих взаимосвязей способна только по косвенной информации, получаемой от измеряемого объекта в процессе измерения с помощью датчиков, восстановить, реконструировать априорную информацию и эта реконструированная априорная информация собственно и является результатом измерения.*

¹¹ Естественно, датчики должны быть связаны с системой обработки измерительной информации с помощью каналов связи.

Ниже приведены два примера использования знания причинно-следственных зависимостей для построения измерительных систем.

Пример 1: пружинные весы. Физики в лице Гука выявили причинно-следственную взаимосвязь между весом тела, положенного на пружинные весы, и степенью сжатия пружины под действием этого веса. Когда вес сравнительно невелик и пружина почти не сжимается, то между ее сжатием и весом существует практически линейная взаимосвязь, знание которой и положено в основу принципа действия пружинных весов, которые *по степени сжатия пружины определяют вес тела*.

Пример 2: ртутный или спиртовой термометр. При небольших изменениях температуры жидкости ее объем изменяется практически линейно от температуры. Знание этой причинно-следственной зависимости положено в основу действия термометров, которые *по степени расширения жидкости определяют ее температуру*. При этом предполагается, что теплоемкость термометра пренебрежимо мала по сравнению с теплоемкостью тела, температура которого измеряется и поэтому в процессе *выравнивания* их температур (достижения теплового *равновесия*) в процессе теплового взаимодействия температура измеряемого тела практически не изменится за счет теплового взаимодействия с термометром, а температура последнего станет равной температуре измеряемого тела. Это в частности означает, что с помощью макротермометра невозможно измерить температуру капельки тумана или бактерии.

Эмпирические данные, используемые для выявления причинно-следственных зависимостей в предметной области и необходимые для синтеза измерительной системы, образуют обучающую выборку, которая является репрезентативной по отношению к некоторой генеральной совокупности, в пределах которой действуют те же причинно-следственные зависимости, что и в обучающей выборке *и в пределах которой применение данной системы методологически корректно*.

В естественнонаучных измерениях генеральная совокупность может быть *глобальной*, т.е. в пространстве может включать и другие галактики, а во времени – миллиарды лет. Когда же измерения производятся в социально-экономических и социально-психологических системах, то границы генеральной совокупности в пространстве могут измеряться километрами, а во времени – годами (так называемые периоды эргодичности), т.е. измерения в этих предметных областях по необходимости являются *локальными*. Для социально-экономических систем не выполняется принцип относительности, подобный принципу относительности Галилея-Эйнштейна [14]. Поэтому в естественных науках возможен и успешно применяется подход к построению измерительных систем на основе заранее известных причинно-следственных связей в предметной области. Однако этот подход неприменим, например, в социально-экономических и социально-психологических эмпирических исследованиях, в которых, по этой причине, *необходимо иметь инструменты для выявления этих причинно-следственных связей непосредственно при синтезе измерительной системы*. Этот инструмент необходим также и при *применении измерительной системы* в адаптивном режиме, т.к. положенные в основу измерительной системы причинно-следственные связи могут изменяться в зависимости от места и времени ее применения. Поэтому необходимо адаптировать и локализовать адаптивные интеллектуальные измерительные системы соответственно к времени и месту их применения. Это наукоемкая и дорогая процедура, одна без нее применение неадаптированных и нелокализованных измерительных технологий превращается в профанацию самой идеи измерений в соответствующих предметных областях. Понятно, что технология стоит на порядки дороже продуктов ее применения. Поэтому на практике как правило пользуются неадаптированными и нелокализованными измерительными инструментами, которые дают неизвестные систематические ошибки измерений или вообще результаты, близкие к случайным. Таким образом, не-

обходимо включить инструментарий синтеза измерительной системы в состав самой этой системы, но для этого надо иметь такой инструментарий и он должен быть достаточно прост в применении, т.е. должен иметь персональный уровень.

Таким универсальным инструментом, обеспечивающим **выявление** причинно-следственных зависимостей в различных предметных областях, является автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос» [3, 4]. Но АСК-анализ и система «Эймос» обеспечивают не только выявление причинно-следственных связей в различных предметных областях, но и **применение** знания этих причинно-следственных зависимостей **для измерения** степени выраженности свойств и идентификации состояний объектов в этих предметных областях. **Причинно-следственные связи**, выявленные в моделируемой предметной области, отражены в базах знаний системы «Эйдос» и могут быть наглядно представлены в разнообразных текстовых и графических формах (которых более 110), в том числе в форме когнитивных функций [15, 16].

В качестве особо важного этапа развития измерительных систем необходимо отметить информационно-измерительные системы.

При построении измерительных систем в естественных науках используются знания причинно-следственных зависимостей, полученные в результате *заранее* проведенных длительных фундаментальных исследований и сохраняющих свое значение *глобально и на очень длительные периоды времени, возможно миллионы и даже миллиарды лет*. Однако в наше время потребности практики измерений часто опережают темпы развития фундаментальной науки, в результате чего возникает проблема синтеза измерительных систем, обеспечивающих измерение состояний сложных нелинейных объектов на основе значений их свойств *при заранее неизвестном виде причинно-следственных зависимостей между свойствами и состояниями*. В наше время, когда эмпирические измерения все более проникают в социально-экономические и психологические исследования, *заранее* знать причинно-следственные зависимости в измеряемой области не представляется возможным [14]. Это значит, что в этих областях неприменим подход, успешно применявшийся в естественных науках.

В автоматизированном системно-когнитивном анализе (АСК-анализ) предлагается сначала построить интеллектуальные модели, отражающие эти причинно-следственные взаимосвязи на основе неполных и зашумленных исходных данных большой размерности, а затем использовать знание этих зависимостей для системных, т.е. многопараметрических нелинейных измерений.

Обратимся к эпиграфу к статье. Итак, современная наука, по мнению ее выдающихся представителей, изучает лишь пространственно-временные совпадения. Неужели уникальные явления, т.е. то, что не совпадает, и не изучается вовсе? И что вообще означает: «совпадение»? Ответом на эти вопросы является системная нечеткая интервальная математика [15]. Ведь ясно, что *абсолютно* точно ничто не совпадает¹², все всегда совпадет лишь *в определенной степени или с определенной точностью* и эту степень всегда можно выбрать такой, что 1) ничего не будет совпадать, или 2) все будет совпадать, или 3) что-то будет, а что-то не будет совпадать. Третий вариант открывает путь к исследованию достаточно уникальных явлений и закономерностей, подчиняющихся принципам относительности лишь локально в определенных доменах (умвелтах) пространства и времени.

Итак, предлагается применить автоматизированный системно-когнитивный анализ как для синтеза, так и для применения адаптивной интеллектуальной измерительной системы с целью измерения не значений параметров объектов, а для *идентифика-*

¹² Не считая, квантовых объектов, для которых действует «принцип тождественности элементарных частиц»

ции состояний измеряемых систем, т.е. для так называемой системной идентификации [7, 13]. Измерительная система должна быть не проще, чем измеряемая система (вариант принципа Эшби), иначе она не может быть адекватной. Значит, для измерения сложных нелинейных систем должны применяться интеллектуальные технологии, обеспечивающие достаточно высокий уровень сложности баз знаний. Когда мы проводим диагностику (квалиметрию) материалов, то стараемся оценить качество каких-либо параметров в определенных шкалах. Это параметрический подход. Параметрический подход корректен только для линейных объектов (материалов), в которых отдельные параметры практически не влияют друг на друга. Для нелинейных материалов качество надо оценивать не по одному параметру, а по всем сразу. Это суть системного подхода к квалиметрии, при котором **качество рассматривается как системное (эмерджентное) свойство системы**. По мнению автора все без исключения свойства объектов и явлений имеют системную эмерджентную природу [17]. Это связано с тем, что структура системы обуславливает на ее макросвойства [18]. Поэтому возникает задача системной идентификации качества материала на основе значений **различных** его параметров, которая решается в АСК-анализе и системе «Эйдос». В качестве примеров системной идентификации можно привести измерение сейсмоопасности микрозоны [19] и измерение качества микрозоны для выращивания пшеницы [20].

Ниже рассмотрим *простой* условный численный пример того, как осуществляется *синтез* измерительной системы в АСК-анализе и его программном инструментарии – универсальной когнитивной аналитической системе «Эйдос» [3, 4] и как при этом осуществляется *метризация* шкал [9], преобразование данных в информацию, а ее в знания, как выявляются причинно-следственные зависимости между результатами измерений и их интерпретацией, и как эта измерительная система *применяется* для системной идентификации состояний сложных систем, в частности для многопараметрического измерения их качества. При этом будем руководствоваться этапами АСК-анализа [3] (когнитивная структуризация и формализация предметной области; синтез и верификация моделей, определение наиболее достоверной модели; решение задач идентификации, прогнозирования, принятия решений и исследования предметной области путем исследования ее модели) и будем последовательно повышать степень формализации создаваемых моделей, преобразуя данные в информацию, а ее в знания:

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ) для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на корпорацию к каким ее изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона [21], состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу.

Знания – это информация, полезная для достижения целей [22].

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).

2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

– вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);

– знания, формализованные в естественном вербальном языке;

– знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);

– знания в форме технологий, организационных, производственных, социально-экономических и политических структур;

– знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно последовательно повышать степень формализации исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

– преобразовать исходные данные в информацию;

– преобразовать информацию в знания;

– использовать знания для решения задач управления, принятия решений и исследования предметной области.

Исходные данные.

В качестве исходных данных, описывающих различные состояния объектов для системной идентификации, рассмотрим правильные тела Платона (таблица 1):

Таблица 1 – ТРЕХМЕРНЫЕ ПРАВИЛЬНЫЕ МНОГОГРАННИКИ

№	Наименование правильного многогранника	Изображение	Количество				
			Сторон у грани	Ребер у вершины	Вершин (всего)	Ребер (всего)	Граней (всего)
1	Тетраэдр		3	3	4	6	4
2	Куб		4	3	8	12	6
3	Октаэдр		3	4	6	12	8
4	Додекаэдр		5	3	20	30	12
5	Икосаэдр		3	5	12	30	20

Далее осуществим синтез интеллектуальной измерительной системы в соответствии с этапами АСК-анализа [3, 4]:

1. Когнитивная структуризация предметной области. Это единственный этап АСК-анализа, осуществляемый не на компьютере. На этом этапе необходимо решить, что мы хотим определять и на основе чего. В данном случае будем идентифицировать тело Платона на основе его признаков, приведенных в таблице 1.

Дальнейшие этапы АСК-анализа выполняются в системе «Эйлос».

2. Формализация предметной области включает: разработку классификационных шкал и градаций; разработку описательных шкал и градаций; разработку обучающей выборки, т.е. кодирование исходных данных с применением справочников классификационных и описательных шкал и градаций. По сути формализация предметной области представляет собой *нормализацию* базы исходных данных.

Таблица 2 – КЛАССИФИКАЦИОННАЯ ШКАЛА И ЕЕ ГРАДАЦИИ

Код	Наименование
1	НАИМЕНОВАНИЕ ТЕЛА ПЛАТОНА-Додекаэдр
2	НАИМЕНОВАНИЕ ТЕЛА ПЛАТОНА-Икосаэдр
3	НАИМЕНОВАНИЕ ТЕЛА ПЛАТОНА-Куб
4	НАИМЕНОВАНИЕ ТЕЛА ПЛАТОНА-Октаэдр
5	НАИМЕНОВАНИЕ ТЕЛА ПЛАТОНА-Тетраэдр

Таблица 3 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ

Код	Наименование
1	СТОРОН У ГРАНИ-3
2	СТОРОН У ГРАНИ-4
3	СТОРОН У ГРАНИ-5
4	РЕБЕР У ВЕРШИНЫ-3
5	РЕБЕР У ВЕРШИНЫ-4
6	РЕБЕР У ВЕРШИНЫ-5
7	ВЕРШИН (ВСЕГО)-12
8	ВЕРШИН (ВСЕГО)-20
9	ВЕРШИН (ВСЕГО)-4
10	ВЕРШИН (ВСЕГО)-6
11	ВЕРШИН (ВСЕГО)-8
12	РЕБЕР (ВСЕГО)-12
13	РЕБЕР (ВСЕГО)-30
14	РЕБЕР (ВСЕГО)-6
15	ГРАНЕЙ (ВСЕГО)-12
16	ГРАНЕЙ (ВСЕГО)-20
17	ГРАНЕЙ (ВСЕГО)-4
18	ГРАНЕЙ (ВСЕГО)-6
19	ГРАНЕЙ (ВСЕГО)-8

Используя классификационные и описательные шкалы и градации (таблицы 2 и 3) закодируем описания тел Платона, приведенные в исходных данных (таблица 1) в результате чего получим таблицу 4.

Таблица 4 – ОБУЧАЮЩАЯ ВЫБОРКА

Наименование объекта обучающей выборки	Код класса	Коды признаков				
Тетраэдр	5	1	4	9	14	17
Куб	3	2	4	11	12	18
Октаэдр	4	1	5	10	12	19
Додекаэдр	1	3	4	8	13	15
Икосаэдр	2	1	6	7	13	16

3. Синтез и верификация моделей, определение наиболее достоверной модели.

На основе результатов формализации предметной области системой «Эйдос» рассчитываются (см. рисунок 1) матрица абсолютных частот (корреляционная матрица), матрицы условных и безусловных процентных распределений, а также на их основе матрицы знаний с различными частными критериями знаний [9].

С применением данных частных критериев знаний рассчитываются модели знаний, отражающие силу и направление причинно-следственной взаимосвязи между значениями факторов и принадлежностью объекта к классам. В таблице 5 приведена одна из 7 моделей знаний:

Таблица 5 – МАТРИЦА ЗНАНИЙ НА ОСНОВЕ ЧАСТНОГО КРИТЕРИЯ ЗНАНИЙ А.ХАРКЕВИЧА (В МИЛЛИБИТАХ)

№	Наименование признака	Наименование класса				
		Додекаэдр	Икосаэдр	Куб	Октаэдр	Тетраэдр
1	СТОРОН У ГРАНИ-3		368		368	368
2	СТОРОН У ГРАНИ-4			1161		
3	СТОРОН У ГРАНИ-5	1161				
4	РЕБЕР У ВЕРШИНЫ-3	368		368		368
5	РЕБЕР У ВЕРШИНЫ-4				1161	
6	РЕБЕР У ВЕРШИНЫ-5		1161			
7	ВЕРШИН (ВСЕГО)-12		1161			
8	ВЕРШИН (ВСЕГО)-20	1161				
9	ВЕРШИН (ВСЕГО)-4					1161
10	ВЕРШИН (ВСЕГО)-6				1161	
11	ВЕРШИН (ВСЕГО)-8			1161		
12	РЕБЕР (ВСЕГО)-12			661	661	
13	РЕБЕР (ВСЕГО)-30	661	661			
14	РЕБЕР (ВСЕГО)-6					1161
15	ГРАНЕЙ (ВСЕГО)-12	1161				
16	ГРАНЕЙ (ВСЕГО)-20		1161			
17	ГРАНЕЙ (ВСЕГО)-4					1161
18	ГРАНЕЙ (ВСЕГО)-6			1161		
19	ГРАНЕЙ (ВСЕГО)-8				1161	

Подбазы знаний наглядно представляются в форме когнитивных функций (рисунки 2):

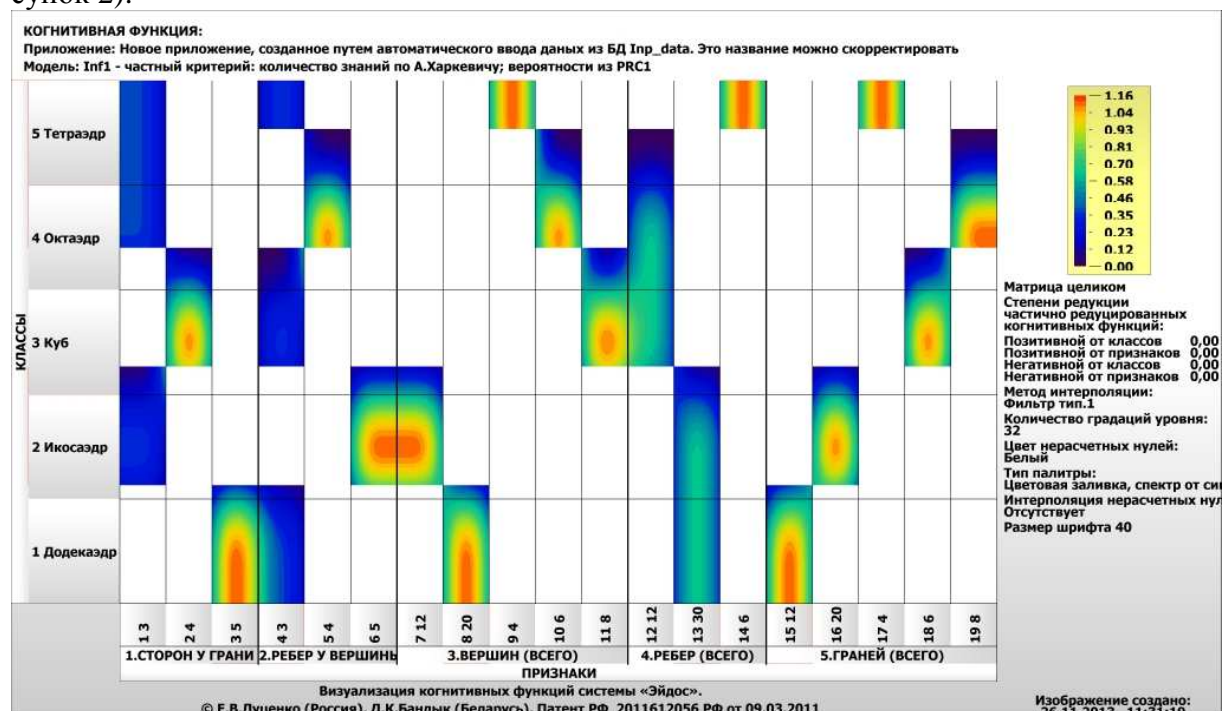


Рисунок 1 – Наглядная визуализация базы знаний Inf1 в форме когнитивной функции

Верификация моделей осуществляется путем решения в них задачи системной идентификации и подсчета количества ошибок 1-го и 2-го рода (ошибок неидентификации и ложной идентификации). Модель знаний, приведенная в таблице 5, показывает 100% достоверность идентификации и неидентификации.

4. Решение задач идентификации, прогнозирования, принятия решений и исследования предметной области.

Рассмотрим интегральные критерии знаний, используемые в настоящее время в АСК-анализе и системе «Эйдос-Х++» [23] для верификации моделей и решения задач идентификации и прогнозирования.

1-й интегральный критерий «Сумма знаний» представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний и имеет вид:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где: M – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j -го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i\text{-й фактор действует;} \\ n, & \text{где } n > 0, \text{ если } i\text{-й фактор действует с истинностью } n; \\ 0, & \text{если } i\text{-й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n , если он присутствует у объекта с интенсивностью n , т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

2-й интегральный критерий «Семантический резонанс знаний» представляет собой нормированное суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний и имеет вид:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

M – количество градаций описательных шкал (признаков);

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору объекта;

σ_j – среднеквадратичное отклонение частных критериев знаний вектора класса;

σ_1 – среднеквадратичное отклонение по вектору распознаваемого объекта.

Результат системной идентификации представляется в следующем виде (рисунок 3):

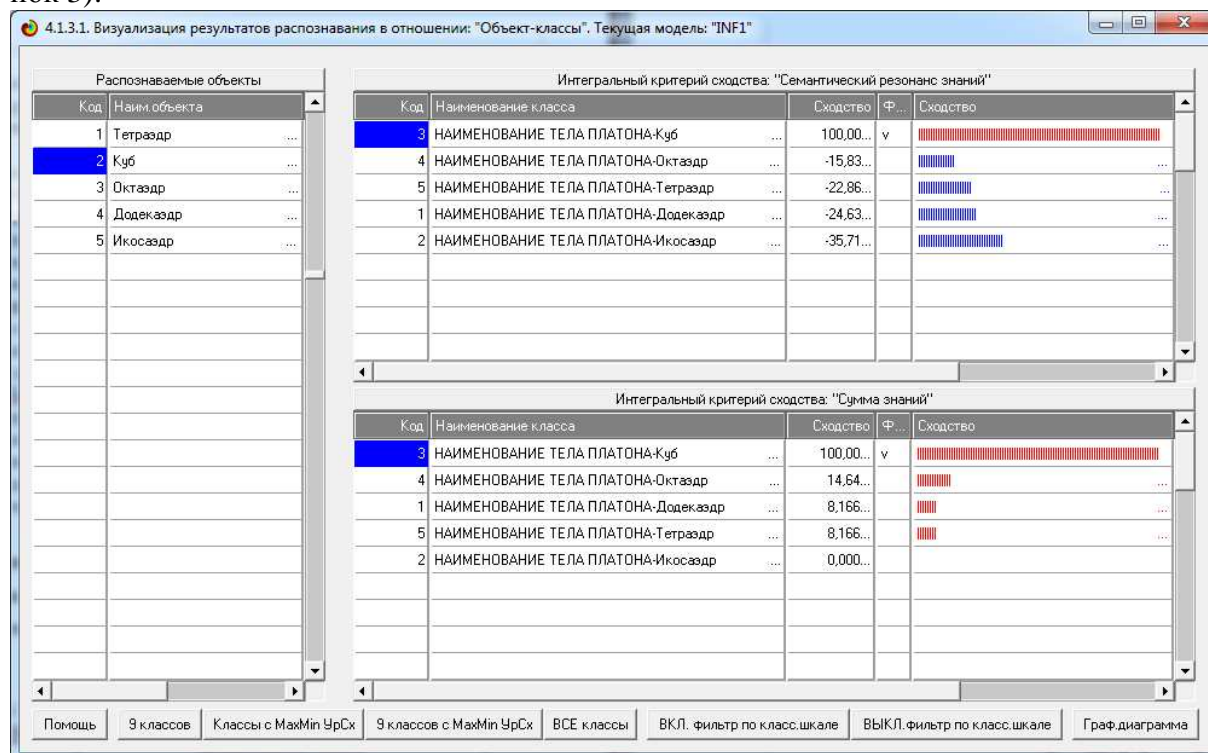


Рисунок 2 – Экранная форма с результатами системной идентификации в модели знаний INF1

На рисунке 3 показано, что по суммарному количеству информации, которое содержится в системе признаков объекта «Куб» он больше всего похож на класс «Куб», а остальные классы не похожи в различной степени.

Таким образом, в АСК-анализе:

1. Рассматривается ряд объектов (фактов), представляющих в совокупности исследуемую выборку.
2. Каждый из объектов исследуемой выборки представляет собой систему, имеющую сложную многоуровневую структуру признаков (экстенциональное описание).
3. Для каждого из объектов исследуемой выборки известно, к каким обобщенным категориям (классам) он относится (интенциональное описание).
4. Необходимо сформировать модель, обеспечивающую идентификацию объектов по их признакам, т.е. определение их принадлежности к обобщенным классам по их признакам.

Если признаки и классы относятся к одному времени, то имеет место задача идентификации (распознавания). Если же признаки (факторы, причины) относятся к прошлому, а классы, характеризующие состояния объектов, – к будущему, то это задача прогнозирования. Математически эти задачи не отличаются.

*Совокупность экстенционального и интенционального описания каждого объекта обучающей выборки, по сути, представляет собой его определение через подведение под более общее понятие и выделение специфических признаков. Иначе говоря, каждый объект обучающей выборки описывается принадлежностью к более общей категории (классу) и наличием у него ряда признаков. Например, так определяется понятие «млекопитающее»: это животное (более общее понятие), выкармливающее своих детей молоком (специфический признак). **На основе ряда определений конкретных объектов (конкретных онтологий) путем их обобщения можно получить определения (обобщающие онтологии) обобщенных образов классов.** Если привести в качестве примеров исследуемой выборки множество различных животных, как млекопитающих, так и других, каждый из таких примеров определить множеством признаков и построить модель, то окажется, что наиболее характерным признаком млекопитающих является не наличие шерсти или когтей, а именно вскармливание детенышей молоком.*

Процедура преобразования исходных данных в информацию – это *анализ* данных, состоящий из трех шагов:

- разработка справочников фактов и событий;
- выявление в исходных данных *фактов* или *событий* и их кодирование;
- выявление причинно-следственных связей (зависимостей) между этими событиями.

Фактически для преобразования исходных данных в информацию необходимо:

1. Разработать классификационные и описательные шкалы и градации.
2. С использованием классификационных и описательных шкал и градаций **закодировать** исходные данные, в результате чего получится обучающая выборка, состоящая из *фактов*, представляющих собой примеры в единстве экстенционального и интенционального описания.
3. Произвести расчет матриц абсолютных частот, условных и безусловных процентных распределений и матрицы информативностей, отражающей причинно-следственные связи между значениями факторов и принадлежностью объектов к классам.

Таким образом, информация по задаче – это исходные данные плюс классификационные и описательные шкалы и градации, обучающая выборка, а также матрицы частот, процентных распределений и информативностей.

Процедура преобразования информации в знания – это оценка полезности информации для достижения *цели*.

Значит знания по задаче – это информация плюс цель и оценка степени полезности информации для достижения этой цели.

Знания получаются из информации, когда мы классифицируем будущие состояния объекта управления как желательные (целевые) и нежелательные.

Банк данных – это базы данных плюс система *управления базами данных* (СУБД) (стандартные термины). СУБД – это, по сути, *система управления данными*.

Информационный банк – это информационные базы плюс информационные системы (предлагается стандартизировать эти термины). Информационная система – это, по сути, *система управления информацией*.

Банк знаний – это базы знаний плюс интеллектуальные системы (стандартные термины). Интеллектуальная система – это, по сути, *система управления знаниями*.

Итак, измерение рассматривается как процесс получения информации об объекте измерения, в частности о степени выраженности тех или иных его свойств или принадлежности состояния объекта измерения к определенным категориям. Предлагается применить системно-когнитивный анализ как для синтеза, так и для применения адаптивной интеллектуальной измерительной системы с целью измерения не значений па-

раметров объектов, а для идентификации состояний измеряемых систем, т.е. для так называемой системной идентификации. Измерительная система должна быть не проще, чем измеряемая система (вариант принципа Эшби), иначе она не может быть адекватной. Значит, для измерения сложных нелинейных систем должны применяться интеллектуальные технологии, обеспечивающие достаточно высокий уровень сложности баз знаний. Когда мы проводим диагностику (квалиметрию) материалов, то стараемся оценить качество каких-либо параметров в определенных шкалах. Это параметрический подход, который корректен только для линейных объектов (материалов), в которых отдельные параметры практически не влияют друг на друга. Для нелинейных материалов качество надо оценивать не по одному параметру, а по всем сразу и в этом суть системного подхода к квалиметрии, когда качество рассматривается как системное (эмерджентное) свойство.

Математические модели АСК-анализа, применяемые при синтезе и применении адаптивных интеллектуальных измерительных систем, а также численные примеры системной идентификации, более подробно рассматриваются в последующих разделах.

6.1.2. Математический метод АСК-анализа – системная теория информации

Первый раздел данной работы посвящен концептуальным основам построения интеллектуальных измерительных систем в АСК-анализе, данный раздел (второй) – математическому методу АСК-анализа, в третьем будет рассмотрен численный пример синтеза интеллектуальной измерительной системы в системе «Эйдос-Х++» и ее применения для системной идентификации состояний сложных систем.

На основе 1-го раздела предлагаются следующие три принципа построения интеллектуальных измерительных систем в АСК-анализе.

1-й принцип состоит на *ясном осознании* того обстоятельства, что когда мы получаем результаты измерения, то по сути мы получаем некоторое количество **информации** о том, в каком состоянии находится измеряемый объект. Однако традиционно результаты измерения выражаются в определенных единицах измерения (в частности, единицах измерения физических величин), а не в единицах измерения информации и этим в определенной степени маскируется или скрывается *смысл самого измерения*, выраженный в 1-м принципе.

2-й принцип, связан с первым и состоит в понимании того, что когда мы получаем результаты измерения то нас интересует не собственно сам этот результат, а количество информации, которое содержится в результате измерения о состоянии объекта измерения, т.е. о том, что нас собственно интересует. Например, когда врач измеряет температуру пациенту то его интересует не эта температура сама по себе как некоторые почему-то думают, а возможность на ее основе сделать выводы о состоянии пациента, т.е. о том болен он или нет, и, если болен, то на сколько серьезно и какой у него диагноз и какой выбрать план лечения при этом диагнозе.

3-й принцип состоит в том, что при *построении* измерительной системы на эмпирических примерах производится *градуировка* или метризация измерительных шкал, т.е. нанесение на них делений, соответствующих различным степеням выраженности измеряемых свойств у объектов измерения. Затем, когда измерительная система *применяется*, т.е. при измерении по ранее полученным шкалам получаются некоторые значения, то на основании этих значений делается вывод о том, что состояние измеряемого объекта близко к состоянию тех примеров, которые давали аналогичный результат измерений при построении шкал. По сути *3-й принцип, отражающий этап построения или синтеза измерительной системы, функционально сходен с этапом обучения сис-*

темы распознавания образов, а этап ее применения сходен с применением системы распознавания для идентификации состояния объекта измерения.

Для того, чтобы реализовать сформулированные принципы в реальной интеллектуальной измерительной системе *необходим математический метод, обеспечивающий преобразование данных, полученных в результате измерений, в информацию о состоянии измеряемого объекта.* Такой метод существует – это математический метод АСК-анализа, основанный на системной нечеткой интервальной математике (СНИМ) [3, 15] и представляющий собой реализацию идей СНИМ в теории информации.

В этой связи необходимо определить соотношение содержания терминов: «данные», «информация» и «знание» (рисунок 4):

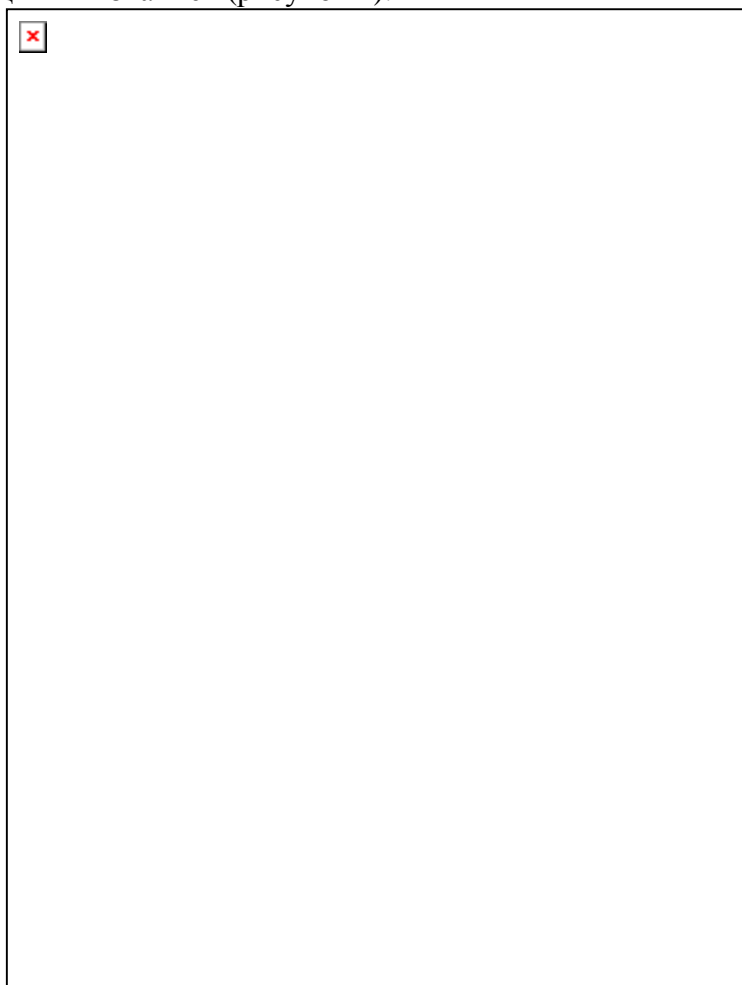


Рисунок 3 – Цикл преобразования эмпирических данных в информацию и знания и их применения для прогнозирования и принятия управленческих решений в АСК-анализе

Данные рассматриваются как информация, записанная на носителях или находящаяся в каналах связи и представленная в определенной системе кодирования или на определенном языке и рассматриваемая безотносительно к ее смысловому содержанию.

Смысл данных согласно концепции смысла Шенка-Абельсона [21] известен и понятен тогда, когда известны причины и следствия между событиями, которые описываются этими данными.

Информация представляет собой осмысленные данные, т.е. данные, описывающие события, между которыми выявлены причинно-следственные связи.

Знания – это информация, полезная для достижения целей, т.е. для управления [22].

В этой связи *возникает вопрос о математической количественной мере причинно-следственных связей, которая бы адекватно отражала их силу и направление*. Из вышесказанного следует, что естественной мерой причинно-следственных связей являются количественные меры информации и в качестве единицы измерения силы и направления причинно-следственных связей могут быть использованы единицы измерения информации. В связи с этой идеей необходимо отметить работу [24], суть которой в применении теории информации для проверки статистических гипотез. Еще в лемме Неймана-Пирсона доказывается, что более вероятна та статистическая гипотеза в пользу которой больше информации. В предисловии к работе [24] А.Н. Колмогоров высоко оценивал это научное направление, но соответствующий поток работ в СССР не возник [25]. По-видимому, АСК-анализ мере можно рассматривать как развитие этого направления прикладной математической статистики, может быть не столько в чистоматематическом теоретическом плане, сколько в прагматически-прикладном [25, 26].

Однако известно довольно много различных количественных мер информации. Поэтому возникает вопрос о том, какая мера информации является наиболее подходящей в нашем случае. По мнению автора это семантическая мера целесообразности информации А.Харкевича [3]. Основным свойством этой меры, предопределяющим ее выбор, является то обстоятельство, что в ее определение *органично* входит понятие *цели*. В соответствии с изложенными выше и в работе [22] представлениями автора о соотношении понятий: «Данные», «Информация» и «Знания» это означает, что *по сути А.Харкевич предложил количественную меру знаний*. Кроме того для вычисления меры А.Харкевича достаточно знать изменение вероятности достижения цели в условиях действия некоторого значения фактора и при его отсутствии, т.е. она вполне может быть рассчитана непосредственно на основе эмпирических данных, что очень важно для практических применений (поэтому и говорят, что эта мера прагматическая).

Операция преобразования данных в информацию называется «анализ данных», представляет собой процедуру выявления *смысла* в данных, т.е. согласно концепции смысла Шенка-Абельсона, выявление причинно-следственных связей между событиями¹³, отражаемыми этими данными, и предполагает выполнение следующих этапов:

1. Разработка справочников, содержащих формальное кодированное описание с одной стороны будущих состояний объекта управления, а с другой стороны – факторов их значений, влияющих на этот объект (классификационных и описательных шкал и градаций в терминологии АСК-анализа).

2. Поиск в исходных данных *событий*, связанных с переходами объекта из одного состояния в другое, и значений факторов, под действием которых эти переходы происходят. При этом в качестве значений факторов могут выступать и переходы объекта из одного состояния в другое в прошлом.

3. Преобразование базы исходных *данных* в базу *событий*, т.е. кодирование исходных данных с использованием справочников классов и факторов.

4. Поиск причинно-следственных связей между прошлыми и будущими событиями в базе событий и формальное представление этих причинно-следственных связей в виде базы информативностей.

Таким образом, если исходные базы данных представляют собой временные ряды, то информационная база включает в себя еще дополнительно:

– базы классификационных и описательных шкал и градаций;

¹³ Причинно-следственные связи, их сила и направление – это вообще не математический термин, а термин, описывающий взаимосвязь событий реальной предметной области. Поэтому выявление причинно-следственных связей непосредственно в данных вообще невозможно, а для этого необходимо предварительно найти в этих данных описания событий.

– базу событий (т.е. обучающую выборку), представляющую собой закодированную с помощью классификационных и описательных шкал и градаций базу исходных данных;

– базу информативностей, содержащую информацию о силе и направлении влияния значений факторов на переход объекта управления в состояния, соответствующие классам.

Основываясь на работах [3, 16, 18] рассмотрим, математический метод АСК-анализа, обеспечивающей решение поставленных задач. Очень краткое и несколько упрощенное описание автоматизированного системно-когнитивного анализа (АСК-анализ) приведено в работе [27].

В работе [15] (и ряде других) развита идея системного обобщения математики и обоснована актуальность этой идеи. Эта идея актуальна по ряду причин разного рода.

Во-первых, потому, что в мире нет ничего кроме систем, а понятие множества является абстракцией от понятия системы: множество – это система без внутренней структуры. Поэтому математика, основанная на понятии системы, имеет некоторые шансы быть более адекватной, чем классическая математика, в очень большой степени основанная на понятии множества.

Во-вторых, идея системного обобщения математики *частично* реализована в теории информации, в результате получены некоторые результаты в области системной теории информации (СТИ), в частности получен вариант выражения для семантической меры целесообразности информации А.Харкевича, удовлетворяющий принципу соответствия с формулой Р.Хартли для равновероятного детерминистского случая. Этим преодолена искусственная пропасть между «Теорией передачи данных по каналам связи», как совершенно справедливо называл свою теорию К.Шеннон, интуитивно понимавший различие между данными и информацией, и семантической теорией информации А.Харкевича¹⁴.

В-третьих, в созданной системной теории информации получены разнообразные формы различных коэффициентов эмерджентности: Хартли, Харкевича, Шеннона, для классических систем, подчиняющихся статистике Л.Больцмана [15] и квантовых систем подчиняющихся статистикам Ферми-Дирака и Бозе-Эйнштейна [15]. Смысл этих коэффициентов раскрыт в работе [15] и других. Если резюмировать, то можно сказать, что эти подходы, по-видимому, открывают новые подходы математического моделирования процессов эволюции систем различного рода и масштаба от микро до макро и мега уровней [8, 15, 16] и другие¹⁵.

Математический метод АСК-анализа основан на системной теории информации (СТИ), которая создана в рамках реализации программной идеи обобщения всех понятий математики, в частности теории информации, базирующихся на теории множеств, путем тотальной замены понятия множества на более общее понятие системы и тщательного отслеживания всех последствий этой замены [15]. Благодаря математическому методу, положенному в основу АСК-анализа, этот метод является непараметрическим и позволяет в реализующей его системе «Эйдос-Х++» сопоставимо обрабатывать десятки и сотни тысяч градаций факторов и будущих состояний нелинейных [28] многопараметрических объектов управления (классов) при неполных (фрагментированных), зашумленных данных числовой и нечисловой природы измеряемых в различных единицах измерения [3, 15].

¹⁴ На наличие этой пропасти, как недостаток прагматической теории информации А.Харкевича, еще в 2003 году указывал д.т.н., проф. В.И.Ключко. Но именно тогда автором и был предложен и обоснован вариант ее преодоления [3].

¹⁵ См., например: <http://lc.kubagro.ru/>, http://elibrary.ru/author_items.asp?authorid=123162

Итак, будем считать, что информация содержится не только в самих базовых элементах системы, но и в ее подсистемах различной сложности, т.е. состоящих из 2, 3, ... m, ... M базовых элементов.

Классическая формула Хартли имеет вид [32]:

$$I = \text{Log}_2 W \quad (1)$$

Будем искать ее системное обобщение в виде [3]:

$$I = \text{Log}_2 W^\varphi \quad (2)$$

где:

W – количество элементов в множестве.

I – количество информации, которое содержится в факте извлечения одного элемента из множества.

φ – коэффициент эмерджентности, названный автором в честь Р.Хартли, коэффициентом эмерджентности Хартли¹⁶.

Суммарное количество таких подсистем для систем, подчиняющихся статистике Ферми-Дирака [15, 29], можно принять равным числу сочетаний. Поэтому *примем*, что системное обобщение формулы Хартли имеет вид:

$$I = \text{Log}_2 \sum_{m=1}^M C_W^m \quad (3)$$

где:

C_W^m – количество подсистем из m элементов;

m – сложность подсистем;

M – *максимальная* сложность подсистем (максимальное число элементов подсистемы).

Так как $C_W^1 = W$, то при $M=1$ система переходит в множество и выражение (3) приобретает вид (1), т.е. для него выполняется *принцип соответствия*, являющийся обязательным для более общей теории.

Учитывая, что при $M=W$:

$$\sum_{m=1}^M C_W^m = 2^W - 1 \quad (4)$$

в этом случае получаем:

$$I = \text{Log}_2 (2^W - 1) \quad (5)$$

Выражение (5) дает *оценку максимального количества информации* в элементе системы. Из выражения (5) видно, что при увеличении числа элементов W количество информации I быстро стремится к W (6) и уже при $W>4$ погрешность выражения (5) не превышает 1%:

¹⁶ См.: В.Вяткин. Групповой плагиат: от студента до министра. – [Электронный ресурс]. Адрес доступа: <http://trv-science.ru/2011/11/08/gruppovoj-plagiat-ot-studenta-do-ministra/>

$$\lim_{W \rightarrow \infty} I/W = 1 \quad (6)$$

Приравняв правые части выражений (2) и (3):

$$I = \text{Log}_2 W^\varphi = \text{Log}_2 \sum_{m=1}^M C_W^m \quad (7)$$

получим выражение для коэффициента эмерджентности Хартли:

$$\varphi = \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (8)$$

Смысл этого коэффициента весьма интересен и раскрыт в работах [3, 15] и ряде других¹⁷. Здесь отметим лишь, что при $M \rightarrow 1$, когда система асимптотически переходит в множество, имеем $\varphi \rightarrow 1$ и (2) \rightarrow (1), как и должно быть согласно принципу соответствия, предложенному Нильсом Бором в 1913 году.

С учетом (8) выражение (2) примет вид:

$$I(W, M) = \text{Log}_2 W \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (9)$$

или при $M=W$ и больших W , учитывая (4) и (5):

$$I(W, M) = \text{Log}_2 W \frac{W}{\text{Log}_2 W} = W \quad (10)$$

Выражение (9) и представляет собой искомое системное обобщение классической формулы Хартли, а выражение (10) – его достаточно хорошее приближение при большом количестве элементов в системе W .

Классическая формула А. Харкевича имеет вид:

$$I_{ij}(W, M) = \text{Log}_2 \frac{P_{ij}}{P_{\Sigma j}} \quad (11)$$

где: – P_{ij} – условная вероятность перехода объекта в j -е состояние *при условии* действия на него i -го значения фактора;

– $P_{\Sigma j}$ – безусловная вероятность перехода объекта в j -е состояние (вероятность самопроизвольного перехода или вероятность перехода, посчитанная по всей выборке, т.е. при действии *любого* значения фактора).

Придадим выражению (11) следующий *эквивалентный* вид (12), который и будем использовать ниже. Вопрос об эквивалентности выражений (11) и (12) рассмотрим позднее.

¹⁷ См., например: <http://lc.kubagro.ru/aidos/Aidos-X.htm>, искать: «эмерджентн»

$$I_{ij}(W, M) = \text{Log}_2 \frac{P_{ij}}{P_{i\Sigma}} \quad (12)$$

где: – индекс i обозначает признак (значение фактора): $1 \leq i \leq M$;
– индекс j обозначает состояние объекта или класс: $1 \leq j \leq W$;
– P_{ij} – условная вероятность наблюдения i -го значения фактора у объектов в j -го класса;
– $P_{i\Sigma}$ – безусловная вероятность наблюдения i -го значения фактора по всей выборке.

Из (12) видно, что формула Харкевича для семантической меры информации по сути является логарифмом от формулы Байеса для апостериорной вероятности (отношение условной вероятности к безусловной).

Известно, что классическая формула Шеннона для количества информации для неравновероятных событий преобразуется в формулу Хартли при условии, что события равновероятны, т.е. удовлетворяет фундаментальному *принципу соответствия*. Поэтому теория информации Шеннона справедливо считается обобщением теории Хартли для неравновероятных событий. Однако, выражения (11) и (12) при подстановке в них реальных численных значений вероятностей P_{ij} , $P_{i\Sigma}$ и $P_{\Sigma j}$ не дает количества информации в **битах**, т.е. для этого выражения не выполняется *принцип соответствия*, обязательный для более общих теорий. Возможно, в этом состоит причина довольно сдержанного, а иногда и скептического отношения специалистов по теории информации Шеннона к семантической теории информации Харкевича.

Причину этого мы видим в том, что в выражениях (11) и (12) отсутствуют глобальные параметры **конкретной** модели W и M , т.е. в том, что А. Харкевич в своем выражении для количества информации не ввел зависимости от мощности пространства будущих состояний объекта W и количества значений факторов M , обуславливающих переход объекта в эти состояния.

Поставим задачу получить такое обобщение формулы Харкевича, которое бы удовлетворяло **тому же самому** *принципу соответствия*, что и формула Шеннона, т.е. преобразовывалось в формулу Хартли в предельном детерминистском равновероятном случае, когда каждому классу (состоянию объекта) соответствует один признак (значение фактора), и каждому признаку – один класс, и эти классы (а, значит и признаки), **равновероятны**, и при этом каждый фактор однозначно, т.е. **детерминистским** образом определяет переход объекта в определенное состояние, соответствующее классу.

В детерминистском случае вероятность P_{ij} наблюдения объекта j -го класса при обнаружении у него i -го признака:

$$P_{ij} = \begin{cases} 1, & \text{при } i = j \\ 0, & \text{при } i \neq j \end{cases}$$

Будем искать это обобщение (12) в виде:

$$I_{ij}(W, M) = \text{Log}_2 \left(\frac{P_{ij}}{P_{i\Sigma}} \right)^\Psi \quad (13)$$

Найдем такое выражение для коэффициента Ψ , названного автором в честь А. Харкевича "коэффициентом эмерджентности Харкевича"¹⁸, которое обеспечивает выполнение для выражения (13) принципа соответствия с классической формулой Хартли (1) и ее системным обобщением (2) и (3) в *равновероятном детерминистском* случае.

Для этого нам потребуется выразить вероятности P_{ij} , P_j и P_i через частоты наблюдения признаков по классам (см. табл. 6). В табл. 1 рамкой обведена область значений, переменные определены ранее.

Таблица 6 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

		Классы					Сумма
		<i>I</i>	...	<i>j</i>	...	<i>W</i>	
Значения факторов	<i>I</i>	N_{11}		N_{1j}		N_{1W}	
	...						
	<i>i</i>	N_{i1}		N_{ij}		N_{iW}	$N_{i\Sigma} = \sum_{j=1}^W N_{ij}$
	...						
	<i>M</i>	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков				$N_{\Sigma j} = \sum_{i=1}^M N_{ij}$			$N_{\Sigma\Sigma} = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$

Алгоритм формирования матрицы абсолютных частот.

Объекты обучающей выборки описываются векторами (массивами) $\vec{L} = \{L_i\}$ имеющихся у них признаков:

$$\vec{L} = \{L_i\} = n, \text{ если у объекта } i\text{-й признак встречается } n \text{ раз.}$$

Первоначально в матрице абсолютных частот все значения равны нулю. Затем организуется цикл по объектам обучающей выборки. Если предъявленного объекта, относящегося к *j*-му классу, есть *i*-й признак, то:

$$N_{ij} = N_{ij} + 1; N_{i\Sigma} = N_{i\Sigma} + 1; N_{\Sigma j} = N_{\Sigma j} + 1; N_{\Sigma\Sigma} = N_{\Sigma\Sigma} + 1$$

Здесь можно провести очень интересную и важную аналогию между способом формирования матрицы абсолютных частот и работой *многоканальной системы выделения полезного сигнала из шума*. Представим себе, что все объекты, предъявляемые для формирования обобщенного образа некоторого класса, в действительности являются различными реализациями одного объекта – "Эйдоса" в смысле Платона [30], поразному зашумленного различными случайными обстоятельствами. И наша задача состоит в том, чтобы подавить этот шум и выделить из него то общее и существенное, что отличает объекты данного класса от объектов других классов. Учитывая, что шум чаще всего является "белым" и имеет свойство при суммировании с самим собой стремиться

¹⁸ См.: В.Вяткин. Групповой плагиат: от студента до министра. – [Электронный ресурс]. Адрес доступа: <http://trv-science.ru/2011/11/08/grupповојј-plagiат-от-studenta-do-ministra/>

к нулю, а сигнал при этом, наоборот, возрастает пропорционально количеству слагаемых, то увеличение объема обучающей выборки приводит ко все лучшему отношению сигнал/шум в матрице абсолютных частот, т.е. к выделению полезной информации из шума. Примерно так мы начинаем постепенно понимать смысл фразы, которую мы сразу не расслышали по телефону и несколько раз переспрашивали. При этом в повторах шум не позволяет понять то одну, то другую часть фразы, но в конце концов за счет использования памяти и интеллектуальной обработки информации мы понимаем ее всю. Так и *объекты, описанные признаками, можно рассматривать как зашумленные фразы, несущие нам информацию об обобщенных образах классов - "Эйдосах" [30], к которым они относятся. И эту информацию мы выделяем из шума при синтезе модели.*

Для выражения (11):

$$P_{ij} = \frac{N_{ij}}{N_{i\Sigma}} \quad (14)$$

Для выражений (12) и (13):

$$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}} \quad (15)$$

Для выражений (11), (12) и (13):

$$\begin{aligned} P_i &= \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}; P_j = \frac{N_{\Sigma j}}{N_{\Sigma\Sigma}}; \\ N_{i\Sigma} &= \sum_{j=1}^W N_{ij}; N_{\Sigma j} = \sum_{i=1}^M N_{ij}; \\ N_{\Sigma\Sigma} &= \sum_{i=1}^M N_{i\Sigma} = \sum_{j=1}^W N_{\Sigma j} = \sum_{i=1}^W \sum_{j=1}^M N_{ij} \end{aligned} \quad (16)$$

В (16) использованы обозначения:

N_{ij} – суммарное количество наблюдений в исследуемой выборке *факта*: "действовало i -е значение фактора и объект перешел в j -е состояние";

$N_{\Sigma j}$ – суммарное по всей выборке количество встреч различных факторов у объектов, перешедших в j -е состояние;

$N_{i\Sigma}$ – суммарное количество встреч i -го фактора у всех объектов исследуемой выборки;

$N_{\Sigma\Sigma}$ – суммарное количество встреч различных значений факторов у всех объектов исследуемой выборки.

Формирование матрицы условных и безусловных процентных распределений.

На основе анализа матрицы частот (табл. 1) классы можно сравнивать по наблюдаемым частотам признаков только в том случае, если количество объектов по всем классам *одинаково*, как и *суммарное количество признаков по классам*. Если же они отличаются, то корректно сравнивать классы можно только по условным и безусловным относительным частотам (оценкам вероятностей) наблюдений признаков, посчитанных на основе матрицы частот (табл. 1) в соответствии с выражениями (14) и (15), в

результате чего получается матрица условных и безусловных процентных распределений (табл. 7).

При расчете матрицы оценок условных и безусловных вероятностей N_j из табл. 1 могут браться либо из предпоследней, либо из последней строки. В 1-м случае N_j представляет собой "Суммарное количество признаков у всех объектов, использованных для формирования обобщенного образа j -го класса", а во 2-м случае - это "Суммарное количество объектов обучающей выборки, использованных для формирования обобщенного образа j -го класса", соответственно получаем различные, хотя и очень сходные семантические информационные модели, которые мы называем СИМ-1 и СИМ-2. Оба этих вида моделей поддерживаются системой "Эйдос".

Таблица 7 – МАТРИЦА УСЛОВНЫХ И БЕЗУСЛОВНЫХ ПРОЦЕНТНЫХ РАСПРЕДЕЛЕНИЙ

		Классы					Безусловная вероятность признака
		1	...	j	...	w	
Значения факторов	1	P_{11}		P_{1j}		P_{1w}	
	...						
	i	P_{i1}		P_{ij}		P_{iw}	$P_{i\Sigma}$
	...						
	M	P_{M1}		P_{Mj}		P_{Mw}	
Безусловная вероятность класса				$P_{\Sigma j}$			

Эквивалентность выражений (11) и (12) устанавливается, если подставить в них выражения относительных частот как оценок вероятностей P_{ij} , $P_{\Sigma j}$ и $P_{i\Sigma}$ через абсолютные частоты наблюдения признаков по классам из (14), (15) и (16). В обоих случаях из выражений (11) и (12) получается одно и то же выражение (17):

$$I_{ij} = \text{Log}_2 \frac{N_{ij} N_{\Sigma\Sigma}}{N_{i\Sigma} N_{\Sigma j}} \quad (17)$$

А из (13) - выражение (18), с которым мы и будем далее работать.

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij} N_{\Sigma\Sigma}}{N_{i\Sigma} N_{\Sigma j}} \right)^\Psi \quad (18)$$

При взаимно-однозначном соответствии классов и признаков в *равновероятном детерминистском* случае имеем (таблица 8):

Таблица 8 – МАТРИЦА ЧАСТОТ В РАВНОВЕРОЯТНОМ ДЕТЕРМИНИСТСКОМ СЛУЧАЕ

		Классы					Сумма
		<i>l</i>	...	<i>j</i>	...	<i>W</i>	
Значения факторов	<i>l</i>	1					1
	...		1				1
	<i>i</i>			1			1
	...				1		1
	<i>M</i>					1	1
Сумма		1	1	1	1	1	$N_{\Sigma\Sigma}$

В этом случае к каждому классу относится один объект, имеющий единственный признак. Откуда получаем для всех *i* и *j* равенства (19):

$$\forall ij: N_{ij} = N_{i\Sigma} = N_{\Sigma j} = 1 \quad (19)$$

Таким образом, обобщенная формула А. Харкевича (18) с учетом (19) в этом случае приобретает вид:

$$I_{ij} = \text{Log}_2 N_{\Sigma\Sigma}^{\Psi} = \text{Log}_2 W^{\varphi} \quad (20)$$

откуда:

$$\Psi = \frac{\text{Log}_2 W^{\varphi}}{\text{Log}_2 N_{\Sigma\Sigma}} \quad (21)$$

или, учитывая выражение для коэффициента эмерджентности Хартли (8):

$$\Psi = \frac{\text{Log}_2 W^{\varphi}}{\frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W}} \quad (22)$$

Выражения (21) и (22) получены автором в 2002 году [3] и названы коэффициентом эмерджентности А.Харкевича, т.к. имеют очевидную связь с его формулой для количества информации. Эти коэффициенты имеют весьма глубокий смысл, который автор попытался раскрыть в работах [3, 15] и ряде других.

Подставив коэффициент эмерджентности А.Харкевича (21) в выражение (18), получим:

Отметим, что 1-я задача получения системного обобщения формул Хартли и Харкевича и 2-я задача получения такого обобщения формулы Харкевича, которая удовлетворяет принципу соответствия с формулой Хартли – это две разные задачи. 1-я задача является более общей и при ее решении, которое приведено выше, *автоматически* решается и 2-я задача, которая является, таким образом, частным случаем 1-й.

Однако представляет самостоятельный интерес и частный случай, в результате которого получается формула Харкевича, удовлетворяющая в *равновероятном детерминистском* случае принципу соответствия с классической формулой Хартли (1), а не с ее системным обобщением (2) и (3). Ясно, что эта формула получается из (23) при $\varphi=1$.

$$\begin{aligned}
I_{ij} &= \text{Log}_2 \left(\frac{N_{ij} N_{\Sigma}}{N_{i\Sigma} N_{\Sigma j}} \right)^{\Psi} = \text{Log}_2 \left(\frac{N_{ij} N_{\Sigma\Sigma}}{N_{i\Sigma} N_{\Sigma j}} \right)^{\frac{\text{Log}_2 W^{\varphi}}{\text{Log}_2 N}} = \\
&= \frac{\text{Log}_2 W^{\varphi}}{\text{Log}_2 N_{\Sigma\Sigma}} \left(\text{Log}_2 \left(\frac{N_{ij}}{N_{i\Sigma} N_{\Sigma j}} \right) + \text{Log}_2 N_{\Sigma\Sigma} \right) = \\
&= \text{Log}_2 \left(\frac{N_{ij}}{N_{i\Sigma} N_{\Sigma j}} \right)^{\frac{\text{Log}_2 W^{\varphi}}{\text{Log}_2 N_{\Sigma\Sigma}}} + \text{Log}_2 W^{\varphi}
\end{aligned}$$

или окончательно:

$$\boxed{I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_{i\Sigma} N_{\Sigma j}} \right)^{\frac{\text{Log}_2 W^{\varphi}}{\text{Log}_2 N_{\Sigma\Sigma}}} + \text{Log}_2 W^{\varphi}} \quad (23)$$

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_{i\Sigma} N_{\Sigma j}} \right)^{\frac{\text{Log}_2 W}{\text{Log}_2 N_{\Sigma\Sigma}}} + \text{Log}_2 W \quad (24)$$

Из выражений (21) и (22) видно, что в этом частном случае, т.е. когда система эквивалентна множеству ($M=1$), коэффициент эмерджентности А.Харкевича приобретает вид:

$$\Psi = \frac{\text{Log}_2 W}{\text{Log}_2 N_{\Sigma\Sigma}} \quad (25)$$

На практике для численных расчетов удобнее пользоваться не выражениями (23) или (24), а **формулой (26)**, которая получается непосредственно из (18) после подстановки в него выражения (25):

$$\boxed{I_{ij} = \frac{\text{Log}_2 W}{\text{Log}_2 N_{\Sigma\Sigma}} \times \text{Log}_2 \frac{N_{ij} N_{\Sigma\Sigma}}{N_{i\Sigma} N_{\Sigma j}}} \quad (26)$$

Используя выражение (26) и данные таблицы 1 непосредственно прямым счетом получаем *матрицу знаний* (таблица 9):

Таблица 9 – МАТРИЦА ЗНАНИЙ (ИНФОРМАТИВНОСТЕЙ)

		Классы					Значимость фактора
		<i>I</i>	...	<i>j</i>	...	<i>W</i>	
Значения факторов	<i>1</i>	I_{11}		I_{1j}		I_{1W}	$s_{1\Sigma} = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	<i>i</i>	I_{i1}		I_{ij}		I_{iW}	$s_{i\Sigma} = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	<i>M</i>	I_{M1}		I_{Mj}		I_{MW}	$s_{M\Sigma} = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$s_{\Sigma 1}$		$s_{\Sigma j}$		$s_{\Sigma W}$	$H = \sqrt{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Здесь – \bar{I}_i ; это среднее количество знаний в *i*-м значении фактора:

$$\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ij}$$

Когда количество информации $I_{ij} > 0$ – *i*-й фактор способствует переходу объекта управления в *j*-е состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это. В векторе *i*-го фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе *j*-го состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, матрица знаний (информативностей), приведенная в таблице 6, является обобщенной таблицей решений, в которой входы (факторы) и выходы (будущие состояния объекта управления) связаны друг с другом не с помощью классических (Аристотелевых) импликаций, принимающих только значения: "истина" и "ложь", а различными значениями истинности, выраженными в битах, и принимающими значения от положительного теоретически-максимально-возможного ("максимальная степень истинности"), до теоретически неограниченного отрицательного ("степень ложности"). Это позволяет автоматически формулировать прямые и опосредованные правдоподобные высказывания с расчетной степенью истинности.

Фактически предложенная модель позволяет осуществить синтез обобщенных таблиц решений для различных предметных областей непосредственно на основе эмпирических исходных данных и продуцировать прямые и обратные правдоподобные (нечеткие) логические рассуждения по неклассическим схемам с различными расчетными значениями истинности, являющимися обобщением классических импликаций.

Таким образом, данная модель позволяет рассчитать, какое количество информации содержится в любом факте о наступлении любого события в любой предметной области, причем для этого не требуется повторности этих фактов и событий. Если дан-

ные повторности осуществляются и при этом наблюдается некоторая вариабельность значений факторов, обуславливающих наступление тех или иных событий, то модель обеспечивает многопараметрическую типизацию, т.е. синтез обобщенных образов классов или категорий наступающих событий с количественной оценкой степени и знака влияния на их наступление различных значений факторов. Причем эти значения факторов могут быть как количественными, так и качественными и измеряться в любых единицах измерения, в любом случае в модели оценивается количество информации, которое в них содержится о наступлении событий, переходе объекта управления в определенные состояния или, просто, о его принадлежности к тем или иным классам. Другие способы метризации приведены в работе [9]. Все они реализованы в системно-когнитивном анализе и интеллектуальной системе «Эйдос» и обеспечивают сопоставление градациям всех видов шкал числовых значений, имеющих смысл количества информации в градации о принадлежности объекта к классу. Поэтому является корректным применение интегральных критериев, включающих операции умножения и суммирования, для обработки числовых значений, соответствующих градациям шкал. Это позволяет единообразно и сопоставимо обрабатывать эмпирические данные, полученные с помощью любых типов шкал, применяя при этом все математические операции.

Информационный портрет класса – это список значений факторов, ранжированных в порядке убывания силы их влияния на переход объекта управления в состояние, соответствующее данному классу. Информационный портрет класса отражает систему его детерминации. Генерация информационного портрета класса представляет собой решение обратной задачи прогнозирования, т.к. при прогнозировании по системе факторов определяется спектр наиболее вероятных будущих состояний объекта управления, в которые он может перейти под влиянием данной системы факторов, а в информационном портрете мы, наоборот, по заданному будущему состоянию объекта управления определяем систему факторов, детерминирующих это состояние, т.е. вызывающих переход объекта управления в это состояние. В начале информационного портрета класса идут факторы, оказывающие положительное влияние на переход объекта управления в заданное состояние, затем факторы, не оказывающие на это существенного влияния, и далее – факторы, препятствующие переходу объекта управления в это состояние (в порядке возрастания силы препятствования). Информационные портреты классов могут быть от *отфильтрованы* по диапазону факторов, т.е. мы можем отобразить влияние на переход объекта управления в данное состояние не всех отраженных в модели факторов, а только тех, коды которых попадают в определенный диапазон, например, относящиеся к определенным описательным шкалам.

Информационный (семантический) портрет фактора – это список классов, ранжированный в порядке убывания силы влияния данного фактора на переход объекта управления в состояния, соответствующие данным классам. Информационный портрет фактора называется также его *семантическим портретом*, т.к. в соответствии с концепцией смысла системно-когнитивного анализа, являющейся обобщением концепции смысла Шенка-Абельсона [21], *смысл фактора состоит в том, какие будущие состояния объекта управления он детерминирует или обуславливает*. Сначала в этом списке идут состояния объекта управления, на переход в которые данный фактор оказывает наибольшее влияние, затем состояния, на которые данный фактор не оказывает существенного влияния, и далее состояния – переходу в которые данный фактор препятствует. Информационные портреты факторов могут быть от *отфильтрованы* по диапазону классов, т.е. мы можем отобразить влияние данного фактора на переход объекта управления не во все возможные будущие состояния, а только в состояния, коды которых попадают в определенный диапазон, например, относящиеся к определенным классификационным шкалам.

Прямые и обратные, непосредственные и опосредованные правдоподобные логические рассуждения с расчетной степенью истинности в системной теории информации.



Перевалов Михаил Ильич,
учитель математики
СШ №31 г.Краснодара,
примерно 1970 год¹⁹

Одним из первых ученых, поднявших и широко обсуждавшим в своих работах проблематику правдоподобных рассуждений, был известный венгерский, швейцарский и американский математик Дьердь Пойа [31], книги которого подарил автору его школьный учитель математики Михаил Ильич Перевалов, за что автор ему очень благодарен (см. также раздел «Формализация логики правдоподобных рассуждений Д. Пойа», глава третья, параграф 7, с.158-163, исходящий из (репрезентативной) теории измерений).

Разве мог он тогда предположить, что через много лет в работе [3]²⁰ *им будет предложена логическая форма представления правдоподобных логических рассуждений с расчетной степенью истинности, которая определяется в соответствии с системной теорией информации непосредственно на основе эмпирических данных.*

В качестве количественной меры влияния факторов, предложено использовать обобщенную формулу А.Харкевича, полученную на основе предложенной эмерджентной теории информации. При этом непосредственно из матрицы абсолютных частот рассчитывается база знаний (табл.4), которая и представляет собой основу содержательной информационной модели предметной области.

Весовые коэффициенты табл.4 непосредственно определяют, какое количество информации I_{ij} система управления получает о наступлении события: "активный объект управления перейдет в j -е состояние", из сообщения: "на активный объект управления действует i -й фактор".

Принципиально важно, что эти весовые коэффициенты не определяются экспертами неформализуемым способом на основе интуиции и профессиональной компетенции (т.е. фактически «на глазок»), а рассчитываются непосредственно на основе эмпирических данных и удовлетворяют всем ранее обоснованным в работе [3] требованиям, т.е. являются сопоставимыми, содержательно интерпретируемыми, отражают понятия "достижение цели управления" и "мощность множества будущих состояний объекта управления" и т.д.

В работе [3] обосновано, что предложенная информационная мера обеспечивает сопоставимость индивидуальных количеств информации, содержащейся в факторах о классах, а также сопоставимость интегральных критериев, рассчитанных для одного объекта и разных классов, для разных объектов и разных классов.

Когда количество информации $I_{ij} > 0$ – i -й фактор способствует переходу объекта управления в j -е состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это. В векторе i -го фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векто-

¹⁹ Фото автора

²⁰ Примерно через 33 года. Все же, наверное, мог предположить. Иначе зачем было бы дарить эти книги.

ре j -го состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, матрица информативностей (табл.4) является обобщенной таблицей решений, в которой входы (факторы) и выходы (будущие состояния активного объекта управления (АОУ) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а **различными значениями истинности, выраженными в битах** и принимающими значения от положительного теоретически-максимально-возможного ("Максимальная степень истинности"), до теоретически неограниченного отрицательного ("Степень ложности").

Фактически предложенная модель позволяет осуществить синтез обобщенных таблиц решений для различных предметных областей непосредственно на основе эмпирических исходных данных и продуцировать на их основе прямые и обратные правдоподобные (нечеткие) логические рассуждения по неклассическим схемам с различными расчетными значениями истинности, являющимся обобщением классических импликаций (табл. 10).

Таблица 10 – Прямые и обратные правдоподобные логические высказывания с расчетной в соответствии с системной теорией информации (СТИ) степенью истинности импликаций

	Прямые высказывания:	Обратные высказывания
1	если A, то B (если действует фактор A , то мы предполагаем с степенью истинности I_{AB} , что АОУ перейдет в состояние B)	если B, то A (если АОУ перешел в состояние B , то мы предполагаем с степенью истинности I_{AB} , что действовал фактор A)
2	если A_1 и A_2 ... и A_M, то B (прогноз влияния системы факторов на поведение АОУ. Степень истинности обобщающей (итоговой) импликации равна алгебраической сумме истинностей составляющих ее элементарных импликаций вида: "если A то B ")	если B, то A_1 и A_2 ... и A_M (информационный портрет класса B , т.е. влияние различных факторов A_i на переход АОУ в будущее состояние B , решение обратной задачи прогнозирования, т.е. выработка управления)
3	если A, то B_1 или B_2 ... или B_W (семантический портрет фактора A , т.е. его влияние на переход АОУ в различные состояния)	
4	если A_1 и A_2 ... и A_M, то B_1 или B_2 ... или B_W (прогноз влияния системы факторов на переход АОУ в различные состояния)	

Приведем пример более сложного высказывания, которое может быть рассчитано непосредственно на основе матрицы информативностей – обобщенной таблицы решений (табл. 4): «Если A , со степенью истинности $\alpha(A,B)$, детерминирует B , и если C , со степенью истинности $\alpha(C,D)$, детерминирует D , и A совпадает по смыслу с C со степенью истинности $\alpha(A,C)$, то это вносит вклад в совпадение B с D , равный степени истинности $\alpha(B,D)$ ».

При этом в прямых рассуждениях как предпосылки рассматриваются факторы, а как заключение – будущие состояния АОУ, а в обратных – наоборот: как предпосылки – будущие состояния АОУ, а как заключение – факторы. Степень истинности i -й предпосылки – это просто количество информации I_{ij} , содержащейся в ней о наступлении j -го будущего состояния АОУ. Если предпосылок несколько, то степень истинности наступления j -го состояния АОУ равна суммарному количеству информации, содержащемуся в них об этом. Количество информации в i -м факторе о наступлении j -го состояния АОУ, рассчитывается в соответствии с выражениями системной теории информации (СТИ).

Прямые правдоподобные логические рассуждения позволяют прогнозировать степень достоверности наступления события по действующим факторам, а обратные – по заданному состоянию восстановить степень необходимости и степень нежелательности каждого фактора для наступления этого состояния, т.е. принимать решение по выбору управляющих воздействий на АОУ, оптимальных для перевода его в заданное целевое состояние.

Приведем простой пример, когда безупречная классическая бинарная логика Аристотеля дает сбой. Рассмотрим высказывания:

А) если студент хорошо сдал экзамен по информационным системам, значит, он умеет хорошо программировать;

Б) если студент умеет хорошо программировать, то он может стать хорошим специалистом в области прикладной информатики.

Откуда средствами логики предикатов получаем вывод:

В) если студент хорошо сдал экзамен по информационным системам, то он может стать хорошим специалистом в области прикладной информатики.

Если при рассмотрении каждого высказывания «А» и «Б» по отдельности у нас не возникает особых возражений, хотя мы сразу чувствуем здесь какой-то подвох, что это не совсем так или не всегда так и легко можем привести вполне реальные примеры, когда эти высказывания могут быть и ложными, то высказывание «В» уже само по себе выглядит очень сомнительным, т.е. проще говоря ложным, тогда как в логике предикатов оно является истинным. Интуитивно мы хорошо понимаем, почему так получается. Дело в том, что в этих высказываниях не отражен *контекст*, т.е. та огромная слабо формализованная и вообще неформализованная информация об объекте моделирования, которой располагает человек, но не располагает логическая система. Например, в этих двух логических высказываниях не отражена информация, которой располагает каждый преподаватель и студент, о том, каким образом иногда сдаются экзамены, когда оценка вообще никак не зависит от знаний. Иначе говоря, чтобы эти высказывания были истинны необходимо, чтобы оценка определялась только знаниями. Но и этого мало. Предполагается, что факт получения хорошей оценки по дисциплине означает *полное* ее освоение, хотя все понимают, что для этого достаточно освоения только тех вопросов, которые были в билете и были заданы преподавателем (кроме того существуют и весьма распространены и другие варианты).

При решении этой же задачи средами АСК-анализа мы формулируем эти высказывания *в форме правдоподобных рассуждений*:

А) если студент хорошо сдал экзамен по информационным системам, то в этом факте содержится I(A) информации о том, что он умеет хорошо программировать;

Б) если студент умеет хорошо программировать, то в этом факте содержится I(B) информации о том он может стать хорошим специалистом в области прикладной информатики.

Откуда средствами АСК-анализа получаем результирующее высказывание:

В) если студент хорошо сдал экзамен по информационным системам, то в этом факте содержится I(V) информации о том он может стать хорошим специалистом в области прикладной информатики.

Это высказывание не выглядит как истинное или ложное и может быть и истинным, и ложным, причем в различной степени, в зависимости от знака и модуля его расчетной степени истинности $I(V)=F(I(A), I(B))$. Для расчета этой величины нужны конкретные эмпирические данные, являющиеся репрезентативными для отражения определенной предметной области (генеральной совокупности), в которой этот вывод и будет иметь эти значения знака и величины степени истинности.

Итак, описанная в данном разделе математическая модель реализована в программном инструментарии АСК-анализа – универсальной когнитивной аналитической системе «Эйдос-Х++». Как следует из самого названия системы это сделано в универсальной постановке не зависящей от предметной области. Поэтому система «Эйдос-Х++» может быть применена, и фактически и была применена, в самых различных предметных областях для построения интеллектуальных измерительных систем и интеллектуальных систем управления, а также для решения задач идентификации, прогнозирования и принятия решений [8, 16, 33-60, 62].

Более подробный численный пример применения данного математического метода и реализующей его системы «Эйдос-Х++» для построения интеллектуальной измерительной системы для системной идентификации состояний сложных систем приводится ниже.

6.1.3. Применение АСК-анализа и системы «Эйдос» для интеллектуальных измерений и идентификации состояний сложных нелинейных систем

В первом разделе данной статьи предлагается применить автоматизированный системно-когнитивный анализ (АСК-анализ) как для синтеза, так и для применения адаптивной интеллектуальной измерительной системы с целью измерения не значений параметров объектов, а для идентификации состояний сложных систем, т.е. для так называемой системной идентификации. Применение данного подхода является корректным для измерения состояний сложных многофакторных нелинейных динамических систем.

Во втором разделе статьи кратко рассматривается математический метод АСК-анализа, реализованный в его программном инструментарии – универсальной когнитивной аналитической системе «Эйдос-Х++». Математический метод АСК-анализа основан на системной теории информации (СТИ), которая создана в рамках реализации программной идеи обобщения всех понятий математики, в частности теории информации, базирующихся на теории множеств, путем тотальной замены понятия множества на более общее понятие системы и тщательного отслеживания всех последствий этой замены [3, 15]. Благодаря математическому методу, положенному в основу АСК-анализа, этот метод является непараметрическим и позволяет сопоставимо обрабатывать десятки и сотни тысяч градаций факторов и будущих состояний нелинейного объекта управления (классов) при неполных (фрагментированных), зашумленных данных числовой и нечисловой природы измеряемых в различных единицах измерения.

В данном разделе статьи приводится численный пример применения АСК-анализа и системы «Эйдос-Х++» как для синтеза системно-когнитивной модели, обеспечивающей многопараметрическую типизацию состояний сложных систем, так и для системной идентификации их состояний, а также для принятия решений об управляющем воздействии, так изменяющем *состав* объекта управления, чтобы его качество (уровень системности) максимально повышалось при минимальных затратах на это²¹. В данном разделе для численного примера в качестве сложной системы выбран коллектив фирмы, а его компонент – сотрудники и кандидаты (персонал). Однако необходимо отметить, что этот пример следует рассматривать шире, т.к. АСК-анализ и система «Эйдос» разрабатывались и реализовались в очень обобщенной постановке, постановке, не зависящей от предметной области, и с успехом могут быть применены и в других областях [4, 33-60, 62].

²¹ Что соответствует принципам Функционально-Стоимостного Анализа и метода Директ-Костинг.

6.1.3.1. РЕШЕНИЕ 1-й ЗАДАЧИ – МНОГОПАРАМЕТРИЧЕСКОЙ ТИПИЗАЦИИ СОСТОЯНИЙ СЛОЖНЫХ ОБЪЕКТОВ

Решение 1-й задачи является стандартным для системы «Эйдос», т.е. она предназначена для решения подобных задач и соответствующие применения описаны в работах автора [33, 60]²². В соответствии с этапами АСК-анализа и порядком преобразования данных в информацию, а ее в знания [22], рассмотрим Excel-таблицу исходных данных (таблица 11).

Таблица 11 – Исходные данные для синтеза системно-когнитивной модели управления качеством системы путем управления ее составом²³

Объект обучающей выборки	Пол	Откуда родом	Успеваемость	Длина волос	Цвет волос	Цвет глаз	Наличие украшений	Одежда	Макияж	Собственность	Посещаемость занятий
Бабенко ПИ-51	мужской	районный центр	менее 25%	Короткие	Русые	Голубые	Нет	джинсы	Отсутствует	телефон, автомобиль, компьютер,	Средняя
Воробьева ПИ-51	женский	поселок	менее 25%	Длинные	Крашенные	Зеленые	Цепочка	джинсы	Отсутствует	телефон	Хорошая
Жеребятёв ПИ51	мужской	поселок	от 50% до 75%	Средние	Очень светлые	Голубые	Нет	джинсы	Незаметный	телефон	Плохая
Заяц ПИ-51	женский	районный центр	менее 25%	Длинные	Каштановые	Зеленые	Серьги	джинсы	Незаметный	компьютер, телефон	Средняя
Иванова ПИ-51	женский	краевой центр	более 75%	Средние	Русые	Зеленые	Цепочка, Перстень, Серьги	джинсы	Незаметный	компьютер, телефон	Хорошая
Котенко ПИ-51	женский	районный центр	более 75%	Короткие	Каштановые	Карие	Цепочка, Серьги	джинсы	Незаметный	компьютер, телефон	Хорошая
Кузина О. ПИ-51	женский	краевой центр	от 50% до 75%	Короткие	Крашенные	Карие	Перстень	джинсы	Заметный	компьютер, телефон	Хорошая
Кузина Я. ПИ-51	женский	краевой центр	от 50% до 75%	Короткие	Крашенные		Перстень	Пиджак	Заметный		Хорошая
Лях ПИ-51	мужской	краевой центр	от 50% до 75%	Короткие	Русые	Серые	Цепочка	Пиджак, джинсы	Отсутствует	квартира, компьютер, телефон	Средняя
Мясников ПИ-51	мужской	краевой центр	от 25% до 50%	Короткие	Русые	Голубые	Цепочка, Браслет	джинсы	Отсутствует	квартира, телефон	Хорошая
Нагапетян ПИ-51	мужской	районный центр	от 25% до 50%	Короткие	Каштановые	Карие	Серьги	джинсы	Незаметный	квартира, компьютер, телефон	Плохая
Черкашина ПИ-51	женский	районный центр	менее 25%	Длинные	Каштановые	Карие	Цепочка	джинсы	Незаметный	квартира, компьютер, телефон	Плохая
Шульгин ПИ-51	мужской	поселок	от 50% до 75%	Короткие	Русые	Серые	Нет	Пиджак	Отсутствует	компьютер	Плохая
Веревкина ПИ-52	женский	краевой центр	от 25% до 50%	Короткие	Очень светлые	Серые	Серьги	джинсы	Незаметный	Ничего нет	Очень хорошая
Григорьева ПИ52	женский	районный центр	от 50% до 75%	Средние	Русые	Серые	Цепочка	джинсы	Заметный	Ничего нет	Очень хорошая
Еременко ПИ-52	женский	районный центр	от 50% до 75%	Средние	Русые	Зеленые, Серые	Цепочка, Серьги	джинсы	Незаметный	компьютер, телефон	Средняя
Иванова ПИ-52	женский	краевой центр	от 25% до 50%	Средние	Очень темные	Голубые	Перстень, Серьги	Пиджак, джинсы	Отсутствует	Ничего нет	Очень хорошая
Крейс ПИ-52	женский	районный центр	от 50% до 75%	Короткие	Русые	Серые	Серьги	Юбка	Незаметный	Ничего нет	Хорошая
Куркина ПИ-52	женский	краевой центр	от 50% до 75%	Длинные	Каштановые	Карие	Цепочка, Серьги	джинсы, Юбка	Заметный	компьютер, телефон	Хорошая
Люлик ПИ-52	женский	поселок	от 50% до 75%	Средние	Крашенные	Зеленые	Серьги	джинсы	Заметный	квартира, компьютер	Хорошая
Мануйлов ПИ-52	мужской	краевой центр	более 75%	Короткие	Русые	Серые	Перстень	джинсы	Отсутствует	квартира, автомобиль, компьютер, телефон	Плохая
Нарижний ПИ-52	мужской	краевой центр	более 75%	Короткие	Русые	Серые	Перстень	джинсы	Отсутствует	квартира, компьютер, телефон	Средняя
Паршакова ПИ-52	женский	село	от 50% до 75%	Средние	Каштановые	Карие	Цепочка	Юбка	Заметный	компьютер	Хорошая
Силенко ПИ-52	мужской	краевой центр	более 75%	Короткие	Каштановые	Зеленые	Цепочка	джинсы	Отсутствует	Ничего нет	Хорошая
Соколова ПИ-52	женский	районный центр	от 50% до 75%	Короткие	Русые	Зеленые	Нет	Пиджак, джинсы	Заметный	квартира, компьютер, телефон	Хорошая
Цисарь ПИ-52	женский	поселок	от 25% до 50%	Средние	Крашенные	Карие	Цепочка	джинсы	Заметный	телефон	Очень хорошая

В этой таблице колонки со 2-й по 4-ю являются классификационными измерительными шкалами, а с 5-й по последнюю – описательными измерительными шкалами. Каждая классификационная шкала представляет собой должность, а ее градации (клас-

²² Для удобства читателей ссылки на эти и другие работы даны с сайта автора:

<http://ic.kubagro.ru/> а также в РИНЦ: http://elibrary.ru/author_items.asp?authorid=123162

²³ Для численной иллюстрации излагаемых подходов в статье используется чрезвычайно упрощенный условный пример малой размерности, связанный с управлением персоналом фирмы. Таблица исходных данных представлена в графической форме с высоким разрешением, что обеспечивает хорошую читабельность при увеличении масштаба просмотра

сы) отражают различную степень успешности работы на этой должности. Описательные шкалы представляют собой личностные и профессиональные свойства сотрудников, а градации (признаки) – степень их выраженности. Отметим, что в приведенном упрощенном численном примере классы не являются профессиональными категориями с указанием степени успешности (например: МЕНЕДЖЕР ТОРГОВОГО ЗАЛА – хорошо подходит), а признаки респондентов не являются их личностными свойствами (например: ФАКТОР А: «ЗАМКНУТОСТЬ – ОБЩИТЕЛЬНОСТЬ» - 8 баллов). Поэтому от читателя требуется некоторая фантазия, чтобы представить себе, что это так. Но суть примера от этого не меняется, и он позволяет нам проиллюстрировать излагаемые в данной работе идеи. Каждому респонденту соответствует строка, в которой он описан и своими свойствами, и принадлежностью к классам. Таким образом, исходные данные в содержательной форме, представленные в таблице 1, эквивалентны такому количеству обучающих выборок, сколько возможно различных способов группировки данных, описанными описательными шкалами и градациями по классам. На основе этих данных и решается 1-я задача *многопараметрической типизации, результаты которой отражены в системно-когнитивной модели*. В системе «Эйдос» осуществляется **нормализация** базы исходных данных путем автоматической разработки классификационных и описательных шкал и градаций и кодирования с их использованием исходных данных и представления их в форме эвентологической базы данных (рисунок 5):

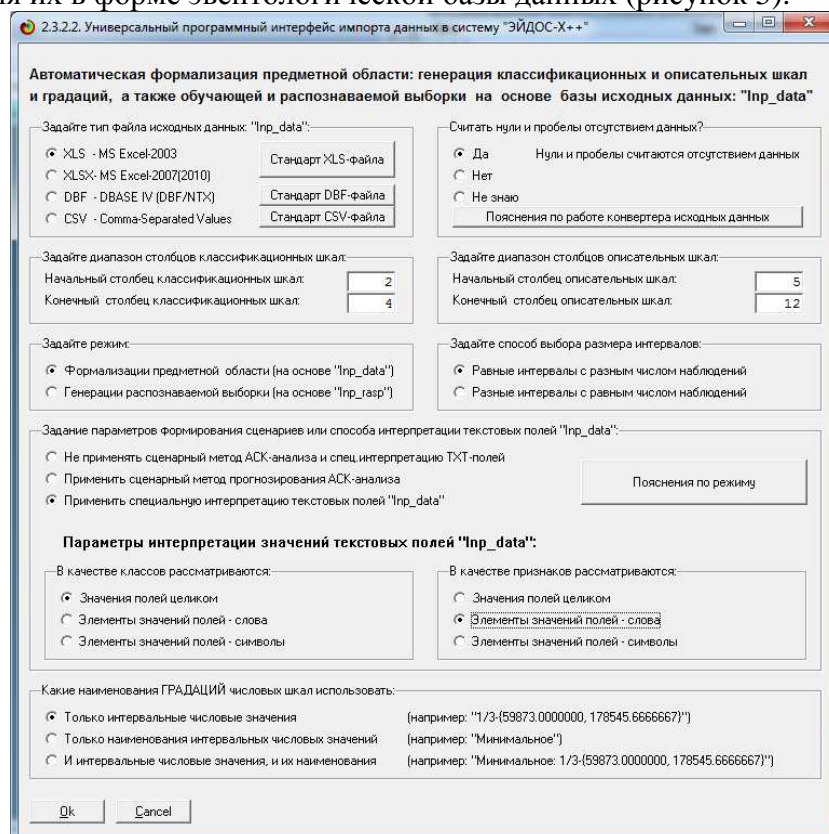


Рисунок 4 – Главная экранная форма для задания параметров импорта данных в систему «Эйдос»

из внешней базы исходных данных, представленной в таблице 1

Затем в соответствии с этапами АСК-анализа и порядком преобразования данных в информацию, а ее в знания [22], в режиме 3.5 системы «Эйдос» выполним синтез и верификацию статистических моделей и моделей знаний (рисунок 6):

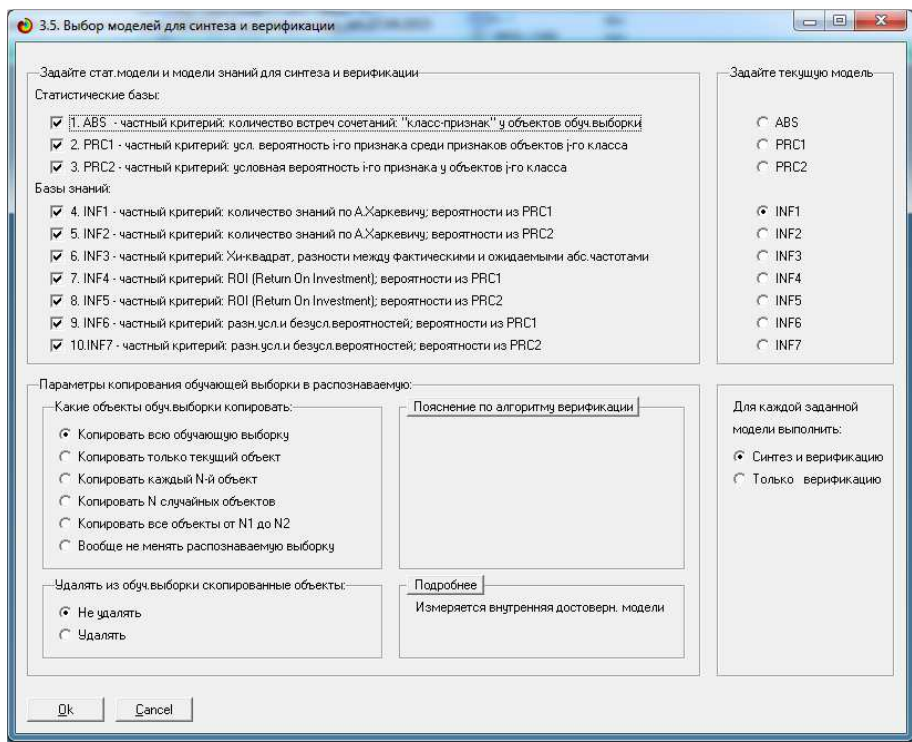


Рисунок 5 – Экранная форма задания параметров синтеза и верификации моделей (параметры по умолчанию)

Созданные модели, наименования которых приведены на рисунке 3, отличающиеся частными критериями, представляют собой результат *многопараметрической типизации* состояний объектов обучающей выборки, описанных в исходных данных. Это и есть решение 1-й задачи. По сути, *это и есть профессиограммы или ключи теста на профессиональную пригодность, созданного в инновационной интеллектуальной технологии «Эйдос»*. В результате работы режима получены статистические модели и модели знаний и проведена их верификация. Модель знаний INF1 приведена на рисунке 7:

Код признака	Наименование описательной шкалы и градации	1. ПОЛ.ЖЕ...	2. ПОЛ.МУ...	3. ОТКУДА РОДОМ.К...	4. ОТКУДА РОДОМ.П...	5. ОТКУДА РОДОМ.Р...	6. ОТКУДА РОДОМ.С...	7. УСПЕВАЕ...	8. УСПЕВАЕ...	9. У...
1	ДЛИНА ВОЛОС-Длинные	0.219		-0.280	0.170	0.176			0.800	
2	ДЛИНА ВОЛОС-Короткие	-0.208	0.243	0.137	-0.462	0.006		0.161	-0.386	
3	ДЛИНА ВОЛОС-Средние	0.152	-0.523	-0.280	0.375	-0.174	0.669	-0.256		
4	ЦВЕТ ВОЛОС-Каштановые	0.050	-0.106	-0.213		0.243	0.736	0.161	0.313	
5	ЦВЕТ ВОЛОС-Очень светлые	-0.130	0.176	0.069	0.520					
6	ЦВЕТ ВОЛОС-Очень темные	0.219		0.419						
7	ЦВЕТ ВОЛОС-Красные	0.219		-0.043	0.612				0.133	
8	ЦВЕТ ВОЛОС-Русые	-0.178	0.220	0.021	-0.340	0.128		0.138	-0.264	
9	ЦВЕТ ГЛАЗ-Карие	0.142	-0.456	-0.213	-0.112	0.098	0.736	-0.188	-0.036	
10	ЦВЕТ ГЛАЗ-Голубые	-0.480	0.381	0.069	0.170	-0.174			0.246	
11	ЦВЕТ ГЛАЗ-Зеленые	0.142	-0.456	-0.213	0.238	0.098		0.161	0.313	
12	ЦВЕТ ГЛАЗ-Серые	-0.130	0.176	0.069	-0.180	0.031		0.094		
13	НАЛИЧИЕ УКРАШЕНИЙ-Браслет		0.526	0.419						
14	НАЛИЧИЕ УКРАШЕНИЙ-Нет	-0.480	0.381		0.520	0.176			0.246	
15	НАЛИЧИЕ УКРАШЕНИЙ-Перстень	0.015	-0.028	0.419				0.444		
16	НАЛИЧИЕ УКРАШЕНИЙ-Серьги	0.166	-0.636	-0.043	-0.292	0.176		-0.019	-0.216	
17	НАЛИЧИЕ УКРАШЕНИЙ-Цепочка	0.074	-0.174	-0.023	-0.034	-0.028	0.464	0.094	0.041	
18	ОДЕЖДА-Джинсы	-0.009	0.015	0.021	0.010	0.015		0.046	0.085	
19	ОДЕЖДА-Пиджак	-0.038	0.064	0.161	0.058	-0.286				

Рисунок 6 – Модель знаний INF1 (фрагмент)

6.1.3.2. РЕШЕНИЕ 2-й ЗАДАЧИ – СИСТЕМНОЙ ИДЕНТИФИКАЦИИ СОСТОЯНИЙ СЛОЖНЫХ ОБЪЕКТОВ

На рисунках 8 и 9 приведены примеры экранных форм с результатами системной идентификации. Рисунок 8 дает информацию для работодателя, проводящего исследование конкретного кандидата на работу, а 9 – проводящего массовое обследование кандидатов:

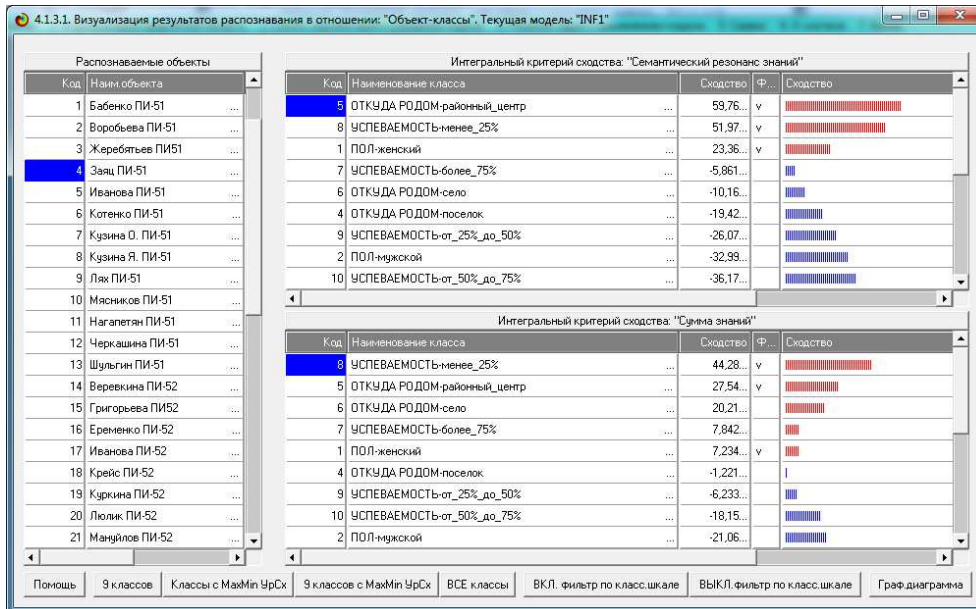


Рисунок 7 – Результаты системной идентификации конкретного респондента с классами

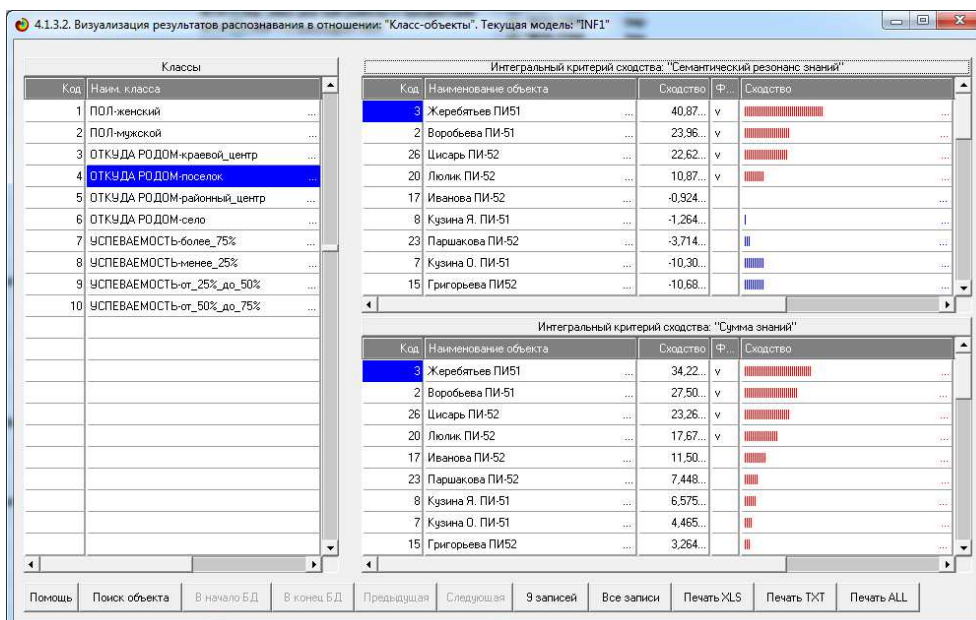


Рисунок 8 – Результаты системной идентификации конкретного класса с респондентами

При этом достоверность системной идентификации с применением различных моделей в модели в системе «Эйдос» оценивается с помощью предложенной автором метрики, по смыслу сходной с F-критерием.

Метризация шкал – решение проблемы сопоставимости при системной идентификации

Как показано выше, в АСК-анализе проводится *последовательное повышение степени формализации исходных данных до уровня, обеспечивающего их обработку на компьютере в программной системе*. После выполнения когнитивной структуризации и формализации предметной области осуществляется синтез статистических моделей и моделей знаний, в которых все шкалы, в которых описаны исходные данные, преобразуются к одному типу: числовому, и к одним единицам измерения: единицам измерения информации, т.е. проводится метризация шкал. В настоящее время в системе «Эйдос» применяется 7 способов метризации шкал [9].

В работе [3] сформулированы **требования** к форме представления данных, информации и знаний, позволяющие оценить *степень их пригодности* для решения задач системной идентификации, прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа).

Прежде всего, результаты решения вышеперечисленных задач должны быть **инвариантны** относительно:

- *единиц измерения* градаций факторов (признаков);
- *типов шкал*, используемых для формализации классов и факторов (номинальные, порядковые и числовые);
- *различных статистических характеристик исходной выборки*: частотных распределений объектов по классам (обобщенным категориям), частотных распределений градаций факторов, различий в количестве признаков в описаниях объектов исследуемой выборки, различий в суммарном количестве признаков по классам.

Кроме того, форма представления должна обеспечивать решение вышеперечисленных задач с минимальными дополнительными затратами ручного труда, а это значит, что *вся предварительная обработка должна быть максимально автоматизирована*.

Эти требования можно рассматривать и как *критерии* выбора наиболее подходящей для решения вышеперечисленных задач формы представления данных, информации и знаний.

Рассмотрим **влияние единиц измерения в исходной выборке на результаты решения задач** прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа).

Если в исходных данных какие-то значения выражены в больших единицах измерения, то их числовые значения будут малыми, и наоборот, если единицы измерения мелкие, то числовые значения – большие. Большие значения оказывают большее влияние на результаты математической обработки, чем малые, и *это приводит к возникновению зависимости результатов решения задач системной идентификации, прогнозирования и принятия решений, а также кластерного анализа, от выбранных размерностей исходных данных, что, на взгляд автора, совершенно неприемлемо и указывает на то, что такое решение нельзя признать корректным и даже вообще решением*. По этой же причине некорректно совместно обрабатывать сами исходные данные, представленные в **различных** единицах измерения (натуральных или ценовых), например, складывать расстояния, представленные в километрах и в метрах, а затем прибавлять к ним тонны и килограммы, а затем еще и безразмерные величины. Вроде это очевидно, но, как это ни удивительно, но как показывает опыт на практике это довольно часто делается, а потом еще на основе подобного «анализа» делаются и выводы. Очень странно, что обычно на это *не обращают никакого внимания* при использовании исходных данных, представленных в различных единицах измерения. Например, даже в таких популярных (причем, совершенно заслуженно) системах, как SPSS и Статистика, в подсисте-

теме кластерного анализа приводятся примеры кластерного анализа над исходными данными, представленными в различных единицах измерения.

В АСК-анализе факторы формально описываются шкалами, а значения факторов – градациями шкал. Существует три основных группы факторов: физические, социально-экономические и психологические (субъективные) и в каждой из этих групп есть много различных видов факторов, т.е. есть много различных физических факторов, много социально-экономических и много психологических, но в АСК-анализе все они рассматриваются *с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта, на который они действуют, в определенное состояние, и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации.* Именно по этой причине вполне корректно складывать (в аддитивных интегральных критериях) силу и направление влияния всех действующих на объект значений факторов, независимо от их природы, и определять результат *совместного* влияния на объект системы значений факторов. При этом в общем случае объект является *нелинейным* и факторы внутри него взаимодействуют друг с другом, т.е. для них не выполняется принцип суперпозиции.

На рисунке 10 приведен пример метризованной номинальной шкалы в модели INF1. По сути это и есть профессиограмма, сформированная в среде инновационной интеллектуальной технологии «Эйдос»:

4.2.1. Информационные портреты классов

Инф.портрет класса: 3 "ОТКУДА РОДОМ-краевой_центр" в модели: 4 "INF1"

Код	Наименование признака	Значимость
1	ПОЛ-женский	
2	ПОЛ-мужской	
3	ОТКУДА РОДОМ-краевой_центр	
4	ОТКУДА РОДОМ-поселок	
5	ОТКУДА РОДОМ-районный_центр	
6	ОТКУДА РОДОМ-село	
7	УСПЕВАЕМОСТЬ-более_75%	
8	УСПЕВАЕМОСТЬ-менее_25%	
9	УСПЕВАЕМОСТЬ-от_25%_до_50%	
10	УСПЕВАЕМОСТЬ-от_50%_до_75%	
6	ЦВЕТ ВОЛОС-Очень_темные	0.419
13	НАЛИЧИЕ УКРАШЕНИЙ-Браслет	0.419
15	НАЛИЧИЕ УКРАШЕНИЙ-Перстень	0.419
23	МАКИЯЖ-Отсутствует	0.215
24	СОБСТВЕННОСТЬ-автомобиль	0.215
19	ОДЕЖДА-Пиджак	0.161
2	ДЛИНА ВОЛОС-Короткие	0.137
5	ЦВЕТ ВОЛОС-Очень_светлые	0.069
10	ЦВЕТ ГЛАЗ-Голубые	0.069
12	ЦВЕТ ГЛАЗ-Серые	0.069
8	ЦВЕТ ВОЛОС-Русые	0.021
18	ОДЕЖДА-Джинсы	0.021
17	НАЛИЧИЕ УКРАШЕНИЙ-Цепочка	-0.023
7	ЦВЕТ ВОЛОС-Крашенные	-0.043
16	НАЛИЧИЕ УКРАШЕНИЙ-Серьги	-0.043
22	МАКИЯЖ-Заметный	-0.076
20	ОДЕЖДА-Юбка	-0.135
4	ЦВЕТ ВОЛОС-Каштановые	-0.213
9	ЦВЕТ ГЛАЗ-Карие	-0.213
11	ЦВЕТ ГЛАЗ-Зеленые	-0.213
1	ДЛИНА ВОЛОС-Длинные	-0.280
3	ДЛИНА ВОЛОС-Средние	-0.280
21	МАКИЯЖ-Незаметный	-0.340

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 MS Excel ВКЛ. фильтр по фактору ВЫКЛ. фильтр по фактору Вписать в окно Показать ВСЕ

Рисунок 9 – Пример метризованной номинальной шкалы (профессиограммы)

6.1.3.3. РЕШЕНИЕ 3-й ЗАДАЧИ – ПРИНЯТИЯ РЕШЕНИЙ ОБ УПРАВЛЯЮЩЕМ ВОЗДЕЙСТВИИ ТАК ИЗМЕНЯЮЩЕМ СОСТАВ ОБЪЕКТА УПРАВЛЕНИЯ, ЧТОБЫ ЕГО КАЧЕСТВО МАКСИМАЛЬНО ПОВЫШАЛОСЬ ПРИ МИНИМАЛЬНЫХ ЗАТРАТАХ НА ЭТО

Для решения 3-й задачи предлагается применить выбор компонент объекта управления по их функциональному назначению с учетом ресурсов, выделенных на реализацию различных функций, затрат, связанных с выбором тех или иных компонентов и степени соответствия различных компонент их функциональному назначению. Фактически предлагается формулировка и решение нового обобщенного варианта задачи о назначениях: «Мультипликативный рюкзак»²⁴, отличающегося от известного тем, что назначения производится не только с учетом ресурсов и затрат, но и с учетом степени соответствия компонент их функциональному назначению, которое предварительно определяется в самой задаче.

Математическая модель, обеспечивающая решение 1-й задачи и отражающая степень соответствия компонент их функциональному назначению, а также весь процесс принятия решений по назначениям, т.е. 2-я задача, реализованы в АСК-анализе и системе «Эйдос-Х++».

6.1.3.3.1. Интегральные критерии системы «Эйдос»

В результате проведения в метризации шкал, т.е. их преобразования независимо от исходного типа к одному типу: числовому, и независимо от исходных единиц измерения к одним единицам измерения: количеству информации, становится возможным корректно совместно обрабатывать результаты формализации описаний исходных данных в этих шкалах и использовать при этом все арифметические операции, в т.ч. сложение [9].

Это позволяет использовать аддитивные интегральные критерии и обоснованно ответить на следующий вопрос. Если нам известно, что объект обладает не одним, а несколькими признаками, то как посчитать их *общий* вклад в сходство с теми или иными классами?

Для этого в системе «Эйдос» используется 2 аддитивных²⁵ интегральных критерия: «Сумма знаний» и «Семантический резонанс знаний».

Интегральный критерий «Семантический резонанс знаний» представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

²⁴ Классический вариант этой задачи описан в работе [61].

²⁵ Т.е. являющихся суммами частных критериев, в отличие от мультипликативных интегральных критериев, которые являются произведениями частных критериев. На самом деле различие аддитивных и мультипликативных интегральных критериев не так велико, как может показаться на первый взгляд, т.к. они аддитивный интегральный критерий по сути является логарифмом мультипликативного. Иначе говоря это один и тот же критерий, но в разных шкалах: линейной и логарифмической.

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где: M – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j -го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } : n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n , если он присутствует у объекта с интенсивностью n , т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Интегральный критерий «Семантический резонанс знаний» представляет собой нормированное суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний имеет вид:

$$I_j = \frac{1}{\sigma_I \sigma_L M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

M – количество градаций описательных шкал (признаков);

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору объекта;

σ_I – среднеквадратичное отклонение частных критериев знаний вектора класса;

σ_L – среднеквадратичное отклонение по вектору распознаваемого объекта.

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j -го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n , если он присутствует у объекта с интенсивностью n , т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}.$$

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j -го класса и состояния распознаваемого объекта.

6.1.3.3.2. Алгоритм решения 3-й задачи

Алгоритм назначения объектов на наиболее подходящие классы с учетом ресурсов классов, затрат на объекты и степени соответствия объектов классам состоит в том, что назначаем текущий объект на тот класс, удельное сходство с которым максимально, при условии, что у данного класса есть для этого ресурсы, и делать это до тех пор, пока есть классы с ресурсами и назначены не все объекты (рисунок 11):

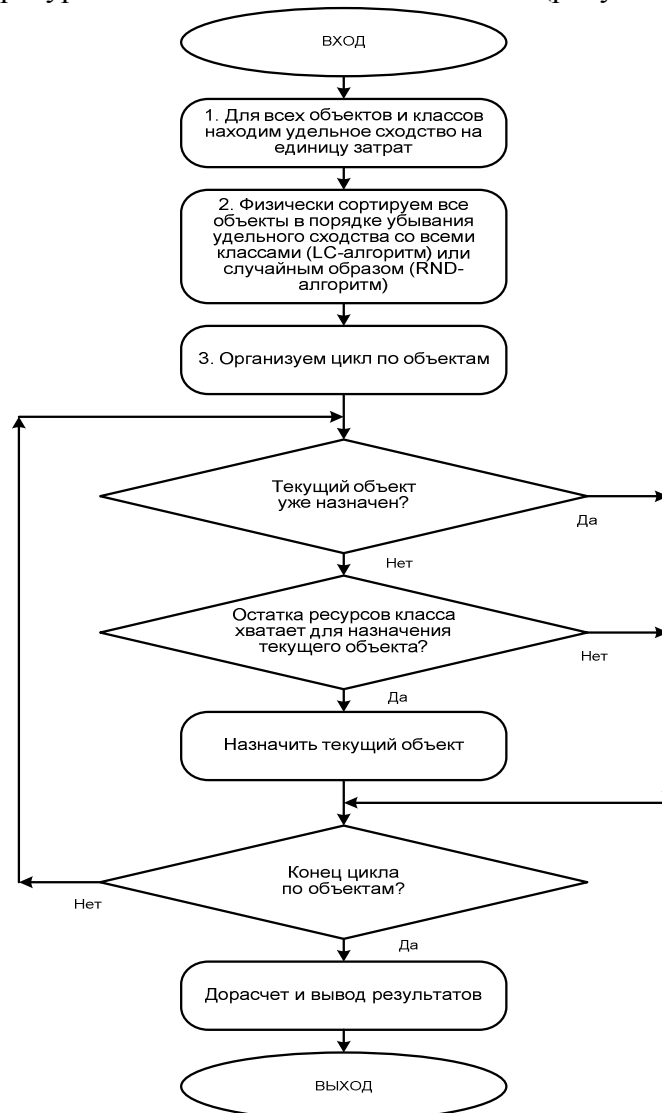


Рисунок 10 – Алгоритм назначения объектов на наиболее подходящие классы с учетом ресурсов классов, затрат на объекты и степени соответствия объектов классам

6.1.3.3.3. Численный пример решения 3-й задачи

Запустим режим 4.1.6 системы «Эйдос»: «Рациональное назначение объектов на классы (задача о ранце)» (Управление персоналом на основе АСК-анализа и функционально-стоимостного анализа (задача о назначениях)) (рисунок 12):

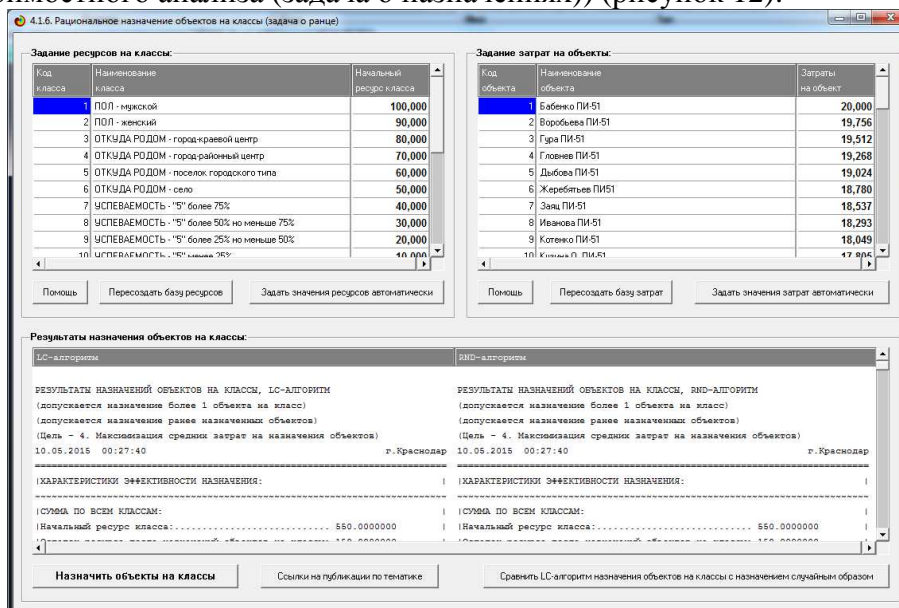


Рисунок 11 – Главная экранная форма режима: 4.1.6. Рациональное назначение объектов на классы (задача о ранце)

В верхнем левом окне пользователь может пересоздать базу ресурсов классов, а также корректировать ресурсы классов. Это возможно либо вручную, либо автоматически. В первом случае ресурсам присваивается значение по умолчанию. Во втором случае есть несколько вариантов присвоения значений ресурсов классов.

В верхнем правом окне пользователь может пересоздать базу затрат на назначение объектов, а также корректировать затраты объектов. Это возможно либо вручную, либо автоматически. В первом случае затратам присваивается значение по умолчанию. Во втором случае есть несколько вариантов присвоения значений затрат на назначение объектов.

При нажатии кнопки «Назначить объекты на классы» появляется окно, позволяющее задать параметры и цель назначения (рисунок 13):

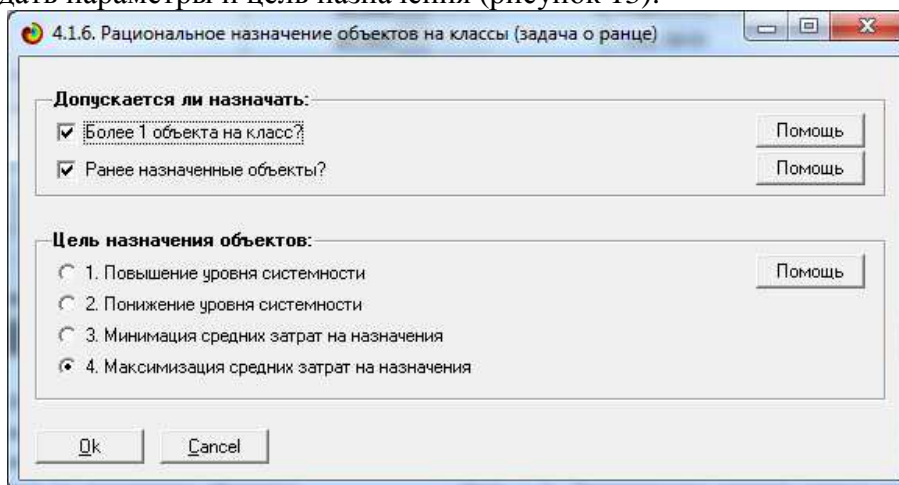


Рисунок 12 – Экранная форма задания параметров и цели назначений объектов на классы

Опцию: "Назначать не более 1 объекта на класс", имеет смысл использовать при разумной комплектации какого-либо сложного изделия, например автомобиля, когда каждый элемент комплектации (объект, деталь) назначается на каждую позицию (класс) 1 раз, например 1 инжектор, 1 левая фара, и т.д. С аналогичной ситуацией мы сталкиваемся при назначении кандидатов на такие должности, например, в спортивной команде, на каждой из которых может быть только один человек.

Опция: «Допускается ли назначать ранее назначенные объекты» позволяет подать на назначение не все объекты, а только не назначенные на классы при предыдущих назначениях. Например, если объектов задано значительно больше, чем классов и была задана опция: «Назначать не более 1 объекта на класс», то при каждом последующем назначении будут получаться автомобили со все более высокой себестоимостью и все более низкого качества, собранные из деталей, *отбракованных* при сборке предыдущих автомобилей. То же самое можно сказать об основном и дополнительном составе сборной: во 2-ю сборную входят игроки, не вошедшие в 1-ю, в 3-ю сборную - не вошедшие в 1-ю и 2-ю, и вообще в N-ю - не вошедшие в 1-ю, 2-ю, ..., (N-1)-ю.

Если данная опция не установлена, то все объекты считаются ранее не назначенными. Признак, что объект был ранее назначен, сбрасывается, при пересоздании базы затрат и при автоматическом задании затрат. При назначении объектов на классы этот признак устанавливается для назначенных объектов независимо от того, установлена ли опция: "Назначать только ранее не назначенные объекты". Но учитывается этот признак при назначении объектов только в случае, если эта опция установлена.

Опция "Задайте цель управления качеством системы:" позволяет выбрать одну из четырех целей работы LC-алгоритма, предложенного автором (рисунок 8):

1. Повышение уровня системности.
2. Понижение уровня системности.
3. Минимизация средних затрат на назначения объектов.
4. Максимизация средних затрат на назначения объектов.

Повышение уровня системности обеспечивает максимальное повышение качества системы с минимальными затратами на это. Понижение уровня системности обеспечивает максимальное понижение качества системы с максимальными затратами на это, что практически означает противодействие системе или даже уничтожение системы (антисистема). Минимизация средних затрат на назначения объектов приводит к назначению максимального количества сотрудников без учета степени их соответствия требованиям должностей с минимальной средней оплатой (всеобщая занятость населения и высокая скрытая безработица, но не очень эффективная экономика). Что-то вроде этого получается при сильной социальной политике. Максимизация средних затрат на назначения объектов приводит к назначению минимального количества сотрудников без учета степени их соответствия требованиям должностей с максимальной средней оплатой (низкая занятость населения и высокая реальная безработица). Аналогичный подход используется руководством при назначении "своих" людей на руководящие и наиболее хорошо оплачиваемые должности.

На практике приходится применять все четыре подхода в различных комбинациях в зависимости от обстоятельств. Например, чтобы коллектив выполнял свою функцию, т.е. вообще работал, сначала используется 1-я цель. Но так производятся назначения не на все должности, а в основном на исполнительские. После этого для назначения на престижные руководящие и хорошо оплачиваемые должности "своих" людей используется 4-я цель. 2-я цель используется военными и в конкурентной борьбе, а 3-я для того, чтобы не возникло социального бунта при повышении уровня реальной безработицы.

В нижнем окне на рисунке 9 приводятся *результаты назначения* объектов на наиболее подходящие классы с учетом ресурсов классов, затрат на объекты и степени соответствия объектов классам, в соответствии с LC-алгоритмом (слева), предложенным автором в работе [6], и классическим для задачи о Мультипликативном рюкзаке RND-алгоритмом (справа), в котором ресурсы классов и затраты на объекты учитываются, а степень соответствия объектов классам не учитывается, т.е. ценность объектов считается независимой от класса и фактически равна затратам на его назначение. Это и есть основной итог расчетов, т.е. результат решения задачи, выбранной в качестве примера. Все результаты расчетов записываются в виде большого количества разнообразных выходных форм, представляющих собой как текстовые файлы, так и Excel-таблицы. Фрагмент одной такой формы приведен ниже:

РЕЗУЛЬТАТЫ НАЗНАЧЕНИЙ ОБЪЕКТОВ НА КЛАССЫ, LC-АЛГОРИТМ

(допускается назначение более 1 объекта на класс)

(допускается назначение ранее назначенных объектов)

(Цель - 1. Повышение уровня системности)

10.05.2015 08:03:35

г.Краснодар

=====

| ХАРАКТЕРИСТИКИ ЭФФЕКТИВНОСТИ НАЗНАЧЕНИЯ: |

| СУММА ПО ВСЕМ КЛАССАМ:

Начальный ресурс класса:.....	550.0000000
Остаток ресурса после назначений объектов на классы:	159.9990000
Всего назначено на классы объектов:.....	26
Суммарное сходство:.....	1142.1670000
Фактические суммарные затраты:.....	390.0010000
Средневзвешенное удельное сходство:.....	27.4660000
Среднее на объект суммарное сходство:.....	399.9600000
Средние на объект фактические суммарные затраты:....	137.5810000

| СРЕДНЕЕ НА КЛАСС:

Начальный ресурс класса:.....	55.0000000
Остаток ресурса после назначений объектов на классы:	15.9999000
В среднем на класс назначено объектов:.....	3
Суммарное сходство:.....	114.2167000
Фактические суммарные затраты:.....	39.0001000
Средневзвешенное удельное сходство:.....	2.7466000
Среднее на объект суммарное сходство:.....	39.9960000
Средние на объект фактические суммарные затраты:....	13.7581000

=====

| КЛАСС НАЗНАЧЕНИЯ:

Код: 1, наименование: ПОЛ - мужской

Начальный ресурс класса:.....	100.0000000
Остаток ресурса после назначений объектов на класс:	19.5120000
Всего на данный класс назначено объектов:.....	5
Суммарное сходство:.....	233.2890000
Фактические суммарные затраты:.....	80.4880000
Средневзвешенное удельное сходство:.....	2.8980000
Среднее на объект суммарное сходство:.....	46.6580000
Средние на объект фактические суммарные затраты:....	16.0980000

Номер по пор.	Код объекта	Наименование объекта	Ур-нь сходст об.с классом	Затраты на назн. объекта	Удельное сход об. с классом
1	35	Мануйлов ПИ-52	64.3320000	11.7070000	5.4950000
2	20	Шульгин ПИ-51	64.9920000	15.3660000	4.2300000
3	12	Лях ПИ-51	57.3090000	17.3170000	3.3090000

4	17	Черкашина ПИ-51	21.0170000	16.0980000	1.3060000
5	1	Бабенко ПИ-51	25.6390000	20.0000000	1.2820000

=====

КЛАСС НАЗНАЧЕНИЯ:

Код: 2, наименование:	ПОЛ - женский
Начальный ресурс класса:.....	90.0000000
Остаток ресурса после назначений объектов на класс:	20.4880000
Всего на данный класс назначено объектов:.....	5
Суммарное сходство:.....	158.5350000
Фактические суммарные затраты:.....	69.5120000
Средневзвешенное удельное сходство:.....	2.2810000
Среднее на объект суммарное сходство:.....	31.7070000
Средние на объект фактические суммарные затраты:...	13.9020000

Номер по пор.	Код объекта	Наименование объекта	Ур-нь сходст об.с классом	Затраты на назн. объекта	Удельное сход об. с классом
1	42	Цисарь ПИ-52	44.6200000	10.0000000	4.4620000
2	24	Григорьева ПИ52	38.3090000	14.3900000	2.6620000
3	33	Люлик ПИ-52	31.7050000	12.1950000	2.6000000
4	23	Веревкина ПИ-52	21.0730000	14.6340000	1.4400000
5	8	Иванова ПИ-51	22.8280000	18.2930000	1.2480000

* * *

ДАННЫЕ ПО НЕНАЗНАЧЕННЫМ ОБЪЕКТАМ (LC-алгоритм):

Ном	Код	Наименование	Код	Наименование	Ур.сход	Затр.на	Уд.сх.
	объект	объекта	клас.	класса	об.с кл	объект	об.кл.
		В СРЕДНЕМ:			0.000	0.000	0.000

ДАННЫЕ ПО КЛАССАМ, НА КОТОРЫЕ НЕ БЫЛО НАЗНАЧЕНИЙ ОБЪЕКТОВ (LC-алгоритм):

Номер	Код	Наименование	Начальный
	класса	класса	ресурс класса
1	10	УСПЕВАЕМОСТЬ - "5" менее 25%.....	10.000
		СУММА :	10.000
		СРЕДНЕЕ:	10.000

=====

Универсальная когнитивная аналитическая система "Эйдос"

По нажатию клавиши выводится более подробная информация о сравнения результатов эффективности LC-алгоритма и RND-алгоритма (рисунок 11):

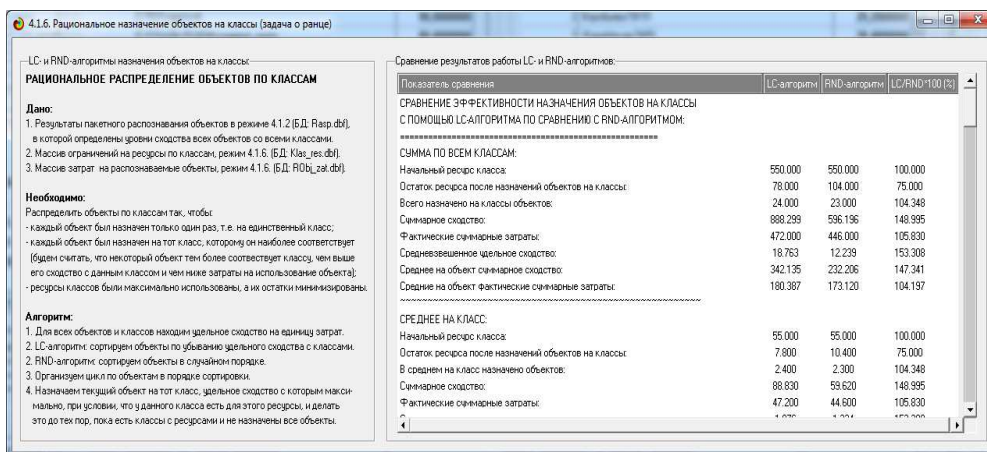


Рисунок 13 – Сравнение эффективности LC-алгоритма и RND-алгоритма

Из рисунка 11 видно, что при использовании LC-алгоритма при экономии ресурсов 25% среднее сходство объектов с классами возрастает примерно на 50%, если за базу сравнения брать RND-алгоритм. При других исходных данных и параметрах назначения эффективность LC-алгоритма может меняться, но всегда остается значительно более высокой. Чем RND-алгоритма. В этом и состоит актуальность постановки задачи и ее решения, предложенных в данной работе.

6.1.3.3.4. Различие в подходах психолога и руководителя к назначению и перемещению персонала

Психологи обычно рекомендуют назначать сотрудников на должности, которым они больше всего соответствуют по своим личностным и профессиональным качествам. Руководители же кроме этого еще учитывают и затраты своих ресурсов на эти назначения, т.е. то, сколько они готовы платить сотруднику за выполнение функциональных обязанностей на этой должности. Фактически руководитель применяет профессиограммы с учетом функционально-стоимостного анализа и метода «Директ-костинг». Поэтому предлагаемый в работе подход соответствует требованиям руководителя. Вместе с тем, психологов (специалистов по персоналу) не интересуют финансовые аспекты назначения персонала, то они могут задать на классы практически неограниченные ресурсы, а затраты на назначение для всех респондентов сделать малыми и одинаковыми (например, равными 1). Тогда система просто назначит сотрудников на должности, которым они больше всего соответствуют без учета затрат на это. Отметим также, что все выходные формы записываются в виде файлов. Обычно целью управления качеством является повышение уровня системности. Однако точно также, т.е. внедряя в определенные систему элементы, можно не повышать, а понижать ее уровень системности, т.е. по сути, разрушать, уничтожать данную систему (так и определяется понятие антисистемы²⁶).

Итак, качество системы рассматривается, как эмерджентное свойство систем, обусловленное их составом и структурой и отражающее их функциональность, надежность и стоимость. Поэтому при управлении качеством, целью управления является формирование у объекта управления заранее заданных системных свойств. Чем ярче у

²⁶ «Антисистемой называется система с отрицательным уровнем системности, т.е. это такое объединение некоторого множества элементов за счет их взаимодействия в целом, которое препятствует достижению целей» [12].

объекта управления выражены системные свойства, тем сильнее у него проявляется нелинейность: и в зависимости самих управляющих факторов друг от друга, и в зависимости результатов действия одних факторов, от действия других. Поэтому проблема управления качеством состоит в том, что в процессе управления сам объект управления изменяется качественно, т.е. изменяются его уровень системности, степень детерминированности и сама передаточная функция. Эта проблема распадается на несколько задач: 1-я состоит в многопараметрической типизации, 2-я в сопоставимой системной идентификации состояния объекта управления, а 3-я – в принятии решений об управляющем воздействии так изменяющем состав объекта управления, чтобы его качество максимально повышалось при минимальных затратах на это. Для решения 3-й задачи предлагается применить выбор компонент объекта управления по их функциональному назначению с учетом ресурсов, выделенных на реализацию различных функций, затрат, связанных с выбором тех или иных компонентов и степени соответствия различных компонент их функциональному назначению. Фактически предлагается формулировка и решение нового обобщения варианта задачи о назначениях: «Мультипликативный рюкзак», отличающееся от известного тем, что назначения производится не только с учетом ресурсов и затрат, но и с учетом степени соответствия компонент их функциональному назначению. Математическая модель, обеспечивающая решение 1-й и 2-й задач и отражающая степень соответствия компонент их функциональному назначению, а также весь процесс принятия решений по назначениям, т.е. 3-я задача, реализованы в АСК-анализе и системе «Эйдос-Х++». Приводится упрощенный численный пример предлагаемого подхода, связанный с назначением персонала.

6.1.4. Выводы

Материалы данного раздела могут быть использованы для синтеза и применения интеллектуальных измерительных систем для системной идентификации в эконометрике, биометрии, экологии, педагогике, психологии, медицине, ветеринарии и в других предметных областях, а также при преподавании дисциплин: интеллектуальные системы; инженерия знаний и интеллектуальные системы; интеллектуальные технологии и представление знаний; представление знаний в интеллектуальных системах; основы интеллектуальных систем; введение в нейроматематику и методы нейронных сетей; основы искусственного интеллекта; интеллектуальные технологии в науке и образовании; управление знаниями; автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»; которые автор ведет в настоящее время²⁷, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее – в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки).

Таким образом, АСК-анализ и система «Эйдос» представляют собой современную инновационную (готовую к внедрению) технологию интеллектуального управления качеством систем путем решения обобщенной задачи о назначениях. Этим и другим применениям должно способствовать и то, что система «Эйдос» находится в полном открытом бесплатном доступе (причем с подробно комментированными открытыми исходными текстами) на сайте автора по адресу: <http://lc.kubagro.ru/aidos/Aidos-X.htm>.

²⁷ http://lc.kubagro.ru/My_training_schedule.doc

6.2. Хиршамания при оценке результатов научной деятельности, ее негативные последствия и попытка их преодоления с применением многокритериального подхода и теории информации

Недавно был начат процесс монетизации оценки результатов научной деятельности, и возникла потребность в методиках количественной и сопоставимой оценки эффективности и качества работы ученого. Появились многочисленные методики материального поощрения за эти результаты. Общим для всех этих методик является завышенная роль индекса Хирша. Сам по себе этот индекс вполне обоснован. Однако в связи с практикой применения индекса Хирша в наших условиях в сознании научного сообщества возникла своеобразная мания, которую автор предлагает называть «Хиршамания». Эта мания характеризуется повышенным нездоровым интересом к самому значению индекса Хирша, особенно к искусственному неадекватному преувеличению этого значения, а также рядом негативных последствий этого интереса. В данной работе делается попытка кратко описать некоторые негативные последствия этой новой психической инфекции, поразившей общественное сознание научного сообщества. А также наметить пути преодоления хотя бы некоторых причин их возникновения. В этом и состоит проблема, решаемая в данной работе. Для решения сформулированной проблемы предлагается применить многокритериальный подход, основанный на теории информации, а именно тот его вариант, который реализован в автоматизированном системно-когнитивном анализе (АСК-анализ) и его программном инструментарии – интеллектуальной системе «Эйдос»

«Индекс Хирша – это наукометрический показатель, который отражает степень понимания того, что такое индекс Хирша»

*Народная мудрость периода
Хиршамании (начало XXI века)*

6.2.1. Проблема, или о том какой урон нанес джин Хирша, выпущенный из бутылки

Недавно научное сообщество лишилось надбавок за ученые степени и звания. Был начат процесс монетизации оценки результатов научной деятельности. Возникла потребность в соответствующих методиках количественной и сопоставимой оценки эффективности и качества работы ученого. Появились многочисленные методики материального поощрения за эти результаты. Эти методики отличаются в разных вузах. Но общим для всех этих методик является большая роль, которая отводится в них так называемому индексу Хирша. Сам по себе этот индекс вполне обоснован²⁸.

Однако в связи с практикой применения индекса Хирша в наших условиях в сознании научного сообщества возникла своеобразная мания, которую автор предлагает называть «Хиршамания». Эта мания характеризуется повышенным нездоровым интересом к самому значению индекса Хирша, особенно к искусственному неадекватному преувеличению этого значения, а также рядом негативных последствий этого интереса. В данной работе мы попытаемся кратко описать некоторые негативные последствия этой новой психической инфекции, поразившей общественное сознание научного сообщества. А также наметить пути преодоления хотя бы некоторых причин их возникновения. В этом и состоит проблема, решаемая в данной работе.

²⁸ <http://yandex.ru/yandsearch?text=индекс%20Хирша=35>

Чтобы наметить удачный план лечения, прежде всего надо поставить правильный диагноз. Мы квалифицируем Хиршаманию как психический вирус, о которых блестяще писал Ричард Броди [1]. Этот психический вирус может рассматриваться как инструмент манипуляции общественным сознанием научного сообщества, что очень хорошо описал в своем бестселлере С.Г. Кара-Мурза [2]. Подобного рода манипуляции, которые особенно облегчились в связи с появлением глобальной информационной среды распространения и адресной доставки агента действия, могут использоваться также для нанесения урона противнику и, по сути, являются информационным оружием [3]. Так что Хиршамания в принципе может быть не таким уж и безобидным явлением.

Так что же собственно произошло? Ученым стали платить надбавки (материальные поощрения) за те или иные значения индекса Хирша. Ученые народ неглупый и быстро сообразил, что имеет прямой смысл эти значения увеличивать. А для этого надо писать научные статьи, монографии, научно-методические работы и т.д., и т.п., размещать их в Российском индексе научного цитирования (РИНЦ)²⁹, и ссылаться на них. Это ясно из самой природы индекса Хирша. Причем не просто писать и ссылаться, а писать как можно больше и ссылаться тоже как можно больше.

Тривиальным является утверждение о том, что статья должна отражать основные результаты какого-то научного исследования, решение той или иной научной или прагматической задачи. Но откуда взять столько научных результатов? Ведь научные исследования требуют инвестиций и имеют длительный цикл проведения.

Так возникает *1-я проблема*, состоящая в том, что писать то, в общем-то, и не о чем, по крайней мере, в желаемом количестве, а писать очень надо. В век интернета решение этой проблемы элементарно. Проводим поиск по ключевым словам, находим источники, в которых об этом уже кем-то написано, и вставляем тексты из этих источников в свои статьи. Если мы корректно ссылаемся на эти заимствования, то они называются «цитированиями», а если нет, то «плагиат». Если ссылаться на все эти заимствования, то может оказаться, что автор не внес никого личного вклада в работу, поэтому часто на них не ссылаются. Таким образом, материально простимулированное применением индекса Хирша для оценки результативности научной деятельности неоправданное стремление много писать, но не для того, чтобы отразить результаты реальных исследований, которых или вообще нет, или недостаточно для удовлетворения амбиций, приводит к распространению плагиата. Что же такое плагиат с правовой точки зрения? Наверное, это просто воровство, в частности воровство авторских текстов и идей, нарушение авторских прав на результаты научной деятельности.

Если есть спрос, то есть и предложение, и вот появляется система «Антиплагиат»³⁰ и много других подобных систем³¹. Подобные системы представляют свои услуги on-line, есть и бесплатные с ограниченными возможностями, и профессиональные, которые, естественно, платные. Практически все вузы уже купили профессиональные системы проверки на антиплагиат, и эта проверка стала нормой. Таким образом, первым отрицательным последствием Хиршамании является небывалое распространение плагиата в научной среде и как реакция на это – борьба с плагиатом (а не с его причинами, т.е. как обычно), причем за деньги самих научных организаций, т.е. косвенно – самих ученых и с большими затратами труда и времени самих ученых.

Началась борьба с плагиатом, началась и борьба с этой борьбой. Как грибы после дождя в информационном пространстве вдруг появились многочисленные ухищрения и «научные рекомендации» для того, чтобы обойти эти системы, т.е. добиться вы-

²⁹ <http://elibrary.ru/>

³⁰ <http://www.antiplagiat.ru/>

³¹ <http://yandex.ru/yandsearch?lr=35&text=антиплагиат>

сокого уровня оригинальности некорректно заимствованного текста³². Прежде всего это различные синонимайзеры³³. Однако системы антиплагиата работают просто с текстами, поэтому можно их обойти, если заимствовать не текст, а идеи, т.е. несколько перефразировать текст, чтобы сам текст стал другим, а его смысл сохранился. Эта процедура называется «Рерайтинг (rewriting)»³⁴, т.е. переписывание и изложение чужих мыслей своими словами. Еще для подобных целей могут быть использованы программы машинного перевода, т.к. они тоже подбирают синонимы и перефразируют³⁵. Конечно, разработчики систем антиплагиата также принимают меры для обнаружения признаков борьбы с ними, т.е. признаков искусственного завышения оригинальности текста (это уже борьба систем антиплагиата с борьбой против них), и т.д., и т.д. почти до бесконечности.

Автор тоже столкнулся с этим явлением (как поставщик текстов и идей для плагиаторов, как высокопоставленных, так и не очень). Лучше всего об этом написано в статье «Групповой плагиат: от студента до министра»³⁶. Чтобы найти многочисленные «труды» плагиаторов, включая диссертации, достаточно в Internet в любой поисковой системе сделать запрос, например: «Коэффициенты эмерджентности Хартли, Харкевича, Шеннона», которые автор системной теории информации (СТИ) проф. Е.В.Луценко назвал так в честь этих выдающихся ученых в области теории информации. При этом автор следовал сложившейся научной традиции называть единицы измерения и математические выражения в честь известных ученых. Причем часто *плагиаторы даже не понимают, что сами основоположники и классики теории информации не предлагали этих коэффициентов, а предложены они были в работах автора*. Наверное, поэтому они и не считают нужным делать ссылки и пишут, например:

1. «По Харкевичу коэффициент эмерджентности определяет степень детерминированности системы...» (подчеркнуто нами, авт., в цитате сохранены орфографические ошибки плагиатора, авт.).

2. «Отсюда строится системная численная мера количества информации в ИС на основе оценки **эмерджентности системы (по Хартли и Харкевичу)**» (выделено плагиатором, а на самом деле «по Луценко», – авт.).

Эти фразы легко найти в Internet. Так что плагиаторская активность не только продолжается, но и набирает обороты.

Однако индекс Хирша отражает не только число статей, но и число их цитирований. Поэтому самих статей для повышения значения индекса Хирша недостаточно, т.е. еще надо, чтобы на них были ссылки. А откуда их взять, если на твои статьи сам, т.е. по собственной инициативе, никто или почти никто не ссылается?

Так возникает 2-я проблема, т.е. проблема увеличения количества ссылок на свои статьи. Но и эту проблему можно решить. Во-первых, можно самому ссылаться на собственные статьи, т.е. заниматься самоцитированием. С правовой точки зрения это, конечно, не плагиат, т.к. нет потерпевшей стороны, т.е. в принципе автор имеет все ав-

³² <http://yandex.ru/yandsearch?text=повышение%20уровня%20оригинальности%20текста&lr=35>

³³ <http://yandex.ru/yandsearch?text=синонимайзеры&lr=35>

³⁴ <http://yandex.ru/yandsearch?text=Рерайтинг&lr=35>

³⁵ Вставка в русские слова английских букв и склеивание слов символами, которые отображаются как пробелы, чтобы система антиплагиат не смогла их идентифицировать – это слишком дешевые и легко распознаваемые искусственные приемы, чтобы к ним можно было серьезно относиться.

³⁶ Вяткин В.Б. Групповой плагиат: от студента до министра. - Троицкий вариант — Наука - <http://trv-science.ru> - [Электронный ресурс]. Адрес доступа: <http://trv-science.ru/2011/11/08/grupповојј-plagiат-ot-studenta-do-ministra/> или: <http://trv-science.ru/2011/11/08/grupповојј-plagiат-ot-studenta-do-ministra/print/>

торские права на свой текст и свои идеи и может распоряжаться ими по своему усмотрению. Но с этической точки зрения чтобы сослаться на себя, надо иметь на это моральное право и необходимо, чтобы эта ссылка была оправдана и обоснованна, а не являлась искусственной, т.е. ссылки ради ссылки, часто даже без сноски на нее из текста. По мнению автора, такое обоснование может состоять в том, что ряд статей образует цикл или систему, т.к. они посвящены описанию различных этапов решения одной проблемы или развитию определенного научного направления, т.е. по сути, являются продолжением друг друга и взаимосвязаны по своему содержанию³⁷. Однако часто у авторов нет возможности для таких обоснованных самоцитирований, то тогда они занимают неоправданными (некорректными) самоцитированиями. Иногда авторов с высоким уровнем самоцитирования обвиняют в том, что они засоряют информационное пространство дублирующей информацией и как бы продают многократно один и тот же информационный продукт. Автор не согласен с этой точкой зрения потому, что если бы она была правильной, то во всем интернет должно бы быть одно место для хранения каждого информационного объекта, а все остальные просто должны были бы на него ссылаться. Но что мы видим на практике? Мы видим огромное количество размещений одного и того же информационного объекта на различных сайтах. Кстати, этим занимаются и различные интеграторы – библиографические базы данных, например НЭИКОН, КиберЛенинка, Agris, Ulrich's Periodicals Directory, DOAJ, OALIB (Open Access Library), Scopus, Web of Science, и даже сам РИНЦ. Смысл таких размещений в новых возможностях в появлении для читателей новых дополнительных возможностей прочтения публикаций, а также возможность их статистической обработки и углубленного анализа в базах интегратора.

Но все же как-то неудобно ссылаться только на самого себя, как будто у тебя и не было предшественников или соавторов. Понятно, что цитирование соавторами это «почти самоцитирование» и также уязвимо для критики. Поэтому авторы, не являющиеся соавторами, часто договариваются о взаимных цитированиях, т.е. я тебя цитирую тебя, а за это – ты меня, что неблагоприятно и с правовой точки зрения является сговором для извлечения дополнительной необоснованной созданием продукта или услуг прибыли, а этической точки зрения достойно морального порицания. В век глобальных коммуникаций и эти вопросы легко решаемы.

И, как всегда, раз наблюдаются искусственные и необоснованные цитирования, то появляются и средства борьбы с этим неблагоприятным явлением (естественно, как всегда только с самим явлением, а не с его причинами). В частности появляется идея использовать для количественного измерения самоцитирований и цитирований соавторами индекс, взятый из экономики, который в ней используется для количественной оценки степени монополизации отрасли, – это индекс Херфиндаля³⁸ и различные его модификации.

Как же научное сообщество среагировало на установленные Минобрнауки РФ «правила игры». Да очень просто: все, даже те, кто уже давно ничего не писал, с готовностью принялись писать научные работы и цитировать их, можно сказать с энтузиазмом принялись повышать свои индексы цитирования и индексы Хирша. Правда это не сопровождалось сколь-нибудь заметным или значительным повышением активности самих научных исследований и инновационных разработок. А это означает, что народ

³⁷ У автора данной статьи много самоцитирований, но он оправдывает их корректность именно таким образом. Практически все работы автора посвящены развитию и многочисленным применениям предложенного им нового научного и инновационного направления искусственного интеллекта: Автоматизированного системно-когнитивного анализа (АСК-анализа) и его программного инструментария – интеллектуальной системы «Эйдос»: <http://lc.kubagro.ru/>

³⁸ <http://yandex.ru/yandsearch?text=индекс%20Херфиндаля&lr=35>

вполне понял, что от него требуется: не сама работа, а лишь показатели отчетности о работе. По сути, речь идет о *фальсификации деятельности*: вместо самой деятельности и ее результатов учитываются и идут в зачет отчеты с растущими показателями о якобы имевшей место деятельности и результатах, и эти отчеты и показатели фактически и принимаются за результаты, а было ли все это на самом деле о чем на самом деле они отражают, фактически никого особенно не интересует³⁹. Налицо явление, которое, в работе [4], названо «виртуализация общества». Фактическая деятельность в реальной области заменяется, *замещается* информацией о ней, а потом выясняется, что информация начинает жить собственной жизнью и *замещает* реальную область, при этом она в принципе может и не отражать процессы в реальной области. В результате масштабной, может быть даже глобальной фальсификации и виртуализации сама деятельность заменяется отчетностью о ней, содержащей различные показатели. И народ быстро сообразил, что нет необходимости в самой деятельности и нет никакого смысла достигать самого результата деятельности в реальной области: вполне *достаточно обеспечить достижение не самого результата деятельности, а нужных показателей отчетности*. Повышения этих показателей все и добиваются.

Но наука в этом плане не оригинальна. В экономике подобная подмена давно стала нормой: практически все осознают как цель своей экономической деятельности не создание благ и услуг для потребления их другими людьми, во что крайне наивно верил Адам Смит и что было беспощадно развенчано Карлом Марксом, а всего лишь личное обогащение (получение прибыли). В экономике давно считается нормой, вполне допустимой и не осуждаемой с морально-этической точки зрения, обогащение без создания реальных благ, например за счет осуществления спекулятивных операций на фондовом и валютном рынке. В результате подобных операций создается ничем реально не обеспеченный чисто инфляционный капитал. И в этом никто не видит ничего аморального или неэтичного. А между тем это прямой и ничем неприкрытый открытый грабёж огромных масс людей во всем мире, покупательная способность которых падает за счет инфляции. Деньги вообще все замещают и все опошляют и не только в экономике, а вообще везде. Так, например, любовь они превращают в проституцию. А науку, которая является общественным институтом и общественной и индивидуальной деятельностью по познанию человека, общества и природы с целью познания истины (фундаментальная наука) и повышения эффективности деятельности человека (прикладная наука), – в искусственное увеличение индекса Хирша путем увеличения потока бессодержательных, переписанных друг у друга, но при этом широко цитируемых публикаций. *В традициях Хиршамании цель достижения высоких результатов научной деятельности подменяется целью достижения высоких значений наукометрических показателей, отражающих эти результаты*. При этом в информационном обществе отражение может быть создано и существовать и без достижения каких-либо реальных результатов и достижений. Таким образом, фальсификация и виртуализация науки, обусловленная Хиршаманией, по сути, приводят к ее профанации. Это настолько очевидно, что можно было бы об этом и не говорить, если бы практически все научное сообщество снизу доверху дружно бы не занималось именно этим, безропотно приняв правила игры, предложенные Минобрнауки, причем не просто занималось, но и

³⁹ Примерно тоже самое происходит и с учебной деятельностью: нужны хорошие учебно-методические комплексы (УМК), различные планы и документация, а был ли сам учебный процесс и каковы его истинные результаты никого особенно не интересует. Причем документации требуется все больше и больше, все время меняются стандарты ее оформления и содержания, так что преподавать становится, в общем-то, и некогда. Примерно также вроде и врачам уже и лечить становится просто некогда из-за оформления все больших объемов документации.

делало при этом вид, что процесс идет нормально, т.е. делая хорошую мину при плохой игре⁴⁰.

Однако, в отличие от экономики, в науке в соответствии этикой научных публикаций⁴¹ подобные действия считаются аморальными, и журналам, допускающим подобные статьи к публикации практически закрыт путь в такие престижные международные библиографические базы данных, как Scopus и Web of Science. Из-за действия подобных этических норм воровать в науке стало так сложно, что иногда начинает казаться, что ты работаешь. Это звучит как юмор, но им не является, т.е. было бы смешно, если бы не было грустно. Так, например, рерайтинг – это действительно сложная и трудоемкая работа, требующая не только профессиональной компетенции в той предметной области, в которой осуществляется плагиат идей, но и определенной филологической подготовки: умения быстро и много писать на правильном русском языке. В работе рерайтера есть почти все, что есть в научной работе, кроме одного: у рерайтера нет своих новых идей и ему *нигде* их брать, кроме как заимствовать их у того, у кого они есть. Но у рерайтера не просто нет своих новых идей, но он и не имеет ни малейшего представления о том, откуда они вообще берутся. Здесь я не могу отказать себе в удовольствии еще раз сослаться на статью В.Б.Вяткина [5], хотя плагиаторы-персонажи этой статьи не удосужились даже переписать мой текст своими словами, т.е. не дотянули до рерайтеров, а просто привели весь текст целиком, включая «авторскую пунктуацию», т.е. изложив все от первого лица, «как было», включая даже мои орфографические и грамматические ошибки.

Таким образом, можно обоснованно констатировать факт, состоящий в том, что решение Минобрнауки РФ о монетизации оценки результатов научной деятельности, в частности придание неоправданно высокой роли в этом процессе индексам публикационной активности (само по себе это вчерашний и даже позавчерашний день), а также индексам цитирования и Хирша, явилось причиной, породившей целый каскад или снежный ком различных негативных последствий, многие из которых имеют криминальный характер, а некоторые «всего лишь» аморальны (рисунок 1).

Подобное в новейшей Российской истории было уже не раз, и, по-видимому, уже есть основания говорить об определенной наметившейся *тенденции или закономерности*. Достаточно упомянуть про позорную эпопею с ЕГЭ, которая начиналась за здравие, а закончилась за упокой и подобную же историю с оценкой эффективности вузов [4]. Очень бы не хотелось, чтобы тоже самое, что случилось с ЕГЭ и с методикой оценки эффективности вузов случилось бы и с РИНЦ. Однако, к большому сожалению и объективно говоря пока все идет к тому, что так и получится, т.е. идея количественной сопоставимой оценки результатов (качества и эффективности) научной деятельности в наших условиях не реализуема. Может быть, это не дело только РИНЦ, а дело всего российского научного сообщества, включая Минобрнауки?

Итак, мы сталкиваемся с ситуацией, когда хорошая идея плохо реализуется, т.к. мы опять впадаем в крайности, относимся к новой (для нас) идее, как панацее от всех проблем. Но такого, к сожалению, не бывает. А, как известно, нет лучшего способа дискредитировать хорошую идею, чем довести ее до абсурда, до крайности, т.к. тогда она становится своей противоположностью. Получается уже в который раз, что за что боролись, на то и напоролись. В этой ситуации перефразируя, похоже, что бессмертные слова В.С.Черномырдина, Минобрнауки может только сказать, что «мы хотели как

⁴⁰ А что остается делать? Ведь рядовые ученые и преподаватели не имеют никакой возможности влиять на принятие решений не только на федеральном уровне управления образованием, но и даже на уровне кафедры, факультета и вуза. Приходится принимать эти условия как объективную реальность, не зависящую от сознания.

⁴¹ <http://ej.kubagro.ru/efh.asp> <http://publicationethics.org/>

лучше», а мы уже имеем все фактические основания констатировать, что опять «получилось как всегда».

Негативные последствия Хиршамании разных уровней иерархии

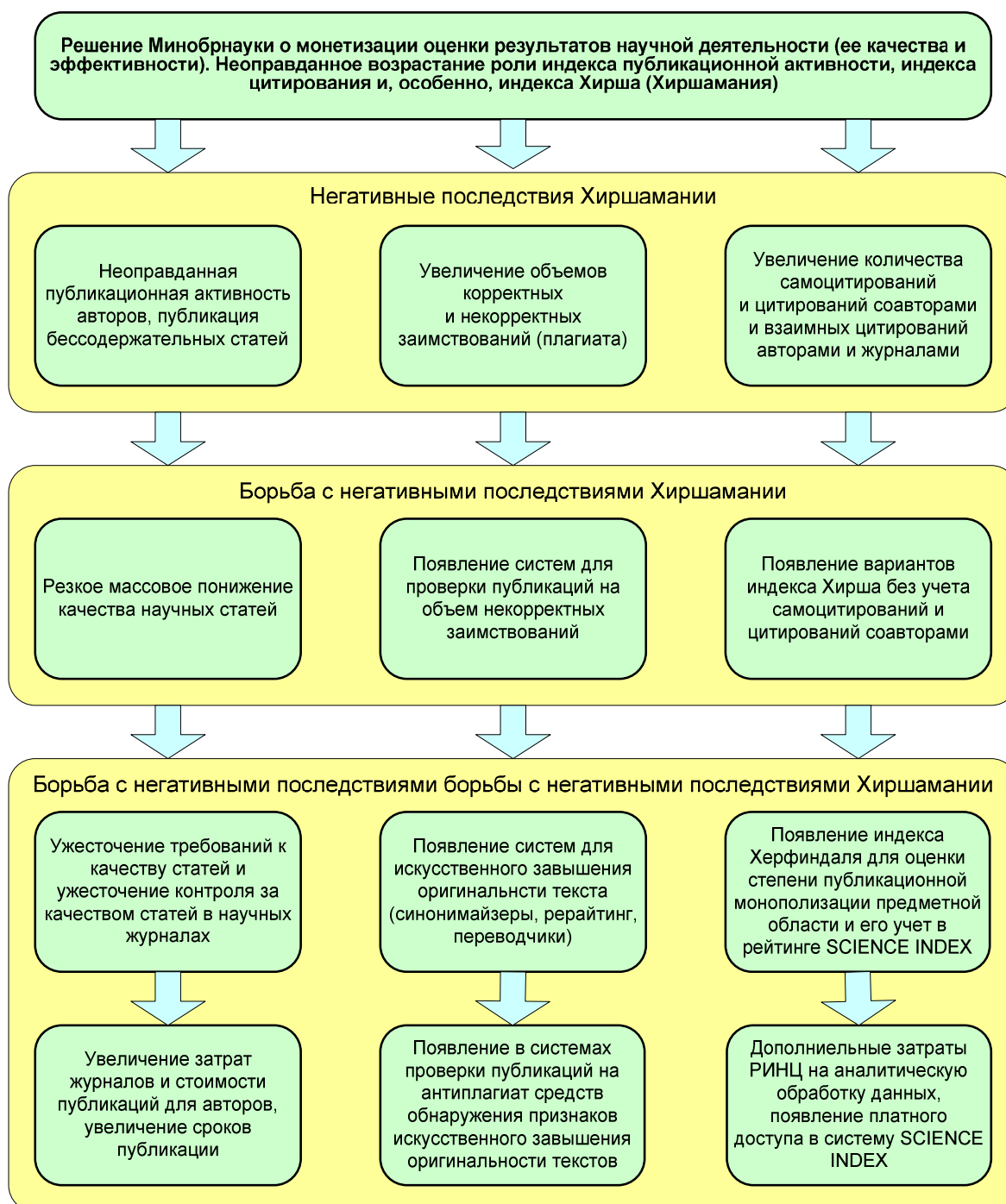


Рисунок 1 – Негативные последствия решения о монетизации оценки результатов научной деятельности (Хиршамании)

Выше уже упоминалось, что в этой новой истории с индексом Хирша, ключевая роль принадлежит Российскому индексу научного цитирования (РИНЦ)⁴². Что такое РИНЦ сегодня? На сегодня это безусловно самая большая в России электронная научная библиотека общего доступа. На момент написания данной статьи в базах РИНЦ со-

⁴² <http://elibrary.ru/>

держалось: журналов: 32991, выпусков журналов: 1193672, полнотекстовых статей: 20325344, издательств: 14976⁴³. Данные РИНЦ используются сегодня для оценки рейтингов научных журналов и результатов научной деятельности, как отдельных ученых, так и вузов и научных институтов. Безусловно, у РИНЦ много достоинств и в настоящее время ему в России нет альтернативы. Но и у РИНЦ есть свои недостатки, которые вызывают совершенно обоснованный и все более громкий ропот в среде научного сообщества. Перечислим наиболее важные них в контексте проблематики данной статьи.

РИНЦ чрезвычайно инерционная система. Обновление его баз данных происходит раз в неделю, а некоторые базы данных, например системы SCIENCE INDEX, обновляются раз в 2 месяца. Это означает, что в РИНЦ не выполняется одно из важнейших требований к базам данных: их актуальность. Проще говоря, если Вы обращаетесь в РИНЦ для получения тех или иных показателей по отдельным ученым или по вузам научным организациям, то нужно иметь в виду, что эта информация может быть сильно устаревшей.

В РИНЦ отсутствует целостность баз данных, т.е. данные в одних графах не подтверждаются данными в других графах. Как профессор по кафедре компьютерных технологий и систем с 2005 года и программист с почти 40-летним стажем автор может предположить, что эта особенность РИНЦ обусловлена тем, что разработчики даталогической и инфологической моделей его баз данных не везде нормализовали его базы данных. Это означает, что информация РИНЦ не очень достоверна или, проще говоря, просто недостоверна.

Программное обеспечение РИНЦ не может идентифицировать авторов и публикации по их некорректным, неполным и нестандартным описаниям в списках литературы. Между тем эта задача решена в общем виде [6]. В результате число работ авторов, число цитирований этих работ и индекс Хирша авторов в системе РИНЦ оказываются систематически заниженными⁴⁴. В качестве выхода из этой проблемной ситуации РИНЦ предлагает нам регистрироваться в системе SCIENCE INDEX и самим привязывать к себе свои работы и ссылки на свои работы из списка, предложенного системой РИНЦ. Иначе говоря, авторам предлагается самим вручную делать работу, которая должна выполняться полностью автоматически или с участием модератора, т.е. автоматизировано, самим РИНЦ.

Но и это еще не все. Дело в том, что список работ и ссылок на них, которые могут принадлежать данному автору, предлагаемый системой РИНЦ неполон, т.е. иначе говоря, некоторые работы, имея права пользователя системы SCIENCE INDEX, привязать вообще невозможно. Но это возможно имея права доступа администратора этой системы. И тут начинается самое интересное. Права администратора системы SCIENCE INDEX *продаются* РИНЦ вузам и научным организациям, причем продаются не дешево. Понятно, что практически все вузы и научные организации фактически были вынуждены купить эту систему, как, кстати, немного ранее и доступ к профессиональной версии системы Антиплагиат (которая, кстати, сейчас интегрируется с системой РИНЦ).

Интересно получается. *Ученые должны провести научные исследования и разработки, написать об этом монографии и статьи, получить авторские свидетельства, разместить их в РИНЦ, привязать их к себе, как авторам, привязать к себе ссылки на них, причем должны сделать это все это сами и еще заплатив журналам, Роспатенту и РИНЦ, за право самим это все сделать. Получается к примеру 5% бюджета*

⁴³ http://elibrary.ru/stat_resources.asp

⁴⁴ Автору статьи администраторы РИНЦ сообщили, что у него около 2000 непривязанных ссылок.

времени ученого уходит на проведение самих научных исследований и разработок, еще 5% – на их оформление в виде монографий, статей и патентов, и еще 90% на их размещение в РИНЦ, а также привязку публикаций и ссылок на них к себе, как автору. Разве так мы представляли себе, на что должен тратить свое творческое время ученый? Причем, как правило, после ненормированного рабочего дня, потраченного на голосовую нагрузку – преподавание? Ну то, что журналам за публикацию статей нужно платить, это еще понять можно. Но почему надо платить еще и РИНЦ за работу, которую мы выполняем сами за его программную систему, это уже не очень понятно.

Понятно, что покупая (и не дешево) права доступа администратора системы SCIENCE INDEX, вузы и научные организации справедливо надеются, что имея эти права, они смогут в комфортной и дружественной информационной среде полностью решить проблемы с привязкой статей и ссылок на них, описанные выше. Однако оказывается, что и это не так.

Прежде всего, надо сказать, что среда далека от комфортной, переходы из режима в режим плохо продуманы и сделаны неудобно. Постоянно надо переходить из одного окна в другое, удобнее бывает открыть два или более окон на большом мониторе, но они мало у кого есть. Реакция системы на клик является чрезвычайно замедленной: иногда несколько минут. При большом объеме работ это резко увеличивает общее время работы. Когда администратор системы SCIENCE INDEX видит список ссылок на работы автора, то те, над которыми ему надо работать, не отмечены двумя красными треугольниками, и таких довольно много, но приходится их искать вручную среди отмеченных. При этом ссылки на непривязанные работы и на привязанные работы, находящиеся на проверке модератора, на экране выглядят одинаково. Это крайне неудобно. О том, что работа находится на проверке, можно догадаться потому, что нет ссылки на возможность корректировки ее описания. Но чтобы узнать есть эта ссылка или нет, надо кликнуть по ссылке на самую ссылающуюся работу, а потом, если ее нет, вернуться назад в список ссылок, возможность чего разработчиками не предусмотрена и приходится возвращаться через авторский указатель. Все это занимает время и раздражает своей непродуманностью, т.к. решение ведь простое. Надо просто отмечать все ссылки в специальной отведенной для этого колонке иконками тер типов, отличающимися по цвету, как сделано с инками, отражающим возможность доступа простого посетителя РИНЦ к полным текстам публикаций. Есть всего три состояния ссылки на публикацию автора: 1) привязанная (неважно кем), 2) привязанная администратором SCIENCE INDEX от организации и находящаяся на проверке у модератора РИНЦ, 3) непривязанная. Я предложил бы непривязанные ссылки отметить красным цветом, находящиеся на проверке – желтым, а привязанные – зеленым. Это интуитивно естественно. Желательно было бы иметь возможность сортировки и фильтрации ссылок по стадии обработки, чтобы было видно только те, над которыми надо работать, или только те, которые уже привязаны или находящиеся на проверке.

После привязки ссылки на работу автора в списке литературы эта корректировка посылается на проверку модератору. Мне так и не удалось узнать в РИНЦ, установлены ли в РИНЦ какие-либо нормативы на максимальную длительность его проверки. Не исключено, что ждать придется годами. Поэтому предлагается ссылки, привязанные администратором SCIENCE INDEX от организации, не посылать на проверку модератору РИНЦ, а сразу включать в базы данных РИНЦ, или накапливать а включать в базы даны раз в неделю. Если РИНЦ заботится о качестве привязки ссылок администратором SCIENCE INDEX от организации, то пусть сам по своему регламенту проверяет качество этих привязок, и если они не соответствуют установленным критериям качества, которые должны быть опубликованы на сайте РИНЦ, то применять к данному администратору и к его организации санкции (меры воздействия), прописанные в догово-

ре. Например, можно лишать данного администратора прав доступа на определенное время или требовать от организации его замены и т.п. Но реакция системы РИНЦ на корректировки администратора SCIENCE INDEX от организации обязательно должна быть резко ускорена.

Но главное все же не это, хотя это и важно. Главное в том, что при привязке ссылок в списках литературы возникают безвыходные ситуации. Даже имея права доступа администратора системы SCIENCE INDEX от организации, далеко не всегда удастся привязать ссылку на работу автора из списка литературы ссылающейся работы. Мы можем скорректировать эту ссылку на правильную, но система не привязывает ее, пока не найдет в базах РИНЦ. Это вообще неприемлемо. Во-первых, потому, что есть работа в базах РИНЦ или ее там нет, не должно влиять на число ссылок на нее, т.к. это работа автора, на которую ссылается другой автор, и он эту работу видел, когда ссылался. Во-вторых, сам поиск не всегда дает положительный результат, т.е. не всегда способен найти работу, даже если она есть в базах РИНЦ и ссылка на нее правильная, в т.ч. сформирована самой системой РИНЦ. В этом случае, который наблюдается примерно в 70% случаев поиска, у администратора SCIENCE INDEX от организации вообще нет способа привязать эту ссылку. Автор предлагал администраторам РИНЦ два решения этой проблемы. Первое подробнейшим образом описано в статье [6]. Работать это решение будет быстро, т.к. у авторов в РИНЦ не более 300 работ, а чаще всего гораздо меньше. Это решение обеспечивает ранжирование работ автора в порядке убывания релевантности ссылке. Проведенные автором численные эксперименты убедительно продемонстрировали, что искомая работа практически всегда будет на первой позиции, т.е. будет иметь наивысшую релевантность. Второе решение вообще примитивное и состоит в том, что если поиск системы РИНЦ не может найти в работах автора, размещенных в РИНЦ ту, на которую сделана ссылка из списка литературы, то надо просто вывести весь список его работ и дать возможность администратору SCIENCE INDEX от организации просто указать в нем нужную работу.

Иначе при работе с РИНЦ слишком часто возникают безвыходные ситуации и заслуженные ученые, имеющие сотни работ, размещенных в РИНЦ и индекс Хирша 5 или 6 и видящие, что невозможно его увеличить не только авторам, но даже имея права администратора SCIENCE INDEX от организации, делают вполне определенные и легко прогнозируемые выводы и о системе РИНЦ, и о Хирше, и о всех, кто все это придумал. Все это дискредитирует и систему РИНЦ и основанную на данных РИНЦ систему оценки результатов научной деятельности.

Автор данной работы представляет политематический (мультидисциплинарный) журнал, издаваемый Кубанским агроуниверситетом [7]⁴⁵. В университете (<http://kubsau.ru/>) в настоящее время работает 26 факультетов, 85 кафедр, 8 докторских диссертационных советов по 21 специальности, около 300 докторов наук, профессоров и 700 кандидатов наук, доцентов, 1500 преподавателей (а в 2003 году, когда создавался журнал, кафедр было около 100 и действовало 12 диссертационных советов). Поэтому естественно, что журнал изначально создавался как мультидисциплинарный (политематический), т.к. именно такой журнал был нужен университету. Еще отмечу, что наш вуз входит в тройку крупнейших патентообладателей России и имеет в 4 раза больше патентов, что все аграрные вузы России ВМЕСТЕ ВЗЯТЫЕ (включая и Темирязевку). За время существования журнала с 2003 года по сентябрь 2016 года в свет вышло 120 номеров, в которых опубликовано 6090 статей. Среди 10770 авторов журнала из России и более 10 стран ближнего и дальнего зарубежья (<http://ej.kubagro.ru/geo.asp>) 3175 докторов наук, 2832 профессоров, 3447 кандидатов

⁴⁵ <http://ej.kubagro.ru/>

наук, 2505 доцентов (<http://ej.kubagro.ru/st.asp>). В среднем ежемесячно в журнале издается 51 статья общим объемом 730 страниц 90 авторов из которых 26 доктора наук, 29 кандидата наук, 24 профессоров, 21 доцент. Но в последние годы объем публикаций резко возрос (см.: <http://ej.kubagro.ru/st.asp> и работу [7]). Например, в 101-м номере Научного журнала КубГАУ опубликовано столько статей, сколько в «других журналах» публикуется примерно за 2 – 2.5 года. Как Вы думаете, где будут публиковаться наши ученые – сотрудники университета: в «других журналах» или в «нашем журнале»? Да у них просто ФИЗИЧЕСКИ практически нет никакой возможности публиковаться где-то еще. Тем более, что публикация в нашем журнале БЕСПЛАТНА для сотрудников университета и аспирантов из любых организаций СНГ. Ну и как это скажется на индексе Херфиндаля? Понятно, что плохо, точнее очень плохо. По этой причине я ПРОТИВ модификации методики расчета рейтинга SCIENCE INDEX с учетом индекса Херфиндаля, т.к. это эквивалентно наказанию крупнейших и наиболее успешных изданий в своей области (и не только в своей области) и крупнейших издателей. При этом я согласен, что для мелких и средних по объему изданий его применение может быть вполне оправдано. Поэтому я предлагаю придавать тем меньший вес индекса Херфиндаля в определении рейтинга SCIENCE INDEX, чем больше *объем номеров* издания в страницах за соответствующий период, за который определяется рейтинг. Для нашего Научного журнала КубГАУ индекс Херфиндаля практически вообще не должен играть никакой роли.

6.2.2. Идея предлагаемого решения проблемы

Недавно все Российское профессиональное научно-педагогическое сообщество стало свидетелем того, как Министерство образования и науки России начало работу по монетизации оценки результатов научной деятельности (ее качества и эффективности).

В этой связи возникает ряд вопросов, аргументированные ответы на которые представляют большой интерес.

Прежде всего, возникает вопрос о том, **что понимается под результатами (качеством и эффективностью) научной деятельности (далее: «результаты»)?** Ведь ясно, что прежде чем оценивать результаты научной деятельности было бы неплохо, а на самом деле совершенно необходимо, разобраться с тем, что же это такое. Ясно, что по этому поводу существует много различных мнений, которые в различной степени аргументированы или не аргументированы и отражают позиции руководителей образования и науки, профессионального научно-педагогического сообщества и различных слоев населения. По мнению автора, с научной точки зрения некорректно и неуместно говорить о каких-то критериях оценки результатов научной деятельности, если не определено само это понятие, т.е. отсутствует консенсус в профессиональной среде по поводу того, что же это такое. Очевидно, для достижения такого консенсуса в наше время **необходимо широкое обсуждение этого вопроса в научной печати, Internet и СМИ.**

Когда консенсус профессионального научно-педагогического сообщества по вопросу о том, что такое «результаты научной деятельности» будет достигнут, на первый план выступает вопрос о том, с **помощью какого метода оценивать эти результаты?** Для автора вполне очевидно, что этот метод должен представлять собой какой-то вариант метода **многокритериальной оценки.** Это обусловлено просто тем, что такие сложные и многофакторные системы как наука в принципе невозможно оценивать по одному показателю или критерию. Хиршамания возникла имена благодаря вольному или невольному, сознательному или несознательному игнорированию этого принципа.

Чтобы обоснованно выбрать метод оценки результатов научной деятельности необходимо сначала научно обосновать требования к нему, а затем составить рейтинг методов по степени соответствия обоснованным требованиям и выбрать метод, наиболее удовлетворяющий обоснованным требованиям.

Когда метод оценки результатов научной деятельности выбран, необходимо ответить на вопрос о том, *на основе каких частных критериев оценивать эти результаты и какой исходной информацией для этого необходимо располагать?* Ясно, что эти критерии в общем случае могут иметь как количественную, так и качественную (нечисловую) природу и могут измеряться в различных единицах измерения. Кроме того эти критерии могут иметь различную силу и направление влияния на оценку результатов научной деятельности. Поэтому предварительно надо бы обосновать требования к частным критериям оценки результатов научной деятельности. Это специальная наукоемкая работа, но для автора и сейчас очевидно, что эти критерии должны быть:

- измеримы, т.е. по ним должна быть исходная информация;
- информативны, т.е. обеспечивать разделение измеряемого объекта по категориям (классам) качества и эффективности научной деятельности;
- не управляемы самим объектом, параметры которого измеряются, т.к. иначе он может влиять на результаты измерения в нужном ему направлении.

Индекс Хирша соответствует первым двум требованиям, но не удовлетворяет третьему, т.е. он вполне управляем потому, что вполне понятно, как он формируется, и авторы в состоянии писать статьи и ссылаться на такие свои работы, чтобы индекс Хирша повышался максимально быстро. Поэтому величина индекса Хирша отражает не только результаты научной деятельности, но и степень понимания автором того, что такое индекс Хирша и как он формируется (см. юмористический эпиграф к данному разделу). В психологии считается, что нельзя пользоваться тестом, ключи интерпретации которого рассекречены (опубликованы), т.к. при желании тестируемый может использовать знание этих ключей для того, чтобы так отвечать на тест, чтобы получить нужные ему результаты тестирования. Это ведь элементарно. Непонятно почему такие простые вещи игнорируются системами, вроде РИНЦ.

Представьте себе мальчишку, который не хочет идти в школу и говорит маме, что у него болит голова. Мама сразу достает термометр, чтобы померить ему температуру, сбивает его, ставит своему отпрыску под мышку и бежит на кухню выключать картошку, которую варит. А мальчишка в это время на одно мгновение окунает термометр в чай и сразу кладет его обратно себе под мышку и тихо сидит с грустным видом. Мама прибегает, смотрит термометр и сразу начинает принимать меры для лечения своего мелкого симулянта, а о школе теперь не может быть и речи. Спрашивается, является ли неисправным измерительный инструмент, т.е. термометр? Нет, конечно, он исправен и совершенно правильно измеряет температуру. *Но объект измерения (симулянт) заинтересован в тех или иных показаниях и не только в принципе может влиять на показания измерительного инструмента, но и фактически делает это.* Примерно тоже самое мы наблюдаем в ситуации с индексом Хирша. Сам по себе это нормальный измерительный инструмент. Но измерительный инструмент, легко управляемый заинтересованной стороной. Поэтому он не пригоден для тех целей, для которых предназначен.

Конечно, возникают вопросы как *о способе определения системы критериев оценки результатов научной деятельности, так и о способе определения силы и направления влияния этих критериев на оценку результатов научной деятельности, т.е. по сути, о модели.* Но еще более существенным является вопрос: *«О способе сопоставимого сведения разнородных по своей природе и измеряемых в различных*

единицах измерения частных критериев эффективности в один количественный интегральный критерий эффективности вуза».

Автоматизированный системно-когнитивный анализ является одним из современных методов, который предоставляет научно обоснованные ответы на все эти вопросы, но самое существенное, что он оснащен широко и успешно апробированным [8] универсальным программным инструментарием, позволяющим решить эти вопросы не только на теоретическом концептуальном уровне, но и **практически**. Очень важно, что этот инструментарий и методики его использования для решения сформулированных задач могут быть доступны всем заинтересованным сторонам не только на федеральном уровне, но в самих вузах и НИИ, а также конкретным ученым, т.к. он находится в полном открытом бесплатном доступе (на сайте автора по адресу: http://lc.kubagro.ru/aidos/_Aidos-X.htm).

Естественно, никто технологию не продает, а если и продает, то так дорого, что купить ее практически невозможно. Поэтому возникает вопрос о разработке или поиске подобной технологии в России.

Таким образом, востребованы теоретическое обоснование, математическая модель, методика численных расчетов (т.е. структуры данных и алгоритмы их обработки) а также реализующие их инструментальные (программные) средства, обеспечивающие создание, поддержку, развитие и применение подобных рейтингов.

Данная статья как раз и посвящена рассмотрению отечественной лицензионной инновационной интеллектуальной технологии, обеспечивающей решение поставленной проблемы. А именно предлагается применить для этой цели автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальную систему «Эйдос».

6.2.3. Автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос» как инструментарий решения проблемы

АСК-анализ представляет собой один из современных инновационных методов искусственно интеллекта, который предоставляет научно обоснованные ответы на все эти вопросы, но самое существенное, что он оснащен широко и успешно апробированным универсальным программным инструментарием, позволяющим решить эти вопросы не только как обычно на теоретическом концептуальном уровне, но и на практике [8]. Модели знаний АСК-анализа основаны на нечеткой декларативной модели представления знаний, предложенной автором в 1983 году и являющейся гибридной моделью, сочетающей в себе преимущества фреймовой, нейросетевой и четкой продукционной моделей и обеспечивающей создание моделей очень больших размерностей до 10 млн. раз превышающих максимальные размерности моделей знаний экспертных систем с четкими продукциями:

– от фреймовой модели модель представления знания системы «Эйдос» отличается существенно упрощенной программной реализацией и более высоким быстродействием без потери функциональности;

– от нейросетевой тем, что обеспечивает хорошо обоснованную теоретически содержательную интерпретацию весовых коэффициентов на рецепторах и обучение методом прямого счета;

– от четкой продукционной модели – нечеткими продукциями, представленными в декларативной форме, что обеспечивает эффективное использование знаний без их многократной генерации для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта.

АСК-анализ является непараметрическим методом, устойчивым к шуму в исходных данных, позволяющий корректно обрабатывать неполные (фрагментированные) исходные данные, описывающие воздействие взаимосвязанных факторов на нелинейный объект моделирования.

Суть метода АСК-анализа в том, что он позволяет рассчитать на основе исходных данных какое *количество информации* содержится в значениях факторов, обуславливающих переходы объекта моделирования в различные будущие состояния, причем как в желательные, так и в нежелательные [9]⁴⁶.

Он состоит в целенаправленном *последовательном повышении степени формализации* исходных данных до уровня, который позволяет ввести исходные данные в компьютерную систему, а затем преобразовать исходные данные в информацию; информацию преобразовать в знания; использовать знания для решения задач прогнозирования, принятия решений и исследования предметной области.

Рассмотрим подробнее вопросы выявления, представления и использования знаний в АСК-анализе и системе «Эйдос».

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ), развиваемой проф. Е.В.Луценко, для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на корпорацию к каким ее изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

⁴⁶ Вопреки тому, как его поняли некоторые авторы

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу.

Знания – это информация, полезная для достижения целей⁴⁷.

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).
2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно **последовательно повышать степень формализации** исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач управления, принятия решений и исследования предметной области.

⁴⁷ Основные публикации автора по вопросам выявления, представления и использования знаний:

– <http://www.twirpx.com/file/793311/>

– Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.

– Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012\0197, IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.

АСК-анализ имеет следующие этапы [8]:

- когнитивно-целевая структуризация предметной области;
- формализация предметной области (формирование классификационных и описательных шкал и градаций и обучающей выборки);
- синтез и верификация статистических и системно-когнитивных моделей;
- решение задач идентификации, прогнозирования, принятия решений и исследования предметной области в наиболее достоверных из созданных моделей.

Единственный неавтоматизированный в системе «Эйдос» этап – это первый, а остальные приведены на рисунке 1.

АСК-анализ имеет ряд особенностей, которые обусловили его выбор в качестве метода решения проблемы:

5. Имеет *теоретическое обоснование*, основой которого является *семантическая* мера целесообразности информации А.Харкевича.

6. Обеспечивает *корректную сопоставимую количественную* обработку *разнородных* по своей природе факторов, измеряемых в *различных единицах* измерения, *высокую точность* и независимость результатов расчетов от единиц измерения исходных данных.

7. Обеспечивает построение *многомерных моделей* объекта моделирования непосредственно на основе *неполных и искаженных* эмпирических данных о нем.

8. Имеет развитую и *доступную программную реализацию* в виде универсальной когнитивной аналитической системы «Эйдос».

6.2.4. Частные критерии и виды моделей системы «Эйдос»

Частные критерии знаний, используемые в настоящее время в АСК-анализе и системе «Эйдос-Х++», приведены в таблице 1.

Таблица 1 – Частные критерии знаний, используемые в настоящее время в АСК-анализе и системе «Эйдос-Х++»

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета относительных частот: N_j – суммарное количество признаков по j -му классу. Относительная частота того, что если у объекта j -го класса обнаружен признак, то это i -й признак	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij} \cdot N}{N_i \cdot N_j}$
INF2 , частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета относительных частот: N_j – суммарное количество объектов по j -му классу. Относительная частота того, что если предъявлен объект j -го класса, то у него будет обнаружен i -й признак.	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij} \cdot N}{N_i \cdot N_j}$
INF3 , частный критерий: Хи-квадрат: разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \frac{N_i \cdot N_j}{N}$

INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета относительных частот: N_j – суммарное количество признаков по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_iN_j} - 1$
INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета относительных частот: N_j – суммарное количество объектов по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_iN_j} - 1$
INF6 , частный критерий: разность условной и безусловной относительных частот, 1-й вариант расчета относительных частот: N_j – суммарное количество признаков по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7 , частный критерий: разность условной и безусловной относительных частот, 2-й вариант расчета относительных частот: N_j – суммарное количество объектов по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Обозначения:

i – значение прошлого параметра;

j – значение будущего параметра;

N_{ij} – количество встреч j -го значения будущего параметра при i -м значении прошлого параметра;

M – суммарное число значений всех прошлых параметров;

W – суммарное число значений всех будущих параметров.

N_i – количество встреч i -м значения прошлого параметра по всей выборке;

N_j – количество встреч j -го значения будущего параметра по всей выборке;

N – количество встреч j -го значения будущего параметра при i -м значении прошлого параметра по всей выборке.

I_{ij} – частный критерий знаний: количество знаний в факте наблюдения i -го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее j -му значению будущего параметра;

Ψ – нормировочный коэффициент (Е.В.Луценко, 1979, впервые опубликовано в 1993 году [15]), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;

P_i – безусловная относительная частота встречи i -го значения прошлого параметра в обучающей выборке;

P_{ij} – условная относительная частота встречи i -го значения прошлого параметра при j -м значении будущего параметра.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (таблица 4)⁴⁸ в матрицы условных и безусловных процентных распределений (таблицы 5 и 6) и матрицы знаний (проф. В.И.Лойко, 2014).

⁴⁸ Которая является также матрицей сопряженности или корреляционной матрицей.

6.2.5. Ценность описательных шкал и градаций для решения задач идентификации текстов и авторов (нормализация текста)

Для любой из моделей системой «Эйдос» рассчитывается *ценность*⁴⁹ градации описательной шкалы, т.е. признака, для идентификации или прогнозирования. ***Количественной мерой ценности признака в той или иной модели является вариабельность по классам частного критерия для этого признака*** (таблица 1) Мер вариабельности может быть много, но наиболее известными является среднее модулей отклонения от среднего, дисперсия и среднеквадратичное отклонение. Последняя мера и используется в АСК-анализе и системе «Эйдос».

В системе «Эйдос» ценность признаков нарастающим итогом выводится в графической форме.

При большом объеме обучающей выборки можно без ущерба для достоверности модели удалить из нее малозначимые признаки (Парето-оптимизация). Для этого в системе «Эйдос» также есть соответствующие инструменты.

6.2.6. Интегральные критерии системы «Эйдос»

Но если нам известно, что объект обладает не одним, а несколькими признаками, то как посчитать их *общий* вклад в сходство с теми или иными классами? Для этого в системе «Эйдос» используется 2 аддитивных интегральных критерия: «Сумма знаний» и «Семантический резонанс знаний».

Интегральный критерий «Семантический резонанс знаний» представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где: M – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j-го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

⁴⁹ Эта ценность в АСК-анализе называется также интегральной информативностью, дифференцирующей или дискриминантной способностью и селективной силой, т.е. эти термины являются синонимами.

$$\bar{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } : n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Интегральный критерий «Семантический резонанс знаний» представляет собой нормированное суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3 и имеет вид:

$$I_j = \frac{1}{\sigma_I \sigma_L M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

M – количество градаций описательных шкал (признаков);

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору объекта;

σ_I – среднее квадратичное отклонение частных критериев знаний вектора класса;

σ_L – среднее квадратичное отклонение по вектору распознаваемого объекта.

$\bar{I}_{ij} = \{I_{ij}\}$ – вектор состояния j-го класса;

$\bar{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\bar{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } : n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}.$$

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j -го класса и состояния распознаваемого объекта.

6.2.7. Выводы

Недавно был начат процесс монетизации оценки результатов научной деятельности и возникла потребность в методиках количественной и сопоставимой оценки эффективности и качества работы ученого. Появились многочисленные методики материального поощрения за эти результаты. Общим для всех этих методик является завышенная роль индекса Хирша. Сам по себе этот индекс вполне обоснован. Однако в связи с практикой применения индекса Хирша в наших условиях в сознании научного сообщества возникла своеобразная *мания*, которую автор предлагает называть «Хиршамания». Эта мания характеризуется повышенным нездоровым интересом к самому *значению* индекса Хирша, особенно к искусственному неадекватному преувеличению этого значения, а также рядом *негативных* последствий этого интереса. В данной работе делается попытка кратко описать некоторые негативные последствия этой новой психической инфекции, поразившей общественное сознание научного сообщества. А также наметить пути преодоления хотя бы некоторых *причин* их возникновения. В этом и состоит проблема, решаемая в данной работе. Для решения сформулированной проблемы предлагается применить многокритериальный подход, основанный на теории информации, а именно тот его вариант, который реализован в автоматизированном системно-когнитивном анализе (АСК-анализ) и его программном инструментарии – интеллектуальной системе «Эйдос»

6.3. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию

Высшая аттестационная комиссия (ВАК) – это своего рода отдел технического контроля (ОТК), оценивающий «качество продукции» не только Министерства образования и науки, но и всех других министерств и ведомств, в которых есть свои вузы и НИИ. В качестве продукции вузов и НИИ выступают не только их разработки, но и сами ученые. ВАК, как и ОТК, обеспечивает обратную связь, информируя управляющую систему о результатах ее работы. Из теории управления известно, что если информация обратной связи неадекватна, то и управляющие решения, принимаемые на ее основе, также будут неадекватными. ***Понятно, что оценивающая структура не должна находиться в подчинении у той структуры, качество работы которой она оценивает. В противном случае нетрудно догадаться, как она будет оценивать.*** В СССР ВАК с 1975 и до самого распада СССР подчинялась не Министерству образования и науки, а непосредственно Совету министров СССР, что соответствует этой логике. Однако с тех пор *существует устойчивая тенденция постепенного снижения статуса ВАК*. Сегодня ВАК уже не просто входит в Минобрнауки, а является всего лишь одним из подразделений одной из его структур: Рособрнадзора. ***Снижение статуса ВАК неизбежно приводит к снижению как статуса, так и адекватности присваиваемых им ученых степеней и научных званий.*** Этот процесс обесценивания традиционных ученых степеней и званий, присваиваемых ВАК, дошел до того, что несколько лет назад отменили надбавки к заработной плате за них. Теперь ***вместо*** традиционных ученых

степеней и званий, присваиваемых ВАК *практически каждым вузом и НИИ разрабатывается свои локальные, т.е. несопоставимые друг с другом наукометрические методики оценки результатов научной и педагогической деятельности.* При всем разнообразии этих методик общим для всех них является несоразмерно большая роль, которая отводится в них индексу Хирша. Значение индекса Хирша начинает играть важную роль при защитах, при рассмотрении конкурсных дел на замещение должностей, а также при определении величины ежемесячного материального поощрения за результаты научной и педагогической деятельности. Сам по себе этот индекс *теоретически* вполне обоснован. Однако в связи с *практикой* применения индекса Хирша в наших условиях в сознании научного сообщества возникла своеобразная мания, которую авторы называют «Хиршамания» [1]. Эта мания характеризуется повышенным нездоровым интересом к самому значению индекса Хирша, а также к **некорректному манипулированию его значением**, т.е. к искусственному неадекватному преувеличению этого значения, а также рядом негативных последствий этого интереса.

Возникают естественные вопросы:

1. Возможно ли как-то количественно оценить степень манипулирования индексом Хирша, т.е. то, в какой степени его значение «целенаправленно организовано»?

2. Возможно ли получить гипотетическое значение индекса Хирша каким оно было бы в случае отсутствия манипулирования им?

В данной работе делается попытка найти конкретные ответы на эти вопросы путем:

– конструирования количественной меры для оценки степени некорректного манипулирования значением индекса Хирша;

– разработки научно-обоснованной модификации индекса Хирша, нечувствительной (устойчивая) к попыткам манипулированию им.

Кроме собственно самих идей предлагается также методика всех численных расчетов, достаточно простая, чтобы ее мог применить каждый автор.

6.3.1. Что такое индекс Хирша

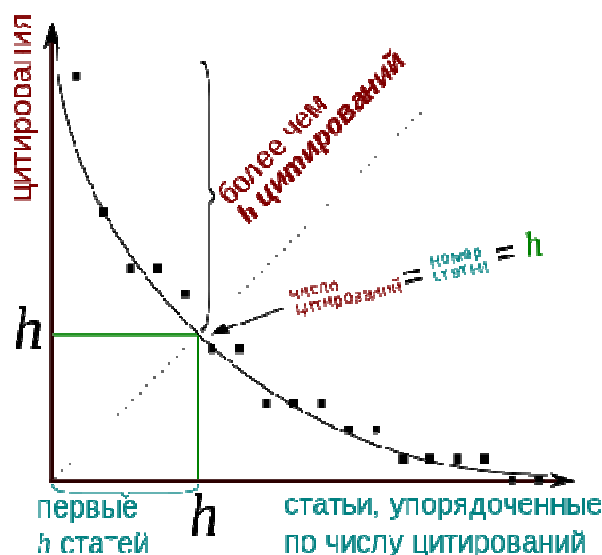


Рисунок 1 – Распространенное в Internet пояснение к понятию: «индекс Хирша»⁵⁰

⁵⁰ Источник: <http://belmapo.by/assets/images/index-hirsha.png>

Если ранжировать все публикации ученого в порядке убывания числа их цитирований («*ранжированный список публикаций*»), то индекс Хирша h – это просто номер публикации в этом списке, процитированной h раз. За этой публикацией идут публикации, процитированные менее h раз, а до нее – более h раз.

Таким образом, индекс Хирша является абсциссой точки пересечения графика числа цитирований для ранжированного списка публикаций с биссектрисой первого квадранта. Пусть $f(h)$ - число цитирований публикации ранга h (т.е. публикации с номером h в ранжированном списке публикаций). Тогда для индекса Хирша h_0 справедливы неравенства

$$f(h) \geq h \text{ при } h \leq h_0 \text{ и } f(h) < h \text{ при } h > h_0.$$

6.3.2. Манипулирование индексом Хирша при малом числе публикаций

6.3.2.1. СПОСОБ СФОРМИРОВАТЬ МАКСИМАЛЬНОЕ ЗНАЧЕНИЕ ИНДЕКСА ХИРША ПРИ МАЛОМ ЧИСЛЕ ПУБЛИКАЦИЙ

Из приведенного выше нехитрого алгоритма вычисления значения индекса Хирша вполне понятно, как получить максимальное значение индекса Хирша h при минимальном числе публикаций $h+1$. Для этого достаточно опубликовать $h+1$ статей, в каждой из которых сослаться на все остальные [2].

6.3.2.2. ПЕРВЫЙ ИНТЕГРАЛЬНЫЙ КРИТЕРИЙ МАНИПУЛИРОВАНИЯ ИНДЕКСОМ ХИРША

Наверное приведенный выше простой и доступный способ сформировать любое заданное значение индекса Хирша первым приходит всем авторам на ум. И это дает нам в руки первый наиболее простой критерий манипулирования индексом Хирша: «*Чем более пологим является линейный тренд числа цитирований, построенный по ранжированному списку публикаций, тем более вероятно, что был применен описанный выше способ максимизации индекса Хирша при малом числе публикаций*».

Максимальный теоретически возможный угол наклона линейного тренда, достижимый лишь асимптотически, равен 90° , а минимальный, естественно, равен нулю: 0° . Количественно этот 1-й частный критерий по сути должен быть какой-то простой функцией от коэффициента наклона линейного тренда ранжированного списка. Естественным было нормировать 1-й частный критерий манипулирования индексом Хирша таким образом, чтобы при наклоне тренда 90° он имел минимальное значение равное 0 (нет манипулирования), а при наклоне 0° имел максимальное значение, равное 1 (полное манипулирование).

Уравнение линейного тренда выгладит следующим образом:

$$y = k \cdot x + b$$

С учетом всех этих соображений предлагается следующее выражение для 1-го частного критерия манипулирования индексом Хирша K_1 при малом числе публикаций:

$$K_1 = \frac{90 - |\text{ArcTg}(k)|}{90},$$

где:

k – коэффициент при x в линейном тренде ранжированного списка публикаций;

$\text{ArcTg}(k)$ – арктангенс коэффициента наклона – угол наклона линейного тренда ранжированного по числу цитирований списка публикаций (в градусах).

Понятно, что чем более пологим является линейный тренд графика числа цитирований, тем ближе коэффициент b в линейном тренде к значению индекса Хирша h :

при : $k \rightarrow 0, b \rightarrow h$

Предлагается следующее выражение для 2-го частного критерия манипулирования индексом Хирша K_2 при малом числе публикаций:

$$K_2 = 1 - \frac{|b - h_e|}{b + h_e},$$

где:

b – свободный член в линейном тренде графика числа цитирований;

h_e – эмпирическое значение индекса Хирша, т.е. полученное непосредственно из ранжированного списка публикаций и построенного по нему графика числа цитирований.

Естественным было нормировать 2-й частный критерий манипулирования индексом Хирша таким образом, чтобы при эмпирическом индексе Хирша $h_e=0$ он был равен нулю (нет манипулирования), при свободном члене b **равном** эмпирическому индексу Хирша h_e он был равен 1 (полное манипулирование), и при увеличении разницы между ними стремился к нулю (уменьшение степени манипулирования) (таблица 1 и рисунок 2):

Таблица 1 – Зависимость 2-го частного критерия манипулирования индексом Хирша от эмпирического значения индекса Хирша при постоянном свободном члене $b=7$

H	2-й частный критерий	H	2-й частный критерий	H	2-й частный критерий
0	0,000000	13	0,700000	26	0,424242
1	0,250000	14	0,666667	27	0,411765
2	0,444444	15	0,636364	28	0,400000
3	0,600000	16	0,608696	29	0,388889
4	0,727273	17	0,583333	30	0,378378
5	0,833333	18	0,560000	31	0,368421
6	0,923077	19	0,538462	32	0,358974
7	1,000000	20	0,518519	33	0,350000
8	0,933333	21	0,500000	34	0,341463
9	0,875000	22	0,482759	35	0,333333
10	0,823529	23	0,466667	36	0,325581
11	0,777778	24	0,451613	37	0,318182
12	0,736842	25	0,437500	38	0,311111

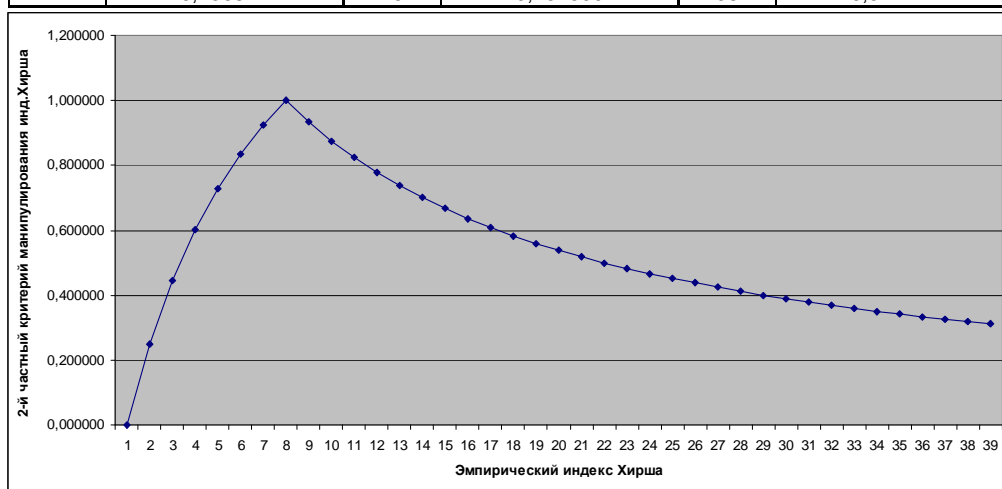


Рисунок 2 – Зависимость 2-го частного критерия манипулирования индексом Хирша от эмпирического значения индекса Хирша при постоянном свободном члене $b=7$

Если считать, что оба эти частных критерия K_1 и K_2 имеют равный вес 0.5, то можно предложить следующее выражение для 1-го интегрального критерия манипулирования индексом Хирша при малом числе публикаций:

$$I_1 = 1 + 0.5 \cdot \left[\frac{90 - |\text{ArcTg}(k)|}{90} - \frac{|b - h_e|}{b + h_e} \right]$$

Все обозначения, использованные в данном выражении, описаны выше.


Данный 1-й интегральный критерий принимает значение равное 0 при отсутствии манипулирования и равное 1 при максимальном, т.е. полном манипулировании. Ниже приведена его вербальная формулировка:

«Чем ближе к нулю коэффициент наклона линейного тренда числа цитирований, построенного по ранжированному списку публикаций и чем ближе свободный член в линейном тренде к эмпирическому значению индекса Хирша, тем более вероятно, что был применен описанный выше способ максимизации индекса Хирша при малом числе публикаций».

Конечно, понятно, что часть цитирований могут естественными, не организованными автором, и они вместе тоже могут формировать достаточно пологий тренд, т.е. понятно, что максимальное значение индекса манипулирования еще не означает самого факта манипулирования, а лишь является его признаком. Аналогично и заимствования сами по себе не означают плагиата, т.к. могут быть снабжены ссылками на источники, а могут быть и заимствованиями из работ самого автора, которые уже по главам порезаны на рефераты и разошлись по всему интернету.

6.3.2.3. ПРИМЕРЫ ПРИМЕНЕНИЯ ПЕРВОГО ИНТЕГРАЛЬНОГО КРИТЕРИЯ МАНИПУЛИРОВАНИЯ ИНДЕКСОМ ХИРША НА ОСНОВЕ БАЗ ДАННЫХ РИНЦ

Для того, чтобы применить этот интегральный критерий к публикациям какого-либо автора выполняем следующие действия:

1. Открываем сайт РИНЦ: <http://elibrary.ru/>.
2. В меню слева выбираем «Авторский указатель», задаем сортировку по числу цитирований по убыванию без фильтра по региону. В результате получаем (на момент написания статьи) (рисунок 3).
3. Выбираем автора, по которому собираемся анализировать индекс Хирша (Новоселов К.С.), кликаем по числу его работ (левее гистограммки: ) , выделяем блоком **вместе с заголовком таблицы** первые его 100 публикаций (или все, если их меньше 100), копируем его в буфер обмена и вставляем в MS Excel (используем копировать: Ctrl+C, и вставить: Ctrl+V или эти пункты в меню, выскакивающему по клику на правой кнопке мыши).
4. Выделяем блоком весь лист отменяем объединение ячеек.
5. Переносим колонку **D** с числом цитирований в колонку **C** (если они не в колонке **C**).
6. Начиная с колонки **D** вставляем следующие значения и формулы для построения графика цитирований и расчета трендов (рисунок 4):

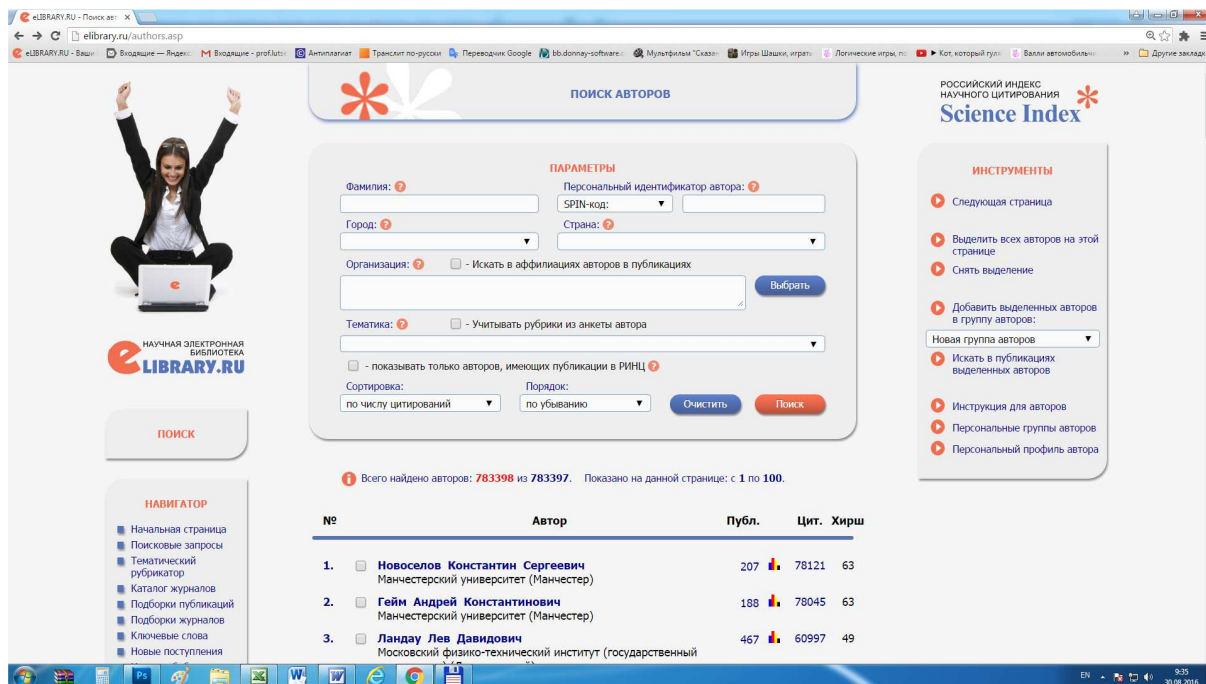


Рисунок 3 – Экранная форма РИНЦ: «Авторский указатель», сортировка по числу цитирований по убыванию без фильтра по региону

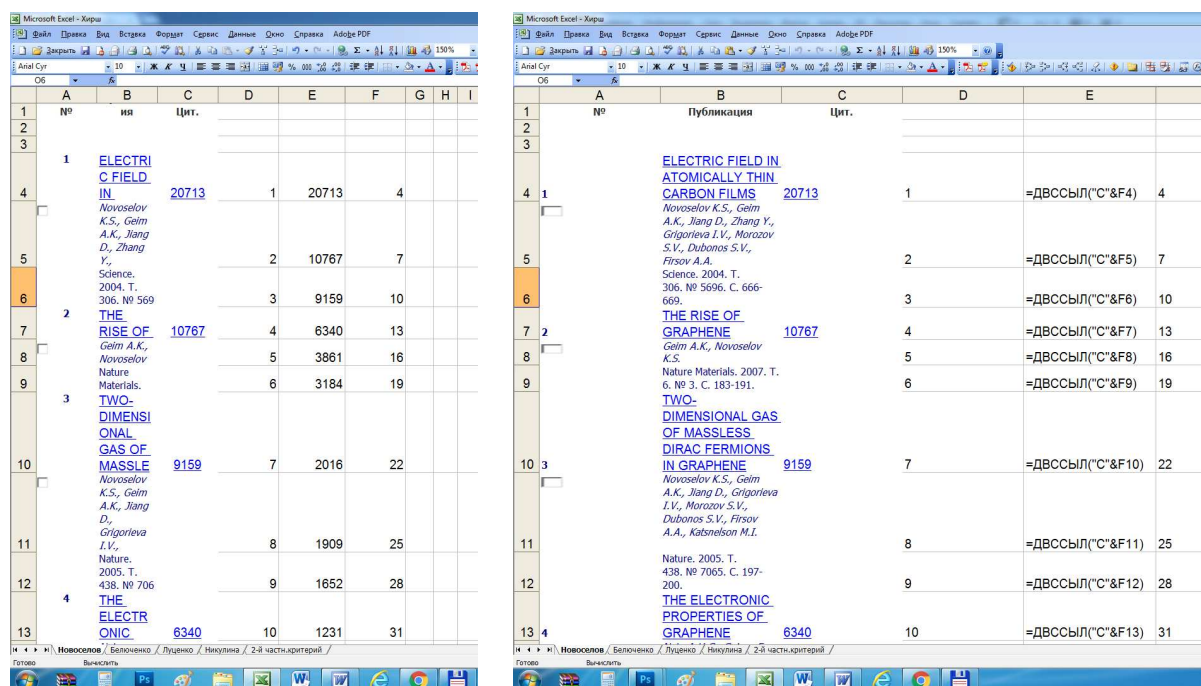


Рисунок 4 – Значения и формулы для построения графика цитирований и расчета трендов

В колонке **D** просто подряд пронумерованы строки с 1 до 100. В колонке **F** подряд идущих строках проставлены номера строк, в которых в колонке **C** приведено число цитирований: 4, 7, 10, 13 и т.д. с шагом 3. В колонке **E** приведены формулы ссылок на ячейки с числом цитирований из колонки **C**. Все это сделано для того, чтобы значения числа цитирований для различных публикаций шли в подряд идущих строках, а не в каждой третьей строке, начиная с 4-й, как это сделано в РИНЦ. Отметим, что и в РИНЦ шаг 3 между строками с числом цитирований может нарушаться, хотя это про-

исходит и редко. Например, у автора: Новоселов К.С. в 64-й публикации (193-я строка в списке РИНЦ) дано не совсем стандартное описание. Поэтому для 65-й публикации вместо 196 строки указана 201-я, в которой фактически находится число цитирований 65-й публикации. Далее и до 10-й публикации они опять идут с стандартным шагом 3. *Чтобы не пропустить подобные ситуации **рекомендуется** проверять значения числа цитирований не только в первых, но и в последних строках списка.*

7. Строим график по числу цитирований. Для этого выделяем блоком ячейки в колонке **Е**, в которых есть число цитирований (удобнее это делать снизу вверх), и строим график (рисунок 5):

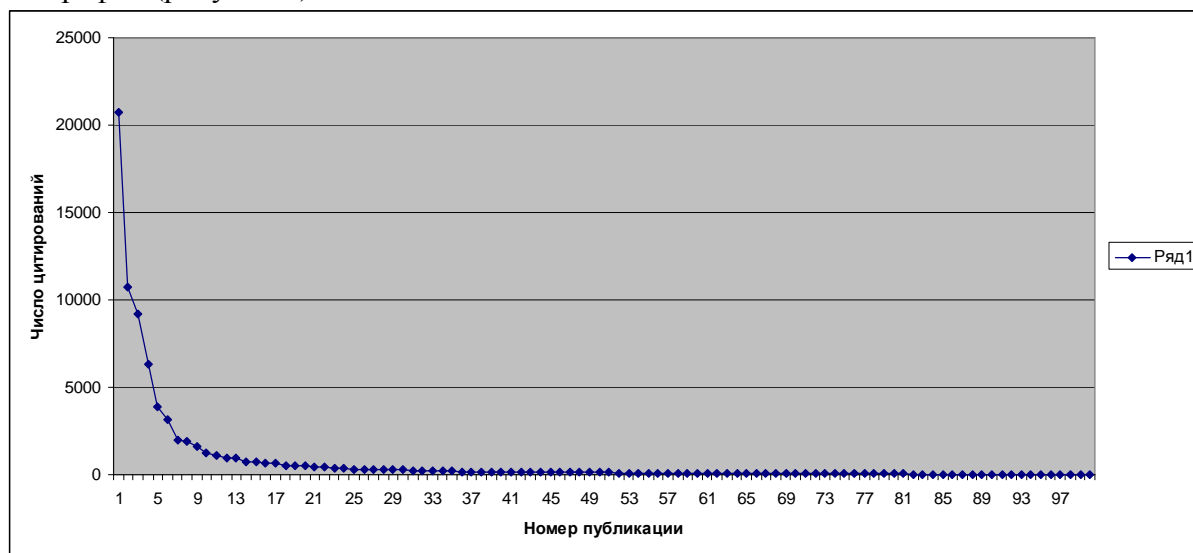


Рисунок 5 – График числа цитирований, построенный по списку публикаций Новоселова К.С., ранжированному по числу цитирований в порядке убывания

8. Строим линейный тренд графика числа цитирований с выводом формулы тренда и критерия качества аппроксимации – коэффициента детерминации R^2 (рисунок 6):

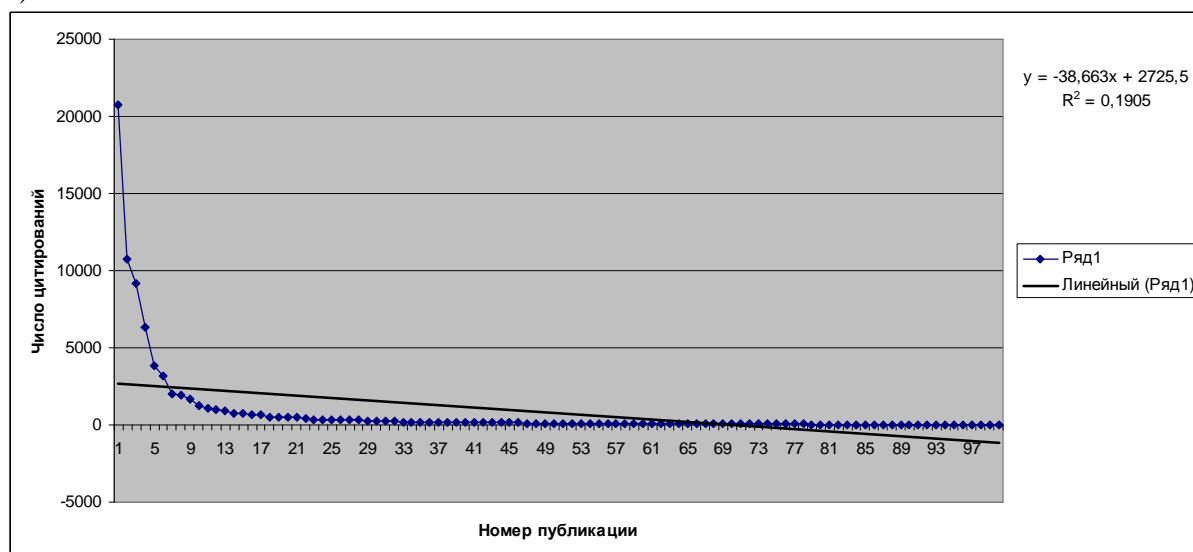


Рисунок 6 – График числа цитирований публикаций Новоселова К.С. с линейным трендом

9. Для расчета частных критериев и интегрального критерия в MS Excel используем формулы, приведенные на рисунке 7:

№	Публикация	Цит.				Козфф. К	Градусы	1-й частный критерий
1						-38.663	=ГРАДУСЫ(ATAN(J2))	=(90-ABS(K2))/90
4	ELECTRIC FIELD IN ATOMICALLY THIN CARBON FILMS Novoselov K.S., Geim A.K., Jiang D., Zhang Y., Grigorieva I.V., Morozov S.V., Dubonos S.V., Firsov A.A. Science. 2004. Т. 306. № 5696. С. 666-669.	20713	1	=ДВССЫЛ("C"&F4)	4	Козфф. В	Эмп. инд. Хирша	2-й частный критерий
5			2	=ДВССЫЛ("C"&F5)	7	2725.5	63	=1-ABS(K5-J5)/(K5+J5)
6			3	=ДВССЫЛ("C"&F6)	10			
7	THE RISE OF GRAPHENE Geim A.K., Novoselov K.S.	10767	4	=ДВССЫЛ("C"&F7)	13	1-й инт. критерий		=(L2+L5)/2
8			5	=ДВССЫЛ("C"&F8)	16			

Рисунок 7 – Формулы для расчета частных критериев и 1-го интегрального критерия манипулирования индексом Хирша при малом числе публикаций

Значения коэффициентов k и b из уравнения линейной регрессии, приведенного на рисунке 6, *вручную* вносим в ячейки **J2** и **J5** соответственно (выделены на рисунке 7 желтым цветом). В результате получим значения частных критериев и интегрального критерия манипулирования индексом Хирша для данного автора (рисунок 8), рассчитанные по приведенным выше формулам.

Из рисунка 8 видно, что все эти значения очень близки к нулю, что означает *полное отсутствие манипулирования* в данном случае.

№	ия	Цит.				Козфф. К	Градусы	1-й критерий
1						-38.663	-88.518402	0.016462196
4	ELECTRIC FIELD IN	20713	1	20713	4	Козфф. В	Эмп. инд. Хирша	2-й критерий
5			2	10767	7	2725.5	63	0.045185584
6			3	9159	10			
7	THE RISE OF	10767	4	6340	13	1-й инт. критерий		0.03082389
8			5	3861	16			

Рисунок 8 – Значения частных критериев и 1-го интегрального критерия манипулирования индексом Хирша для автора: Новоселов К.С.

Рассмотрим применение предлагаемого интегрального критерия на примере **2-го автора**, рейтинг, Ф.И.О. и место работы которого мы не указываем из этических соображений.

На рисунке 9 приведен график числа цитирований с линейным трендом этого 2-го автора, а в таблице 2 результаты расчета частных критериев и интегрального критерия :

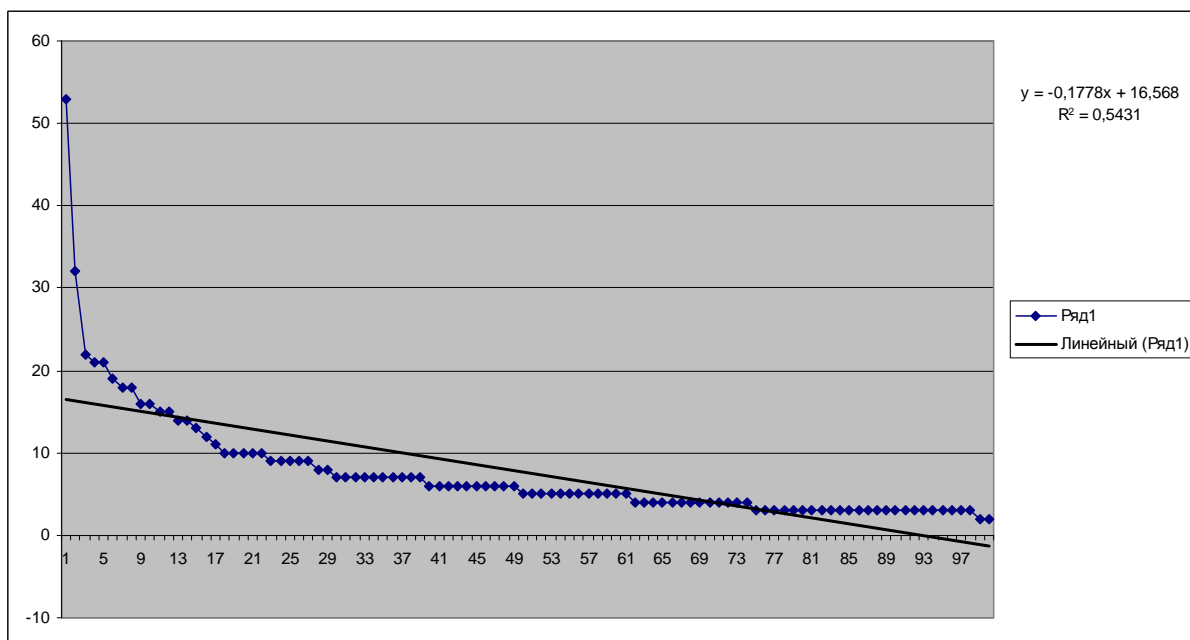


Рисунок 9 – График числа цитирований публикаций 2-го автора с линейным трендом

Таблица 2 – Результаты расчетов частных критериев и 1-го интегрального критерия манипулирования индексом Хирша при малом числе публикаций для 2-го автора

Коэфф. К	Градусы	1-й частный критерий
-0,1778	-10,081832	0,887979642
Коэфф. В	Эмп. инд. Хирша	2-й частный критерий
16,568	14	0,915990578
1-й инт.критерий		0,90198511

Из таблицы 2 видно, что доля манипулирования индексом Хирша в данном случае значительно выше, т.к. значение 1-го интегрального критерия близко к 0,9.

Из приведенных графиков и таблиц мы видим, что для лидера рейтинга РИНЦ по числу цитирований предлагаемый 1-й критерий манипулирования индексом Хирша дает значительно меньшую величину, чем у 2-го автора. Видно, что этот результат получается за счет того, что *у лидера различие между числом цитирований наиболее и наименее цитируемых работ первой сотни работ, значительно больше, чем у обычного автора.*

6.3.3. Манипулирование индексом Хирша при большом числе публикаций

6.3.3.1. СПОСОБ УВЕЛИЧИТЬ ЗНАЧЕНИЕ ИНДЕКСА ХИРША ПРИ БОЛЬШОМ ЧИСЛЕ ПУБЛИКАЦИЙ

Если у автора большое число публикаций, то очевидно, использовать способ формирования максимального значения индекса Хирша, который использовался при малом числе публикаций, т.е. сослаться во всех публикациях на все, *не представляется возможным* по ряду причин. Понятно, что статья, у которой в списке литературы приведено десятки источников и в основном автора самой этой статьи, будет выглядеть

несколько странно⁵¹. Во многих журналах просто есть ограничение и на суммарное число источников в списке литературы и на число источников автора публикации. Но цитирование всех публикаций данного автора в каждой его публикации не только невозможно технически⁵², но и *не имеет особого смысла*, т.к. увеличение числа цитирований статей, находящихся в ранжированном списке далеко от значения индекса Хирша, не окажет влияния на его значение ни в ближайшее время, ни в перспективе (за исключением может быть каких-то научных «бестселлеров», которые сразу становятся очень цитируемыми и сохраняют популярность длительное время).

Поэтому многие авторы, у которых большое количество публикаций, приходят к тому, чтобы *увеличивать число ссылок не на все публикации, а только на те, которые оказывают самое непосредственное влияние на значение индекса Хирша, т.е. на публикации в окрестности индекса Хирша в ранжированном списке публикаций*. В результате вблизи значения индекса Хирша, причем как текущего, так и перспективного с точки зрения этих авторов, формируется характерная «ступенька» или «полочка», показанная на рисунке 10 красным цветом.

В результате такого манипулирования индекс Хирша приобретает вместо значения h некоторое большее значение h_2 . При этом площадь под кривой числа цитирований, соответствующая суммарному числу цитирований автора, увеличивается совершенно незначительно, а значение индекса Хирша за счет этого возрастает довольно заметно, т.е. затраты на это повышение оказываются весьма эффективными.

Вот как выглядит подобная «полочка» на реальном графике числа цитирований, построенном по данным РИНЦ⁵³ 3-го автора (рисунок 11):

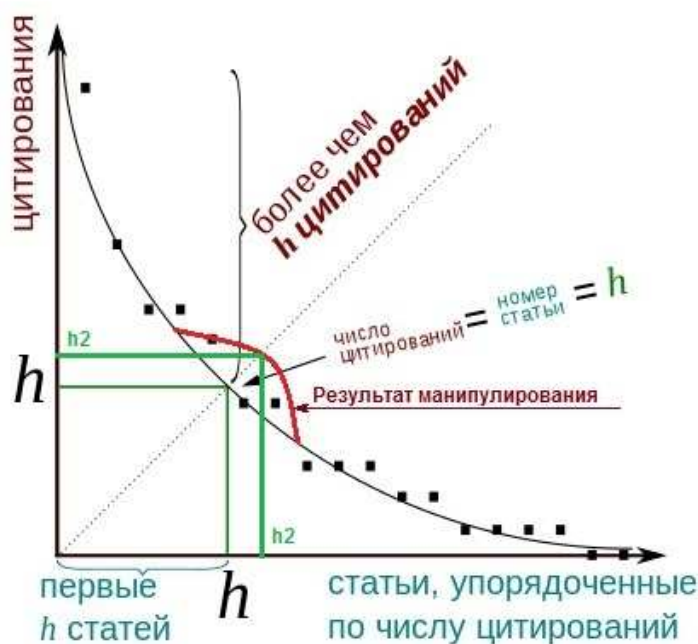


Рисунок 10. Результат манипулирования индексом Хирша при большом числе публикаций: характерная «полочка» в окрестности индекса Хирша в ранжированном списке публикаций (*теория*)

⁵¹ Авторы считают, что это может быть корректным в случае, если автор статьи является главой или активным исследователем определенной научной школы и статья посвящена развитию научного направления этой научной школы.

⁵² За исключением монографий и учебных пособий

⁵³ Фамилия инициалы, место работы автора, цитирования на работы которого анализируются в этом примере, не указывается из этических соображений

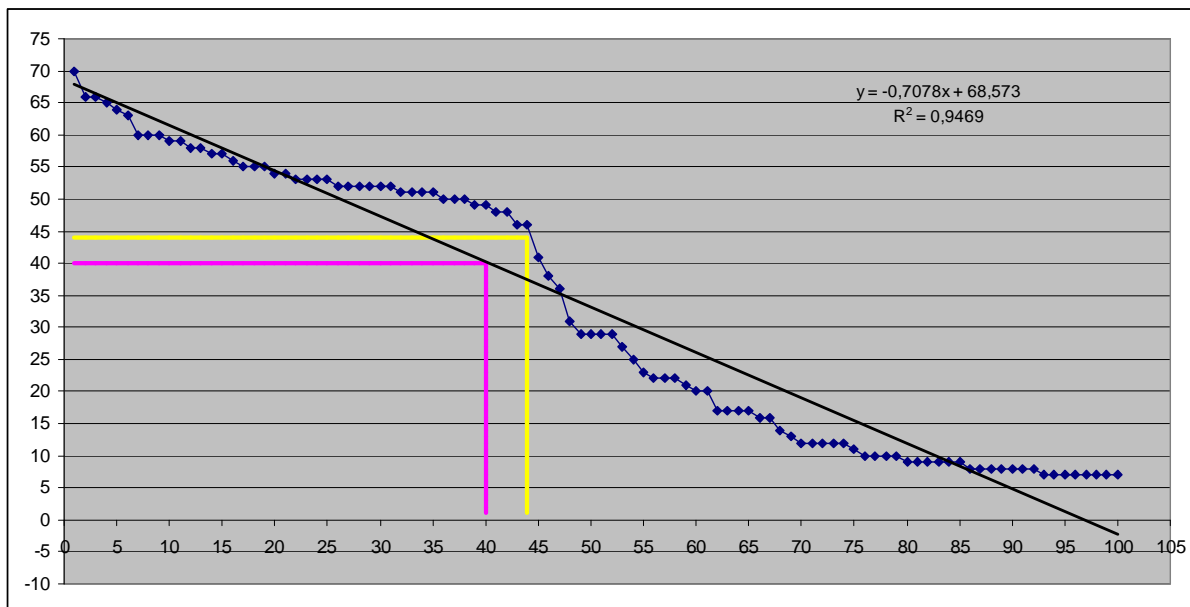


Рисунок 11 – Результат манипулирования индексом Хирша при большом числе публикаций: характерная «полочка» в окрестности индекса Хирша в ранжированном списке публикаций 3-го автора (*факт*)

6.3.3.2. НАУЧНО-ОБОСНОВАННАЯ МОДИФИКАЦИЯ ИНДЕКСА ХИРША, НЕЧУВСТВИТЕЛЬНАЯ (УСТОЙЧИВАЯ) К ПОПЫТКАМ МАНИПУЛИРОВАНИЮ ИМ

Идея второго критерия манипулирования индексом Хирша, применяемого при большом числе публикаций, основана на том, что при цитировании статей в окрестностях текущего значения индекса Хирша площадь под кривой числа цитирований, соответствующая суммарному числу цитирований автора, увеличивается очень незначительно. А это в свою очередь означает, что, по-видимому, *если аппроксимировать эту кривую с использованием метода наименьших квадратов (МНК), то эта аппроксимация окажется малочувствительной или устойчивой к появлению в результате манипулирования этой небольшой «полочки».*

Это позволяет сформулировать гипотезу о том, что значение индекса Хирша, определенное не по классическому алгоритму, а посчитанное на основе аппроксимации кривой числа цитирований, окажется менее чувствительным и более устойчивым к попыткам манипулирования, чем классический индекс Хирша.

Но откуда взять эту аппроксимацию кривой числа цитирований и как определить значение индекса Хирша на ее основе? В общем виде все это довольно -просто. Непосредственно из самого определения классического индекса Хирша следует, что если аппроксимации кривой числа цитирований выражается в виде уравнения:

$$y = f(h)$$

то теоретическим значением индекса Хирша h будет корень уравнения:

$$h = f(h).$$

Такого рода уравнения обычно легко решаются численно итерационным методом, реализованным в частности, в MS Excel.

Сам вид функции $f()$ предлагается определять с использованием аппарата аппроксимации трендов функциями различных видов в MS Excel.

В принципе можно было бы *каждый раз выбирать для аппроксимации тот вид монотонной⁵⁴ функции, который обеспечивает наивысший коэффициент детерминации R^2 , т.е. наиболее хорошее приближение (наилучший тренд)*. В данном случае для аппроксимации графика числа цитирований ранжированного списка публикаций уместно использовать лишь монотонно возрастающие или убывающие функции: линейную, логарифмическую, степенную, экспоненциальную, но не полиномиальную, т.к. она может иметь точки перегиба и даже нарушения монотонности и является чувствительной к особенностям графика, обусловленными манипулированием индексом Хирша.

Но можно выбрать какой-то один вид функции, который чаще других обеспечивает наилучшее приближение. В результате многочисленных численных экспериментов по аппроксимации кривых числа цитирований различных авторов, проведенных по данным РИНЦ, было выявлено, что наилучшее приближение с коэффициентом детерминации около 0,9 и выше, как правило обеспечивается трендом в виде степенной функции:

$$y = a \cdot x^b$$

Поэтому предлагается находить теоретическое значение индекса Хирша h путем решения уравнения:

$$h = a \cdot h^b$$

При этом само уравнение тренда предлагается формировать в MS Excel непосредственно на основе данных РИНЦ, как описано выше в разделе 2.3 при формировании линейной регрессии (примеры приведены ниже).

Решение этого уравнения легко находится аналитически:

$$h = a \cdot h^b$$

$$h \cdot h^{-b} = a \cdot h^{-b} \cdot h^b$$

$$h^{1-b} = a$$

$$h = a^{1/(1-b)}$$

$$h = a^{1/(1-b)}$$

6.3.3.3. ВТОРОЙ ИНТЕГРАЛЬНЫЙ КРИТЕРИЙ МАНИПУЛИРОВАНИЯ ИНДЕКСОМ ХИРША

И это дает нам в руки второй более сложный второй критерий манипулирования индексом Хирша:

«Чем больше отличаются друг от друга эмпирический индекс Хирша, определенный по классическому алгоритму, и теоретический индекс Хирша, найденный путем решения наилучшего уравнения тренда, тем больше вероятность того, что классический индекс Хирша получен в результате манипулирования (хотя возможны и другие варианты: шум и несовершенство алгоритма)».

⁵⁴ линейную, логарифмическую, степенную, экспоненциальную

Аналитически 2-й интегральный критерий манипулирования индексом Хирша, т.е. относительное превышение эмпирического значения индекса Хирша над теоретическим, может быть выражен по-разному. Авторы предлагают измерять это превышение в долях от теоретического значения, как более близкого к истинному:

$$I_2 = \left[\frac{h_e - h_t}{h_t} \right]$$

где:

h_e – классическое эмпирическое значение индекса Хирша;

h_t – классическое эмпирическое значение индекса Хирша.

6.3.3.4. ПРимеры ОПРЕДЕЛЕНИЯ ТЕОРЕТИЧЕСКИХ ЗНАЧЕНИЙ ИНДЕКСА ХИРША ПУТЕМ РЕШЕНИЯ УРАВНЕНИЙ ТРЕНДОВ

Как и в разделе 2.3 примеры рассмотрим на примере тех же авторов:

– Новоселов Константин Сергеевич, имеющий 1-й рейтинг по числу цитирований по данным РИНЦ⁵⁵;

– 2-й и 3-й авторы, рейтинг и Ф.И.О. и место работы которых мы не указываем из этических соображений.

Новоселов Константин Сергеевич.

На графике числа цитирований, приведенном на рисунке 5, построим тренд в виде степенной функции (рисунок 12):

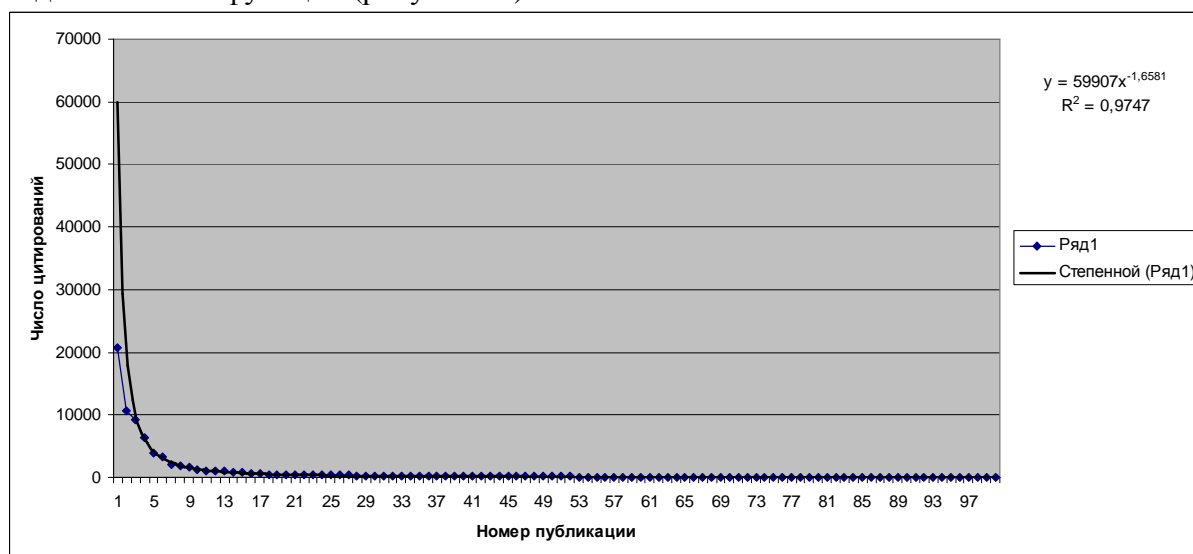


Рисунок 12 – График числа цитирований публикаций Новоселова К.С. и тренд в виде степенной функции

Мы видим, что уравнение тренда имеет вид:

$$y = 59907 \cdot h^{-1,6581}$$

С очень хорошим качеством аппроксимации: $R^2 = 0,9747$.

Для нахождения теоретического значения индекса Хирша необходимо решить уравнение тренда:

$$h = 59907 \cdot h^{-1,6581}$$

⁵⁵ На момент написания статьи

Для решения этого уравнения воспользуемся on-line сервисом Вольфрам-математики по адресу: <http://www.wolframalpha.com/>. Введя решаемое уравнение (заменив в нем запятые на точки, добавив знаки операций и скобки) в окно сервиса, представленное на рисунке 13, получим: $h=62.7$, что после округления с точностью до целых совпадает с эмпирическим значением $h=63$:

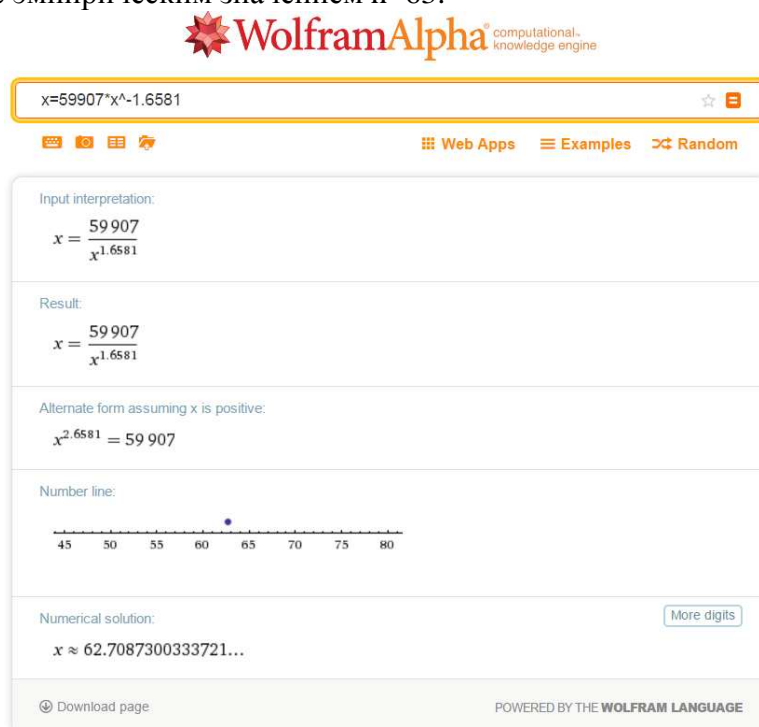


Рисунок 13 – Выходной экран on-line сервиса Вольфрам-математики с решением уравнения тренда графика числа цитирований публикаций Новоселова К.С.

Найденное on-line решение точно совпадает с полученным аналитически:

$$h = a^{1/(1-b)}$$

При решении в MS Excel по этой формуле со значениями коэффициентов: $a=59907$; $b=-1,6581$ получаем $h = 59907^{1/(1+1.6581)}$ или $h=62.7087300333721$, что совпадает по всем знакам после запятой с решением, полученным on-line с помощью Вольфрам-математики.

В разделе 3.3. мы видели, что 2-й интегральный критерий манипулирования индексом Хирша рассчитывается по формуле:

$$I_2 = \left[\frac{h_e - h_t}{h_t} \right]$$

где:

h_e – классическое эмпирическое значение индекса Хирша;

h_t – классическое эмпирическое значение индекса Хирша.

Для Новоселова К.С. это дает значение, весьма близкое к нулю (десятые доли процента):

$$I_2 = \left[\frac{63 - 62.7087300333721}{62.7087300333721} \right] = 0.00464481.$$

Фрагменты Excel-файла, в которых проводятся расчеты по приведенным выше формулам, приведены на рисунках 14 (результаты расчетов) и 15 (формулы):

	A	B	C	D	E	F	G	H	I	J	K	L	M
2		THE											
7		RISE OF	10767	4	6340	13				1-й инт. критерий		0,03082389	
8		Gelm A.K.,		5	3861	16							
9		Novoselov		6	3184	19							
10		Nature								Эмпирический индекс Хирша			
11		Materials.								he=	63		
12		2005. Т.		9	1652	28				Теоретический индекс Хирша			
13		438. № 706								a=	59907		
14		4 THE	6340	10	1231	31				b=	-1,6581		
15		ELECTR		11	1116	34				ht=	62,70873	$h = a^{1/(1-b)}$	
16		ONIC		12	965	37				2-й инт. критерий			
17		Novoselov	3861	13	927	40						0,0046448	
18		K.S., Jiang		14	769	43							
19		D., Schedin											
20		F., Booth											
21		Proceedings											
22		of the											

Рисунок 14 – Фрагмент Excel-файла с расчетами, представленными в таблице 3 (результаты расчетов)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		№	Публикации	Цит.						Кэфф. К	Градусы	1-й частный критерий	
2										=38,663	=ГРАДУСЫ(ATAN(J2))	=(90-ABS(K2))/90	
3													
4		1	ELECTRIC FIELD IN	20713	1	=ДВССЫЛ("C"&F4)	4			Кэфф. В	Эмп. инд. Хирша	2-й частный критерий	
5			ATOMICALLY THIN		2	=ДВССЫЛ("C"&F5)	7			2725,5	63	=1-ABS(K5-J5)/(K5+J5)	
6			CARBON FILMS		3	=ДВССЫЛ("C"&F6)	10						
7			Novoselov K.S., Gelm		4	=ДВССЫЛ("C"&F7)	13						
8			A.K., Jiang D., Zhang T.,		5	=ДВССЫЛ("C"&F8)	16						
9			Grigorieva I.V., Morozov		6	=ДВССЫЛ("C"&F9)	19						
10			S.V., Dubonos S.V.,										
11			Firsov A.A.,		7	=ДВССЫЛ("C"&F10)	22						
12			Science, 2004. Т.		8	=ДВССЫЛ("C"&F11)	25						
13			306, № 6596, С. 666-		9	=ДВССЫЛ("C"&F12)	28						
14			669.		10	=ДВССЫЛ("C"&F13)	31						
15			THE RISE OF	10767	11	=ДВССЫЛ("C"&F14)	34						
16			GRAPHENE		12	=ДВССЫЛ("C"&F15)	37						
17			Gelm A.K., Novoselov		13	=ДВССЫЛ("C"&F16)	40						
18			K.S.,		14	=ДВССЫЛ("C"&F17)	43						
19			Nature Materials, 2007.		15	=ДВССЫЛ("C"&F18)	46						
20			T. 6, № 3, С. 183-191.										
21			TWO-DIMENSIONAL										
22			GAS OF										
23			MASSLESS DIRAC										
24			FERMIONS IN										
25			GRAPHENE	9159	7	=ДВССЫЛ("C"&F10)	22						
26			Novoselov K.S., Gelm		8	=ДВССЫЛ("C"&F11)	25						
27			A.K., Jiang D.,		9	=ДВССЫЛ("C"&F12)	28						
28			Grigorieva I.V., Morozov		10	=ДВССЫЛ("C"&F13)	31						
29			S.V., Dubonos S.V.,		11	=ДВССЫЛ("C"&F14)	34						
30			Firsov A.A.,		12	=ДВССЫЛ("C"&F15)	37						
31			Nature, 2005. Т.		13	=ДВССЫЛ("C"&F16)	40						
32			438. № 7065, С. 197-		14	=ДВССЫЛ("C"&F17)	43						
33			200.		15	=ДВССЫЛ("C"&F18)	46						
34			THE ELECTRONIC										
35			PROPERTIES OF										
36			GRAPHENE	6340	10	=ДВССЫЛ("C"&F13)	31						
37			Neto A.H.C., Guinea F.,		11	=ДВССЫЛ("C"&F14)	34						
38			Peres N.M.R., Novoselov		12	=ДВССЫЛ("C"&F15)	37						
39			K.S., Gelm A.K.		13	=ДВССЫЛ("C"&F16)	40						
40			Reviews of Modern		14	=ДВССЫЛ("C"&F17)	43						
41			Physics, 2009. Т.		15	=ДВССЫЛ("C"&F18)	46						
42			81, № 1, С. 109-162.										
43			TWO-DIMENSIONAL										
44			ATOMIC CRYSTALS	3861	13	=ДВССЫЛ("C"&F16)	40						
45			Novoselov K.S., Jiang		14	=ДВССЫЛ("C"&F17)	43						
46			D., Schedin F., Booth		15	=ДВССЫЛ("C"&F18)	46						
47			T.J., Khotkevich V.V.,										
48			Gelm A.K., Morozov S.V.										
49			Proceedings of the										
50			National Academy of										
51			Sciences of the United										
52			States of America, 2005.										
53			T. 102, № 30, С. 10451-										
54			10453.										
55			PITECTATION OF										

Рисунок 15 – Фрагмент Excel-файла с расчетами, представленными в таблице 3 (расчетные формулы)

Для 2-го автора график числа цитирований публикаций и тренд в виде степенной функции представлены на рисунке 16:

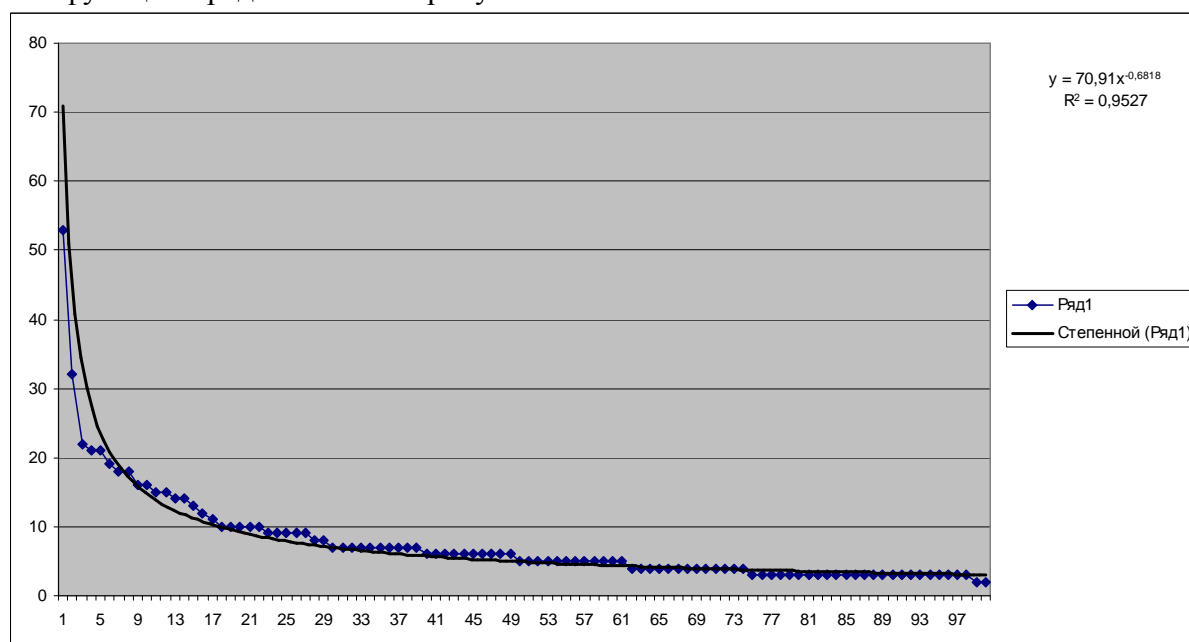


Рисунок 16 – График числа цитирований публикаций 2-го автора и тренд в виде степенной функции

Таблица 3 – Эмпирический и теоретический индексы Хирша и 2-й инт. критерий манипулирования индексом Хирша для 2-го автора

Эмпирический индекс Хирша	
$h_e =$	14
Теоретический индекс Хирша	
$a =$	70,91
$b =$	-0,6818
$h_t =$	12,6017994
2-й инт. критерий	
	0,11095245

Для 2-го автора 2-й интегральный критерий имеет значение порядка 10%.

Для 3-го автора график числа цитирований публикаций и тренд в виде степенной функции представлены на рисунке 17:.

Для 3-го автора эмпирическое значение индекса Хирша равно 44, а теоретическое 40, что дает значение 2-го интегрального критерия манипулирования индексом Хирша: $I_2 = (44 - 40) / 40 = 0.1$. Это значит, что в данном случае манипулирование привело к увеличению индекса Хирша примерно на 10%.

Интересно, что у некоторых авторов теоретическое значение индекса Хирша получается не меньше, а больше эмпирического, т.е. эмпирическое значение «недооценено».

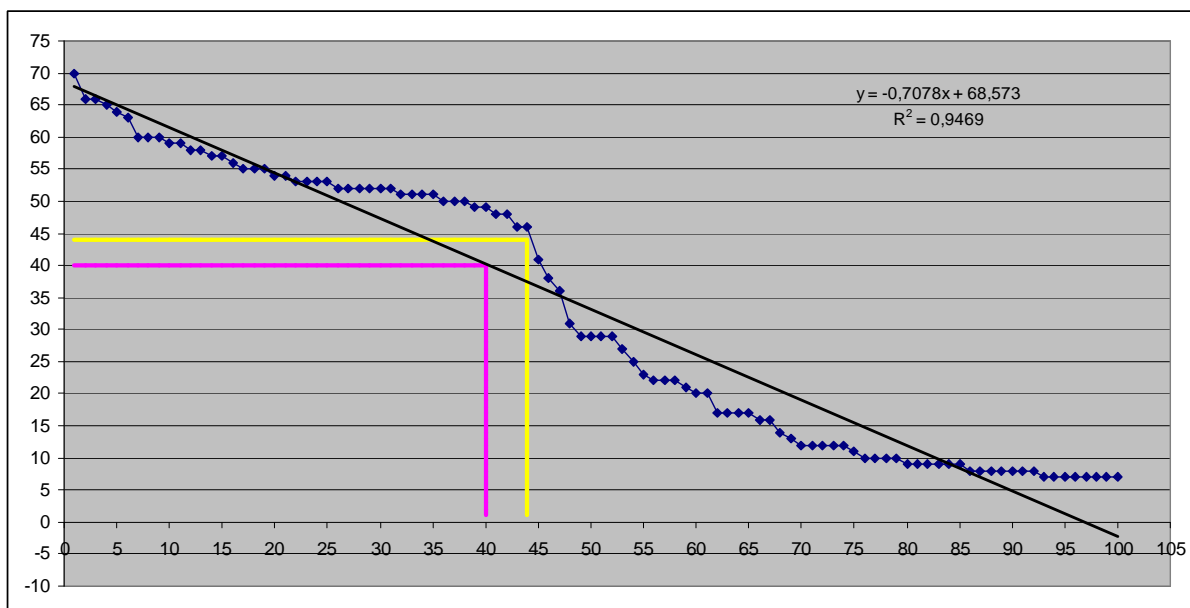


Рисунок 17. График числа цитирований публикаций 3-го автора и тренд в виде степенной функции

6.3.4. Согласованность 1-го и 2-го интегральных критериев манипулирования индексом Хирша

Рассмотрим сводную таблицу 4, в которой приведем все просчитанные в данной работе частные и интегральные критерии по всем авторам:

Таблица 4 – Частные и интегральные критерии по всем авторам

Автор	1-й частный критерий	2-й частный критерий	1-й интегральный критерий	Эмпирический индекс Хирша	Теоретический индекс Хирша	2-й интегральный критерий
Новоселов К.С.	0,01646	0,04518	0,03082	63	62,70873	0,00464
2-й автор	0,88797	0,91599	0,90198	14	12,60179	0,11095
3-й автор	0,60787	0,78171	0,69479	44	40	0,10000

Мы видим, что и частные критерии, и оба интегральных критерия манипулирования индексом Хирша дают согласованные, совпадающие по смыслу результаты, т.е. когда мы не видим манипулирования по 1-му частному критерию, то не видим его и по 2-му, т.е. эмпирический индекс Хирша практически совпадает с теоретическим. Возможно это объясняется тем, что авторы, не занимавшиеся манипулированием индексом Хирша, когда у них было мало публикаций, не начинают заниматься этим и когда публикаций у них становится большое количество. Это повышает степень обоснованности и достоверности этих критериев.

6.3.5. Выводы и рекомендации

Итак, на основе вышеизложенного можно считать, что:

- 1) существует некое неизвестное «истинное значение индекса Хирша»;
- 2) есть «эмпирическое (классическое) значение индекса Хирша», которое является истинным значением, измененным в результате совместного действия факторов

манипулирования (рассматривались в данной статье) а также естественного шума и несовершенства алгоритма Хирша (в данном разделе эти факторы только упоминаются);

3) есть «теоретическое значение индекса Хирша», – это решение уравнения наилучшего тренда графика числа цитирований ранжированного списка публикаций.

«Теоретическое значение индекса Хирша» – это новое научное понятие из области наукометрии, которое авторы предлагают ввести в научный оборот и практику наукометрии по следующим причинам:

– теоретическое значение индекса Хирша является устойчивым к манипулированию и другим факторам, искажающим истинное значение индекса Хирша и может обоснованно считаться значительно более близким к истинному значению индекса Хирша, чем классическое эмпирическое значение;

– технология получения теоретического значения индекса Хирша (путем решение уравнения наилучшего тренда графика числа цитирований ранжированного списка публикаций) проста и доступна авторам и организациям.

В статье предлагаются два убедительных количественных частных критерия манипулирования индексом Хирша при малом числе статей и основанный на них аддитивный интегральный критерий, основанные на линейном тренде графика числа цитирований ранжированного списка публикаций.

Степень различия между эмпирическим и теоретическим значениями индекса Хирша можно считать устойчивым интегральным критерием манипулирования индексом Хирша при любом числе публикаций.

Предлагается:

1. Применить результаты данной статьи при расчетах в РИНЦ и строить рейтинги авторов, журналов и организаций (подразделений) не только на основе эмпирического классического индекса Хирша, но и на основе теоретического индекса Хирша, а также по критериям манипулирования.

2. Не придавать излишне и неоправданно большого значения классическому эмпирическому значению индекса Хирша при оценках и принятии решений.

Excel-файл с расчетами по описанным методикам по авторам статьи приведен по ссылке: <http://ej.kubagro.ru/2016/07/upload/05.zip>.

6.4. Синтез и верификация многокритериальной системно-когнитивной модели университетского рейтинга Гардиан и ее применение для сопоставимой оценки эффективности российских вузов с учетом направления подготовки

Раздел посвящен решению проблемы, заключающейся в том, что с одной стороны рейтинг российских вузов востребован, а с другой стороны пока он не создан. Предлагаемая идея решения проблемы состоит в применении отечественной лицензионной инновационной интеллектуальной технологии для этих целей: а именно предлагается применить автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальную систему «Эйдос». Эти методы подробно описываются в этом контексте. Предлагается рассмотреть возможности применения данного инструментария на примере университетского рейтинга Гардиан, и рассматриваются его частные критерии (показатели вузов). Указываются источники данных и методика их подготовки для обработки в системе «Эйдос». В соответствии с методологией АСК-анализа описывается установка системы «Эйдос», ввод исходных данных в нее и формализация предметной области, синтез и верификация модели, их отображение и применение для решения задач оценки рейтинга Гардиан.

ан для российских вузов и исследования объекта моделирования. Рассматриваются перспективы и пути создания интегрированного рейтинга российских вузов и эксплуатации рейтинга в адаптивном режиме. Указываются ограничения предлагаемого подхода и перспективы его развития

6.4.1. Формулировка проблемы

Университетские рейтинги давно стали общепринятым в мире методом оценки эффективности вузов⁵⁶.

Этими рейтингами для решения различных задач пользуются и потенциальные студенты, и их родители, и ученые, и руководители. Таким образом, они востребованы практически всем обществом.

Недавно и министерство образования и науки РФ обратилось к идее создания подобного рейтинга для российских вузов, и это в общем нельзя не приветствовать.

Однако первый опыт создания подобного рейтинга, по-видимому, приходится признать неудачным, т.к. он вызвал большой поток совершенно справедливой и хорошо обоснованной критики со стороны научно-педагогического сообщества. Возражения вызвали, прежде как сами критерии оценки эффективности вузов⁵⁷, так и полная непрозрачность процедуры формирования этих критериев, а также то, что за бортом широкого обсуждения (которого, вообще не было) осталось и само понятие эффективности вузов, т.е. их основное назначение. А ведь именно тем, что понимается под эффективностью вузов, определяются и критерии ее оценки. Но предложенные критерии оказались таковы, что у многих возникло вполне обоснованное подозрение, что под эффективностью вузов при их формировании понималось вовсе не качество образования, а нечто другое не свойственное вузам.

Эта критика звучит и на научных конференциях,⁵⁸ и в научных публикациях [1]. А то, о чем не принято говорить на научных конференциях и писать в научных публикациях, высказывается на форумах и на личных страницах ученых и педагогов. Например, на своем личном сайте доктор педагогических наук профессор А.А.Остапенко пишет: «Основных критериев, как мы помним пять: средний балл ЕГЭ принятых на обучение студентов; объём научных работ на одного сотрудника; количество иностранцев-выпускников; доходы вуза в расчёте на одного сотрудника, а также общая площадь учебно-лабораторных зданий в расчёте на одного студента. Как они связаны с эффективностью вуза и что такое эффективность вообще понять, мысля рационально, непросто. Даже всерьёз обсуждать эти критерии как-то странно»⁵⁹. Но мы все же выскажем одно соображение. На наш взгляд довольно странно выглядит попытка сравнения друг с другом вузов разных направленности подготовки, т.е. например аграрных вузов и вузов, готовящих специалистов для атомной и ракетно-космической промышленности. Иначе говоря, для вузов разной направленности должны быть свои рейтинги.

Правда со временем, наверное, в какой-то степени и под влиянием этой критики, позиция Минобрнауки РФ стала меняться. А то, что к тому времени уже успели закрыть несколько вузов, как говорят: «имеющих признаки неэффективности»⁶⁰, – это как бы и не так важно. Динамику этих изменений позиции профильного министерства

⁵⁶ См., например: <http://www.hotcourses.ru/study-in-the-uk/choosing-a-university/university-rankings-guide/>

⁵⁷ <http://yandex.ru/yandsearch?lr=35&text=критерии%20оценки%20эффективности%20вузов&lr=35>

⁵⁸ Достаточно сделать запрос: «научные конференции форумы по эффективности вузов»

⁵⁹ <http://ost101.livejournal.com/> <http://profdavidson.ucoz.ru/>

⁶⁰ На наш взгляд такие признаки имеют все вузы. Поэтому дело не в том, имеют они такие признаки или нет, а в том, на сколько эти признаки существенны в совокупности.

можно проследить по Нормативно-правовым документам Минобрнауки РФ, устанавливающим критерии оценки эффективности деятельности вузов⁶¹.

Таким образом, налицо **проблема**, которая состоит в том, что с одной стороны рейтинг российских вузов востребован, а с другой стороны как-то пока не очень получается его сформировать. То есть, как обычно желаемое не совпадает с действительным, и «хотели как лучше, а вышло как всегда» (В.С.Черномырдин).

6.4.2. Авторский подход к решению проблемы

6.4.2.1. ИДЕЯ ПРЕДЛАГАЕМОГО РЕШЕНИЯ ПРОБЛЕМЫ

Идея решения проблемы проста: обратиться к мировому опыту в этой области, творчески его переосмыслить применительно к российским реалиям и разработать свои научно-обоснованные подходы, с учетом всего лучшего, что есть в мировом опыте.

Существует несколько популярных и авторитетных рейтингов вузов¹:

- Университетский рейтинг The Guardian⁶²;
- Университетский рейтинг Times⁶³;
- Мировой рейтинг Times Higher Education⁶⁴;
- Рейтинг мировых вузов Шанхайского Университета⁶⁵.

Мы не будем их здесь описывать, т.к. по ним достаточно информации в общем доступе, в т.ч. по приведенным ссылкам.

Но хотели бы отметить, что для поддержки любого подобного рейтинга необходима соответствующая инфраструктура, оснащенная различными видами обеспечения ее деятельности (финансовое, кадровое, организационное, техническое, математическое, программное, информационное и т.д.). Все эти виды обеспечения в совокупности представляют собой технологию ведения и применения данного рейтинга.

Естественно, никто технологию не продает, а если и продает, то так дорого, что купить ее практически невозможно. Поэтому возникает вопрос о разработке или поиске подобной технологии в России.

Таким образом, востребованы теоретическое обоснование, математическая модель, методика численных расчетов (т.е. структуры данных и алгоритмы их обработки) а также реализующие их инструментальные (программные) средства, обеспечивающие создание, поддержку, развитие и применение подобных рейтингов.

Данная статья как раз и посвящена рассмотрению отечественной лицензионной инновационной интеллектуальной технологии, обеспечивающей решение поставленной проблемы. А именно предлагается применить для этой цели автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальную систему «Эйдос».

6.4.2.2. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ И ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС» КАК ИНСТРУМЕНТАРИЙ РЕШЕНИЯ ПРОБЛЕМЫ

Этот подход кратко описан в статье [2]. Здесь рассмотрим его подробнее.

Прежде всего, возникает вопрос о том, **что понимается под эффективностью вузов?** Ведь ясно, что прежде чем оценивать эффективность вузов было бы неплохо, а на самом деле совершенно необходимо, разобраться с тем, что же это такое. Причина этого ясна: выбор критериев оценки во многом обуславливается тем, что именно оценивается.

⁶¹ <http://uup.samgtu.ru/node/211>

⁶² <http://www.theguardian.com/education/table/2011/may/17/university-league-table-2012>

⁶³ http://extras.thetimes.co.uk/public/good_university_guide_landing?CMP=KNGvccp1-university%20rankings

⁶⁴ http://extras.thetimes.co.uk/public/good_university_guide_landing?CMP=KNGvccp1-university%20rankings

⁶⁵ http://www.educationindex.ru/article_ranking-shanghai-2014.aspx

Ясно, что по этому поводу существует много различных мнений, которые в различной степени аргументированы или не аргументированы и отражают позиции руководителей образования и науки, профессионального научно-педагогического сообщества и различных слоев населения. По мнению автора, с научной точки зрения некорректно и неуместно говорить о каких-то критериях оценки эффективности вузов, если не определено само это понятие эффективности, т.е. отсутствует консенсус в профессиональной среде по поводу того, что же это такое.

Очевидно, для достижения такого консенсуса в наше время необходимо широкое обсуждение этого вопроса в научной печати, Internet и СМИ. Однако такое обсуждение не было организовано и критерии оценки эффективности или признаков неэффективности практически *неожиданно* «свалились научно-педагогическому сообществу как снег на голову».

Уже после этого, как это произошло, началось обсуждение этого вопроса на различных научных конференциях, в научной и периодической прессе, на личных сайтах, формах и т.п. Но пока шло это обсуждение и пока оно не пришло к какому-либо консенсусу в этом вопросе, ряд вузов были закрыты, филиалы сокращены и т.д.

По мнению автора, цель вуза в том, чтобы формировать компетентных и творчески мыслящих специалистов в соответствии с прогнозом социального заказа, т.е. таких, которые будут востребованы обществом в будущем периоде профессиональной деятельности этих специалистов, который составляет 30-40 лет. А должен ли вуз зарабатывать, должен ли он иметь те или иные площади в расчете на одного учащегося – это все нужно знать только для того, чтобы спрогнозировать, сможет ли он выполнить свою основную задачу, т.е. подготовку специалистов. Ни в коем случае нельзя рассматривать эти показатели как самоцель, т.к. достижение тех или иных их значений, вообще говоря, может и ничего не говорить о достижении цели вуза. Несут ли эти критерии какую-либо информацию о достижении цели вуза, и какую именно по величине и знаку, – это еще надо определить в процессе специального исследования, которое, скорее всего не было проведено. Странно, что об этом приходится писать, но приходится, т.к. похоже, об этом стали забывать.

Когда консенсус профессионального научно-педагогического сообщества по вопросу о том, что же понимать под «эффективностью вуза» будет достигнут, на первый план выступает вопрос о том, *с помощью какого метода оценивать эту эффективность*, т.е. *как ее измерить*.

Для автора вполне очевидно, что этот метод должен представлять собой какой-то вариант метода многокритериальной оценки. Это обусловлено просто тем, что такие сложные и многофакторные системы как вузы в принципе невозможно оценивать по одному показателю или критерию. Чтобы обоснованно выбрать метод оценки эффективности вузов необходимо сначала научно обосновать требования к нему, а затем составить рейтинг методов по степени соответствия обоснованным требованиям и выбрать метод, наиболее удовлетворяющий обоснованным требованиям.

Применение метода факторного анализа для этих целей, по-видимому, некорректно, т.к. этот метод, предъявляющий настолько жесткие требования к исходным данным об объекте моделирования, что их практически невозможно выполнить. Во-первых, факторный анализ – это **параметрический** метод, предполагающий, что исходные данные подчиняются многомерным нормальным распределениям. Во-вторых, это метод **неустойчивый**, т.е. небольшие изменения исходных данных могут привести к значительным изменениям в модели. Поэтому исходные данные для факторного анализа должны быть абсолютно точными, что невозможно не только фактически, но даже в принципе. В-третьих, **перед** началом факторного анализа необходимо определить **наиболее важные факторы**, которые и будут исследоваться в создаваемой модели. Но

при этом в руководствах по факторному анализу не уточняется, каким способом это предлагается сделать. А между тем при большом количестве факторов, что является обычным для большинства реальных задач, это не тривиальная задача, которую вручную решить невозможно.

Когда метод оценки эффективности вузов выбран, необходимо ответить на вопрос о том, *на основе каких частных критериев оценивать эффективность вузов и какой исходной информацией о вузах для этого необходимо располагать?*

Ясно, что эти критерии в общем случае могут иметь как количественную, так и качественную природу и могут измеряться в различных единицах измерения. Кроме того эти критерии могут иметь различную силу и направление влияния на интегральную оценку эффективности вузов. Конечно, возникают вопросы как *о способе определения системы критериев эффективности вуза*, так и *о способе определения силы и направления влияния критериев на оценку эффективности вузов*.

Но еще более существенным является вопрос: *«О способе сопоставимого сведения разнородных по своей природе и измеряемых в различных единицах измерения частных критериев эффективности в один количественный интегральный критерий эффективности вуза»*.

Отметим, что в материалах Минобрнауки РФ и о критериях оценки эффективности вузов⁶⁶ даже не упоминается вопрос о том, что когда значения частных критериев для того или иного вуза установлены, то необходимо каким-то образом на их основе получить обобщающую количественную оценку его эффективности в виде **одного** числа, т.е. надо как-то объединить значения всех частных критериев в одной формуле, в одном математическом выражении, которое и называется «Интегральный критерий».

Поэтому, наверное, и говорят не об эффективности или неэффективности вуза, а всего лишь «о признаках неэффективности», а признаками являются значения отдельных частных критериев. Если таких признаков неэффективности много, то делают вывод о том, что вуз неэффективен. Фактически такой подход, который может быть и применялся, можно назвать неосознанным применением частных критериев и интегрального критерия, т.е. «неосознанным многокритериальным подходом». При таком подходе все частные критерии имеют одинаковый вес, например принимающий значения 0 (неэффективен) и 1 (эффективен). Когда значения всех частных критериев для вуза установлены, то эти веса суммируются и сумма сравнивается с минимальными и максимальными оценками, полученными для всех вузов. Допустим, в Минобрнауки РФ из каких-то своих соображений решили, что в результате оценки эффективности вузов должно быть закрыто из-за низкой эффективности 1.5% вузов. Тогда все вузы сортируются по убыванию этой суммы и 1.5% с конца рейтинга помещаются в «черный список».

Но такой «неосознанный многокритериальный подход» очень и очень уязвим для критики.

Во-первых, возникает законный вопрос о том, почему все критерии имеют одинаковый вес, хотя даже интуитивно ясно, что они имеют разное значение и по-разному влияют на эффективность вуза (которая, кстати, непонятно в чем заключается).

Во-вторых, непонятно, как можно складывать средний балл ЕГЭ принятых на обучение студентов, объем научных работ на одного сотрудника, количество иностранцев-выпускников, доходы вуза в расчёте на одного сотрудника и общую площадь учебно-лабораторных зданий в расчёте на одного студента. За подобные математические операции ставят двойку по физике в 7-м классе средней школы. Там школьников учат, что перед тем как складывать величины, измеренные в разных единицах измерения,

⁶⁶ См., например: <http://uup.samgtu.ru/node/211>

например рост учащихся, выраженный в метрах (1.72) и выраженный в сантиметрах (160), нужно перевести эти величины в одну единицу измерения, например в метры или в сантиметры. А иначе получится: $1.72+160=161.72$, т.е. некий результат, не поддающийся разумной содержательной интерпретации⁶⁷. Как бы нечто похожее и на таком же научном уровне не получилось при оценке наличия у вуза «признаков неэффективности». Но научно-педагогическую общественность не поставили в известность о том, каким образом вычисляется интегральная оценка эффективности вуза на основе установленных для него значений частных критериев. Поэтому высказанное опасение остается не снятым.

В развитом осознанном многокритериальном подходе для вычисления значения интегрального критерия нужно знать силу и направление влияния каждого значения частных критериев на величину этого интегрального критерия. Интегральные критерии бывают трех видов: аддитивные, мультипликативные и общего вида. Чаще всего используются аддитивные интегральные критерии, в которых значение интегрального критерия равно просто сумме значений частных критериев. Но чтобы значения частных критериев можно было корректно суммировать необходимо, чтобы они были значениями на числовых измерительных шкалах [3], и чтобы они измерялись в одних и тех же единицах измерения или были безразмерными.

Оба эти требования выполняются в Автоматизированном системно-когнитивном анализе (АСК-анализ), в котором все значения всех факторов, независимо от того количественные они или качественные и в каких единицах они измеряются в исходных данных, в моделях системы «Эйдос» (системно-когнитивных моделях) они все измеряются в одних и тех же единицах измерения – единицах количества информации [2, 3]. Поэтому метод АСК-анализа и предлагается для решения поставленной проблемы.

АСК-анализ представляет собой один из современных методов искусственно интеллекта, который предоставляет научно обоснованные ответы на все эти вопросы, но самое существенное, что он оснащен широко и успешно апробированным универсальным программным инструментарием, позволяющим решить эти вопросы не только как обычно на теоретическом концептуальном уровне, но и на практике [2]. Модели знаний АСК-анализа основаны на нечеткой декларативной модели представления знаний, предложенной автором в 1983 году и являющейся гибридной моделью, сочетающей в себе преимущества фреймовой, нейросетевой и четкой продукционной моделей и обеспечивающей создание моделей очень больших размерностей до 10 млн. раз превышающих максимальные размерности моделей знаний экспертных систем с четкими продуктами:

- от фреймовой модели модель представления знания системы «Эйдос» отличается существенно упрощенной программной реализацией и более высоким быстродействием без потери функциональности;
- от нейросетевой тем, что обеспечивает хорошо обоснованную теоретически содержательную интерпретацию весовых коэффициентов на рецепторах и обучение методом прямого счета [8];
- от четкой продукционной модели – нечеткими продуктами, представленными в декларативной форме, что обеспечивает эффективное использование знаний без их многократной генерации для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта.

АСК-анализ является непараметрическим методом, устойчивым к шуму в исходных данных, позволяющий корректно обрабатывать неполные (фрагментирован-

⁶⁷ Проще говоря «ерундой».

ные) исходные данные, описывающие воздействие взаимозависимых факторов на нелинейный [7] объект моделирования.

Суть метода АСК-анализа в том, что он позволяет рассчитать на основе исходных данных какое количество информации содержится в значениях факторов, обуславливающих переходы объекта моделирования в различные будущие состояния, причем как в желательные, так и в нежелательные [3].

Он состоит в целенаправленном *последовательном повышении степени формализации* исходных данных до уровня, который позволяет ввести исходные данные в компьютерную систему, а затем преобразовать исходные данные в информацию; информацию преобразовать в знания; использовать знания для решения задач прогнозирования, принятия решений и исследования предметной области.

Рассмотрим подробнее вопросы выявления, представления и использования знаний в АСК-анализе и системе «Эйдос».

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ), развиваемой проф. Е.В.Луценко, для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на корпорацию к каким ее изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу.

Знания – это информация, полезная для достижения целей.

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).
2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно **последовательно повышать степень формализации** исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач управления, принятия решений и исследования предметной области.

АСК-анализ имеет следующие этапы [2]:

- когнитивно-целевая структуризация предметной области;
- формализация предметной области (формирование классификационных и описательных шкал и градаций и обучающей выборки);
- синтез и верификация статистических и системно-когнитивных моделей;
- решение задач идентификации, прогнозирования, принятия решений и исследования предметной области в наиболее достоверных из созданных моделей.

Единственный неавтоматизированный в системе «Эйдос» этап – это первый, а остальные приведены на рисунке 1.

АСК-анализ имеет ряд особенностей, которые обусловили его выбор в качестве метода решения проблемы:

9. Имеет *теоретическое обоснование*, основой которого является *семантическая мера целесообразности информации А.Харкевича*.

10. Обеспечивает *корректную сопоставимую количественную* обработку *разнородных* по своей природе факторов, измеряемых в *различных единицах* измерения, *высокую точность* и независимость результатов расчетов от единиц измерения исходных данных.

11. Обеспечивает построение *многомерных моделей* объекта моделирования непосредственно на основе *неполных и искаженных* эмпирических данных о нем.

12. Имеет развитую и *доступную программную реализацию* в виде универсальной когнитивной аналитической системы «Эйдос».

Очень важно, что этот инструментарий и методики его использования для решения сформулированных задач могут быть доступны всем заинтересованным сторонам не только на федеральном уровне, но и в самих вузах, что позволит им осуществлять аудиторскую самооценку и видеть свое место и динамику среди других вузов. Это позволит руководителям вузов принимать более осознанные и научно обоснованные решения, направленные на повышение эффективности и рейтинга их вуза. Конечно, для реализации на практике регулярного рейтингового анализа вузов необходимо создание соответствующей достаточно разветвленной инфраструктуры.

Более подробному и конкретному исследованию связанных с этим вопросов и посвящена данная работа, в которой далее кратко расстраивается университетский рейтинг Гардиан (который выбран просто в качестве примера), а затем приводится численный пример его реализации в форме приложения интеллектуальной системы «Эйдос». Отметим, что создание этого приложения *не требует программирования* [4-6], т.е. система «Эйдос» анализирует исходные данные рейтинга и строит модель, в которой отражено как влияют значения частных критериев на значение интегрального критерия, т.е. на итоговую общую оценку рейтинга вуза.

6.4.2.3. ЧАСТНЫЕ КРИТЕРИИ УНИВЕРСИТЕТСКОГО РЕЙТИНГА ГАРДИАН

Университетский рейтинг Гардиан⁶⁸ выгодно отличается от других тем, что измеряет качество преподавания, использования учебных ресурсов, а также оценивает уровень исследовательской деятельности, что очень полезно для тех, кто интересуется послевузовскими программами – магистратурой, докторантурой и проч.

Как указано на официальном сайте рейтинга¹⁰ в нем используются следующие **частные критерии**:

1. Качество преподавания, которое оценивается национальным студенческим исследованием (NSS): процент удовлетворенных студентов.
2. Получение обратной связи от преподавателя и качество заданий. Оценивается опросом NSS, в котором устанавливается процент удовлетворенных студентов.
3. Результаты опроса NSS, в котором оценивается процент студентов, удовлетворенных общим качеством выбранной программы.
4. Затраты на студента – оценка по 10-балльной шкале.
5. Соотношение студент – работник вуза: количество студентов на штатную единицу университета.
6. Карьерные перспективы: процент выпускников, сумевших найти работу или продолжить обучение в течение полугода после окончания вуза.
7. Уровень прогресса студентов на основе сравнения университетских результатов с оценками предыдущего сертификата (обычно, школьного или университетского): оценка по 10-балльной шкале. Данный показатель демонстрирует, насколько преподавательский состав способен повлиять на улучшение успеваемости студентов.
8. Проходной балл при поступлении в вуз на основе оценок предыдущего сертификата обучения (школьный или университетский сертификат).

⁶⁸ <http://www.theguardian.com/education/table/2011/may/17/university-league-table-2012>

Отметим, что считаем важным достоинством данного рейтинга то, что он ведется по различным направлениям подготовки, которых 45 (таблица 1):

Таблица 1 – Направления подготовки, по которым проводился университетский рейтинг Гардиан

№	Наименование	24	Engineering: mechanical
1	Agriculture, forestry and food	25	English
2	American studies	26	Geography and environmental studies
3	Anatomy and physiology	27	History and history of art
4	Anthropology	28	Law
5	Archaeology and Forensics	29	Mathematics
6	Architecture	30	Media studies, communications and librarianship
7	Art and design	31	Medicine
8	Biosciences	32	Modern languages and linguistics
9	Building and town and country planning	33	Music
10	Business and management studies	34	Nursing and paramedical studies
11	Chemistry	35	Pharmacy and pharmacology
12	Classics	36	Philosophy
13	Computer sciences and IT	37	Physics
14	Dentistry	38	Politics
15	Drama and dance	39	Psychology
16	Earth and marine sciences	40	Religious studies and theology
17	Economics	41	Social policy and administration
18	Education	42	Sociology
19	Engineering: chemical	43	Sports science
20	Engineering: civil	44	Tourism, transport and travel
21	Engineering: electronic and electrical	45	Veterinary science
22	Engineering: general		
23	Engineering: materials and mineral		

В университетском рейтинге Гардиан содержатся рейтинги следующих 155 вузов (таблица 2):

Таблица 2 – Вузы, по которым есть информация в университетском рейтинге Гардиан

№	Наименование	29	Chichester
1	Aberdeen	30	City
2	Abertay Dundee	31	Conservatoire for Dance and Drama
3	Aberystwyth	32	Courtauld Institute
4	Anglia Ruskin	33	Coventry
5	Arts UC, Bournemouth	34	Cumbria
6	Aston	35	De Montfort
7	Bangor	36	Derby
8	Bath	37	Dundee
9	Bath Spa	38	Durham
10	Bedfordshire	39	East London
11	Birmingham	40	Edge Hill
12	Birmingham City	41	Edinburgh
13	Bishop Grosseteste UC	42	Edinburgh College of Art
14	Bolton	43	Edinburgh Napier
15	Bournemouth	44	Edinburgh School of Architecture
16	Bradford	45	Essex
17	Brighton	46	Exeter
18	Brighton Sussex Medical School	47	Glamorgan
19	Bristol	48	Glasgow
20	Brunel	49	Glasgow Caledonian
21	Buckingham	50	Glasgow School of Art
22	Bucks New University	51	Gloucestershire
23	Cambridge	52	Glyndwr
24	Canterbury Christ Church	53	Goldsmiths
25	Cardiff	54	Greenwich
26	Central Lancashire	55	Guildhall School of Music and Drama
27	Central School of Speech and Drama	56	Harper Adams UC
28	Chester	57	Heriot-Watt

58	Hertfordshire
59	Heythrop College
60	Huddersfield
61	Hull
62	Hull York Medical School
63	Imperial College
64	Keele
65	Kent
66	King's College London
67	Kingston
68	Lancaster
69	Leeds
70	Leeds College of Music
71	Leeds Met
72	Leeds Trinity University College
73	Leicester
74	Lincoln
75	Liverpool
76	Liverpool John Moores
77	London Met
78	London School of Economics
79	London South Bank
80	Loughborough
81	Manchester
82	Manchester Met
83	Manchester School of Architecture
84	Marjon (St Mark and St John)
85	Middlesex
86	Newcastle
87	Newman University College
88	Newport
89	Northampton
90	Northumbria
91	Norwich UC of the Arts
92	Nottingham
93	Nottingham Trent
94	Oxford
95	Oxford Brookes
96	Peninsula Medical School
97	Plymouth
98	Portsmouth
99	Queen's, Belfast
100	Queen Margaret
101	Queen Mary
102	Ravensbourne
103	Reading
104	Robert Gordon
105	Roehampton
106	Rose Bruford College
107	Royal Academy of Music
108	Royal Agricultural College

109	Royal College of Music
110	Royal Holloway
111	Royal Northern College of Music
112	Royal Scottish Academy of Music and Drama
113	Royal Veterinary College
114	Salford
115	School of Pharmacy
116	Sheffield
117	Sheffield Hallam
118	SOAS
119	Southampton
120	Southampton Solent
121	St Andrews
122	St George's Medical School
123	St Mary's UC, Belfast
124	St Mary's UC, Twickenham
125	Staffordshire
126	Stirling
127	Stranmillis UC
128	Strathclyde
129	Sunderland
130	Surrey
131	Sussex
132	Swansea
133	Teesside
134	Thames Valley
135	The Liverpool Institute for Performing Arts
136	Trinity Laban Conservatoire
137	Trinity Saint David
138	UC Falmouth
139	UC Suffolk
140	UCL
141	UEA
142	Ulster
143	University College Birmingham
144	University for the Creative Arts
145	University of the Arts, London
146	UWE Bristol
147	UWIC
148	Warwick
149	West of Scotland
150	Westminster
151	Winchester
152	Worcester
153	Writtle College
154	York
155	York St John

Однако интегральный критерий, позволяющий получить рейтинговую оценку вуза на основе установленных для него значений частных критериев, на официальном сайте рейтинга Гардиан⁶⁹ не приводится. Поэтому для того, чтобы применить данный рейтинг на практике необходимо реконструировать его интегральный критерий и создать модель, отражающую силу и знак связи между значениями частных критериев и значениями интегрального критерия. Решим эту задачу в системе «Эйдос» на численном примере на основе реальных данных рейтинга Гардиан.

⁶⁹ <http://www.theguardian.com/education/table/2011/may/17/university-league-table-2012>

6.4.3. Численный пример

6.4.3.1. Источники исходных данных

В нижней части одной из страниц официального сайта университетского рейтинга Гардиан⁷⁰ есть ссылка на Excel-таблицу, которую мы использовали в качестве исходных данных:

Download the data

- [DATA: download the full spreadsheet.](#)

Кликнув по этой ссылке, мы получаем on-line доступ к этой Excel-таблице (рисунки 2):

Unit group	2012 max	2009/10 full-time rate diff	2008/09 rate diff	Average Teaching Score	NSS Teaching (%)	NSS Overall (%)	Expenditure per student/10	Student:staff ratio	Career prospects (%)	Value added score/10	Entry Tariff	NSS Feedback (%)	
Millennium	47	8,300	8,730	8.9	47.4	77	78	7.7	22.7	58	6.2	252	
Millennium	70	8,300	7,650	13.3	59.7	83	81	6.7	15.8	51	6.2	244	
University Alliance	74	8,500	15,340	7.6	49.9	85	84	4.9	20.5	52	3.3	202	
Gold HE	46	8,100	4,141	8.6	44.2	81	78	4.4	21.9	53	6.7	262	
Millennium	56	9,000	12,655	8.9	54.5	80	78	3.3	15.7	56	7.2	294	
University Alliance	82	8,500	16,460	8.1	46.3	82	82	4.6	20.3	65	6.0	238	
University Alliance	89	8,500	6,938	8.5	53.1	81	82	4.1	20.7	51	6.1	266	
University Alliance	89	8,500	16,460	8.1	46.3	82	79	4.4	19.8	57	7.2	286	
Microsoft Excel (xlsx)	8,500	6,355	9.4	54.5	85	82	5.2	22.9	49	6.5	236	70	
OpenDocument Format (ods)	8,000	8,885	9.3	58.5	83	78	3.3	17.2	58	4.8	270	68	
OpenDocument Format (ods)	8,000	7,185	9.2	52.2	82	82	3.9	17.5	57	5.3	286	72	
Document PDF	8,000	12,233	10.7	47.3	80	82	3.9	17.5	47	4.1	256	68	
OpenDocument Format (ods)	8,000	10,910	5.3	47.1	84	84	3.9	16.5	56	4.3	311	64	
OpenDocument Format (ods)	8,000	3,161	11.5	39.1	79	72	3.4	22.5	51	3.0	249	49	
OpenDocument Format (ods)	8,000	5,257	14.5	59.2	73	67	7.7	16.7	69	4.4	217	62	
OpenDocument Format (ods)	8,000	12,341	7.1	41.7	80	78	3.5	23.8	51	4.3	279	64	
OpenDocument Format (ods)	8,000	3,088	6.7	59.2	87	84	3.3	19.6	51	4.4	286	70	
OpenDocument Format (ods)	8,000	5,244	9.1	47.5	81	80	4.5	19.9	48	7.2	281	62	
OpenDocument Format (ods)	8,000	7,950	10,564	12.4	58.8	81	78	4.4	18.4	64	5.6	266	72
OpenDocument Format (ods)	8,000	7,280	7.9	42.1	84	82	3.5	18.6	49	4.4	246	64	
OpenDocument Format (ods)	8,000	7,902	12.0	44.5	81	76	6.2	19.1	57	4.1	187	64	
OpenDocument Format (ods)	8,000	10,770	15.4	42.3	79	75	3.2	20.5	52	4.8	270	63	
OpenDocument Format (ods)	8,000	1,872	10.0	41.7	86	83	2.2	26.2	59	4.1	237	87	
OpenDocument Format (ods)	8,000	3,075	7.7	53.4	83	79	2.6	17.7	49	6.1	274	67	
OpenDocument Format (ods)	8,000	4,720	10.4	48.7	78	76	5.6	15.3	43	4.3	209	65	
OpenDocument Format (ods)	8,000	16,302	8.6	44.5	78	77	4.0	20.6	51	4.4	267	62	
OpenDocument Format (ods)	8,000	9,297	11.3	42.3	82	78	4.1	20.5	46	4.9	247	67	
OpenDocument Format (ods)	8,000	5,375	8.9	52.9	83	82	4.0	18.5	54	8.0	242	65	
OpenDocument Format (ods)	8,000	9,297	5.3	56.3	84	83	3.4	19.7	53	6.0	288	68	
OpenDocument Format (ods)	8,000	11,813	13.3	49.9	75	73	4.0	17.1	49	7.5	271	68	
OpenDocument Format (ods)	8,000	2,078	12.4	44.5	81	74	3.6	22.5	62	4.3	219	69	
OpenDocument Format (ods)	8,000	6,653	12.7	58.9	86	83	3.3	18.5	59	3.4	258	74	
OpenDocument Format (ods)	8,000	3,257	7.1	43.2	82	79	2.7	19.4	55	4.6	242	68	
OpenDocument Format (ods)	8,000	3,257	7.1	43.2	82	79	3.4	18.8	54	5.4	258	68	
OpenDocument Format (ods)	8,000	3,881	0.0	45.5	85	81	5.8	21.2	48	4.0	200	72	
OpenDocument Format (ods)	8,000	15,271	16.5	52.2	79	78	7.2	21.0	51	6.4	236	66	
OpenDocument Format (ods)	8,000	14,889	11.1	48.6	77	77	4.0	20.9	48	5.7	258	61	

Рисунок 2 – Excel-таблица исходных данных по университетскому рейтингу Гардиан с официального сайта рейтинга (фрагмент)

Чтобы скачать эту таблицу на локальном компьютере нужно кликнуть слева вверху по пункту меню «Файл», а затем выбрать: «Сохранить как» и указать тип файла.

6.4.3.2. ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ ДЛЯ СИСТЕМЫ «ЭЙДОС»

Однако в соответствии с 1-м и единственным не автоматизированным в системе «Эйдос» этапом АСК-анализа, который называется: «Когнитивно-целевая структуризация предметной области» перед созданием интеллектуального приложения мы должны определить, что мы хотим определить с помощью модели и на основе чего.

В данной задаче для каждого университета по значениям его показателей мы бы хотели определить:

- обобщающий рейтинг Гардиан (**Guardian score/100**);
- рейтинг по каждому из направлений подготовки (**Rank**), перечисленных в таблице 1;
- основное (профилирующее) направление подготовки (**Field of study**).
- само наименование университета (**Name of Institution**).

Наименования показателей университета:

1. % Satisfied with Teaching.
2. % Satisfied overall with course.
3. Expenditure per student (FTE).
4. Student:staff ratio.
5. Career prospects.
6. Value added score/10.

⁷⁰ <http://www.theguardian.com/news/datablog/2011/may/17/university-guide-2012-data-guardian>

7. Average Entry Tariff.
8. % Satisfied with Assessment.

Перевод этих показателей на русский язык приведен в разделе 2.3.

Учитывая эти результаты выполнения 1-го этапа АСК-анализа, перед для вводом данных в систему «Эйдос», таблицу, скачанную на предыдущем шаге с официального сайта рейтинга Гардиан и приведенную на рисунке 2, необходимо преобразовать в такую форму, которая бы отражала те способы группировки данных по университетам, которые нас интересуют и соответствовала бы требованиям системы «Эйдос» к внешним базам исходных данных (рисунок 3):



Рисунок 3 – Экранная форма системы «Эйдос» с описанием требований к внешним базам исходных данных

Для этого преобразуем таблицу, приведенную на рисунке 2, следующим образом:

1. Добавим лист (вкладку) с наименованием: «Inp_data» на 1-ю позицию. На этом листе будет формироваться результат для ввода данных всех данных по рейтингу в систему «Эйдос».

2. Переименуем наименования всех вкладок с рейтингами по направлений подготовки, полностью убирая текстовое наименование направления подготовки и оставляя только его номер (код). Это нужно для того, чтобы проще было писать формулы со ссылками на листы с информацией о рейтингах по направлениям подготовки.

3. Добавим отладочную страницу «Р», на которой апробируем способ отображения абсолютного рейтинга в относительный (нормированный). Дело в том, что в таблице на рисунке 2 в каждом рейтинге по направлению подготовки участвует разное число университетов, а рейтингом является просто порядковый номер в списке. В результате рейтинги по направлениям подготовки изменяются в различных пределах от 1 до числа университетов, имеющих данное направление подготовки. В результате такие рейтинги оказываются **несопоставимыми**, что нас не устраивает. Чтобы преодолеть эту проблему мы нормировали абсолютные рейтинги по направлениям подготовки к 10-бальной числовой шкале, т.е. преобразовали их в относительные. Можно было взять и любое другое число градаций шкалы, но мы посчитали, что такая шкала обеспечивает необходимую и достаточную для практики точность. Кроме того этот лист мы затем используем для модификации листов с рейтингами по направлениям подготовки. В таблице 3

приведены результат нормирования абсолютного рейтинга с 27 градациями и формулы, с помощью которых это делается.

Таблица 3 – Способ и результат нормирования абсолютного рейтинга по направлению подготовки

Результат нормирования абсолютного рейтинга

	A	B	C
1	X1=	1	
2	X2=	27	
3	Y1=	1	
4	Y2=	10	
5			
6	Абсолют. рейтинг		Относит. рейтинг
7	1		1,00
8	2		1,35
9	3		1,69
10	4		2,04
11	5		2,38
12	6		2,73
13	7		3,08
14	8		3,42
15	9		3,77
16	10		4,12
17	11		4,46
18	12		4,81
19	13		5,15
20	14		5,50
21	15		5,85
22	16		6,19
23	17		6,54
24	18		6,88
25	19		7,23
26	20		7,58
27	21		7,92
28	22		8,27
29	23		8,62
30	24		8,96
31	25		9,31
32	26		9,65
33	27		10,00

Способ (формулы) нормирования абсолютного рейтинга

A	B	C
X1=	=A7	
X2=	=A33	
Y1=	1	
Y2=	10	
Абсолютный рейтинг		Относительный рейтинг
1		$=\frac{B3+(A7-B1)}{(B2-B1)}*(B4-B3)$
2		$=\frac{B3+(A8-B1)}{(B2-B1)}*(B4-B3)$
3		$=\frac{B3+(A9-B1)}{(B2-B1)}*(B4-B3)$
4		$=\frac{B3+(A10-B1)}{(B2-B1)}*(B4-B3)$
5		$=\frac{B3+(A11-B1)}{(B2-B1)}*(B4-B3)$
6		$=\frac{B3+(A12-B1)}{(B2-B1)}*(B4-B3)$
7		$=\frac{B3+(A13-B1)}{(B2-B1)}*(B4-B3)$
8		$=\frac{B3+(A14-B1)}{(B2-B1)}*(B4-B3)$
9		$=\frac{B3+(A15-B1)}{(B2-B1)}*(B4-B3)$
10		$=\frac{B3+(A16-B1)}{(B2-B1)}*(B4-B3)$
11		$=\frac{B3+(A17-B1)}{(B2-B1)}*(B4-B3)$
12		$=\frac{B3+(A18-B1)}{(B2-B1)}*(B4-B3)$
13		$=\frac{B3+(A19-B1)}{(B2-B1)}*(B4-B3)$
14		$=\frac{B3+(A20-B1)}{(B2-B1)}*(B4-B3)$
15		$=\frac{B3+(A21-B1)}{(B2-B1)}*(B4-B3)$
16		$=\frac{B3+(A22-B1)}{(B2-B1)}*(B4-B3)$
17		$=\frac{B3+(A23-B1)}{(B2-B1)}*(B4-B3)$
18		$=\frac{B3+(A24-B1)}{(B2-B1)}*(B4-B3)$
19		$=\frac{B3+(A25-B1)}{(B2-B1)}*(B4-B3)$
20		$=\frac{B3+(A26-B1)}{(B2-B1)}*(B4-B3)$
21		$=\frac{B3+(A27-B1)}{(B2-B1)}*(B4-B3)$
22		$=\frac{B3+(A28-B1)}{(B2-B1)}*(B4-B3)$
23		$=\frac{B3+(A29-B1)}{(B2-B1)}*(B4-B3)$
24		$=\frac{B3+(A30-B1)}{(B2-B1)}*(B4-B3)$
25		$=\frac{B3+(A31-B1)}{(B2-B1)}*(B4-B3)$
26		$=\frac{B3+(A32-B1)}{(B2-B1)}*(B4-B3)$
27		$=\frac{B3+(A33-B1)}{(B2-B1)}*(B4-B3)$

Получим выражение для линейного отображения абсолютной шкалы, с числом градаций X_2 в относительную шкалу, с заданным числом градаций Y_2 (рисунок 4):

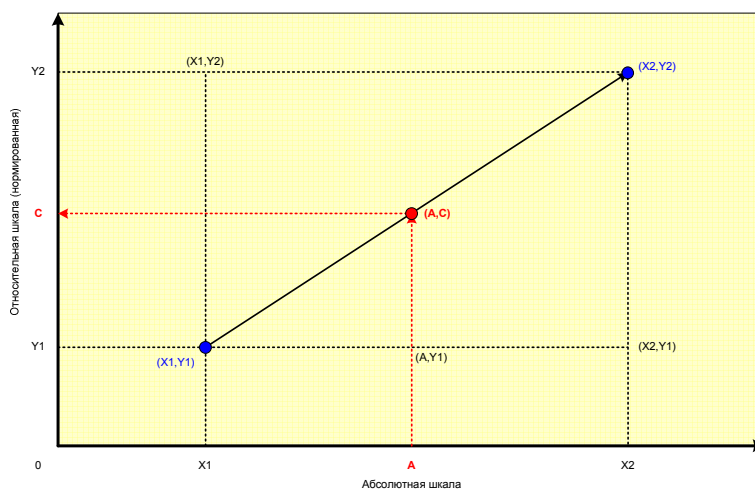


Рисунок 4 – К выводу выражения для линейного отображения абсолютной шкалы в относительную шкалу (линейная нормировка абсолютной шкалы)

Из рисунка 4 видно, что:

$$\frac{C - Y_1}{A - X_1} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

Откуда получаем искомое выражение для нормировки:

$$C = Y_1 + \frac{A - X_1}{X_2 - X_1} (Y_2 - Y_1)$$

где:

C – относительный рейтинг;

A – абсолютный рейтинг;

Y_1 – значение начальной градации шкалы относительного рейтинга;

Y_2 – значение конечной градации шкалы относительного рейтинга (число градаций, если $Y_1 = 1$);

X_1 – значение начальной градации шкалы абсолютного рейтинга;

X_2 – значение конечной градации шкалы абсолютного рейтинга (число градаций, если $X_1 = 1$).

Этому выражению можно придать вид линейного уравнения, но нам в этом нет необходимости.

4. На следующем шаге:

– копируем таблицу для нормировки абсолютных шкал на все листы с рейтингами по направлениям подготовки;

– корректируем значение X_2 на фактическое в данной абсолютной шкале.

В результате и получаем такие листы (таблица 4):

Таблица 4 – Преобразование абсолютного рейтинга по направлению подготовки: «Медицина» в относительный (нормированный)

Medicine										X1=1		
Rank	Name of Institution	Guardian score/100	% Satisfied with Teaching	% Satisfied overall with course	Expenditure per student (FTE)	Student:staff ratio	Career prospects	Value added score/10	Average Entry Tariff	% Satisfied with Assessment		Rank
1	Oxford	100,0	99	96	7,0	100	7	552	83			
2	Cambridge	92,8	92	92	10	6,7	98	2	578	59	1,00	1
3	Edinburgh	90,3	94	92	10	5,7	99	4	531	52	1,31	1
4	Dundee	86,2	95	94	10	5,1	100	6	498	64	1,62	2
5	UCL	80,7	91	91	6	4,8	100	9	532	58	1,93	2
6	Imperial College	67,1	92	90	7	6,9	100	3	507	44	2,24	2
7	Leicester	65,2	92	90	5	6,4	99	5	497	50	2,55	3
8	Newcastle	64,0	96	98	4	7,7	100	5	488	55	2,86	3
8	Peninsula Medical School	64,0	92	89	7	9,1	100	4	487	62	3,17	3
10	Nottingham	63,7	92	89	3	7,0	99	4	518	48	3,17	3
11	King's College London	62,3	89	85	5	4,9	100	6	486	43	3,79	4
12	Warwick	61,2	90	85	8	8,7	100	9	49	42	4,10	4
13	Leeds	55,2	88	85	6	8,5	100	6	490	50	4,41	4
14	Hull York Medical School	54,2	89	87	4	9,9	100	8	492	54	4,72	5
15	Manchester	53,8	80	67	8	8,7	99	9	496	35	5,03	5
15	Sheffield	53,8	91	89	3	8,5	100	4	495	55	5,34	5
17	Aberdeen	52,0	92	93	4	9,0	99	6	444	66	5,34	5
18	Brighton Sussex Medical School	51,9	90	86	4	12,4	100	7	55	55	5,97	6
18	Queen Mary	51,9	82	77	4	9,0	100	8	499	44	6,28	6
20	St George's Medical School	50,7	89	90	6	10,9	99	6	473	54	6,28	6
21	Southampton	49,5	67	63	4	6,8	99	7	404	43	6,90	7
22	St Andrews	48,8	96	97	2	12,7	96	1	504	59	7,21	7
23	Glasgow	47,5	79	65	5	7,6	99	5	510	33	7,52	8
24	UEA	45,0	90	87	3	8,6	100	9	438	54	7,83	8
25	Birmingham	44,2	91	83	3	10,1	99	3	516	30	8,14	8
26	Queen's, Belfast	42,3	89	86	4	10,1	100	3	460	50	8,45	8
27	Liverpool	38,9	75	66	4	8,9	100	4	485	31	8,76	9
28	Bristol	38,0	85	68	6	11,0	100	4	496	24	9,07	9
29	Keele	32,0	84	62	3	10,1	100	7	450	27	9,38	9
30	Cardiff	28,8	79	58	3	11,1	100	8	487	19	9,69	10
											10,00	10

Отметим, что значение $Y_2=10$ во всех листах берется с листа «Р» с исходной таблицей для преобразования абсолютных шкал в относительные, и, если его изменить там, то оно сразу меняется на всех листах с рейтингами по направлениям подготовки.

5. Затем формируем лист для ввода данных в систему «Эйдос». Для этого мы собираем на одном листе данные со всех листов с рейтингами по направлениям подготовки (таблица 5).

Таблица 5 – Исходные данные по рейтингу Гардиан, подготовленные для ввода в систему «Эйдос» (фрагмент)

The object of training sample	Guardian score/100	Rank	Field of study	Name of Institution	% Satisfied with Teaching	% Satisfied overall with course	Expenditure per student (fTE)	Student:staff ratio	Career prospects	Value added score/10	Average Entry Tariff	% Satisfied with Assessment
Medicine-Oxford, 2012	100.00	01 Medicine	Medicine	Oxford	98.51	96.00	7.00	100.00	7.00	652.00	7.00	552.00
Medicine-Cambridge, 2012	99.00	01 Medicine	Medicine	Cambridge	92.44	92.00	10.00	67.00	98.00	2.00	678.00	59.38
Medicine-Edinburgh, 2012	90.30	02 Medicine	Medicine	Edinburgh	93.92	92.00	10.00	5.70	99.00	4.00	531.00	51.79
Medicine-Dundee, 2012	86.20	02 Medicine	Medicine	Dundee	95.25	94.00	10.00	5.10	100.00	6.00	488.00	64.40
Medicine-UCL, 2012	80.70	02 Medicine	Medicine	UCL	91.04	91.00	6.00	4.80	100.00	9.00	532.00	58.43
Medicine-Imperial College, 2012	87.10	03 Medicine	Medicine	Imperial College	82.20	86.00	7.00	0.90	100.00	3.00	507.00	45.21
Medicine-Leicester, 2012	85.20	03 Medicine	Medicine	Leicester	91.79	90.00	5.00	6.40	99.00	5.00	497.00	50.10
Medicine-Newcastle, 2012	64.00	03 Medicine	Medicine	Newcastle	95.72	97.98	4.00	7.70	100.00	5.00	488.00	54.71
Medicine-Penninsula Medical School, 2012	64.00	03 Medicine	Medicine	Penninsula Medical School	91.78	89.08	7.00	9.10	100.00	4.00	487.00	61.91
Medicine-Nottingham, 2012	63.70	04 Medicine	Medicine	Nottingham	91.30	89.03	3.00	7.00	99.03	4.00	319.00	49.19
Medicine-King's College London, 2012	62.30	04 Medicine	Medicine	King's College London	88.75	85.00	5.00	4.90	100.00	6.00	486.00	42.60
Medicine-Warwick, 2012	61.20	04 Medicine	Medicine	Warwick	89.55	84.85	8.00	8.70	100.00	9.00	0.00	42.28
Medicine-Leeds, 2012	55.20	05 Medicine	Medicine	Leeds	87.58	85.00	6.00	8.50	100.00	6.00	480.00	53.26
Medicine-Hull York Medical School, 2012	54.20	05 Medicine	Medicine	Hull York Medical School	86.69	87.00	4.00	9.90	100.00	8.00	492.00	53.68
Medicine-Manchester, 2012	53.80	05 Medicine	Medicine	Manchester	79.95	67.00	8.00	8.70	99.00	9.00	496.00	35.48
Medicine-Sheffield, 2012	53.80	05 Medicine	Medicine	Sheffield	91.29	89.11	3.00	8.50	100.00	4.00	485.00	55.40
Medicine-Aberdeen, 2012	52.90	06 Medicine	Medicine	Aberdeen	91.71	93.11	4.00	9.00	99.03	6.00	444.00	68.19
Medicine-Brighton Sussex Medical School, 2012	51.90	06 Medicine	Medicine	Brighton Sussex Medical School	89.97	86.00	4.00	12.40	100.00	7.00	0.00	55.49
Medicine-Queen Mary, 2012	51.90	06 Medicine	Medicine	Queen Mary	82.41	77.00	4.00	9.00	100.00	8.00	499.00	44.14
Medicine-St George's Medical School, 2012	50.70	07 Medicine	Medicine	St George's Medical School	88.94	86.90	6.00	10.90	99.00	6.00	473.00	63.82
Medicine-Southampton, 2012	49.60	07 Medicine	Medicine	Southampton	87.00	82.83	4.00	8.90	99.00	7.00	484.00	43.17
Medicine-St Andrews, 2012	49.60	08 Medicine	Medicine	St Andrews	95.98	97.00	2.00	12.70	96.00	1.00	504.00	55.72
Medicine-Glasgow, 2012	47.60	08 Medicine	Medicine	Glasgow	79.86	84.65	5.00	7.60	99.00	5.00	510.00	32.80
Medicine-UEA, 2012	45.00	08 Medicine	Medicine	UEA	89.50	87.00	3.00	8.60	100.00	9.00	438.00	53.60
Medicine-Birmingham, 2012	44.20	08 Medicine	Medicine	Birmingham	90.99	83.00	3.00	10.10	99.00	3.00	516.00	29.94
Medicine-Queen's, Belfast, 2012	42.30	09 Medicine	Medicine	Queen's, Belfast	89.86	86.00	4.00	10.00	100.00	3.00	480.00	59.10
Medicine-Liverpool, 2012	38.90	09 Medicine	Medicine	Liverpool	75.31	65.66	4.00	8.90	100.00	4.00	485.00	31.14
Medicine-Bristol, 2012	38.00	09 Medicine	Medicine	Bristol	85.28	88.00	6.00	11.00	100.00	4.00	486.00	24.40
Medicine-Keel, 2012	29.20	10 Medicine	Medicine	Keel	92.32	83.07	10.10	9.00	99.00	7.00	470.00	65.47
Medicine-Cardiff, 2012	28.90	10 Medicine	Medicine	Cardiff	78.55	58.00	3.00	11.10	100.00	8.00	487.00	18.88
Dentistry-King's College London, 2012	100.00	01 Dentistry	Dentistry	King's College London	90.78	86.14	8.00	7.10	99.00	7.00	484.00	65.67
Dentistry-Glasgow, 2012	84.10	02 Dentistry	Dentistry	Glasgow	97.25	95.03	3.00	8.60	100.00	6.00	463.00	74.40
Dentistry-Cardiff, 2012	81.30	03 Dentistry	Dentistry	Cardiff	94.01	97.00	8.00	6.40	100.00	1.00	471.00	51.20
Dentistry-Queen's, Belfast, 2012	75.90	03 Dentistry	Dentistry	Queen's, Belfast	97.90	96.00	9.00	6.90	100.00	5.00	441.00	62.93
Dentistry-Birmingham, 2012	70.60	04 Dentistry	Dentistry	Birmingham	87.60	86.00	9.00	9.00	99.00	6.00	473.00	50.60
Dentistry-Bristol, 2012	65.00	05 Dentistry	Dentistry	Bristol	91.52	94.96	6.00	12.20	100.00	5.00	473.00	76.60
Dentistry-Dundee, 2012	60.70	06 Dentistry	Dentistry	Dundee	88.97	90.91	4.00	6.60	98.00	7.00	477.00	52.81
Dentistry-Sheffield, 2012	51.10	06 Dentistry	Dentistry	Sheffield	93.72	89.03	3.00	8.60	99.00	6.00	472.00	61.00
Dentistry-Liverpool, 2012	46.90	07 Dentistry	Dentistry	Liverpool	88.19	81.00	7.00	9.20	98.00	7.00	472.00	63.75
Dentistry-Manchester, 2012	43.50	08 Dentistry	Dentistry	Manchester	82.77	84.00	5.00	8.40	100.00	7.00	470.00	52.99
Dentistry-Newcastle, 2012	38.50	09 Dentistry	Dentistry	Newcastle	89.58	89.00	1.00	9.00	99.00	6.00	470.00	67.50
Dentistry-Queen Mary, 2012	29.90	09 Dentistry	Dentistry	Queen Mary	80.45	77.23	4.00	7.50	100.00	7.00	482.00	58.81
Dentistry-Leeds, 2012	29.80	10 Dentistry	Dentistry	Leeds	94.72	91.82	9.00	11.50	100.00	6.00	481.00	41.45
Veterinary science-Edinburgh, 2012	100.00	01 Veterinary science	Veterinary science	Edinburgh	90.22	89.00	8.00	11.00	99.00	3.00	528.00	59.20
Veterinary science-Liverpool, 2012	79.00	03 Veterinary science	Veterinary science	Liverpool	93.50	84.00	9.00	6.70	92.00	9.00	496.00	45.81
Veterinary science-Glasgow, 2012	66.30	04 Veterinary science	Veterinary science	Glasgow	94.01	94.00	3.00	6.10	92.00	4.00	477.00	48.00
Veterinary science-Nottingham, 2012	59.10	07 Veterinary science	Veterinary science	Nottingham	0.00	0.00	4.00	7.20	6.00	0.00	477.00	0.00
Veterinary science-Royal Veterinary College, 2012	56.10	09 Veterinary science	Veterinary science	Royal Veterinary College	89.25	80.00	8.00	8.50	87.00	6.00	487.00	37.72
Anatomy and physiology-Oxford, 2012	100.00	01 Anatomy and physiology	Anatomy and physiology	Oxford	93.18	92.00	7.00	12.00	99.00	8.00	443.00	74.44
Anatomy and physiology-Gloucestershire, 2012	94.20	01 Anatomy and physiology	Anatomy and physiology	Gloucestershire	90.68	89.90	4.00	9.80	96.00	10.00	350.00	75.35
Anatomy and physiology-Cardiff, 2012	91.20	02 Anatomy and physiology	Anatomy and physiology	Cardiff	92.40	92.00	8.00	12.00	99.00	8.00	426.00	66.90
Anatomy and physiology-Plymouth, 2012	91.20	02 Anatomy and physiology	Anatomy and physiology	Plymouth	95.50	93.00	8.00	8.70	82.00	8.00	335.00	86.83
Anatomy and physiology-Brunel, 2012	89.90	02 Anatomy and physiology	Anatomy and physiology	Brunel	96.98	96.00	3.00	12.60	81.00	9.00	337.00	63.80
Anatomy and physiology-Liverpool, 2012	80.10	02 Anatomy and physiology	Anatomy and physiology	Liverpool	92.56	91.48	3.00	8.00	81.00	4.00	374.00	67.30
Anatomy and physiology-Sussex, 2012	80.10	02 Anatomy and physiology	Anatomy and physiology	Sussex	94.25	95.05	6.00	12.50	99.00	9.00	386.00	61.08
Anatomy and physiology-Newcastle, 2012	75.40	03 Anatomy and physiology	Anatomy and physiology	Newcastle	93.12	95.55	4.00	13.70	77.00	5.00	414.00	64.62
Anatomy and physiology-Aston, 2012	75.30	03 Anatomy and physiology	Anatomy and physiology	Aston	85.11	84.71	3.00	14.40	95.00	2.00	463.00	75.98
Anatomy and physiology-Sheffield Hallam, 2012	62.20	04 Anatomy and physiology	Anatomy and physiology	Sheffield Hallam	96.33	90.90	7.00	14.20	78.00	5.00	388.00	56.63
Anatomy and physiology-Manchester, 2012	60.90	05 Anatomy and physiology	Anatomy and physiology	Manchester	91.26	88.83	5.00	12.80	77.00	4.00	418.00	50.10
Anatomy and physiology-Glasgow Caledonian, 2012	57.80	05 Anatomy and physiology	Anatomy and physiology	Glasgow Caledonian	92.01	91.00	4.00	15.20	75.00	7.00	377.00	59.80
Anatomy and physiology-Edinburgh, 2012	57.80	05 Anatomy and physiology	Anatomy and physiology	Edinburgh	92.01	93.00	9.00	17.60	95.00	5.00	437.00	43.95
Anatomy and physiology-Robert Gordon, 2012	57.70	06 Anatomy and physiology	Anatomy and physiology	Robert Gordon	94.00	93.84	3.00	20.30	80.00	7.00	414.00	64.39
Anatomy and physiology-De Montfort, 2012	57.60	06 Anatomy and physiology	Anatomy and physiology	De Montfort	85.78	86.07	5.00	15.60	97.00	5.00	292.00	61.08
Anatomy and physiology-Hertfordshire, 2012	56.40	06 Anatomy and physiology	Anatomy and physiology	Hertfordshire	89.82	88.86	3.00	16.80	87.00	7.00	299.00	67.60
Anatomy and physiology-UEA, 2012	56.10	06 Anatomy and physiology	Anatomy and physiology	UEA	96.18	97.61	0.00	0.90	89.00	3.00	363.00	58.28
Anatomy and physiology-Queen's, Belfast, 2012	55.30	07 Anatomy and physiology	Anatomy and physiology	Queen's, Belfast	91.00	90.00	9.00	10.10	60.00	7.00	362.00	45.78
Anatomy and physiology-Aberdeen, 2012	55.10	07 Anatomy and physiology	Anatomy and physiology	Aberdeen	89.13	85.11	4.00	14.40	70.00	6.00	320.00	73.98
Anatomy and physiology-Bradford, 2012	54.70	07 Anatomy and physiology	Anatomy and physiology	Bradford	89.42	95.80	4.00	16.70	89.00	3.00	362.00	61.92
Anatomy and physiology-Birmingham City, 2012	46.70	07 Anatomy and physiology	Anatomy and physiology	Birmingham City	94.90	89.99	3.00	20.90	76.00	8.00	325.00	52.41
Anatomy and physiology-Leeds, 2012	40.30	08 Anatomy and physiology	Anatomy and physiology	Leeds	89.27	88.13	5.00	23.10	98.00	6.00	419.00	63.94
Anatomy and physiology-Northampton, 2012	39.30	08 Anatomy and physiology	Anatomy and physiology	Northampton	93.75	80.00	3.00	17.80	0.00	0.00	232.00	67.54
Anatomy and physiology-Anglia Ruskin, 2012	36.80	08 Anatomy and physiology	Anatomy and physiology	Anglia Ruskin	88.78	94.00	5.00	27.80	60.00	8.00	310.00	78.53
Anatomy and physiology-Manchester Met, 2012	36.10	09 Anatomy and physiology	Anatomy and physiology	Manchester Met	85.40	80.00	2.00	15.60	82.00	3.00	352.00	63.99
Anatomy and physiology-City, 2012	33.00	09 Anatomy and physiology	Anatomy and physiology	City	86.33	78.83	4.00	15.40	92.00	1.00	382.00	45.41
Anatomy and physiology-Cumbria, 2012	33.00	09 Anatomy and physiology	Anatomy and physiology	Cumbria	81.94	84.65	3.00	19.90	96.00	2.00	306.00	60.90
Anatomy and physiology-St Mary's LIC, Twickenham, 2012	32.60	09 Anatomy and physiology	Anatomy and physiology	St Mary's LIC, Twickenham	95.24	93.04	1.00	9.00	100.00	1.00	426.00	71.60
Anatomy and physiology-King's College London, 2012	30.20	10 Anatomy and physiology	Anatomy and physiology	King's College London	87.22	89.00	7.00	28.50	66.00	3.00	421.00	54.80
Nursing and paramedical studies-Ulster, 2012	29.90	10 Nursing and paramedical studies	Nursing and paramedical studies	Ulster	81.13	64.67	3.00	15.10	57.00	8.00	352.00	56.16
Nursing and paramedical studies-Edinburgh, 2012	29.00	01 Nursing and paramedical studies	Nursing and paramedical studies	Edinburgh	96.03	96.00	3.00	16.00	100.00	6.00	388.00	67.00
Nursing and paramedical studies-Glasgow, 2012	28.70	01 Nursing and paramedical studies	Nursing and paramedical studies	Glasgow	97.94	96.91	7.00	16.90	6.00	6.00	377.00	95.15
Nursing and paramedical studies-UEA, 2012	28.10	01 Nursing and paramedical studies	Nursing and paramedical studies	UEA	91.24	90.41	10.00	14.10	97.00	8.00	322.00	62.03
Nursing and paramedical studies-Leeds, 2012	27.90	01 Nursing and paramedical studies	Nursing and paramedical studies	Leeds	91.14	89.14	10.00	15.20	100.00	5.00	327.00	67.00
Nursing and paramedical studies-Staffordshire, 2012	27.00	01 Nursing and paramedical studies	Nursing and paramedical studies	Staffordshire	91.79	100.00	7.00	14.40	97.00	8.00	251.00	78.80
Nursing and paramedical studies-Portsmouth, 2012	26.70	02 Nursing and paramedical studies	Nursing and paramedical studies	Portsmouth	80.04	80.01	0.00	12.40	80.00	4.00	320.00	62.01
Nursing and paramedical studies-City, 2012	25.10	02 Nursing and paramedical studies	Nursing and paramedical studies	City	82.05	80.00	1.00	17.40	80.00	5.00	280.00	62.81
Nursing and paramedical studies-Keels, 2012	25.00	02 Nursing and paramedical studies	Nursing and paramedical studies	Keels	96.44	95.29	6.00	16.50	87.00	7.00	282.00	76.15
Nursing and paramedical studies-Southampton, 2012	25.00	02 Nursing and paramedical studies	Nursing and paramedical studies	Southampton	88.12	86.89	8.00	17.20	89.00	9.00	386.00	54.40
Nursing and paramedical studies-Birmingham, 2012	23.20	02 Nursing and paramedical studies	Nursing and paramedical studies	Birmingham	92.74	92.74	2.00	18.10	100.00	4.00	372.00	67.10
Nursing and paramedical studies-Bedfordshire, 2012	24.30	02 Nursing and paramedical studies	Nursing and paramedical studies	Bedfordshire	90.45	82.00	4.00	15.00	6.00	10.00	244.00	67.60
Nursing and paramedical studies-Liverpool, 2012</												

В таблице 5 приводится лишь фрагмент исходных данных, т.к. их распечатка составляет 25 листов.

Отметим, что исходный файл и файл исходных данных находится на сайте автора в полном открытом бесплатном доступе на страничке: http://lc.kubagro.ru/ej_data/1071503001/Downloads.rar. Ниже приведена таблица 6 с формулами для расчета таблицы 5:

Таблица 6 – Формулы для расчета исходных данных по рейтингу Гардиан, для их подготовки к для ввода в систему «Эйдос» (фрагмент)

The object of training sample	Guardian score/100	Rank	Field of study	Name of Institution	% Satisfied with Teaching	% Satisfied overall with course	Expenditure per student (FTE)	Student:staff ratio	Career prospects	Value added score/10	Average Entry larif	% Satisfied with Assessment
"1'SAS1&"&"8'1'B3&"2012"	"=1'SC3	"=TEKCT('1'R3;"00."&"8'1'SAS1	"=1'SAS1	"=1'B3	"=1'D3	"=1'E3	"=1'F3	"=1'G3	"=1'H3	"=1'I3	"=1'J3	"=1'K3
"1'SAS1&"&"8'1'B4&"2012"	"=1'SC4	"=TEKCT('1'R4;"00."&"8'1'SAS1	"=1'SAS1	"=1'B4	"=1'D4	"=1'E4	"=1'F4	"=1'G4	"=1'H4	"=1'I4	"=1'J4	"=1'K4
"1'SAS1&"&"8'1'B5&"2012"	"=1'SC5	"=TEKCT('1'R5;"00."&"8'1'SAS1	"=1'SAS1	"=1'B5	"=1'D5	"=1'E5	"=1'F5	"=1'G5	"=1'H5	"=1'I5	"=1'J5	"=1'K5
"1'SAS1&"&"8'1'B6&"2012"	"=1'SC6	"=TEKCT('1'R6;"00."&"8'1'SAS1	"=1'SAS1	"=1'B6	"=1'D6	"=1'E6	"=1'F6	"=1'G6	"=1'H6	"=1'I6	"=1'J6	"=1'K6
"1'SAS1&"&"8'1'B7&"2012"	"=1'SC7	"=TEKCT('1'R7;"00."&"8'1'SAS1	"=1'SAS1	"=1'B7	"=1'D7	"=1'E7	"=1'F7	"=1'G7	"=1'H7	"=1'I7	"=1'J7	"=1'K7
"1'SAS1&"&"8'1'B8&"2012"	"=1'SC8	"=TEKCT('1'R8;"00."&"8'1'SAS1	"=1'SAS1	"=1'B8	"=1'D8	"=1'E8	"=1'F8	"=1'G8	"=1'H8	"=1'I8	"=1'J8	"=1'K8
"1'SAS1&"&"8'1'B9&"2012"	"=1'SC9	"=TEKCT('1'R9;"00."&"8'1'SAS1	"=1'SAS1	"=1'B9	"=1'D9	"=1'E9	"=1'F9	"=1'G9	"=1'H9	"=1'I9	"=1'J9	"=1'K9
"1'SAS1&"&"8'1'B10&"2012"	"=1'SC10	"=TEKCT('1'R10;"00."&"8'1'SAS1	"=1'SAS1	"=1'B10	"=1'D10	"=1'E10	"=1'F10	"=1'G10	"=1'H10	"=1'I10	"=1'J10	"=1'K10
"1'SAS1&"&"8'1'B11&"2012"	"=1'SC11	"=TEKCT('1'R11;"00."&"8'1'SAS1	"=1'SAS1	"=1'B11	"=1'D11	"=1'E11	"=1'F11	"=1'G11	"=1'H11	"=1'I11	"=1'J11	"=1'K11
"1'SAS1&"&"8'1'B12&"2012"	"=1'SC12	"=TEKCT('1'R12;"00."&"8'1'SAS1	"=1'SAS1	"=1'B12	"=1'D12	"=1'E12	"=1'F12	"=1'G12	"=1'H12	"=1'I12	"=1'J12	"=1'K12
"1'SAS1&"&"8'1'B13&"2012"	"=1'SC13	"=TEKCT('1'R13;"00."&"8'1'SAS1	"=1'SAS1	"=1'B13	"=1'D13	"=1'E13	"=1'F13	"=1'G13	"=1'H13	"=1'I13	"=1'J13	"=1'K13
"1'SAS1&"&"8'1'B14&"2012"	"=1'SC14	"=TEKCT('1'R14;"00."&"8'1'SAS1	"=1'SAS1	"=1'B14	"=1'D14	"=1'E14	"=1'F14	"=1'G14	"=1'H14	"=1'I14	"=1'J14	"=1'K14
"1'SAS1&"&"8'1'B15&"2012"	"=1'SC15	"=TEKCT('1'R15;"00."&"8'1'SAS1	"=1'SAS1	"=1'B15	"=1'D15	"=1'E15	"=1'F15	"=1'G15	"=1'H15	"=1'I15	"=1'J15	"=1'K15
"1'SAS1&"&"8'1'B16&"2012"	"=1'SC16	"=TEKCT('1'R16;"00."&"8'1'SAS1	"=1'SAS1	"=1'B16	"=1'D16	"=1'E16	"=1'F16	"=1'G16	"=1'H16	"=1'I16	"=1'J16	"=1'K16
"1'SAS1&"&"8'1'B17&"2012"	"=1'SC17	"=TEKCT('1'R17;"00."&"8'1'SAS1	"=1'SAS1	"=1'B17	"=1'D17	"=1'E17	"=1'F17	"=1'G17	"=1'H17	"=1'I17	"=1'J17	"=1'K17
"1'SAS1&"&"8'1'B18&"2012"	"=1'SC18	"=TEKCT('1'R18;"00."&"8'1'SAS1	"=1'SAS1	"=1'B18	"=1'D18	"=1'E18	"=1'F18	"=1'G18	"=1'H18	"=1'I18	"=1'J18	"=1'K18
"1'SAS1&"&"8'1'B19&"2012"	"=1'SC19	"=TEKCT('1'R19;"00."&"8'1'SAS1	"=1'SAS1	"=1'B19	"=1'D19	"=1'E19	"=1'F19	"=1'G19	"=1'H19	"=1'I19	"=1'J19	"=1'K19
"1'SAS1&"&"8'1'B20&"2012"	"=1'SC20	"=TEKCT('1'R20;"00."&"8'1'SAS1	"=1'SAS1	"=1'B20	"=1'D20	"=1'E20	"=1'F20	"=1'G20	"=1'H20	"=1'I20	"=1'J20	"=1'K20
"1'SAS1&"&"8'1'B21&"2012"	"=1'SC21	"=TEKCT('1'R21;"00."&"8'1'SAS1	"=1'SAS1	"=1'B21	"=1'D21	"=1'E21	"=1'F21	"=1'G21	"=1'H21	"=1'I21	"=1'J21	"=1'K21
"1'SAS1&"&"8'1'B22&"2012"	"=1'SC22	"=TEKCT('1'R22;"00."&"8'1'SAS1	"=1'SAS1	"=1'B22	"=1'D22	"=1'E22	"=1'F22	"=1'G22	"=1'H22	"=1'I22	"=1'J22	"=1'K22
"1'SAS1&"&"8'1'B23&"2012"	"=1'SC23	"=TEKCT('1'R23;"00."&"8'1'SAS1	"=1'SAS1	"=1'B23	"=1'D23	"=1'E23	"=1'F23	"=1'G23	"=1'H23	"=1'I23	"=1'J23	"=1'K23
"1'SAS1&"&"8'1'B24&"2012"	"=1'SC24	"=TEKCT('1'R24;"00."&"8'1'SAS1	"=1'SAS1	"=1'B24	"=1'D24	"=1'E24	"=1'F24	"=1'G24	"=1'H24	"=1'I24	"=1'J24	"=1'K24
"1'SAS1&"&"8'1'B25&"2012"	"=1'SC25	"=TEKCT('1'R25;"00."&"8'1'SAS1	"=1'SAS1	"=1'B25	"=1'D25	"=1'E25	"=1'F25	"=1'G25	"=1'H25	"=1'I25	"=1'J25	"=1'K25
"1'SAS1&"&"8'1'B26&"2012"	"=1'SC26	"=TEKCT('1'R26;"00."&"8'1'SAS1	"=1'SAS1	"=1'B26	"=1'D26	"=1'E26	"=1'F26	"=1'G26	"=1'H26	"=1'I26	"=1'J26	"=1'K26
"1'SAS1&"&"8'1'B27&"2012"	"=1'SC27	"=TEKCT('1'R27;"00."&"8'1'SAS1	"=1'SAS1	"=1'B27	"=1'D27	"=1'E27	"=1'F27	"=1'G27	"=1'H27	"=1'I27	"=1'J27	"=1'K27
"1'SAS1&"&"8'1'B28&"2012"	"=1'SC28	"=TEKCT('1'R28;"00."&"8'1'SAS1	"=1'SAS1	"=1'B28	"=1'D28	"=1'E28	"=1'F28	"=1'G28	"=1'H28	"=1'I28	"=1'J28	"=1'K28
"1'SAS1&"&"8'1'B29&"2012"	"=1'SC29	"=TEKCT('1'R29;"00."&"8'1'SAS1	"=1'SAS1	"=1'B29	"=1'D29	"=1'E29	"=1'F29	"=1'G29	"=1'H29	"=1'I29	"=1'J29	"=1'K29
"1'SAS1&"&"8'1'B30&"2012"	"=1'SC30	"=TEKCT('1'R30;"00."&"8'1'SAS1	"=1'SAS1	"=1'B30	"=1'D30	"=1'E30	"=1'F30	"=1'G30	"=1'H30	"=1'I30	"=1'J30	"=1'K30
"1'SAS1&"&"8'1'B31&"2012"	"=1'SC31	"=TEKCT('1'R31;"00."&"8'1'SAS1	"=1'SAS1	"=1'B31	"=1'D31	"=1'E31	"=1'F31	"=1'G31	"=1'H31	"=1'I31	"=1'J31	"=1'K31
"1'SAS1&"&"8'1'B32&"2012"	"=1'SC32	"=TEKCT('1'R32;"00."&"8'1'SAS1	"=1'SAS1	"=1'B32	"=1'D32	"=1'E32	"=1'F32	"=1'G32	"=1'H32	"=1'I32	"=1'J32	"=1'K32
"2'SAS1&"&"8'2'B3&"2012"	"=2'SC3	"=TEKCT('2'R3;"00."&"8'2'SAS1	"=2'SAS1	"=2'B3	"=2'D3	"=2'E3	"=2'F3	"=2'G3	"=2'H3	"=2'I3	"=2'J3	"=2'K3
"2'SAS1&"&"8'2'B4&"2012"	"=2'SC4	"=TEKCT('2'R4;"00."&"8'2'SAS1	"=2'SAS1	"=2'B4	"=2'D4	"=2'E4	"=2'F4	"=2'G4	"=2'H4	"=2'I4	"=2'J4	"=2'K4
"2'SAS1&"&"8'2'B5&"2012"	"=2'SC5	"=TEKCT('2'R5;"00."&"8'2'SAS1	"=2'SAS1	"=2'B5	"=2'D5	"=2'E5	"=2'F5	"=2'G5	"=2'H5	"=2'I5	"=2'J5	"=2'K5
"2'SAS1&"&"8'2'B6&"2012"	"=2'SC6	"=TEKCT('2'R6;"00."&"8'2'SAS1	"=2'SAS1	"=2'B6	"=2'D6	"=2'E6	"=2'F6	"=2'G6	"=2'H6	"=2'I6	"=2'J6	"=2'K6
"2'SAS1&"&"8'2'B7&"2012"	"=2'SC7	"=TEKCT('2'R7;"00."&"8'2'SAS1	"=2'SAS1	"=2'B7	"=2'D7	"=2'E7	"=2'F7	"=2'G7	"=2'H7	"=2'I7	"=2'J7	"=2'K7
"2'SAS1&"&"8'2'B8&"2012"	"=2'SC8	"=TEKCT('2'R8;"00."&"8'2'SAS1	"=2'SAS1	"=2'B8	"=2'D8	"=2'E8	"=2'F8	"=2'G8	"=2'H8	"=2'I8	"=2'J8	"=2'K8
"2'SAS1&"&"8'2'B9&"2012"	"=2'SC9	"=TEKCT('2'R9;"00."&"8'2'SAS1	"=2'SAS1	"=2'B9	"=2'D9	"=2'E9	"=2'F9	"=2'G9	"=2'H9	"=2'I9	"=2'J9	"=2'K9
"2'SAS1&"&"8'2'B10&"2012"	"=2'SC10	"=TEKCT('2'R10;"00."&"8'2'SAS1	"=2'SAS1	"=2'B10	"=2'D10	"=2'E10	"=2'F10	"=2'G10	"=2'H10	"=2'I10	"=2'J10	"=2'K10
"2'SAS1&"&"8'2'B11&"2012"	"=2'SC11	"=TEKCT('2'R11;"00."&"8'2'SAS1	"=2'SAS1	"=2'B11	"=2'D11	"=2'E11	"=2'F11	"=2'G11	"=2'H11	"=2'I11	"=2'J11	"=2'K11
"2'SAS1&"&"8'2'B12&"2012"	"=2'SC12	"=TEKCT('2'R12;"00."&"8'2'SAS1	"=2'SAS1	"=2'B12	"=2'D12	"=2'E12	"=2'F12	"=2'G12	"=2'H12	"=2'I12	"=2'J12	"=2'K12
"2'SAS1&"&"8'2'B13&"2012"	"=2'SC13	"=TEKCT('2'R13;"00."&"8'2'SAS1	"=2'SAS1	"=2'B13	"=2'D13	"=2'E13	"=2'F13	"=2'G13	"=2'H13	"=2'I13	"=2'J13	"=2'K13
"2'SAS1&"&"8'2'B14&"2012"	"=2'SC14	"=TEKCT('2'R14;"00."&"8'2'SAS1	"=2'SAS1	"=2'B14	"=2'D14	"=2'E14	"=2'F14	"=2'G14	"=2'H14	"=2'I14	"=2'J14	"=2'K14
"2'SAS1&"&"8'2'B15&"2012"	"=2'SC15	"=TEKCT('2'R15;"00."&"8'2'SAS1	"=2'SAS1	"=2'B15	"=2'D15	"=2'E15	"=2'F15	"=2'G15	"=2'H15	"=2'I15	"=2'J15	"=2'K15
"3'SAS1&"&"8'3'B&"2012"	"=3'SC	"=TEKCT('3'R;"00."&"8'3'SAS1	"=3'SAS1	"=3'B	"=3'D	"=3'E	"=3'F	"=3'G	"=3'H	"=3'I	"=3'J	"=3'K
"3'SAS1&"&"8'3'B4&"2012"	"=3'SC4	"=TEKCT('3'R4;"00."&"8'3'SAS1	"=3'SAS1	"=3'B4	"=3'D4	"=3'E4	"=3'F4	"=3'G4	"=3'H4	"=3'I4	"=3'J4	"=3'K4
"3'SAS1&"&"8'3'B5&"2012"	"=3'SC5	"=TEKCT('3'R5;"00."&"8'3'SAS1	"=3'SAS1	"=3'B5	"=3'D5	"=3'E5	"=3'F5	"=3'G5	"=3'H5	"=3'I5	"=3'J5	"=3'K5
"3'SAS1&"&"8'3'B6&"2012"	"=3'SC6	"=TEKCT('3'R6;"00."&"8'3'SAS1	"=3'SAS1	"=3'B6	"=3'D6	"=3'E6	"=3'F6	"=3'G6	"=3'H6	"=3'I6	"=3'J6	"=3'K6
"3'SAS1&"&"8'3'B7&"2012"	"=3'SC7	"=TEKCT('3'R7;"00."&"8'3'SAS1	"=3'SAS1	"=3'B7	"=3'D7	"=3'E7	"=3'F7	"=3'G7	"=3'H7	"=3'I7	"=3'J7	"=3'K7
"3'SAS1&"&"8'3'B8&"2012"	"=3'SC8	"=TEKCT('3'R8;"00."&"8'3'SAS1	"=3'SAS1	"=3'B8	"=3'D8	"=3'E8	"=3'F8	"=3'G8	"=3'H8	"=3'I8	"=3'J8	"=3'K8
"3'SAS1&"&"8'3'B9&"2012"	"=3'SC9	"=TEKCT('3'R9;"00."&"8'3'SAS1	"=3'SAS1	"=3'B9	"=3'D9	"=3'E9	"=3'F9	"=3'G9	"=3'H9	"=3'I9	"=3'J9	"=3'K9
"4'SAS1&"&"8'4'B&"2012"	"=4'SC	"=TEKCT('4'R;"00."&"8'4'SAS1	"=4'SAS1	"=4'B	"=4'D	"=4'E	"=4'F	"=4'G	"=4'H	"=4'I	"=4'J	"=4'K
"4'SAS1&"&"8'4'B4&"2012"	"=4'SC4	"=TEKCT('4'R4;"00."&"8'4'SAS1	"=4'SAS1	"=4'B4	"=4'D4	"=4'E4	"=4'F4	"=4'G4	"=4'H4	"=4'I4	"=4'J4	"=4'K4
"4'SAS1&"&"8'4'B5&"2012"	"=4'SC5	"=TEKCT('4'R5;"00."&"8'4'SAS1	"=4'SAS1	"=4'B5	"=4'D5	"=4'E5	"=4'F5	"=4'G5	"=4'H5	"=4'I5	"=4'J5	"=4'K5
"4'SAS1&"&"8'4'B6&"2012"	"=4'SC6	"=TEKCT('4'R6;"00."&"8'4'SAS1	"=4'SAS1	"=4'B6	"=4'D6	"=4'E6	"=4'F6	"=4'G6	"=4'H6	"=4'I6	"=4'J6	"=4'K6
"4'SAS1&"&"8'4'B7&"2012"	"=4'SC7	"=TEKCT('4'R7;"00."&"8'4'SAS1	"=4'SAS1	"=4'B7	"=4'D7	"=4'E7	"=4'F7	"=4'G7	"=4'H7	"=4'I7	"=4'J7	"=4'K7
"4'SAS1&"&"8'4'B8&"2012"	"=4'SC8	"=TEKCT('4'R8;"00."&"8'4'SAS1	"=4'SAS1	"=4'B8	"=4'D8	"=4'E8	"=4'F8	"=4'G8	"=4'H8	"=4'I8	"=4'J8	"=4'K8
"4'SAS1&"&"8'4'B9&"2012"	"=4'SC9	"=TEKCT('4'R9;"00."&"8'4'SAS1	"=4'SAS1	"=4'B9	"=4'D9	"=4'E9	"=4'F9	"=4'G9	"=4'H9	"=4'I9	"=4'J9	"=4'K9
"4'SAS1&"&"8'4'B10&"2012"	"=4'SC10	"=TEKCT('4'R10;"00."&"8'4'SAS1	"=4'SAS1	"=4'B10	"=4'D10	"=4'E10	"=4'F10	"=4'G10	"=4'H10	"=4'I10	"=4'J10	"=4'K10
"4'SAS1&"&"8'4'B11&"2012"	"=4'SC11	"=TEKCT('4'R11;"00."&"8'4'SAS1	"=4'SAS1	"=4'B11	"=4'D11	"=4'E11	"=4'F11	"=4'G11	"=4'H11	"=4'I11	"=4'J11	"=4'K11
"4'SAS1&"&"8'4'B12&"2012"	"=4'SC12	"=TEKCT('4'R12;"00."&"8'4'SAS1	"=4'SAS1	"=4'B12	"=4'D12	"=4'E12	"=4'F12	"=4'G12	"=4'H12	"=4'I12	"=4'J12	"=4'K12
"4'SAS1&"&"8'4'B13&"2012"	"=4'SC13	"=TEKCT('4'R13;"00."&"8'4'SAS1	"=4'SAS1	"=4'B13	"=4'D13	"=4'E13	"=4'F13	"=4'G13	"=4'H13	"=4'I13	"=4'J13	"=4'K13
"4'SAS1&"&"8'4'B14&"2012"	"=4'SC14	"=TEKCT('4'R14;"00."&"8'4'SAS1	"=4'SAS1	"=4'B14	"=4'D14	"=4'E14	"=4'F14	"=4'G14	"=4'H14	"=4'I14	"=4'J14	"=4'K14
"4'SAS1&"&"8'4'B15&"2012"	"=4'SC15	"=TEKCT('4'R15;"00."&"8'4'SAS1	"=4'SAS1	"=4'B15	"=4'D15	"=4'E15	"=4'F15	"=4'G15	"=4'H15	"=4'I15	"=4'J15	"=4'K15
"4'SAS1&"&"8'4'B16&"2012"	"=4'SC16	"=TEKCT('4'R16;"00."&"8'4'SAS1	"=4'SAS1	"=4'B16	"=4'D16	"=4'E16	"=4'F16	"=4'G16	"=4'H16	"=4'I16	"=4'J16	"=4'K16
"4'SAS1&"&"8'4'B17&"2012"	"=4'SC17	"=TEKCT('4'R17;"00."&"8'4'SAS1	"=4'SAS1	"=4'B17	"=4'D17	"=4'E17	"=4'F17	"=4'G17	"=4'H17	"=4'I17	"=4'J17	"=4'K17
"4'SAS1&"&"8'4'B18&"2012"	"=4'SC18	"=TEKCT('4'R18;"00."&"8'4'SAS1	"=4'SAS1	"=4'B18	"=4'D18	"=4'E18	"=4'F18	"=4'G18	"=4'H18	"=4'I18	"=4'J18	"=4'K18
"4'SAS1&"&"8'4'B19&"2012"	"=4'SC19	"=TEKCT('4'R19;"00."&"8'4'SAS1	"=4'SAS1	"=4'B19	"=4'D19	"=4'E19	"=4'F19	"=4'G19	"=4'H19	"=4'I19	"=4'J19	"=4'K19
"4'SAS1&"&"8'4'B20&"2012"	"=4'SC20	"=TEKCT('4'R20;"00."&"8'4'SAS1	"=4'SAS1	"=4'B20	"=4'D20	"=4'E20	"=4'F20	"=4'G20	"=4'H20	"=4'I20	"=4'J20	"=4'K20
"4'SAS1&"&"8'4'B21&"2012"	"=4'SC21	"=TEKCT('4'R21;"00."&"8'4'SAS1	"=4'SAS1	"=4'B21	"=4'D21	"=4'E21	"=4'F21	"=4'G21	"=4'H21	"=4'I21	"=4'J21	"=4'K21
"4'SAS1&"&"8'4'B22&"2012"	"=4'SC22	"=TEKCT('4'R22;"00."&"8'4'SAS1	"=4'SAS1	"=4'B22	"=4'D22	"=4'E22	"=4'F22	"=4'G22	"=4'H22	"=4'I22	"=4'J22	"=4'K22
"4'SAS1&"&"8'4'B23&"2012"	"=4'SC23	"=TEKCT('4'R23;"00."&"8'4'SAS1	"=4'SAS1	"=4'B23	"=4'D23	"=4'E23	"=4'F23	"=4'G23	"=4'H23	"=4'I23	"=4'J23	"=4'K23
"4'SAS1&"&"8'4'B24&"2012"	"=4'SC24	"=TEKCT('4'R24;"00."&"8'4'SAS1	"=4'SAS1</									


6.4.3.3. УСТАНОВКА СИСТЕМЫ «ЭЙДОС»


Скачиваем и устанавливаем систему «Эйдос». Это наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с исходными текстами, находящаяся в полном открытом бесплатном доступе (около 50 Мб). Обновление имеет объем около 3 Мб.⁷¹

ИНСТРУКЦИЯ



по скачиванию и установке системы «Эйдос» (объем около 50 Мб)


Система не требует инсталляции, не меняет никаких системных файлов и содержимого папок операционной системы, т.е. является портативной (portable) программой. Но чтобы она работала необходимо аккуратно выполнить следующие пункты.

1. Скачать самую новую на текущий момент версию системы «Эйдос-X++» по ссылкам: <http://lc.kubagro.ru/a.rar> или: <http://lc.kubagro.ru/Aidos-X.exe> (ссылки для обновления системы даны в режиме 6.2).
2. Разархивировать этот архив в любую папку с правами на запись с коротким латинским именем и путем доступа, включающим только папки с такими же именами (лучше всего в корневой каталог какого-нибудь диска).
3. Запустить систему. Файл запуска: _AIDOS-X.exe *
4. Задать имя: 1 и пароль: 1 (потом их можно поменять в режиме 1.2).
5. Перед тем как запустить новый режим НЕОБХОДИМО ЗАВЕРШИТЬ предыдущий (Help можно не закрывать). Окна закрываются в порядке, обратном порядку их открытия.

* Разработана программа: «_START_AIDOS.exe», полностью снимающая с пользователя системы «Эйдос-X++» заботу о проверке наличия и скачивании обновлений. Эту программу надо просто скачать по ссылке: http://lc.kubagro.ru/Install_Aidos-X/_START_AIDOS.exe, поместить в папку с исполнимым модулем системы и всегда запускать систему с помощью этого файла.

При запуске программы _START_AIDOS.EXE система Эйдос не должна быть запущена, т.к. она содержится в файле обновлений и при его разархивировании возникнет конфликт, если система будет запущена.

1. Программа _START_AIDOS.exe определяет дату системы Эйдос в текущей папке, и дату обновлений на FTP-сервере не скачивая их, и, если система Эйдос в текущей папке устарела, скачивает обновления. (Если в текущей папке нет исполнимого модуля системы Эйдос, то программа пытается скачать полную инсталляцию системы, но не может этого сделать из-за ограниченной функциональности демо-версии библиотеки Xb2NET.DLL).
2. После этого появляется диалоговое окно с сообщением, что надо сначала разархивировать систему, заменяя все файлы (опция: «Yes to All» или «OverWrite All»), и только после этого закрыть данное окно.
3. Потом программа _START_AIDOS.exe запускает обновления на разархивирование. После окончания разархивирования окно архиватора с отображением стадии процесса исчезает.
4. После закрытия диалогового окна с инструкцией (см. п.2), происходит запуск обновленной версии системы Эйдос на исполнение.

Для работы программы _START_AIDOS.exe необходима библиотека: Xb2NET.DLL, которую можно скачать по ссылке: http://lc.kubagro.ru/Install_Aidos-X/Xb2NET.DLL. Перед первым запуском этой программы данную библиотеку необходимо скачать и поместить либо в папку с этой программой, а значит и исполнимым модулем системы «Эйдос-X++», либо в любую другую папку, на которую в операционной системе прописаны пути поиска файлов, например в папку: c:\Windows\System32\. Эта библиотека стоит около 500\$ и у меня ее нет, поэтому я даю только

⁷¹ http://lc.kubagro.ru/aidos/_Aidos-X.htm

бесплатную демо-версию, которая выдает сообщение об ограниченной функциональности, но для наших целей ее достаточно.

Лицензия:

Автор отказывается от какой бы то ни было ответственности за последствия применения или не применения Вами системы «Эйдос».

Проще говоря, пользуйтесь если понравилось, а если не понравилось – сотрите и забудьте, а лучше вообще не скачивайте.

6.4.3.4. ВВОД ИСХОДНЫХ ДАННЫХ В СИСТЕМУ «ЭЙДОС» С ПОМОЩЬЮ ОДНОГО И ЕЕ ПРОГРАММНЫХ ИНТЕРФЕЙСОВ

Записываем файл исходных данных, приведенный в таблице 5, с именем: Inp_data.xls в папку с системой (если она на диске C: в корневом каталоге) по пути:

c:\Aidos-X\AID_DATA\Inp_data\Inp_data.xls

Запускаем систему «Эйдос» и задаем режим 2.3.2.2 с параметрами, указанными на экранной форме (рисунок 5):

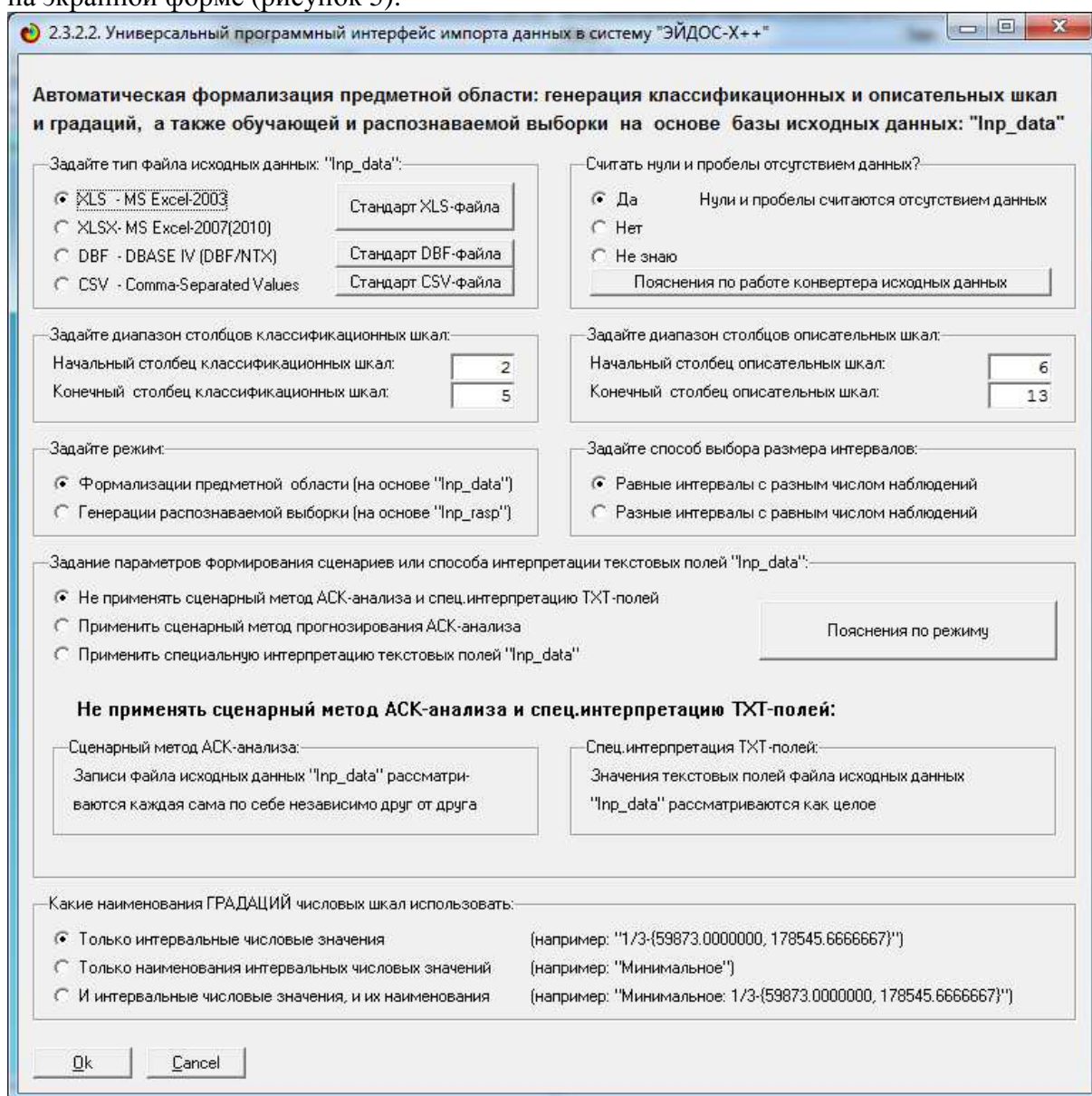


Рисунок 5 – Экранная форма универсального программного интерфейса импорта данных из внешних баз данных в систему «Эйдос»

Через несколько секунд **на заднем фоне**⁷² появляется окно (рисунок 6) на котором нажимаем «Сохранить», после чего появляется экранный калькулятор (рисунок 7):

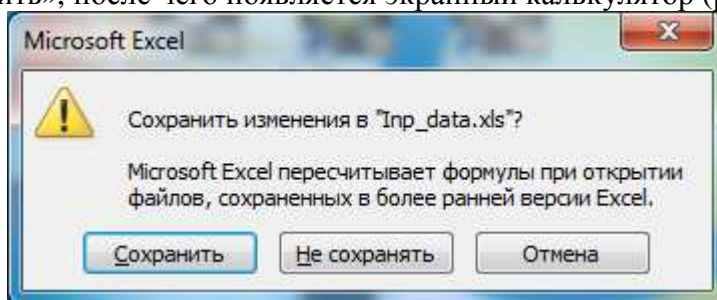


Рисунок 6 – Экранная форма, выдаваемая MS Excel, т.к. в файле исходных данных есть расчетные ячейки

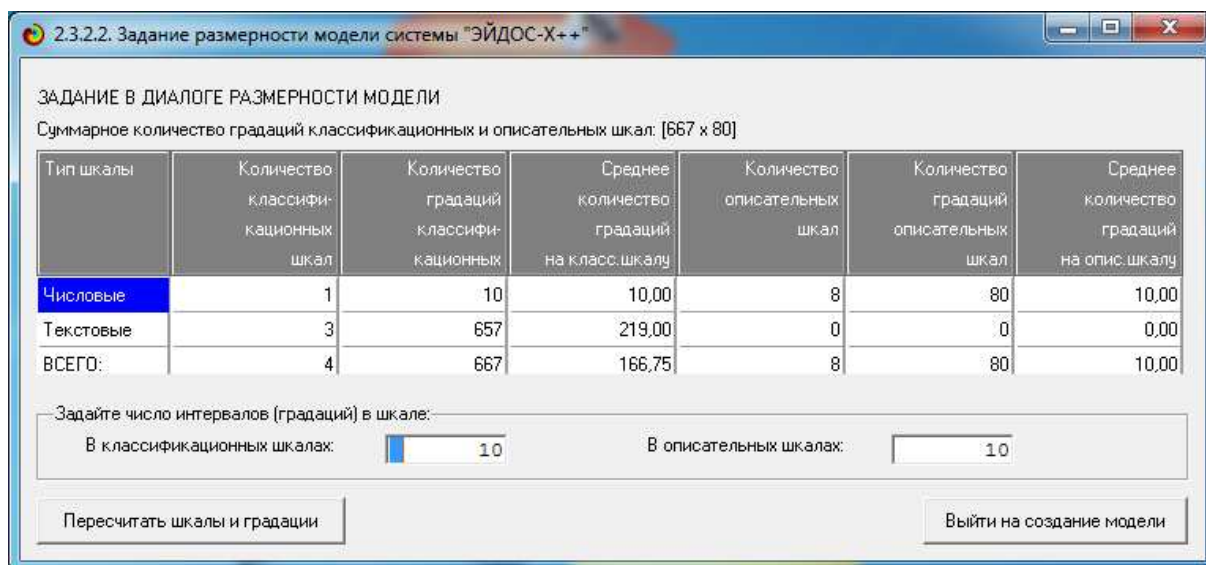


Рисунок 7 – Экранного калькулятора универсального программного интерфейса импорта данных из внешних баз данных в систему «Эйдос»

На этом калькуляторе а данном случае задано по 10 интервальных числовых значений в числовых классификационных и описательных шкалах. Можно задать другие их количество, затем пересчитать шкалы и градации и выйти на создание модели.

За 41 секунду происходит формирование классификационных и описательных шкал и градаций и обучающей выборки по 2559 примерам вузов, описанных в исходных данных (рисунок 8):

⁷² А значит, чтобы его увидеть надо свернуть все окна

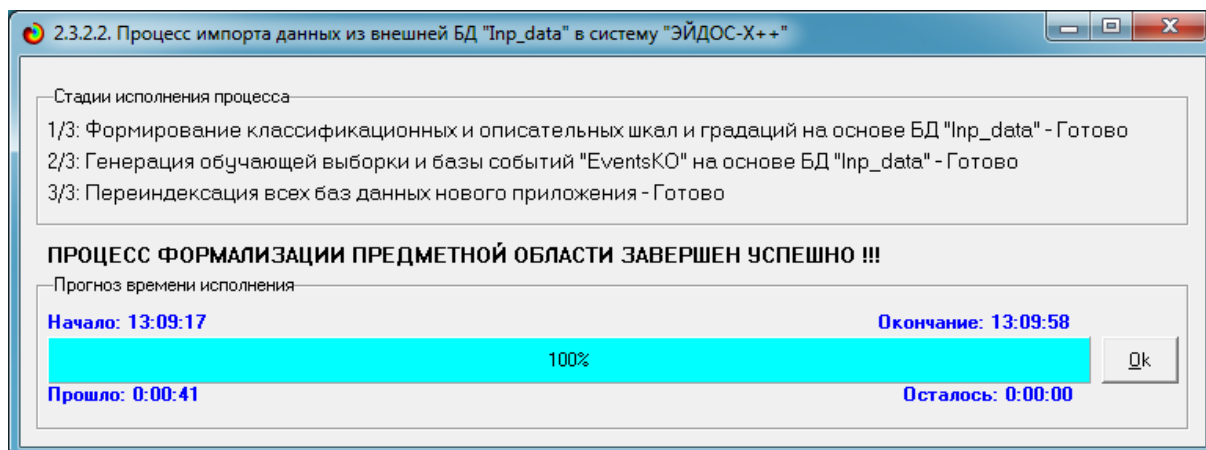


Рисунок 8. Экранная форма отображения стадии и прогноза времени исполнения

В результате автоматически формируются классификационные и описательные шкалы и градации и обучающая выборка, приведенные в таблицах 7, 8, 9:

Таблица 7 – Классификационные шкалы и градации (фрагмент)

Код	Наименование
1	GUARDIAN SCORE/100-1/10-{25.9000000, 33.3100000}
2	GUARDIAN SCORE/100-2/10-{33.3100000, 40.7200000}
3	GUARDIAN SCORE/100-3/10-{40.7200000, 48.1300000}
4	GUARDIAN SCORE/100-4/10-{48.1300000, 55.5400000}
5	GUARDIAN SCORE/100-5/10-{55.5400000, 62.9500000}
6	GUARDIAN SCORE/100-6/10-{62.9500000, 70.3600000}
7	GUARDIAN SCORE/100-7/10-{70.3600000, 77.7700000}
8	GUARDIAN SCORE/100-8/10-{77.7700000, 85.1800000}
9	GUARDIAN SCORE/100-9/10-{85.1800000, 92.5900000}
10	GUARDIAN SCORE/100-10/10-{92.5900000, 100.0000000}
11	RANK-01.Agriculture, forestry and food
12	RANK-01.American studies
13	RANK-01.Anatomy and physiology
14	RANK-01.Anthropology
15	RANK-01.Archaeology and Forensics
16	RANK-01.Architecture
17	RANK-01.Art and design
18	RANK-01.Biosciences
19	RANK-01.Building and town and country planning
20	RANK-01.Business and management studies
21	RANK-01.Chemistry
22	RANK-01.Classics
23	RANK-01.Computer sciences and IT
24	RANK-01.Dentistry
25	RANK-01.Drama and dance
26	RANK-01.Earth and marine sciences
27	RANK-01.Economics
28	RANK-01.Education
29	RANK-01.Engineering: chemical
30	RANK-01.Engineering: civil
31	RANK-01.Engineering: electronic and electrical
32	RANK-01.Engineering: general
33	RANK-01.Engineering: materials and mineral
34	RANK-01.Engineering: mechanical
35	RANK-01.English
36	RANK-01.Geography and environmental studies
37	RANK-01.History and history of art
38	RANK-01.Law
39	RANK-01.Mathematics
40	RANK-01.Media studies, communications and librarianship
41	RANK-01.Medicine
42	RANK-01.Modern languages and linguistics
43	RANK-01.Music
44	RANK-01.Nursing and paramedical studies
45	RANK-01.Pharmacy and pharmacology

46	RANK-01.Philosophy
47	RANK-01.Physics
48	RANK-01.Politics
49	RANK-01.Psychology
50	RANK-01.Religious studies and theology
51	RANK-01.Social policy and administration
52	RANK-01.Social work
53	RANK-01.Sociology
54	RANK-01.Sports science
55	RANK-01.Tourism, transport and travel
56	RANK-01.Veterinary science
57	RANK-02.Agriculture, forestry and food
58	RANK-02.American studies
59	RANK-02.Anatomy and physiology
60	RANK-02.Anthropology
61	RANK-02.Archaeology and Forensics
62	RANK-02.Architecture
63	RANK-02.Art and design
64	RANK-02.Biosciences
65	RANK-02.Building and town and country planning
66	RANK-02.Business and management studies
67	RANK-02.Chemistry
68	RANK-02.Classics
69	RANK-02.Computer sciences and IT
70	RANK-02.Dentistry
71	RANK-02.Drama and dance
72	RANK-02.Earth and marine sciences
73	RANK-02.Economics
74	RANK-02.Education
75	RANK-02.Engineering: chemical
76	RANK-02.Engineering: civil
77	RANK-02.Engineering: electronic and electrical
78	RANK-02.Engineering: general
79	RANK-02.Engineering: materials and mineral
80	RANK-02.Engineering: mechanical
81	RANK-02.English
82	RANK-02.Geography and environmental studies
83	RANK-02.History and history of art
84	RANK-02.Law
85	RANK-02.Mathematics
86	RANK-02.Media studies, communications and librarianship
87	RANK-02.Medicine
88	RANK-02.Modern languages and linguistics
89	RANK-02.Music
90	RANK-02.Nursing and paramedical studies
91	RANK-02.Pharmacy and pharmacology

92	RANK-02.Philosophy
93	RANK-02.Physics
94	RANK-02.Politics
95	RANK-02.Psychology
96	RANK-02.Religious studies and theology
97	RANK-02.Social policy and administration
98	RANK-02.Social work
99	RANK-02.Sociology
100	RANK-02.Sports science
101	RANK-02.Tourism, transport and travel
102	RANK-03.Agriculture, forestry and food
103	RANK-03.American studies
104	RANK-03.Anatomy and physiology
105	RANK-03.Anthropology
106	RANK-03.Archaeology and Forensics
107	RANK-03.Architecture
108	RANK-03.Art and design
109	RANK-03.Biosciences
110	RANK-03.Building and town and country planning
111	RANK-03.Business and management studies
112	RANK-03.Chemistry
113	RANK-03.Classics
114	RANK-03.Computer sciences and IT
115	RANK-03.Dentistry
116	RANK-03.Drama and dance
117	RANK-03.Earth and marine sciences
118	RANK-03.Economics
119	RANK-03.Education
120	RANK-03.Engineering: chemical
121	RANK-03.Engineering: civil
122	RANK-03.Engineering: electronic and electrical
123	RANK-03.Engineering: general
124	RANK-03.Engineering: materials and mineral
125	RANK-03.Engineering: mechanical
126	RANK-03.English
127	RANK-03.Geography and environmental studies
128	RANK-03.History and history of art
129	RANK-03.Law
130	RANK-03.Mathematics
131	RANK-03.Media studies, communications and librarianship
132	RANK-03.Medicine
133	RANK-03.Modern languages and linguistics
134	RANK-03.Music
135	RANK-03.Nursing and paramedical studies
136	RANK-03.Pharmacy and pharmacology
137	RANK-03.Philosophy
138	RANK-03.Physics
139	RANK-03.Politics
140	RANK-03.Psychology
141	RANK-03.Religious studies and theology
142	RANK-03.Social policy and administration
143	RANK-03.Social work
144	RANK-03.Sociology
145	RANK-03.Sports science
146	RANK-03.Tourism, transport and travel
147	RANK-03.Veterinary science
148	RANK-04.Agriculture, forestry and food
149	RANK-04.American studies
150	RANK-04.Anatomy and physiology
151	RANK-04.Anthropology
152	RANK-04.Archaeology and Forensics
153	RANK-04.Architecture
154	RANK-04.Art and design
155	RANK-04.Biosciences
156	RANK-04.Building and town and country planning
157	RANK-04.Business and management studies
158	RANK-04.Chemistry
159	RANK-04.Classics
160	RANK-04.Computer sciences and IT
161	RANK-04.Dentistry
162	RANK-04.Drama and dance
163	RANK-04.Earth and marine sciences

164	RANK-04.Economics
165	RANK-04.Education
166	RANK-04.Engineering: chemical
167	RANK-04.Engineering: civil
168	RANK-04.Engineering: electronic and electrical
169	RANK-04.Engineering: general
170	RANK-04.Engineering: materials and mineral
171	RANK-04.Engineering: mechanical
172	RANK-04.English
173	RANK-04.Geography and environmental studies
174	RANK-04.History and history of art
175	RANK-04.Law
176	RANK-04.Mathematics
177	RANK-04.Media studies, communications and librarianship
178	RANK-04.Medicine
179	RANK-04.Modern languages and linguistics
180	RANK-04.Music
181	RANK-04.Nursing and paramedical studies
182	RANK-04.Pharmacy and pharmacology
183	RANK-04.Philosophy
184	RANK-04.Physics
185	RANK-04.Politics
186	RANK-04.Psychology
187	RANK-04.Religious studies and theology
188	RANK-04.Social policy and administration
189	RANK-04.Social work
190	RANK-04.Sociology
191	RANK-04.Sports science
192	RANK-04.Tourism, transport and travel
193	RANK-04.Veterinary science
194	RANK-05.Agriculture, forestry and food
195	RANK-05.American studies
196	RANK-05.Anatomy and physiology
197	RANK-05.Anthropology
198	RANK-05.Archaeology and Forensics
199	RANK-05.Architecture
200	RANK-05.Art and design
201	RANK-05.Biosciences
202	RANK-05.Building and town and country planning
203	RANK-05.Business and management studies
204	RANK-05.Chemistry
205	RANK-05.Classics
206	RANK-05.Computer sciences and IT
207	RANK-05.Dentistry
208	RANK-05.Drama and dance
209	RANK-05.Earth and marine sciences
210	RANK-05.Economics
211	RANK-05.Education
212	RANK-05.Engineering: chemical
213	RANK-05.Engineering: civil
214	RANK-05.Engineering: electronic and electrical
215	RANK-05.Engineering: general
216	RANK-05.Engineering: materials and mineral
217	RANK-05.Engineering: mechanical
218	RANK-05.English
219	RANK-05.Geography and environmental studies
220	RANK-05.History and history of art
221	RANK-05.Law
222	RANK-05.Mathematics
223	RANK-05.Media studies, communications and librarianship
224	RANK-05.Medicine
225	RANK-05.Modern languages and linguistics
226	RANK-05.Music
227	RANK-05.Nursing and paramedical studies
228	RANK-05.Pharmacy and pharmacology
229	RANK-05.Philosophy
230	RANK-05.Physics
231	RANK-05.Politics
232	RANK-05.Psychology
233	RANK-05.Religious studies and theology
234	RANK-05.Social policy and administration
235	RANK-05.Social work

236	RANK-05.Sociology
237	RANK-05.Sports science
238	RANK-05.Tourism, transport and travel
239	RANK-06.Agriculture, forestry and food
240	RANK-06.American studies
241	RANK-06.Anatomy and physiology
242	RANK-06.Anthropology
243	RANK-06.Archaeology and Forensics

244	RANK-06.Architecture
245	RANK-06.Art and design
246	RANK-06.Biosciences
247	RANK-06.Building and town and country planning
248	RANK-06.Business and management studies
249	RANK-06.Chemistry
250	RANK-06.Classics
251	RANK-06.Computer sciences and IT

Таблица 8 – Описательные шкалы и градации (показатели)

Код	Наименование
80	% SATISFIED WITH ASSESSMENT-10/10-{88.8275502, 96.6000000}
79	% SATISFIED WITH ASSESSMENT-9/10-{81.0551004, 88.8275502}
78	% SATISFIED WITH ASSESSMENT-8/10-{73.2826506, 81.0551004}
77	% SATISFIED WITH ASSESSMENT-7/10-{65.5102008, 73.2826506}
76	% SATISFIED WITH ASSESSMENT-6/10-{57.7377510, 65.5102008}
75	% SATISFIED WITH ASSESSMENT-5/10-{49.9653012, 57.7377510}
74	% SATISFIED WITH ASSESSMENT-4/10-{42.1928514, 49.9653012}
73	% SATISFIED WITH ASSESSMENT-3/10-{34.4204016, 42.1928514}
72	% SATISFIED WITH ASSESSMENT-2/10-{26.6479518, 34.4204016}
71	% SATISFIED WITH ASSESSMENT-1/10-{18.8755020, 26.6479518}
70	AVERAGE ENTRY TARIFF-10/10-{551.7000000, 598.0000000}
69	AVERAGE ENTRY TARIFF-9/10-{505.4000000, 551.7000000}
68	AVERAGE ENTRY TARIFF-8/10-{459.1000000, 505.4000000}
67	AVERAGE ENTRY TARIFF-7/10-{412.8000000, 459.1000000}
66	AVERAGE ENTRY TARIFF-6/10-{366.5000000, 412.8000000}
65	AVERAGE ENTRY TARIFF-5/10-{320.2000000, 366.5000000}
64	AVERAGE ENTRY TARIFF-4/10-{273.9000000, 320.2000000}
63	AVERAGE ENTRY TARIFF-3/10-{227.6000000, 273.9000000}
62	AVERAGE ENTRY TARIFF-2/10-{181.3000000, 227.6000000}
61	AVERAGE ENTRY TARIFF-1/10-{135.0000000, 181.3000000}
60	VALUE ADDED SCORE/10-10/10-{9.1000000, 10.0000000}
59	VALUE ADDED SCORE/10-9/10-{8.2000000, 9.1000000}
58	VALUE ADDED SCORE/10-8/10-{7.3000000, 8.2000000}
57	VALUE ADDED SCORE/10-7/10-{6.4000000, 7.3000000}
56	VALUE ADDED SCORE/10-6/10-{5.5000000, 6.4000000}
55	VALUE ADDED SCORE/10-5/10-{4.6000000, 5.5000000}
54	VALUE ADDED SCORE/10-4/10-{3.7000000, 4.6000000}
53	VALUE ADDED SCORE/10-3/10-{2.8000000, 3.7000000}
52	VALUE ADDED SCORE/10-2/10-{1.9000000, 2.8000000}
51	VALUE ADDED SCORE/10-1/10-{1.0000000, 1.9000000}
50	CAREER PROSPECTS-10/10-{91.6000000, 100.0000000}
49	CAREER PROSPECTS-9/10-{83.2000000, 91.6000000}
48	CAREER PROSPECTS-8/10-{74.8000000, 83.2000000}
47	CAREER PROSPECTS-7/10-{66.4000000, 74.8000000}
46	CAREER PROSPECTS-6/10-{58.0000000, 66.4000000}
45	CAREER PROSPECTS-5/10-{49.6000000, 58.0000000}
44	CAREER PROSPECTS-4/10-{41.2000000, 49.6000000}
43	CAREER PROSPECTS-3/10-{32.8000000, 41.2000000}
42	CAREER PROSPECTS-2/10-{24.4000000, 32.8000000}
41	CAREER PROSPECTS-1/10-{16.0000000, 24.4000000}
40	STUDENT:STAFF RATIO-10/10-{46.0900000, 50.7000000}
39	STUDENT:STAFF RATIO-9/10-{41.4800000, 46.0900000}
38	STUDENT:STAFF RATIO-8/10-{36.8700000, 41.4800000}
37	STUDENT:STAFF RATIO-7/10-{32.2600000, 36.8700000}
36	STUDENT:STAFF RATIO-6/10-{27.6500000, 32.2600000}
35	STUDENT:STAFF RATIO-5/10-{23.0400000, 27.6500000}
34	STUDENT:STAFF RATIO-4/10-{18.4300000, 23.0400000}
33	STUDENT:STAFF RATIO-3/10-{13.8200000, 18.4300000}
32	STUDENT:STAFF RATIO-2/10-{9.2100000, 13.8200000}
31	STUDENT:STAFF RATIO-1/10-{4.6000000, 9.2100000}
30	EXPENDITURE PER STUDENT (FTE)-10/10-{9.1000000, 10.0000000}
29	EXPENDITURE PER STUDENT (FTE)-9/10-{8.2000000, 9.1000000}
28	EXPENDITURE PER STUDENT (FTE)-8/10-{7.3000000, 8.2000000}
27	EXPENDITURE PER STUDENT (FTE)-7/10-{6.4000000, 7.3000000}
26	EXPENDITURE PER STUDENT (FTE)-6/10-{5.5000000, 6.4000000}
25	EXPENDITURE PER STUDENT (FTE)-5/10-{4.6000000, 5.5000000}
24	EXPENDITURE PER STUDENT (FTE)-4/10-{3.7000000, 4.6000000}
23	EXPENDITURE PER STUDENT (FTE)-3/10-{2.8000000, 3.7000000}

22	EXPENDITURE PER STUDENT (FTE)-2/10-{1.9000000, 2.8000000}
21	EXPENDITURE PER STUDENT (FTE)-1/10-{1.0000000, 1.9000000}
20	% SATISFIED OVERALL WITH COURSE-10/10-{93.7515677, 100.0000000}
19	% SATISFIED OVERALL WITH COURSE-9/10-{87.5031353, 93.7515677}
18	% SATISFIED OVERALL WITH COURSE-8/10-{81.2547030, 87.5031353}
17	% SATISFIED OVERALL WITH COURSE-7/10-{75.0062706, 81.2547030}
16	% SATISFIED OVERALL WITH COURSE-6/10-{68.7578383, 75.0062706}
15	% SATISFIED OVERALL WITH COURSE-5/10-{62.5094060, 68.7578383}
14	% SATISFIED OVERALL WITH COURSE-4/10-{56.2609736, 62.5094060}
13	% SATISFIED OVERALL WITH COURSE-3/10-{50.0125413, 56.2609736}
12	% SATISFIED OVERALL WITH COURSE-2/10-{43.7641089, 50.0125413}
11	% SATISFIED OVERALL WITH COURSE-1/10-{37.5156766, 43.7641089}
10	% SATISFIED WITH TEACHING-10/10-{95.0000000, 100.0000000}
9	% SATISFIED WITH TEACHING-9/10-{90.0000000, 95.0000000}
8	% SATISFIED WITH TEACHING-8/10-{85.0000000, 90.0000000}
7	% SATISFIED WITH TEACHING-7/10-{80.0000000, 85.0000000}
6	% SATISFIED WITH TEACHING-6/10-{75.0000000, 80.0000000}
5	% SATISFIED WITH TEACHING-5/10-{70.0000000, 75.0000000}
4	% SATISFIED WITH TEACHING-4/10-{65.0000000, 70.0000000}
3	% SATISFIED WITH TEACHING-3/10-{60.0000000, 65.0000000}
2	% SATISFIED WITH TEACHING-2/10-{55.0000000, 60.0000000}
1	% SATISFIED WITH TEACHING-1/10-{50.0000000, 55.0000000}

Таблица 9 – Обучающая выборка (фрагмент)

The object of training sample	Guardian score/100	Rank	Field of study	Name of Institution	% Satisfied with Teaching	% Satisfied overall with course	Expenditure per student (FTE)	Student:staff ratio	Career prospects	Value added score/10	Average Entry Tariff	% Satisfied with Assessment
Medicine-Oxford, 2012	10	41	498	606	10	20		31	50	57	70	79
Medicine-Cambridge, 2012	10	41	498	535	9	19	30	31	50	52	70	76
Medicine-Edinburgh, 2012	9	87	498	553	9	19	30	31	50	54	69	75
Medicine-Dundee, 2012	9	87	498	549	10	20	30	31	50	56	68	76
Medicine-UCL, 2012	8	87	498	652	9	19	26	31	50	59	69	76
Medicine-Imperial College, 2012	6	132	498	575	9	19	27	31	50	53	69	74
Medicine-Leicester, 2012	6	132	498	585	9	19	25	31	50	55	68	75
Medicine-Newcastle, 2012	6	132	498	598	10	20	24	31	50	55	68	75
Medicine-Peninsula Medical School, 2012	6	132	498	608	9	19	27	31	50	54	68	76
Medicine-Nottingham, 2012	6	178	498	604	9	19	23	31	50	54	69	74
Medicine-King's College London, 2012	5	178	498	578	8	18	25	31	50	56	68	74
Medicine-Warwick, 2012	5	178	498	660	8	18	28	31	50	59		74
Medicine-Leeds, 2012	4	224	498	581	8	18	26	31	50	56	68	75
Medicine-Hull York Medical School, 2012	4	224	498	574	8	18	24	32	50	58	68	75
Medicine-Manchester, 2012	4	224	498	593	6	15	28	31	50	59	68	73
Medicine-Sheffield, 2012	4	224	498	628	9	19	23	31	50	54	68	75
Medicine-Aberdeen, 2012	4	269	498	513	9	19	24	31	50	56	67	77
Medicine-Brighton Sussex Medical School, 2012	4	269	498	530	8	18	24	32	50	57		75
Medicine-Queen Mary, 2012	4	269	498	613	7	17	24	31	50	58	68	74
Medicine-St George's Medical School, 2012	4	315	498	634	8	19	26	32	50	56	68	75
Medicine-Southampton, 2012	4	315	498	631	8	18	24	31	50	57	68	74
Medicine-St Andrews, 2012	4	361	498	633	10	20	22	32	50	51	68	76
Medicine-Glasgow, 2012	3	361	498	560	6	15	25	31	50	55	69	72
Medicine-UEA, 2012	3	361	498	653	8	18	23	31	50	59	67	75
Medicine-Birmingham, 2012	3	361	498	523	9	18	23	32	50	53	69	72
Medicine-Queen's, Belfast, 2012	3	406	498	611	8	18	24	32	50	53	68	75
Medicine-Liverpool, 2012	2	406	498	587	6	15	24	31	50	54	68	72
Medicine-Bristol, 2012	2	406	498	531	8	15	26	32	50	54	68	71
Medicine-Keele, 2012	1	452	498	576	7	14	23	32	50	57	67	72
Medicine-Cardiff, 2012	1	452	498	537	6	14	23	32	50	58	68	71
Dentistry-King's College London, 2012	10	24	481	578	9	18	28	31	50	57	68	77
Dentistry-Glasgow, 2012	8	70	481	560	10	20	23	32	50	56	68	78
Dentistry-Cardiff, 2012	8	115	481	537	9	20	28	31	50	51	68	75
Dentistry-Queen's, Belfast, 2012	7	115	481	611	10	20	29	31	50	55	67	76
Dentistry-Birmingham, 2012	7	161	481	523	10	18	25	31	50	56	68	76
Dentistry-Bristol, 2012	6	207	481	531	9	20	26	32	50	55	68	78
Dentistry-Dundee, 2012	5	252	481	549	8	19	24	31	50	57	68	75
Dentistry-Sheffield, 2012	4	252	481	628	9	19	23	31	50	56	68	76
Dentistry-Liverpool, 2012	3	298	481	587	8	17	27	31	50	57	68	76
Dentistry-Manchester, 2012	3	344	481	593	9	18	25	31	50	57	68	75
Dentistry-Newcastle, 2012	2	389	481	598	8	19	21	32	50	56	68	76
Dentistry-Queen Mary, 2012	1	389	481	613	7	17	24	31	50	57	68	76
Dentistry-Leeds, 2012	1	435	481	581	9	19	29	32	50	56	68	73
Veterinary science-Cambridge, 2012	10	56	512	535	9	19	28	31	49	53	69	77
Veterinary science-Edinburgh, 2012	8	147	512	553	9	18	29	31	50	59	68	74
Veterinary science-Liverpool, 2012	6	193	512	587	9	20	23	31	50	54	68	74
Veterinary science-Glasgow, 2012	5	284	512	560	9	20	23	31	50	58	68	75
Veterinary science-Nottingham, 2012	5	330	512	604			24	31				68
Veterinary science-Royal Veterinary College, 2012	5	421	512	625	8	17	28	31	49	56	68	73
Veterinary science-Bristol, 2012	1	467	512	531	9	19	23	32	49	53	67	74
Anatomy and physiology-Oxford, 2012	10	13	470	606	10	17	30	32		56	69	75
Anatomy and physiology-Glamorgan, 2012	10	13	470	559	9	19	24	32	50	60	65	78
Anatomy and physiology-Cardiff, 2012	10	59	470	537	9	20	30	32	50	55	67	75
Anatomy and physiology-Plymouth, 2012	9	59	470	609	10	19	28	31	48	58	65	79
Anatomy and physiology-Brunel, 2012	8	59	470	532	10	20	23	32	48	59	65	76
Anatomy and physiology-Liverpool, 2012	8	59	470	587	9	19	29	32	48	54	66	77
Anatomy and physiology-Sussex, 2012	8	59	470	643	9	20	26	32		59	66	76
Anatomy and physiology-Newcastle, 2012	7	104	470	598	9	20	24	32	48	55	67	76
Anatomy and physiology-Aston, 2012	7	104	470	518	8	18	26	33	50	52	66	78
Anatomy and physiology-Bristol, 2012	7	150	470	531	9	19	27	32	47	54	67	76
Anatomy and physiology-Nottingham, 2012	6	150	470	604	10	19	27	33		54	67	77
Anatomy and physiology-Birmingham, 2012	6	150	470	523	9	18		32	47	57	67	75
Anatomy and physiology-Sheffield Hallam, 2012	5	150	470	629	10	19	23	33	48	55	66	77

Anatomy and physiology-Manchester, 2012	5	196	470	593	9	19	25	32	48	54	67	75
Anatomy and physiology-Glasgow Caledonian, 2012	5	196	470	561	9	19	24	33	48	57	66	76
Anatomy and physiology-Edinburgh, 2012	5	196	470	553	9	19	29	33	45	55	67	74
Anatomy and physiology-Robert Gordon, 2012	5	241	470	616	9	20	23	34	48	57	67	76
Anatomy and physiology-De Montfort, 2012	5	241	470	547	8	19	25	33	50	55	64	76
Anatomy and physiology-Hertfordshire, 2012	5	241	470	570	8	19	23	33	49	57	63	77
Anatomy and physiology-UEA, 2012	5	241	470	653	10	20			49	53	65	75
Anatomy and physiology-Queen's, Belfast, 2012	4	287	470	611	9	19	29	32	46	57	65	74
Anatomy and physiology-Aberdeen, 2012	4	287	470	513	8	20	24	33	47	56	64	78
Anatomy and physiology-Bradford, 2012	4	287	470	528	8	20	24	33	49	53	65	76
Anatomy and physiology-Birmingham City, 2012	2	287	470	524	9	15	23	34	48	58	65	75
Anatomy and physiology-Leeds, 2012	2	333	470	581	8	19	25	35	45	56	67	76
Anatomy and physiology-Northampton, 2012	2	333	470	601	9	17	23	33			63	77
Anatomy and physiology-Anglia Ruskin, 2012	2	333	470	516	8	20	25	36	46	58	64	78
Anatomy and physiology-Manchester Met, 2012	2	378	470	594	8	17	22	33	48	53	65	75
Anatomy and physiology-City, 2012	1	378	470	542	8	17	24	33	50	51	66	74
Anatomy and physiology-Cumbria, 2012	1	378	470	546	7	18	23	34	50	52	64	76
Anatomy and physiology-St Mary's UC, Twickenham, 2012	1	378	470	636	10	20	22	35	47	56	62	77
Anatomy and physiology-King's College London, 2012	1	424	470	578	8	19	27	36	46	53	67	75
Anatomy and physiology-Ulster, 2012	1	424	470	654	7	15	23	33	45	58	65	75
Nursing and paramedical studies-Edinburgh, 2012	10	44	501	553	10	20	29	32		58	66	78
Nursing and paramedical studies-Glasgow, 2012	10	44	501	560	10	20	27	33		56	66	80
Nursing and paramedical studies-UEA, 2012	9	44	501	653	9	18	30	33	50	58	65	76
Nursing and paramedical studies-Leeds, 2012	7	44	501	581	8	17	30	33	50	53	65	77
Nursing and paramedical studies-Staffordshire, 2012	7	44	501	637	9	20	27	33	50	58	63	78
Nursing and paramedical studies-Portsmouth, 2012	7	90	501	610	8	18	28	32	50	54	65	76
Nursing and paramedical studies-City, 2012	7	90	501	542	8	18	30	33	49	55	64	77
Nursing and paramedical studies-Keele, 2012	7	90	501	576	10	20	26	33	49	57	64	78
Nursing and paramedical studies-Southampton, 2012	7	90	501	631	8	18	28	33	49	59	66	75
Nursing and paramedical studies-Birmingham, 2012	7	90	501	523	9	19		32	49	54	66	76
Nursing and paramedical studies-Bedfordshire, 2012	7	90	501	522	9	18	24	33		60	63	77
Nursing and paramedical studies-Liverpool, 2012	7	90	501	587	8	17	29	32	50	52	65	76
Nursing and paramedical studies-Oxford Brookes, 2012	7	90	501	607	9	19	24	33	50	55	64	78
Nursing and paramedical studies-Nottingham, 2012	7	135	501	604	8	18	27	33	49	59	65	76
Nursing and paramedical studies-Surrey, 2012	7	135	501	642	8	19	28	35	50	58	65	77
Nursing and paramedical studies-Manchester, 2012	7	135	501	593	9	18	27	33	50	55	65	76
Nursing and paramedical studies-Brighton, 2012	6	135	501	529	8	17	24	33	49	58	65	77
Nursing and paramedical studies-Thames Valley, 2012	6	135	501	646	8	16	28	33	50	60	62	77
Nursing and paramedical studies-Middlesex, 2012	6	135	501	597	8	17	29	33	50	56	63	77
Nursing and paramedical studies-Edge Hill, 2012	6	135	501	552	9	19	23	33	49	54	63	79
Nursing and paramedical studies-Bangor, 2012	6	135	501	519	8	18	26	33	50	52	64	77
Nursing and paramedical studies-Coventry, 2012	6	135	501	545	9	19	23	33	49	58	64	77
Nursing and paramedical studies-Northampton, 2012	6	181	501	601	8	18	24	33	49	59	63	76

Полностью обучающая выборка в статье не может быть приведена, т.к. файл исходных данных содержит 2559 строк.

Этим завершается 2-й этап АСК-анализа, который называется «Формализация предметной области» и создаются все необходимые и достаточные предпосылки для выполнения следующего этапа, т.е. синтеза и верификации (измерения достоверности) модели.

6.4.3.5. Синтез и верификация многокритериальной системно-когнитивной модели университетского рейтинга Гардиан, учитывающей направления подготовки

Синтез и верификация многокритериальной системно-когнитивной модели университетского рейтинга Гардиан, учитывающей направления подготовки, представляет собой задачу, требующую довольно значительных вычислительных ресурсов. Решение этой задачи на компьютере с процессором i7 и 16 Гб оперативной памяти с размещением задачи на SSD, потребовало около 13 часов счета (рисунок 9).

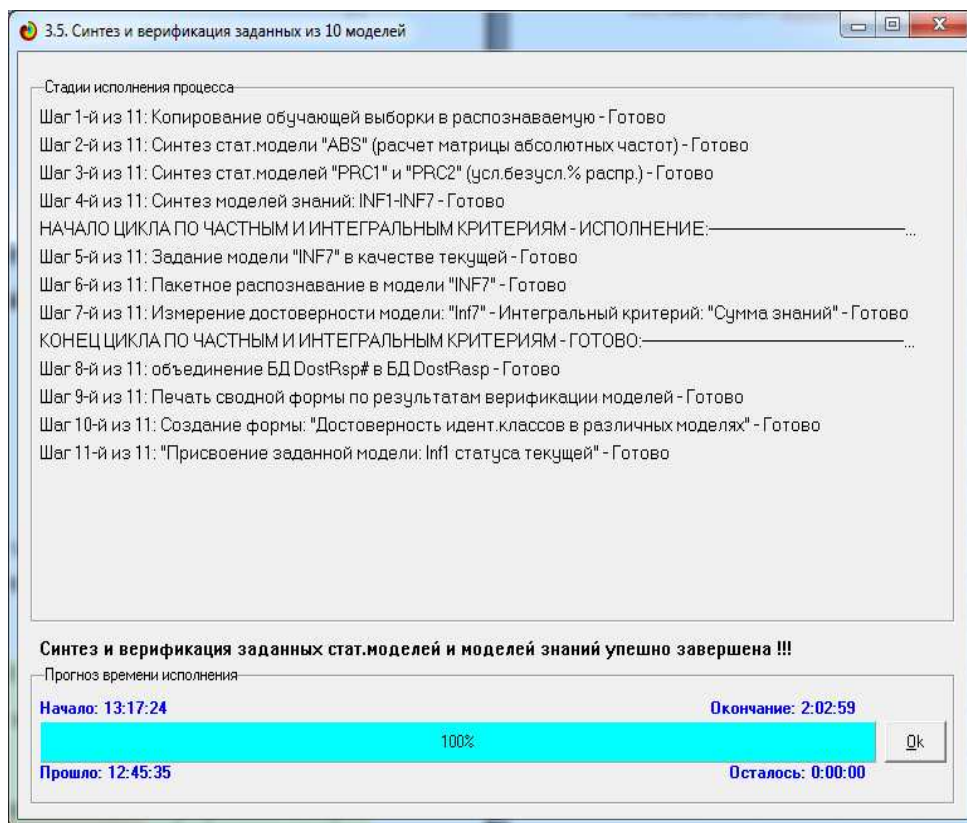


Рисунок 9 – Экранная форма с отображением стадии синтеза и верификации моделей и прогнозом времени исполнения

Такая большая длительность расчетов обусловлена тем, что для измерения достоверности 10 моделей была использована вся обучающая выборка, включающая 2559 примеров.

Математические аспекты формирования системно-когнитивных моделей описаны в ряде работ автора [3] и здесь их подробно освещать нет необходимости. Отметим лишь, что для преобразования матрицы абсолютных частот в другие модели используются формулы преобразования, приведенные в таблице 10:

Таблица 10 – Частные критерии знаний, используемые в настоящее время в АСК-анализе и системе «Эйдос-Х++»

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета относительных частот: N_j – суммарное количество признаков по j -му классу. Относительная частота того, что если у объекта j -го класса обнаружен признак, то это i -й признак	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_i N_j}$
INF2 , частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета относительных частот: N_j – суммарное количество объектов по j -му классу. Относительная частота того, что если предъявлен объект j -го класса, то у него будет	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_i N_j}$

обнаружен i -й признак.		
INF3 , частный критерий: Хи-квадрат: разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \frac{N_i N_j}{N}$
INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета относительных частот: N_j – суммарное количество признаков по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij} N}{N_i N_j} - 1$
INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета относительных частот: N_j – суммарное количество объектов по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij} N}{N_i N_j} - 1$
INF6 , частный критерий: разность условной и безусловной относительных частот, 1-й вариант расчета относительных частот: N_j – суммарное количество признаков по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7 , частный критерий: разность условной и безусловной относительных частот, 2-й вариант расчета относительных частот: N_j – суммарное количество объектов по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Обозначения:

i – значение прошлого параметра;

j - значение будущего параметра;

N_{ij} – количество встреч j -го значения будущего параметра при i -м значении прошлого параметра;

M – суммарное число значений всех прошлых параметров;

W - суммарное число значений всех будущих параметров.

N_i – количество встреч i -м значения прошлого параметра по всей выборке;

N_j – количество встреч j -го значения будущего параметра по всей выборке;

N – количество встреч j -го значения будущего параметра при i -м значении прошлого параметра по всей выборке.

I_{ij} – частный критерий знаний: количество знаний в факте наблюдения i -го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее j -му значению будущего параметра;

Ψ – нормировочный коэффициент (Е.В.Луценко, 1979, впервые опубликовано в 1993 году [15]), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;

P_i – безусловная относительная частота встречи i -го значения прошлого параметра в обучающей выборке;

P_{ij} – условная относительная частота встречи i -го значения прошлого параметра при j -м значении будущего параметра.

В результате сформированы 10 моделей: 3 статистических и 7 системно-когнитивных моделей (моделей знаний). Фрагменты трех из них приведены ниже (таблицы 11, 12, 13):

Таблица 11 – Матрица абсолютных частот, модель ABS (фрагмент)

Код	Наименование показателя	1/10-{25.90, 33.31}	2/10-{33.31, 40.72}	3/10-{40.72, 48.13}	4/10-{48.13, 55.54}	5/10-{55.54, 62.95}	6/10-{62.95, 70.36}	7/10-{70.36, 77.77}	8/10-{77.77, 85.18}	9/10-{85.18, 92.59}	10/10-{92.59, 100.00}
1	% SATISFIED WITH TEACHING-1/10-{50.0000000, 55.0000000}	2	2	1	0	2	0	1	0	0	0
2	% SATISFIED WITH TEACHING-2/10-{55.0000000, 60.0000000}	4	8	0	1	1	0	1	0	0	0
3	% SATISFIED WITH TEACHING-3/10-{60.0000000, 65.0000000}	3	8	7	5	2	4	2	2	1	0
4	% SATISFIED WITH TEACHING-4/10-{65.0000000, 70.0000000}	7	23	17	18	9	4	6	3	1	0
5	% SATISFIED WITH TEACHING-5/10-{70.0000000, 75.0000000}	6	22	31	43	30	29	9	5	0	0
6	% SATISFIED WITH TEACHING-6/10-{75.0000000, 80.0000000}	17	29	63	72	79	43	22	12	5	3
7	% SATISFIED WITH TEACHING-7/10-{80.0000000, 85.0000000}	15	22	65	93	108	89	53	43	20	12
8	% SATISFIED WITH TEACHING-8/10-{85.0000000, 90.0000000}	6	21	55	96	121	121	101	41	30	27
9	% SATISFIED WITH TEACHING-9/10-{90.0000000, 95.0000000}	6	9	25	54	87	104	97	71	38	35
10	% SATISFIED WITH TEACHING-10/10-{95.0000000, 100.0000000}	1	2	3	12	23	18	39	23	21	32
11	% SATISFIED OVERALL WITH COURSE-1/10-{37.5156766, 43.7641089}	3	3	2	1	0	0	0	1	1	0
12	% SATISFIED OVERALL WITH COURSE-2/10-{43.7641089, 50.0125413}	3	5	1	1	2	1	0	0	1	0
13	% SATISFIED OVERALL WITH COURSE-3/10-{50.0125413, 56.2609736}	4	7	5	4	6	2	2	1	0	0
14	% SATISFIED OVERALL WITH COURSE-4/10-{56.2609736, 62.5094060}	6	16	17	11	13	6	5	3	0	0
15	% SATISFIED OVERALL WITH COURSE-5/10-{62.5094060, 68.7578383}	9	20	20	31	21	19	5	6	0	0
16	% SATISFIED OVERALL WITH COURSE-6/10-{68.7578383, 75.0062706}	9	23	41	45	45	33	19	11	4	0
17	% SATISFIED OVERALL WITH COURSE-7/10-{75.0062706, 81.2547030}	16	28	65	81	96	73	45	16	9	2
18	% SATISFIED OVERALL WITH COURSE-8/10-{81.2547030, 87.5031353}	8	26	66	109	118	110	80	42	25	12
19	% SATISFIED OVERALL WITH COURSE-9/10-{87.5031353, 93.7515677}	5	15	40	84	118	114	111	73	43	44
20	% SATISFIED OVERALL WITH COURSE-10/10-{93.7515677, 100.0000000}	4	3	10	27	43	54	64	47	33	51
21	EXPENDITURE PER STUDENT (FTE)-1/10-{1.0000000, 1.9000000}	2	4	3	0	1	0	0	0	0	0
22	EXPENDITURE PER STUDENT (FTE)-2/10-{1.9000000, 2.8000000}	9	28	44	33	23	13	9	0	0	0
23	EXPENDITURE PER STUDENT (FTE)-3/10-{2.8000000, 3.7000000}	31	48	84	114	102	64	20	13	5	2
24	EXPENDITURE PER STUDENT (FTE)-4/10-{3.7000000, 4.6000000}	17	35	63	111	121	86	51	21	2	2
25	EXPENDITURE PER STUDENT (FTE)-5/10-{4.6000000, 5.5000000}	6	15	38	47	67	63	58	19	7	4
26	EXPENDITURE PER STUDENT (FTE)-6/10-{5.5000000, 6.4000000}	0	9	18	35	54	59	42	28	14	6
27	EXPENDITURE PER STUDENT (FTE)-7/10-{6.4000000, 7.3000000}	2	6	8	28	39	52	47	29	14	8
28	EXPENDITURE PER STUDENT (FTE)-8/10-{7.3000000, 8.2000000}	0	5	13	14	35	40	48	32	19	15
29	EXPENDITURE PER STUDENT (FTE)-9/10-{8.2000000, 9.1000000}	1	0	0	12	22	25	33	35	23	11
30	EXPENDITURE PER STUDENT (FTE)-10/10-{9.1000000, 10.0000000}	0	0	1	3	4	9	24	23	31	56
31	STUDENT:STAFF RATIO-1/10-{4.6000000, 9.2100000}	1	2	4	11	16	13	17	17	18	20
32	STUDENT:STAFF RATIO-2/10-{9.2100000, 13.8200000}	7	7	9	29	44	62	73	69	46	50
33	STUDENT:STAFF RATIO-3/10-{13.8200000, 18.4300000}	7	24	53	101	160	150	131	69	43	37
34	STUDENT:STAFF RATIO-4/10-{18.4300000, 23.0400000}	15	36	90	144	167	122	93	33	12	6
35	STUDENT:STAFF RATIO-5/10-{23.0400000, 27.6500000}	22	34	74	79	59	47	16	11	2	2
36	STUDENT:STAFF RATIO-6/10-{27.6500000, 32.2600000}	8	24	20	23	19	8	1	1	1	0
37	STUDENT:STAFF RATIO-7/10-{32.2600000, 36.8700000}	4	12	10	6	2	4	1	0	0	0
38	STUDENT:STAFF RATIO-8/10-{36.8700000, 41.4800000}	2	5	9	3	1	0	0	0	0	0
39	STUDENT:STAFF RATIO-9/10-{41.4800000, 46.0900000}	2	3	2	0	0	1	0	0	0	0
40	STUDENT:STAFF RATIO-10/10-{46.0900000, 50.7000000}	0	1	0	0	0	0	0	0	0	0
41	CAREER PROSPECTS-1/10-{16.0000000, 24.4000000}	2	5	5	5	2	0	0	0	0	0
42	CAREER PROSPECTS-2/10-{24.4000000, 32.8000000}	5	14	18	21	21	4	0	0	0	1
43	CAREER PROSPECTS-3/10-{32.8000000, 41.2000000}	12	15	41	61	57	21	14	0	1	1
44	CAREER PROSPECTS-4/10-{41.2000000, 49.6000000}	6	32	36	65	58	36	31	7	3	0
45	CAREER PROSPECTS-5/10-{49.6000000, 58.0000000}	8	19	32	66	67	89	51	19	5	1
46	CAREER PROSPECTS-6/10-{58.0000000, 66.4000000}	3	7	24	40	65	51	55	21	14	8
47	CAREER PROSPECTS-7/10-{66.4000000, 74.8000000}	1	3	10	30	28	55	47	45	27	14
48	CAREER PROSPECTS-8/10-{74.8000000, 83.2000000}	0	3	11	10	26	30	42	36	21	29
49	CAREER PROSPECTS-9/10-{83.2000000, 91.6000000}	2	1	1	4	14	10	19	19	14	19
50	CAREER PROSPECTS-10/10-{91.6000000, 100.0000000}	6	3	9	20	14	16	16	10	7	9
51	VALUE ADDED SCORE/10-1/10-{1.0000000, 1.9000000}	15	21	20	24	12	3	2	2	1	0

Таблица 11 – Матрица условных и безусловных процентных распределений, модель PRC2 (фрагмент)

Код	Наименование показателя	1/10-{25.90, 33.31}	2/10-{33.31, 40.72}	3/10-{40.72, 48.13}	4/10-{48.13, 55.54}	5/10-{55.54, 62.95}	6/10-{62.95, 70.36}	7/10-{70.36, 77.77}	8/10-{77.77, 85.18}	9/10-{85.18, 92.59}	10/10-{92.59, 100.00}
1	% SATISFIED WITH TEACHING-1/10-{50.0000000, 55.0000000}	3	1	0	0	0	0	0	0	0	0
2	% SATISFIED WITH TEACHING-2/10-{55.0000000, 60.0000000}	6	5	0	0	0	0	0	0	0	0
3	% SATISFIED WITH TEACHING-3/10-{60.0000000, 65.0000000}	4	5	3	1	0	1	1	1	1	0
4	% SATISFIED WITH TEACHING-4/10-{65.0000000, 70.0000000}	10	15	6	5	2	1	2	1	1	0
5	% SATISFIED WITH TEACHING-5/10-{70.0000000, 75.0000000}	9	15	11	11	6	7	3	2	0	0
6	% SATISFIED WITH TEACHING-6/10-{75.0000000, 80.0000000}	25	19	23	18	17	10	7	6	4	3
7	% SATISFIED WITH TEACHING-7/10-{80.0000000, 85.0000000}	22	15	24	23	23	21	16	21	16	10
8	% SATISFIED WITH TEACHING-8/10-{85.0000000, 90.0000000}	9	14	20	24	26	29	30	20	24	23
9	% SATISFIED WITH TEACHING-9/10-{90.0000000, 95.0000000}	9	6	9	14	18	25	29	35	31	30
10	% SATISFIED WITH TEACHING-10/10-{95.0000000, 100.0000000}	1	1	1	3	5	4	12	11	17	28
11	% SATISFIED OVERALL WITH COURSE-1/10-{37.5156766, 43.7641089}	4	2	1	0	0	0	0	0	1	0
12	% SATISFIED OVERALL WITH COURSE-2/10-{43.7641089, 50.0125413}	4	3	0	0	0	0	0	0	1	0
13	% SATISFIED OVERALL WITH COURSE-3/10-{50.0125413, 56.2609736}	6	5	2	1	1	0	1	0	0	0
14	% SATISFIED OVERALL WITH COURSE-4/10-{56.2609736, 62.5094060}	9	11	6	3	3	1	1	1	0	0
15	% SATISFIED OVERALL WITH COURSE-5/10-{62.5094060, 68.7578383}	13	13	7	8	4	5	1	3	0	0
16	% SATISFIED OVERALL WITH COURSE-6/10-{68.7578383, 75.0062706}	13	15	15	11	10	8	6	5	3	0
17	% SATISFIED OVERALL WITH COURSE-7/10-{75.0062706, 81.2547030}	23	19	24	20	20	18	13	8	7	2
18	% SATISFIED OVERALL WITH COURSE-8/10-{81.2547030, 87.5031353}	12	17	24	27	25	27	24	21	20	10
19	% SATISFIED OVERALL WITH COURSE-9/10-{87.5031353, 93.7515677}	7	10	15	21	25	27	33	36	35	38
20	% SATISFIED OVERALL WITH COURSE-10/10-{93.7515677, 100.0000000}	6	2	4	7	9	13	19	23	27	44
21	EXPENDITURE PER STUDENT (FTE)-1/10-{1.0000000, 1.9000000}	3	3	1	0	0	0	0	0	0	0
22	EXPENDITURE PER STUDENT (FTE)-2/10-{1.9000000, 2.8000000}	13	19	16	8	5	3	3	0	0	0
23	EXPENDITURE PER STUDENT (FTE)-3/10-{2.8000000, 3.7000000}	45	32	31	29	22	15	6	6	4	2
24	EXPENDITURE PER STUDENT (FTE)-4/10-{3.7000000, 4.6000000}	25	23	23	28	26	21	15	10	2	2
25	EXPENDITURE PER STUDENT (FTE)-5/10-{4.6000000, 5.5000000}	9	10	14	12	14	15	17	9	6	3
26	EXPENDITURE PER STUDENT (FTE)-6/10-{5.5000000, 6.4000000}	0	6	7	9	11	14	13	14	11	5
27	EXPENDITURE PER STUDENT (FTE)-7/10-{6.4000000, 7.3000000}	3	4	3	7	8	13	14	14	11	7
28	EXPENDITURE PER STUDENT (FTE)-8/10-{7.3000000, 8.2000000}	0	3	5	4	7	10	14	16	15	13
29	EXPENDITURE PER STUDENT (FTE)-9/10-{8.2000000, 9.1000000}	1	0	0	3	5	6	10	17	19	9
30	EXPENDITURE PER STUDENT (FTE)-10/10-{9.1000000, 10.0000000}	0	0	0	1	1	2	7	11	25	48
31	STUDENT:STAFF RATIO-1/10-{4.6000000, 9.2100000}	1	1	1	3	3	3	5	8	15	17
32	STUDENT:STAFF RATIO-2/10-{9.2100000, 13.8200000}	10	5	3	7	9	15	22	34	37	43
33	STUDENT:STAFF RATIO-3/10-{13.8200000, 18.4300000}	10	16	19	25	34	36	39	34	35	32
34	STUDENT:STAFF RATIO-4/10-{18.4300000, 23.0400000}	22	24	33	36	35	29	28	16	10	5
35	STUDENT:STAFF RATIO-5/10-{23.0400000, 27.6500000}	32	23	27	20	12	11	5	5	2	2
36	STUDENT:STAFF RATIO-6/10-{27.6500000, 32.2600000}	12	16	7	6	4	2	0	0	1	0
37	STUDENT:STAFF RATIO-7/10-{32.2600000, 36.8700000}	6	8	4	2	0	1	0	0	0	0
38	STUDENT:STAFF RATIO-8/10-{36.8700000, 41.4800000}	3	3	3	1	0	0	0	0	0	0
39	STUDENT:STAFF RATIO-9/10-{41.4800000, 46.0900000}	3	2	1	0	0	0	0	0	0	0
40	STUDENT:STAFF RATIO-10/10-{46.0900000, 50.7000000}	0	1	0	0	0	0	0	0	0	0
41	CAREER PROSPECTS-1/10-{16.0000000, 24.4000000}	3	3	2	1	0	0	0	0	0	0
42	CAREER PROSPECTS-2/10-{24.4000000, 32.8000000}	7	9	7	5	4	1	0	0	0	1
43	CAREER PROSPECTS-3/10-{32.8000000, 41.2000000}	17	10	15	15	12	5	4	0	1	1
44	CAREER PROSPECTS-4/10-{41.2000000, 49.6000000}	9	21	13	16	12	9	9	3	2	0
45	CAREER PROSPECTS-5/10-{49.6000000, 58.0000000}	12	13	12	17	14	21	15	9	4	1
46	CAREER PROSPECTS-6/10-{58.0000000, 66.4000000}	4	5	9	10	14	12	16	10	11	7
47	CAREER PROSPECTS-7/10-{66.4000000, 74.8000000}	1	2	4	8	6	13	14	22	22	12
48	CAREER PROSPECTS-8/10-{74.8000000, 83.2000000}	0	2	4	3	5	7	13	18	17	25
49	CAREER PROSPECTS-9/10-{83.2000000, 91.6000000}	3	1	0	1	3	2	6	9	11	16
50	CAREER PROSPECTS-10/10-{91.6000000, 100.0000000}	9	2	3	5	3	4	5	5	6	8
51	VALUE ADDED SCORE/10-1/10-{1.0000000, 1.9000000}	22	14	7	6	3	1	1	1	1	0
52	VALUE ADDED SCORE/10-2/10-{1.9000000, 2.8000000}	14	13	10	10	7	6	4	1	0	3
53	VALUE ADDED SCORE/10-3/10-{2.8000000, 3.7000000}	13	15	17	15	11	8	7	3	3	3
54	VALUE ADDED SCORE/10-4/10-{3.7000000, 4.6000000}	6	17	19	16	12	13	10	11	11	3
55	VALUE ADDED SCORE/10-5/10-{4.6000000, 5.5000000}	6	7	8	10	12	13	14	8	7	9
56	VALUE ADDED SCORE/10-6/10-{5.5000000, 6.4000000}	10	11	8	8	14	13	11	14	11	9
57	VALUE ADDED SCORE/10-7/10-{6.4000000, 7.3000000}	7	6	8	10	15	16	19	24	25	28
58	VALUE ADDED SCORE/10-8/10-{7.3000000, 8.2000000}	6	6	8	9	10	12	18	18	24	25
59	VALUE ADDED SCORE/10-9/10-{8.2000000, 9.1000000}	1	2	3	7	7	7	9	12	10	11
60	VALUE ADDED SCORE/10-10/10-{9.1000000, 10.0000000}	0	1	1	1	3	4	4	3	4	6

61	AVERAGE ENTRY TARIFF-1/10-{135.0000000, 181.3000000}	4	5	4	2	0	1	1	0	0	0
62	AVERAGE ENTRY TARIFF-2/10-{181.3000000, 227.6000000}	30	29	21	10	9	5	4	1	2	1
63	AVERAGE ENTRY TARIFF-3/10-{227.6000000, 273.9000000}	38	37	35	35	26	18	13	5	7	1
64	AVERAGE ENTRY TARIFF-4/10-{273.9000000, 320.2000000}	10	19	25	27	29	25	12	10	10	2
65	AVERAGE ENTRY TARIFF-5/10-{320.2000000, 366.5000000}	6	3	7	12	15	16	19	15	14	9
66	AVERAGE ENTRY TARIFF-6/10-{366.5000000, 412.8000000}	3	2	4	6	10	17	23	19	20	13
67	AVERAGE ENTRY TARIFF-7/10-{412.8000000, 459.1000000}	4	2	1	3	6	12	21	24	22	19
68	AVERAGE ENTRY TARIFF-8/10-{459.1000000, 505.4000000}	4	2	1	3	2	3	6	16	15	14
69	AVERAGE ENTRY TARIFF-9/10-{505.4000000, 551.7000000}	0	0	1	0	1	1	1	5	7	25

Таблица 12 – Матрица информативностей в модели модель INF1,
мера информации по А.Харкевичу в миллибитах (фрагмент)

Код	Наименование показателя	1/10-{25.90, 33.31}	2/10-{33.31, 40.72}	3/10-{40.72, 48.13}	4/10-{48.13, 55.54}	5/10-{55.54, 62.95}	6/10-{62.95, 70.36}	7/10-{70.36, 77.77}	8/10-{77.77, 85.18}	9/10-{85.18, 92.59}	10/10-{92.59, 100.00}
1	% SATISFIED WITH TEACHING-1/10-{50.0000000, 55.0000000}	1876	1219	140		252		-55			
2	% SATISFIED WITH TEACHING-2/10-{55.0000000, 60.0000000}	1930	1850		-718	-850		-579			
3	% SATISFIED WITH TEACHING-3/10-{60.0000000, 65.0000000}	1009	1168	556	-59	-954	-272	-683	-254	-394	
4	% SATISFIED WITH TEACHING-4/10-{65.0000000, 70.0000000}	922	1256	503	216	-493	-1064	-560	-709	-1186	
5	% SATISFIED WITH TEACHING-5/10-{70.0000000, 75.0000000}	221	646	431	369	-63	14	-795	-856		
6	% SATISFIED WITH TEACHING-6/10-{75.0000000, 80.0000000}	523	310	456	233	178	-224	-616	-692	-984	-1349
7	% SATISFIED WITH TEACHING-7/10-{80.0000000, 85.0000000}	77	-262	140	104	97	41	-225	29	-170	-536
8	% SATISFIED WITH TEACHING-8/10-{85.0000000, 90.0000000}	-832	-446	-144	-15	47	151	167	-156	22	-5
9	% SATISFIED WITH TEACHING-9/10-{90.0000000, 95.0000000}	-696	-1016	-665	-359	-93	161	269	438	355	347
10	% SATISFIED WITH TEACHING-10/10-{95.0000000, 100.0000000}	-1267	-1348	-1510	-690	-279	-379	432	420	782	1194
11	% SATISFIED OVERALL WITH COURSE-1/10-{37.5156766, 43.7641089}	1949	1291	453	-460				108	546	
12	% SATISFIED OVERALL WITH COURSE-2/10-{43.7641089, 50.0125413}	1748	1516	-326	-661	-215	-688			345	
13	% SATISFIED OVERALL WITH COURSE-3/10-{50.0125413, 56.2609736}	1325	1134	353	-168	38	-772	-606	-755		
14	% SATISFIED OVERALL WITH COURSE-4/10-{56.2609736, 62.5094060}	905	1065	614	-83	-75	-615	-601	-598		
15	% SATISFIED OVERALL WITH COURSE-5/10-{62.5094060, 68.7578383}	800	808	307	337	-119	-97	-1043	-463		
16	% SATISFIED OVERALL WITH COURSE-6/10-{68.7578383, 75.0062706}	331	455	436	179	47	-106	-400	-427	-832	
17	% SATISFIED OVERALL WITH COURSE-7/10-{75.0062706, 81.2547030}	287	96	297	145	155	32	-205	-638	-679	-1872
18	% SATISFIED OVERALL WITH COURSE-8/10-{81.2547030, 87.5031353}	-561	-236	39	123	57	103	4	-104	-98	-649
19	% SATISFIED OVERALL WITH COURSE-9/10-{87.5031353, 93.7515677}	-1021	-763	-446	-163	-11	65	209	288	285	365
20	% SATISFIED OVERALL WITH COURSE-10/10-{93.7515677, 100.0000000}	-661	-1558	-1055	-563	-306	-12	296	467	611	1034
21	EXPENDITURE PER STUDENT (FTE)-1/10-{1.0000000, 1.9000000}	1690	1610	870		-512					
22	EXPENDITURE PER STUDENT (FTE)-2/10-{1.9000000, 2.8000000}	639	927	803	228	-204	-575	-715			
23	EXPENDITURE PER STUDENT (FTE)-3/10-{2.8000000, 3.7000000}	743	450	416	335	111	-173	-976	-906	-1264	-1967
24	EXPENDITURE PER STUDENT (FTE)-4/10-{3.7000000, 4.6000000}	199	143	132	269	210	30	-239	-550	-2071	-2011
25	EXPENDITURE PER STUDENT (FTE)-5/10-{4.6000000, 5.5000000}	-292	-187	87	-70	94	147	244	-257	-651	-1057
26	EXPENDITURE PER STUDENT (FTE)-6/10-{5.5000000, 6.4000000}		-445	-368	-149	81	260	143	234	94	-552
27	EXPENDITURE PER STUDENT (FTE)-7/10-{6.4000000, 7.3000000}	-933	-675	-936	-227	-83	262	344	370	201	-205
28	EXPENDITURE PER STUDENT (FTE)-8/10-{7.3000000, 8.2000000}		-783	-488	-761	-129	87	405	496	500	363
29	EXPENDITURE PER STUDENT (FTE)-9/10-{8.2000000, 9.1000000}	-1208			-631	-257	-46	352	830	918	364
30	EXPENDITURE PER STUDENT (FTE)-10/10-{9.1000000, 10.0000000}			-2308	-1727	-1619	-838	145	538	1225	1778
31	STUDENT:STAFF RATIO-1/10-{4.6000000, 9.2100000}	-951	-1031	-954	-446	-265	-334	56	485	971	1119
32	STUDENT:STAFF RATIO-2/10-{9.2100000, 13.8200000}	-331	-989	-1280	-640	-244	-34	269	650	751	881
33	STUDENT:STAFF RATIO-3/10-{13.8200000, 18.4300000}	-891	-522	-362	-160	92	143	197	91	135	70
34	STUDENT:STAFF RATIO-4/10-{18.4300000, 23.0400000}	-192	-120	143	199	191	35	-25	-460	-865	-1382
35	STUDENT:STAFF RATIO-5/10-{23.0400000, 27.6500000}	736	441	588	307	-67	-152	-884	-767	-1750	-1689
36	STUDENT:STAFF RATIO-6/10-{27.6500000, 32.2600000}	886	1144	491	273	-18	-634	-2200	-1772	-1334	
37	STUDENT:STAFF RATIO-7/10-{32.2600000, 36.8700000}	1134	1392	739	-21	-1068	-386	-1375			
38	STUDENT:STAFF RATIO-8/10-{36.8700000, 41.4800000}	1113	1219	1208	-43	-1089					
39	STUDENT:STAFF RATIO-9/10-{41.4800000, 46.0900000}	1876	1556	718			-221				
40	STUDENT:STAFF RATIO-10/10-{46.0900000, 50.7000000}		2374								
41	CAREER PROSPECTS-1/10-{16.0000000, 24.4000000}	1156	1261	761	426	-469					
42	CAREER PROSPECTS-2/10-{24.4000000, 32.8000000}	681	881	589	383	252	-1025				-1087
43	CAREER PROSPECTS-3/10-{32.8000000, 41.2000000}	596	125	462	458	270	-457	-629		-1961	-1901
44	CAREER PROSPECTS-4/10-{41.2000000, 49.6000000}	-153	584	182	339	113	-180	-138	-949	-1217	
45	CAREER PROSPECTS-5/10-{49.6000000, 58.0000000}	-133	-70	-137	132	13	354	56	-338	-1012	-2293
46	CAREER PROSPECTS-6/10-{58.0000000, 66.4000000}	-772	-724	-197	-107	166	69	298	-76	25	-381
47	CAREER PROSPECTS-7/10-{66.4000000, 74.8000000}	-1602	-1344	-842	-261	-450	217	252	645	657	170
48	CAREER PROSPECTS-8/10-{74.8000000, 83.2000000}		-1158	-576	-991	-326	-102	345	645	634	963
49	CAREER PROSPECTS-9/10-{83.2000000, 91.6000000}	-253	-1488	-1989	-1169	-256	-432	269	698	882	1196
50	CAREER PROSPECTS-10/10-{91.6000000, 100.0000000}	608	-628	-213	118	-311	-95	71	108	249	519
51	VALUE ADDED SCORE/10-1/10-{1.0000000, 1.9000000}	1451	1073	532	349	-360	-1410	-1582	-1153	-1293	

52	VALUE ADDED SCORE/10-2/10-{1.9000000, 2.8000000}	637	557	337	320	-45	-87	-379	-1291		-793
53	VALUE ADDED SCORE/10-3/10-{2.8000000, 3.7000000}	222	310	423	310	59	-166	-314	-912	-940	-1120
54	VALUE ADDED SCORE/10-4/10-{3.7000000, 4.6000000}	-648	221	331	182	-59	1	-219	-116	-153	-1075
55	VALUE ADDED SCORE/10-5/10-{4.6000000, 5.5000000}	-481	-376	-182	-56	136	167	200	-201	-391	-145
56	VALUE ADDED SCORE/10-6/10-{5.5000000, 6.4000000}	-63	-32	-267	-264	195	104	-52	196	28	-192
57	VALUE ADDED SCORE/10-7/10-{6.4000000, 7.3000000}	-595	-762	-557	-355	3	35	163	365	439	551
58	VALUE ADDED SCORE/10-8/10-{7.3000000, 8.2000000}	-630	-612	-331	-340	-184	-29	273	275	562	594
59	VALUE ADDED SCORE/10-9/10-{8.2000000, 9.1000000}	-1300	-1043	-628	-17	-11	-44	180	423	283	410
60	VALUE ADDED SCORE/10-10/10-{9.1000000, 10.0000000}		-1130	-1053	-624	-26	264	373	96	382	723
61	AVERAGE ENTRY TARIFF-1/10-{135.0000000, 181.3000000}	894	943	818	316	-1068	-200	-797			
62	AVERAGE ENTRY TARIFF-2/10-{181.3000000, 227.6000000}	978	936	651	42	-70	-504	-776	-1569	-1131	-1986
63	AVERAGE ENTRY TARIFF-3/10-{227.6000000, 273.9000000}	444	426	374	360	121	-216	-471	-1198	-1025	-2697
64	AVERAGE ENTRY TARIFF-4/10-{273.9000000, 320.2000000}	-576	-49	185	196	290	161	-477	-585	-614	-2046
65	AVERAGE ENTRY TARIFF-5/10-{320.2000000, 366.5000000}	-668	-1326	-528	-73	127	173	289	86	51	-252
66	AVERAGE ENTRY TARIFF-6/10-{366.5000000, 412.8000000}	-1171	-1491	-988	-594	-183	259	528	357	447	81
67	AVERAGE ENTRY TARIFF-7/10-{412.8000000, 459.1000000}	-690	-1348	-2186	-1028	-484	100	569	695	654	544
68	AVERAGE ENTRY TARIFF-8/10-{459.1000000, 505.4000000}	-134	-792	-1053	-545	-676	-371	133	913	917	834
69	AVERAGE ENTRY TARIFF-9/10-{505.4000000, 551.7000000}			-1002	-1914	-1130	-448	-1197	652	923	1958
70	AVERAGE ENTRY TARIFF-10/10-{551.7000000, 598.0000000}								374		2494
71	% SATISFIED WITH ASSESSMENT-1/10-{18.8755020, 26.6479518}	2454	1796								

Достоверность этих моделей различна (рисунок 10):

Наименование модели и частного критерия	Интегральный критерий	Вероятность правильной идентифика...	Вероятность правильной не идентиф...	Средняя вероятност...	Дата получения результата	Время получения результ...
ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс. частот с обр...	99.512	13.360	56.436	08.03.2015	14:27:38
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс. частот по признак...	100.000	0.706	50.353	08.03.2015	14:27:54
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн. частот с о...	99.512	13.310	56.411	08.03.2015	15:42:12
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн. частот по приз...	100.000	0.706	50.353	08.03.2015	15:42:32
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн. частот с о...	99.512	13.360	56.436	08.03.2015	16:51:20
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн. частот по приз...	100.000	0.706	50.353	08.03.2015	16:51:34
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	91.422	56.026	73.724	08.03.2015	18:11:40
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	88.492	48.641	68.566	08.03.2015	18:11:54
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	91.422	55.979	73.701	08.03.2015	19:30:47
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	88.560	48.363	68.462	08.03.2015	19:31:01
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	93.474	51.362	72.418	08.03.2015	20:48:54
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	93.474	51.363	72.418	08.03.2015	20:49:08
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	85.053	64.570	74.811	08.03.2015	22:07:00
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	93.767	35.210	64.489	08.03.2015	22:07:15
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	85.043	64.474	74.758	08.03.2015	23:25:29
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	93.845	34.991	64.418	08.03.2015	23:25:44
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	91.803	49.500	70.652	09.03.2015	00:45:50
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	93.474	35.342	64.408	09.03.2015	00:46:06
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	91.745	49.365	70.555	09.03.2015	02:02:44
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	93.621	35.150	64.385	09.03.2015	02:02:58

Рисунок 10 – Экранная форма отчета по достоверности моделей

Для количественной оценки достоверности моделей применена метрика, предложенная автором и по смыслу сходная с известным F-критерием (рисунок 11):

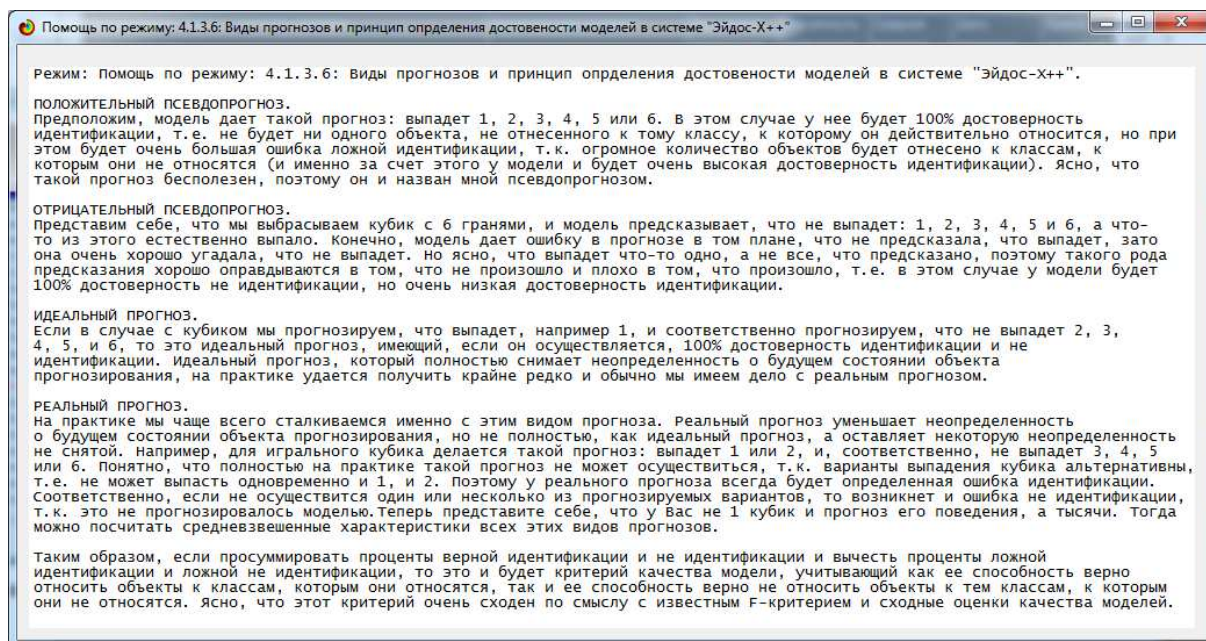


Рисунок 11 – Экранная форма пояснения по достоверности моделей

Обращает на себя внимание, что системно-когнитивные модели (INF1 – INF7) имеют значительно более высокую среднюю достоверность, чем статистические. Такая картина по опыту автора наблюдается в подавляющем большинстве приложений. В этом и состоит обоснование целесообразности применения системно-когнитивных (интеллектуальных) моделей.

6.4.3.6. НАГЛЯДНОЕ ОТОБРАЖЕНИЕ ПОДМАТРИЦ СИСТЕМНО-КОГНИТИВНЫХ МОДЕЛЕЙ УНИВЕРСИТЕТСКОГО РЕЙТИНГА ГАРДИАН В ВИДЕ КОГНИТИВНЫХ ФУНКЦИЙ

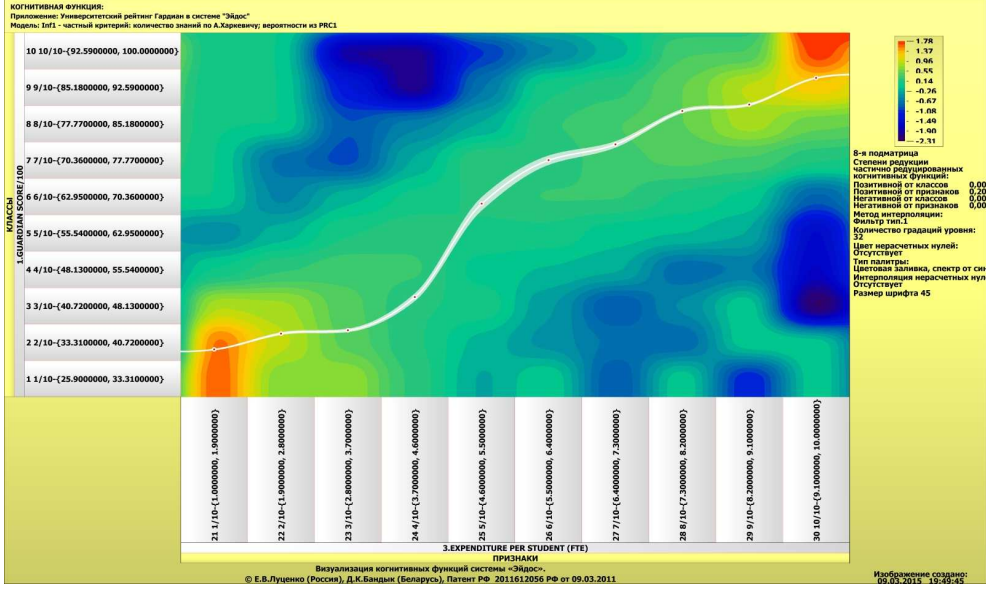
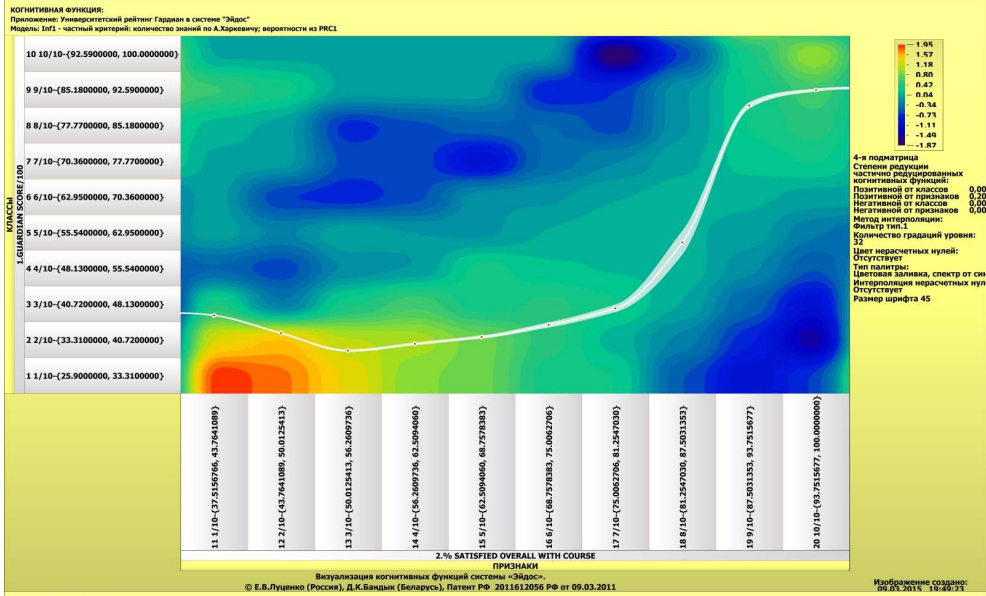
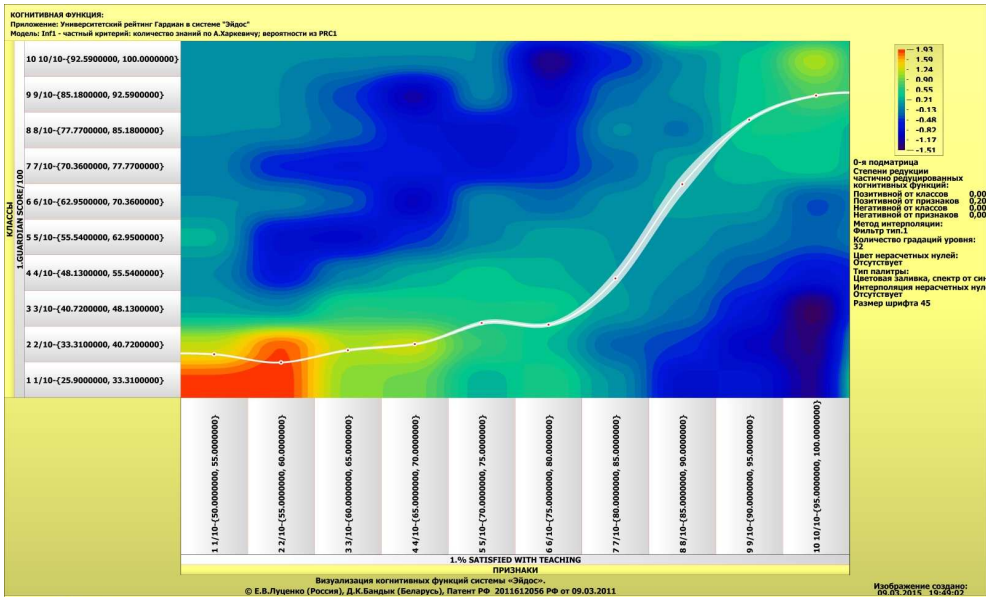
Применительно к задаче, рассматриваемой в данной работе, когнитивная функция показывает в наглядной графической форме, какое количество информации содержится в различных значениях показателей вузов о том, что у них будет определенный рейтинг по напылению подготовки и общий рейтинг Гардиан.

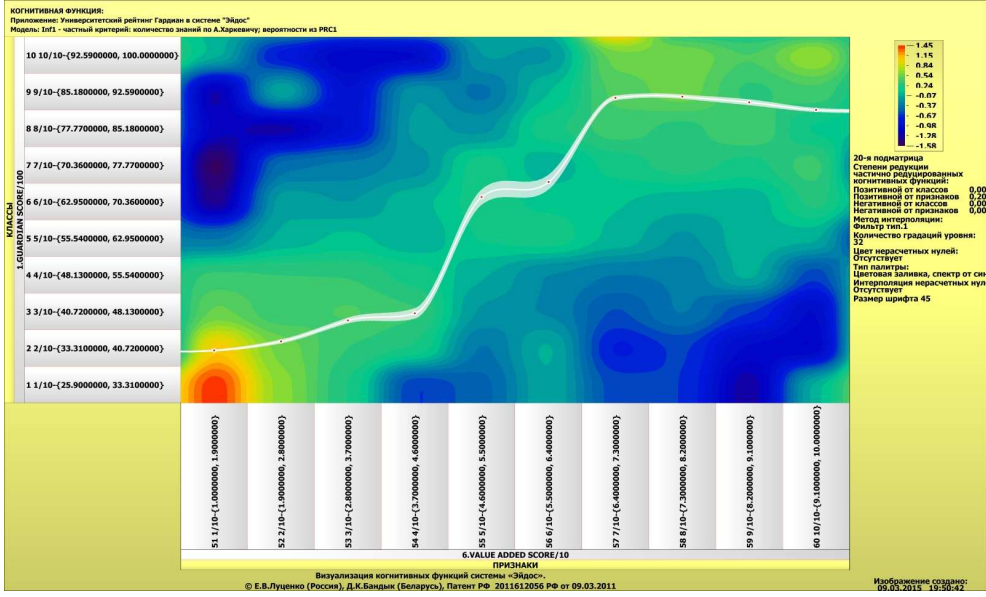
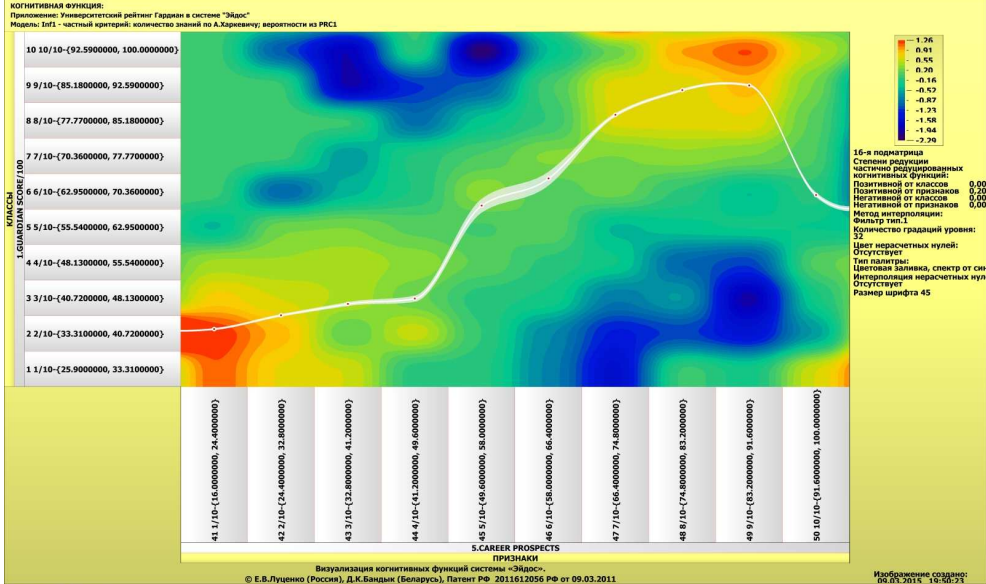
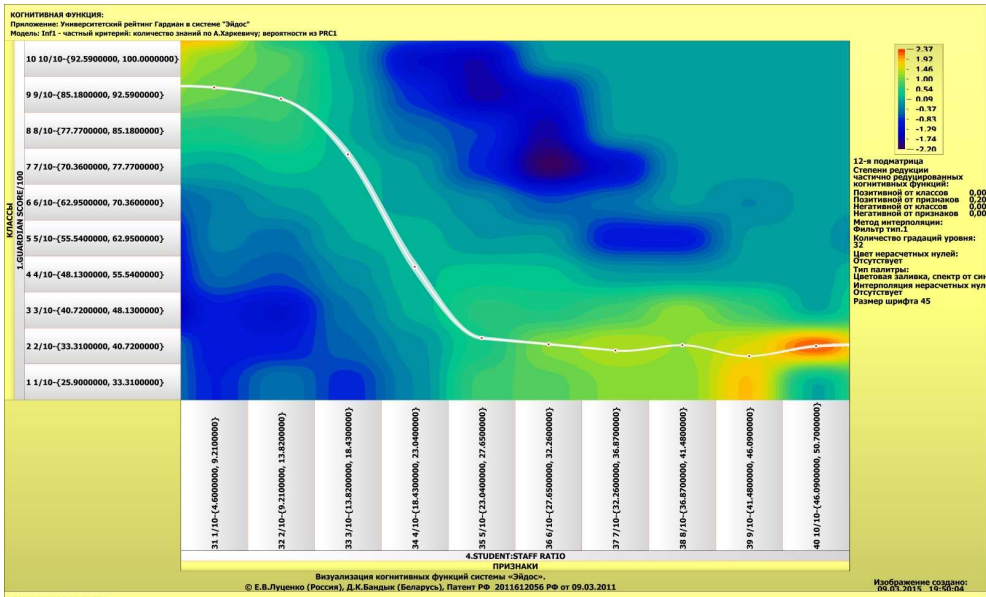
Когнитивным функциям посвящено много работ автора⁷³, но наиболее новой и обобщающей из них является работа [9]. Поэтому здесь не будем останавливаться на описании того, что представляют собой когнитивные функции в АСК-анализе.

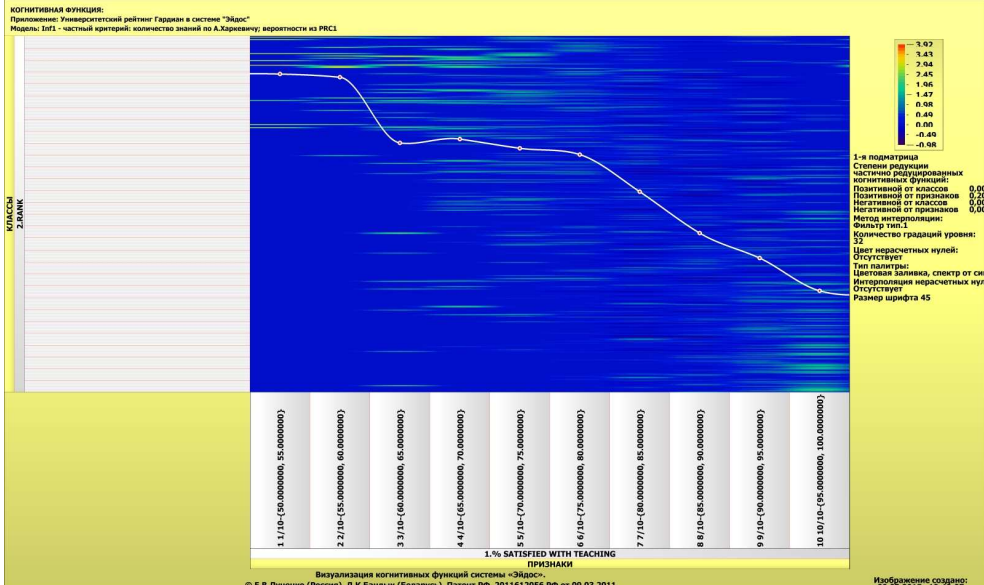
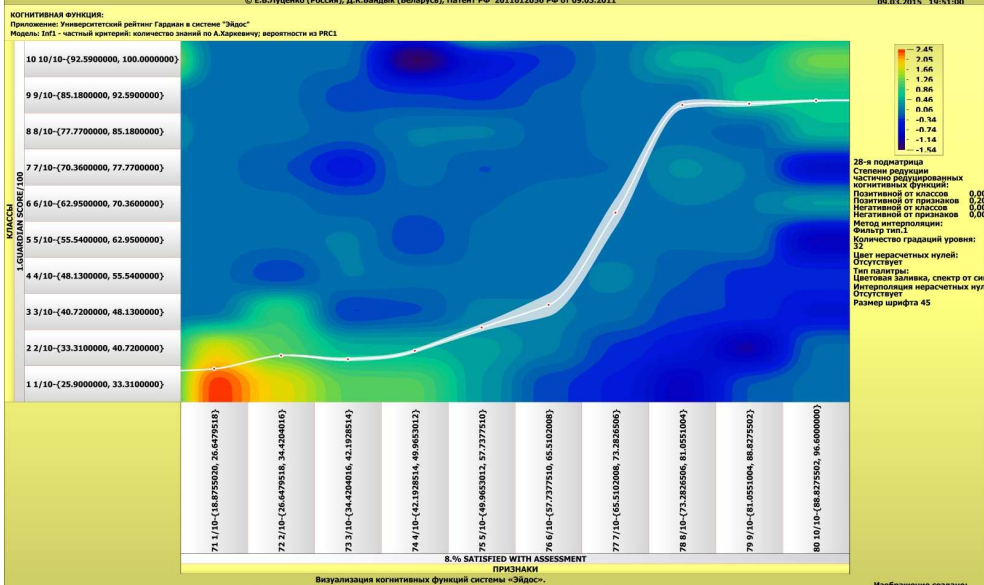
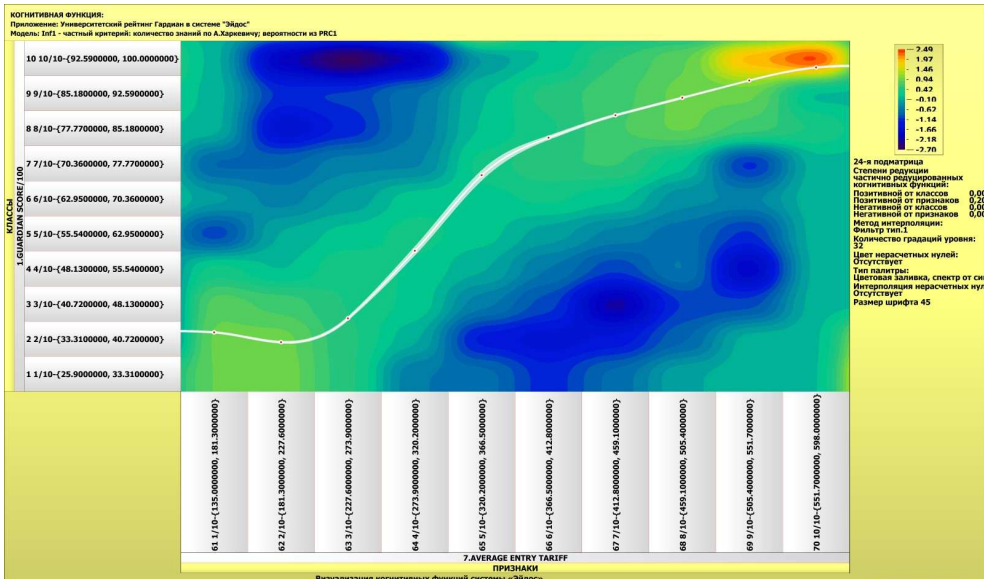
Отметим, что при построении средневзвешенных трендов применены математические методы, предложенные и описанные в работах [10, 11, 12], в частности применен метод взвешенных наименьших квадратов, модифицированный путем использования в качестве весовых коэффициентов количества информации в наблюдениях.

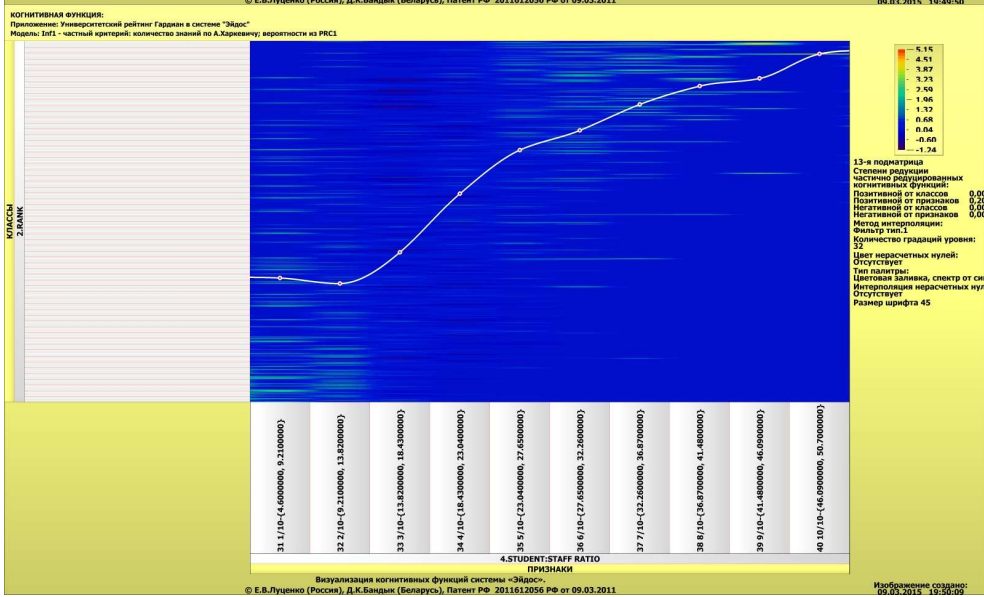
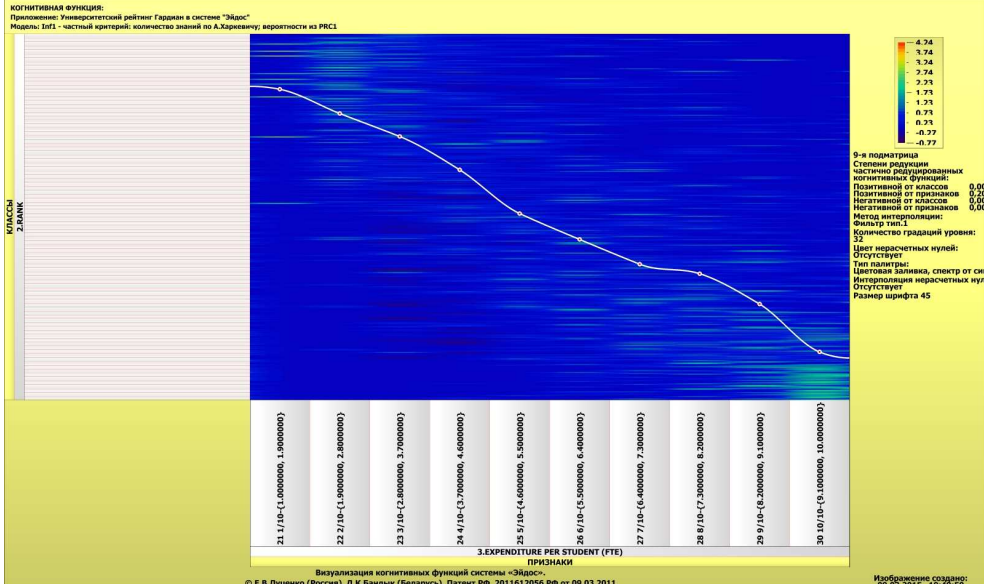
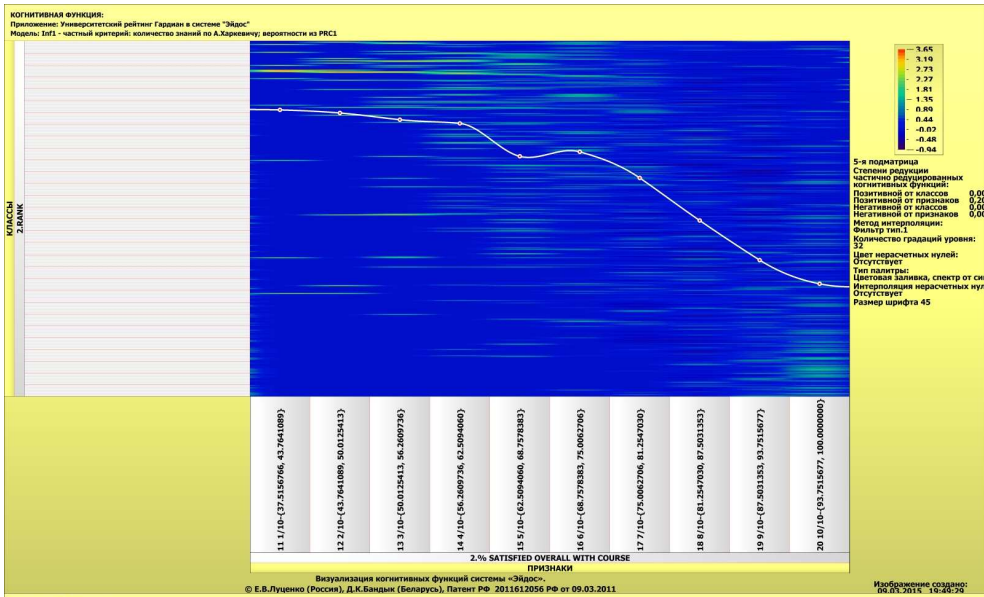
На рисунке 12 приведены визуализации некоторых когнитивных функций данного приложения для модели INF1:

⁷³ См., например: <http://www.twirpx.com/file/775236/>









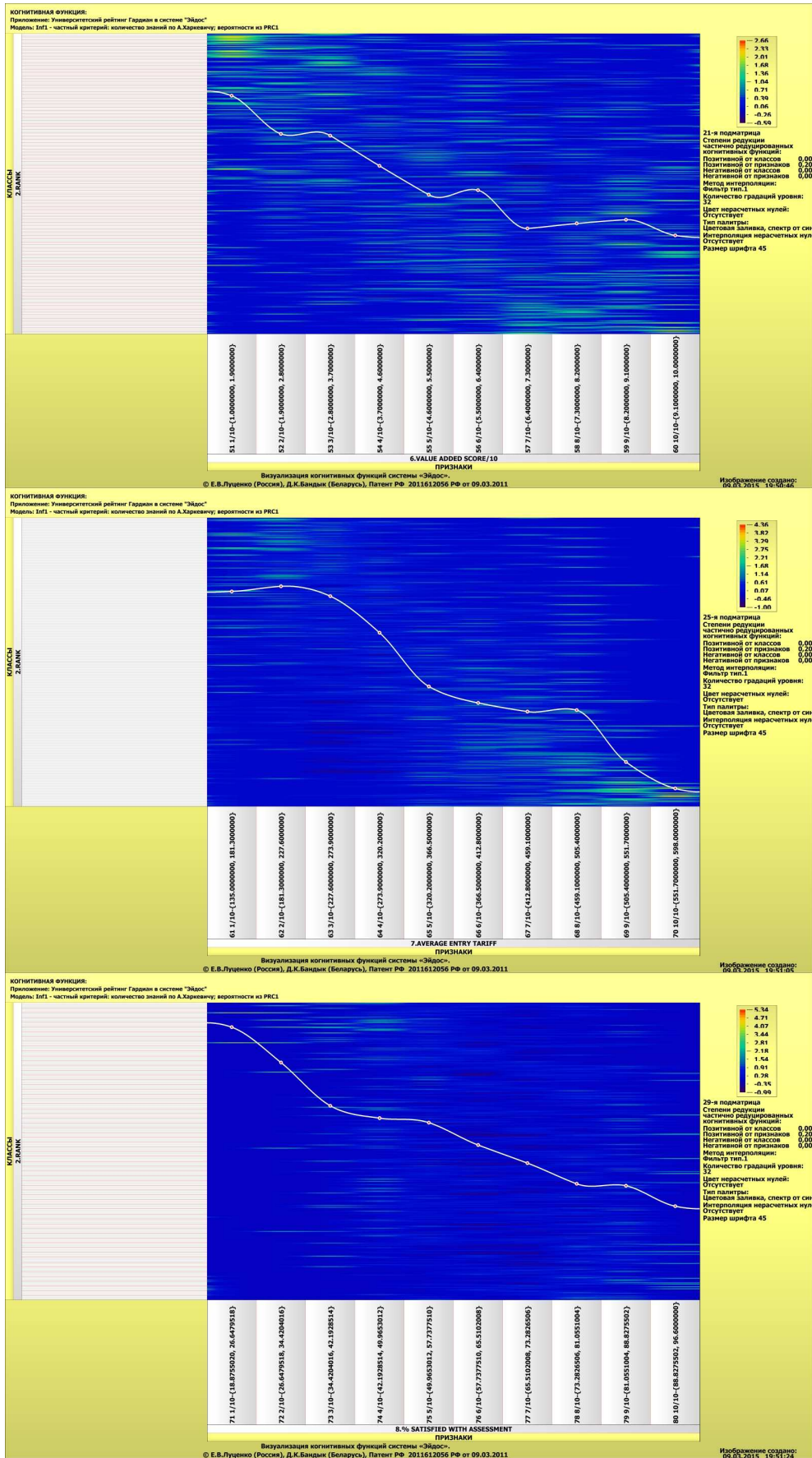


Рисунок 12 – Визуализация когнитивных функций зависимостей рейтинга Гардиан от значений показателей в системно-когнитивной модели INF1

Из приведенных когнитивных функций видно, что увеличение или уменьшение значений показателей вузов влияет на рейтинг Гардиан по направлению подготовки и общий рейтинг Гардиан, примерно пропорционально или обратно пропорционально. Отметим, что об этом можно говорить потому, что в системно-когнитивных моделях используются интервальные числовые и порядковые измерительные шкалы.

Это подтверждает разумность и корректность построения университетского рейтинга Гардиан его разработчиками.

6.4.3.7. ИНТЕГРАЛЬНЫЙ КРИТЕРИЙ И РЕШЕНИЕ ЗАДАЧИ ОЦЕНКИ РЕЙТИНГА ВУЗА В СИСТЕМНО-КОГНИТИВНОЙ МОДЕЛИ УНИВЕРСИТЕТСКОГО РЕЙТИНГА ГАРДИАН

Из модели INF1 мы видим, какое количество информации содержится в том или ином значении каждого показателя о том, что вуз с этим значением показателя имеет тот или иной рейтинг по направлению подготовки и общий рейтинг Гардиан.

Но если нам известно не одно, а несколько значений показателей вузов, то как посчитать их *общий* вклад в сходство с теми или иными классами? Для этого в системе «Эйдос» используется 2 аддитивных интегральных критерия: «Сумма знаний» и «Семантический резонанс знаний».

Интегральный критерий «Семантический резонанс знаний» представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где: M – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j-го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } : n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-X++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Интегральный критерий «Семантический резонанс знаний» представляет собой *нормированное* суммарное количество знаний, содержащееся в системе факторов раз-

личной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3 и имеет вид:

$$I_j = \frac{1}{\sigma_I \sigma_L M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

M – количество градаций описательных шкал (признаков);

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору объекта;

σ_I – среднее квадратичное отклонение частных критериев знаний вектора класса;

σ_L – среднее квадратичное отклонение по вектору распознаваемого объекта.

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j -го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i\text{-й фактор действует;} \\ n, & \text{где } n > 0, \text{ если } i\text{-й фактор действует с истинностью } n; \\ 0, & \text{если } i\text{-й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n , если он присутствует у объекта с интенсивностью n , т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}.$$

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j -го класса и состояния распознаваемого объекта.

Пример решения задачи идентификации для вузов рейтинга Гардиан по направлению подготовки и общего рейтинга Гардиан приведен на рисунке 12:

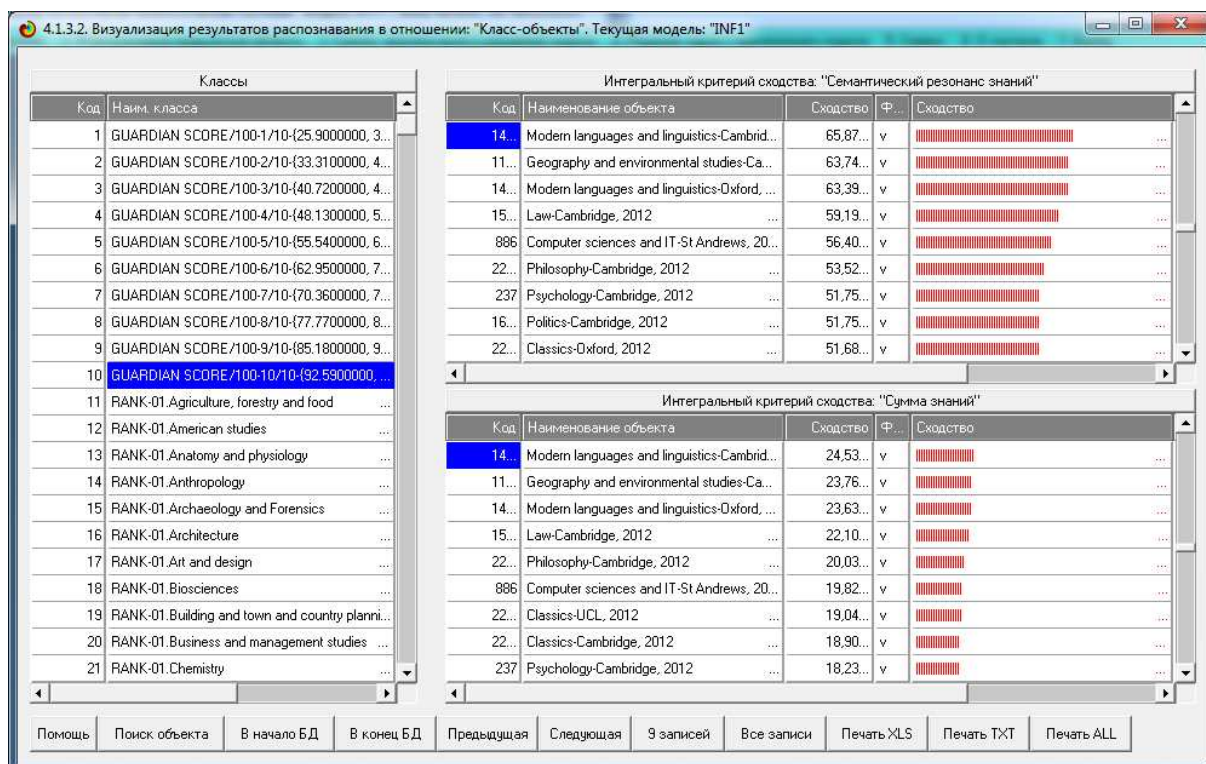
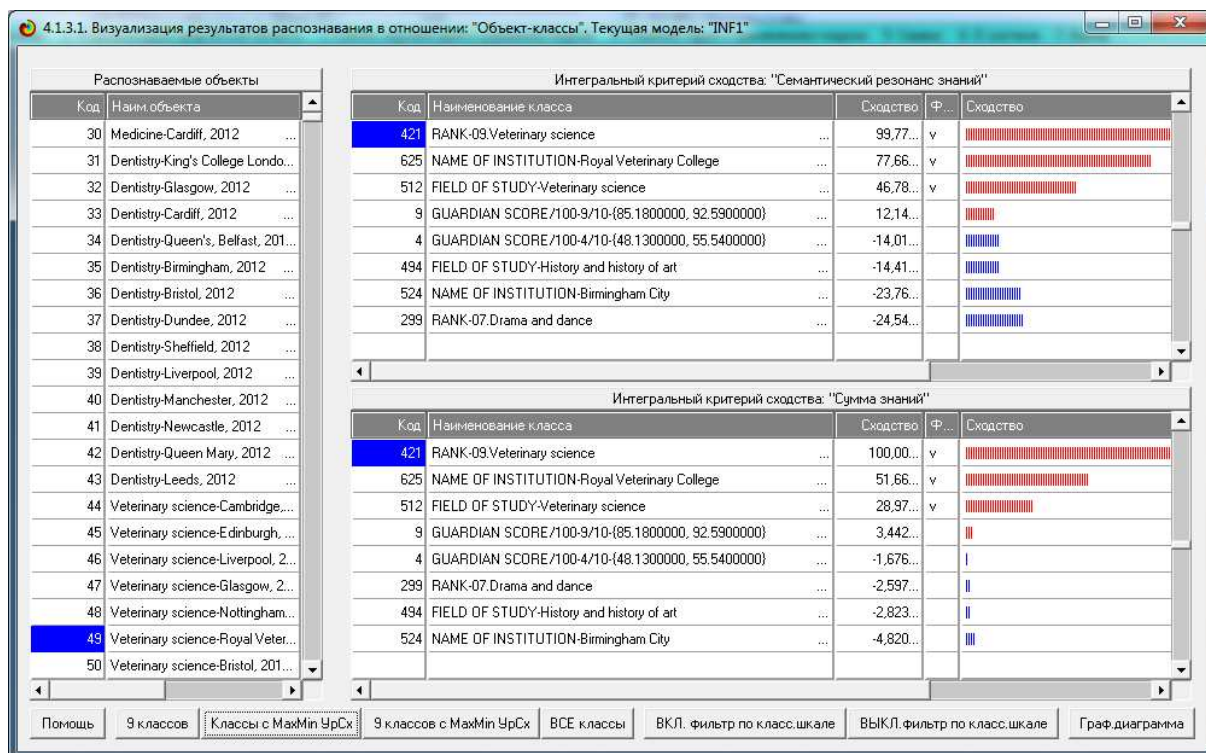


Рисунок 12 – Экранная форма с результатами идентификации рейтинга Гардиан по направлению подготовки и общего рейтинга Гардиан

6.4.3.8. ИССЛЕДОВАНИЕ МНОГОКРИТЕРИАЛЬНОЙ СИСТЕМНО-КОГНИТИВНОЙ МОДЕЛИ УНИВЕРСИТЕТСКОГО РЕЙТИНГА ГАРДИАН, УЧИТЫВАЮЩЕЙ НАПРАВЛЕНИЯ ПОДГОТОВКИ

6.4.3.8.1. Автоматизированный количественный SWOT-анализ университетского рейтинга Гардиан

В системе «Эйдос» реализован Автоматизированный количественный SWOT-анализ [13]. Его можно применить для исследования того, какие значения показателей способствуют, а какие препятствуют присвоению вузу тех или иных рейтингов Гардиан. Например, высокому общему рейтингу Гардиан способствуют и препятствуют значения показателей, приведенные на SWOT-диаграмме (рисунок 13), соответствующей SWOT-матрице (рисунок 14) и нелокальном нейроне (рисунок 15):

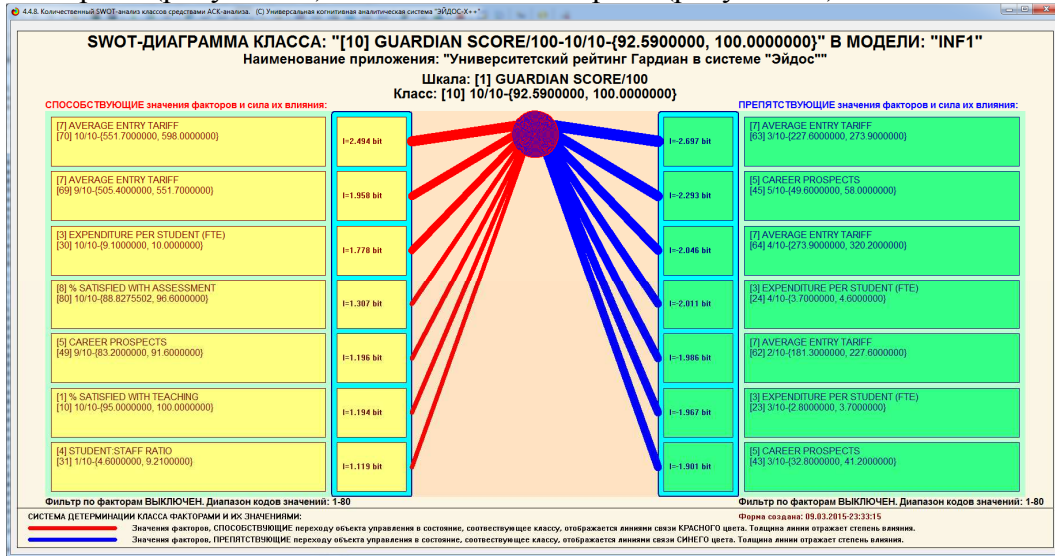


Рисунок 13 – SWOT-диаграмма высокого рейтинга Гардиан

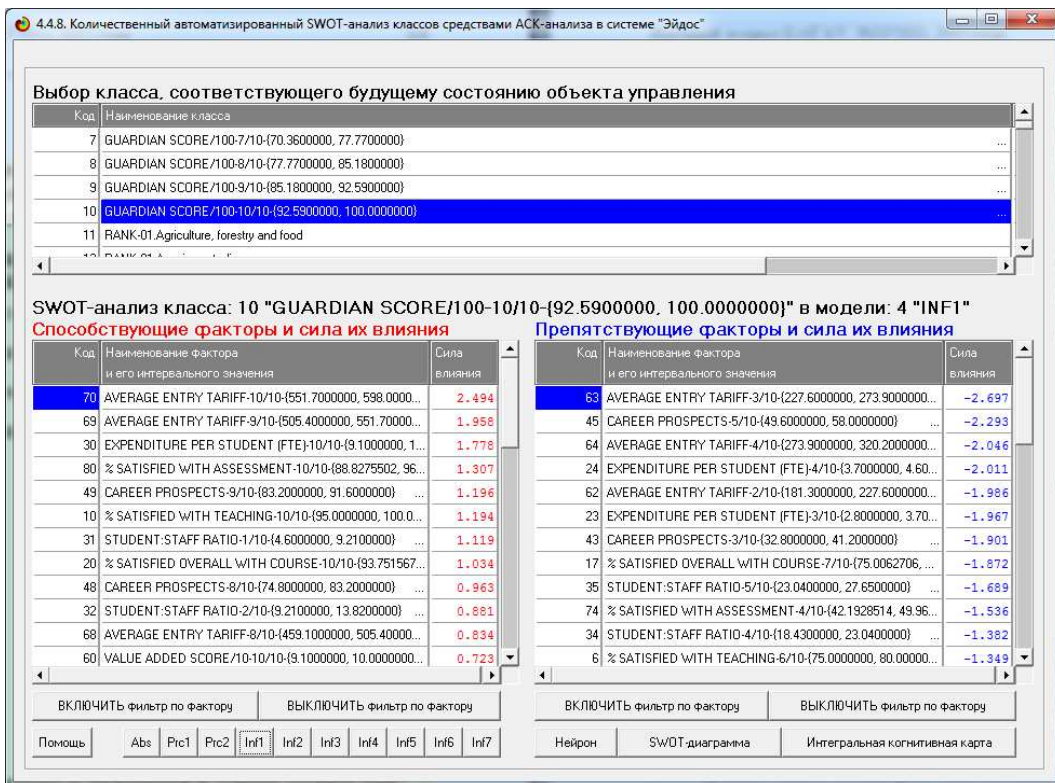


Рисунок 14 – SWOT-матрица высокого рейтинга Гардиан

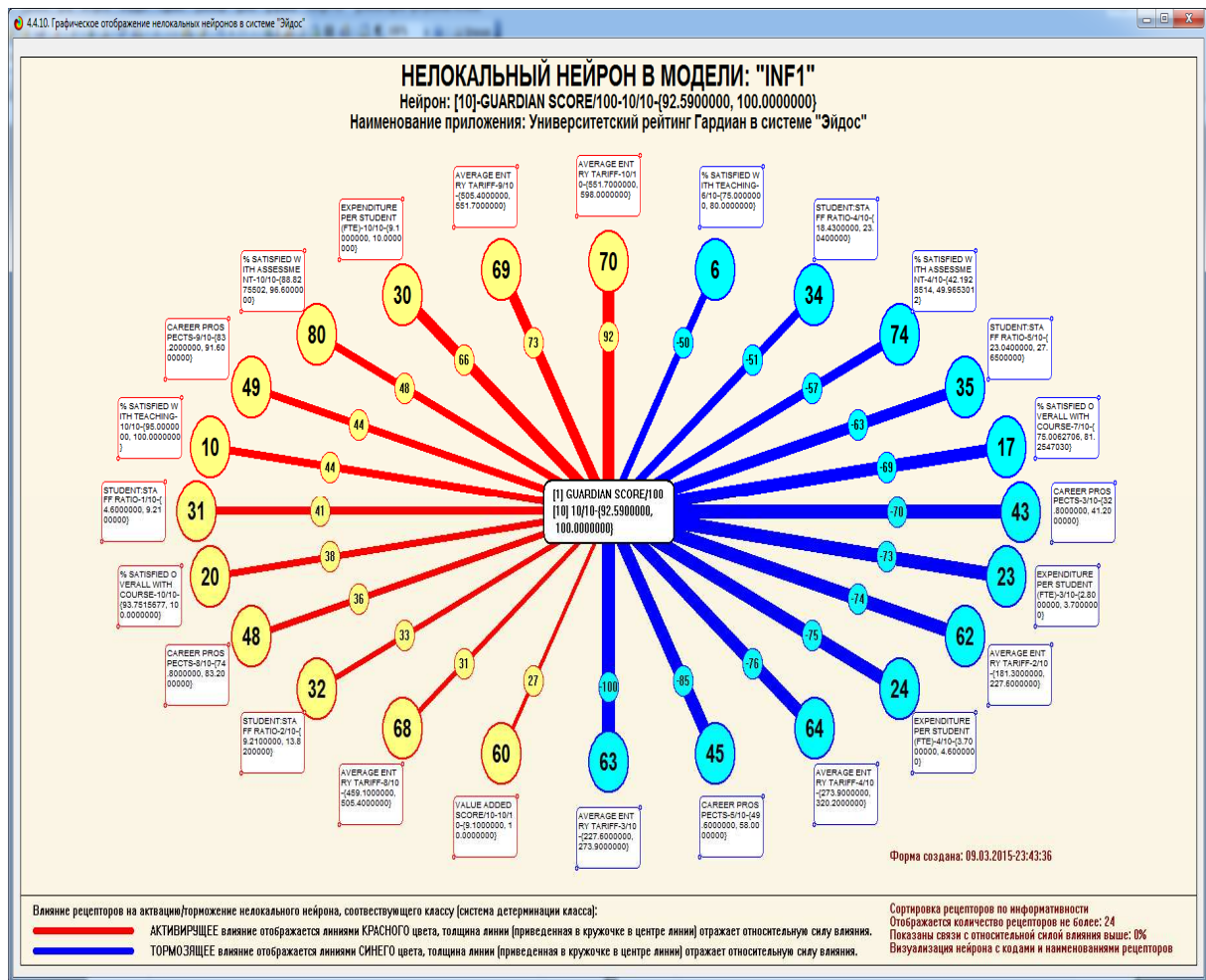


Рисунок 15 – Нелокальный нейрон высокого рейтинга Гардиан

6.4.3.8.2. Информационные портреты классов и значений показателей университетского рейтинга Гардиан

Информационный портрет класса – это список факторов, ранжированных в порядке убывания силы их влияния на переход объекта управления в состояние, соответствующее данному классу. Информационный портрет класса отражает систему его детерминации. Генерация информационного портрета класса представляет собой решение обратной задачи прогнозирования, т.к. при прогнозировании по системе факторов определяется спектр наиболее вероятных будущих состояний объекта управления, в которые он может перейти под влиянием данной системы факторов, а в информационном портрете мы наоборот, по заданному будущему состоянию объекта управления определяем систему факторов, детерминирующих это состояние, т.е. вызывающих переход объекта управления в это состояние. В начале информационного портрета класса идут факторы, оказывающие положительное влияние на переход объекта управления в заданное состояние, затем факторы, не оказывающие на это существенного влияния, и далее – факторы, препятствующие переходу объекта управления в это состояние (в порядке возрастания силы препятствования). Информационные портреты классов могут быть от *отфильтрованы* по диапазону факторов, т.е. мы можем отобразить влияние на переход объекта управления в данное состояние не всех отраженных в модели факторов, а только тех, коды которых попадают в определенный диапазон, например, относящиеся к определенным описательным шкалам.

Пример информационного портрета класса приведен на рисунке 16:

Код	Наименование признака	Значимость
1	GUARDIAN SCORE/100-1/10-(25.9000000, 33.31...	
2	GUARDIAN SCORE/100-2/10-(33.3100000, 40.72...	
3	GUARDIAN SCORE/100-3/10-(40.7200000, 48.13...	
4	GUARDIAN SCORE/100-4/10-(48.1300000, 55.54...	
5	GUARDIAN SCORE/100-5/10-(55.5400000, 62.95...	
6	GUARDIAN SCORE/100-6/10-(62.9500000, 70.36...	
7	GUARDIAN SCORE/100-7/10-(70.3600000, 77.77...	
8	GUARDIAN SCORE/100-8/10-(77.7700000, 85.18...	
9	GUARDIAN SCORE/100-9/10-(85.1800000, 92.59...	
10	GUARDIAN SCORE/100-10/10-(92.5900000, 100...	
11	RANK-01 Agriculture, forestry and food ...	
12	RANK-01 American studies ...	
13	RANK-01 Anatomy and physiology ...	
14	RANK-01 Anthropology ...	
15	RANK-01 Archaeology and Forensics ...	
16	RANK-01 Architecture ...	
17	RANK-01 Art and design ...	
18	RANK-01 Biosciences ...	
19	RANK-01 Building and town and country planning ...	
20	RANK-01 Business and management studies ...	
21	RANK-01 Chemistry ...	
22	RANK-01 Classics ...	
23	RANK-01 Computer sciences and IT ...	
30	EXPENDITURE PER STUDENT (FTE)/10/10-(9.1000000, 10.0000000)	1.178
10	% SATISFIED WITH TEACHING-10/10-(95.0000000, 100.0000000)	1.194
31	STUDENT STAFF RATIO-1/10-(4.6000000, 9.2100000)	1.119
20	% SATISFIED OVERALL WITH COURSE-10/10-(93.7515677, 100.0000000)	1.034
32	STUDENT STAFF RATIO-2/10-(9.2100000, 13.8200000)	0.881
19	% SATISFIED OVERALL WITH COURSE-9/10-(87.5031953, 93.7515677)	0.365
29	EXPENDITURE PER STUDENT (FTE)/9/10-(8.2000000, 9.1000000)	0.364
28	EXPENDITURE PER STUDENT (FTE)/8/10-(7.3000000, 8.2000000)	0.363
9	% SATISFIED WITH TEACHING-9/10-(90.0000000, 95.0000000)	0.347
33	STUDENT STAFF RATIO-3/10-(13.8200000, 18.4300000)	0.070
8	% SATISFIED WITH TEACHING-8/10-(85.0000000, 90.0000000)	-0.005
27	EXPENDITURE PER STUDENT (FTE)/7/10-(6.4000000, 7.3000000)	-0.205
7	% SATISFIED WITH TEACHING-7/10-(80.0000000, 85.0000000)	-0.536
26	EXPENDITURE PER STUDENT (FTE)/6/10-(5.5000000, 6.4000000)	-0.552
18	% SATISFIED OVERALL WITH COURSE-8/10-(81.2547030, 87.5031953)	-0.649
25	EXPENDITURE PER STUDENT (FTE)/5/10-(4.6000000, 5.5000000)	-1.057
42	CAREER PROSPECTS-2/10-(24.4000000, 32.8000000)	-1.087
6	% SATISFIED WITH TEACHING-6/10-(75.0000000, 80.0000000)	-1.349
34	STUDENT STAFF RATIO-4/10-(18.4300000, 23.0400000)	-1.382
35	STUDENT STAFF RATIO-5/10-(23.0400000, 27.6500000)	-1.689
17	% SATISFIED OVERALL WITH COURSE-7/10-(75.0062706, 81.2547030)	-1.872
23	EXPENDITURE PER STUDENT (FTE)/3/10-(2.8000000, 3.7000000)	-1.967
24	EXPENDITURE PER STUDENT (FTE)/4/10-(3.7000000, 4.6000000)	-2.011

Рисунок 16 – Экранная форма с информационным портретом класса: «Наивысший общий рейтинг Гардиан»

Информационный (семантический) портрет фактора – это список классов, ранжированный в порядке убывания силы влияния данного фактора на переход объекта управления в состояния, соответствующие данным классам. Информационный портрет фактора называется также его *семантическим портретом*, т.к. в соответствии с концепцией смысла системно-когнитивного анализа, являющейся обобщением концепции смысла Шенка-Абельсона, *смысл фактора состоит в том, какие будущие состояния объекта управления он детерминирует*. Сначала в этом списке идут состояния объекта управления, на переход в которые данный фактор оказывает наибольшее влияние, затем состояния, на которые данный фактор не оказывает существенного влияния, и далее состояния – переходу в которые данный фактор препятствует. Информационные портреты факторов могут быть от *отфильтрованы* по диапазону классов, т.е. мы можем отобразить влияние данного фактора на переход объекта управления не во все возможные будущие состояния, а только в состояния, коды которых попадают в определенный диапазон, например, относящиеся к определенным классификационным шкалам.

Пример информационного портрета значения фактора (показателя) приведен на рисунке 17:

Код	Наименование признака	Значимость
1	% SATISFIED WITH TEACHING-1/10-(50.0000000, 55.0000000)	
2	% SATISFIED WITH TEACHING-2/10-(55.0000000, 60.0000000)	
3	% SATISFIED WITH TEACHING-3/10-(60.0000000, 65.0000000)	
4	% SATISFIED WITH TEACHING-4/10-(65.0000000, 70.0000000)	
5	% SATISFIED WITH TEACHING-5/10-(70.0000000, 75.0000000)	
6	% SATISFIED WITH TEACHING-6/10-(75.0000000, 80.0000000)	
7	% SATISFIED WITH TEACHING-7/10-(80.0000000, 85.0000000)	
8	% SATISFIED WITH TEACHING-8/10-(85.0000000, 90.0000000)	
9	% SATISFIED WITH TEACHING-9/10-(90.0000000, 95.0000000)	
10	% SATISFIED WITH TEACHING-10/10-(95.0000000, 100.0000000)	
11	% SATISFIED OVERALL WITH COURSE-1/10-(37.0000000, 43.0000000)	
12	% SATISFIED OVERALL WITH COURSE-2/10-(43.0000000, 49.0000000)	
13	% SATISFIED OVERALL WITH COURSE-3/10-(49.0000000, 55.0000000)	
14	% SATISFIED OVERALL WITH COURSE-4/10-(55.0000000, 61.0000000)	
15	% SATISFIED OVERALL WITH COURSE-5/10-(61.0000000, 67.0000000)	
16	% SATISFIED OVERALL WITH COURSE-6/10-(67.0000000, 73.0000000)	
17	% SATISFIED OVERALL WITH COURSE-7/10-(73.0000000, 79.0000000)	
18	% SATISFIED OVERALL WITH COURSE-8/10-(79.0000000, 85.0000000)	
19	% SATISFIED OVERALL WITH COURSE-9/10-(85.0000000, 91.0000000)	
20	% SATISFIED OVERALL WITH COURSE-10/10-(91.0000000, 97.0000000)	
21	EXPENDITURE PER STUDENT (FTE)/1/10-(1.0000000, 1.1000000)	
22	EXPENDITURE PER STUDENT (FTE)/2/10-(1.1000000, 1.2000000)	
23	EXPENDITURE PER STUDENT (FTE)/3/10-(1.2000000, 1.3000000)	
657	NAME OF INSTITUTION-University of the Arts, London	3.606
551	NAME OF INSTITUTION-East London	2.370
641	NAME OF INSTITUTION-Sunderland	2.358
524	NAME OF INSTITUTION-Birmingham City	2.297
589	NAME OF INSTITUTION-London Met	2.126
654	NAME OF INSTITUTION-Ulster	1.962
587	NAME OF INSTITUTION-Liverpool	1.896
553	NAME OF INSTITUTION-Edinburgh	1.807

Рисунок 17 – Экранная форма с информационным портретом значения показателя с установленным фильтром по наименованиям вузов

6.4.3.8.3. Кластерно-конструктивный анализ университетского рейтинга Гардиан

Кластерно-конструктивный анализ – это новый математический метод анализа знаний, реализованный в АСК-анализе и системе «Эйдос» [14], обеспечивающий:

- выявление классов, наиболее сходных по системе их детерминации и объединение их в кластеры;
- выявление кластеров классов, наиболее сильно отличающиеся по системе их детерминации и построение из них полюсов конструкторов классов, при этом остальные кластеры включаются в конструкторы в качестве промежуточных между полюсами;
- выявление факторов, наиболее сходных по детерминируемым ими классам и объединение их в кластеры;
- выявление кластеров факторов, наиболее сильно отличающиеся по детерминируемым ими классам и построение из них полюсов конструкторов факторов, при этом остальные кластеры включаются в конструкторы в качестве промежуточных между полюсами.

Состояния объекта управления, соответствующие классам, включенным в один кластер, могут быть достигнуты одновременно, т.е. являются *совместимыми (коалиционными)* по детерминирующим их факторам. Состояния объекта управления, соответствующие классам, образующим полюса конструктора, не могут быть достигнуты одновременно, т.е. являются *противоположными* по детерминирующим их факторам (*антагонистическими*).

Факторы, включенные в один кластер, оказывают сходное влияние на поведение объекта управления и могут, при необходимости, быть использованы для замены друг друга. Факторы, образующие полюса конструктора, оказывают противоположное влияние на поведение объекта управления.

Кластерно-конструктивный анализ классов позволяет сравнить их по сходству системы детерминации и отобразить эту информацию в наглядной графической форме семантической сети классов.

Кластерно-конструктивный анализ факторов позволяет сравнить факторы по сходству их влияния на переход объекта в будущие состояния и отобразить эту информацию в наглядной графической форме семантической сети факторов.

Примеры когнитивных диаграмм, отражающих некоторые результаты кластерно-конструктивного анализа модели университетского рейтинга Гардиан, приведены на рисунках 18, 19, 20:

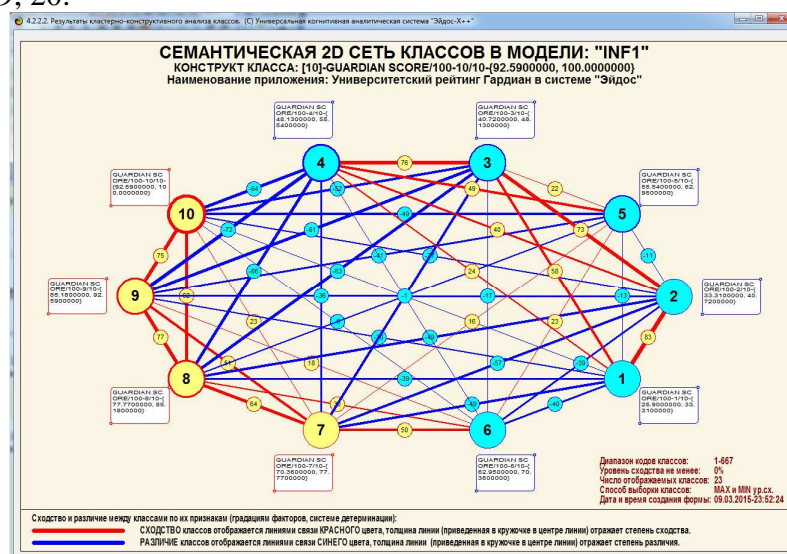


Рисунок 18. Пример конструктора класса университетского рейтинга Гардиан

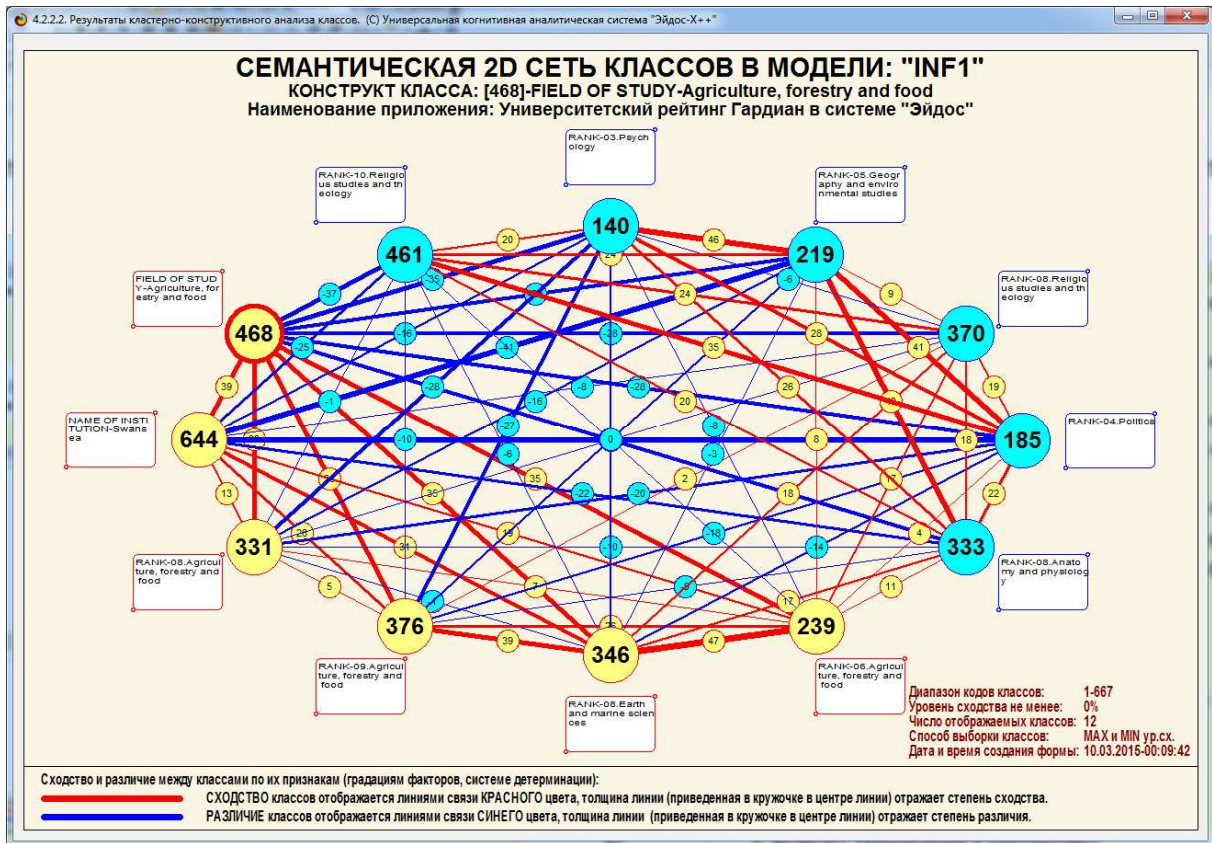


Рисунок 19. Пример конструкта класса университетского рейтинга Гардиан

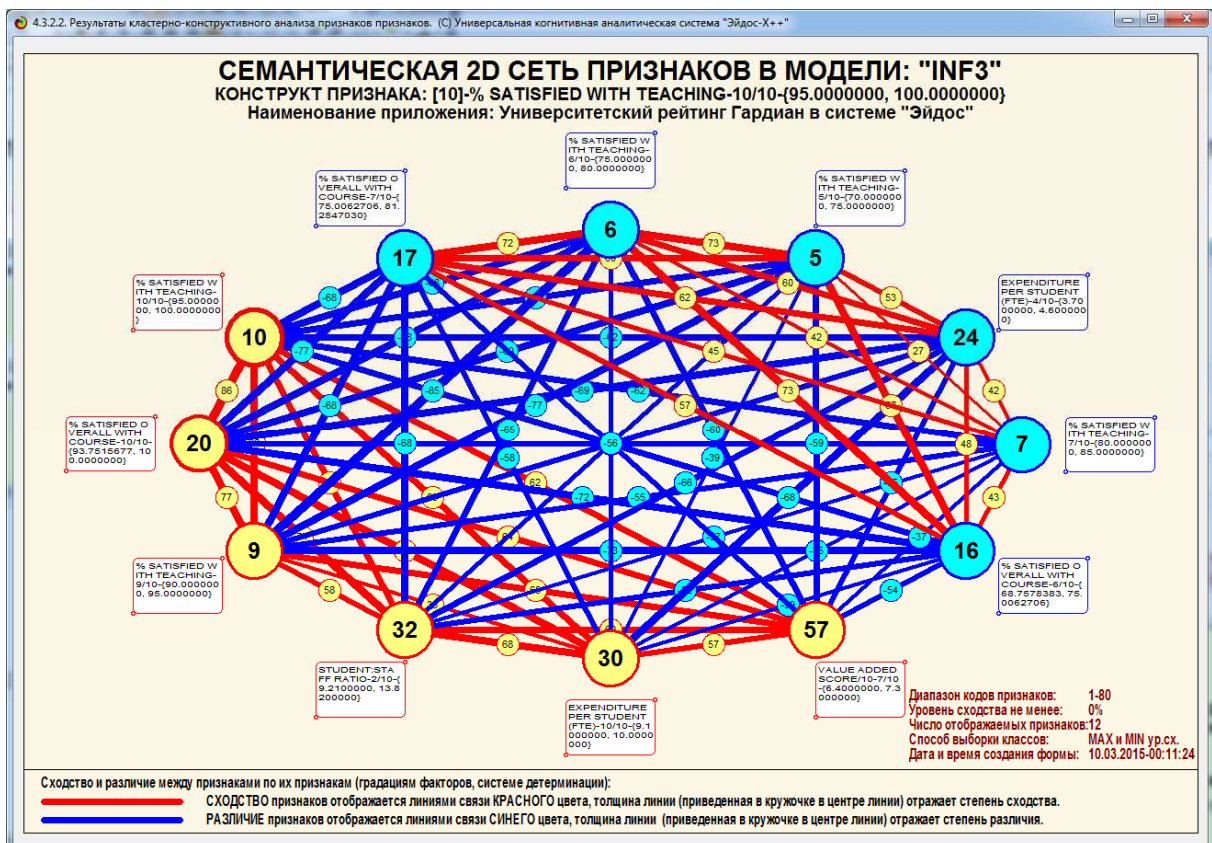


Рисунок 20. Пример конструкта значения показателя университетского рейтинга Гардиан

Как видно из приведенных когнитивных диаграмм, все классы и значения показателей являются взаимозависимыми, что исключает применение факторного анализа, как метода моделирования линейных систем.

6.4.4. Интеграция различных рейтингов в одном «супер рейтинге» – путь к использованию рейтинга Гардиан для оценки российских вузов

6.4.4.1. Пилотное исследование и Парето-оптимизация

Минобрнауки РФ в своих регламентирующих документах предлагает очень много частных критериев⁷⁴. Ясно, что собрать информацию по всем этим показателям очень сложно, дорого и трудоемко. Поэтому представляет интерес, выявить из них минимальное количество таких критериев, которых было бы достаточно для надежного решения задачи определения рейтинга вуза.

Системно-когнитивные модели позволяют выявить показатели, оказывающие наиболее существенное влияние на объекты моделирования, что позволяет удалить из моделей не существенные показатели, т.е. провести Паретто-оптимизацию, в результате которой в моделях остаются только существенные показатели.

Таким образом, решается задача, аналогичная задаче разработки системы стандартизированных показателей, но конкретно для данного предприятия.

В результате можно сократить размерность моделей без потери их достоверности, а значит существенно сократить затраты труда и времени на сбор, ввод в компьютер и обработку исходных данных, т.е. эффективность их использования.

В таблице 13 приведен список значений факторов системно-когнитивной модели INF1 (см. табл.) университетского рейтинга Гардиан, в котором эти значения проранжированы в порядке убывания вариабельности информативности, которая в АСК-анализе рассматривается как значимость (дифференцирующая способность) этого значения. Вариабельность информативности измеряется как ее среднеквадратичное отклонение по всем классам. Но в данном случае она посчитана только по первым 10 классам, т.е. по общему рейтингу.

Таблица 13 – Ранжированная таблица значений показателей для построения Парето-диаграммы университетского рейтинга Гардиан

Код	Значение показателя	Значимость	Паретто
70	AVERAGE ENTRY TARIFF-10/10-{551.7000000, 598.0000000}	1499,07	1499,07
30	EXPENDITURE PER STUDENT (FTE)-10/10-{9.1000000, 10.0000000}	1493,90	2992,96
2	% SATISFIED WITH TEACHING-2/10-{55.0000000, 60.0000000}	1430,66	4423,63
69	AVERAGE ENTRY TARIFF-9/10-{505.4000000, 551.7000000}	1315,20	5738,83
36	STUDENT:STAFF RATIO-6/10-{27.6500000, 32.2600000}	1196,80	6935,63
51	VALUE ADDED SCORE/10-1/10-{1.0000000, 1.9000000}	1153,71	8089,34
37	STUDENT:STAFF RATIO-7/10-{32.2600000, 36.8700000}	1074,37	9163,71
49	CAREER PROSPECTS-9/10-{83.2000000, 91.6000000}	1052,72	10216,43
62	AVERAGE ENTRY TARIFF-2/10-{181.3000000, 227.6000000}	1034,30	11250,73
38	STUDENT:STAFF RATIO-8/10-{36.8700000, 41.4800000}	1026,14	12276,87
21	EXPENDITURE PER STUDENT (FTE)-1/10-{1.0000000, 1.9000000}	1020,24	13297,11
63	AVERAGE ENTRY TARIFF-3/10-{227.6000000, 273.9000000}	1010,23	14307,33
12	% SATISFIED OVERALL WITH COURSE-2/10-{43.7641089, 50.0125413}	1009,23	15316,56
67	AVERAGE ENTRY TARIFF-7/10-{412.8000000, 459.1000000}	994,19	16310,75
43	CAREER PROSPECTS-3/10-{32.8000000, 41.2000000}	994,12	17304,87
10	% SATISFIED WITH TEACHING-10/10-{95.0000000, 100.0000000}	948,77	18253,64
39	STUDENT:STAFF RATIO-9/10-{41.4800000, 46.0900000}	939,42	19193,06
35	STUDENT:STAFF RATIO-5/10-{23.0400000, 27.6500000}	909,13	20102,18
24	EXPENDITURE PER STUDENT (FTE)-4/10-{3.7000000, 4.6000000}	905,30	21007,49
23	EXPENDITURE PER STUDENT (FTE)-3/10-{2.8000000, 3.7000000}	899,29	21906,78
72	% SATISFIED WITH ASSESSMENT-2/10-{26.6479518, 34.4204016}	882,78	22789,56

⁷⁴ См., например: <http://uup.samgtu.ru/node/211>

4	% SATISFIED WITH TEACHING-4/10-{65.0000000, 70.0000000}	879,48	23669,05
11	% SATISFIED OVERALL WITH COURSE-1/10-{37.5156766, 43.7641089}	857,09	24526,14
61	AVERAGE ENTRY TARIFF-1/10-{135.0000000, 181.3000000}	832,47	25358,60
1	% SATISFIED WITH TEACHING-1/10-{50.0000000, 55.0000000}	827,04	26185,65
80	% SATISFIED WITH ASSESSMENT-10/10-{88.8275502, 96.6000000}	826,29	27011,93
13	% SATISFIED OVERALL WITH COURSE-3/10-{50.0125413, 56.2609736}	818,74	27830,68
42	CAREER PROSPECTS-2/10-{24.4000000, 32.8000000}	812,94	28643,62
20	% SATISFIED OVERALL WITH COURSE-10/10-{93.7515677, 100.0000000}	804,92	29448,54
47	CAREER PROSPECTS-7/10-{66.4000000, 74.8000000}	794,88	30243,42
31	STUDENT:STAFF RATIO-1/10-{4.6000000, 9.2100000}	782,26	31025,68
45	CAREER PROSPECTS-5/10-{49.6000000, 58.0000000}	775,12	31800,80
48	CAREER PROSPECTS-8/10-{74.8000000, 83.2000000}	758,31	32559,11
32	STUDENT:STAFF RATIO-2/10-{9.2100000, 13.8200000}	746,78	33305,89
68	AVERAGE ENTRY TARIFF-8/10-{459.1000000, 505.4000000}	743,35	34049,24
3	% SATISFIED WITH TEACHING-3/10-{60.0000000, 65.0000000}	738,54	34787,78
66	AVERAGE ENTRY TARIFF-6/10-{366.5000000, 412.8000000}	736,20	35523,97
29	EXPENDITURE PER STUDENT (FTE)-9/10-{8.2000000, 9.1000000}	726,84	36250,82
64	AVERAGE ENTRY TARIFF-4/10-{273.9000000, 320.2000000}	700,26	36951,07
41	CAREER PROSPECTS-1/10-{16.0000000, 24.4000000}	696,72	37647,79
14	% SATISFIED OVERALL WITH COURSE-4/10-{56.2609736, 62.5094060}	696,11	38343,90
74	% SATISFIED WITH ASSESSMENT-4/10-{42.1928514, 49.9653012}	674,77	39018,68
17	% SATISFIED OVERALL WITH COURSE-7/10-{75.0062706, 81.2547030}	672,82	39691,50
22	EXPENDITURE PER STUDENT (FTE)-2/10-{1.9000000, 2.8000000}	667,46	40358,97
60	VALUE ADDED SCORE/10-10/10-{9.1000000, 10.0000000}	666,34	41025,30
6	% SATISFIED WITH TEACHING-6/10-{75.0000000, 80.0000000}	657,57	41682,88
52	VALUE ADDED SCORE/10-2/10-{1.9000000, 2.8000000}	644,10	42326,98
15	% SATISFIED OVERALL WITH COURSE-5/10-{62.5094060, 68.7578383}	631,57	42958,55
59	VALUE ADDED SCORE/10-9/10-{8.2000000, 9.1000000}	608,18	43566,72
44	CAREER PROSPECTS-4/10-{41.2000000, 49.6000000}	584,91	44151,64
53	VALUE ADDED SCORE/10-3/10-{2.8000000, 3.7000000}	583,81	44735,44
5	% SATISFIED WITH TEACHING-5/10-{70.0000000, 75.0000000}	555,04	45290,48
28	EXPENDITURE PER STUDENT (FTE)-8/10-{7.3000000, 8.2000000}	528,69	45819,17
9	% SATISFIED WITH TEACHING-9/10-{90.0000000, 95.0000000}	524,30	46343,48
34	STUDENT:STAFF RATIO-4/10-{18.4300000, 23.0400000}	517,10	46860,57
73	% SATISFIED WITH ASSESSMENT-3/10-{34.4204016, 42.1928514}	508,31	47368,89
27	EXPENDITURE PER STUDENT (FTE)-7/10-{6.4000000, 7.3000000}	507,02	47875,91
65	AVERAGE ENTRY TARIFF-5/10-{320.2000000, 366.5000000}	498,92	48374,82
79	% SATISFIED WITH ASSESSMENT-9/10-{81.0551004, 88.8275502}	496,61	48871,44
19	% SATISFIED OVERALL WITH COURSE-9/10-{87.5031353, 93.7515677}	477,60	49349,03
57	VALUE ADDED SCORE/10-7/10-{6.4000000, 7.3000000}	468,80	49817,83
71	% SATISFIED WITH ASSESSMENT-1/10-{18.8755020, 26.6479518}	465,09	50282,92
58	VALUE ADDED SCORE/10-8/10-{7.3000000, 8.2000000}	451,24	50734,16
16	% SATISFIED OVERALL WITH COURSE-6/10-{68.7578383, 75.0062706}	443,77	51177,93
54	VALUE ADDED SCORE/10-4/10-{3.7000000, 4.6000000}	424,13	51602,06
78	% SATISFIED WITH ASSESSMENT-8/10-{73.2826506, 81.0551004}	423,10	52025,17
25	EXPENDITURE PER STUDENT (FTE)-5/10-{4.6000000, 5.5000000}	401,72	52426,89
50	CAREER PROSPECTS-10/10-{91.6000000, 100.0000000}	373,87	52800,76
33	STUDENT:STAFF RATIO-3/10-{13.8200000, 18.4300000}	361,39	53162,15
46	CAREER PROSPECTS-6/10-{58.0000000, 66.4000000}	358,11	53520,26
26	EXPENDITURE PER STUDENT (FTE)-6/10-{5.5000000, 6.4000000}	308,74	53829,01
8	% SATISFIED WITH TEACHING-8/10-{85.0000000, 90.0000000}	306,02	54135,02
75	% SATISFIED WITH ASSESSMENT-5/10-{49.9653012, 57.7377510}	292,73	54427,75
18	% SATISFIED OVERALL WITH COURSE-8/10-{81.2547030, 87.5031353}	272,45	54700,20
55	VALUE ADDED SCORE/10-5/10-{4.6000000, 5.5000000}	243,46	54943,66
7	% SATISFIED WITH TEACHING-7/10-{80.0000000, 85.0000000}	219,55	55163,21
77	% SATISFIED WITH ASSESSMENT-7/10-{65.5102008, 73.2826506}	180,25	55343,46
56	VALUE ADDED SCORE/10-6/10-{5.5000000, 6.4000000}	170,49	55513,95
76	% SATISFIED WITH ASSESSMENT-6/10-{57.7377510, 65.5102008}	143,31	55657,26

На рисунке 21 приведена Парето-диаграмма, построенная по таблице 13:

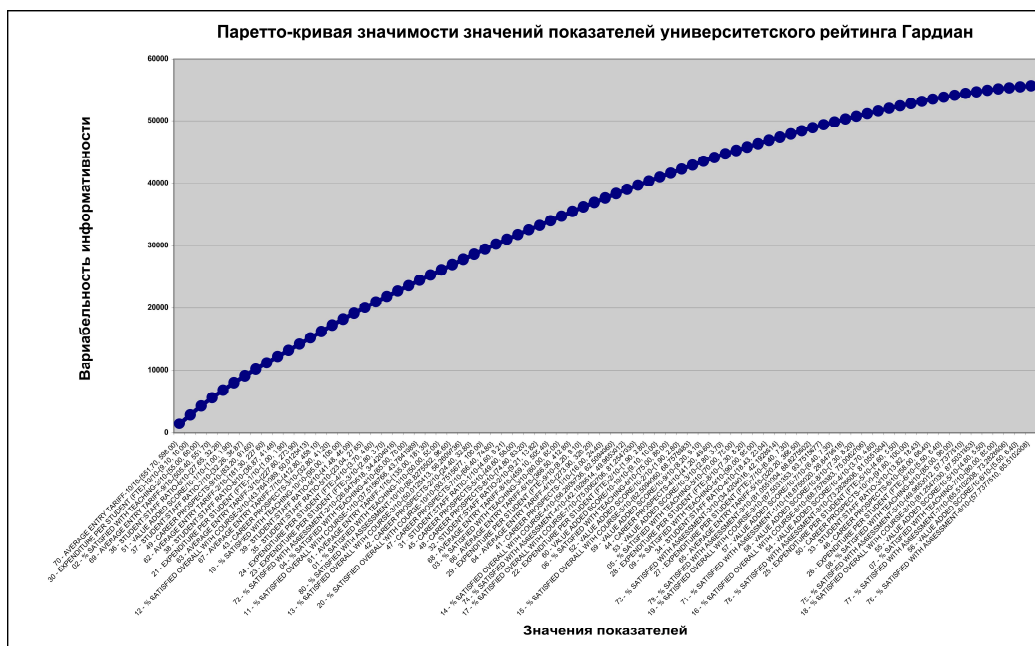


Рисунок 13. Паретто-кривая значимости значений показателей университетского рейтинга Гардиан

Из приведенной Паретто-кривой можно сделать вывод о том, что Паретто-оптимизация была проведена разработчиками университетского рейтинга Гардиан на этапе его создания, т.к. все используемые в нем значения показателей имеют достаточно высокую значимость. Когда в модели есть малозначимые факторы, то Паретто-кривая поднимается гораздо резче и потом идет более полого (рисунок 14).

Но при разработке отечественного рейтинга, по-видимому, сначала должно быть проведено пилотное исследование на всех мыслимых показателях, информацию по которым возможно собрать, на не очень большом количестве вузов, участвующих в эксперименте (при этом важно, чтобы вузы должны быть разных направлений подготовки). При этом при пилотном исследовании используется **максимальная** система показателей, которую можно взять из многих известных рейтингов и материалов Минобрнауки РФ.

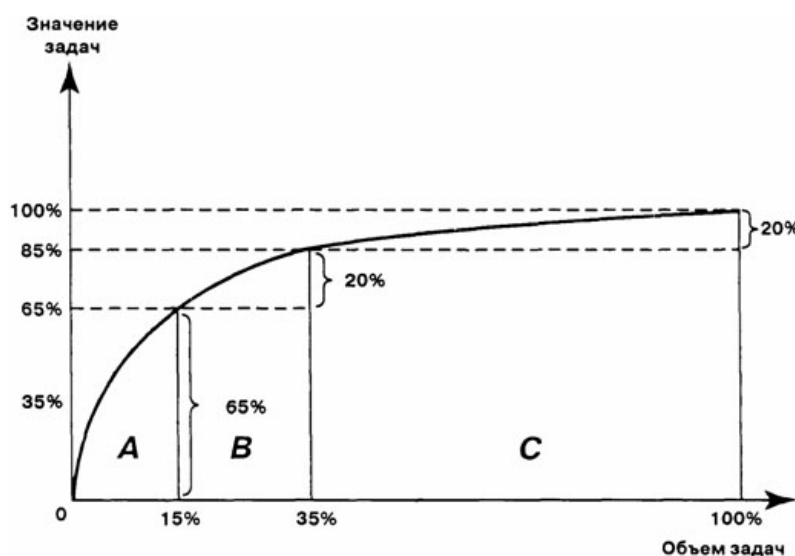


Рисунок 14. Классическая Паретто-кривая⁷⁵

⁷⁵ См., например: <http://yandex.ru/yandsearch?lr=35&text=Паретто-кривая>

Затем необходимо провести Паретто-оптимизацию и разработать минимальную по количеству систему показателей, дающих максимум информации для определения рейтинга вуза (конфигуратор вузовского рейтинга). Таким образом, созданная по этой технологии наукометрическая методика определения рейтинга вуза будет представлять собой методику, интегрирующую многие известные рейтинги, используемые при ее разработке.

После тестирования и сертификации системно-когнитивной модели, построенной на этой системе показателей, ее можно применять в адаптивном режиме.

6.4.4.2. ЭКСПЛУАТАЦИЯ МЕТОДИКИ В АДАПТИВНОМ РЕЖИМЕ

АСК-анализ и система «Эйдос» представляют собой с одной стороны инструмент **разработки**, а с другой стороны среду или оболочку (Run-time system) **эксплуатации** создаваемого интеллектуального приложения.

Это открывает уникальные возможности, которые полностью отсутствуют, когда мы используем приобретаемые у сторонних разработчиков продукты подобных технологий.

Возникает закономерный вопрос о том, в какой степени эти продукты применимы в наших условиях и что они будут измерять, если их применить для российских вузов? Не столкнемся ли мы с ситуацией, когда из-за того, что не могут найти линейку, измеряют размеры предметов с помощью шкалы от наружного термометра, т.е. применяют непригодный для наших целей измерительный инструмент, даже и может быть и качественный, но предназначенный для других целей и других условий. Используя университетский рейтинг Гардиан для оценки российских вузов мы сравниваем их не только друг с другом, но и с зарубежными вузами и как бы отвечаем на вопрос о том, как бы оценивались наши вузы, если бы они оказались за рубежом. Но дело в том, что они находятся у нас и поэтому модели и методов принятия решений, заложенные его разработчиками в этом рейтинге, могут быть **не адекватными** для наших условий, и для приведения их в соответствие с нашими реалиями может быть необходима **локализация** этих моделей и методов.

Имея инструментарий разработки измерительного инструмента мы получаем возможность периодически, например, ежегодно, использовать его для пересоздания модели, с целью учета изменений в моделируемом объекте и других факторов [15].

6.4.5. Выводы. Ограничения и перспективы

Таким образом, АСК-анализ и система «Эйдос» представляют собой современную инновационную (готовую к внедрению) технологию решения задач статистики методами теории информации.

Данная статья может быть использована как описание лабораторной работы по дисциплинам:

- Интеллектуальные системы;
- Инженерия знаний и интеллектуальные системы;
- Интеллектуальные технологии и представление знаний;
- Представление знаний в интеллектуальных системах;
- Основы интеллектуальных систем;
- Введение в нейроматематику и методы нейронных сетей;
- Основы искусственного интеллекта;
- Интеллектуальные технологии в науке и образовании;
- Управление знаниями;
- Автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»;

которые автор ведет в настоящее время⁷⁶, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки).

Этим и другим применениям должно способствовать и то, что *данное приложение вместе с системой «Эйдос»* размещено автором в полном открытом бесплатном доступе по адресу: <https://cloud.mail.ru/public/a5b22d65bc88/Aidos-X-1071503001.rar>. Для установки системы с данным приложением на компьютере достаточно развернуть архив в корневом каталоге на диске С:.

Таким образом, в статье предлагается решение **проблемы**, заключающейся в том, что с одной стороны рейтинг российских вузов востребован, а с другой стороны пока он не создан. Предлагаемая идея решения проблемы состоит в применении отечественной лицензионной инновационной интеллектуальной технологии для этих целей: а именно предлагается применить автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальную систему «Эйдос». Эти методы подробно описываются в этом контексте. Предлагается рассмотреть возможности применения данного инструментария на примере университетского рейтинга Гардиан и рассматриваются его частные критерии (показатели вузов). Указываются источники данных и методика их подготовки для обработки в системе «Эйдос». В соответствии с методологией АСК-анализа описывается установка системы «Эйдос», ввод исходных данных в нее и формализация предметной области, синтез и верификация модели, их отображение и применение для решения задач оценки рейтинга Гардиан для российских вузов и исследования объекта моделирования. Рассматриваются перспективы и пути создания интегрированного рейтинга российских вузов и эксплуатации рейтинга в адаптивном режиме. Указываются ограничения предлагаемого подхода и перспективы его развития.

Конечно, рассматриваемая проблема требует к себе очень серьезного отношения и большого объема работ по совершенствованию инструментария, созданию и исследованию моделей на российских данных. Поэтому предлагаемые в статье решения можно рассматривать не более как идею решения поставленной проблемы и численную иллюстрацию этой идеи, но ни в коем случае не как готовое решение.

ГЛАВА 7. ПРИМЕНЕНИЕ АСК-АНАЛИЗА И ТЕОРИИ ИНФОРМАЦИИ ДЛЯ РЕШЕНИЯ НЕКОТОРЫХ ЗАДАЧ СТАТИСТИКИ

7.1. Исследование статистических распределений методами теории информации

Предлагается теоретическое обоснование, методика численных расчетов и программная реализация решения задач статистики, в частности исследования статистических распределений, методами теории информации. При этом непосредственно на основе эмпирических данных расчетным путем определяется количество информации в наблюдениях, которое используется для анализа статистических распределений. Предлагаемый способ расчета количества информации не основан на предположениях о независимости наблюдений и их нормальном распределении, т.е. является непараметрическим и обеспе-

⁷⁶ http://lc.kubagro.ru/My_training_schedule.doc

чивает корректное моделирование нелинейных систем, а также позволяет сопоставимо обрабатывать разнородные (измеряемые в шкалах различных типов) данные числовой и нечисловой природы, измеряемые в различных единицах измерения. Таким образом, АСК-анализ и система «Эйдос» представляют собой современную инновационную (готовую к внедрению) технологию решения задач статистики методами теории информации. Данный раздел может быть использован как описание лабораторной работы по дисциплинам: интеллектуальные системы; инженерия знаний и интеллектуальные системы; интеллектуальные технологии и представление знаний; представление знаний в интеллектуальных системах; основы интеллектуальных систем; введение в нейроматематику и методы нейронных сетей; основы искусственного интеллекта; интеллектуальные технологии в науке и образовании; управление знаниями; автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»; которые автор ведет в настоящее время, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки)

«... навыки мысли и аналитический аппарат теории информации должны, по-видимому, привести к заметной перестройке здания математической статистики»

А.Н. Колмогоров [2,3,4]

7.1.1. Формулировка проблемы

В статистике существует **проблема** определения закона распределения наблюдений, а затем и определения параметров этого распределения. **Традиционно** эта проблема решается путем проверки *статистических гипотез*, на основе специально разработанных довольно многочисленных *статистических тестов и критериев* с учетом ошибок первого и второго рода.

Эта теория детально разработана и общеизвестна. Однако необходимо отметить, что по ряду причин на практике ей довольно редко пользуются, а когда все же пользуются, то часто делают это некорректно. Довольно многие просто пользуются теми или иными возможностями MS Excel или статистических пакетов и при этом даже не задумываются, что применяемые ими методы являются **параметрическими**, т.е. существенным образом **основаны** на предположении о выполнении для исследуемых наблюдений гипотезы о нормальности распределения. Естественно они и не пытаются проверить, так ли это. Детальный и исчерпывающий на сегодняшний день анализ причин некорректного использования статистических технологий приведен в классической работе [1] и здесь мы не будем на нем останавливаться. Отметим лишь, что этих причин так много и они настолько разнообразны, что на наш взгляд ученый, собирающийся применить статистику в своих исследованиях или при решении задач в своей области науки, по-видимому, практически обречен на ее некорректное использование.

7.1.2. Общая идея предлагаемого решения проблемы

Общая идея решения сформулированной проблемы состоит в предложении *применить* непараметрические методы, в частности *теорию информации*, для решения тех задач, которые традиционно решаются в параметрической статистике.

Конечно, применение теории информации для решения проблем и развития статистики не является абсолютно новой идеей⁷⁷. Как указывает в своих работах [2, 3] профессор А.И.Орлов, сходные идеи развивал еще в середине XX века С.Кульбак [4], а в эпиграф данного раздела вынесено программное высказывание выдающегося российского математика А.Н. Колмогорова: «... навыки мысли и аналитический аппарат теории информации должны, по-видимому, привести к заметной перестройке здания математической статистики», которое содержится в его предисловии к той же книге С.Кульбака и также приведено в работах [2, 3]. В наше время в этом направлении продуктивно работают Дуглас Хаббард [5], а также известный российский математик, разработчик синергетической теории информации В.Б.Вяткин [6-13]⁷⁸.

7.1.3. Опыт применения теории информации в статистике

В работе [1] в разделе «1.1. Проблема множественных проверок статистических гипотез» профессор А.И.Орлов указывает, что регрессионный анализ является параметрическим методом и при множественных проверках статистических гипотез им пользоваться некорректно, т.к. однородные группы, полученные с помощью какого-либо алгоритма классификации (кластеризации), подчиняются не нормальному распределению, а усеченному нормальному.

Имеется определенный положительный опыт решения поставленной проблемы путем применения теории информации.

В статье [14] метод наименьших квадратов (МНК) широко известен и пользуется заслуженной популярностью. Вместе с тем не прекращаются попытки усовершенствования этого метода. Результатом одной из таких попыток является взвешенный метод наименьших квадратов (ВМНК), суть которого в том, чтобы придать наблюдениям вес обратно пропорциональный погрешностям их аппроксимации. Этим самым, фактически, наблюдения игнорируются тем в большей степени, чем сложнее их аппроксимировать. В результате такого подхода формально погрешность аппроксимации снижается, но фактически это происходит путем частичного отказа от рассмотрения «проблемных» наблюдений, вносящих большую ошибку. Если эту идею, лежащую в основе ВМНК довести до крайности (и тем самым до абсурда), то в пределе такой подход приведет к тому, что из всей совокупности наблюдений останутся только те, которые практически точно ложатся на тренд, полученный методом наименьших квадратов, а остальные просто будут проигнорированы. Однако, по мнению автора, фактически это не решение проблемы, а отказ от ее решения, хотя внешне и выглядит как решение. В работе предлагается именно решение, основанное на теории информации: считать весом наблюдения количество информации в аргументе о значении функции. Этот подход был обоснован в рамках нового инновационного метода искусственного интеллекта: метода автоматизированного системно-когнитивного анализа (АСК-анализа) и реализован еще 30 лет назад в его программном инструментарии – интеллектуальной системе «Эйдос» в виде так называемых «когнитивных функций». В данном разделе приводится алгоритм и программная реализация данного подхода, проиллюстрированные на подробном численном примере.

В статье [15] кратко рассматриваются математическая сущность предложенной автором модификации взвешенного метода наименьших квадратов (ВМНК), в котором в качестве весов наблюдений применяется количество информации в них. Предлагается

⁷⁷ Наверное, абсолютно новых идей вообще не существует.

⁷⁸ Может быть синергетическая теория информации, созданная трудами В.Б.Вяткина, также может быть использована для расчета количества информации в наблюдениях, что приведет к еще одному варианту статистики, основанной на теории информации.

два варианта данной модификации ВМНК. В первом варианте взвешивание наблюдений производится путем замены одного наблюдения с определенным количеством информации в нем соответствующим количеством наблюдений единичного веса, а затем к ним применяется стандартный метод наименьших квадратов (МНК). Во втором варианте взвешивание наблюдений производится для каждого значения аргумента путем замены всех наблюдений с определенным количеством информации в них одним наблюдением единичного веса, полученным как средневзвешенное от них, а затем к ним применяется стандартный МНК. Подробно описана методика численных расчетов количества информации в наблюдениях, основанная на теории автоматизированного системно-когнитивного анализа (АСК-анализ) и реализованная в его программном инструментарии – интеллектуальной системе «Эйдос». Приводится иллюстрация предлагаемого подхода на простом численном примере.

Отметим также, что в статье [16] на небольшом численном примере рассматриваются новые математическая модель, алгоритм и результаты агломеративной кластеризации, основные отличия которых от ранее известных стоят в том, что: а) в них параметры обобщенного образа кластера не вычисляются как средние от исходных объектов (классов) или центры тяжести, а определяются с помощью той же самой базовой когнитивной операции АСК-анализа, которая применяется и для формирования обобщенных образов классов на основе примеров объектов и которая действительно обеспечивает обобщение; б) в качестве критерия сходства используется не евклидово расстояние или его варианты, а интегральный критерий неметрической природы: «суммарное количество информации», применение которого теоретически корректно и дает хорошие результаты в неортонормированных пространствах, которые обычно и встречаются на практике; в) кластерный анализ проводится не на основе исходных переменных или матрицы сопряженности, зависящих от единиц измерения по осям, а в когнитивном пространстве, в котором по всем осям (описательным шкалам) используется одна единица измерения: количество информации, и поэтому результаты кластеризации не зависят от исходных единиц измерения признаков объектов. Имеется и ряд других менее существенных отличий. Все это позволяет получить результаты кластеризации, понятные специалистам и поддающиеся содержательной интерпретации, хорошо согласующиеся с оценками экспертов, их опытом и интуитивными ожиданиями, что часто представляет собой проблему для классических методов кластеризации. Описанные методы теоретически обоснованы в системно-когнитивном анализе (СК-анализ) и реализованы в его программном инструментарии – интеллектуальной системе «Эйдос».

Таким образом, в работах автора [14, 15 и 16] по сути, намечается путь решения проблемы построения непараметрического регрессионного анализа, основанного на теории информации, в том числе и для его применения в относительно однородных группах, полученных путем когнитивной кластеризации.

7.1.4. Некоторые задачи статистики, которые могли бы быть решены методами теории информации

Задача № 1: *проверка статистических гипотез.* По сути, эта задача является частным вариантом задачи распознавания образов, т.к. в ней по первичным и вторичным (расчетным) признакам наблюдений необходимо определить вид статистического распределения и его параметры. А теория информации хорошо позволяет решать подобные задачи распознавания, в том числе и в условиях зашумленности исходных данных.

Задача № 2: *исследование влияния уровня системности действующих на объекты наблюдения факторов на степень отклонения статистического распределения их характеристик от нормального.* Данная задача тесно связана с системным обобщени-

ем математики, в частности системной теорией информации, которые были предложены автором в ряде работ [см., например: 17, 35, 36]. Решение этой задачи может заложить основы системного обобщения статистики (системной статистики) в результате применения идей системного обобщения математики в статистике. Эта задача тесно связана с **Центральными предельными теоремами (ЦПТ) или законом больших чисел** теории вероятностей, утверждающих, что сумма достаточно большого количества слабо зависимых случайных величин, имеющих примерно одинаковые масштабы (ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), имеет статистическое распределение, стремящееся (сходящееся) к нормальному распределению. С позиций системного обобщения математики независимые зависимые случайные величины представляют собой *множество* случайных величин. Если же между ними есть зависимости, то их уже нельзя (вернее можно, но это некорректно) рассматривать как множество и более адекватным является представление о них, как о *системе* случайных величин [17]. Система имеет эмерджентные свойства, которых не было у ее элементов и эти свойства тем ярче выражены, чем выше уровень системности. Автором предложено несколько разных вариантов коэффициентов эмерджентности, которые представляют собой количественные информационные меры уровня системности и степени детерминированности систем⁷⁹. Система факторов влияет на систему не так, как их сумма, т.е. нелинейно. В результате статистическое распределение системы случайных величин отклоняется от нормального тем в большей степени, чем выше уровень системности и нелинейность. Таким образом, вся параметрическая статистика описывает только линейные системы, а для нелинейных систем она является неадекватной. Отметим, что к нелинейным системам, имеющим высокий уровень системности и ярко выраженные эмерджентные (синергетические) свойства, относятся все живые системы, искусственные и естественные экосистемы, биоценозы, системы с участием людей (социально-экономические, психологические, культурные, политические), вообще все сложные и большие системы.

Задача № 3: *нахождение информативных подмножеств признаков в регрессионном анализе и в автоматизированных системах управления.* Данная задача сформулирована профессором А.И.Орловым в работе [1] следующим образом: «...в большинстве важных для практики случаев статистические свойства процедур анализа данных, основанных на множественных проверках, остаются пока неизвестными. Примерами являются процедуры нахождения информативных подмножеств признаков (коэффициенты для таких и только таких признаков отличны от 0) в регрессионном анализе или выявления отклонений параметров в автоматизированных системах управления». Решение этой задачи давно (еще в 1979 году) предложено автором в теории АСК-анализа и реализовано в его программном инструментарии – интеллектуальной системе «Эйдос» и представляет собой базовую когнитивную операцию «Абстрагирование» [18]. Это решение основано на использовании варибельности количества информации в значении аргумента (в признаке) о значении функции (классе) в качестве меры информативности (ценности, дискриминантной, дифференцирующей способности) данного значения аргумента (признака), т.е. его полезности для различения классов.

Задача № 4: *"стыковка" статистических процедур.* Данная задача также сформулирована профессором А.И.Орловым в работе [1]: «Проблема множественных про-

⁷⁹ Отметим, что идеи оказались лакомым кусочком для плагиаторов. Об этом хорошо написано в статье В.Б.Вяткина «Групповой плагиат: от студента до министра» <http://trv-science.ru/2011/11/08/gruppovojj-plagiat-ot-studenta-do-ministra/> и других его статьях на эту тему. Чтобы убедиться в этом достаточно сделать запрос: «Коэффициенты эмерджентности».

верок статистических гипотез – часть более общей проблемы "стыковки" (сопряжения) статистических процедур. Дело в том, что каждая процедура может применяться лишь при некоторых условиях, а в результате применения предыдущих процедур эти условия могут нарушаться». Решению очень сходной задачи посвящен АСК-анализ, в котором с единых позиций теории информации рассматривается полная необходимая и достаточная система (конфигуратор) базовых когнитивных операций [19, 20, 21]. По сути можно сказать, что грандиозное здание статистики построено без единого плана, т.е. не системно, и в результате отдельные его конструкции не всегда гармонично сочетаются друг с другом и не образуют единого целого. Можно, конечно, попытаться все это упорядочить и расписать на языке непосвященных, а также снабдить их программным инструментарием, но эта задача в настоящее время, похоже, никем не ставится. Автор предлагает другое, как это ни парадоксально, но возможно более простое решение: не реформировать старое, а построить рядом новое здание системной статистики и сделать это по единому проекту, единой теоретической и методологической основе теории информации. По крайней мере, в совершенно аналогичной ситуации с автоматизацией системного анализа второй вариант решения оказался более эффективным, чем другие [19]. В частности оказалось возможным создать и единую систему, основанную на этой единой теоретической и методологической основе теории информации: интеллектуальную систему «Эйдос». Это вселяет надежду на решение проблемы, о которой в работе [1] профессор А.И.Орлов писал: «Математическая статистика демонстрирует ... виртуозную математическую технику для анализа частных случаев и полную беспомощность при выдаче практических рекомендаций».

Задача № 5: *конструирование системной информационной меры взаимосвязи двух векторов, аналогичной коэффициенту корреляции.* Это сделано в АСК-анализе и реализовано в системе «Эйдос» и описано автором в монографии [37] еще в 1996 году в режиме «Содержательное сравнение двух классов». Суть идеи состоит в том, что:

- а) при расчете коэффициента корреляции учитываются не сами значения аргумента, а количество информации о значениях функции, которое в них содержится;
- б) учитываются не только вклад в сходство-различие значений аргумента с одинаковыми индексами, но и все их сочетания⁸⁰.

Разумеется, этим перечень задач статистики, которые на взгляд автора могли бы быть решены с методами теории информации, в частности АСК-анализа и системы «Эйдос», далеко не исчерпывается. Конечно, здесь возникает естественный вопрос о том, какие вообще задачи статистики могут быть решены с помощью теории информации. На это вопрос можно было бы ответить другим вопросом: «А какие задачи статистики не могут быть решены с помощью теории информации?» На наш взгляд *любая* наука, а не только статистика, в процессе исследования и как его результат получает определенную информацию об объекте исследования. Поэтому теория информации в определенном смысле является метанаукой имеющей не меньшую общность, чем философия, но в отличие от нее являющаяся естественной высокоматематизированной наукой, имеющей свой программный инструментарий. Даже мысленный эксперимент Альберта Эйнштейна с движущимся поездом и источниками света на платформе, на основе которого в теории относительности формируется представление об одновременности и времени, фактически является не более чем описанием системы передачи информации в пространстве-времени с помощью световых сигналов. Даже когда мы узнаем, чему равен предел функции или интеграл, то даже если мы об этом и не знаем, то все равно на самом деле мы тоже получаем об этом информацию, количество которой можно посчитать и выразить в битах. Поэтому любая задача в любой области науки

⁸⁰ См. главу III: <http://lc.kubagro.ru/aidos/aidos96/3.htm>

требует для своего решения практического применения теории информации, которое чаще осуществляется всего неосознанно и на качественном уровне. АСК-анализ и его программный инструментарий – система «Эйдос», реализованная в универсальной постановке, не зависящей от предметной области, позволяют осуществить это на осознанном уровне.

Задача № 6: *исследование информационных моделей статистических распределений.* Решение этой задачи включает кластерный и конструктивный анализ распределений, их информационные портреты и многие другие исследования с использованием возможностей АСК-анализа и системы «Эйдос» [21].

Далее в данном разделе кратко рассмотрим возможный вариант применения теории информации для получения определенной информации о статистическом распределении наблюдений. Это можно рассматривать как подготовку к решению сформулированной выше задачи №1 и некоторых других сформулированных задач.

7.1.5. Когнитивные функции, как необходимый элемент решения проблемы

Данная идея состоит в том, чтобы *рассматривать статистические распределения как когнитивные функции.* Это открывает перспективы использования теории информации для анализа функций, в т.ч. и статистических распределений.

В АСК-анализе предложено новое понятие **когнитивных функций**, которое рассмотрено и развито в ряде работ автора и соавторов [17, 22–34] и поэтому здесь нет смысла подробно останавливаться на этом понятии. Отметим лишь суть. В работе [17] кратко рассматриваются классическое понятие функциональной зависимости в математике, определяются ограничения применимости этого понятия для адекватного моделирования реальности и формулируется проблема, состоящая в поиске такого обобщения понятия функции, которое было бы более пригодно для адекватного отражения причинно-следственных связей в реальной области. Далее рассматривается теоретическое и практическое решения поставленной проблемы, состоящие в том, что:

а) предлагается универсальный не зависящий от предметной области способ вычисления количества информации в значении аргумента о значении функции, т.е. когнитивные функции;

б) предлагается программный инструментарий: интеллектуальная система «Эйдос», позволяющая на практике осуществлять эти расчеты, т.е. строить когнитивные функции на основе фрагментированных зашумленных эмпирических данных большой размерности.

Предлагаются понятия нередуцированных, частично и полностью редуцированных прямых и обратных, позитивных и негативных когнитивных функций и метод формирования редуцированных когнитивных функций, являющийся вариантом известного взвешенного метода наименьших квадратов, отличающимся от стандартного **ВМНК** *учетом в качестве весов наблюдений количества информации в значениях аргумента о значениях функции.*

7.1.6. Математическая сущность предлагаемого решения проблемы

Идея применения теории информации для исследования статистических распределений проста и базируется на двух вполне очевидных предпосылках.

Предпосылка № 1. Любое исследование, в т.ч. исследование статистических распределений, представляет собой процесс получения информации об объекте исследования. Этот процесс включает: источник информации, канал передачи информации и получатель информации, т.е. представляет собой информационно-измерительную сис-

тему. Поэтому применение теории информации для построения такой системы является совершенно естественным и очевидным⁸¹.

Предпосылка № 2. Понятия «Информация» и «Статистического распределение» тесно взаимосвязаны. Широко известен хрестоматийный пример определения количественной меры информации Хартли через понятие равномерного распределения случайной величины: количество информации $I = \log_2 N$ по Хартли равно количеству информации, которое мы получаем, когда узнаем, что равномерно распределенная случайная величина попала в некоторый определенный i -й интервал: один из N равных интервалов. На подобных соображениях основан известный метод Монте-Карло. В работах [35, 36] предложено развитие этих идей с использованием классической статистики Больцмана, также ее квантовых обобщений статистик Ферми-Дирака и Бозе-Эйнштейна. Оказалось, что эти статистики математически тесно связаны с системным обобщением теории информации, предложенным автором.

В данной работе ставится задача дать ответ на несколько простых вопросов:

Вопрос 1-й. Какое количество информации о принадлежности случайной величины к нескольким нормальным распределениям с разными параметрами мы получаем, когда узнаем, что она попала в некоторый определенный i -й интервал: один из N равных интервалов?

Вопрос 2-й. Какое суммарное количество информации о степени сходства эмпирического распределения наблюдений с нормальными распределениями с различными параметрами мы получаем, зная частоты попадания случайной величины в каждый из интервалов при N равных интервалов?

Вопрос 3-й. Попадание случайной величины в какие интервалы при N равных интервалах является более характерным и в какие менее характерным для нормальных распределений с различными параметрами?

Для ответа на эти вопросы нам потребуется универсальный не зависящий от предметной области способ вычисления количества информации в наблюдениях, который мы рассмотрим ниже.

7.1.7. Математическая модель и методика численных расчетов количества информации в наблюдениях

Ниже в наиболее упрощенном виде приводится методика численных расчетов количества информации в наблюдениях, основанная на теории автоматизированного системно-когнитивного анализа (АСК-анализ) и реализованная в его программном инструментарии – интеллектуальной системе «Эйдос» [21].

Для удобства рассмотрения введем следующие обозначения:

i – индекс значения аргумента;

j – индекс значения функции;

M – количество значений аргумента;

W – количество значений функции;

N_{ij} – количество встреч j -го значения функции при i -м значении аргумента;

$N_{i\Sigma}$ – суммарное количество наблюдений при i -м значении аргумента по всей выборке;

$N_{\Sigma j}$ – суммарное количество наблюдений j -го значения функции по всей выборке;

$N_{\Sigma\Sigma}$ – суммарное количество наблюдений по всей выборке;

⁸¹ и очень странно, почему это до сих пор широко не осознанно

I_{ij} – количество информации в i -м значении аргумента о том, что функция имеет j -е значение, т.е. это количество информации в наблюдении (i, j) ;

Ψ – нормировочный коэффициент (Е.В.Луценко, 1979), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли в равновероятном детерминистском случае;

$P_{i\Sigma}$ – безусловная относительная частота встречи i -го значения аргумента в обучающей выборке;

P_{ij} – условная относительная частота встречи j -го значения функции при i -м значении аргумента.

Используя исходную выборку эмпирических наблюдений посчитаем матрицу абсолютных частот (таблица 1):

Таблица 1 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

		Классы					Сумма
		I	...	j	...	W	
Значения факторов	I	N_{11}		N_{1j}		N_{1W}	
	...						
	i	N_{i1}		N_{ij}		N_{iW}	$N_{i\Sigma} = \sum_{j=1}^W N_{ij}$
	...						
	M	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков				$N_{\Sigma j} = \sum_{i=1}^M N_{ij}$			$N_{\Sigma\Sigma} = \sum_{i=1}^M \sum_{j=1}^W N_{ij}$

Алгоритм формирования матриц абсолютных частот и условных и безусловных процентных распределений.

Объекты обучающей выборки описываются векторами (массивами) $\vec{L} = \{L_i\}$ имеющих у них признаков:

$\vec{L} = \{L_i\} = n$, если у объекта i -й признак встречается n раз.

Первоначально в матрице абсолютных частот все значения равны нулю. Затем организуется цикл по объектам обучающей выборки. Если у предъявленного объекта, относящегося к j -му классу, есть i -й признак, то:

$$N_{ij} = N_{ij} + 1; N_{i\Sigma} = N_{i\Sigma} + 1; N_{\Sigma j} = N_{\Sigma j} + 1; N_{\Sigma\Sigma} = N_{\Sigma\Sigma} + 1 \quad (1)$$

На основе анализа матрицы частот (табл. 1) классы можно сравнивать по наблюдаемым частотам признаков только в том случае, если количество объектов по всем классам **одинаково**, как и **суммарное количество признаков по классам**. Если же они отличаются, то корректно сравнивать классы можно только по условным и безусловным относительным частотам (оценкам вероятностей) наблюдений признаков, посчитанных на основе матрицы частот (табл. 1) в соответствии с выражениями (2), в резуль-

тате чего получается матрица условных и безусловных процентных распределений (табл. 2):

$$\begin{aligned}
 P_{ij} &= \frac{N_{ij}}{N_{\Sigma j}}; P_i = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}; P_j = \frac{N_{\Sigma j}}{N_{\Sigma\Sigma}}; \\
 N_{i\Sigma} &= \sum_{j=1}^W N_{ij}; N_{\Sigma j} = \sum_{i=1}^M N_{ij}; \\
 N_{\Sigma\Sigma} &= \sum_{i=1}^M N_{i\Sigma} = \sum_{j=1}^W N_{\Sigma j} = \sum_{i=1}^M \sum_{j=1}^M N_{ij}
 \end{aligned}
 \tag{2}$$

Таблица 2 – МАТРИЦА УСЛОВНЫХ И БЕЗУСЛОВНЫХ ПРОЦЕНТНЫХ РАСПРЕДЕЛЕНИЙ

		Классы					Безусловная вероятность признака
		<i>1</i>	...	<i>j</i>	...	<i>W</i>	
Значения факторов	<i>1</i>	P_{11}		P_{1j}		P_{1W}	
	...						
	<i>i</i>	P_{i1}		P_{ij}		P_{iW}	$P_{i\Sigma}$
	...						
	<i>M</i>	P_{M1}		P_{Mj}		P_{MW}	
Безусловная вероятность класса				$P_{\Sigma j}$			

Далее произведем расчет количества информации в наблюдениях в соответствии с выражением (3):

$$I_{ij} = \text{Log}_2 \left(\frac{P_{ij}}{P_{i\Sigma}} \right)^\Psi
 \tag{3}$$

С учетом (2) преобразуем (3) к виду (4):

$$I_{ij} = \Psi \cdot \text{Log}_2 \frac{N_{ij} N_{\Sigma\Sigma}}{N_{i\Sigma} N_{\Sigma j}}
 \tag{4}$$



А.А.Харкевич

Здесь Ψ – упрощенная форма коэффициента эмерджентности А.Харкевича (10), предложенный автором в 1979 году и названный так в честь известного советского ученого, внесшего большой вклад в теорию информации, на работах которого основана излагаемая методика численных расчетов количества информации в наблюдениях.

$$\Psi = \frac{\text{Log}_2 W}{\text{Log}_2 N_{\Sigma\Sigma}} \quad (5)$$

Используя выражения (3) и (5) на основе таблицы 2 рассчитывается *матрицу информативностей* (таблица 3). Она также может быть получена :непосредственно из таблицы 1 с использованием выражений (4) и (5):

Таблица 3 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ

		Классы					Значимость фактора
		<i>I</i>	...	<i>j</i>	...	<i>W</i>	
Значения факторов	<i>1</i>	I_{11}		I_{1j}		I_{1W}	$\sigma_{1\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	<i>i</i>	I_{i1}		I_{ij}		I_{iW}	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	<i>M</i>	I_{M1}		I_{Mj}		I_{MW}	$\sigma_{M\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Здесь – \bar{I}_i это *среднее* количество информации в *i-м* значении фактора:

$$\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ij}$$

Когда количество информации $I_{ij} > 0$ – *i-й* фактор способствует переходу объекта управления в *j-е* состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это. В векторе *i-го* фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе *j-го* состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, данная модель позволяет рассчитать, какое количество информации содержится в любом факте о наступлении любого события в любой предметной области, причем для этого не требуется повторности этих фактов и событий. Если данные повторности осуществляются и при этом наблюдается некоторая вариабельность значений факторов, обуславливающих наступление тех или иных событий, то модель обеспечивает многопараметрическую типизацию, т.е. синтез обобщенных образов классов или категорий наступающих событий с количественной оценкой степени и знака влияния на их наступление различных значений факторов. Причем эти значения факторов могут быть как количественными, так и качественными и измеряться в любых единицах измерения, в любом случае в модели оценивается количество информации, которое в них содержится о наступлении событий, переходе объекта управления в определенные состояния или, просто, о его принадлежности к тем или иным классам. Другие способы метризации приведены в работе [20].

Ниже на простом численном примере мы кратко рассмотрим технологию, позволяющую на практике и в любой предметной области посчитать, какое количество информации содержится в наблюдении. В связи с ограничениями на объем статьи автор не имеет возможности полностью раскрыть все позиции на приведенных ниже скриншотах и рисунках, т.е. фактически предполагается некоторое предварительное знакомство читателя с системой «Эйдос». Если же такое знакомство недостаточно полное, то автор отсылает автора к публикациям в списке литературы и к сайту: <http://lc.kubagro.ru/>.


7.1.8. Краткий численный пример

Скачиваем и устанавливаем систему «Эйдос». Это наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с исходными текстами, находящаяся в полном открытом бесплатном доступе (около 50 Мб). Обновление имеет объем около 3 Мб.⁸²


ИНСТРУКЦИЯ

по скачиванию и установке системы «Эйдос» (объем около 50 Мб)


Система не требует инсталляции, не меняет никаких системных файлов и содержимого папок операционной системы, т.е. является портативной (portable) программой. Но чтобы она работала необходимо аккуратно выполнить следующие пункты.

1. Скачать самую новую на текущий момент версию системы «Эйдос-X++» по ссылке: <http://lc.kubagro.ru/a.rar> или: <http://lc.kubagro.ru/Aidos-X.exe> (ссылки для обновления системы даны в режиме 6.2).
2. Разархивировать этот архив в любую папку с правами на запись с коротким латинским именем и путем доступа, включающим только папки с такими же именами (лучше всего в корневой каталог какого-нибудь диска).
3. Запустить систему. Файл запуска: _AIDOS-X.exe *
4. Задать имя: 1 и пароль: 1 (потом их можно поменять в режиме 1.2).
5. Перед тем как запустить новый режим НЕОБХОДИМО ЗАВЕРШИТЬ предыдущий (Help можно не закрывать). Окна закрываются в порядке, обратном порядку их открытия.


⁸² <http://lc.kubagro.ru/aidos/Aidos-X.htm>

* Разработана программа: «_START_AIDOS.exe», полностью снимающая с пользователя системы «Эйдос-X++» заботу о проверке наличия и скачивании обновлений. Эту программу надо просто скачать по ссылке: http://lc.kubagro.ru/Install_Aidos-X/_START_AIDOS.exe, поместить в папку с исполнимым модулем системы и всегда запускать систему с помощью этого файла.


При запуске программы _START_AIDOS.EXE система Эйдос не должна быть запущена, т.к. она содержится в файле обновлений и при его разархивировании возникнет конфликт, если система будет запущена.

1. Программа _START_AIDOS.exe определяет дату системы Эйдос в текущей папке, и дату обновлений на FTP-сервере не скачивая их, и, если система Эйдос в текущей папке устарела, скачивает обновления. (Если в текущей папке нет исполнимого модуля системы Эйдос, то программа пытается скачать полную инсталляцию системы, но не может этого сделать из-за ограниченной функциональности демо-версии библиотеки Xb2NET.DLL).

2. После этого появляется диалоговое окно с сообщением, что надо сначала разархивировать систему, заменяя все файлы (опция: «Yes to All» или «OverWrite All»), и только после этого закрыть данное окно.

3. Потом программа _START_AIDOS.exe запускает обновления на разархивирование. После окончания разархивирования окно архиватора с отображением стадии процесса исчезает.

4. После закрытия диалогового окна с инструкцией (см. п.2), происходит запуск обновленной версии системы Эйдос на исполнение.

Для работы программы _START_AIDOS.exe необходима библиотека: Xb2NET.DLL, которую можно скачать по ссылке: http://lc.kubagro.ru/Install_Aidos-X/Xb2NET.DLL. Перед первым запуском этой программы данную библиотеку необходимо скачать и поместить либо в папку с этой программой, а значит и исполнимым модулем системы «Эйдос-X++», либо в любую другую папку, на которую в операционной системе прописаны пути поиска файлов, например в папку: c:\Windows\System32\. Эта библиотека стоит около 500\$ и у меня ее нет, поэтому я даю только бесплатную демо-версию, которая выдает сообщение об ограниченной функциональности, но для наших целей ее достаточно.

Лицензия:

Автор отказывается от какой бы то ни было ответственности за последствия применения или не применения Вами системы «Эйдос».

Проще говоря, пользуйтесь если понравилось, а если не понравилось – сотрите и забудьте, а лучше вообще не скачивайте.

В диспетчере приложений системы «Эйдос» (режим 1.3) кликаем по кнопке «Добавить учебное приложение» и выбираем лабораторную работу 2.05 «Исследование нормального распределения» (рисунок 1):

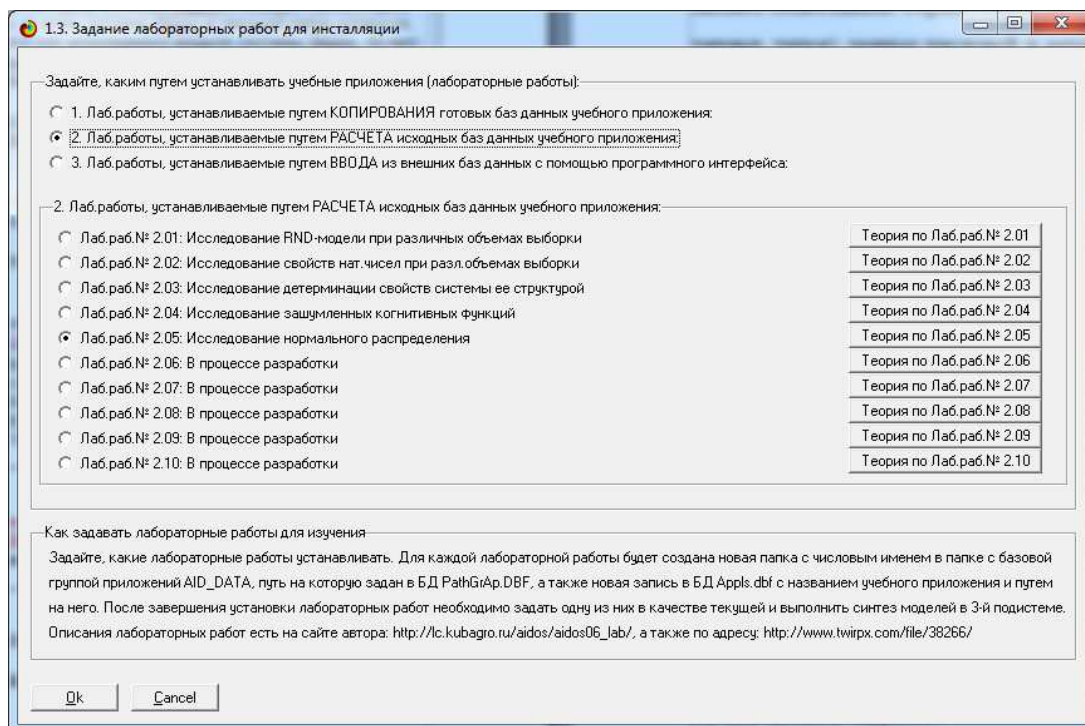


Рисунок 1 – Выбор режима инсталляции лабораторной работы

На экранной форме (рисунок 2) задаем параметры создаваемых распределений случайной величины:

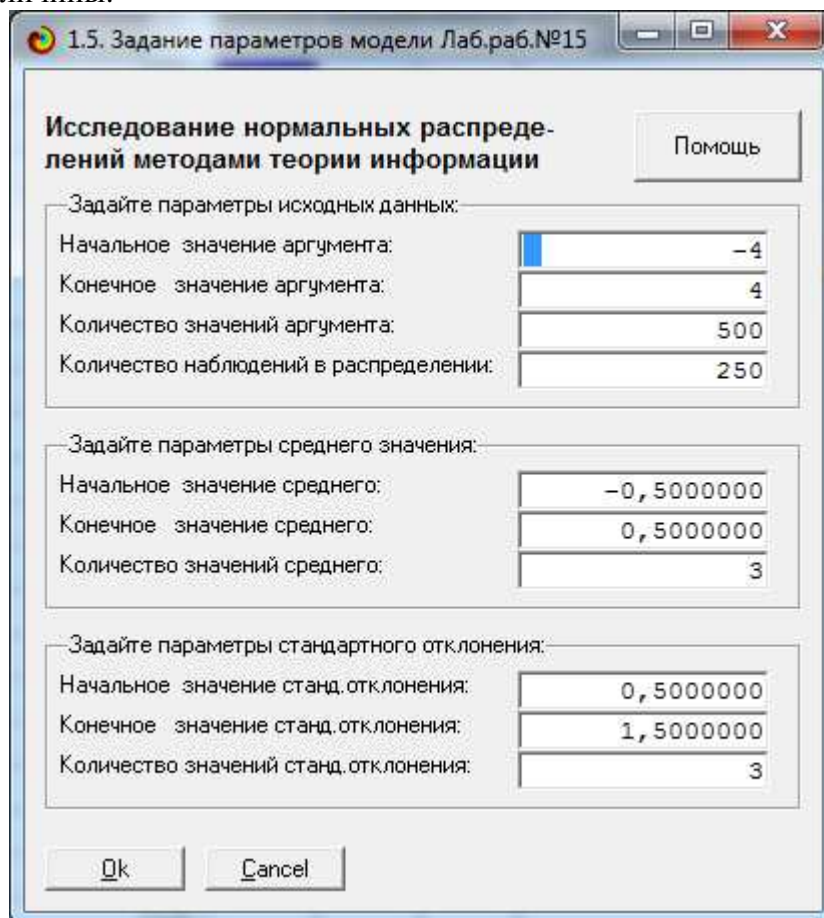


Рисунок 2 – Экранная форма задания параметров генерируемых распределений

Этапы генерации распределений приведены на рисунке 3:

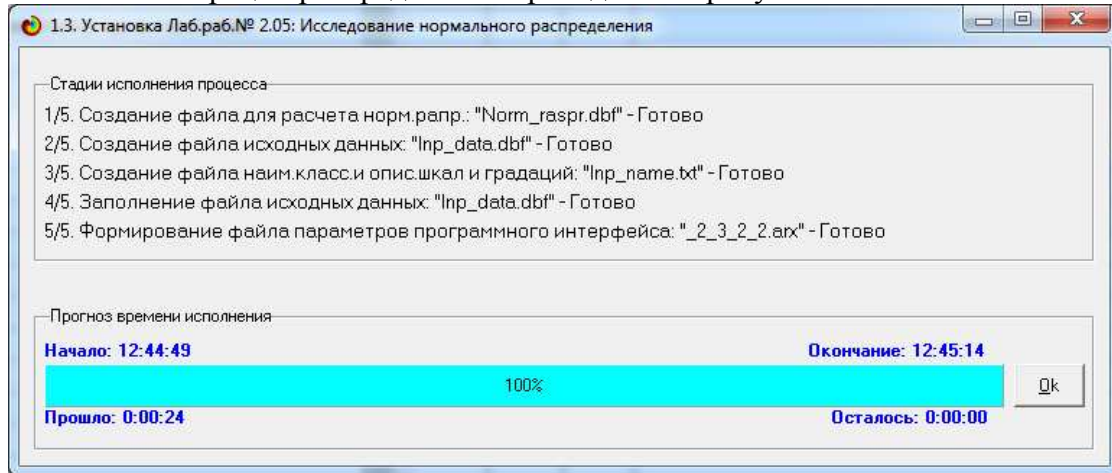


Рисунок 3 – Экранная форма отображения стадии процесса генерации исследуемых распределений случайной величины

В результате созданы следующие распределения (рисунок 4), отличающиеся сочетаниями параметров среднего и стандартного отклонения, заданными в режиме, экранная форма которого приведена на рисунке 2:

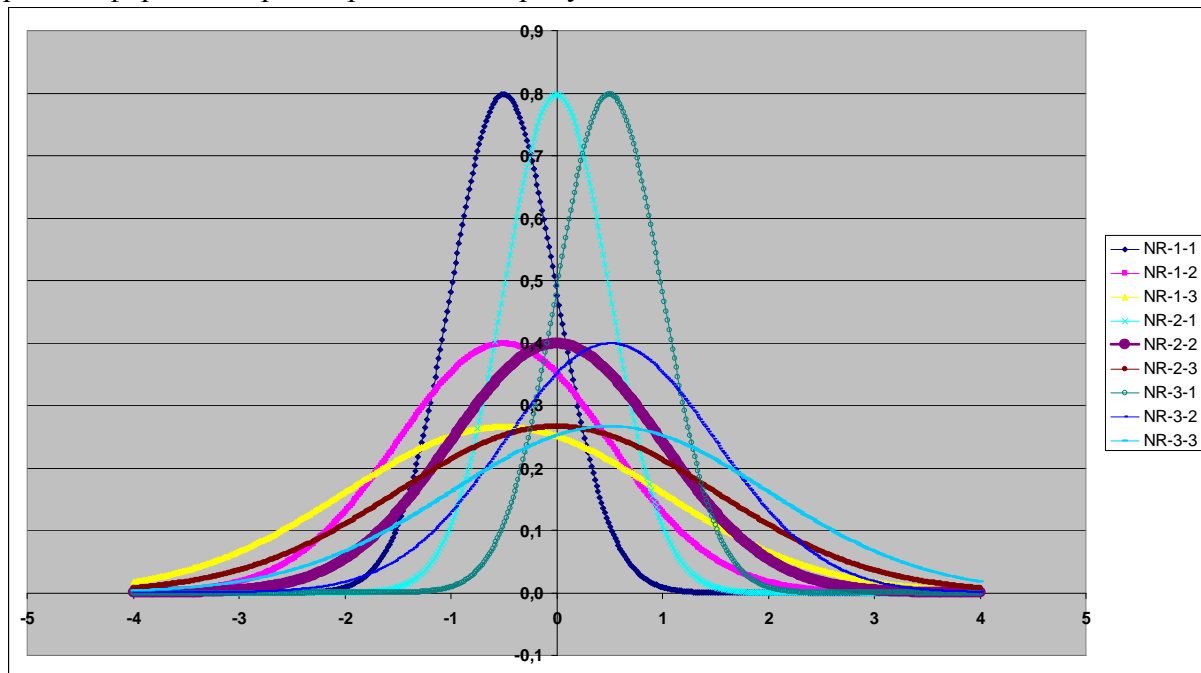


Рисунок 4. Исследуемые нормальные распределения, полученные при различных сочетаниях параметров среднего и стандартного отклонения

В таблице 1 приведены параметры созданных нормальных распределений (на рисунке 4 и в таблице 1 выделено стандартное нормальное распределение со средним значением 0 и стандартным отклонением 1):

Таблица 1 – Параметры созданных нормальных распределений

	GAUSS 1-1	GAUSS 1-2	GAUSS 1-3	GAUSS 2-1	GAUSS 2-2	GAUSS 2-3	GAUSS 3-1	GAUSS 3-2	GAUSS 3-3
Среднее	-0,50	-0,50	-0,50	0,00	0,00	0,00	0,50	0,50	0,50
Ст.отклонение	0,50	1,00	1,50	0,50	1,00	1,50	0,50	1,00	1,50

На рисунке 5 приведена экранная форма задания параметров генерации интервальных моделей, формируемые по умолчанию:

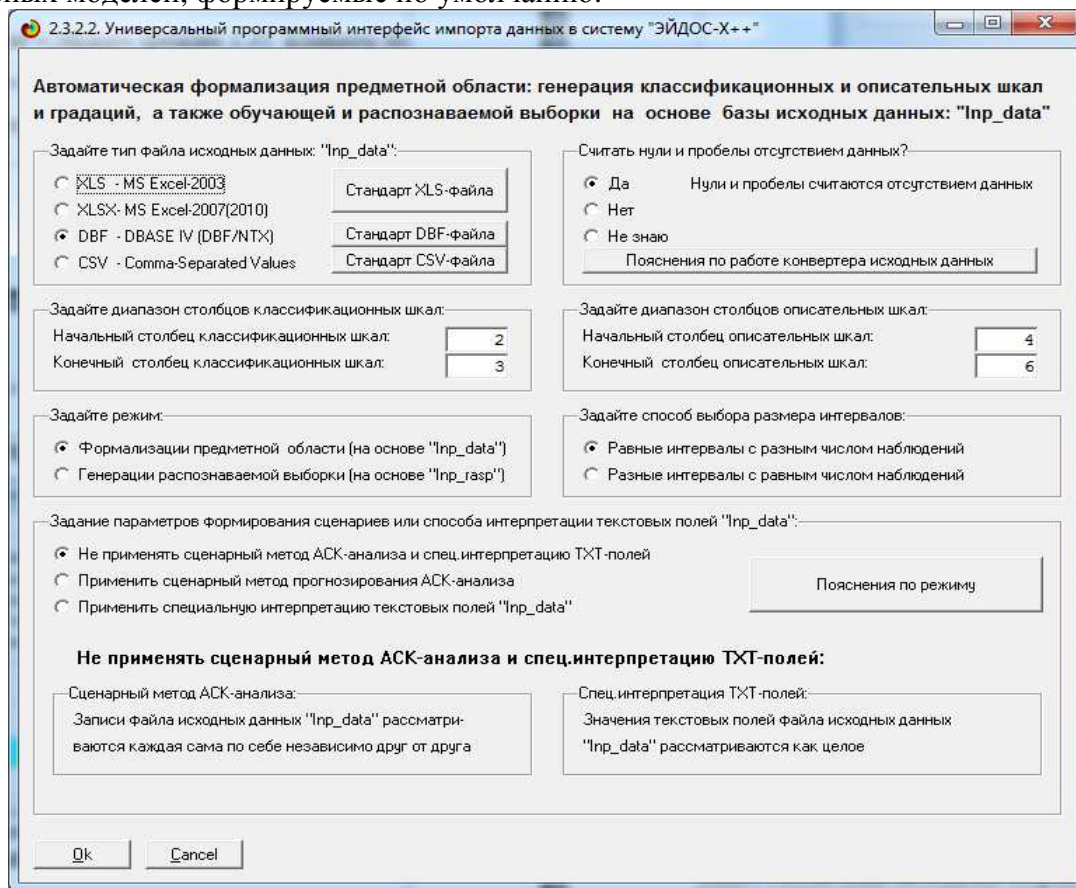


Рисунок 5 – Экранная форма задания параметров генерации интервальных моделей, формируемых по умолчанию

На рисунке 6 приведен внутренний калькулятор параметров интервальных моделей:

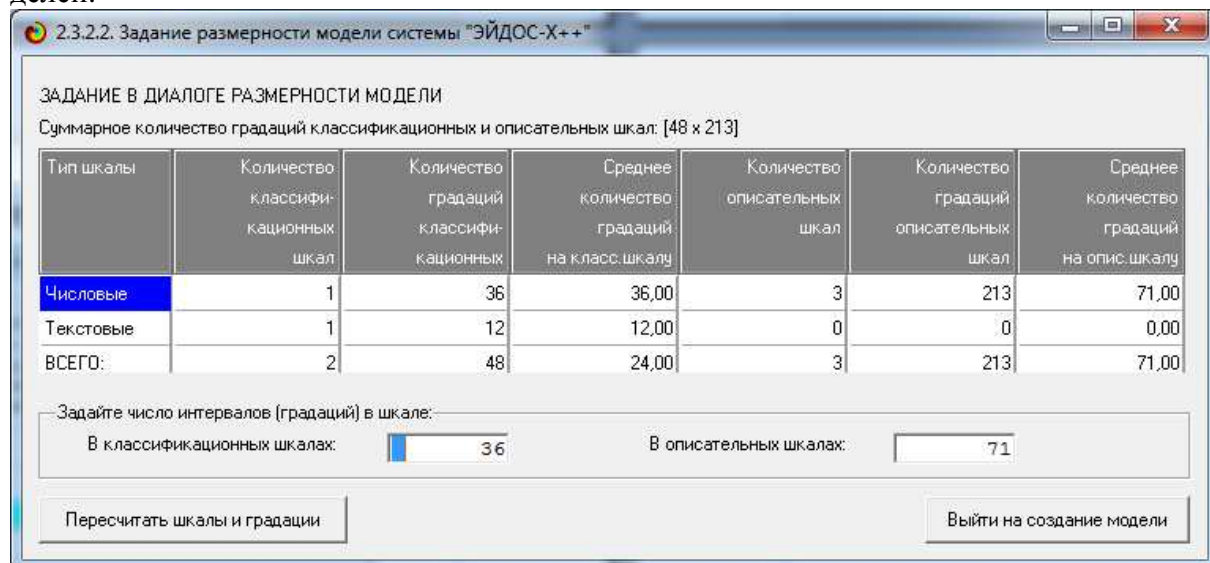


Рисунок 6 – Экранная форма внутреннего калькулятора параметров интервальных моделей

На рисунке 7 приведена форма отображения стадии процесса импорта исходных данных интервальных моделей:

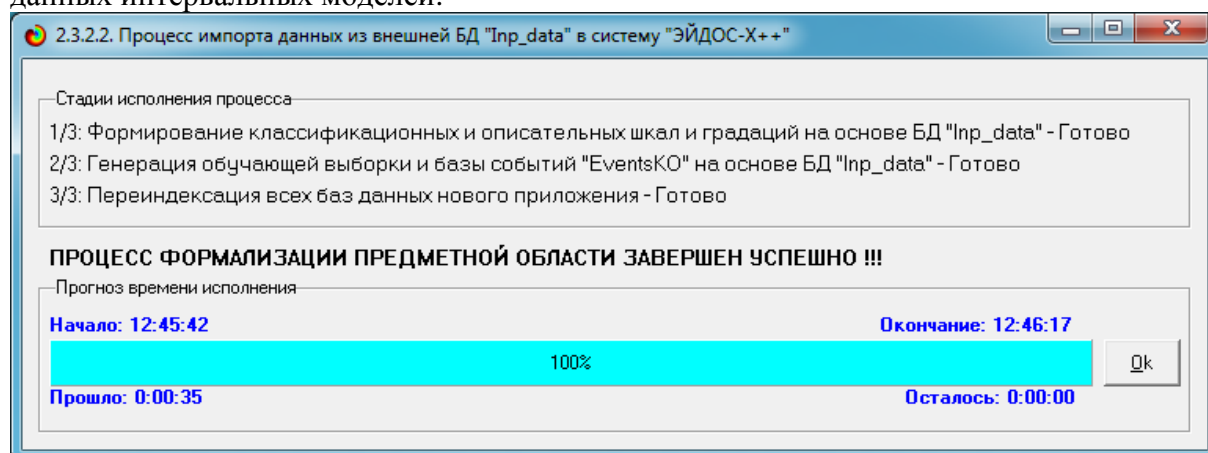


Рисунок 7 – Экранная форма отображения стадии процесса импорта исходных данных интервальных моделей

На рисунке 8 приведена экранная форма задания параметров расчета статистических и информационных интервальных моделей, а на рисунке 8 – стадии процесса их расчета:

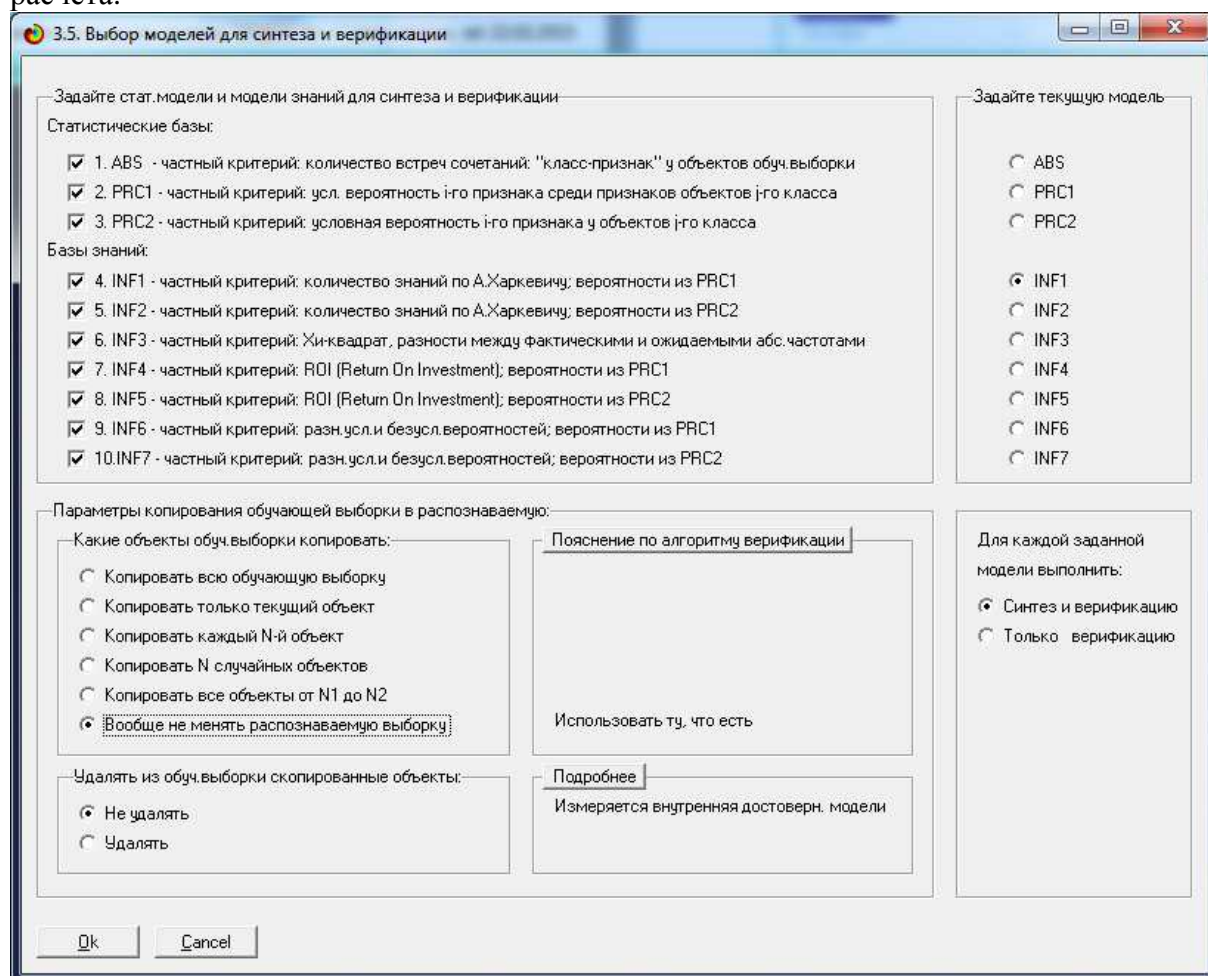


Рисунок 8 – Экранная форма задания параметров расчета статистических и информационных интервальных моделей

КОГНИТИВНАЯ ФУНКЦИЯ:

Приложение: Лаб.раб.№ 2.05: Исследование нормального распределения
Модель: Inf1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRCI

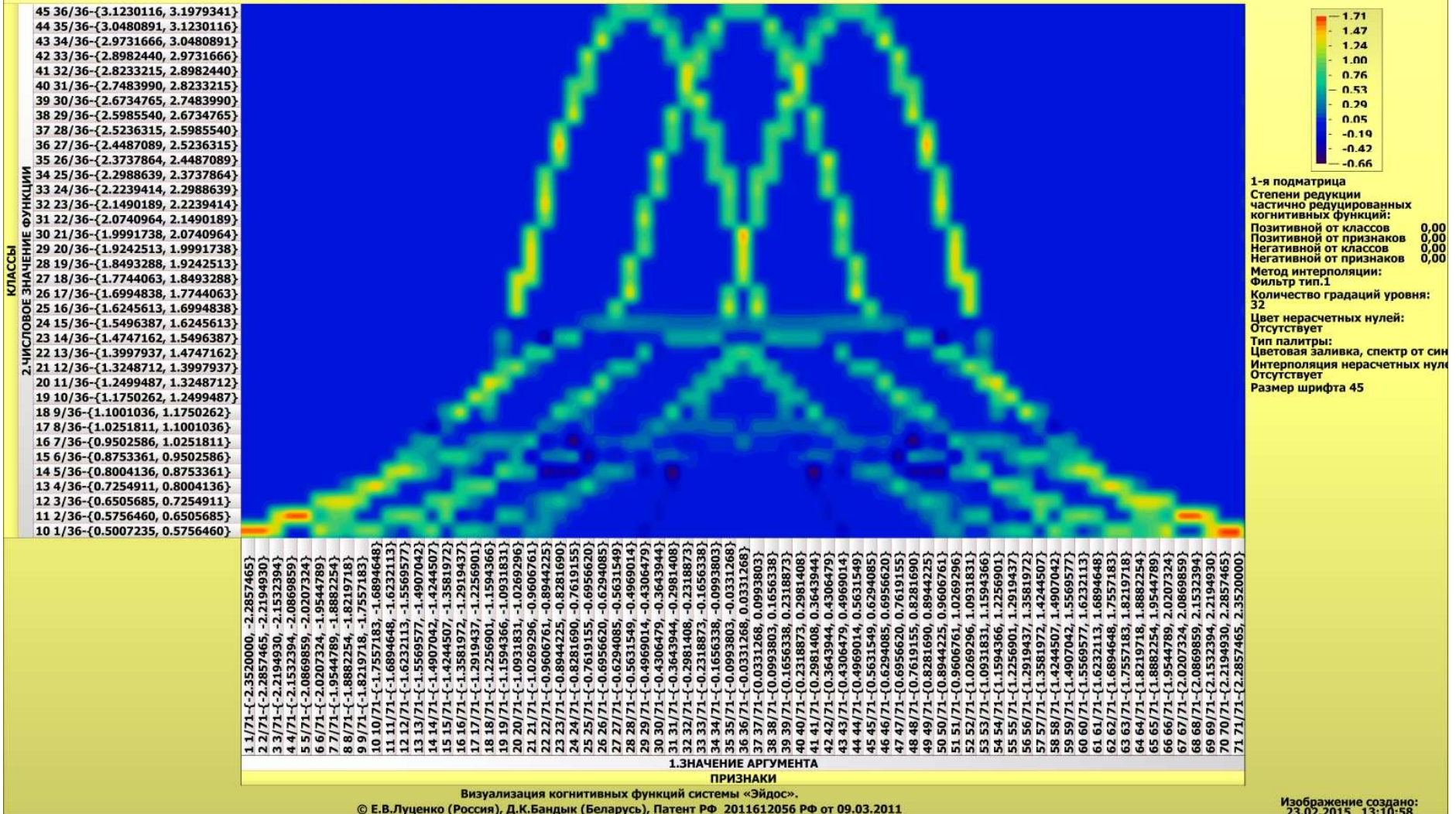


Рисунок 10 – Визуализация статистических распределений в виде когнитивных функций

Из экранной формы, представленной на рисунке 11, мы *видим*, что в различных интервальных значениях аргумента содержится различное количество информации о значениях функции статистического распределения. Таким образом, мы получили ответ на первый вопрос, ответ на который хотели получить в статье: «*Вопрос 1-й. Какое количество информации о принадлежности случайной величины к нескольким нормальным распределениям с разными параметрами мы получаем, когда узнаем, что она попала в некоторый определенный i-й интервал: один из N равных интервалов?*»

Но система «Эйдос» позволяет получить ответ на этот вопрос в самых разнообразных формах, которых очень много, например, в форме инвертированных SWOT-диаграмм (предложены автором в работе [38]). Экранная форма инвертированной SWOT-матрицы для одного из интервальных значений аргумента приведена на рисунке 11, а соответствующая ей инвертированная SWOT-диаграмма – на рисунке 12:

4.4.9 Количественный автоматизированный SWOT-анализ значений факторов средствами АСК-анализа в системе "Эйдос"

Выбор значения фактора, оказывающего влияние на переход объекта управления в будущее состояния

Код	Наименование значения фактора
26	ЗНАЧЕНИЕ АРГУМЕНТА-26/71-{-0.6956620, -0.6294085}
27	ЗНАЧЕНИЕ АРГУМЕНТА-27/71-{-0.6294085, -0.5631549}
28	ЗНАЧЕНИЕ АРГУМЕНТА-28/71-{-0.5631549, -0.4969014}
29	ЗНАЧЕНИЕ АРГУМЕНТА-29/71-{-0.4969014, -0.4306479}
30	ЗНАЧЕНИЕ АРГУМЕНТА-30/71-{-0.4306479, -0.3643944}

SWOT-анализ значения фактора: 27 "ЗНАЧЕНИЕ АРГУМЕНТА-27/71-{-0.6294085, -0.5631549}" в модели: 4...

СПОСОБСТВУЕТ:

Код	Состояния объекта управления, переходу в которые данное значение фактора СПОСОБСТВУЕТ	Сила влияния
1	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1	0.472
2	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2	0.248
4	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-1	0.159

ПРЕПЯТСТВУЕТ:

Код	Состояния объекта управления, переходу в которые данное значение фактора ПРЕПЯТСТВУЕТ	Сила влияния
8	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-2	-0.154
5	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-2	-0.154
9	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-3	-0.144
3	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-3	-0.144
6	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-3	-0.142

ВКЛЮЧИТЬ фильтр по кл.шкале ВЫКЛЮЧИТЬ фильтр по кл.шкале ВКЛЮЧИТЬ фильтр по кл.шкале ВЫКЛЮЧИТЬ фильтр по кл.шкале

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 SWOT-диаграмма

Рисунок 11 – Пример экранной формы инвертированной SWOT-матрицы

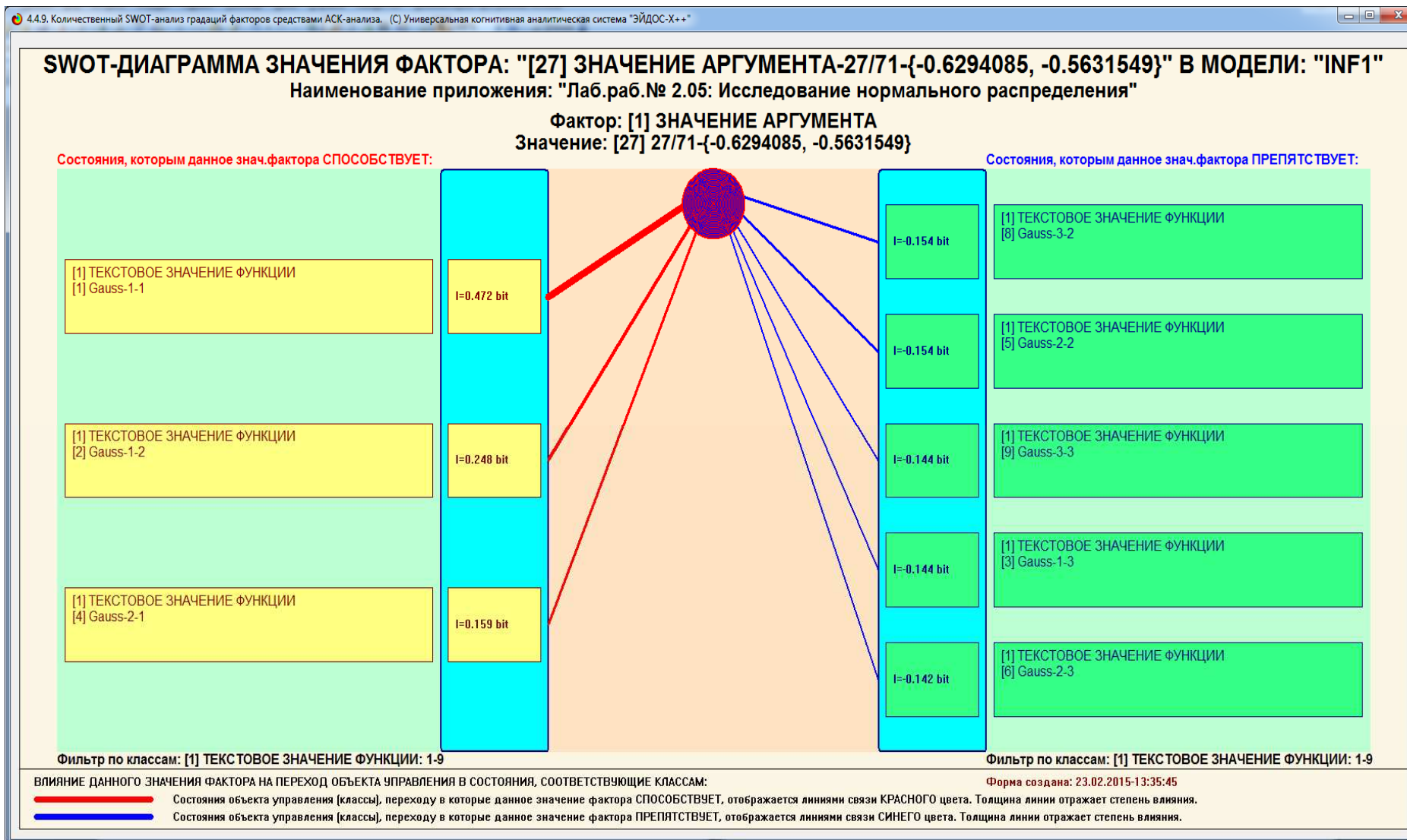


Рисунок 12 – Пример экранной формы инвертированной SWOT-диаграммы

Рассмотрим теперь варианты различных по форме ответов на второй вопрос: «Вопрос 2-й. Какое суммарное количество информации о степени сходства эмпирического распределения наблюдений с нормальными распределениями с различными параметрами мы получаем, зная частоты попадания случайной величины в каждый из интервалов при N равных интервалах?», которые предоставляет система «Эйдос».

На рисунке 13 представлена экранная форма с отображением фрагмента эмпирического распределения частот наблюдений №1, а на рисунке 14, – результаты идентификации этого эмпирического распределения по суммарному количеству информации, содержащемуся в нем.

4.1.1. Ручной ввод-корректировка распознаваемой выборки (режим сидамина). Текущая модель: "INF1"

Код объекта	Наименование объекта	Дата	Время
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9

Код объекта	Класс 1	Класс 2	Класс 3	Класс 4
1	1	10	10	11
1	11	12	12	13
1	13	14	14	15
1	15	16	17	17
1	18	19	20	20
1	21	22	23	23
1	23	24	24	25
1	25	26	26	27
1	27	27	28	28
1	28	29	29	30

Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	14	72	143	14	72	143	15
1	72	143	15	72	143	15	72
1	143	15	72	143	15	72	143
1	16	72	143	16	72	143	16
1	72	143	16	72	143	17	72
1	143	17	72	143	17	72	143
1	17	72	143	18	72	143	18
1	72	143	18	72	143	18	72
1	143	19	72	143	19	72	143
1	19	72	143	19	72	143	19

Рисунок 13 – Экранная форма с отображением эмпирического распределения частот наблюдений (фрагмент)

4.1.3.1. Визуализация результатов распознавания в отношении: "Объект-классы". Текущая модель: "INF1"

Код	Наим объекта
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9

Код	Наименование класса	Сходство	Ф...	Сходство
1	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1	99,99...	v	...
2	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2	49,06...
4	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-1	49,06...
3	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-3	48,16...
7	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-1	48,12...
5	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-2	-1,432...		...
6	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-3	-1,819...		...
9	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-3	-1,885...		...
8	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-2	-1,912...		...

Код	Наименование класса	Сходство	Ф...	Сходство
1	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1	100,00...	v	...
2	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2	49,91...
4	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-1	49,83...
3	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-3	49,22...
7	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-1	49,04...
5	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-2	0,082...
8	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-2	-0,339...
6	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-3	-0,362...
9	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-3	-0,389...

Рисунок 14 – Результаты определения вида и параметров эмпирического статистического распределения №1 по максимальному количеству информации в нем о других распределениях, содержащихся в моделях

Мы видим, что в эмпирическом распределении №1 больше всего информации содержится о его сходстве со стандартным нормальным распределением Gauss-1-1, что и соответствует действительности.

Возникает вопрос: «А чем эта технология отличается от расчета простой корреляции между эмпирическим и теоретическим распределениями?». Есть несколько очень существенных отличий:

1. При расчете корреляции мы используем математический аппарат параметрической статистики, что является корректным только в случае, когда исследуемые данные подчиняются нормальному распределению. В данном же случае использован математический аппарат, не основанный на этом предположении.

2. При расчете корреляции все координаты векторов имеют одинаковый вес, равный 1, тогда как в данном случае вес наблюдений различен, причем он может быть как положительным, так и отрицательным и равен количеству информации в них [14, 15].

3. Расчет корреляции производится для одной группы, тогда как количество информации рассчитывается на основе анализа модели, отражающей условные и безусловные частотные распределения во всех группах и на основе их сравнения. Таким образом, данный метод является обобщением метода контрольных групп (на большое число групп), являющегося общепринятым в науке для выявления влияния факторов.

Обратимся к ответу на «*Вопрос 3-й. Попадание случайной величины в какие интервалы при N равных интервалах является более характерным и в какие менее характерным для нормальных распределений с различными параметрами?*»

На качественном уровне ответ на этот вопрос дает визуализация когнитивной функции на рисунке 10. На количественном – сама база знаний, сформированная в системе, а также другие формы, полученные на ее основе. В таблице 2 приведен фрагмент базы знаний с количеством информации в битах по А.Харкевичу в интервальных значениях аргумента о том, что это значение принадлежит статистическому распределению с определенными параметрами:

Таблица 2 – Количество информации в битах по А.Харкевичу в интервальных значениях аргумента о том, что это значение принадлежит нормальному распределению с определенными параметрами

Аргумент	G-1-1	G-1-2	G-1-3	G-2-1	G-2-2	G-2-3	G-3-1	G-3-2	G-3-3
1/71-{-2.3520000, -2.2857465}			1,28688						
2/71-{-2.2857465, -2.2194930}			1,28688						
3/71-{-2.2194930, -2.1532394}			1,28688						
4/71-{-2.1532394, -2.0869859}			1,28688						
5/71-{-2.0869859, -2.0207324}			1,28688						
6/71-{-2.0207324, -1.9544789}		0,87429	0,88422						
7/71-{-1.9544789, -1.8882254}		0,87429	0,88422						
8/71-{-1.8882254, -1.8219718}		0,76838	0,77831			0,24853			
9/71-{-1.8219718, -1.7557183}		0,63875	0,64868			0,65119			
10/71-{-1.7557183, -1.6894648}		0,63875	0,64868			0,65119			
11/71-{-1.6894648, -1.6232113}		0,63875	0,64868			0,65119			
12/71-{-1.6232113, -1.5569577}		0,63875	0,64868			0,65119			
13/71-{-1.5569577, -1.4907042}		0,54920	0,55913		0,14654	0,56164			
14/71-{-1.4907042, -1.4244507}	-0,01163	0,40321	0,41314		0,40321	0,41565			
15/71-{-1.4244507, -1.3581972}	0,32982	0,34200	0,35193		0,34200	0,35444			
16/71-{-1.3581972, -1.2919437}	0,22391	0,23609	0,24602		0,23609	0,24853			0,24602
17/71-{-1.2919437, -1.2256901}	0,22391	0,23609	0,24602		0,23609	0,24853			0,24602
18/71-{-1.2256901, -1.1594366}	0,22391	0,23609	0,24602		0,23609	0,24853			0,24602
19/71-{-1.1594366, -1.0931831}	0,32982	0,21237	0,22230		0,21237	0,22481			0,22230
20/71-{-1.0931831, -1.0269296}	0,53702	0,14654	0,15647		0,14654	0,15898			0,15647
21/71-{-1.0269296, -0.9606761}	0,45945	0,06897	0,07890		0,06897	0,08141		0,06897	0,07890
22/71-{-0.9606761, -0.8944225}	0,39103	0,00055	0,01048	-0,01163	0,00055	0,01299		0,00055	0,01048
23/71-{-0.8944225, -0.8281690}	0,45945	0,17488	-0,05073	-0,07284	-0,06066	-0,04822		-0,06066	-0,05073

24/71-{-0.8281690, -0.7619155}	0,51000	0,28664	-0,10610	-0,12821	-0,11603	-0,10359		-0,11603	-0,10610
25/71-{-0.7619155, -0.6956620}	0,51000	0,28664	-0,10610	-0,12821	-0,11603	-0,10359		-0,11603	-0,10610
26/71-{-0.6956620, -0.6294085}	0,51000	0,28664	-0,10610	-0,12821	-0,11603	-0,10359		-0,11603	-0,10610
27/71-{-0.6294085, -0.5631549}	0,47168	0,24832	-0,14441	0,15857	-0,15434	-0,14190		-0,15434	-0,14441
28/71-{-0.5631549, -0.4969014}	0,45945	0,23609	-0,15664	0,22391	-0,16657	-0,15413		-0,16657	-0,15664
29/71-{-0.4969014, -0.4306479}	0,44040	0,21704	-0,17569	0,20486	-0,18562	-0,17318	-0,73009	-0,18562	-0,17569
30/71-{-0.4306479, -0.3643944}	0,41295	0,18959	-0,20314	0,17741	-0,21307	-0,20063	-0,22525	-0,21307	-0,20314
31/71-{-0.3643944, -0.2981408}	0,33959	0,11622	-0,27651	0,28904	0,11622	-0,27400	-0,29862	-0,28644	-0,27651
32/71-{-0.2981408, -0.2318873}	0,32982	0,10646	-0,28627	0,32982	0,10646	-0,28376	-0,30838	-0,29620	-0,28627
33/71-{-0.2318873, -0.1656338}	0,32982	0,10646	-0,28627	0,32982	0,10646	-0,28376	-0,30838	-0,29620	-0,28627
34/71-{-0.1656338, -0.0993803}	0,20278	-0,02058	-0,24619	0,36990	0,14654	-0,24368	-0,13867	-0,25612	-0,24619
35/71-{-0.0993803, -0.0331268}	0,13436	-0,25612	-0,24619	0,36990	0,14654	-0,24368	0,13436	-0,25612	-0,24619
36/71-{-0.0331268, 0.0331268}	0,13436	-0,25612	-0,24619	0,36990	0,14654	-0,24368	0,13436	-0,25612	-0,24619
37/71-{0.0331268, 0.0993803}	0,13436	-0,25612	-0,24619	0,36990	0,14654	-0,24368	0,13436	-0,25612	-0,24619
38/71-{0.0993803, 0.1656338}	-0,13867	-0,25612	-0,24619	0,36990	0,14654	-0,24368	0,20278	-0,02058	-0,24619
39/71-{0.1656338, 0.2318873}	-0,30838	-0,29620	-0,28627	0,32982	0,10646	-0,28376	0,32982	0,10646	-0,28627
40/71-{0.2318873, 0.2981408}	-0,30838	-0,29620	-0,28627	0,32982	0,10646	-0,28376	0,32982	0,10646	-0,28627
41/71-{0.2981408, 0.3643944}	-0,29862	-0,28644	-0,27651	0,28904	0,11622	-0,27400	0,33959	0,11622	-0,27651
42/71-{0.3643944, 0.4306479}	-0,22525	-0,21307	-0,20314	0,17741	-0,21307	-0,20063	0,41295	0,18959	-0,20314
43/71-{0.4306479, 0.4969014}	-0,73009	-0,18562	-0,17569	0,20486	-0,18562	-0,17318	0,44040	0,21704	-0,17569
44/71-{0.4969014, 0.5631549}		-0,16657	-0,15664	0,22391	-0,16657	-0,15413	0,45945	0,23609	-0,15664
45/71-{0.5631549, 0.6294085}		-0,15434	-0,14441	0,15857	-0,15434	-0,14190	0,47168	0,24832	-0,14441
46/71-{0.6294085, 0.6956620}		-0,11603	-0,10610	-0,12821	-0,11603	-0,10359	0,51000	0,28664	-0,10610
47/71-{0.6956620, 0.7619155}		-0,11603	-0,10610	-0,12821	-0,11603	-0,10359	0,51000	0,28664	-0,10610
48/71-{0.7619155, 0.8281690}		-0,11603	-0,10610	-0,12821	-0,11603	-0,10359	0,51000	0,28664	-0,10610
49/71-{0.8281690, 0.8944225}		-0,06066	-0,05073	-0,07284	-0,06066	-0,04822	0,45945	0,17488	-0,05073
50/71-{0.8944225, 0.9606761}		0,00055	0,01048	-0,01163	0,00055	0,01299	0,39103	0,00055	0,01048
51/71-{0.9606761, 1.0269296}		0,06897	0,07890		0,06897	0,08141	0,45945	0,06897	0,07890
52/71-{1.0269296, 1.0931831}			0,15647		0,14654	0,15898	0,53702	0,14654	0,15647
53/71-{1.0931831, 1.1594366}			0,22230		0,21237	0,22481	0,32982	0,21237	0,22230
54/71-{1.1594366, 1.2256901}			0,24602		0,23609	0,24853	0,22391	0,23609	0,24602
55/71-{1.2256901, 1.2919437}			0,24602		0,23609	0,24853	0,22391	0,23609	0,24602
56/71-{1.2919437, 1.3581972}			0,24602		0,23609	0,24853	0,22391	0,23609	0,24602
57/71-{1.3581972, 1.4244507}					0,34200	0,35444	0,32982	0,34200	0,35193
58/71-{1.4244507, 1.4907042}					0,40321	0,41565	-0,01163	0,40321	0,41314
59/71-{1.4907042, 1.5569577}					0,14654	0,56164		0,54920	0,55913
60/71-{1.5569577, 1.6232113}						0,65119		0,63875	0,64868
61/71-{1.6232113, 1.6894648}						0,65119		0,63875	0,64868
62/71-{1.6894648, 1.7557183}						0,65119		0,63875	0,64868
63/71-{1.7557183, 1.8219718}						0,65119		0,63875	0,64868
64/71-{1.8219718, 1.8882254}						0,24853		0,76838	0,77831
65/71-{1.8882254, 1.9544789}								0,87429	0,88422
66/71-{1.9544789, 2.0207324}								0,87429	0,88422
67/71-{2.0207324, 2.0869859}									1,28688
68/71-{2.0869859, 2.1532394}									1,28688
69/71-{2.1532394, 2.2194930}									1,28688
70/71-{2.2194930, 2.2857465}									1,28688
71/71-{2.2857465, 2.3520000}									1,28688

Параметры, соответствующие распределениям, приведены в таблице 1. Изображение несимметрично относительно аргумента из-за использования небольшого количества интервальных значений аргумента.

Характеристика любого заданного распределения, полученная на основе этой базы данных, приведена в SWOT-матрице (рисунок 15) и соответствующей диаграмме (рисунок 16):

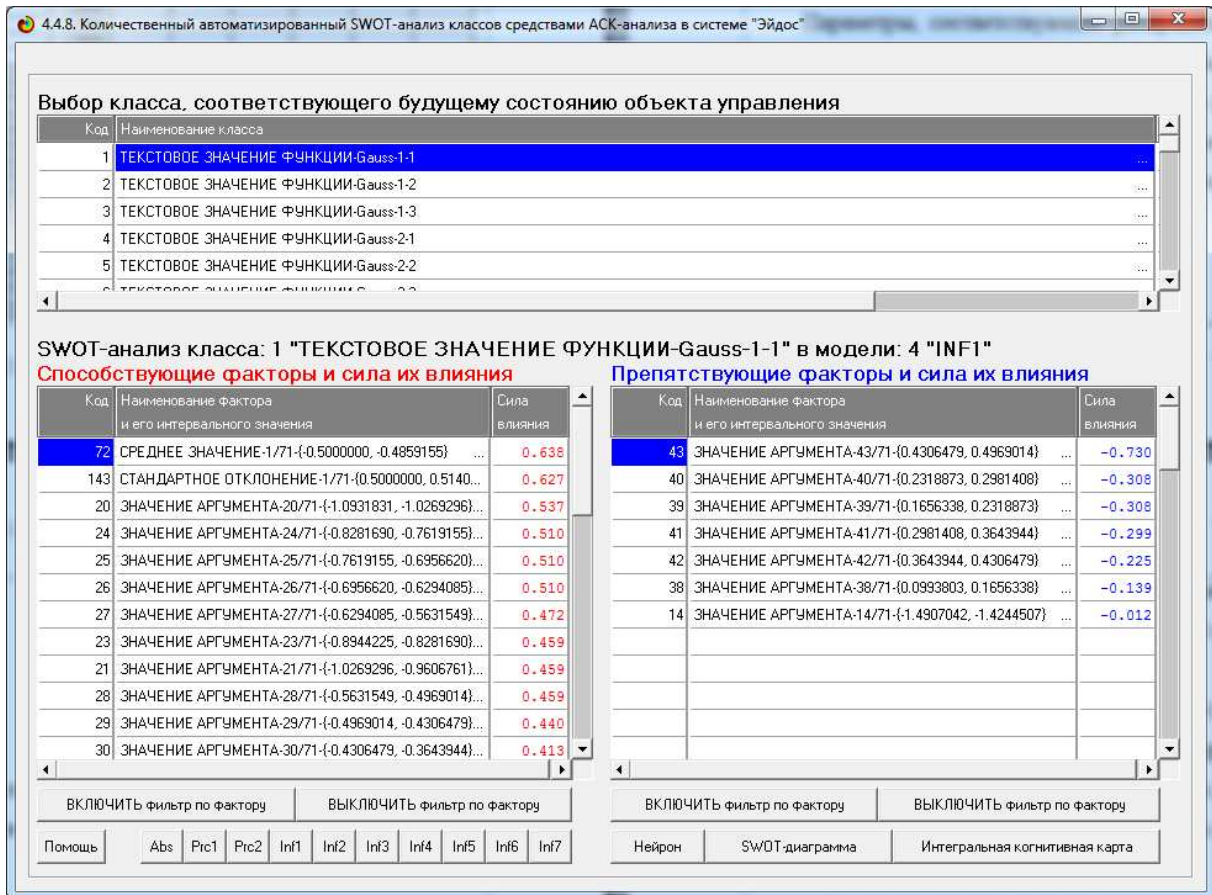


Рисунок 15 – Экранная форма с характеристикой распределения Gauss-1-1 в виде SWOT-матрицы

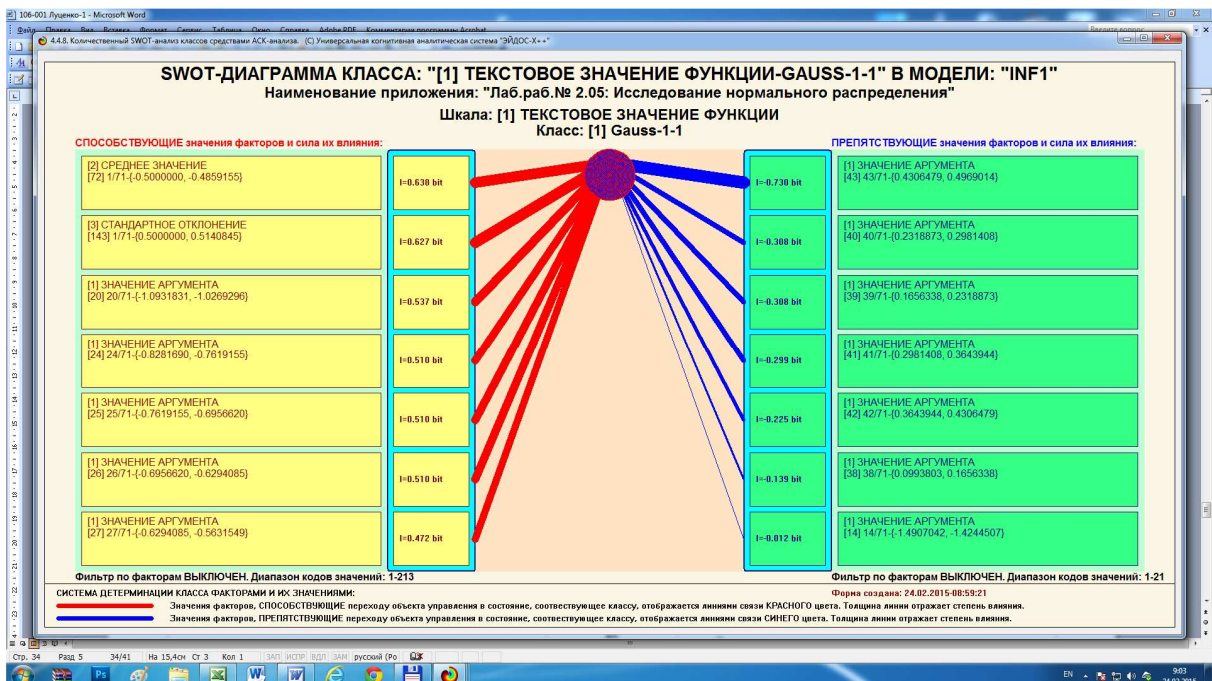


Рисунок 16 – Экранная форма с характеристикой нормального распределения Gauss-1-1 в виде SWOT-диаграммы

Экранная форма задания на выполнение содержательного сравнения двух заданных распределений приведена на рисунке 17, а результат сравнения – в диаграмме на рисунке 18.

Отметим, что эти формы представляют собой результат решения **задачи № 5**: одной из сформулированных в данном разделе задач статистики, которую предположительно можно решить методами теории информации.

4.2.3. Когнитивные диаграммы классов. Задание параметров генерации выходных форм

Выбор классов для когнитивной диаграммы:

Задайте коды двух классов, для левого и правого информационных портретов когнитивной диаграммы по очереди выбирая их курсором в таблице и кликая на соответствующей кнопке ниже нее

Код	Наименование класса
0	ВСЕ КЛАССЫ
1	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1
2	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2
3	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-3
4	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-1
5	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-2

Выбор кода класса левого инф.портрета Выбор кода класса правого инф.портрета

Выбор способа фильтрации признаков в информационных портретах когнитивной диаграммы:

Задайте коды двух описательных шкал, для левого и правого информационных портретов когнитивной диаграммы по очереди выбирая их курсором в таблице и кликая на соответствующей кнопке ниже нее

Код	Наименование описательной шкалы	Минимальный код градации	Максимальный код градации
0	ВСЕ ОПИСАТЕЛЬНЫЕ ШКАЛЫ	1	213
1	ЗНАЧЕНИЕ АРГУМЕНТА	1	71
2	СРЕДНЕЕ ЗНАЧЕНИЕ	72	142
3	СТАНДАРТНОЕ ОТКЛОНЕНИЕ	143	213

Выбор кода описательной шкалы левого инф.портрета Выбор кода описательной шкалы правого инф.портрета

Задайте модели, в которых проводить расчеты когнитивных диаграмм:

Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

Задайте max количество отображаемых связей:

В диалоге заданы следующие параметры расчета когнитивных диаграмм:

Класс для левого инф.портрета: [1] ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1
Класс для правого инф.портрета: [2] ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2
Описательная шкала для левого инф.портрета: [0] ВСЕ ОПИСАТЕЛЬНЫЕ ШКАЛЫ
Описательная шкала для правого инф.портрета: [0] ВСЕ ОПИСАТЕЛЬНЫЕ ШКАЛЫ
Модели, заданные для расчета: Inf1

Рисунок 17 – Экранная форма задания на выполнение содержательного сравнения двух заданных распределений

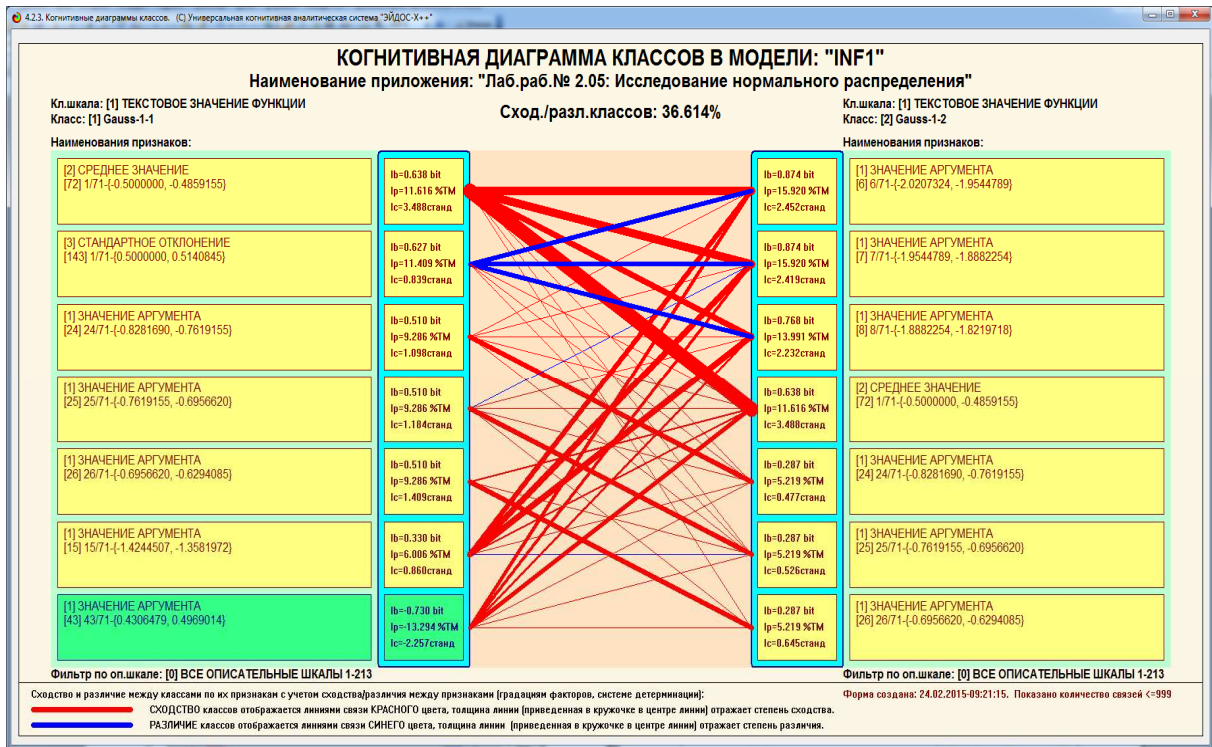


Рисунок 18 – Результат содержательного сравнения двух распределений

На рисунке 19 приведена визуализация таблицы 2:

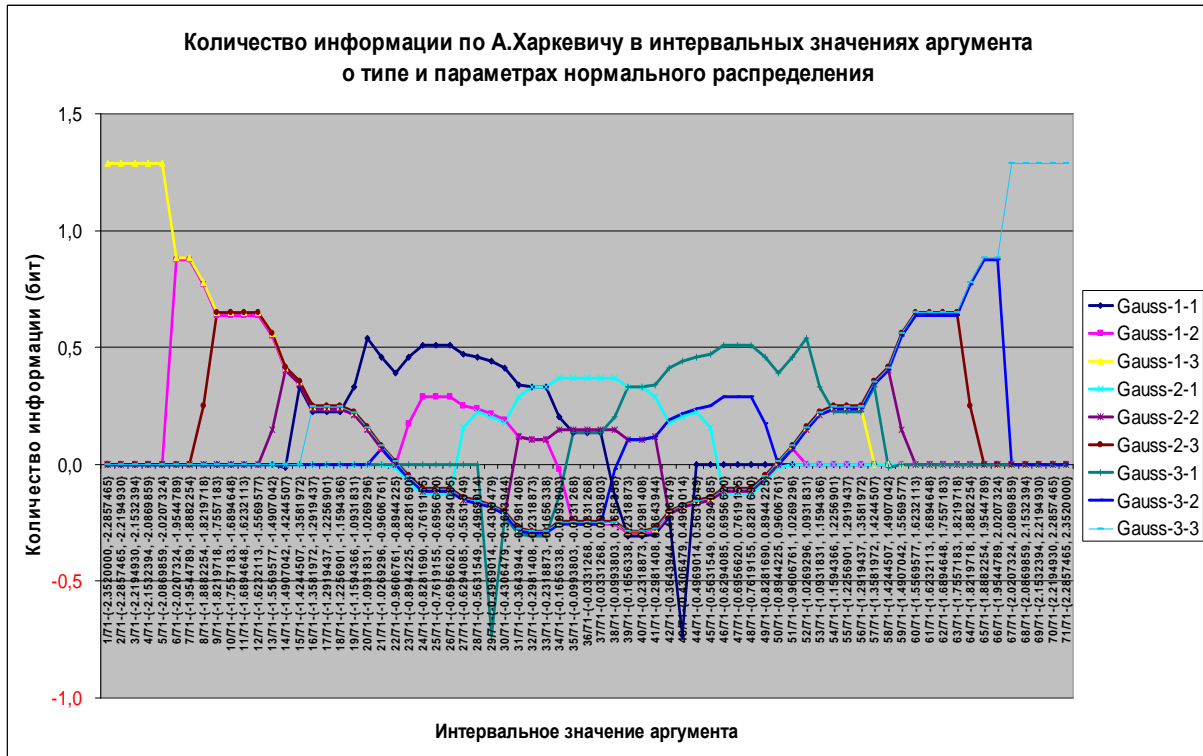


Рисунок 19 – Количество информации по А.Харкевичу в интервальных значениях аргумента о том, что это значение принадлежит нормальному распределению с определенными параметрами (визуализация таблицы 2)

Приведем две выходные формы, полученные в результате решения **Задачи № 6:** «исследование информационных моделей статистических распределений» (решение

этой задачи включает кластерный и конструктивный анализ распределений, их информационные портреты и многие другие исследования с использованием возможностей АСК-анализа и системы «Эйдос» [21]).

На рисунке 20 приведена экранная форма, а на рисунке 21 – соответствующая диаграмма, отражающие результаты кластерно-конструктивного анализа смоделированных нормальных распределений:

Код	Наименование класса	№	Наименование класса	Сходство
1	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1	1	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-1	100.000
2	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2	2	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-2	36.614
3	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-3	27	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-1	9.540
4	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-1	33	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-2	0.914
5	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-2	35	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-1-3	0.373
6	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-3	40	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-2-3	-5.663
7	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-1	42	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-3	-5.915
8	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-2	44	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-1	-27.883
9	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-3	45	ТЕКСТОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-Gauss-3-2	-27.979
10	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-1/36-10.50...			
11	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-2/36-10.57...			
12	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-3/36-10.65...			
13	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-4/36-10.72...			
14	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-5/36-10.80...			
15	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-6/36-10.87...			
16	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-7/36-10.95...			
17	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-8/36-11.02...			
18	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-9/36-11.10...			
19	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-10/36-11.1...			
20	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-11/36-11.2...			
21	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-12/36-11.3...			
22	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-13/36-11.3...			
23	ЧИСЛОВОЕ ЗНАЧЕНИЕ ФУНКЦИИ-14/36-11.4...			

Рисунок 20 – Экранная форма с результатами кластерно-конструктивного анализа смоделированных нормальных распределений в модели с количеством информации по А.Харкевичу (INF1) (установлен фильтр по типу распределения)

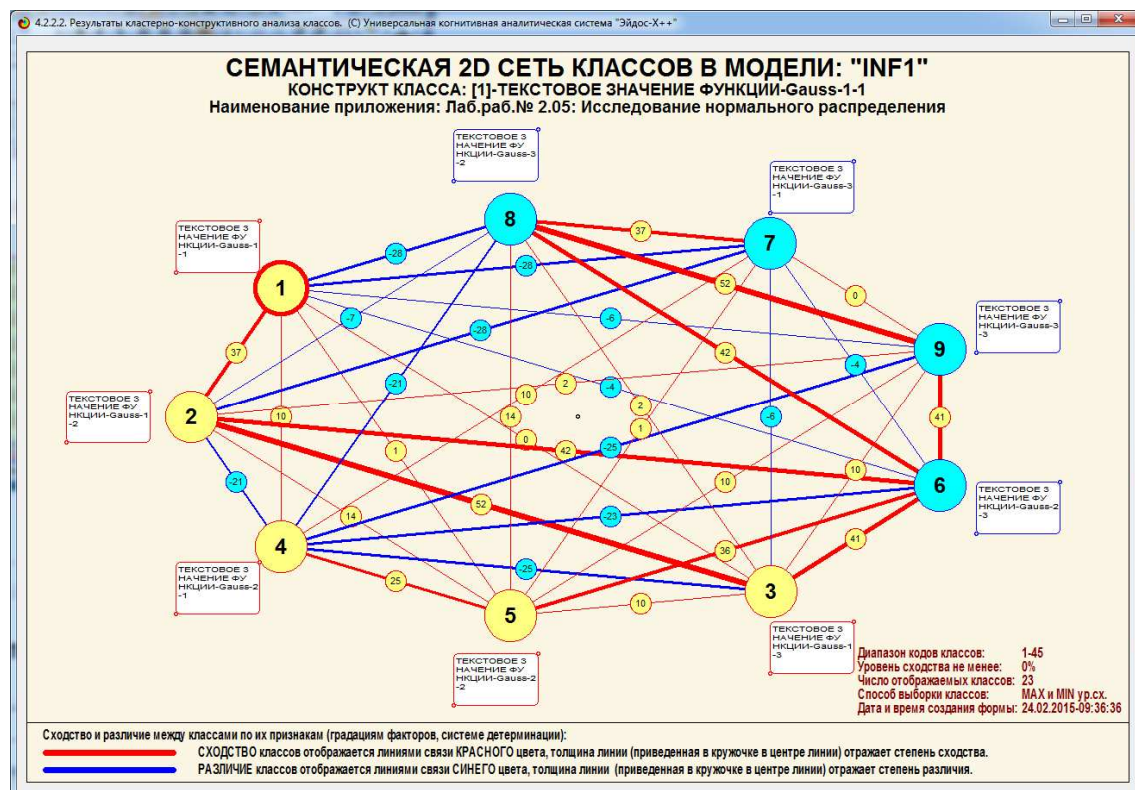


Рисунок 21 – Когнитивная диаграмма с результатами кластерно-конструктивного анализа смоделированных нормальных распределений в модели с количеством информации по А.Харкевичу (INF1) (установлен фильтр по типу распределения)

7.1.9. Выводы

Предлагается теоретическое обоснование, методика численных расчетов и программная реализация решения задач статистики, в частности исследования статистических распределений, методами теории информации. При этом непосредственно на основе эмпирических данных расчетным путем определяется количество информации в наблюдениях, которое используется для анализа статистических распределений. Предлагаемый способ расчета количества информации не основан на предположениях о независимости наблюдений и их нормальном распределении, т.е. является непараметрическим и обеспечивает корректное моделирование нелинейных систем, а также позволяет сопоставимо обрабатывать разнородные (измеряемые в шкалах различных типов) данные числовой и нечисловой природы, измеряемые в различных единицах измерения.

Таким образом, АСК-анализ и система «Эйдос» представляют собой современную инновационную (готовую к внедрению) технологию решения задач статистики методами теории информации.

Данный раздел может быть использована как описание лабораторной работы по дисциплинам:

- Интеллектуальные системы;
- Инженерия знаний и интеллектуальные системы;
- Интеллектуальные технологии и представление знаний;
- Представление знаний в интеллектуальных системах;
- Основы интеллектуальных систем;
- Введение в нейроматематику и методы нейронных сетей;
- Основы искусственного интеллекта;
- Интеллектуальные технологии в науке и образовании;
- Управление знаниями;
- Автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»;

которые автор ведет в настоящее время⁸³, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки).

7.1.10. Ограничения и перспективы

В данном разделе лишь намечены некоторые пути применения теории информации для решения задач статистики. Для реального решения сформулированных выше и других связанных с этим направлением задач необходимы обширные научные исследования и разработки инструментальных средств, что является делом будущего. Планируется описать уже решенные задачи, а также решить некоторые из сформулированных выше задач, в частности использовать статистические критерии в качестве вторичных признаков статистических распределений и исследовать различные статистические распределения.

⁸³ http://lc.kubagro.ru/My_training_schedule.doc

7.2. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в них

7.2.1. Математические аспекты

В данном разделе кратко рассматриваются математическая сущность предложенной автором модификации взвешенного метода наименьших квадратов (ВМНК), в котором в качестве весов наблюдений применяется количество информации в них. Предлагается два варианта данной модификации ВМНК. В первом варианте взвешивание наблюдений производится путем замены одного наблюдения с определенным количеством информации в нем соответствующим количеством наблюдений единичного веса, а затем к ним применяется стандартный метод наименьших квадратов (МНК). Во втором варианте взвешивание наблюдений производится для каждого значения аргумента путем замены всех наблюдений с определенным количеством информации в них одним наблюдением единичного веса, полученным как средневзвешенное от них, а затем к ним применяется стандартный МНК. Подробно описана методика численных расчетов количества информации в наблюдениях, основанная на теории автоматизированного системно-когнитивного анализа (АСК-анализ) и реализованная в его программном инструментарии – интеллектуальной системе «Эйдос». Приводится иллюстрация предлагаемого подхода на простом численном примере. В будущем планируется дать более развернутое математическое обоснование метода взвешенных наименьших квадратов, модифицированного путем применения в качестве весов наблюдений количества информации в них, а также исследовать его свойства

«... навыки мысли и аналитический аппарат теории информации должны, по-видимому, привести к заметной перестройке здания математической статистики»

А.Н. Колмогоров [1, 2, 19]

7.2.1.1. ФОРМУЛИРОВКА ПРОБЛЕМЫ

Данный раздел посвящен математическим аспектам нового варианта взвешенного метода наименьших квадратов (ВМНК), модифицированного путем применения в качестве весов наблюдений количества информации в них. Данный подход предложен автором, в теоретическом плане основан на автоматизированном системно-когнитивном анализе (АСК-анализ) и реализован в его программном инструментарии – системе «Эйдос» [36].

В работе [36] подробно описаны проблемы стандартного (классического) метода наименьших квадратов (МНК), состоящей в том, что *в исходных данных обычно есть такие, которые хуже, чем остальные вписываются в регрессионную модель, т.е. описываются ей с большей погрешностью*. По мнению автора, **причина** этого состоит не только в самих данных, но и в **способе** их отражения в модели.

Иначе говоря, по-видимому, в принципе **возможно** построение разных моделей, отражающих одни и те же эмпирические данные, причем количество этих моделей не ограничено, и в одних моделях эта погрешность будет больше, а в других, более удачных – меньше. Но фактически, т.е. на практике, часто выбор возможных моделей ограничен одной. Поэтому **актуальным** является каждый новый метод построения моделей, который может иметь некоторые преимущества перед уже известными.

Традиционным решением этой проблемы является взвешенный метод наименьших квадратов. В той же работе [36] обосновывается, что подход, реализованный в ВМНК, на самом деле лишь создает видимость решения, а фактически основан просто на игнорировании данных, причем тем в большей степени, чем хуже они вписываются в регрессионную модель.

Рассмотрим *еще* две проблемы, дополнительно к уже описанным в [36], которые обуславливают актуальность предложенной модификации взвешенного метода наименьших квадратов.

Первая проблема ВМНК состоит в том, что на практике ошибки наблюдений являются неизвестными, поэтому их обычно принимают пропорциональными значениям переменных. «Суть взвешенного метода наименьших квадратов состоит в том, что остаткам обобщенной модели регрессии придаются определённые веса, которые равны обратным величинам соответствующих дисперсий $G^2(e_i)$. Однако на практике значения дисперсий являются величинами неизвестными, поэтому для вычисления наиболее подходящих весов используется предположение о том, что они пропорциональны значениям факторных переменных x_i »⁸⁴ (курсив мой, авт.).

Вторая проблема ВМНК состоит в применении *евклидовой меры* расстояния при определении ошибки наблюдений. Но эта мера адекватна только для *ортонормированных пространств*, которые на практике вообще никогда не встречаются, как, кстати, и *линейные системы*. «Если случайные ошибки модели регрессии подвержены гетероскедастичности (но являются неавтокоррелированными), то для оценивания неизвестных коэффициентов модели регрессии применяется взвешенный метод наименьших квадратов»¹.

7.2.1.2. ИДЕЯ ПРЕДЛАГАЕМОГО РЕШЕНИЯ ПРОБЛЕМЫ

В качестве возможного решения поставленной проблемы в работе [36] и предлагается модификация ВМНК, в которой:

- в качестве весов наблюдений используется количество информации в них;
- в качестве меры расстояния применяется суммарное количество информации (т.е. по сути свертка или скалярное произведение), т.е. информационное расстояние, мера расстояния неметрической природы, вообще не предполагающая ортонормированность пространства.

Кроме того очень важно, что АСК-анализе все факторы рассматриваются *с одной единственной точки зрения*: сколько информации содержится в их значениях о переходе объекта, на который они действуют, в определенные будущие состояния, и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: **единицах количества информации** [3]. Именно по этой причине вполне корректно складывать силу и направление влияния всех действующих на объект значений факторов, независимо от их природы, и определять результат *совместного* влияния на объект *системы* значений факторов. При этом в общем случае объект является *нелинейным* и факторы внутри него взаимодействуют друг с другом, т.е. для них не выполняется принцип суперпозиции. *Если же разные факторы измеряются в различных единицах измерения, то результаты сравнения объектов будут зависеть от этих единиц измерения, что совершенно недопустимо с теоретической точки зрения* [3].

Введем определение когнитивной функции: когда функция используется для отображения причинно-следственной зависимости, т.е. информации (согласно концепции Шенка-Абельсона [34]), или *знаний*, если эта информация полезна для достижения

⁸⁴http://www.srines.com/book_1535_chapter_66_66_Dostupnyjjobobshhjonnyj_metod_naimenshikh_kvadratov_Vzv_eshennyj_metod_naimenshikh_kvadratov.html

целей [35], то будем называть такую функцию *когнитивной функцией*, от англ. «*cognition*»⁸⁵ [3].

Смысл когнитивной функциональной зависимости в том, что в значении аргумента содержится определенное количество информации о том, какое значение примет функция, т.е. когнитивная функция отражает знания о степени соответствия значений функции значениям аргумента [3].

Очень важно, что этот подход позволяет автоматически решить проблему сопоставимой обработки многих факторов, измеряемых в различных единицах измерения, т.к. в этом подходе рассматриваются не сами факторы, какой бы природы они не были и какими бы шкалами не формализовались, а количество информации, которое в них содержится о поведении моделируемого объекта [3].

Необходимо также отметить, что представление о полностью линейных объектах (системах) является *абстракцией* и реально все объекты являются принципиально нелинейными. Вместе с тем для большинства систем нелинейные эффекты можно считать эффектами второго и более высоких порядков и такие системы *в первом приближении* можно считать линейными. Возможны различные модели *взаимодействия факторов*, в частности, развиваемые в форме системного обобщения теории множеств. Этот подход в перспективе может стать одним из вариантов развития теории нелинейных систем [3].

Отметим, что математическая модель АСК-анализа (системная теория информации) *органично* учитывает принципиальную нелинейность всех объектов. Это проявляется в нелокальности нейронной сети системы «Эйдос» [46], приводящей к зависимости *всех* информативностей от *любого* изменения в исходных данных, а не как в методе обратного распространения ошибки. В результате *значения матрицы информативностей количественно отражают факторы не как множество, а как систему.*

В АСК-анализе ставится задача метризации шкал, т.е. преобразования к наиболее формализованному виду, и предлагается 7 способов метризации всех типов шкал, обеспечивающих совместную сопоставимую количественную обработку разнородных факторов, измеряемых в различных единицах измерения за счет преобразования всех шкал к одним универсальным единицам измерения в качестве которых выбраны единицы измерения количества информации. Все эти способы метризации реализованы в АСК-анализе и системе «Эйдос» [3]. В работах [4, 5, 6] кратко описаны суть и история появления и развития метода АСК-анализа и его программного инструментария – интеллектуальной системы «Эйдос», поэтому здесь мы их излагать не будем. Отметим лишь, что эти методы созданы довольно давно и уже в 1987 году были акты внедрения интеллектуальных приложений, в которых формировались информационные портреты классов и значений факторов [7]⁸⁶.

Поэтому для нас является вполне естественным предположить, что в качестве весов наблюдений целесообразно использовать количество информации, которое содержится в этих наблюдениях о том, что интересующие нас выходные параметры объекта моделирования примут те или иные значения или сам объект моделирования перейдет в состояния, соответствующие тем или иным классам или окажется принадлежащим к определенным обобщающим категориям (группам). В этом и состоит основная идея предлагаемого решения поставленной проблемы.

В АСК-анализе на основе **системной теории информации** [7, 17] развит математический аппарат, обеспечивающий формальное описание поведения сложных нелинейных объектов моделирования под воздействием систем управляющих факторов и

⁸⁵ <http://lingvo.yandex.ru/cognition/c%20английского/>

⁸⁶ <http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>

окружающей среды, а также созданы инструментальные средства, реализующие этот математический аппарат.

В частности в АСК-анализе предложено понятие **когнитивных функций**, которое рассмотрено и развито в ряде работ автора и соавторов [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] и поэтому здесь нет смысла подробно останавливаться на этом понятии. Отметим лишь суть. В работе [16] кратко рассматриваются классическое понятие функциональной зависимости в математике, определяются ограничения применимости этого понятия для адекватного моделирования реальности и формулируется проблема, состоящая в поиске такого обобщения понятия функции, которое было бы более пригодно для адекватного отражения причинно-следственных связей в реальной области. Далее рассматривается теоретическое и практическое решения поставленной проблемы, состоящие в том, что:

а) предлагается универсальный не зависящий от предметной области способ вычисления количества информации в значении аргумента о значении функции, т.е. когнитивные функции;

б) предлагается программный инструментарий: интеллектуальная система «Эйдос», позволяющая на практике осуществлять эти расчеты, т.е. строить когнитивные функции на основе фрагментированных зашумленных эмпирических данных большой размерности.

Предлагаются понятия нередуцированных, частично и полностью редуцированных прямых и обратных, позитивных и негативных когнитивных функций и метод формирования редуцированных когнитивных функций, являющийся вариантом известного взвешенного метода наименьших квадратов, отличающимся от стандартного ВМНК *учетом в качестве весов наблюдений количества информации в значениях аргумента о значениях функции*.

Конечно, применение теории информации для решения проблем и развития статистики не является абсолютно новой идеей⁸⁷. Как указывает в своих работах [1, 2] профессор А.И.Орлов, сходные идеи развивал еще в середине XX века С.Кульбак [19], а в эпиграф данного раздела вынесено программное высказывание выдающегося русского математика А.Н. Колмогорова: «... навыки мысли и аналитический аппарат теории информации должны, по-видимому, привести к заметной перестройке здания математической статистики», которые содержатся в его предисловии к той же книге С.Кульбака и также приведенное в работах [1, 2]. В наше время в этом направлении продуктивно работают Дуглас Хаббард [20], а также российский математик В.Б.Вяткин [21-28]⁸⁸.

Кроме того, иногда авторы, излагающие в частности взвешенный метод наименьших квадратов, может быть не вполне осознанно используют слово «информация» не как научный термин, а в обиходном разговорном смысле. Например, в работе, приведенной на сайте: http://lib.alnam.ru/book_prs2.php?id=38, автор пишет: «Чтобы учесть *разницу в информации, которую несет каждое наблюдение*, для нахождения оценки необходимо минимизировать взвешенную сумму квадратов отклонений» (отмечено мной, авт.). Казалось бы, остается «лишь» посчитать это количество информации и вариант взвешенного метода наименьших квадратов, основанный на теории информации, готов, но, однако мы видим, что ниже идет изложение стандартного ВМНК.

⁸⁷ Наверное абсолютно новых идей вообще не существует.

⁸⁸ Может быть синергетическая теория информации, созданная трудами В.Б.Вяткина, также может быть использована для расчета количества информации в наблюдениях, что приведет к еще одному варианту взвешенного метода наименьших квадратов.

В работе [37] автор пишет: «...по схеме скользящей средней оценкой текущего уровня является взвешенное среднее всех предшествующих уровней, причем веса при наблюдениях убывают по мере удаления от последнего (текущего) уровня, т. е. **информационная ценность наблюдений тем больше, чем ближе они к концу периода наблюдений**» (отмечено мной, авт.). Здесь мы тоже видим пример применения слова «**информация**» и сочетания «**информационная ценность наблюдений**» в каком-то бытовом смысле, а не в качестве научных терминов. Этот вывод можно сделать на основе подхода, примененного для их расчета или оценки. Казалось бы, нужные слова произнесены и даже написаны и опубликованы, и остается «**только**» а) *прочитать* их, б) *понять*, что буквально сказано и в) *сделать* это. Однако почему-то это никому не приходит в голову, т.е. никто не собирается действительно взять да и посчитать это количество информации. Ведь ясно, что эти подходы, описанные в приведенных выше статьях, не основаны на теории информации. Примерно также на бытовом уровне все понимают, что когда мы спрашиваем о том, какая температура на улице и нам отвечают, то этим самым сообщают нам определенное количество информации. Но никому не приходит в голову посчитать, какое именно количество информации нам сообщают в этом случае, как и в других случаях.

Таким образом, даже если принять в принципе изложенные выше идеи о применении количества информации в наблюдении в качестве веса наблюдения во взвешенном методе наименьших квадратов, то все равно остается очень существенный и принципиальный вопрос о том, **каким способом возможно реально посчитать это количество информации**. Этот вопрос разбивается на две части:

– с помощью какого математического аппарата возможно посчитать количество информации в наблюдении?

– с помощью какого программного инструментария, реализующего этот математический аппарат, возможно реально посчитать количество информации в наблюдении?

Основная идея решения проблемы и предложение автора состоит в том, что для этой цели вполне подходят Автоматизированный системно-когнитивный анализ (АСК-анализ), его математическая модель (системная теория информации), а также реализующий их программный инструментарий АСК-анализа – система «Эйдос». АСК-анализ и система «Эйдос» представляют собой современную **интеллектуальную инновационную** (полностью готовую к внедрению) технологию взвешенного метода наименьших квадратов, модифицированного путем применения в качестве весов наблюдений количества информации в них.

У интеллектуальных технологий есть одно слабое место: их никто не понимает, по крайней мере, почти никто из тех, для кого они предназначены и кому они объективно необходимы. А это значит, что для того, чтобы довести их до практики необходимо придать им такую форму, в которой их и не надо понимать, а можно сразу применять. Это пытается сделать автор в своих разработках, ведущихся в течение многих лет [5], т.е. пытается создать универсальную инновационную (готовую к внедрению) интеллектуальную технологию персонального уровня, т.е. не требующую от пользователя специальной квалификации в области технологий искусственного интеллекта. Результатом этих усилий и являются АСК-анализ и система «Эйдос».

При принятии решений о применении для решения поставленной проблемы этой интеллектуальной инновационной технологии естественно возникает вопрос о степени точности восстановления в создаваемых с помощью нее моделях исследуемых эмпирических зависимостей в АСК-анализе и системе «Эйдос».

Традиционно точность восстановления зависимости оценивается дисперсиями и доверительными интервалами. В АСК-анализе **смысловым** аналогом доверительного интервала, в определенной степени, конечно, является количество информации в аргу-

менте о значении функции. Поэтому *необходимо* исследовать соотношение смыслового содержания этих понятий: доверительного интервала и количества информации.

На математическом уровне это планируется сделать в будущем, а в данном разделе отметим лишь, что *чем больше доверительный интервал, тем выше неопределенность наших знаний о значении функции, соответствующем значению аргумента, а чем он меньше, тем эта определенность выше*. Но информация и определяется как количественная мера степени снятия неопределенности. Учитывая это можно утверждать, что *чем больше доверительный интервал, тем меньше информации о значении функции, соответствующем значению аргумента мы получаем, а чем он меньше, тем это количество информации больше*. Забегая вперед, отметим, что в *частично-редуцированных* когнитивных функциях количество информации в значениях аргумента о значениях функции наглядно изображено шириной полосы функции, что не только по смыслу, но внешне очень сходно с доверительным интервалом. При этом отметим еще один интересный момент, который состоит в том, что если традиционный доверительный интервал при экстраполяции при удалении от эмпирических значений ко все далее отстоящим от них в будущем все время увеличивается, то в степень редукции когнитивной функции то увеличивается, то уменьшается. Это связано с тем, что *АСК-анализ и система «Эйдос» позволяют не только прогнозировать будущие события, но и прогнозировать достоверность или риски этих прогнозов* [7]⁸⁹, т.е. прогнозировать продолжительность периодов эргодичности и точки бифуркации (качественного изменения закономерностей в моделируемой предметной области), что наглядно и отображается в такой форме.

В частности при этом при нулевом доверительном интервале **формально** получается, что мы имеем бесконечное количество информации о значении функции, но **на практике** это вообще невозможно [17] и даже в теории возможно только для отдельных точек **целых** значений аргумента и функции. При бесконечном доверительном интервале в значении аргумента функции содержится ноль информации о значении функции.

В переписке по содержанию статьи профессор А.И.Орлов пишет: «Погрешность средства измерения в ряде случаев меняется с изменением значения измеряемой величины. Если закон изменения характеристик погрешностей известен (например, внесен в паспорт средства измерения), то он дает обоснованные веса. Из подобных соображений вытекает предложение Копяева изменить минимизируемый функционал – вместо суммы квадратов абсолютных расхождений минимизировать сумму квадратов относительных отклонений [40]» (курсив мой. авт.).

Это очень глубокое замечание, из которого вытекают интересные выводы, некоторые из которых мы кратко рассмотрим ниже.

В статье [41] предлагается применить автоматизированный системно-когнитивный анализ как для синтеза адаптивной интеллектуальной измерительной системы, так и для ее использования не с целью измерения параметров объектов, а для идентификации состояний измеряемых систем, т.е. для так называемой системной идентификации. При этом задача измерения рассматривается как предельно упрощенный вариант задачи идентификации или распознавания образов, а задача синтеза измерительной системы – как предельно упрощенный вариант синтеза системы распознавания образов. Программный инструментарий автоматизированного системно-когнитивного анализа – *интеллектуальную систему «Эйдос» предлагается применить как универсальное средство для синтеза и эксплуатации адаптивных интеллектуаль-*

⁸⁹ Подробнее об этом см., например, раздел: <http://lc.kubagro.ru/aidos/aidos02/7.4.htm>.

ных измерительных систем в различных предметных областях. Эта система позволяет вычислять количество информации, содержащейся в результатах измерения, о том, что измеряемая величина примет то или иное значение или объект системной идентификации находится в том или ином состоянии. Применение данного подхода является корректным для измерения состояния сложных многофакторных нелинейных динамических систем.

Упрощенно говоря, система «Эйдос» является интеллектуальной измерительной системой и может рассматриваться в этом качестве. При этом для нее *«закон изменения характеристик погрешностей известен»*, так как в ней роль погрешностей выполняет количество информации, а количество информации тесно связано с понятиями неопределенности и погрешности. Общепринятым является представление об информации, как количественной мере степени снятия неопределенности. Погрешность также является мерой неопределенности наших знаний об истинном значении измеряемой величины. Чем больше погрешность измерения, тем меньше информации мы получаем в процессе измерения о значении измеряемой величины, чем меньше погрешность – тем больше информации в наблюдении (измерении). Поэтому подход, реализованный в предлагаемом варианте ВМНК, находится в *согласии* с предложением работы [40]. Подобные аргументы создают теоретическое обоснование корректности использования количества информации в наблюдениях в качестве их *«обоснованных весов»* в предлагаемом варианте взвешенного метода наименьших квадратов.

7.2.1.3. МАТЕМАТИЧЕСКАЯ СУЩНОСТЬ ПРЕДЛАГАЕМОГО РЕШЕНИЯ ПРОБЛЕМЫ

В описании математического аппарата стандартного метода наименьших квадратов (МНК) в данном разделе нет никакой необходимости, т.к. этому посвящено большое количество общедоступных работ⁹⁰.

Поэтому в данном разделе мы рассмотрим только ключевые моменты, позволяющие *так преобразовать исходные данные о наблюдениях, чтобы они учитывали количество информации в них, рассчитанное по методике численных расчетов АСК-анализа, и чтобы к ним было возможно применить стандартный МНК и при этом учитывалось количество информации в наблюдениях.*

В работе [9] предлагается два варианта данной модификации взвешенного метода наименьших квадратов.

В первом варианте взвешивание наблюдений производится путем замены одного наблюдения с определенным количеством информации в нем соответствующим количеством наблюдений единичного веса, а затем к ним применяется стандартный метод наименьших квадратов (МНК). Фактически в этом варианте решение задачи взвешивания наблюдений решается самим методом наименьших квадратов. Алгоритм и программная реализация данного подхода подробно описаны в статье [36].

В данной же работе, как и планировалось в [36], кратко рассмотрим математические аспекты предлагаемого решения.

В стандартном методе наименьших квадратов минимизируется сумма квадратов отклонений эмпирических значений аппроксимируемой величины y_i от расчетных значений \bar{y}_i , вычисленных в соответствии с моделью (1):

⁹⁰ См., например: [http://yandex.ru/yandsearch?text=математическое%20описание%20метода%20наименьших%20квадратов%20\(МНК\)](http://yandex.ru/yandsearch?text=математическое%20описание%20метода%20наименьших%20квадратов%20(МНК))

$$\sum_i (y_i - \bar{y}_i)^2 \rightarrow \min \quad (1)$$

Во взвешенном методе наименьших квадратов минимизируется сумма квадратов отклонений эмпирических значений аппроксимируемой величины y_i от расчетных значений \bar{y}_i , вычисленных в соответствии с моделью, причем разным наблюдениям приписывается разный вес w_i (2):

$$\sum_i w_i (y_i - \bar{y}_i)^2 \rightarrow \min \quad (2)$$

Ключевым моментом при применении взвешенного МНК является способ выбора и задания весов наблюдений.

Традиционно считается, что разумным вариантом является выбор весов пропорционально ошибкам не взвешенной регрессии [38, 39]. Предполагается, что этим самым более надежным наблюдениям придается больший вес, а сомнительным – меньший. Вроде выглядит разумно. Но проблема в том, что к более надежными и или к сомнительными эмпирические наблюдения относятся путем их сравнения с расчетными значениями, полученными с применением создаваемой модели. Получается, что если модель хорошо описывает эмпирические данные, то они считаются надежными, а если нет, то ненадежными. Как говорится «если факты не соответствуют теории, то тем хуже для фактов». Автор не склонен придерживаться подобной логики и поэтому видит возможность сделать из этого и другой вывод: *если модель хорошо описывает эмпирические данные, то эта модель надежная, а если нет, то ненадежная*, и этот вывод выглядит гораздо более убедительным и разумным.

Подбор этих весов наблюдений вручную может являться сложной и практически неразрешимой задачей, как из-за сложной структуры данных (например, непостоянства дисперсии и среднего ошибок наблюдений), так и из-за возможной очень большой размерности данных. Таким образом, **возникает задача автоматического определения весов наблюдений и разработка алгоритмов и программного инструментария, обеспечивающего автоматизацию определения и взвешивания весов наблюдений в МНК.**

Предлагается новое, ранее не встречавшееся в литературе, решение этой задачи и соответствующее обобщение метода наименьших квадратов (МНК), в котором **точки (наблюдения) имеют вес, равный количеству информации в значении аргумента о значении функции.** Ясно, что по сути, **речь идет о применении когнитивных функций [8-18] в взвешенном МНК.**

$$\sum_i I_i (y_i - \bar{y}_i)^2 \rightarrow \min \quad (3)$$

Здесь I_i – количество информации в i -м наблюдении, т.е. точнее говоря в i -м значении аргумента x_i о том, что i -е функции примет значение $y(x_i)$.

В выражениях (1), (2) и (3) не уточняется, могут ли эмпирические значения функции y_i относиться к одному значению аргумента и это не существенно для МНК. Но если точно известно, что существует M значений аргумента и одному значению аргумента x_i соответствует N_i значений функции, то для дальнейшего изложения нам удобнее записать выражения (1), (2) и (3) в следующей форме, явно учитывающей это обстоятельство:

$$\sum_{i=1}^M \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_i)^2 \rightarrow \min \quad (1')$$

$$\sum_{i=1}^M \sum_{j=1}^{N_i} w_{ij} (y_{ij} - \bar{y}_i)^2 \rightarrow \min \quad (2')$$

$$\sum_{i=1}^M \sum_{j=1}^{N_i} I_{ij} (y_{ij} - \bar{y}_i)^2 \rightarrow \min \quad (3')$$

Отметим, что в случае, когда вес эмпирического наблюдения w_{ij} является целым числом, то выражение (2') эквивалентно выражению:

$$\sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{k=1}^{w_{ij}} (y_{ij} - \bar{y}_i)^2 \rightarrow \min \quad (2'')$$

Этим мы и воспользовались в статье [36], когда заменили одно наблюдение с весом w_{ij} этим количеством наблюдений с единичным весом.

Во втором варианте взвешивание наблюдений производится для каждого значения аргумента путем замены всех наблюдений с определенным количеством информации в них одним наблюдением единичного веса, полученным как средневзвешенное от них, а затем к ним применяется стандартный МНК. В данном варианте ВМНК решение задачи взвешивания наблюдений решается **до** применения стандартного метода наименьших квадратов с помощью другого инструментария, в качестве которого в частности может применяться и интеллектуальная система «Эйдос».

*Перед применением стандартного МНК для каждого значения аргумента **предварительно** рассчитывается средневзвешенное значение функции из всех ее значений с их весами.*

Рассмотрим, как по предлагаемой методике рассчитывается средневзвешенное значение функции с учетом количества информации в аргументе о значении функции для одного значения аргумента.

Для двух точек выбор координаты средневзвешенной точки y соответствует «правилу рычага», т.е. ее положение выбирается таким, чтобы рычаг, образованный двумя точками с координатами y_1 и y_2 и весами I_1 и I_2 , находился в равновесии, если его опора будет в средневзвешенной точке с координатой \bar{y} :

$$(y_2 - y)I_2 = (y - y_1)I_1 \quad (4)$$

Откуда находим y . При двух точках, соответствующих одному значению аргумента, координата y средневзвешенной точки, имеет вид:

$$\bar{y} = \frac{y_1 I_1 + y_2 I_2}{I_1 + I_2} \quad (5)$$

Если же для i -го значения аргумента x_i таких точек N_i , то средневзвешенное значение функции \bar{y}_i выражение (5) принимает вид (6):

$$\bar{y}_i = \frac{\sum_{j=1}^{N_i} y_{ij} I_{ij}}{\sum_{j=1}^{N_i} I_{ij}} \quad (6)$$

В результате средневзвешенная точка находится тем ближе к некоторой точке, чем больше количество информации в значении аргумента о том, что функция примет значение, соответствующее этой точке.

После этого преобразования можно применять стандартный МНК.

В модуле визуализации когнитивных функций [11] этот метод реализован программно по постановке автора разработчиком интеллектуальных систем из Белоруссии Д.К.Бандык и обеспечивает отображение частично и полностью редуцированных когнитивных функций.

7.2.1.4. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДИКА ЧИСЛЕННЫХ РАСЧЕТОВ КОЛИЧЕСТВА ИНФОРМАЦИИ В НАБЛЮДЕНИЯХ

Как говорилось выше, ключевым моментом предлагаемой модификации ВМНК является способ определения количества информации в наблюдениях. Поэтому далее в наиболее упрощенном виде приводится методика численных расчетов количества информации в наблюдениях, основанная на теории автоматизированного системно-когнитивного анализа (АСК-анализ) и реализованная в его программном инструментарии – интеллектуальной системе «Эйдос» [7, 17].

Для удобства рассмотрения введем следующие обозначения:

i – индекс значения аргумента;

j – индекс значения функции;

M – количество значений аргумента;

W – количество значений функции;

N_{ij} – количество встреч j -го значения функции при i -м значении аргумента;

$N_{i\Sigma}$ – суммарное количество наблюдений при i -м значении аргумента по всей выборке;

$N_{\Sigma j}$ – суммарное количество наблюдений j -го значения функции по всей выборке;

$N_{\Sigma\Sigma}$ – суммарное количество наблюдений по всей выборке;

I_{ij} – количество информации в i -м значении аргумента о том, что функция имеет j -е значение, т.е. это количество информации в наблюдении (i, j) ;

Ψ – нормировочный коэффициент (Е.В.Луценко, 1979), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли в равновероятном детерминистском случае;

$P_{i\Sigma}$ – безусловная относительная частота встречи i -го значения аргумента в обучающей выборке;

P_{ij} – условная относительная частота встречи j -го значения функции при i -м значении аргумента.

Используя исходную выборку эмпирических наблюдений посчитаем матрицу абсолютных частот (таблица 1):

Таблица 1 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

		Классы					Сумма
		<i>1</i>	...	<i>j</i>	...	<i>W</i>	
Значения факторов	<i>1</i>	N_{11}		N_{1j}		N_{1W}	
	...						
	<i>i</i>	N_{i1}		N_{ij}		N_{iW}	$N_{i\Sigma} = \sum_{j=1}^W N_{ij}$
	...						
	<i>M</i>	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков				$N_{\Sigma j} = \sum_{i=1}^M N_{ij}$			$N_{\Sigma\Sigma} = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$

Алгоритм формирования матриц абсолютных частот и условных и безусловных процентных распределений.

Объекты обучающей выборки описываются векторами (массивами) $\vec{L} = \{L_i\}$ имеющих у них признаков:

$\vec{L} = \{L_i\} = n$, если у объекта *i*-й признак встречается *n* раз.

Первоначально в матрице абсолютных частот все значения равны нулю. Затем организуется цикл по объектам обучающей выборки. Если у предъявленного объекта, относящегося к *j*-му классу, есть *i*-й признак, то:

$$N_{ij} = N_{ij} + 1; N_{i\Sigma} = N_{i\Sigma} + 1; N_{\Sigma j} = N_{\Sigma j} + 1; N_{\Sigma\Sigma} = N_{\Sigma\Sigma} + 1 \quad (7)$$

На основе анализа матрицы частот (табл. 1) классы можно сравнивать по наблюдаемым частотам признаков только в том случае, если количество объектов по всем классам **одинаково**, как и **суммарное количество признаков по классам**. Если же они отличаются, то корректно сравнивать классы можно только по условным и безусловным относительным частотам (оценкам вероятностей) наблюдений признаков, рассчитанных на основе матрицы частот (табл. 1) в соответствии с выражениями (8), в результате чего получается матрица условных и безусловных процентных распределений (табл. 2):

$$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}; P_i = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}; P_j = \frac{N_{\Sigma j}}{N_{\Sigma\Sigma}};$$

$$N_{i\Sigma} = \sum_{j=1}^W N_{ij}; N_{\Sigma j} = \sum_{i=1}^M N_{ij}; \quad (8)$$

$$N_{\Sigma\Sigma} = \sum_{i=1}^M N_{i\Sigma} = \sum_{j=1}^W N_{\Sigma j} = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$$

Таблица 2 – МАТРИЦА УСЛОВНЫХ И БЕЗУСЛОВНЫХ ПРОЦЕНТНЫХ РАСПРЕДЕЛЕНИЙ

		Классы					Безусловная вероятность признака
		<i>l</i>	...	<i>j</i>	...	<i>w</i>	
Значения факторов	<i>l</i>	P_{ll}		P_{lj}		P_{lw}	
	...						
	<i>i</i>	P_{il}		P_{ij}		P_{iw}	$P_{i\Sigma}$
	...						
	<i>m</i>	P_{ml}		P_{mj}		P_{mw}	
Безусловная вероятность класса				$P_{\Sigma j}$			

Далее произведем расчет количества информации в наблюдениях в соответствии с выражением (9):

$$I_{ij} = \text{Log}_2 \left(\frac{P_{ij}}{P_{i\Sigma}} \right)^\Psi \quad (9)$$

С учетом (8) преобразуем (9) к виду (10):

$$I_{ij} = \Psi \cdot \text{Log}_2 \frac{N_{ij} N_{\Sigma\Sigma}}{N_{i\Sigma} N_{\Sigma j}} \quad (10)$$



А.А.Харкевич

Здесь Ψ – упрощенная форма коэффициента эмерджентности А.Харкевича (10), предложенный автором в 1979 году и названный так в честь известного советского ученого, внесшего большой вклад в теорию информации, на работах которого основана излагаемая методика численных расчетов количества информации в наблюдениях.

$$\Psi = \frac{\text{Log}_2 W}{\text{Log}_2 N_{\Sigma\Sigma}} \quad (11)$$

Используя выражения (9) и (11) на основе таблицы 2 рассчитывается *матрицу информативностей* (таблица 3). Она также может быть получена :непосредственно из таблицы 1 с использованием выражений (10) и (11):

Таблица 3 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ

		Классы					Значимость фактора
		I	...	j	...	W	
Значения факторов	I	I_{11}		I_{1j}		I_{1W}	$\sigma_{1\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	i	I_{i1}		I_{ij}		I_{iW}	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	M	I_{M1}		I_{Mj}		I_{MW}	$\sigma_{M\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Здесь – \bar{I}_i ; это *среднее* количество информации в i -м значении фактора:

$$\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ij}$$

Когда количество информации $I_{ij} > 0$ – i -й фактор способствует переходу объекта управления в j -е состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это. В векторе i -го фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе j -го состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, данная модель позволяет рассчитать, какое количество информации содержится в любом факте о наступлении любого события в любой предметной области, причем для этого не требуется повторности этих фактов и событий. Если данные повторности осуществляются и при этом наблюдается некоторая вариабельность значений факторов, обуславливающих наступление тех или иных событий, то модель обеспечивает многопараметрическую типизацию, т.е. синтез обобщенных образов классов или категорий наступающих событий с количественной оценкой степени и знака влияния на их наступление различных значений факторов. Причем эти значения факторов могут быть как количественными, так и качественными и измеряться в любых единицах измерения, в любом случае в модели оценивается количество информации, которое в них содержится о наступлении событий, переходе объекта управления в определенные состояния или, просто, о его принадлежности к тем или иным классам. Другие способы метризации приведены в работе [3].

Ниже на простом численном примере мы кратко рассмотрим технологию, позволяющую на практике и в любой предметной области посчитать, какое количество информации содержится в наблюдении. В связи с ограничениями на объем статьи автор не име-

ет возможности полностью раскрыть все позиции на приведенных ниже скриншотах и рисунках, т.е. фактически предполагается некоторое предварительное знакомство читателя с системой «Эйдос». Если же такое знакомство недостаточно полное, то автор отсылает автора к публикациям в списке литературы и к сайту: <http://lc.kubagro.ru/>.

7.2.1.5. ЧИСЛЕННЫЙ ПРИМЕР

Для иллюстрации предлагаемых подходов используем тот же численный пример, что и в статье [36], но рассмотрим только второй вариант предлагаемой модификации ВМНК, т.к. первый вариант был подробно рассмотрен в [36].

Запустим режим 4.6 системы «Эйдос», реализующий данный метод, с параметрами, приведенными на рисунке 1:

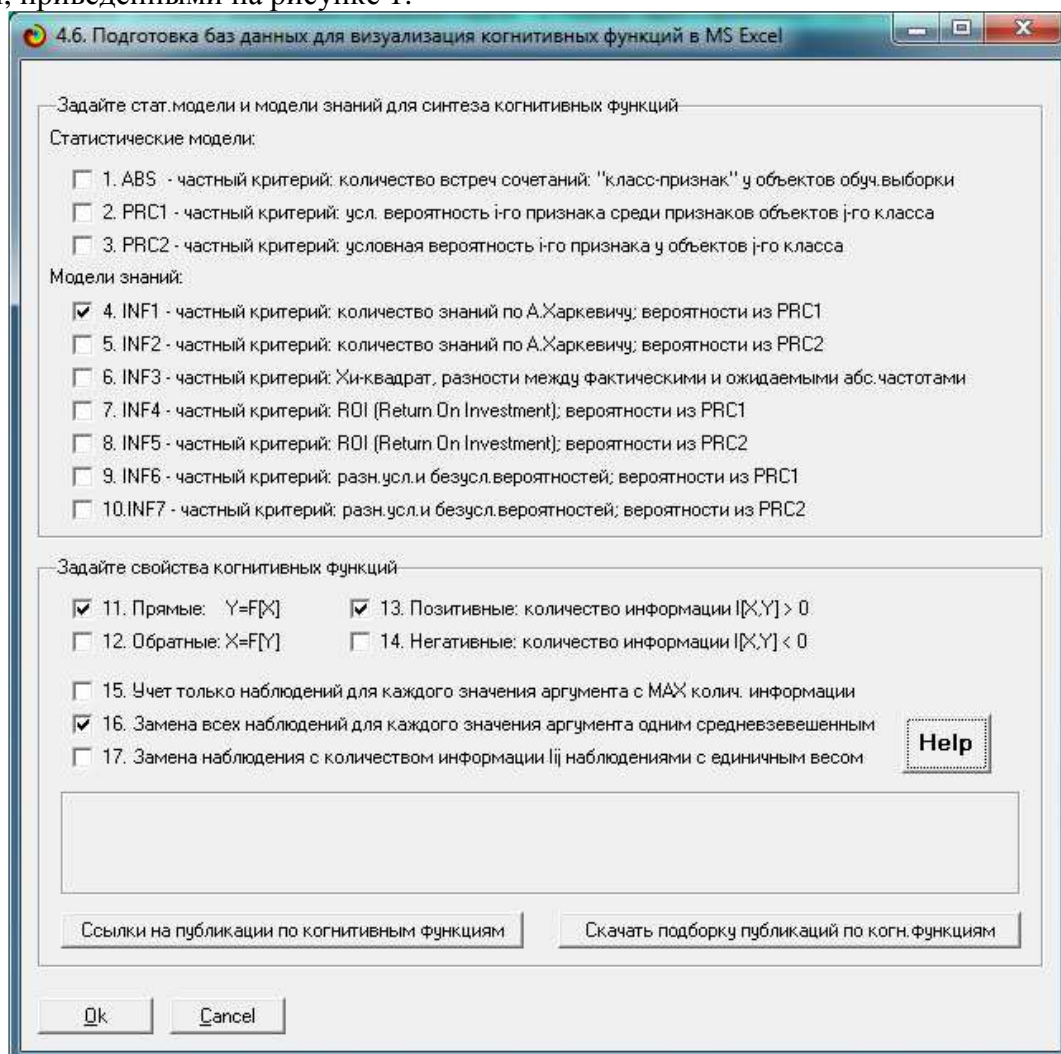


Рисунок 1 – Экранная форма задания параметров режима 4.6 системы «Эйдос»

В результате выполнения режима создаются базы данных, непосредственно считываемые MS Excel и содержащие данные для визуализации когнитивных функций. Виды этих баз данных и способ формирования их имен приведены в таблице 4.

Рассмотрим рисунок 22 из статьи [36] с результатами применения первого варианта предлагаемого метода, приведенный ниже под номером 2:

Таблица 4 – Виды этих баз данных для визуализации когнитивных функций и способ формирования их имен

Прямые и обратные	Позитивные и негативные	Вариант способа учета количества информации в наблюдениях для одного значения аргумента	Имена баз данных для MS Excel
Прямые: $Y=F[X]$	Позитивные: количество информации $I[X,Y] > 0$	Учет только наблюдений для каждого значения аргумента с МАХ колич. информации	####-Y(X)-Pos-One_point-##-##.dbf
		Замена всех наблюдений для каждого значения аргумента одним средневзвешенным	####-Y(X)-Pos-All_points_Avr-##-##.dbf
		Замена наблюдения с количеством информации I_{ij} наблюдениями с единичным весом	####-Y(X)-Pos-All_points_N1-##-##.dbf
	Негативные: количество информации $I[X,Y] < 0$	Учет только наблюдений для каждого значения аргумента с МАХ колич. информации	####-Y(X)-Pos-One_point-##-##.dbf
		Замена всех наблюдений для каждого значения аргумента одним средневзвешенным	####-Y(X)-Pos-All_points_Avr-##-##.dbf
		Замена наблюдения с количеством информации I_{ij} наблюдениями с единичным весом	####-Y(X)-Pos-All_points_N1-##-##.dbf
Обратные: $X=F[Y]$	Позитивные: количество информации $I[X,Y] > 0$	Учет только наблюдений для каждого значения аргумента с МАХ колич. информации	####-Y(X)-Pos-One_point-##-##.dbf
		Замена всех наблюдений для каждого значения аргумента одним средневзвешенным	####-Y(X)-Pos-All_points_Avr-##-##.dbf
		Замена наблюдения с количеством информации I_{ij} наблюдениями с единичным весом	####-Y(X)-Pos-All_points_N1-##-##.dbf
	Негативные: количество информации $I[X,Y] < 0$	Учет только наблюдений для каждого значения аргумента с МАХ колич. информации	####-Y(X)-Pos-One_point-##-##.dbf
		Замена всех наблюдений для каждого значения аргумента одним средневзвешенным	####-Y(X)-Pos-All_points_Avr-##-##.dbf
		Замена наблюдения с количеством информации I_{ij} наблюдениями с единичным весом	####-Y(X)-Pos-All_points_N1-##-##.dbf

Примечание: В начале имени идет обозначение модели, в которой получена когнитивная функция, а непосредственно перед расширением имен баз данных через тире указываются коды описательной и классификационной шкал.

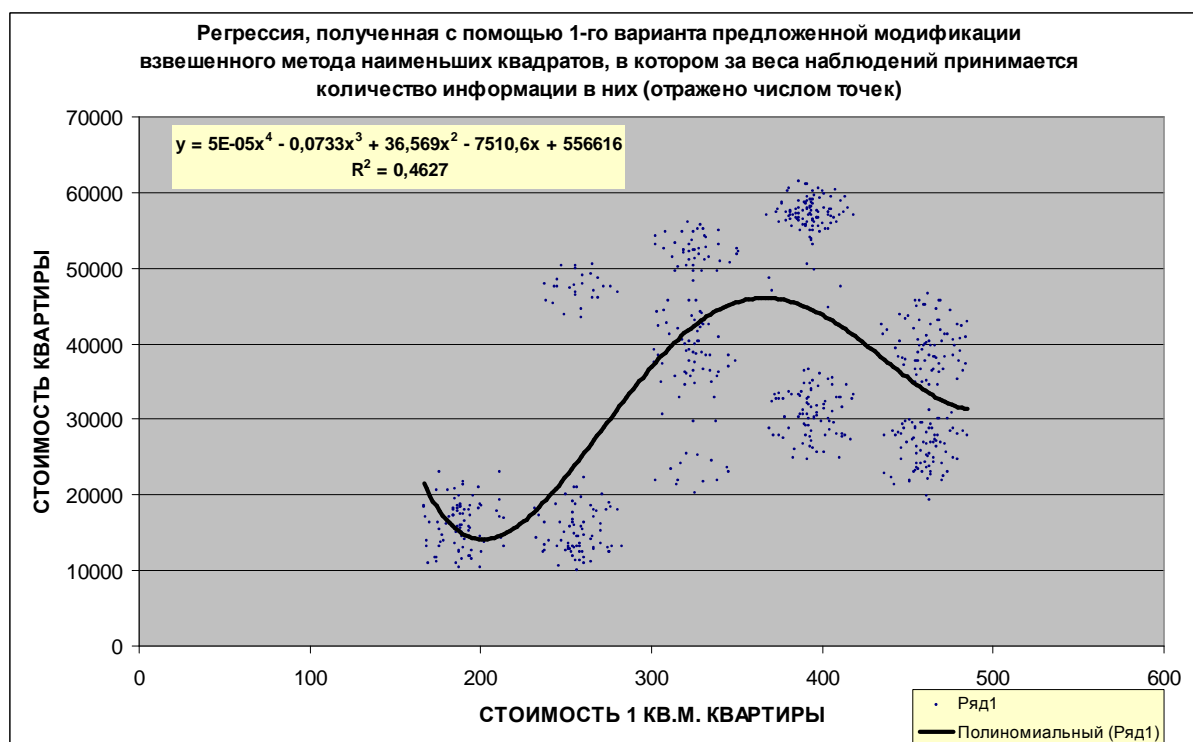


Рисунок 2 – Регрессия, построенная на основе всех наблюдений с учетом количества информации в них с использованием 1-го варианта предлагаемой модификации ВМНК

В таблице 5 приводятся результаты взвешивания наблюдений с учетом количества информации в них с использованием 2-го варианта предлагаемой модификации ВМНК, а на рисунке 3 показаны соответствующие регрессии, построенные по этим данным:

Таблица 5 – Результаты взвешивания наблюдений с учетом количества информации в них с использованием 2-го варианта предлагаемой модификации ВМНК

Наименование аргумента	Наименование значения функции	Значение аргумента	Значение функции
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	188,3629000	16260,8366534
2/5-{222.5048000, 290.7886000}	3/10-{20523.4000000, 25429.4000000}	256,6467000	23509,8510850
3/5-{290.7886000, 359.0724000}	7/10-{40147.4000000, 45053.4000000}	324,9305000	42225,3300638
4/5-{359.0724000, 427.3562000}	8/10-{45053.4000000, 49959.4000000}	393,2143000	45297,9398623
5/5-{427.3562000, 495.6400000}	5/10-{30335.4000000, 35241.4000000}	461,4981000	33211,6434714

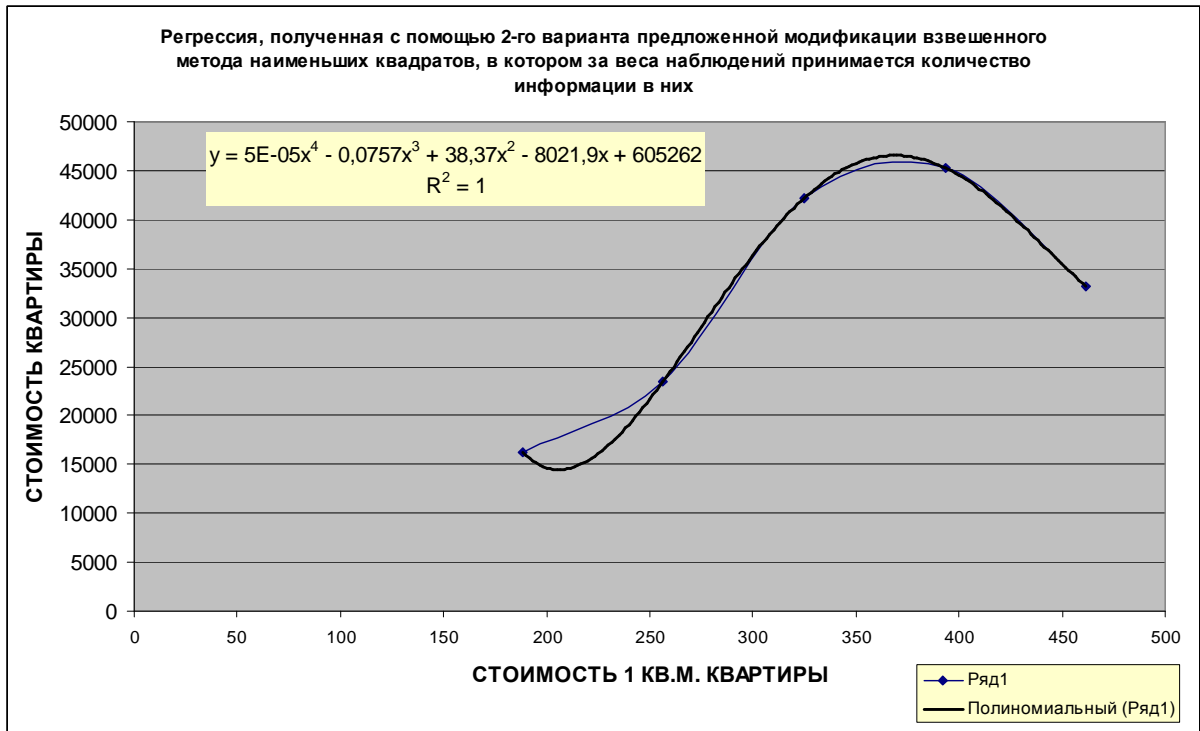


Рисунок 3 – Регрессия, построенная на основе всех наблюдений с учетом количества информации в них с использованием 2-го варианта предлагаемой модификации ВМНК

На рисунке 4 для удобства их сравнения совмещены изображения с рисунков 2 и 3.

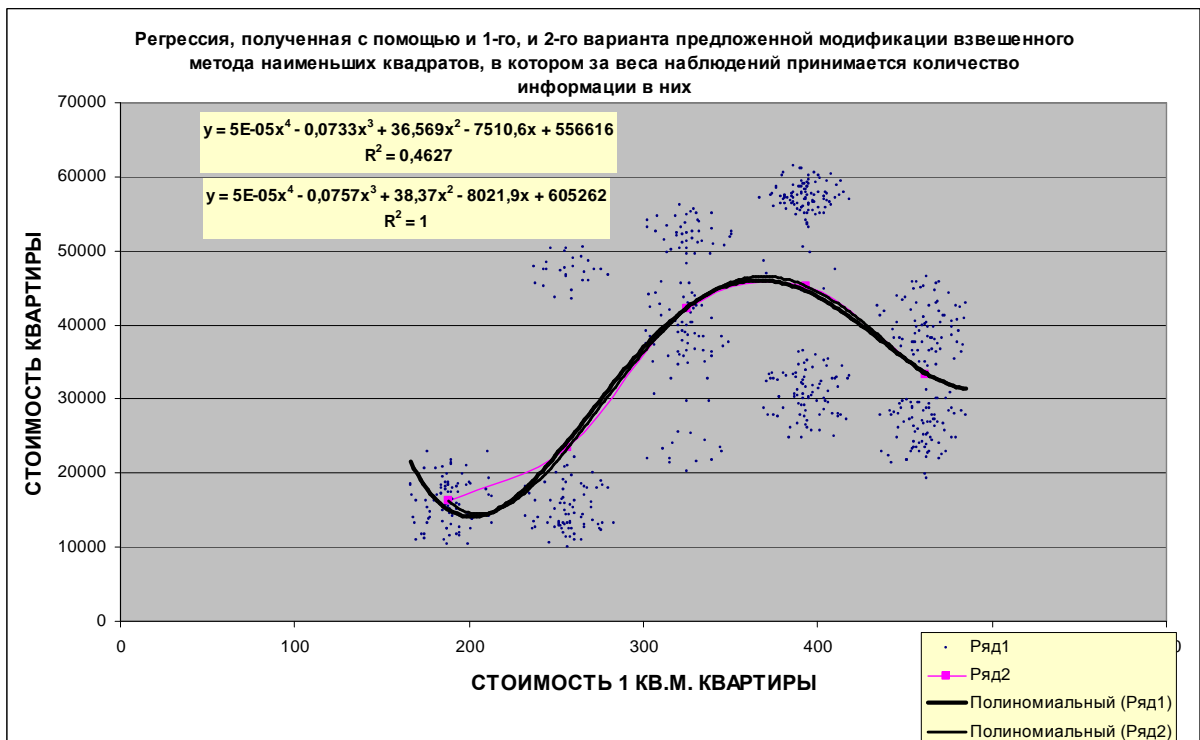


Рисунок 4 – Регрессии, построенные на основе всех наблюдений с учетом количества информации в них с использованием и 1-го, и 2-го вариантов предлагаемой модификации ВМНК

Из сравнения по рисункам 2, 3 и 4 и приведенным на них уравнениям регрессий 1-го и 2-го вариантов взвешивания наблюдений с использованием в качестве весов количества информации в наблюдениях мы можем сделать вывод, что отличаются они весьма незначительно.

7.2.1.6. Выводы

В данном разделе кратко рассмотрена математическая сущность предложенной автором модификации взвешенного метода наименьших квадратов (ВМНК), в котором в качестве весов наблюдений применяется количество информации в них. Предлагается два варианта данной модификации ВМНК. В первом варианте взвешивание наблюдений производится путем замены одного наблюдения с определенным количеством информации в нем соответствующим количеством наблюдений единичного веса, а затем к ним применяется стандартный метод наименьших квадратов (МНК). Во втором варианте взвешивание наблюдений производится для каждого значения аргумента путем замены всех наблюдений с определенным количеством информации в них одним наблюдением единичного веса, полученным как средневзвешенное от них, а затем к ним применяется стандартный МНК. Подробно описана методика численных расчетов количества информации в наблюдениях, основанная на теории автоматизированного системно-когнитивного анализа (АСК-анализ) и реализованная в его программном инструментарии – интеллектуальной системе «Эйдос». Приводится иллюстрация предлагаемого подхода на простом численном примере.

Главный вывод, который можно сделать по материалам статьи, состоит в том, что предлагается обоснованное решение двух дополнительных проблем, сформулированных в начале статьи, т.е. предлагается теоретическое обоснование, методика численных расчетов и программная реализация модификации взвешенного метода наименьших квадратов, в котором в качестве весов наблюдений применяется количество информации в них. Если в ВМНК принимается *гипотеза*, что веса наблюдений тем больше (более надежны), чем меньше ошибка, в качестве которой используется дисперсия, то в предлагаемой модификации ВМНК непосредственно на основе эмпирических данных расчетным путем определяется количество информации в наблюдениях, которое используется в качестве весов наблюдений, вместо традиционной погрешности. Необходимо подчеркнуть, что предлагаемый способ расчета количества информации не основан на предположениях о независимости наблюдений и их нормальном распределении, т.е. является непараметрическим и обеспечивает корректное моделирование нелинейных систем, а также позволяет сопоставимо обрабатывать разнородные (измеряемые в шкалах различных типов) данные числовой и нечисловой природы, измеряемые в различных единицах измерения.

Таким образом, АСК-анализ и система «Эйдос» представляют собой современную инновационную (готовую к внедрению) технологию взвешенного метода наименьших квадратов, модифицированного путем применения в качестве весов наблюдений количества информации в них.

Данный раздел может быть использован как описание лабораторной работы по дисциплинам:

- Интеллектуальные системы;
- Инженерия знаний и интеллектуальные системы;
- Интеллектуальные технологии и представление знаний;
- Представление знаний в интеллектуальных системах;
- Основы интеллектуальных систем;
- Введение в нейроматематику и методы нейронных сетей;
- Основы искусственного интеллекта;

- Интеллектуальные технологии в науке и образовании;
- Управление знаниями;
- Автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»;

которые автор ведет в настоящее время⁹¹, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки).

7.2.1.7. ОГРАНИЧЕНИЯ И ПЕРСПЕКТИВЫ

В данном разделе не ставилась задача исследовать математические и прагматические свойства предлагаемой модификации ВМНК, основанной на использовании в качестве весов наблюдений количества информации в них. Это предполагается сделать в будущих статьях, посвященных данному методу.

Профессор А.И.Орлов в переписке по поводу статьи отмечает, что в будущем «...желательно иметь вероятностно-статистическую теорию, в которой доказаны теоремы о состоятельности оценок параметров зависимости, построены доверительные интервалы для зависимости, как это сделано в классическом случае линейной зависимости в моих книгах (см., например, п.5.1 в "Эконометрике" <http://ibm.bmstu.ru/nil/biblio.html#books-13-econ>). К сожалению, вряд ли такую теорию можно быстро построить».

7.2.2. Алгоритм и программная реализация

Метод наименьших квадратов (МНК) широко известен и пользуется заслуженной популярностью. Вместе с тем не прекращаются попытки усовершенствования этого метода. Результатом одной из таких попыток является взвешенный метод наименьших квадратов (ВМНК), суть которого в том, чтобы придать наблюдениям вес обратно пропорциональный погрешностям их аппроксимации. Этим самым, фактически, наблюдения игнорируются тем в большей степени, чем сложнее их аппроксимировать. В результате такого подхода формально погрешность аппроксимации снижается, но фактически это происходит путем частичного отказа от рассмотрения «проблемных» наблюдений, вносящих большую ошибку. Если эту идею, лежащую в основе ВМНК довести до крайности (и тем самым до абсурда), то в пределе такой подход приведет к тому, что из всей совокупности наблюдений останутся только те, которые практически точно ложатся на тренд, полученный методом наименьших квадратов, а остальные просто будут проигнорированы. Однако, по мнению автора, фактически это не решение проблемы, а отказ от ее решения, хотя внешне и выглядит как решение. В работе предлагается именно решение, основанное на теории информации: считать весом наблюдения количество информации в аргументе о значении функции. Этот подход был обоснован в рамках нового инновационного метода искусственного интеллекта: метода автоматизированного системно-когнитивного анализа (АСК-анализа) и реализован еще 30 лет назад в его программном инструментарии – интеллектуальной системе «Эйдос» в виде так называемых «когнитивных функций». В данном разделе приводится алгоритм и программная реализация данного подхода, проиллюстрированные на подробном численном примере. В будущем планируется дать развернутое

⁹¹ http://lc.kubagro.ru/My_training_schedule.doc

математическое обоснование метода взвешенных наименьших квадратов, модифицированного путем применения теории информации для расчета весовых коэффициентов наблюдений, а также исследовать его свойства

«... навыки мысли и аналитический аппарат теории информации должны, по-видимому, привести к заметной перестройке здания математической статистики»

А.Н. Колмогоров [1, 2, 19]

7.2.2.1. ПРОБЛЕМА ВОССТАНОВЛЕНИЯ АНАЛИТИЧЕСКОЙ ФОРМЫ ФУНКЦИИ ПО ЕЕ ГРАФИКУ ИЛИ ТАБЛИЧНО ЗАДАНЫМ ЗНАЧЕНИЯМ



René Descartes

31.03.1596 – 11.02.1650

После ряда основополагающих работ Рене Декарта стало понятно, что любой функции соответствует график, а любому графику – функция. Построение графика по аналитически заданной функции не представляет собой проблемы, т.к. известен способ, как это сделать, т.е. это задача.

Решается эта задача путем:

- расчета с использованием аналитического выражения для функции таблицы ее значений (таблица 1), соответствующих различным значениям аргумента;

- построения графика параметрически заданной функции (1). Если функциональная зависимость y от x не задана непосредственно $y = f(x)$, а через промежуточную величину – t , то формулы (1)

$$x = \varphi(t); y = \psi(t) \quad (1)$$

задают параметрическое представление функции одной переменной.

Таблица 1 представляет собой таблицу значений функции y и ее аргумента x (1) для различных значений параметра t .

Таблица 1 – Параметрическое задание функции в виде таблицы

	$t = 1$	$t = 2$...	$t = n$
x	$\varphi(t)$	$\varphi(t)$...	$\varphi(t)$
y	$\psi(t)$	$\psi(t)$...	$\psi(t)$

Однако решение **обратной задачи**, т.е. задачи восстановления аналитической формы представления функции, т.е. формулы вида: $y = f(x)$, по ее графику или таблично заданным значениям до сих пор представляет собой проблему, не решенную в общем виде.

Решению этой проблемы посвящен регрессионный анализ [32, 33], в котором широко применяется метод наименьших квадратов (МНК), а также его взвешенный вариант. Однако этот метод позволяет точно восстановить неизвестный истинный вид

функции лишь в редких частных случаях, а в общем виде решает лишь задачу поиска и подбора такого вида функции из заранее определенного набора, которая в определенном смысле или по определенным критериям наилучшим образом совпадает с этой неизвестной истинной функцией.

Одним из общепринятых и действительно наиболее убедительных критериев качества подбора функции, аппроксимирующей эмпирические данные (типа таблицы 1), является минимизация суммы квадратов отклонений эмпирических значений от этой аппроксимирующей их функции.

Однако исследование этих отклонений при аппроксимации различных эмпирических данных показало, что далеко не всегда эти отклонения равномерно зависят от значения функции. Иначе говоря, качество аппроксимации эмпирических данных может изменяться для различных значений аргумента, т.е. качество аппроксимации различно для различных фрагментов функции и эмпирических данных.

Ясно, что качество аппроксимирующей функции не может быть выше качества ее фрагмента, наиболее плохо аппроксимирующего эти эмпирические данные. Вполне понятно и стремление математиков-практиков повысить качество аппроксимации. Но что предлагают в этом плане математики-теоретики?

Если эмпирических данных, выпадающих из закономерности, отражаемой аппроксимирующей функцией, не очень много, то их объявляют «артефактами» и это дает теоретические основания просто игнорировать их путем **удаления** из исследуемой выборки. Ясно, что после этой операции качество аппроксимации заметно улучшается.

Но является ли это решением проблемы? По мнению автора формально является, т.к. вроде как качество модели возрастает, но конечно фактически это не решение, т.к. основано на порочном принципе: «Если факты не вписываются в теорию (в нашем случае аналитическую модель), то тем хуже для фактов». Фактически это «страусиный» способ решения проблем, который состоит просто в том, чтобы не видеть их или делать вид, что их не существует. При этом исследователь часто не отдает себе в этом отчет и впадает в иллюзию (гипостазирование), что он моделирует саму реальность и исследует ее путем исследования созданной им ее модели, тогда как в действительности он исследует только ту часть реальности, которую смог смоделировать при своих ограниченных возможностях моделирования. Профессор А.И.Орлов пишет, что это равносильно тому, чтобы «искать под фонарем, а не там, где потеряли» [1].

Конечно, разработка таких более мощных методов моделирования ведется [2]. Но ознакомление с ними математиков-практиков, и даже руководителей науки, далеко отстает от фактической потребности применения этих методов [2].

Приведем простейший пример, иллюстрирующий высказанные мысли. Если данные не вписываются в линейную модель, то можно игнорировать или удалить из исследуемой те из них, которые вносят основной вклад в суммарную ошибку, а можно использовать квадратичную модель, которая точно описывает эти данные во всей их полноте (таблица 2, рисунки 1, 2 и 3):

Таблица 2 – Исходные данные для примера

Значение аргумента	1	2	3	4	5	6	7
Значение функции	1	4	9	16	25	36	49

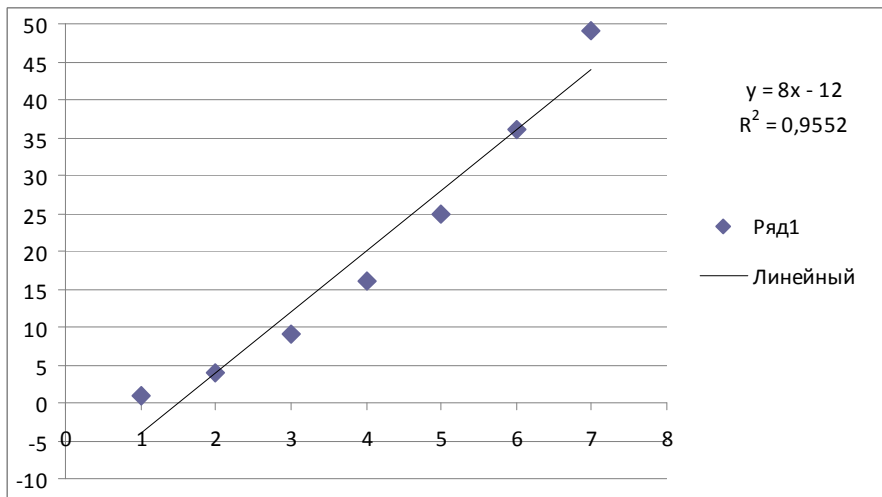


Рисунок 1 – Линейная модель не адекватно отражает исходные данные

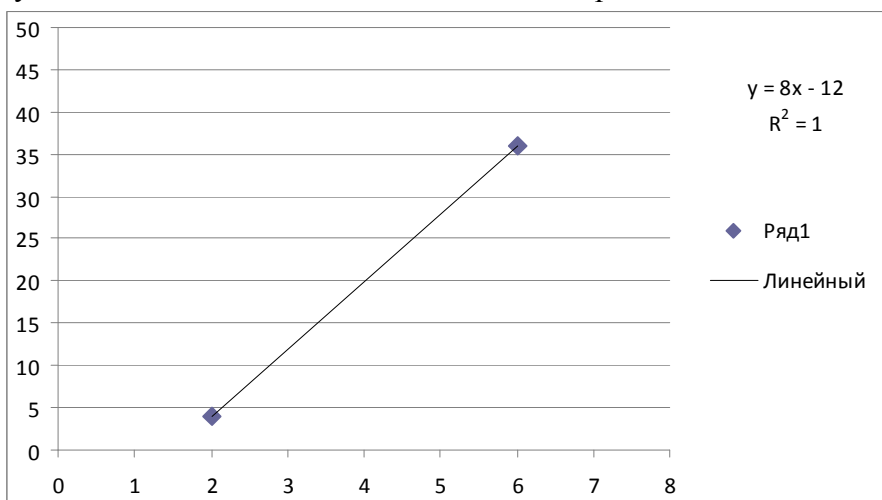


Рисунок 2 – Линейная модель адекватно отражает исходные данные, Из которых удалены все наблюдения, кроме 2-го и 6-го

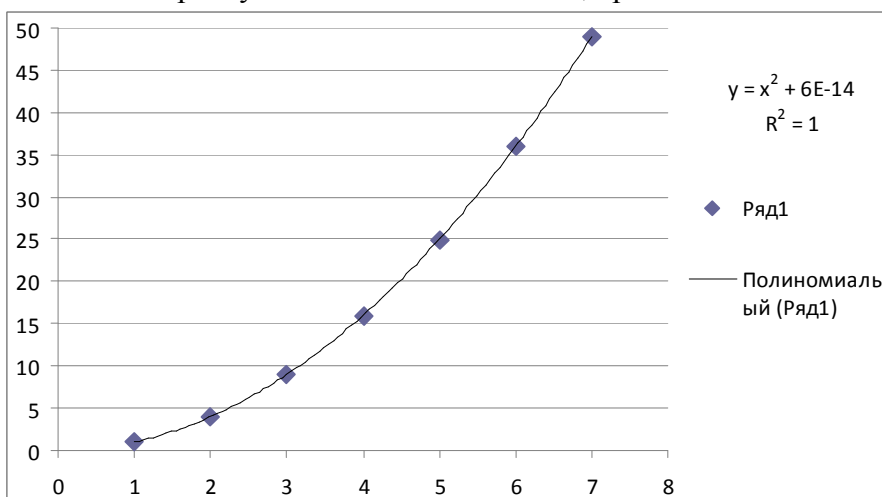


Рисунок 3 – Квадратичная модель адекватно отражает все исходные данные

7.2.2.2. ВЗВЕШЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (ВМНК) КАК ТРАДИЦИОННЫЙ ПУТЬ РЕШЕНИЯ ПРОБЛЕМЫ

Но есть и более развитые идеи и методы улучшения модели по формальным критериям качества: не вообще удалять неудобные данные, а просто уменьшать их значение или вес и делать это тем в большей степени, чем более эти данные неудобны, т.е. с чем большей ошибкой они отражаются в модели. На этой идее основан взвешенный метод наименьших квадратов (ВМНК), который является традиционным путем решения поставленной проблемы. Фактически в этом методе данные сначала преобразуются взвешиванием наблюдений (делением на величину, пропорциональную предполагаемому стандартному отклонению случайных ошибок), а к предварительно взвешенным данным уже применяется обычный стандартный метод наименьших квадратов.

Профессор А.И.Орлов пишет⁹² о том, что:

- на подавление выбросов нацелены робастные методы;
- вероятностно-статистическая модель порождения данных – первична, а метод оценивания параметров качества модели – вторичен;
- точность восстановления зависимости традиционно оценивается дисперсиями и доверительным интервалами;
- если в разные моменты времени проводится различное количество наблюдений, вследствие чего их надежность, погрешности и другие характеристики, вообще говоря, оказываются зависящими от времени⁹³, то *взвешивание данных действительно необходимо*.

Тем ни менее к взвешенному методу наименьших квадратов также может быть адресован ряд критических замечаний, которые мы кратко рассмотрим ниже.

7.2.2.3. НЕДОСТАТКИ ТРАДИЦИОННОГО РЕШЕНИЯ ПРОБЛЕМЫ (ВМНК)

Все те возражения, которые были высказаны выше в отношении процедуры удаления из исходных данных «артефактов» полностью сохраняют силу и для взвешенного метода наименьших квадратов.

Но здесь появляются и дополнительные возражения.

Прежде всего, возникают взаимосвязанные вопросы о цели моделирования и цели повышения качества моделирования.

Если целью моделирования является наиболее полное и адекватное отражение реальности в моделях, а так по наивности обычно все и думают, то повышение качества моделирования должно осуществляться не путем выбора наиболее легко и просто моделируемой предметной области, а путем совершенствования математического аппарата и программного инструментария моделирования.

Но если исходить из этой логики, то в методе взвешенных наименьших квадратов вес наблюдений должен быть принят не обратно пропорциональным вносимым этими наблюдениями ошибкам аппроксимации простым МНК, а **наоборот** пропорциональным этим ошибкам. Проще говоря, чем сложнее некоторые данные отразить в модели, тем более пристальное внимание должно быть им уделено, а не наоборот, как в ВМНК, где фактически от таких данных просто отмахиваются игнорируя их и теоретически обосновывая их якобы «несущественность».

Но в чем **фактически** состоит причина, по которой эти данные вдруг стали считаться несущественными? Да просто в том, что «они портят всю картину», такую

⁹² В переписке по статье

⁹³ а это практически всегда так, но иногда этим можно корректно пренебречь, а иногда нельзя.

стройную и удобную, т.е. ухудшают **формальное** качество модели. Поэтому **если цель (точнее ее называть самоцелью) моделирования состоит не в адекватном отражении реальности, а в повышении формального качества модели**, то от таких данных надо избавиться, но уже не просто удалив их из исследуемой выборки как «артефакты», а более цивилизованным способом, т.е. приписав им меньший вес, в т.ч. вес, равный нулю.

Более того, в статистических пакетах предоставлена возможность задавать веса вручную, позволяет регулировать вклад тех или иных данных в результаты построения моделей. Иначе говоря, предоставляется возможность **вручную практически произвольно по своему усмотрению** влиять на модель путем подбора нужных весовых коэффициентов. Но если так, то может быть проще использовать не статистические пакеты, а просто взять и сразу написать в аналитическом отчете, что «компьютер посчитал так...» и нарисовать в графическом редакторе нужные выходные формы. С аналогичными подходами мы сталкиваемся и при проведении кластерного анализа [30].

7.2.2.4. ПРЕДЛАГАЕМОЕ РЕШЕНИЕ ПРОБЛЕМЫ: МЕТОД ВЗВЕШЕННЫХ НАИМЕНЬШИХ КВАДРАТОВ, МОДИФИЦИРОВАННЫЙ ПУТЕМ ПРИМЕНЕНИЯ ТЕОРИИ ИНФОРМАЦИИ ДЛЯ РАСЧЕТА ВЕСОВЫХ КОЭФФИЦИЕНТОВ НАБЛЮДЕНИЙ

В работах [1, 2] рассматриваются точки роста и перспективы статистических методов, и дается положительная оценка методу автоматизированного системно-когнитивного анализа (АСК-анализ) и его программному инструментарию – интеллектуальной системе «Эйдос».

В АСК-анализе факторы формально описываются шкалами, а значения факторов – градациями шкал [3]. Существует три основных группы факторов: физические, социально-экономические и психологические (субъективные) и в каждой из этих групп есть много различных видов факторов, т.е. есть много различных физических факторов, много социально-экономических и много психологических, но в АСК-анализе все факторы рассматриваются *с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта, на который они действуют, в определенные будущие состояния, и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации*. Именно по этой причине вполне корректно складывать силу и направление влияния всех действующих на объект значений факторов, независимо от их природы, и определять результат *совместного* влияния на объект системы значений факторов. При этом в общем случае объект является *нелинейным* и факторы внутри него взаимодействуют друг с другом, т.е. для них не выполняется принцип суперпозиции.

Если же разные факторы измеряются в различных единицах измерения, то результаты сравнения объектов будут зависеть от этих единиц измерения, что совершенно недопустимо с теоретической точки зрения [3].

Введем определение когнитивной функции: когда функция используется для отображения причинно-следственной зависимости, т.е. информации (согласно концепции Шенка-Абельсона [34]), или *знаний*, если эта информация полезна для достижения целей [35], то будем называть такую функцию *когнитивной функцией*, от англ. «*cognition*»⁹⁴ [3].

Смысл когнитивной функциональной зависимости в том, что в значении аргумента содержится определенное количество информации о том, какое значение примет

⁹⁴ <http://lingvo.yandex.ru/cognition/c%20английского/>

функция, т.е. когнитивная функция отражает знания о степени соответствия значений функции значениям аргумента [3].

Очень важно, что этот подход позволяет автоматически решить проблему сопоставимой обработки многих факторов, измеряемых в различных единицах измерения, т.к. в этом подходе рассматриваются не сами факторы, какой бы природы они не были и какими бы шкалами не формализовались, а количество информации, которое в них содержится о поведении моделируемого объекта [3].

Необходимо также отметить, что представление о полностью линейных объектах (системах) является *абстракцией* и реально все объекты являются принципиально нелинейными. Вместе с тем для большинства систем нелинейные эффекты можно считать эффектами второго и более высоких порядков и такие системы *в первом приближении* можно считать линейными. Возможны различные модели *взаимодействия факторов*, в частности, развиваемые в форме системного обобщения теории множеств. Этот подход в перспективе может стать одним из вариантов развития теории нелинейных систем [3].

Отметим, что математическая модель АСК-анализа (системная теория информации) *органично* учитывает принципиальную нелинейность всех объектов. Это проявляется в нелокальности нейронной сети системы «Эйдос» [46], приводящей к зависимости *всех* информативностей от *любого* изменения в исходных данных, а не как в методе обратного распространения ошибки. В результате *значения матрицы информативностей количественно отражают факторы не как множество, а как систему*.

В АСК-анализе ставится задача метризации шкал, т.е. преобразования к наиболее формализованному виду, и предлагается 7 способов метризации всех типов шкал, обеспечивающих совместную сопоставимую количественную обработку разнородных факторов, измеряемых в различных единицах измерения за счет преобразования всех шкал к одним универсальным единицам измерения в качестве которых выбраны единицы измерения количества информации. Все эти способы метризации реализованы в АСК-анализе и системе «Эйдос» [3]. В работах [4, 5, 6] кратко описаны суть и история появления и развития метода АСК-анализа и его программного инструментария – интеллектуальной системы «Эйдос», поэтому здесь мы их излагать не будем. Отметим лишь, что эти методы созданы довольно давно и уже в 1987 году были акты внедрения интеллектуальных приложений, в которых формировались информационные портреты классов и значений факторов [7]⁹⁵.

Поэтому для нас является вполне естественным предположить, что в качестве весов наблюдений целесообразно использовать количество информации, которое содержится в этих наблюдениях о том, что интересующие нас выходные параметры объекта моделирования примут те или иные значения или сам объект моделирования перейдет в состояния, соответствующие тем или иным классам или окажется принадлежащим к определенным обобщающим категориям (группам). В этом и состоит основная идея предлагаемого решения поставленной проблемы.

В АСК-анализе на основе **системной теории информации** [7, 17] развит математический аппарат, обеспечивающий формальное описание поведения сложных нелинейных объектов моделирования под воздействием систем управляющих факторов и окружающей среды, а также созданы инструментальные средства, реализующие этот математический аппарат.

В частности в АСК-анализе предложено понятие **когнитивных функций**, которое рассмотрено и развито в ряде работ автора и соавторов [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] и поэтому здесь нет смысла подробно останавливаться на этом понятии. От-

⁹⁵ <http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>

метим лишь суть. В работе [16] кратко рассматриваются классическое понятие функциональной зависимости в математике, определяются ограничения применимости этого понятия для адекватного моделирования реальности и формулируется проблема, состоящая в поиске такого обобщения понятия функции, которое было бы более пригодно для адекватного отражения причинно-следственных связей в реальной области. Далее рассматривается теоретическое и практическое решения поставленной проблемы, состоящие в том, что:

а) предлагается универсальный не зависящий от предметной области способ вычисления количества информации в значении аргумента о значении функции, т.е. когнитивные функции;

б) предлагается программный инструментарий: интеллектуальная система «Эйдос», позволяющая на практике осуществлять эти расчеты, т.е. строить когнитивные функции на основе фрагментированных зашумленных эмпирических данных большой размерности.

Предлагаются понятия нередуцированных, частично и полностью редуцированных прямых и обратных, позитивных и негативных когнитивных функций и метод формирования редуцированных когнитивных функций, являющийся вариантом известного взвешенного метода наименьших квадратов, отличающимся от стандартного **ВМНК *учетом в качестве весов наблюдений количества информации в значениях аргумента о значениях функции.***

Конечно, применение теории информации для решения проблем и развития статистики не является абсолютно новой идеей⁹⁶. Как указывает в своих работах [1, 2] профессор А.И.Орлов, сходные идеи развивал еще в середине XX века С.Кульбак [19], а в эпиграф данной статьи вынесено программное высказывание выдающегося российского математика А.Н. Колмогорова: «... навыки мысли и аналитический аппарат теории информации должны, по-видимому, привести к заметной перестройке здания математической статистики», которые содержится в его предисловии к той же книге С.Кульбака и также приведенное в работах [1, 2]. В наше время в этом направлении продуктивно работают Дуглас Хаббард [20], а также российский математик В.Б.Вяткин [21-28]⁹⁷.

Кроме того, иногда авторы, излагающие в частности взвешенный метод наименьших квадратов, может быть не вполне осознанно используют слово «информация» не как научный термин, а в обиходном разговорном смысле. Например, в работе, приведенной на сайте: http://lib.alnam.ru/book_prs2.php?id=38, автор пишет: «Чтобы учесть ***разницу в информации, которую несет каждое наблюдение***, для нахождения оценки необходимо минимизировать взвешенную сумму квадратов отклонений» (отмечено мной, авт.). Казалось бы, остается «лишь» посчитать это количество информации и вариант взвешенного метода наименьших квадратов, основанный на теории информации, готов, но, однако мы видим, что ниже идет изложение стандартного ВМНК.

Таким образом, даже если принять в принципе изложенные выше идеи о применении количества информации в наблюдении в качестве веса наблюдения во взвешенном методе наименьших квадратов, то все равно остается очень существенный и принципиальный вопрос о том, ***каким способом возможно реально посчитать это количество информации.*** Этот вопрос разбивается на две части:

⁹⁶ Наверное абсолютно новых идей вообще не существует.

⁹⁷ Может быть синергетическая теория информация, созданная трудами В.Б.Вяткина, также может быть использована для расчета количества информации в наблюдениях, что приведет к еще одному варианту взвешенного метода наименьших квадратов.

– с помощью какого математического аппарата возможно посчитать количество информации в наблюдении?

– с помощью какого программного инструментария, реализующего этот математический аппарат, возможно посчитать количество информации в наблюдении?

Автоматизированный системно-когнитивный анализ (АСК-анализ) и его математическая модель (системная теория информации), а также реализующий их программный инструментарий АСК-анализа – система «Эйдос» – это и есть ответы на этот вопрос. Таким образом, АСК-анализ и система «Эйдос» представляют собой современную **интеллектуальную инновационную** (полностью готовую к внедрению) технологию взвешенного метода наименьших квадратов, модифицированного путем применения в качестве весов наблюдений количества информации в них.

При этом при принятии решений о применении для решения поставленной проблемы этой интеллектуальной инновационной технологии естественно возникает вопрос о степени точности восстановления в создаваемых с помощью нее моделях исследуемых эмпирических зависимостей в АСК-анализе и системе «Эйдос».

Традиционно точность восстановления зависимости оценивается дисперсиями и доверительным интервалами. В АСК-анализе **смысловым** аналогом доверительного интервала, в определенной степени, конечно, является количество информации в аргументе о значении функции. Поэтому **необходимо** исследовать соотношение смыслового содержания этих понятий: доверительного интервала и количества информации.

На математическом уровне это планируется сделать в будущем, а в данном разделе отметим лишь, что **чем больше доверительный интервал, тем выше неопределенность наших знаний о значении функции, соответствующем значению аргумента, а чем он меньше, тем эта определенность выше**. Но информация и определяется как количественная мера степени снятия неопределенности. Учитывая это можно утверждать, что **чем больше доверительный интервал, тем меньше информации о значении функции, соответствующем значению аргумента мы получаем, а чем он меньше, тем это количество информации больше**. Забегая вперед, отметим, что в **частично-редуцированных** когнитивных функциях, например изображенных на рисунке 15, количество информации в значениях аргумента о значениях функции наглядно изображено шириной полосы функции, что не только по смыслу, но внешне очень сходно с доверительным интервалом. При этом отметим еще один интересный момент, который состоит в том, что если традиционный доверительный интервал при экстраполяции при удалении от эмпирических значений ко все более отстоящим от них в будущем все время увеличивается, то в степень редукции когнитивной функции то увеличивается, то уменьшается. Это связано с тем, что **АСК-анализ и система «Эйдос» позволяют не только прогнозировать будущие события, но и прогнозировать достоверность или риски этих прогнозов** [7]⁹⁸, т.е. прогнозировать продолжительность периодов эргодичности и точки бифуркации (качественного изменения закономерностей в моделируемой предметной области), что наглядно и отображается в такой форме.

В частности при этом при нулевом доверительном интервале **формально** получается, что мы имеем бесконечное количество информации о значении функции, но **на практике** это вообще невозможно [17] и даже в теории возможно только для отдельных точек **целых** значений аргумента и функции. При бесконечном доверительном интервале в значении аргумента функции содержится ноль информации о значении функции.

⁹⁸ Подробнее об этом см., например, раздел: <http://lc.kubagro.ru/aidos/aidos02/7.4.htm>.

Ниже на простом численном примере мы кратко рассмотрим технологию, позволяющую на практике и в любой предметной области посчитать, какое количество информации содержится в наблюдении. В связи с ограничениями на объем статьи автор не имеет возможности полностью раскрыть все позиции на приведенных ниже скриншотах и рисунках, т.е. фактически предполагается некоторое предварительное знакомство читателя с системой «Эйдос». Если же такое знакомство недостаточно полное, то автор отсылает автора к публикациям в списке литературы и к сайту: <http://lc.kubagro.ru/>.

7.2.2.5. ЧИСЛЕННЫЙ ПРИМЕР

В АСК-анализе и системе «Эйдос» реализован аппарат когнитивных функций, который может быть применен для иллюстрации варианта взвешенного метода наименьших квадратов. На важность подобных наглядных примеров также указывал А.Н. Колмогоров: «По-видимому, внедрение предлагаемых методов⁹⁹ в практическую статистику будет облегчено, если тот же материал будет изложен более доступно и проиллюстрирован на подробно разобранных содержательных примерах» [1, 2, 19].

Для этой цели рассмотрим численный пример, основанный на исходных данных, приведенных в работе (таблица 3) [29]¹⁰⁰.

Необходимо отметить, что данные в таблице 3 достаточно *условные*, поскольку не содержат полного (адекватного) набора исходных данных, от которых зависит цена предложения квартиры (которая, кстати, в них и не содержится). В частности в таблице 3 нет числа комнат, указаны не все возможные типы домов, не учтена инфраструктура, как на сайте: http://1bezposrednikov.ru/krasnodar/kalkulyator_stoimosti/, не указано, входит ли площадь кухни в площадь квартиры, т.е. что это за площадь: общая или жилая, и т.д. Вместе с тем для целей данной статьи, т.е. для иллюстрации излагаемых в ней идей и методов, они достаточны (после некоторых корректировок, о которых сказано ниже).

Таблица 3 – Исходные данные для эконометрического моделирования¹⁰¹

№ наблюдения	Стоимость 1 кв.м. квартиры (руб./м ²)	Жилая площадь квартиры (м ²)	Тип дома	Наличие балкона	Площадь кухни (м ²)	Тип жилья
1	360,000	80	0	0	25	0
2	388,015	110	0	1	23	0
3	328,393	127	0	0	30	0
4	319,000	135	0	1	20	0
5	343,600	76	0	0	16	0
6	360,000	75	0	1	16	0
7	315,499	107	0	0	12	0
8	470,000	62	0	0	16	0
9	305,006	137	0	0	20	0
10	338,398	72	0	1	20	0
11	309,632	147	1	0	50	0
12	396,660	45	1	1	11,3	0
13	300,400	120	0	1	14	0
14	390,400	70	0	1	14	0
15	257,151	154	0	1	25	0
16	342,000	58	0	1	15	0

⁹⁹ Имеются в виду методы статистики, основанные на теории информации

¹⁰⁰ Очень хорошие описательные шкалы и градации, а также обучающую выборку можно сгенерировать с использованием калькулятора стоимости квартиры, приведенного на сайте: http://1bezposrednikov.ru/krasnodar/kalkulyator_stoimosti/.

¹⁰¹ Источник информации [3]: <http://www.scienceforum.ru/2014/489/626>

17	348,840	58	0	1	15,3	0
18	360,000	64	0	1	18	0
19	355,000	108	0	0	13	0
20	330,060	113	0	1	15	0
21	315,904	99	0	1	25	0
22	303,100	136	0	0	18	0
23	317,152	120	0	1	30	0
24	290,500	156	0	1	20	1
25	374,000	105	0	1	25	1
26	288,000	110	0	1	10,8	1
27	298,200	63	1	1	12	1
28	177,419	97	1	0	10	1
29	201,100	80	1	0	10	1
30	212,470	50	1	1	9	1
31	330,000	63	0	0	15	1
32	258,000	66	1	1	13	1
33	200,300	87	1	0	11	1
34	206,940	104	1	0	10	1
35	313,000	43	1	1	13	1
36	213,600	74	1	0	18	1
37	257,140	70	1	1	10	1
38	308,440	77	0	1	10,4	1
39	315,860	104	0	1	25	0
40	354,200	90	0	1	23	0
41	402,000	86	0	1	31	0
42	360,300	158	0	1	18	1
43	240,600	180	0	0	20	1
44	350,270	83	0	0	16	1
45	390,000	80	0	1	10	1
46	430,000	54	0	0	20	0
47	290,800	138	0	0	14	0
48	315,800	110	1	0	35	0
49	253,013	76	1	1	12	1
50	154,221	102	1	0	12,5	1
51	183,025	103	1	1	10,2	1
52	253,187	65	1	1	10	1
53	275,000	79	1	1	14	1
54	290,231	65	1	0	10	1
55	219,700	86	1	1	12	1
56	296,270	125	0	1	25	1
57	224,800	82	1	1	14	1
58	241,260	54	1	1	9,6	1
59	308,000	118	0	1	22,2	1
60	180,263	118	1	1	15	1
61	300,000	140	0	1	20	1
62	364,602	93	0	1	14	1
63	485,400	75	0	1	18	0
64	221,400	180	0	1	30	1
65	208,600	49	1	0	10	1
66	307,850	75	1	1	13	1
67	263,600	55	1	0	6,5	1
68	307,260	51	0	1	10	0
69	264,600	108	0	0	15	0
70	255,430	46	1	1	12	1
71	294,290	53	1	0	15	0
72	327,800	61	0	0	9	1
73	333,600	74	0	0	15	1
74	200,200	90	1	1	9	0

Факторы, от которых зависит стоимость квартиры, делятся на 2 типа:

1. Количественные:
 - жилая площадь квартиры (без площади кухни)¹⁰²;
 - площадь кухни.
2. Качественные:
 - тип дома: 0 – монолитный, 1 – панельный;
 - наличие балкона: 0 – нет; 1 – есть;
 - тип жилья: 0 – новостройка, 1 – вторичное жилье.

В таблице 3 произведена замена числовых кодов качественных факторов на лингвистические переменные. Это обеспечивает более высокую наглядность и читаемость выходных форм, а система «Эйдос» обеспечивает такую возможность, поэтому эта замена и была произведена. Кроме того добавлена расчетная колонка «Стоимость квартиры», равная произведению стоимости одного квадратного метра квартиры на ее *общую* площадь, а общая площадь (в явном виде не указанная в таблице) равна сумме жилой площади квартиры и площади кухни.

В результате этих операций получена таблица 4, которая является исходной для ввода в систему «Эйдос» с помощью одного и ее стандартных программных интерфейсов с внешними базами данных (режим 2.3.2.2).

Таблица 4 – Исходные данные для разработки интеллектуального приложения, иллюстрирующего модификацию взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции

№ наблюдения	Стоимость квартиры (руб.)	Стоимость 1 кв.м. квартиры (руб./м ²)	Жилая площадь квартиры (м ²)	Тип дома	Наличие балкона	Площадь кухни (м ²)	Тип жилья
1	30800,000	360,000	80	монолитный	нет	25,0	новостройка
2	45211,650	388,015	110	монолитный	есть	23,0	новостройка
3	45515,911	328,393	127	монолитный	нет	30,0	новостройка
4	45765,000	319,000	135	монолитный	есть	20,0	новостройка
5	27329,600	343,600	76	монолитный	нет	16,0	новостройка
6	28200,000	360,000	75	монолитный	есть	16,0	новостройка
7	35042,393	315,499	107	монолитный	нет	12,0	новостройка
8	30132,000	470,000	62	монолитный	нет	16,0	новостройка
9	44525,822	305,006	137	монолитный	нет	20,0	новостройка
10	25804,656	338,398	72	монолитный	есть	20,0	новостройка
11	52865,904	309,632	147	панельный	нет	50,0	новостройка
12	18358,200	396,660	45	панельный	есть	11,3	новостройка
13	37728,000	300,400	120	монолитный	есть	14,0	новостройка
14	28308,000	390,400	70	монолитный	есть	14,0	новостройка
15	43451,254	257,151	154	монолитный	есть	25,0	новостройка
16	20706,000	342,000	58	монолитный	есть	15,0	новостройка
17	21120,120	348,840	58	монолитный	есть	15,3	новостройка
18	24192,000	360,000	64	монолитный	есть	18,0	новостройка
19	39744,000	355,000	108	монолитный	нет	13,0	новостройка
20	38991,780	330,060	113	монолитный	есть	15,0	новостройка
21	33749,496	315,904	99	монолитный	есть	25,0	новостройка
22	43669,600	303,100	136	монолитный	нет	18,0	новостройка

¹⁰² Нами конкретизировано, что здесь имеется в виду именно жилая площадь, а не общая. Возможны и другие варианты, но для наших целей (численной иллюстрации метода) не играет роли, какой из них выбран.

23	41658,240	317,152	120	монолитный	есть	30,0	новостройка
24	48438,000	290,500	156	монолитный	есть	20,0	вторичное жилье
25	41895,000	374,000	105	монолитный	есть	25,0	вторичное жилье
26	32868,000	288,000	110	монолитный	есть	10,8	вторичное жилье
27	19542,600	298,200	63	панельный	есть	12,0	вторичное жилье
28	18179,643	177,419	97	панельный	нет	10,0	вторичное жилье
29	16888,000	201,100	80	панельный	нет	10,0	вторичное жилье
30	11073,500	212,470	50	панельный	есть	9,0	вторичное жилье
31	21735,000	330,000	63	монолитный	нет	15,0	вторичное жилье
32	17886,000	258,000	66	панельный	есть	13,0	вторичное жилье
33	18383,100	200,300	87	панельный	нет	11,0	вторичное жилье
34	22561,760	206,940	104	панельный	нет	10,0	вторичное жилье
35	14018,000	313,000	43	панельный	есть	13,0	вторичное жилье
36	17138,400	213,600	74	панельный	нет	18,0	вторичное жилье
37	18699,800	257,140	70	панельный	есть	10,0	вторичное жилье
38	24550,680	308,440	77	монолитный	есть	10,4	вторичное жилье
39	35449,440	315,860	104	монолитный	есть	25,0	новостройка
40	33948,000	354,200	90	монолитный	есть	23,0	новостройка
41	37238,000	402,000	86	монолитный	есть	31,0	новостройка
42	59771,400	360,300	158	монолитный	есть	18,0	вторичное жилье
43	46908,000	240,600	180	монолитный	нет	20,0	вторичное жилье
44	30400,410	350,270	83	монолитный	нет	16,0	вторичное жилье
45	32000,000	390,000	80	монолитный	есть	10,0	вторичное жилье
46	24300,000	430,000	54	монолитный	нет	20,0	новостройка
47	42062,400	290,800	138	монолитный	нет	14,0	новостройка
48	38588,000	315,800	110	панельный	нет	35,0	новостройка
49	20140,988	253,013	76	панельный	есть	12,0	вторичное жилье
50	17005,542	154,221	102	панельный	нет	12,5	вторичное жилье
51	19902,175	183,025	103	панельный	есть	10,2	вторичное жилье
52	17107,155	253,187	65	панельный	есть	10,0	вторичное жилье
53	22831,000	275,000	79	панельный	есть	14,0	вторичное жилье
54	19515,015	290,231	65	панельный	нет	10,0	вторичное жилье
55	19926,200	219,700	86	панельный	есть	12,0	вторичное жилье
56	40158,750	296,270	125	монолитный	есть	25,0	вторичное жилье
57	19581,600	224,800	82	панельный	есть	14,0	вторичное жилье
58	13546,440	241,260	54	панельный	есть	9,6	вторичное жилье
59	38963,600	308,000	118	монолитный	есть	22,2	вторичное жилье
60	23041,034	180,263	118	панельный	есть	15,0	вторичное жилье
61	44800,000	300,000	140	монолитный	есть	20,0	вторичное жилье
62	35209,986	364,602	93	монолитный	есть	14,0	вторичное жилье
63	37755,000	485,400	75	монолитный	есть	18,0	новостройка
64	45252,000	221,400	180	монолитный	есть	30,0	вторичное жилье
65	10711,400	208,600	49	панельный	нет	10,0	вторичное жилье
66	24063,750	307,850	75	панельный	есть	13,0	вторичное жилье
67	14855,500	263,600	55	панельный	нет	6,5	вторичное жилье
68	16180,260	307,260	51	монолитный	есть	10,0	новостройка
69	30196,800	264,600	108	монолитный	нет	15,0	новостройка
70	12301,780	255,430	46	панельный	есть	12,0	вторичное жилье
71	16392,370	294,290	53	панельный	нет	15,0	новостройка
72	20544,800	327,800	61	монолитный	нет	9,0	вторичное жилье
73	25796,400	333,600	74	монолитный	нет	15,0	вторичное жилье
74	18828,000	200,200	90	панельный	есть	9,0	новостройка
75	40999,920	495,640	78	монолитный	есть	30,0	новостройка

По условиям задачи, рассматриваемой в данной работе в качестве численного примера применения предлагаемого метода, на основе исходных данных, приведенных


в таблице 4, необходимо найти *зависимости* стоимости квартиры от всех ее характеристик, приведенных в этих исходных данных.


Для решения этой задачи прежде всего необходимо скачать и установить систему «Эйдос». Скачать систему «Эйдос-Х++» (самую новую на текущий момент версию) или обновление системы до текущей версии, можно на сайте: <http://lc.kubagro.ru/> по адресу: http://lc.kubagro.ru/aidos/_Aidos-X.htm. По этой ссылке всегда находится наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с исходными текстами, находящаяся в полном открытом бесплатном доступе (около 50 Мб) (инструкция).

ИНСТРУКЦИЯ


по скачиванию и установке системы «Эйдос» (объем около 50 Мб)

Система не требует инсталляции, не меняет никаких системных файлов и содержимого папок операционной системы, т.е. является портативной (portable) программой. Но чтобы она работала необходимо аккуратно выполнить следующие пункты.


1. Скачать самую новую на текущий момент версию системы «Эйдос-Х++» по ссылке: <http://lc.kubagro.ru/a.rar> (ссылки для обновления системы даны в режиме 6.2)
2. Разархивировать этот архив в любую папку с правами на запись с коротким латинским именем и путем доступа, включающим только папки с такими же именами (лучше всего в корневой каталог какого-нибудь диска).
3. Запустить систему. Файл запуска: _AIDOS-X.exe *
4. Задать имя: 1 и пароль: 1 (потом их можно поменять в режиме 1.2).
5. Перед тем как запустить новый режим НЕОБХОДИМО ЗАВЕРШИТЬ предыдущий (Help можно не закрывать). Окна закрываются в порядке, обратном порядку их открытия.

* Разработана программа: «_START_AIDOS.exe», полностью снимающая с пользователя системы «Эйдос-Х++» заботу о проверке наличия и скачивании обновлений. Эту программу надо просто скачать по ссылке: http://lc.kubagro.ru/Install_Aidos-X/_START_AIDOS.exe, поместить в папку с исполнимым модулем системы и всегда запускать систему с помощью этого файла.


При запуске программы _START_AIDOS.EXE система Эйдос не должна быть запущена, т.к. она содержится в файле обновлений и при его разархивировании возникнет конфликт, если система будет запущена.

1. Программа _START_AIDOS.exe определяет дату системы Эйдос в текущей папке, и дату обновлений на FTP-сервере не скачивая их, и, если система Эйдос в текущей папке устарела, скачивает обновления. (Если в текущей папке нет исполнимого модуля системы Эйдос, то программа пытается скачать полную инсталляцию системы, но не может этого сделать из-за ограниченной функциональности демо-версии библиотеки Xb2NET.DLL).

2. После этого появляется диалоговое окно с сообщением, что надо сначала разархивировать систему, заменяя все файлы (опция: «Yes to All» или «OverWrite All»), и только после этого закрыть данное окно.

3. Потом программа _START_AIDOS.exe запускает обновления на разархивирование. После окончания разархивирования окно архиватора с отображением стадии процесса исчезает.


4. После закрытия диалогового окна с инструкцией (см. п.2), происходит запуск обновленной версии системы Эйдос на исполнение.

Для работы программы _START_AIDOS.exe необходима библиотека: Xb2NET.DLL, которую можно скачать по ссылке: http://lc.kubagro.ru/Install_Aidos-X/Xb2NET.DLL. Перед первым запуском этой программы данную библиотеку необходимо скачать и поместить либо в папку с этой программой, а значит и исполнимым модулем системы «Эйдос-Х++», либо в любую другую папку, на которую в операционной системе прописаны пути поиска файлов, например в папку: c:\Windows\System32\. Эта библиотека стоит около 500\$ и у меня ее нет, поэтому я даю только бесплатную демо-версию, которая выдает сообщение об ограниченной функциональности, но для наших целей ее достаточно.

Лицензия:

Автор отказывается от какой бы то ни было ответственности за последствия применения или не применения Вами системы «Эйдос».
Проще говоря, пользуйтесь если понравилось, а если не понравилось – сотрите и забудьте, а лучше вообще не скачивайте.

Необходимо отметить, что на папку с системой у пользователя должны быть все права доступа, иначе система не сможет корректировать свои базы данных и индексные массивы, что необходимо для ее нормальной работы.

Затем записываем таблицу 4 в виде Excel-файла с именем Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\Inp_data.xls и запускаем систему (файл запуска: _AIDOS-X.exe).

При запуске системы появляется окно авторизации:

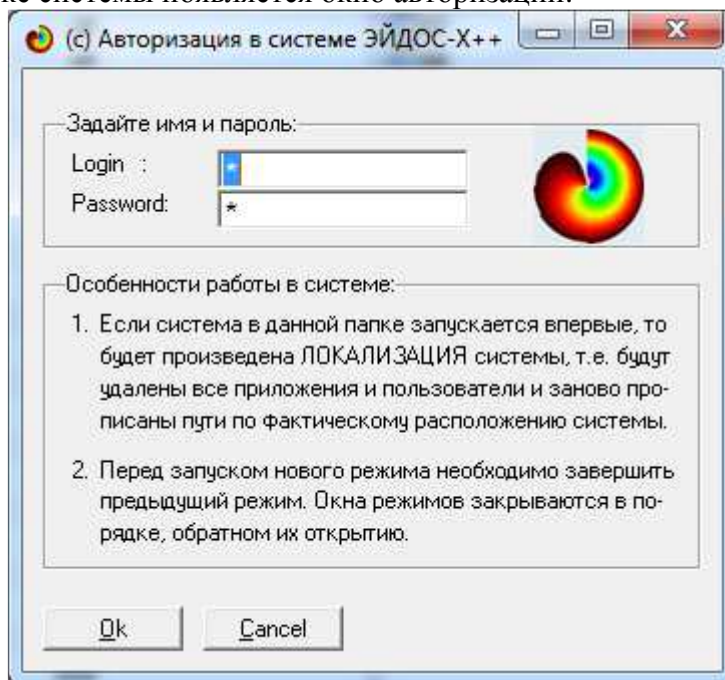


Рисунок 4 – Окно авторизации системы «Эйдос»

Вводим начальные имя 1 и пароль 1, которые в последующем можно изменить в режиме 1.2.

Отметим, что система «Эйдос» является программным инструментарием АСК-анализа и автоматизирует все его этапы, кроме первого:

1. Когнитивная структуризация предметной области (неформализованный этап).

На этом этапе решается, что мы хотим прогнозировать и на основе чего.

2. Формализация предметной области. На этом этапе разрабатываются классификационные и описательные шкалы и градации, а затем с их использованием исходные данные кодируются и представляются в форме баз событий, между которыми могут быть выявлены причинно-следственные связи.

3. Синтез и верификация моделей (оценка достоверности, адекватности). Повышение качества модели. Выбор наиболее достоверной модели для решения в ней задач.

4. Решение задач идентификации и прогнозирования.

5. Решение задач принятия решений и управления.

6. Решение задач исследования моделируемой предметной области путем исследования ее модели.

На рисунке 3 приведены автоматизированные в системе «Эйдос» этапы АСК-анализа, которые обеспечивают последовательное повышение степени формализации модели путем преобразования исходных данных в информацию, а далее в знания:

Для выполнения 2-го этапа АСК-анализа запускаем универсальный программный интерфейс ввода данных из внешних баз данных (режим 2.3.2.2) (рисунок 6):

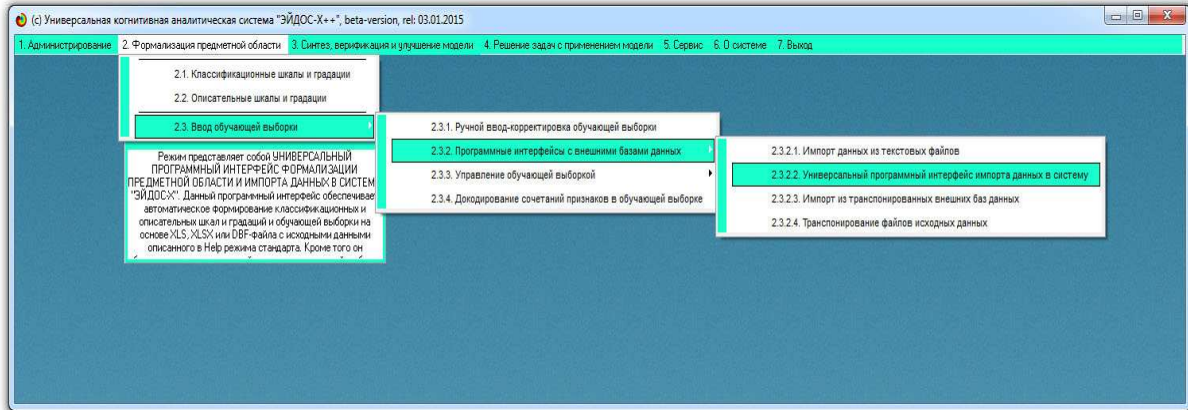


Рисунок 6 – Запуск универсального программного интерфейса ввода данных из внешних баз данных

Появляется следующая экранная форма (рисунок 7):

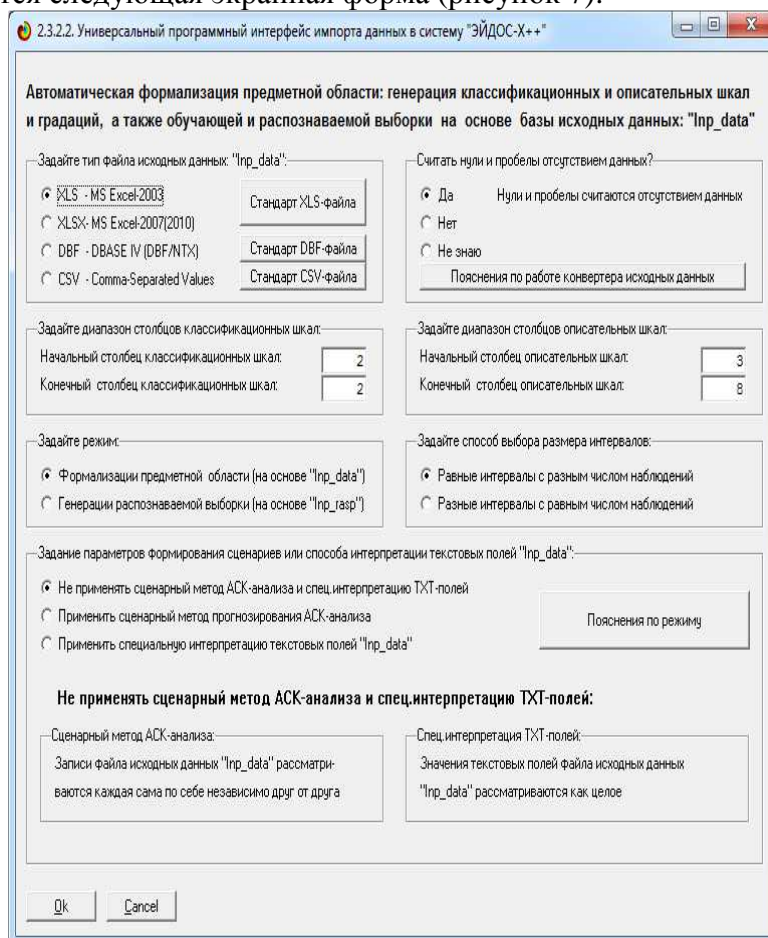


Рисунок 7 – Экранная форма задания параметров универсального программного интерфейса ввода данных из внешних баз данных

На рисунке 6 показаны нужные в данном случае значения задаваемых параметров.

Help данного режима приведен на рисунке 8:



Рисунок 8 – Экранная форма Help универсального программного интерфейса ввода данных из внешних баз данных

Таблица 4 соответствует требованиям системы «Эйдос» к внешним базам данных, приведенным на рисунке 8.

Если кликнуть ОК на экранной форме, приведенной на рисунке 6, то начинается автоматический процесс формализации предметной области, который начинается с конвертирования Excel-файла в dbf-файл. При этом на заднем фоне может возникнуть окно, приведенное на рисунке 9:

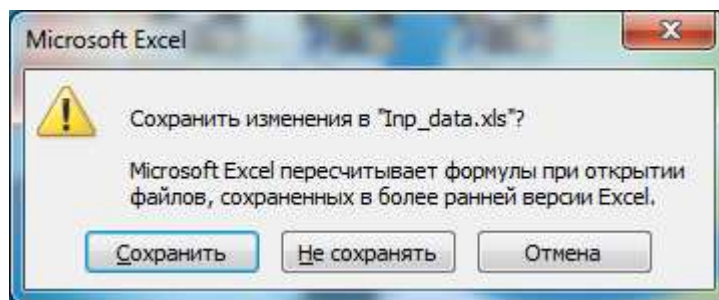


Рисунок 9 – Окно на заднем фоне, возникающее при пересчете Excel-файла в процессе его преобразования в dbf-файл

Чтобы увидеть это окно надо кликнуть по иконке системы «Эйдос» на панели задач при всех свернутых окнах других приложений или их отсутствии. На этом окне можно выбрать любой вариант, кроме отмены.

Сразу же после этого система находит классификационные и описательные шкалы и градации, определяет тип данных в шкалах и отображает окно, приведенное на рисунке 10:

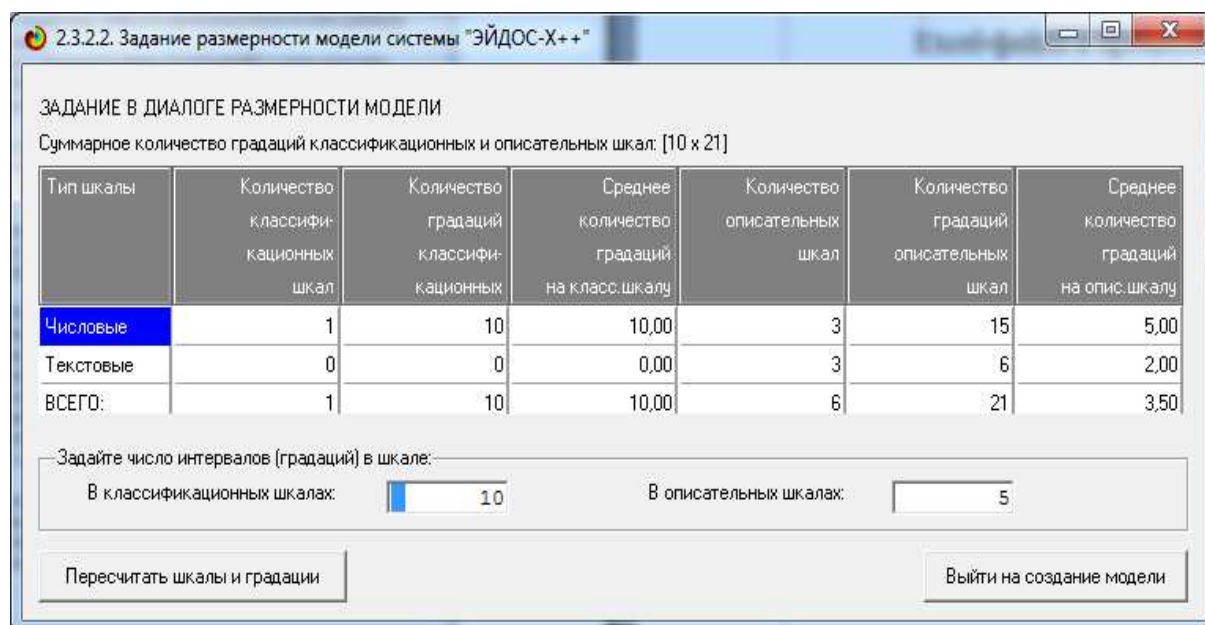


Рисунок 10 – Внутреннего калькулятора универсального программного интерфейса импорта данных из внешних баз данных

Если в таблице исходных данных есть числовые шкалы, то появляется возможность задать количество интервальных числовых значений (интервалов в числовых шкалах) в них отдельно для классификационных и описательных шкал. Принцип определения разумного количества интервалов такой. Если их задать очень много, то в некоторых интервалах вообще не будет данных или будет очень мало (меньше 5), что нежелательно. Если задать интервалов очень мало, то они будут очень большого размера и точность модели будет не высока. Таким образом, можно сделать такой вывод, что чем больше объем выборки, тем меньшего размера мы можем позволить себе задавать интервалы. Но не нужно этим особенно увлекаться, т.е. если есть возможность сделать очень маленькие интервалы, но нам не нужна такая точность, то лучше делать интервалы такого размера, чтобы они обеспечили необходимую точность, но не меньшего размера. В режиме 2.3.2.2 есть возможность задавать либо равные интервалы с разным

числом наблюдений, либо разные интервалы с примерно одинаковым числом измерений. Это может иметь смысл, если в исходных данных в числовых шкалах представлен широкий спектр частот, и мы не хотим терять высокочастотные гармоники, которые могут оказаться не оцифрованными при равных интервалах. Это позволяет автоматически ставить точки тем чаще, чем выше кривизна кривых, построенных на шкалах. Все эти рассуждения напоминает какие-то следствия теоремы Котельникова об отсчетах.

В данной экранной форме задаем количество интервалов в классификационных и описательных шкалах. Если оно изменяется, то необходимо кликнуть по кнопке «Пересчитать шкалы и градации», а затем, когда будет выбран окончательный вариант, выйти на создание модели.

Сразу же начинается процесс импорта данных в систему «Эйдос», этапы и прогноз времени исполнения которого отображаются на экранной форме (рисунок 11):

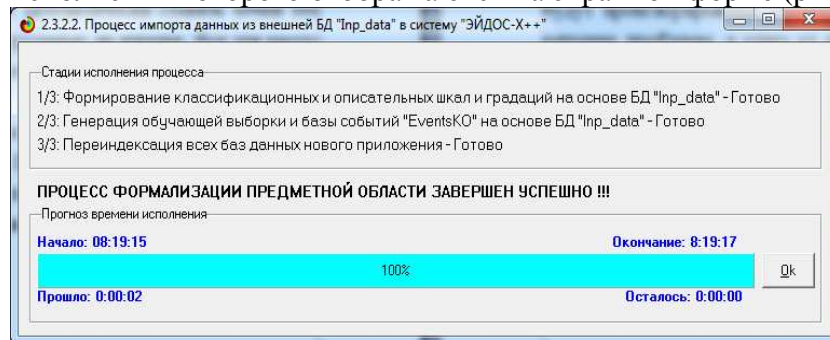


Рисунок 11 – Внутреннего калькулятора универсального программного интерфейса импорта данных из внешних баз данных

Затем в режиме 3.5 системы «Эйдос» с параметрами по умолчанию (рисунок 12) выполняется 3-й этап АСК-анализа, т.е. синтез и верификация модели:

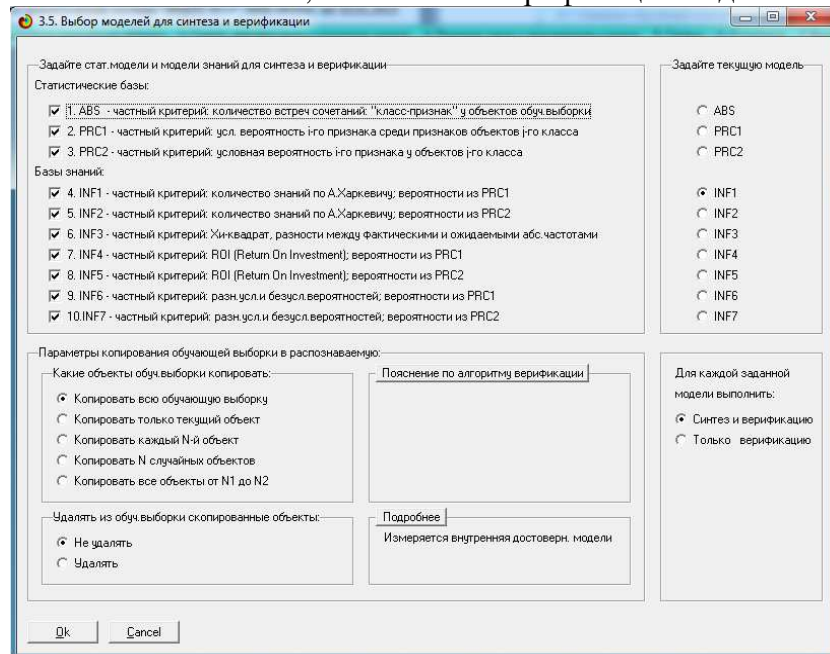


Рисунок 12 – Экранная форма задания параметров режима синтеза и верификации модели системы «Эйдос»

Этапы выполнения данного режима и прогноз времени исполнения отображаются на экранной форме (рисунок 13):

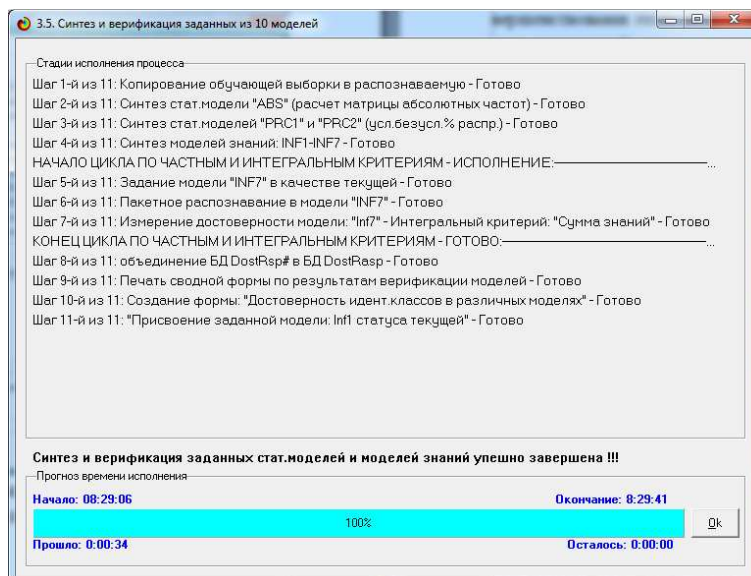


Рисунок 13 – Экранная форма с отображением этапов прогнозом времени исполнения режима синтеза и верификации модели системы «Эйдос»

Перейдем теперь в режим 4.5 «Визуализация когнитивных функций» (рисунок 14):

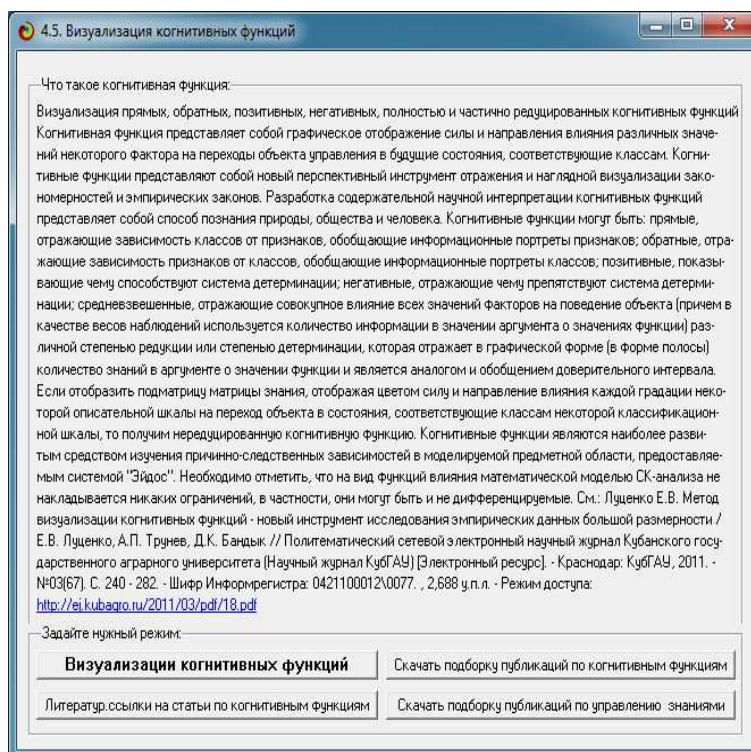


Рисунок 14 – Начальная экранная форма режима визуализации когнитивных функций системы «Эйдос»

На рисунке 15 приведены визуализации когнитивной функции (КФ) зависимости стоимости квартиры от стоимости одного квадратного метра ее площади при разных способах определения и визуализации частично редуцированных когнитивных функций.

Программная реализация данного режима визуализации когнитивных функций разработан по постановке автора разработчиком интеллектуальных, графических и музыкальных систем из Белоруссии Дмитрием Константиновичем Бандык [30].

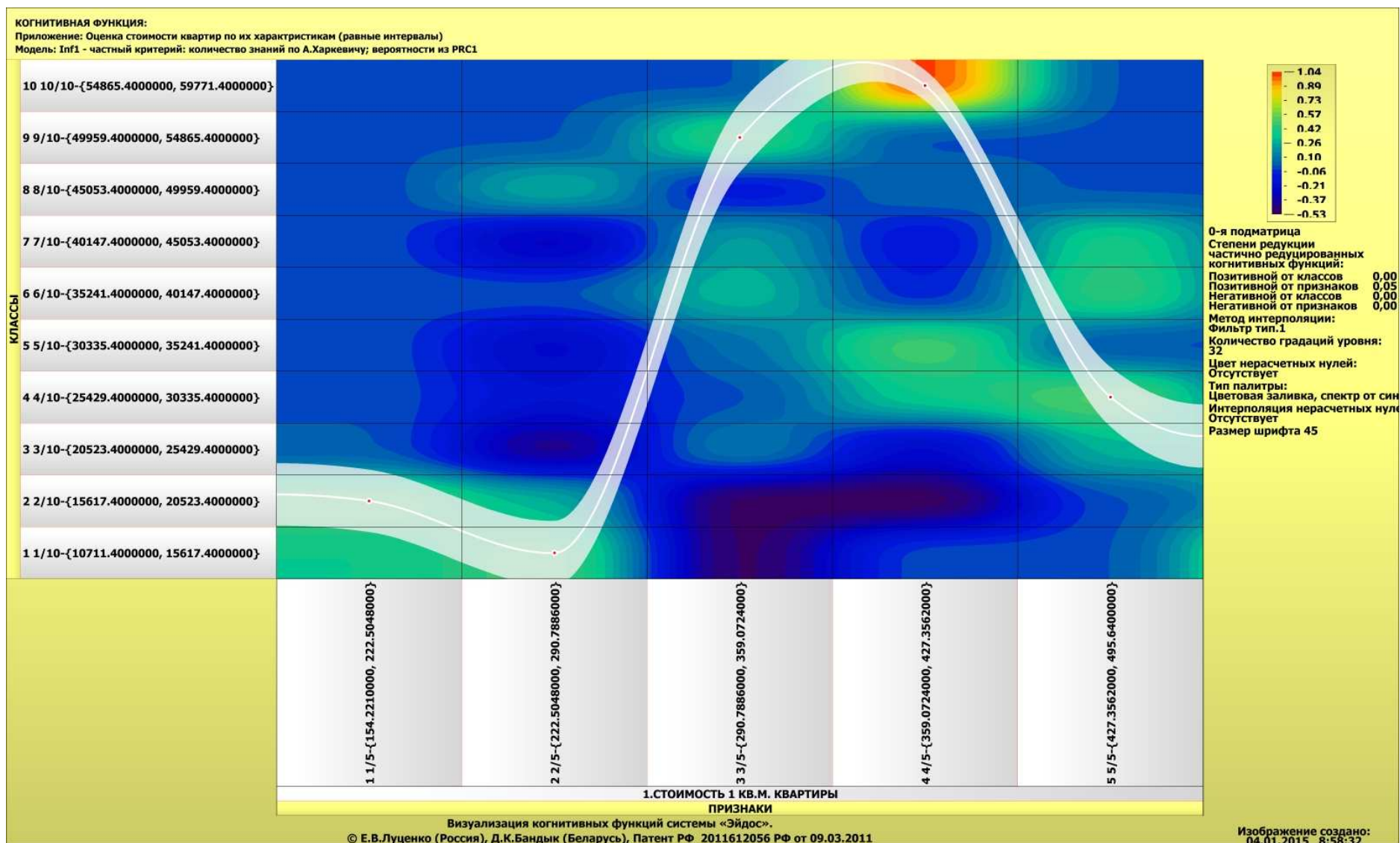


Рисунок 15-а – Визуализации когнитивной функции зависимости стоимости квартиры от стоимости 1 кв.метра: частично-редуцированная КФ проведена по значениям функции, о которых в аргументе содержится максимальное количество информации

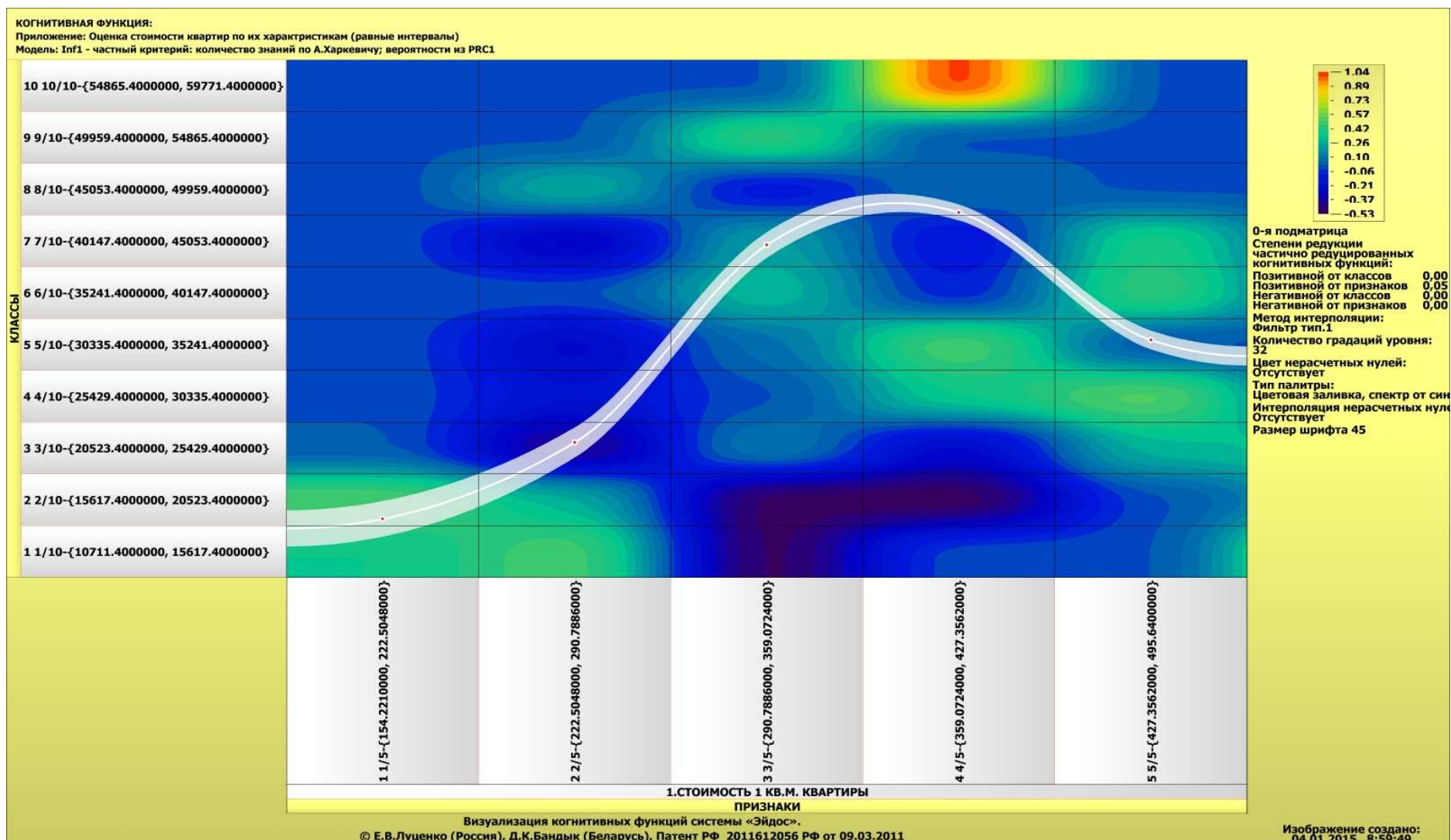


Рисунок 15-б – Визуализации когнитивной функции зависимости стоимости квартиры от стоимости 1 кв.метра: частично-редуцированная КФ проведена по точкам, полученным путем применения предложенной модификации взвешенного метода наименьших квадратов, в котором в качестве весов наблюдений используется количество информации в аргументе о значении функции.

При этом применены настройки параметров отображения когнитивных функций, приведенные, приведенные на рисунке 16:

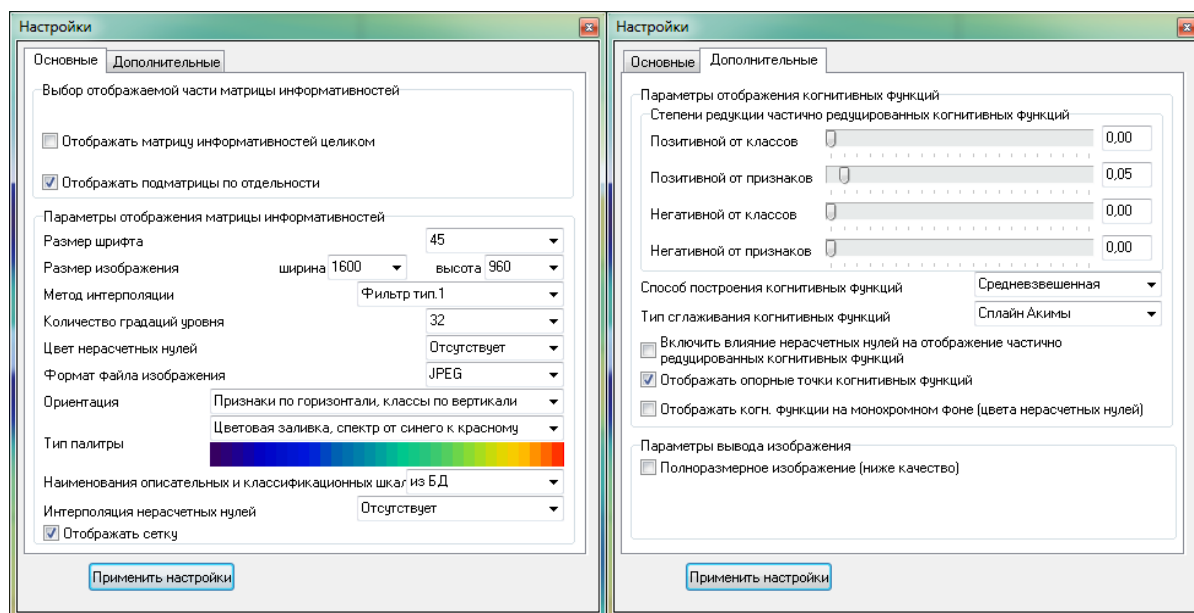


Рисунок 16 – Настройки параметров отображения когнитивных функций в режиме 4.5 системы «Эйдос»

Из рисунка 15 мы видим, что у дешевых квартир минимальная стоимость 1 кв.метра, а максимальной она вопреки ожиданиям является у просто дорогих квартир, а не у самых дорогих.

По соям координат приведены интервальные числовые значения:

- по оси X: стоимости 1 квадратного метра жилья;
- по оси Y: стоимости квартиры.

Графики оцененной зависимости, полученные с помощью предложенной модификации взвешенного метода наименьших квадратов, основанного на применении в качестве весов наблюдений количества информации в аргументе о значении функции, будут приведены ниже.

Естественно возникает вопрос о степени точности восстановления исследуемых эмпирических зависимостей в моделях, созданных с применением АСК-анализе и системе «Эйдос».

Традиционно точность восстановления зависимости оценивается дисперсиями и доверительным интервалами. В АСК-анализе **смысловым** аналогом доверительного интервала, в определенной степени, конечно, является количество информации в аргументе о значении функции. Поэтому **необходимо** исследовать соотношение смыслового содержания этих понятий: доверительного интервала и количества информации.

На математическом уровне это планируется сделать в будущем, а в данном разделе отметим лишь, что **чем больше доверительный интервал, тем выше неопределенность наших знаний о значении функции, соответствующем значению аргумента, а чем он меньше, тем эта определенность выше**. Но информация и определяется как количественная мера степени снятия неопределенности. Учитывая это можно утверждать, что **чем больше доверительный интервал, тем меньше информации о значении функции, соответствующем значению аргумента мы получаем, а чем**

он меньше, тем это количество информации больше. Забегая вперед, отметим, что в *частично-редуцированных* когнитивных функциях, например изображенных на рисунке 15, количество информации в значениях аргумента о значениях функции наглядно изображено шириной полосы функции, что не только по смыслу, но внешне очень сходно с доверительным интервалом. При этом отметим еще один интересный момент, который состоит в том, что если традиционный доверительный интервал при экстраполяции при удалении от эмпирических значений ко все более отстоящим от них в будущем все время увеличивается, то в степень редукции когнитивной функции то увеличивается, то уменьшается. Это связано с тем, что *АСК-анализ и система «Эйдос» позволяют не только прогнозировать будущие события, но и прогнозировать достоверность или риски этих прогнозов* [7]¹⁰³, т.е. прогнозировать продолжительность периодов эргодичности и точки бифуркации (качественного изменения закономерностей в моделируемой предметной области), что наглядно и отображается в такой форме.

В частности при этом при нулевом доверительном интервале **формально** получается, что мы имеем бесконечное количество информации о значении функции, но **на практике** это вообще невозможно [17] и даже в теории возможно только для отдельных точек **целых** значений аргумента и функции. При бесконечном доверительном интервале в значении аргумента функции содержится ноль информации о значении функции.

Когнитивные функции, приведенные на рисунке 15, получены на основе модели знаний, основанной на мере А.Харкевича, в которой учтены все переменные, т.е. факторы или описательные шкалы модели и отражено их взаимное влияние друг на друга и выходные параметры. Это влияние отражено в результатах кластерно-конструктивного анализа, отображенных в форме семантических сетей на рисунках 17 и 18:

Итак, из рисунка 17 мы видим, что классификационные шкалы, являющиеся осями в когнитивном пространстве классов, зависят друг от друга, т.е. неортонормированы. Из рисунка 18 мы видим, что описательные шкалы (факторы), являющиеся осями в когнитивном пространстве факторов, также зависят друг от друга, т.е. неортонормированы.

Таким образом, когнитивное (фазовое) пространство модели знаний системы «эйдос» является неортонормированным, а модель, следовательно, является нелинейной. Поэтому очень важно, что в АСК-анализе и системе «Эйдос» используется неметрический интегральный критерий, не основанный на предположении об ортонормированности пространства [7].

Рассмотрим теперь применение предложенной модификации взвешенного метода наименьших квадратов, в котором в качестве весов наблюдений используется количество информации в аргументе о значении функции. Для этой цели разработан режим 4.6 «Подготовка баз данных для визуализации когнитивных функций в MS Excel».

¹⁰³ Подробнее об этом см., например, раздел: <http://lc.kubagro.ru/aidos/aidos02/7.4.htm>.

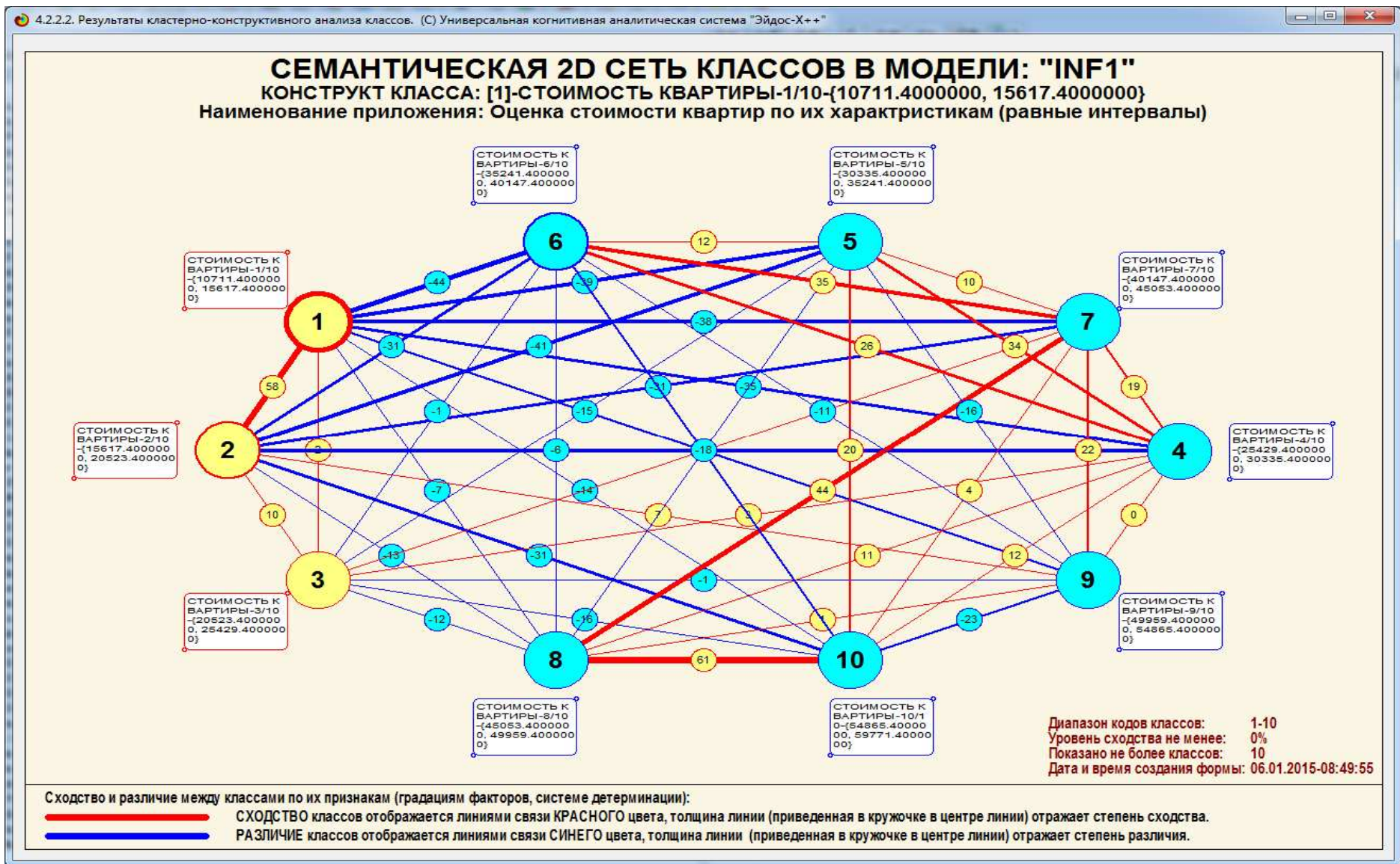


Рисунок 17 – Результаты кластерно-конструктивного анализа классов, т.е. их сходство и различие по системе детерминации

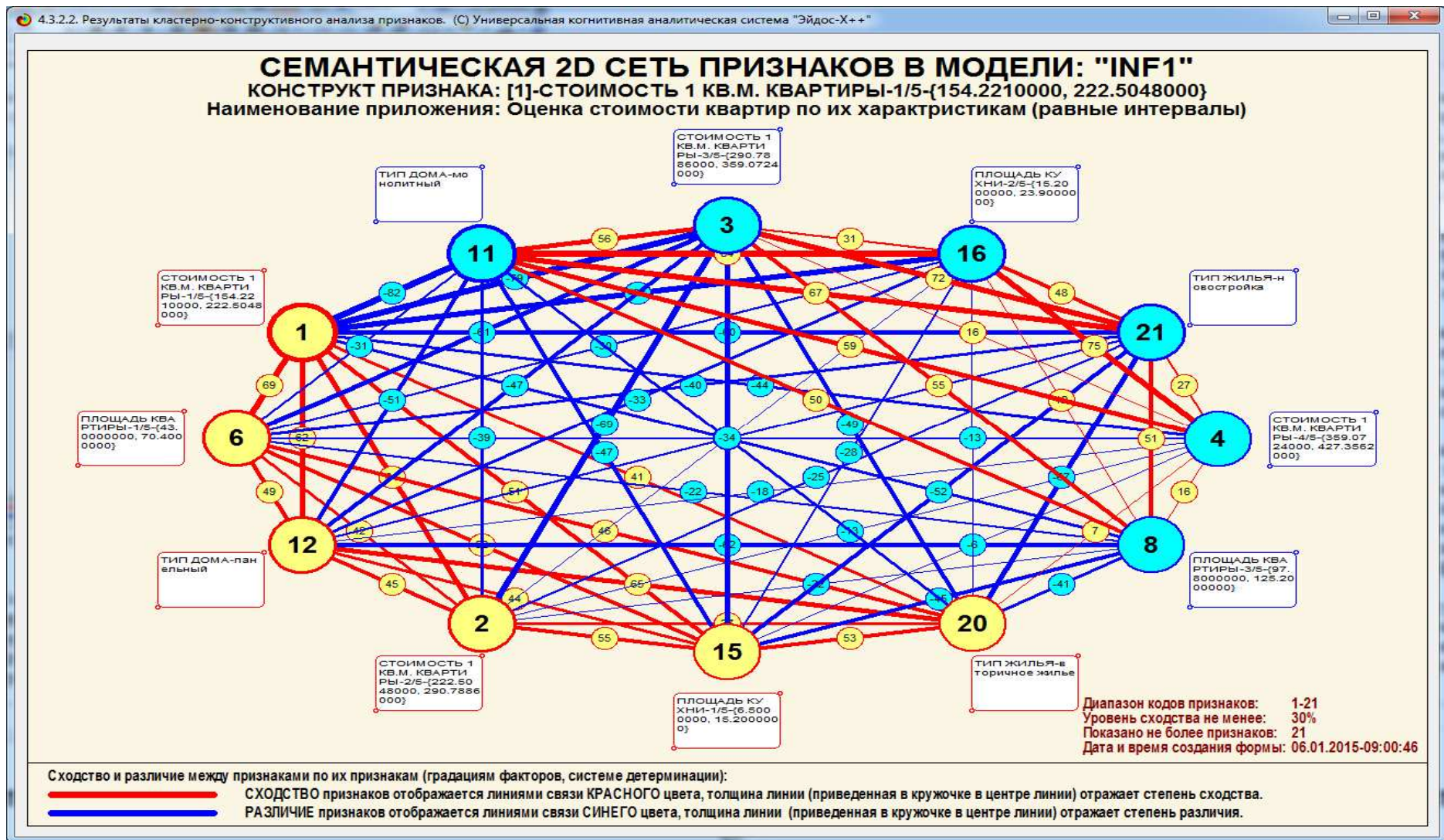


Рисунок 18. Результаты кластерно-конструктивного анализа значений факторов, т.е. их сходство и различие по классам, принадлежность и не принадлежность к которым они обуславливают

При разработке этого режима использованы следующие *идеи*.

1. MS Excel (особенно версии начиная с 2007) обладает очень удобными средствами регрессионного анализа, использующими *стандартный* метод наименьших квадратов, которые целесообразно использовать.

2. Однако, поскольку MS Excel в регрессионном анализе использует лишь *стандартный* метод наименьших квадратов, в котором все наблюдения имеют одинаковый (единичный) вес, то предлагается отражать вес наблюдения количеством точек.

3. Поскольку вес наблюдения в предлагаемой нами модификации взвешенного метода наименьших квадратов равен количеству информации в аргументе о значении функции, то для того, чтобы посчитать это количество точек для каждого наблюдения необходимо приписать точке определенное количество информации.

4. Это можно сделать расчетным путем для каждого наблюдения зная количество информации в данном наблюдении и количество точек в наблюдении с максимальным количеством информации. Количество информации в данном наблюдении определяется при синтезе и верификации моделей в системе «Эйдос», а количество точек в наблюдении с максимальным количеством информации необходимо задать в диалоге.

5. Если для каждого наблюдения все точки, количество которых отражает количество информации в данном наблюдении, отображать с их точными координатами, то они все попадают в одну точку на изображении. Чтобы было видно, сколько этих точек в данном наблюдении предлагается задавать небольшое случайное рассеяние этих точек вокруг точки с точными значениями координат. Величину этого рассеяния можно задавать в диалоге в процентах от диапазона значений описательной и классификационной шкалы отображаемой подматрицы.

6. Стандартный режим регрессионного анализа MS Excel будет строит регрессии с учетом всех точек каждого наблюдения, сгенерированных в количестве, пропорциональном количеству информации в этом наблюдении. Поэтому полученная регрессия будет соответствовать предлагаемой модификации взвешенного метода наименьших квадратов.

При запуске режима 4.6 «Подготовка баз данных для визуализации когнитивных функций в MS Excel» отображается окно настройки параметров (рисунок 19):

Выполняется этот режим довольно быстро (несколько секунд), т.к. его алгоритм сводится к выборкам данных из ранее посчитанных статистических баз и баз знаний, представленных в системе «Эйдос» в нечеткой декларативной форме. Если бы в системе «Эйдос» использовалась четкая процедурная модель представления знаний, при котором генерация знаний производилась бы с различными степенями нечеткости непосредственно перед их использованием, то данный режим работал бы на много порядков медленнее и был бы непригоден для реального практического применения.

По окончании работы режима выводится экранная форма, представленная на рисунке 20:

На рисунке 21 приведен скриншот, на котором показано содержимое папки: `..\AID_DATA\A#####\System\Cogn_fun\` с базами данных для визуализации когнитивных функций и регрессий, созданных в режиме 4.6 в соответствии с параметрами, приведенными на рисунке 19.

Из всех созданных баз данных рассмотрим только те, которые позволяют отобразить те же подматрицы баз знаний (сочетания классификационных и описательных шкал), что и на рисунке 15, это базы данных с именами:

- Inf1-Y(X)-Pos-One_point-1-1.dbf;
- Inf1-Y(X)-Pos-All_points-1-1.dbf.

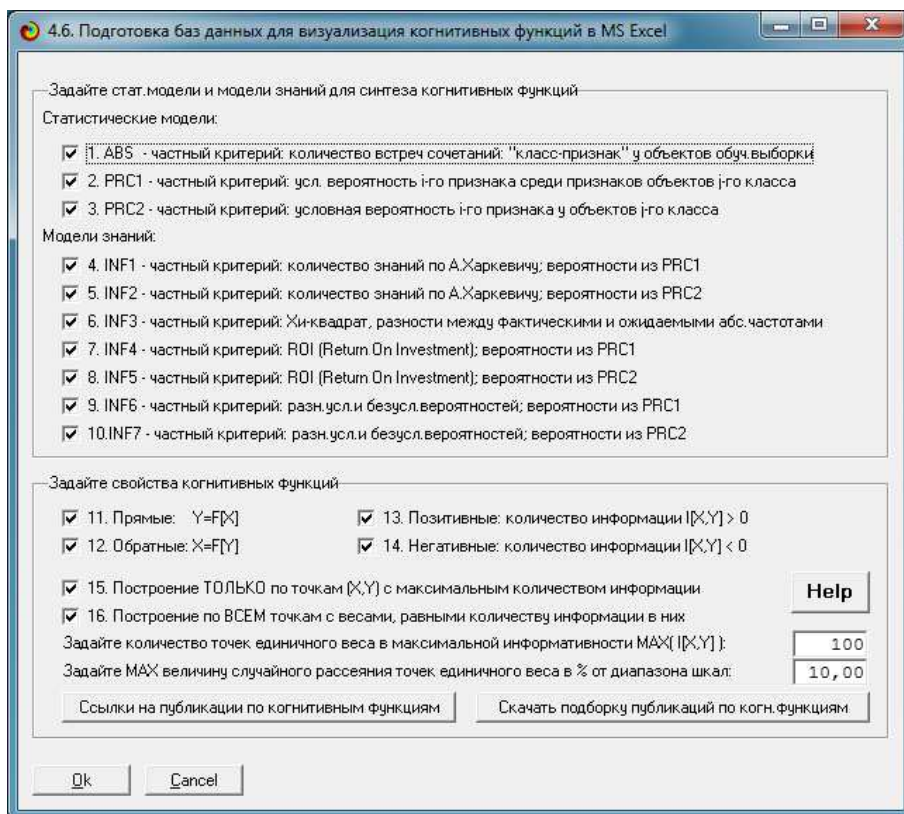


Рисунок 19. Настройки параметров создаваемых баз данных для визуализации когнитивных функций в режиме 4.6 системы «Эйдос»

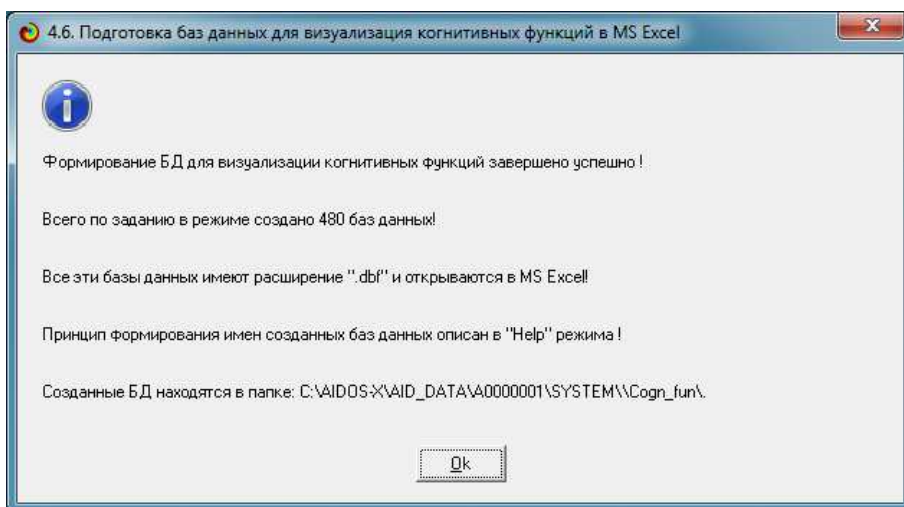


Рисунок 20. Экранная форма, отображаемая по окончании выполнения режима 4.6 системы «Эйдос»

В результате работы режима 4.6 формируются базы данных для визуализации редуцированных когнитивных функций, имена которых формируются способом, который поясняет таблица 5.

Таблица 5 – Виды и имена баз данных для визуализации когнитивных функций, формируемые в режиме 4.6

Прямые и обратные	Позитивные и негативные	Построенные по точкам с максимальным количеством информации или по ВСЕМ точкам с весами, равными количеству информации в них	Имена баз данных для визуализации КФ в MS Excel
Прямые: $Y=F[X]$	Позитивные: количество информации $I[X, Y] > 0$	Построение ТОЛЬКО по точкам (X, Y) с максимальным количеством информации	КФ-1: #####-Y(X)-Pos-One_point-##-##.dbf
		Построение по ВСЕМ точкам с весами, равными количеству информации в них	КФ-2: #####-Y(X)-Pos-All_points-##-##.dbf
	Негативные: количество информации $I[X, Y] < 0$	Построение ТОЛЬКО по точкам (X, Y) с максимальным количеством информации	КФ-3: #####-Y(X)-Neg-One_point-##-##.dbf
		Построение по ВСЕМ точкам с весами, равными количеству информации в них	КФ-4: #####-Y(X)-Neg-All_points-##-##.dbf
Обратные: $X=F[Y]$	Позитивные: количество информации $I[X, Y] > 0$	Построение ТОЛЬКО по точкам (X, Y) с максимальным количеством информации	КФ-5: #####-X(Y)-Pos-One_point-##-##.dbf
		Построение по ВСЕМ точкам с весами, равными количеству информации в них	КФ-6: #####-X(Y)-Pos-All_points-##-##.dbf
	Негативные: количество информации $I[X, Y] < 0$	Построение ТОЛЬКО по точкам (X, Y) с максимальным количеством информации	КФ-7: #####-X(Y)-Neg-One_point-##-##.dbf
		Построение по ВСЕМ точкам с весами, равными количеству информации в них	КФ-8: #####-X(Y)-Neg-All_points-##-##.dbf

Эти базы данных формируются для всех моделей (в начале имен БД наименования моделей): { **Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7** } и для всех сочетаний классификационных и описательных шкал (в конце имен БД коды шкал) и записываются в папку: **..\AID_DATA\A#####\System\Cogn_fun**.

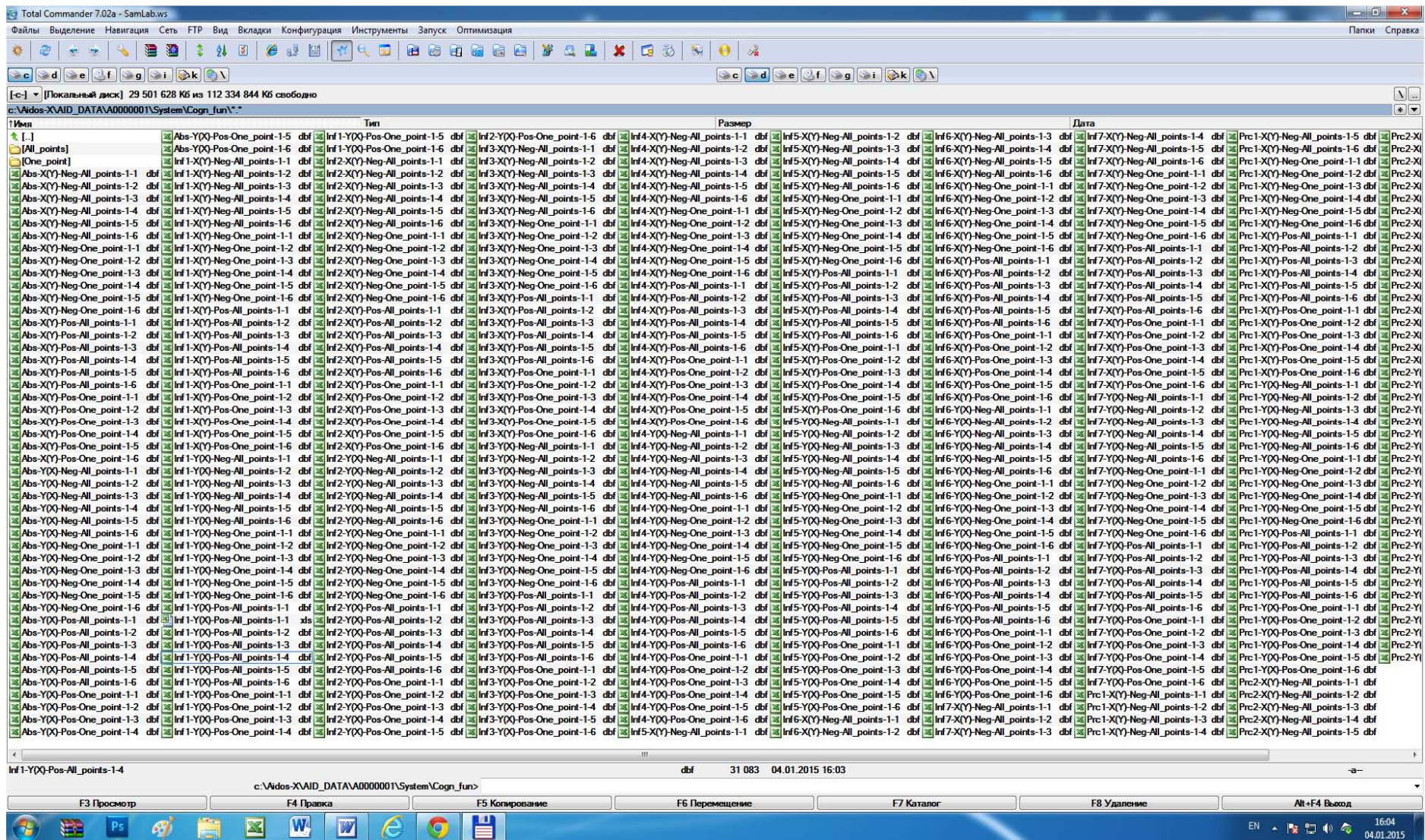


Рисунок 21 – Скриншот, на котором показано содержимое папки: ..\AID_DATA\A#####\System\Cogn_fun\ с базами данных для визуализации когнитивных функций и регрессий, созданных в режиме 4.6 в соответствии с параметрами, приведенными на рисунке 19

В таблице 6 приведена база данных Inf1-Y(X)-Pos-One_point-1-1.dbf, а в таблице 7 – фрагмент базы данных Inf1-Y(X)-Pos-All_points-1-1.dbf.

Таблица 6 – База данных «Inf1-Y(X)-Pos-One_point-1-1.dbf»
для визуализации когнитивных функций по точкам
с максимальным количеством информации в наблюдениях

Наименование градации описательной шкалы	Наименование градации классификационной шкалы	Градация опис.шкалы	Градация класс.шкалы
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	188,3629000	18070,4000000
2/5-{222.5048000, 290.7886000}	1/10-{10711.4000000, 15617.4000000}	256,6467000	13164,4000000
3/5-{290.7886000, 359.0724000}	9/10-{49959.4000000, 54865.4000000}	324,9305000	52412,4000000
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	393,2143000	57318,4000000
5/5-{427.3562000, 495.6400000}	4/10-{25429.4000000, 30335.4000000}	461,4981000	27882,4000000

Таблица 7 – База данных «Inf1-Y(X)-Pos-All_points-1-1.dbf» для визуализации когнитивных функций по всем наблюдениям с весами, равными количеству информации в наблюдениях (пример, когда макс. количество информации отражено 10 точками)

Наименование градации описательной шкалы	Наименование градации классификационной шкалы	Градация опис.шкалы	Градация класс.шкалы	№ точки	Кол-во Информации (бит)
1/5-{154.2210000, 222.5048000}	1/10-{10711.4000000, 15617.4000000}	181,3450917	13836,3479983	1	0,3555752
1/5-{154.2210000, 222.5048000}	1/10-{10711.4000000, 15617.4000000}	184,1479638	11120,4317504	2	0,3555752
1/5-{154.2210000, 222.5048000}	1/10-{10711.4000000, 15617.4000000}	214,2460175	13164,3999991	3	0,3555752
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	194,2899689	19021,1910145	1	0,5120035
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	177,4300036	18394,0312272	2	0,5120035
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	186,5090552	17059,1910253	3	0,5120035
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	190,7028171	16524,8359564	4	0,5120035
1/5-{154.2210000, 222.5048000}	2/10-{15617.4000000, 20523.4000000}	215,2246262	18070,3999972	5	0,5120035
2/5-{222.5048000, 290.7886000}	1/10-{10711.4000000, 15617.4000000}	263,1092915	15879,1332606	1	0,4982368
2/5-{222.5048000, 290.7886000}	1/10-{10711.4000000, 15617.4000000}	239,6950459	15189,3525096	2	0,4982368
2/5-{222.5048000, 290.7886000}	1/10-{10711.4000000, 15617.4000000}	238,4616221	12770,1180014	3	0,4982368
2/5-{222.5048000, 290.7886000}	1/10-{10711.4000000, 15617.4000000}	261,9259384	11842,9811896	4	0,4982368
2/5-{222.5048000, 290.7886000}	1/10-{10711.4000000, 15617.4000000}	266,1218665	13164,3999952	5	0,4982368
2/5-{222.5048000, 290.7886000}	2/10-{15617.4000000, 20523.4000000}	249,5484524	19604,3972972	1	0,2777635
2/5-{222.5048000, 290.7886000}	2/10-{15617.4000000, 20523.4000000}	248,3602013	15067,0561877	2	0,2777635
2/5-{222.5048000, 290.7886000}	2/10-{15617.4000000, 20523.4000000}	282,0905272	18070,3999955	3	0,2777635
2/5-{222.5048000, 290.7886000}	8/10-{45053.4000000, 49959.4000000}	244,8125363	50085,6627054	1	0,2777635
2/5-{222.5048000, 290.7886000}	8/10-{45053.4000000, 49959.4000000}	246,2892018	45281,3162081	2	0,2777635
2/5-{222.5048000, 290.7886000}	8/10-{45053.4000000, 49959.4000000}	267,2780252	47506,3999981	3	0,2777635
3/5-{290.7886000, 359.0724000}	3/10-{20523.4000000, 25429.4000000}	330,3995391	22976,3999957	1	0,1335549
3/5-{290.7886000, 359.0724000}	5/10-{30335.4000000, 35241.4000000}	336,2903636	32788,3999976	1	0,0862421
3/5-{290.7886000, 359.0724000}	6/10-{35241.4000000, 40147.4000000}	311,5579867	38177,8804718	1	0,3067154
3/5-{290.7886000, 359.0724000}	6/10-{35241.4000000, 40147.4000000}	322,0267662	34029,3776279	2	0,3067154
3/5-{290.7886000, 359.0724000}	6/10-{35241.4000000, 40147.4000000}	345,0736688	37694,3999979	3	0,3067154
3/5-{290.7886000, 359.0724000}	7/10-{40147.4000000, 45053.4000000}	303,6228369	42600,4000020	1	0,2426704
3/5-{290.7886000, 359.0724000}	7/10-{40147.4000000, 45053.4000000}	351,0653550	42600,3999999	2	0,2426704
3/5-{290.7886000, 359.0724000}	9/10-{49959.4000000, 54865.4000000}	324,9305000	55754,6396374	1	0,4631437
3/5-{290.7886000, 359.0724000}	9/10-{49959.4000000, 54865.4000000}	318,2936322	52412,4000015	2	0,4631437
3/5-{290.7886000, 359.0724000}	9/10-{49959.4000000, 54865.4000000}	324,9305000	48191,0420545	3	0,4631437
3/5-{290.7886000, 359.0724000}	9/10-{49959.4000000, 54865.4000000}	333,6471197	52412,3999997	4	0,4631437
4/5-{359.0724000, 427.3562000}	4/10-{25429.4000000, 30335.4000000}	387,1796851	31381,3608947	1	0,3625915
4/5-{359.0724000, 427.3562000}	4/10-{25429.4000000, 30335.4000000}	385,3991840	27427,9773700	2	0,3625915
4/5-{359.0724000, 427.3562000}	4/10-{25429.4000000, 30335.4000000}	408,7770130	27882,3999987	3	0,3625915
4/5-{359.0724000, 427.3562000}	5/10-{30335.4000000, 35241.4000000}	397,3995679	33782,5078177	1	0,5104565
4/5-{359.0724000, 427.3562000}	5/10-{30335.4000000, 35241.4000000}	382,6745617	33842,3004182	2	0,5104565
4/5-{359.0724000, 427.3562000}	5/10-{30335.4000000, 35241.4000000}	373,9232543	31680,3234310	3	0,5104565
4/5-{359.0724000, 427.3562000}	5/10-{30335.4000000, 35241.4000000}	396,2378521	31184,3967189	4	0,5104565
4/5-{359.0724000, 427.3562000}	5/10-{30335.4000000, 35241.4000000}	394,3929650	32788,3999992	5	0,5104565
4/5-{359.0724000, 427.3562000}	8/10-{45053.4000000, 49959.4000000}	407,4816485	47506,3999964	1	0,0695099
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	395,3594729	58358,6771743	1	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	399,7423185	58346,4692167	2	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	390,3208291	58284,1852927	3	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	386,6324428	58108,7857112	4	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	368,9816470	57318,4000001	5	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	378,8259286	56283,9047843	6	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	387,5061798	54242,0098802	7	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	401,1296802	56338,7738351	8	1,0437864
4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	411,7426326	54808,1590574	9	1,0437864

4/5-{359.0724000, 427.3562000}	10/10-{54865.4000000, 59771.4000000}	414,4819519	57318,3999978	10	1,0437864
5/5-{427.3562000, 495.6400000}	3/10-{20523.4000000, 25429.4000000}	452,2921759	26071,3518690	1	0,2899832
5/5-{427.3562000, 495.6400000}	3/10-{20523.4000000, 25429.4000000}	459,3299898	21906,1611866	2	0,2899832
5/5-{427.3562000, 495.6400000}	3/10-{20523.4000000, 25429.4000000}	465,1941402	22976,3999955	3	0,2899832
5/5-{427.3562000, 495.6400000}	4/10-{25429.4000000, 30335.4000000}	467,8300951	28755,7846561	1	0,5357519
5/5-{427.3562000, 495.6400000}	4/10-{25429.4000000, 30335.4000000}	458,4054364	28646,9649134	2	0,5357519
5/5-{427.3562000, 495.6400000}	4/10-{25429.4000000, 30335.4000000}	443,7581067	27314,1840382	3	0,5357519
5/5-{427.3562000, 495.6400000}	4/10-{25429.4000000, 30335.4000000}	466,2508313	24877,3928943	4	0,5357519
5/5-{427.3562000, 495.6400000}	4/10-{25429.4000000, 30335.4000000}	466,8937851	27882,3999948	5	0,5357519
5/5-{427.3562000, 495.6400000}	6/10-{35241.4000000, 40147.4000000}	461,4981000	39595,7426313	1	0,4631437
5/5-{427.3562000, 495.6400000}	6/10-{35241.4000000, 40147.4000000}	440,8689611	37694,4000022	2	0,4631437
5/5-{427.3562000, 495.6400000}	6/10-{35241.4000000, 40147.4000000}	461,4981000	33375,7515244	3	0,4631437
5/5-{427.3562000, 495.6400000}	6/10-{35241.4000000, 40147.4000000}	472,2978060	37694,3999977	4	0,4631437
5/5-{427.3562000, 495.6400000}	7/10-{40147.4000000, 45053.4000000}	461,4981000	44621,2671468	1	0,3990987
5/5-{427.3562000, 495.6400000}	7/10-{40147.4000000, 45053.4000000}	455,0021052	42600,4000026	2	0,3990987
5/5-{427.3562000, 495.6400000}	7/10-{40147.4000000, 45053.4000000}	461,4981000	39932,4565020	3	0,3990987
5/5-{427.3562000, 495.6400000}	7/10-{40147.4000000, 45053.4000000}	468,1049593	42600,3999976	4	0,3990987

Стандартными средствами MS Excel на основе таблиц 6 и 7 построены регрессии, изображенные на рисунке 22.

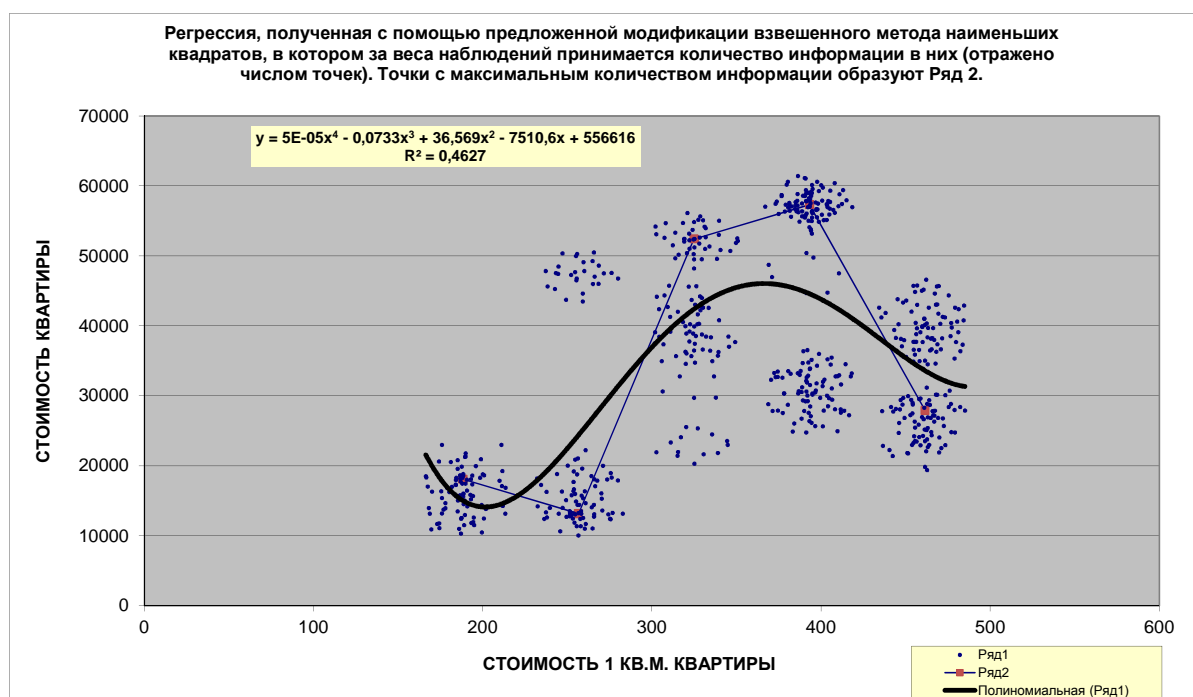


Рисунок 22 – Регрессия, построенная на основе всех наблюдений с учетом количества информации в них (в ряду 2 показаны также наблюдения с максимальным количеством информации)

На рисунке 22 число точек, на которых строится регрессия, значительно превосходит число параметров, т.к. каждая точка модели, соответствующая наблюдению, представляется в форме ряда точек, количество которых соответствует количеству информации в этом наблюдении.

Сравнивая когнитивные функции зависимости стоимости квартиры от стоимости 1 кв.метра, приведенные на рисунке 15, с аппроксимацией на рисунке 22 мы видим, что они совпадают. Это и не удивительно, т.е. так и должно быть, т.к. они построены на основе одной и той же модели знаний.

Но здесь важно не только это, но и то, что *режим 4.6 позволяет привлечь для построения и исследования когнитивных функций в виде регрессий весь хорошо*

разработанный аппарат регрессионного анализа, в том числе и аппарат оценки качества регрессий с помощью дисперсий и доверительных интервалов.

Программная реализация режима подготовки баз данных для визуализации когнитивных функций в MS Excel (режим 4.6) системы «Эйдос» и исходный текст всей системы «Эйдос» приведена по ссылке: <ftp://lc.kubagro.ru/Downloads.exe>. Для того, чтобы в исходном тексте системы «Эйдос», т.е. в файле _AIDOS-X.doc, найти исходный текст программы описанного в данном разделе режима необходимо найти в этом файле контекст: «N F4_6()».

Ниже приведен **алгоритм режима** подготовки баз данных для визуализации когнитивных функций в MS Excel, приведен ниже. Но с целью экономии места сделано это не в традиционной форме блок-схемы, а в более компактной форме списка шагов.

Шаг-0. Вход режима подготовки баз данных для визуализации когнитивных функций в MS Excel.

Шаг-1. Определение массивов и переменных, используемых в режиме.

Шаг-2. Выйти из режима, если нет авторизации в системе.

Шаг-3. Если не запущен режим, работающий с БД, то перейти в папку выбранного приложения

Шаг-4. Создать папку для баз данных визуализации когнитивных функций "Cogn_fun" в папке текущего приложения, если ее не было.

Шаг-5. Проверить созданы ли в папке текущего приложения базы данных статистических моделей и моделей знаний: {Abs.dbf, Prc1.dbf, Prc2.dbf, Inf1.dbf, Inf2.dbf, Inf3.dbf, Inf4.dbf, Inf5.dbf, Inf6.dbf, Inf7.dbf}. Если их нет, то выдать сообщение о том, что для того, чтобы создать их необходимо выполнить режим 3.4 или 3.5 и выйти из режима, а иначе продолжить.

Шаг-6. Проверить, существует ли файл с заданием на создание баз данных для визуализации когнитивных функций. Если он существует, загрузить его и присвоить значения из него массиву параметров диалога. Если же не существует – то присвоить значения по умолчанию массиву параметров и записать его в виде файла.

Шаг-7. Организовать экранную форму для задания параметров создания баз данных для визуализации когнитивных функций с параметрами из массива с параметрами.

Шаг-8. Проверить, задана ли хотя бы одна стат.модель или модель знаний для создания БД для визуализации КФ. Если нет – выдать сообщение и выйти, иначе продолжить.

Шаг-9. Записать файл с информацией о параметрах создания БД для визуализации КФ.

Шаг-10. Удалить все dbf-файлов из папки: "Cogn_fun».

Шаг-11. Открыть базы данных классификационных и описательных шкал и градаций.

Шаг-12. Определить максимальную длину наименования градации описательной шкалы.

Шаг-13. Определить максимальную длину наименования градации классификационной шкалы.

Шаг-14. Занести в БД описательных и классификационных шкал информацию о начальной и конечной градации каждой шкалы

Шаг-15. Открыть все базы данных статистических моделей и моделей знаний: {Abs.dbf, Prc1.dbf, Prc2.dbf, Inf1.dbf, Inf2.dbf, Inf3.dbf, Inf4.dbf, Inf5.dbf, Inf6.dbf, Inf7.dbf}.

Шаг-16. Определение число операций, необходимых для создания БД для визуализации КФ. Это необходимо для отображения стадии исполнения режима.

Шаг-17. Организовать отображение стадии исполнения режима.

Шаг-18. Начало цикла по моделям: {Abs.dbf, Prc1.dbf, Prc2.dbf, Inf1.dbf, Inf2.dbf, Inf3.dbf, Inf4.dbf, Inf5.dbf, Inf6.dbf, Inf7.dbf}.

Шаг-19. Создавать КФ по данной модели? Если да, то на следующий шаг, а иначе – на конец цикла по моделям.

Шаг-20. Создавать КФ-1: прямые, позитивные, построенные ТОЛЬКО по точкам с максимальным количеством информации? Если да, то на следующий шаг, иначе на проверку создания других видов КФ (Шаг-39).

Шаг-21. Начало цикла по подматрицам текущей модели.

Шаг-22. Начало цикла по классификационным шкалам.

Шаг-23. Определить диапазон градаций текущей классификационной шкалы.

Шаг-24. Начало цикла по описательным шкалам.

Шаг-25. Создать БД для визуализации КФ с нужным именем и открыть ее.

Шаг-26. Определить диапазон градаций текущей описательной шкалы.

Шаг-27. Начало цикла по градациям описательной шкалы текущей модели.

Шаг-28. Для каждой градации описательной шкалы найти градацию классификационной шкалы с Max информативностью и занести их в БД КФ.

Шаг-29. Если градация найдена, то на следующий шаг, а иначе на проверку, создавать ли следующий вид когнитивных функций (Шаг-35).

Шаг-30. Извлечь наименование градации описательной шкалы

Шаг-31. Если описательная шкала числовая, то посчитать среднее значение числового интервала градации, а иначе значением градации считать ее код.

Шаг-32. Если классификационная шкала числовая, то посчитать среднее значение числового интервала градации, а иначе значением градации считать ее код.

Шаг-33. Записать в БД визуализации КФ новую запись с именами градаций описательной и классификационной шкал и значениями этих градаций.

Шаг-34. Конец проверки на наличие градации (Шаг-29).

Шаг-35. Конец цикла по градациям описательной шкалы текущей модели (Шаг-27).

Шаг-36. Закрыть БД визуализации КФ.

Шаг-37. Конец цикла по описательным шкалам (Шаг-24).

Шаг-38. Конец цикла по классификационным шкалам (Шаг-22).

Шаг-39. Конец проверки на создание 1-го вида когнитивных функций (Шаг-20).

Шаг-40. Создавать КФ-2: прямые, позитивные, построение по ВСЕМ точкам с весами, равными количеству информации в них? Если да, то на следующий шаг, иначе на проверку создания других видов КФ (Шаг-60).

Шаг-41. Начало цикла по классификационным шкалам текущей модели.

Шаг-42. Определить диапазон градаций текущей классификационной шкалы.

Шаг-43. Начало цикла по описательным шкалам.

Шаг-44. Создать БД для визуализации КФ с нужным именем и открыть ее.

Шаг-45. Определить диапазон градаций текущей описательной шкалы.

Шаг 46. Найти максимальную и минимальную информативность в подматрице БД INF# и использовать ее для расчета весового коэффициента и определения количества точек с единичным весом в единице информации для $lij > 0$. Заодно определить диапазоны изменения значений градаций классификационных и описательных шкал и градаций для подматрицы функции.

Для каждой градации описательной шкалы найти все градации классификационной шкалы и для каждой из них занести в БД визуализации КФ количество точек единичного веса, соответствующее количеству информации в значении аргумента о значении функции.

Шаг-47. Начало цикла по градациям описательной шкалы текущей модели.

Шаг-48. Начало цикла по градациям классификационной шкалы текущей модели.

Шаг-49. Извлечь из БД текущей модели количество информации в текущей градации описательной шкалы о текущей градации классификационной шкалы.

Шаг-50. Если это количество информации положительное, то перейти на следующий шаг, а иначе – на проверку следующего элемента матрицы текущей модели (Шаг-56).

Шаг-51. Определить диапазон градаций текущей описательной шкалы.

Шаг-52. Определить диапазон градаций текущей классификационной шкалы.

Шаг-53. Посчитать количество точек, соответствующее количеству информации в градации.

Шаг-54. Посчитать угол в градусах между соседними точками рассеяния.

Шаг-55. Занести в БД визуализации КФ количество точек единичного веса, соответствующее количеству информации в значении аргумента о значении функции (для каждой точки создать запись в БД).

Шаг-56. Конец проверки на положительность количества информации в элементе матрицы модели (Шаг-50).

Шаг-57. Конец цикла по градациям классификационной шкалы текущей модели (Шаг-48).

Шаг-58. Конец цикла по градациям описательной шкалы текущей модели (Шаг-47).

Шаг-59. Закрыть БД визуализации КФ.

Шаг-60. Конец цикла по описательным шкалам (Шаг-43).

Шаг-61. Конец цикла по классификационным шкалам (Шаг-41).

Шаг-62. Конец проверки на создание 2-го вида когнитивных функций (Шаг-40).

* * * *

Остальные 6 видов когнитивных функций, классифицированные в таблице 5, рассчитываются аналогично КФ-1 и КФ-2 с небольшими изменениями в алгоритмах их расчета по сравнению с приведенными выше.

Шаг-63. Конец проверки на расчет БД для данной модели.

Шаг-64. Конец цикла по моделям.

Шаг-65. Закрытие структуры отображения стадии исполнения.

Шаг-66. Закрытие всех баз данных.

Шаг-67. Отображение окна с информацией об окончании работы режима.

Шаг-68. Выход из режима подготовки БД для визуализации КФ.

Конец алгоритма режима 4.6 системы «Эйдос».

7.2.2.6. Выводы

Метод наименьших квадратов (МНК) широко известен и пользуется заслуженной популярностью. Вместе с тем не прекращаются попытки усовершенствования этого метода. Результатом одной из таких попыток является взвешенный метод наименьших квадратов (ВМНК), суть которого в том, чтобы придать наблюдениям вес обратно пропорциональный погрешностям их аппроксимации. Этим самым фактически наблюдения игнорируются тем в большей степени, чем сложнее их аппроксимировать. В результате такого подхода формально погрешность аппроксимации снижается, но фактически это происходит путем частичного отказа от рассмотрения «проблемных» наблюдений, вносящих большую ошибку. Если эту идею, лежащую в основе ВМНК довести до крайности (и тем самым до абсурда), то в пределе такой подход приведет к тому, что из всей совокупности наблюдений останутся только те, которые практически точно ложатся на тренд, полученный методом наименьших квадратов, а остальные просто будут проигнорированы. Однако, по мнению автора, фактически это не решение проблемы, а отказ от ее решения, хотя внешне и выглядит как решение. В работе предлагается

именно решение, основанное на теории информации: считать весом наблюдения количество информации в аргументе о значении функции. Этот подход был обоснован в рамках нового инновационного метода искусственного интеллекта: метода автоматизированного системно-когнитивного анализа (АСК-анализа) и реализован еще 30 лет назад в его программном инструментарии – интеллектуальной системе «Эйдос» в виде так называемых «когнитивных функций». В данном разделе приводится алгоритм и программная реализация данного подхода, проиллюстрированные на подробном численном примере.

Таким образом, автоматизированный системно-когнитивный анализ (АСК-анализ) и его математическая модель (системная теория информации), а также реализующий их программный инструментарий АСК-анализа – система «Эйдос» – это и есть ответы на этот вопрос. Таким образом, АСК-анализ и система «Эйдос» представляют собой современную инновационную (готовую к внедрению) технологию взвешенного метода наименьших квадратов, модифицированного путем применения в качестве весов наблюдений количества информации в них.

Данный раздел может быть использован как описание лабораторной работы по дисциплинам:

- Интеллектуальные системы;
- Инженерия знаний и интеллектуальные системы;
- Интеллектуальные технологии и представление знаний;
- Представление знаний в интеллектуальных системах;
- Основы интеллектуальных систем;
- Введение в нейроматематику и методы нейронных сетей;
- Основы искусственного интеллекта;
- Интеллектуальные технологии в науке и образовании;
- Управление знаниями;
- Автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос»;

которые автор ведет в настоящее время¹⁰⁴, а также и в других дисциплинах, связанных с преобразованием данных в информацию, а ее в знания и применением этих знаний для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области (а это практически все дисциплины во всех областях науки).

7.2.2.7. ОГРАНИЧЕНИЯ И ПЕРСПЕКТИВЫ

В данном разделе не ставилась задача описать математический метод АСК-анализа, обеспечивающий расчет количества информации в наблюдениях, т.к. этому посвящено много монографий и статей автора, размещенных в полном открытом бесплатном доступе:

- <http://lc.kubagro.ru/>;
- <http://lc.kubagro.ru/aidos/index.htm>;
- <http://ej.kubagro.ru/a/viewaut.asp?id=11>;
- <http://www.twirpx.com/user/858406/>;
- http://elibrary.ru/author_items.asp?authorid=123162.

В будущем планируется дать развернутое математическое обоснование метода взвешенных наименьших квадратов, модифицированного путем применения в качестве весов наблюдений количества информации в них и применения теории информации

¹⁰⁴ http://lc.kubagro.ru/My_training_schedule.doc

для расчета этих весовых коэффициентов наблюдений, а также исследовать свойства данной модификации метода взвешенных наименьших квадратов.

7.3. Асимптотический информационный критерий качества шума

Интуитивно все понимают, что шум, это сигнал, в котором нет информации или в котором на практике не удастся выявить информацию. Точнее, понятно, что некая последовательность элементов (ряд) тем в большей степени является шумом, чем меньше информации содержится в значениях одних элементов о значениях других. Тем более странно, что никто не предложил не только способа, но даже идеи измерения количества информации в одних фрагментах сигнала о других его фрагментах и его использования в качестве критерия оценки степени близости данного сигнала к шуму. Авторами предложен асимптотический информационный критерий качества шума, а также метод, технология и методика его применения на практике. В качестве метода применения асимптотического информационного критерия качества шума на практике предлагается автоматизированный системно-когнитивный анализ (АСК-анализ), в качестве технологии – программный инструментарий АСК-анализа: универсальная когнитивная аналитическая система «Эйдос», в качестве методика – методика создания приложений в данной системе, а также их использования для решения задач идентификации, прогнозирования, принятия решений и исследования предметной области путем исследования ее модели. Приводится наглядный численный пример, иллюстрирующий излагаемые идеи и подтверждающий работоспособность предлагаемого асимптотического информационного критерия качества шума, а также метода, технологии и методика его применения на практике

*«В начале было Слово»
Евангелие от Иоанна*

*«...законы природы являются лишь высказываниями о
пространственно-временных совпадениях...»
Альберт Эйнштейн*

7.3.1. Формулировка проблемы

Данный раздел является продолжением серии работ авторов, посвященных системной нечеткой интервальной математике [1, 2, 3] и применению теории информации для решения задач математической статистики [4, 5, 6], в частности анализа текстов и рядов объектов числовой и нечисловой природы (слов, чисел, символов, цифр).

Шум есть везде или выражаясь более точным языком математики «почти везде». Любой сигнал, получаемый нами, может рассматриваться как сумма истинного сигнала и шума. Понятие шума является одним из основополагающих понятий в теории связи, в которой решается важнейшая задача подавления шума и повышения отношения сигнал/шум [7, 8, 9]. Но смысл понятия «Шум» гораздо шире, что будет видно из следующего обсуждения. Поэтому очень важно уметь исследовать шум, выявлять его, идентифицировать тип шума, **оценивать качество шума**, подавлять (отфильтровывать) шум, генерировать шум с заранее заданными характеристиками и качеством и т.д.

В теоретическом исследовании и практических решениях всех этих вопросов, связанных с шумом, большую роль играют физические и численные эксперименты с шумом. В численных экспериментах в качестве источников шума используются различные генераторы псевдослучайных последовательностей, основанные на различных

алгоритмах. На использовании этих генераторов основано целое научное направление: «Метод статистических испытаний или статистического моделирования Монте-Карло»¹⁰⁵ [10].

Понятно, что результаты статистического моделирования напрямую зависят от качества используемых генераторов шума. Вопросам исследования шума посвящено огромное количество научных работ. Надо отметить, что используются разные терминологические системы. При рассмотрении пары сигнал/шум говорят о выделении (или оценке) сигнала. Если есть зависимость от времени, то обсуждают временные ряды (когда время дискретно) и случайные процессы (когда время непрерывно). Решают задачи выделения тренда [49], оценки периода [50] и др.

Однако, количественное *измерение* качества шума остается недостаточно исследованной **проблемой**, которую необходимо решать и теоретически, и практически. Данная статья посвящена поиску подходов к решению этой проблемы с применением методов, основных на *теории информации*.

7.3.2. Идея решения проблемы – Асимптотический информационный критерий качества шума

В 60-х годах XX века А.Н.Колмогоров связывал случайность с алгоритмизацией. Он считал, что последовательность чисел является случайной, если ее нельзя задать с помощью алгоритма, заметно более короткого по сравнению с длиной самой последовательности [44 - 48]. Очевидно, все генераторы случайных чисел задаются сравнительно короткими алгоритмами, а потому по А.Н.Колмогорову получить с их помощью действительно случайную последовательность невозможно. По этой причине будем использовать термин «псевдослучайный» и для самих генераторов, и для получаемых с их помощью последовательностей.

Теоретически возможность создания эффективных алгоритмов генерации псевдослучайных чисел обычно обосновывается с помощью теорем теории чисел. Но с появлением в распоряжении исследователей мощных компьютеров возрастает роль и численных экспериментов в исследованиях шума.

В 1985 г. известным журналом «Заводская лаборатория. Диагностика материалов»¹⁰⁶ была развернута научная дискуссия по поводу статистических свойств генерации псевдослучайных последовательностей. Все началось с того, что И.Г.Журбенко (МГУ им. М.В. Ломоносова) обнаружил, что в рядах, полученных с помощью распространенного в те годы генератора псевдослучайных последовательностей, *три последовательных значения довольно точно связаны линейной зависимостью*. Такие аномалии для конкретных датчиков обнаруживают и сейчас¹⁰⁷ Но что значит «Связаны»?

Интуитивно все понимают, что шум, это сигнал, в котором нет информации или в котором на практике не удастся выявить информацию. Точнее, понятно, что некая последовательность элементов (ряд) тем в большей степени является шумом, чем меньше информации содержится в значениях одних элементов о значениях других. Тем более странно, что никто не предложил не только способа, но даже лежащей на поверхности идеи измерения количества информации в одних фрагментах сигнала о других его фрагментах и использования этого количества информации в качестве критерия

¹⁰⁵ Датой рождения метода Монте-Карло принято считать 1949 г., когда появилась статья под названием «The Monte Carlo method». Создателями этого метода считают американских математиков Дж. Фон Неймана и С. Улама.

¹⁰⁶ <http://www.zldm.ru/>

¹⁰⁷ См., например, <http://forum.orlovs.pp.ru/viewtopic.php?f=1&t=426&p=9206#p9206>

оценки степени близости данного сигнала к шуму. При этом сходные подходы к приятию решений хорошо известны [7, 8, 9].

Авторы предлагают *асимптотический информационный критерий качества шума, представляющий собой вариабельность количества информации в значениях одних элементов последовательности (ряда) о значениях других его элементов.*

Отметим, что в работе [17] еще в 2002 году на стр. 290¹⁰⁸ одним из авторов было предложено использовать аналогичный критерий в качестве *количественной меры степени выраженности закономерностей в предметной области.* написано: «Из этого следует возможность использования в качестве *количественной меры степени выраженности закономерностей в предметной области* использовать не матрицу абсолютных частот и меру X^2 , а новую меру, основанную на *матрице информативностей и системном обобщении формулы Харкевича для количества информации:*

$$H = \frac{1}{\sqrt{(A \cdot W - I)}} \sum_{j=1}^W \sum_{i=1}^A (I_{ij} - \bar{I})^2 \quad (3.81)$$

где:

$$\bar{I} = \frac{1}{A \cdot W} \sum_{j=1}^W \sum_{i=1}^A I_{ij} \quad \text{— средняя информативность признаков по матрице информативностей.}$$

Значение данной меры показывает среднее отличие количества информации в факторах о будущих состояниях активного объекта управления от среднего количества информации в факторе (которое при больших выборках близко к 0). По своей математической форме эта мера сходна с мерами для значимости факторов и степени сформированности образов классов и коррелирует с объемом пространства классов и пространства атрибутов». В данной же статье предлагается *количественную меру степени выраженности закономерностей в предметной области использовать в качестве критерия близости этой предметной области к шуму.*

Данный критерий является *асимптотическим*, т.к. результаты измерения с помощью этого критерия, по-видимому, должны сходиться к истинному значению при увеличении количества исследуемых элементов последовательности. Математические формулировки этого утверждения о состоятельности критерия будут обсуждаться в дальнейшем.

Количество информации может вычисляться для различных элементов последовательности: например в значении каждого элемента о значении последующего элемента, в паре элементов о паре последующих, в тройке значений последовательных элементов о значении последующего и т.п., и т.д.

Количество информации может рассчитываться с помощью различных количественных мер измерения информации: Найквиста, Хартли, Больцмана, Шеннона, Харкевича, алгоритмических подходов к измерению информации и др.

Вариабельность может рассчитываться с помощью различных мер вариабельности: среднего модуля отклонения от среднего, среднеквадратичного отклонения и др.

Поэтому существует много различных вариантов применения предложенного критерия.

¹⁰⁸ См, например: <http://elibrary.ru/download/62451150.pdf>

7.3.3. Обсуждение понятия «Шум» и теоретическое обоснование Асимптотического информационного критерия качества шума

7.3.3.1. СООТНОШЕНИЕ СОДЕРЖАНИЯ ПОНЯТИЙ: «ШУМ», «ДАННЫЕ», «ИНФОРМАЦИЯ» И «ЗНАНИЯ»

Как мы видели выше, понятие «Шум» тесно связано с понятием «Информация», точнее с отсутствием информации в сигнале или невозможностью ее извлечения из сигнала. Но как связано содержание понятий: «Данные», «Информация», «Знания»? В рассмотрении этого вопроса будем основываться на статье [11]¹⁰⁹.

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ) [12, 17], для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на объект управления к каким его изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона [13], состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных» (этот термин используется и в иных смыслах, например, как синоним термина "прикладная статистика" [41, 42]), которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал [15, 41] и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу). По сути, этот этап является *нормализацией* базы исходных данных.

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения показателей) [1].

Какие же математические меры могут быть использованы для количественного измерения силы и направления *причинно-следственных* зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно,

109 См., например: «Подборка публикаций по вопросам выявления, представления и использования знаний»: <http://www.twirpx.com/file/793311/>

что это совершенно не так (пассивный эксперимент дает возможности выявить связи, но не причины). Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу [15, 17].

Знания – это информация, полезная для достижения целей.

Значит, для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).
2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные. Отметим, что это делается, в частности, при SWOT и PEST анализе [16].

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно последовательно повышать степень формализации исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач управления, принятия решений и исследования предметной области.

Таким образом, понятие «Шум» по своему содержанию наиболее близко к понятию «Данные». Для того, чтобы выяснить являются ли данные просто шумом или содержат информацию, нужно выявить в них причинно-следственные зависимости.

7.3.3.2. ПРОЦЕДУРЫ ПРЕОБРАЗОВАНИЯ ДАННЫХ В ИНФОРМАЦИЮ, А ЕЕ В ЗНАНИЯ

Процедуры преобразования данных в информацию, а ее в знания, реализованные в системе «Эйдос», приведены на рисунке 2:

О соотношении содержания понятий: «Данные», «Информация» и «Знания»

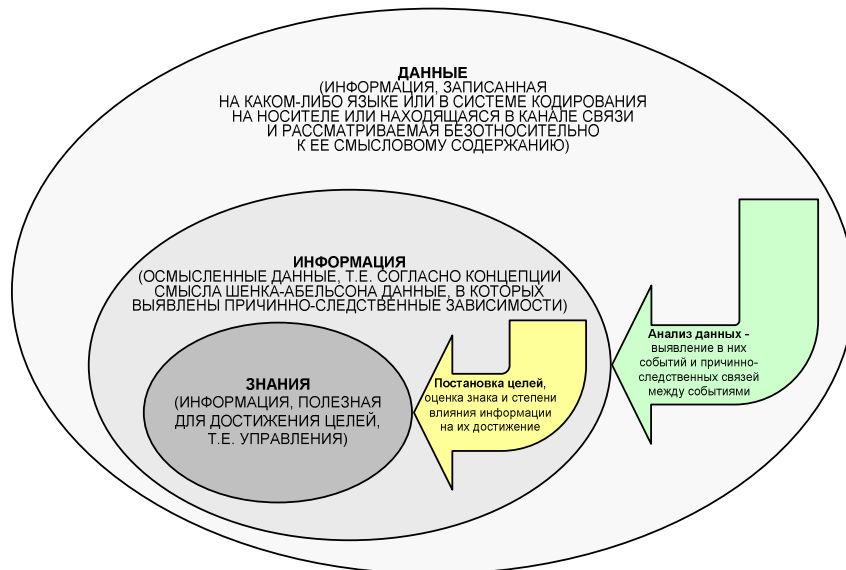


Рисунок 2 –Процедуры преобразования данных в информацию, а ее в знания

Отметим в этой связи известное высказывание Альберта Эйнштейна, приведенное в качестве эпиграфа к статье: «...законы природы являются лишь высказываниями о пространственно-временных совпадениях...» [43]. Учитывая вышесказанное, можно сказать, что законы природы отражают реально существующие причинно-следственные зависимости, т.е. содержат информацию о них. Отметим также, что расчет этого количества информации основывается на матрице абсолютных частот, т.е. на предварительном определении абсолютного количества этих совпадений (фактов), о которых говорит Альберт Эйнштейн. Фактом в АСК-анализе является совпадение действия на моделируемый объект определенного значения фактора и перехода этого объекта в определенное состояние.

7.3.3.3. Шум как данные, которые не удастся преобразовать в информацию имеющимися средствами (абракадабра)

Шум можно определить, как сигнал, в котором нет закономерностей. Но как доказать что их нет? Возможно ли даже в принципе доказать, что чего-то нет? Здесь необходимо вспомнить о критерии Поппера и принципе Эшби [14]. По мнению авторов, это невозможно даже в принципе. На практике возможно доказать лишь, что с помощью имеющихся в нашем распоряжении методов обнаружения закономерностей их выявить не удалось. Это позволяет провести различие между абстрактным теоретическим понятием шума и понятием «практически шума». Реально мы всегда исследуем лишь практически шум. В определение шума входит не только характеристика самого сигнала (отсутствие закономерностей), но и характеристика нас самих, точнее наших возможностей обнаружения закономерностей в этом сигнале. А они, во-первых, ограничены, во-вторых, изменяются от места к месту (доступ к вычислительным ресурсам и средствам обнаружения закономерностей), в третьих, изменяются со временем (вычислительные технологии быстро эволюционируют). Поэтому то, что еще вчера считалось где-то шумом, сегодня где-то уже им не будет признаваться.

Если в результате применения процедур выявления причинно-следственных закономерностей в данных, например, реализованных в системе «Эйдос», не удастся выявить эти закономерности, т.е. не удастся преобразовать эти данные в информацию, то можно говорить о том, что эти данные являются «практически шумом», т.е. на данном

этапе развития *для нас* неотличимы от шума. Вопрос о том, являются ли «на самом деле» эти данные шумом, имеет скорее философско-методологический характер [14].

7.3.3.4. СООБЩЕНИЕ, КАК СИСТЕМА, СМЫСЛ КАК ЭМЕРДЖЕНТНОЕ СВОЙСТВО СИСТЕМ, ШУМ КАК ДЕСТРУКТИРОВАННОЕ СООБЩЕНИЕ, Т.Е. СООБЩЕНИЕ, В КОТОРОМ УТРАЧЕН СМЫСЛ

Все свойства систем имеют эмерджентную природу [1, 37]. Не является исключением и свойство текстов иметь смысл. Любое сообщение на естественном языке является *системой* символов некоторого алфавита, образующих иерархическую систему с многими уровнями иерархии (например: слова, предложения, абзацы, параграфы, главы, книги), и между элементами всех этих уровней существует множество горизонтальных и вертикальных взаимосвязей, в результате чего у этой системы появляется новое эмерджентное свойство, отсутствовавшее у элементов: *смысл*. *Смысл – это эмерджентное свойство символических систем*. Уровень системной организации, количественно измеряемый предложенным автором [1, 2, 27]¹¹⁰ коэффициентом эмерджентности Хартли, своего рода «плотность смысла на символ» у стихов выше, чем у прозы, а у песен, еще выше, чем у стихов.

С этой точки зрения *шум представляет собой бессмысленное сообщение*. Но как на практике сообщение может стать бессмысленным? Это оказывается возможным, если нарушить или разрушить его внутреннюю иерархическую структуру и взаимосвязи элементов в этой структуре. Нечто подобное происходит с древними текстами, подвергшимся разрушительному действию факторов окружающей среды в течение длительного времени. В качестве других примеров можно привести костры из книг, а также действие уничтожителя бумаги. Когда люди не хотят, чтобы содержимое записки стало кому-либо известным, они просто разрывают ее на мелкие кусочки и кидают их в разные урны.

7.3.3.5. ШУМ КАК АРХИВ, КОТОРЫЙ УЖЕ НЕ УДАЕТСЯ ЗААРХИВИРОВАТЬ ИМЕЮЩИМИСЯ СРЕДСТВАМИ

Если в данных выявлены закономерности, то их можно использовать для сжатия данных, т.е. создания архива. При этом если в архиве обнаружены закономерности, то его можно еще сжать, но уже в меньшей степени. Чем лучше степень сжатия в архиве (чем лучше архиватор), тем меньше в нем можно *обнаружить* закономерностей, т.е. тем больше он становится похожим на шум, а *наилучший архив вообще не отличим от шума*. Идеальный архив вообще невозможно сжать, как и шум, в котором вообще нет закономерностей. При архивировании плотность записи информации на символ увеличивается.

Так может быть шум, – это не бессмысленный сигнал, т.е. сигнал, в котором нет информации, а наоборот, сигнал, с наивысшей в принципе возможной плотностью записью информации?

Это значит, что *архиваторы можно считать генераторами шума*.

По-видимому, можно доказать теорему: *при итерационном применении архиватора к архиву этот архив сходится к шуму*.

В этом подходе, в отличие от подхода А.Н. Колмогорова [44 - 48], шум создается не просто программой, но программой, использующей внешние данные, причем дан-

¹¹⁰ Есть много попыток прямого плагиата (см., например: Виктор Вяткин. Групповой плагиат: от студента до министра. <http://trv-science.ru/2011/11/08/gruppovojj-plagiat-ot-studenta-do-ministra/>) и просто непонимания того, что этот коэффициент эмерджентности предложен не Хартли, а Е.В.Луценко в работе [17].

ные, даже возможно, содержащие закономерности. Необходимо отметить, что эти исходные данные могут быть весьма велики по объему, что делает шум более качественным, несмотря на то, что алгоритм работы программы может быть очень коротким. В этом важное отличие нашего подхода от подхода А.Н. Колмогорова.

Во многих языках программирования (например, на Паскале) для улучшения псевдослучайного сигнала перед запуском функции, возвращающей случайное число, можно задать некоторое числовое значение, изменяющее ее работу. Ясно, что желательно, чтобы и само это значение тоже не было постоянным, т.к. иначе работа генератора псевдослучайных чисел будет изменяться одинаково. Поэтому часто в качестве такого значения часто используется как-либо функция от текущего значения компьютерного таймера.

7.3.3.6. Шум как зашифрованный информационный сигнал, который не удается расшифровать имеющимися средствами. Ключ дешифрования как способ извлечения смысла

Чтение, можно рассматривать как извлечение информации из книги, т.е. дешифрованием записанного в книге сигнала. Познание представляет собой извлечение информации из объекта познания. Если провести аналогию между объектом познания и книгой, то можно считать, что познание представляет собой чтение объекта познания, пользуясь термином К. Маркса, его распрямление. С этой точки зрения вся природа представляет собой зашифрованное послание, а познание представляет собой нечто иное как чтение книги природы. Научный метод с этой точки зрения представляет собой проверенный и доказавший свою высокую эффективность ключ дешифрования книги природы, даже может быть своего рода отмычку, позволяющую вскрыть «тайну за семью печатями». Конечно, при такой точке зрения возникает закономерный вопрос об авторстве книги природы. Здесь ничего не приходит на ум, кроме первых слов Евангелия от Иоанна, приведенных в качестве эпиграфа к статье: «В начале было Слово» и далее по тексту. Слово, или выражаясь современным языком – информация, преобразует первозданный Хаос во Вселенную, полную чудес, информация способна структурировать бессмысленный набор символов в гениальное высокоорганизованное литературное произведение [39].

Но человек не только может читать книгу природы, он может и вносить в нее некоторые не очень большие правки и дополнения. Труд представляет собой процессе записи информации, содержащейся в субъективном образе будущего продукта труда, в предмет труда [40], выражаясь термином К. Маркса: опредмечивание. С этой точки зрения можно рассматривать труд и его результат – антропоморфное общество, как внесение поправок и дополнений в книгу природы.

Шум – это текст на неизвестном языке. Чтобы его расшифровать – надо перевести его на известный язык.

В принципе не взламываемый шифр – это шифр, в котором каждый символ встречается лишь один раз и нет никаких «пространственно-временных совпадений», хотя бы в принципе позволяющих выявить смысл. Примером такого шифра является замена каждого символа исходного сообщения на номер этого символа в некотором очень большом тексте (можно псевдослучайном), при этом каждый символ из большого текста используется лишь один раз или не используется ни разу.

Таким образом, *системы шифрования можно рассматривать как генераторы шума, и чем сложнее взломать шифр, тем ближе зашифрованный сигнал к идеальному шуму.*

Так может быть шум, – это не бессмысленный сигнал, т.е. сигнал, в котором нет информации, а наоборот, сигнал, с очень важной информацией, зашифрованный очень стойким шифром?

По-видимому, можно доказать теорему: *при итерационном применении шифрования (одного метода или различных методов в определенном порядке) к уже зашифрованному сигналу результат шифрования сходится к шуму*. Для получения математических утверждений нужно тем или иным способом дать строгие определения понятиям "шум" и "шифрование".

7.3.3.7. Шум и Хаос в Древнегреческой и Древнеиндийской космогонии

По-видимому, в момент большого взрыва (если принять эту распространенную космогоническую модель, предложенную иезуитом Жоржем Леметром в 1927 г.) мир был менее структурирован, чем сейчас, и имел более низкий, чем сейчас, уровень системности [1].

Индусы в своих учениях говорили, что дифференцированная вселенная периодически возникает и опять и переходит в не проявленное состояние (день и ночь Браммы). Это очень напоминает архивирование (или шифрование) и разархивирование (дешифрование) и опять архивирование и т.д. Дух дифференцирует материю, затем материя одухотворяется (круг Сансары). Из хаоса, утверждали древние греки, Вселенная родилась, и в Хаос же возвратится. Этому же учил Пифагор, об этом же глубокомысленно и красноречиво молчит Дао. Вселенная периодически становится более доступной для познания, как если бы Изида под нашим пристальным взором иногда немного приоткрывала свое покрывало, скрывающее ее прекрасное (как говорят) лицо. Можно утверждать, что сходные космогонические концепции пронизывают все древние мировоззренческие системы и, похоже, в современной науке получают еще одну интерпретацию, которая в чем-то глубже, а в чем-то более поверхностна, чем древние¹¹¹.

7.3.4. Метод, технология и методика применения асимптотического информационного критерия качества шума на практике

7.3.4.1. АСК-анализ как метод применения асимптотического информационного критерия качества шума

Системный анализ представляет собой современный метод научного познания, общепризнанный метод решения проблем. Однако возможности практического применения системного анализа ограничиваются отсутствием развернутого программного инструментария, обеспечивающего его автоматизацию. Существуют программные системы, автоматизирующие отдельные этапы или функции системного анализа в конкретных предметных областях. Автоматизированный системно-когнитивный анализ (АСК-анализ) представляет собой системный анализ, структурированный по базовым когнитивным операциям (БКО), благодаря чему удалось разработать для него математическую модель, методику численных расчетов (структуры данных и алгоритмы их обработки), а также реализующую их программную систему – систему Эйдос [17, 18]. Система Эйдос разработана в постановке, не зависящей от предметной области, и имеет ряд программных интерфейсов с внешними данными различных типов. АСК-анализ может быть применен как инструмент, многократно усиливающий возможности естественного интеллекта во всех областях, где используется естественный интеллект. АСК-анализ был успешно применен для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта путем исследования его модели во многих предметных областях, в частности в экономике, технике, социо-

¹¹¹ Сегодня изложение этих учений легко найти в Internet

логии, педагогике, психологии, медицине, экологии, ампелографии (см. [14, 17, 31] и др.).

7.3.4.1.1. Истоки

Известно, что системный анализ является одним из общепризнанных в науке методов решения проблем и многими учеными рассматривается вообще как метод научного познания. Однако как заметил еще в 1984 году проф. И. П. Стабин [19], на практике применение системного анализа наталкивается на проблему. Суть этой проблемы в том, что обычно системный анализ успешно применяется в сравнительно простых случаях, в которых в принципе можно обойтись и без него, тогда как в действительно сложных ситуациях, когда он действительно чрезвычайно востребован и у него нет альтернатив, сделать это удается гораздо реже. Проф. И. П. Стабин предложил и путь решения этой проблемы, который он видел в автоматизации системного анализа [19].

Путь от идеи до создания программной системы включает ряд этапов:

- выбрать теоретический математический метод;
- разработать методику численных расчетов, включающую структуры данных в оперативной памяти и внешних баз данных (даталогическую и инфологическую модели) и алгоритмы обработки этих данных;
- разработать программную систему, реализующую эти математические методы и методики численных расчетов.

7.3.4.1.2. Методика

Предпосылки решения проблемы

Перегудов Ф. И. и Тарасенко Ф. П. в своих основополагающих работах 1989 и 1997 годов [20, 21] подробно рассмотрели математические методы, которые в принципе могли бы быть применены для автоматизации отдельных этапов системного анализа. Однако даже самые лучшие математические методы не могут быть применены на практике без реализующих их программных систем, а путь от математического метода к программной системе долог и сложен. Для этого необходимо разработать численные методы или методики численных расчетов (алгоритмы и структуры данных), реализующие математический метод, а затем разработать программную реализацию системы, основанной на этом численном методе.

В числе первых попыток реальной автоматизации системного анализа следует отметить докторскую диссертацию проф. Симанкова В. С. (2001) [22]. Эта попытка была основана на высокой детализации этапов системного анализа и подборе уже существующих программных систем, автоматизирующих эти этапы. Идея была в том, что чем выше детализация системного анализа, тем мельче этапы, тем проще их автоматизировать. Эта попытка была реализована, однако, лишь для специального случая исследования в области возобновляемой энергетики, т.к. объединяемые системы оказались различных разработчиков, созданные с помощью различного инструментария и не имеющие программных интерфейсов друг с другом, т.е. не образующие единой автоматизированной системы. Эта попытка, безусловно, явилась большим шагом по пути, предложенному проф. И. П. Стабиным, но и ее нельзя признать обеспечившей достижение поставленной цели, сформулированной Стабиным И.П. (т.е. создание автоматизированного системного анализа), т.к. она не привела к созданию единой универсальной программной системы, автоматизирующей системный анализ, которую можно было бы применять в различных предметных областях. Парадоксально, но эта попытка автоматизации системного анализа была несистемна.

Необходимо отметить работы Дж. Клира по системологии и автоматизации решения системных задач [23], которые внесли большой вклад в автоматизацию систем-

ного анализа путем создания и применения универсального решателя системных задач (УРСЗ), реализованного в рамках оригинальной экспертной системы. Обсуждение проблем развития системного анализа продолжается (см., например, [51]).

АСК-анализ, как решение проблемы

Автоматизированный системно-когнитивный анализ разработан профессором Е. В. Луценко и предложен в 2002 году [12, 17, 18]. Основная идея, позволившая сделать это, состоит в рассмотрении системного анализа как метода познания (отсюда и «когнитивный» от «*cognitio*» – знание, познание, лат.). Это позволило структурировать системный анализ не по этапам, как пытались сделать ранее, а по базовым когнитивным операциям системного анализа (БКОСА), т.е. таким операциям, к комбинациям которых сводятся остальные. Эти операции образуют минимальную систему, достаточную для описания системного анализа, как метода познания, т.е. когнитивный конфигуратор (Лефевр В.А., 1962) [24] и их оказалось не очень много, всего 10:

- 1) присвоение имен;
- 2) восприятие (описание конкретных объектов в форме отнологий, т.е. их признаками и принадлежностью к обобщающим категориям - классам);
- 3) обобщение (синтез, индукция);
- 4) абстрагирование;
- 5) оценка адекватности модели;
- 6) сравнение, идентификация и прогнозирование;
- 7) дедукция и абдукция;
- 8) классификация и генерация конструкторов;
- 9) содержательное сравнение;
- 10) планирование и поддержка принятия управленческих решений.

Каждая из этих операций оказалась достаточно элементарна для формализации и программной реализации.

Компоненты АСК-анализа [12, 17, 18]:

- формализуемая когнитивная концепция и следующий из нее когнитивный конфигуратор;
- теоретические основы, методология, технология и методика АСК-анализа;
- математическая модель АСК-анализа, основанная на системном обобщении теории информации;
- методика численных расчетов, в универсальной форме реализующая математическую модель АСК-анализа, включающая иерархическую структуру данных и 24 детальных алгоритма 10 БКОСА;
- специальное инструментальное программное обеспечение, реализующее математическую модель и численный метод АСК-анализа – Универсальная когнитивная аналитическая система "Эйдос".

Этапы АСК-анализа:

- 1) когнитивная структуризация предметной области;
- 2) формализация предметной области (конструирование классификационных и описательных шкал и градаций и подготовка обучающей выборки);
- 3) синтез системы моделей предметной области (в настоящее время система Эйдос поддерживает 3 статистические модели и 7 моделей знаний);
- 4) верификация (оценка достоверности) системы моделей предметной области;
- 5) повышение качества системы моделей;
- 6) решение задач идентификации, прогнозирования и поддержки принятия решений;
- 7) исследование моделируемого объекта путем исследования его моделей: кластерно-конструктивный анализ классов и факторов; содержательное сравнение классов

и факторов; изучение системы детерминации состояний моделируемого объекта, нелокальные нейроны и интерпретируемые нейронные сети прямого счета; построение классических когнитивных моделей (когнитивных карт); построение интегральных когнитивных моделей (интегральных когнитивных карт).

В АСК-анализе все факторы рассматриваются с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта, на который они действуют, в определенное состояние, и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации. Это напоминает подход Дугласа Хаббарда [25], но, в отличие от него, имеет универсальный программный инструментарий, разработанный в постановке, не зависящей от предметной области и находящийся в полном открытом бесплатном доступе (даже с исходными текстами) на сайте автора: <http://lc.kubagro.ru/aidos/Aidos-X.htm>. Поэтому АСК-анализ обеспечивает корректную сопоставимую обработку числовых и нечисловых данных, представленных в разных типах измерительных шкал и разных единицах измерения. Метод АСК-анализа является устойчивым непараметрическим методом, обеспечивающим создание моделей больших размерностей при неполных и зашумленных исходных данных о сложном нелинейном динамичном объекте управления. Этот метод является чуть ли не единственным, обеспечивающим многопараметрическую типизацию и системную идентификацию методов, инструментарий которого (интеллектуальная система Эйдос) находится в полном открытом бесплатном доступе.

В развитии различных теоретических и практических аспектов АСК-анализа приняли участие многие ученые: Луценко Е.В.¹¹², Лойко В.И., Трунев А.П. (Канада)¹¹³, Орлов А.И.¹¹⁴, Коржаков В.Е., Барановская Т.П., Ермоленко В.В., Наприев И.Л., Некрасов С.Д., Лаптев В.Н., Третьяк В.Г., Шукин Т.Н., Симанков В.С., Ткачев А.Н., Сафронова Т.И., Макаревич О.А., Макаревич Л.О., Сергеева Е.В. (Фомина Е.В.), Бандык Д.К., Артемов А.А., Крохмаль В.В., Рябцев В.Г. и другие.

7.3.4.1.3. Результаты

Метод системно-когнитивного анализа и его программный инструментарий интеллектуальная система "Эйдос" были успешно применены при проведении научных исследований, по результатам которых защищено довольно много докторских и кандидатских диссертаций в различных направлениях науки: 3 доктора экономических наук (+1 в стадии подтверждения в ВАК РФ, +1 в стадии подготовки к защите), 2 доктора технических наук, 4 кандидата психологических наук, 1 кандидат технических наук, 1 кандидат экономических наук, 1 кандидат медицинских наук:

АСК-анализ был успешно применены при выполнении десятков грантов РФФИ и РГНФ различной направленности за длительный период с 2002 года по настоящее время (2016 год).

По проблематике АСК-анализа издана 21 монография, получено 28 патентов на системы искусственного интеллекта, их подсистемы, режимы и приложения, опубликовано 196 статей в изданиях, входящих в Перечень ВАК РФ (по данным РИНЦ). В одном только Научном журнале КубГАУ (входит в Перечень ВАК РФ с 26-го марта 2010 года) автором АСК-анализа Луценко Е.В. опубликовано 175 статей по различным тео-

¹¹² <http://lc.kubagro.ru/>

¹¹³ <http://chaosandcorrelation.org/>

¹¹⁴ <http://orlovs.pp.ru/>

ретическим и практическим аспектам АСК-анализа, общим объемом 297,246 у.п.л., в среднем 1,699 у.п.л. на одну статью¹¹⁵.

По этим публикациям, грантам и диссертационным работам видно, что АСК-анализ уже был успешно применен в следующих предметных областях и научных направлениях: экономика (региональная, отраслевая, предприятий, прогнозирование фондовых рынков), социология, эконометрика, биометрия, педагогика (создание педагогических измерительных инструментов и их применение), психология (личности, экстремальных ситуаций, профессиональных и учебных достижений, разработка и применение профессиограмм), сельское хозяйство (прогнозирование результатов применения агротехнологий, принятие решений по выбору рациональных агротехнологий и микроразнообразия), экология, ампелография, геофизика (глобальное и локальное прогнозирование землетрясений, параметров магнитного поля Земли, движения полюсов Земли), климатология (прогнозирование Эль-Ниньо и Ла-Нинья), возобновляемая энергетика, мелиорация и управление мелиоративными системами, и ряд других областей.

АСК-анализ вызывает большой интерес во всем мире. Сайт автора АСК-анализа посетило около 500 тыс. посетителей с уникальными IP-адресами со всего мира¹¹⁶. Еще около 500 тыс. посетителей (в расчете на фамилию автора) открывали статьи по АСК-анализу в Научном журнале КубГАУ¹¹⁷.

Все это позволяет говорить о том, что АСК-анализ представляет собой современную инновационную технологию искусственного интеллекта и постепенно превращается в новое междисциплинарное научное направление.

7.3.4.2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДИКА ЧИСЛЕННЫХ РАСЧЕТОВ (АЛГОРИТМЫ И СТРУКТУРЫ ДАННЫХ) АСИМПТОТИЧЕСКОГО ИНФОРМАЦИОННОГО КРИТЕРИЯ КАЧЕСТВА ШУМА

В этом качестве и для этой цели предлагается применить математическую модель и методику численных расчетов (алгоритмы и структуры данных) АСК-анализа, которая подробно описана в ряде работ [1, 2, 14, 17] и других. Суть этой математической модели состоит в том, что сначала рассчитывается матрица абсолютных частот, отражающая «пространственно-временные совпадения», т.е. факты, содержащиеся в исходных данных, а затем на основе нее рассчитываются матрицы условных и безусловных процентных распределений (относительные частоты), матрица информативностей и другие модели знаний [15].

Методика численных расчетов включает структуры внешних баз данных и данных в оперативной памяти, а также алгоритмы их обработки, реализующие математическую модель [17]. Основными алгоритмами являются алгоритмы, реализующие базовые когнитивные операции системного анализа (БКОСА).

7.3.4.3. СИСТЕМА «ЭЙДОС», КАК ТЕХНОЛОГИЯ И МЕТОДИКА ПРИМЕНЕНИЯ АСИМПТОТИЧЕСКОГО ИНФОРМАЦИОННОГО КРИТЕРИЯ КАЧЕСТВА ШУМА

Система «Эйдос» является программным *инструментарием* АСК-анализа и подробно описана в ряде работ [18]¹¹⁸, поэтому здесь приведем лишь графическую схему преобразования данных в информацию, а ее в знания и решения задач идентифика-

¹¹⁵ <http://ej.kubagro.ru/a/viewaut.asp?id=11>

¹¹⁶ <http://lc.kubagro.ru/>

¹¹⁷ <http://ej.kubagro.ru/a/viewaut.asp?id=11&order=viewed>

¹¹⁸ См., например: <http://lc.kubagro.ru/aidos/Aidos-X.htm> (список работ в нижней части страницы)

ции, прогнозирования, принятия решений и исследования моделируемого объекта путем исследования его модели в системе «Эйдос» (рисунок 2):

Последовательность обработки данных, информации и знаний в системе «Эйдос-Х++»

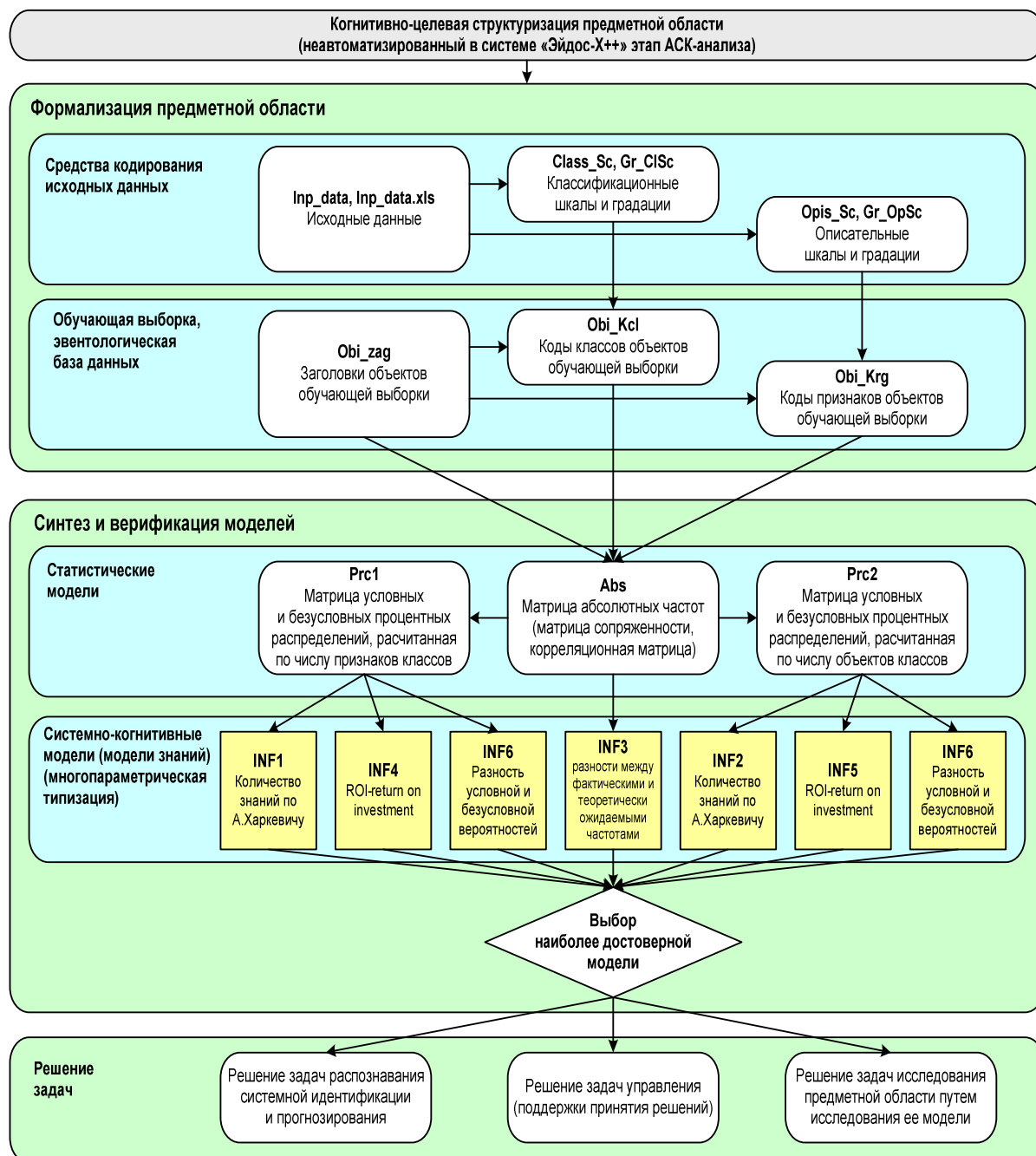


Рисунок 2 – Схема преобразования данных в информацию, а ее в знания и решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта путем исследования его модели в системе «Эйдос»

7.3.5. Описание численного эксперимента и его результатов

7.3.5.1. ПРОГРАММА ПОДГОТОВКИ ИСХОДНЫХ ДАННЫХ

При работе над данной статьей авторами разработана специальная программа, скриншоты окна которой при выборе различных пунктов приведены на рисунке 3:

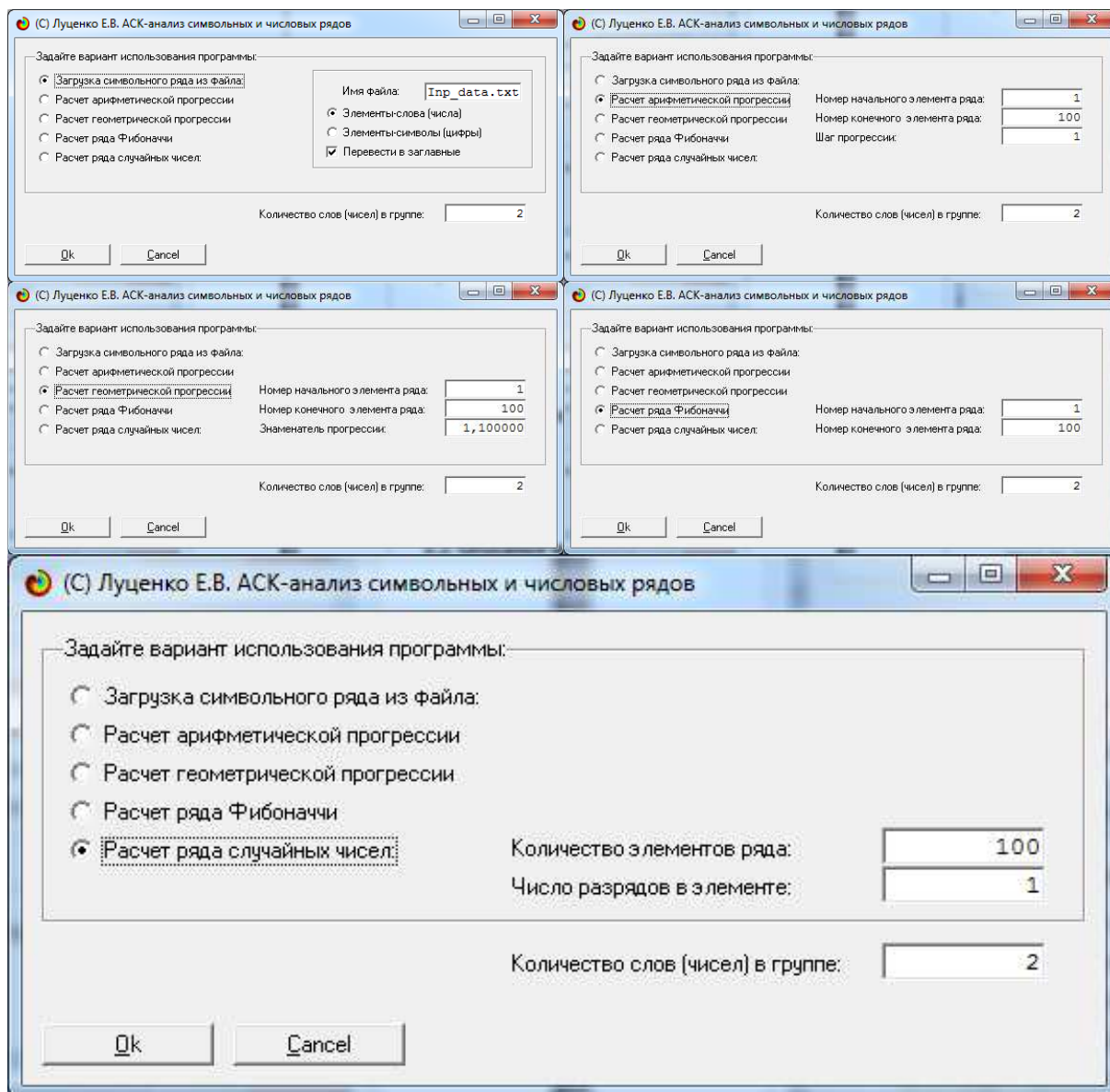


Рисунок 3 – Скриншоты окна программы подготовки исходных данных

Эта программа предназначена для подготовки исходных данных для системы «Эйдос» и обеспечивает

- ввод из внешнего DOS-TXT-файла числовой или символьной последовательности с заданными параметрами (в качестве элементов последовательности можно рассматривать числа-слова или цифра-символы, можно переводить или не переводить символы в верхний регистр) и преобразования ее в форму базы данных, непосредственно воспринимаемой одним из программных интерфейсов системы «Эйдос» с внешними базами данных (стандарт этой базы данных описан на рисунке 4);
- генерации известных неслучайных последовательностей: арифметической и геометрической прогрессий и ряда Фибоначчи с заданными параметрами;
- генерации псевдослучайных последовательностей с использованием стандартного генератора использованного языка программирования (xBase++).

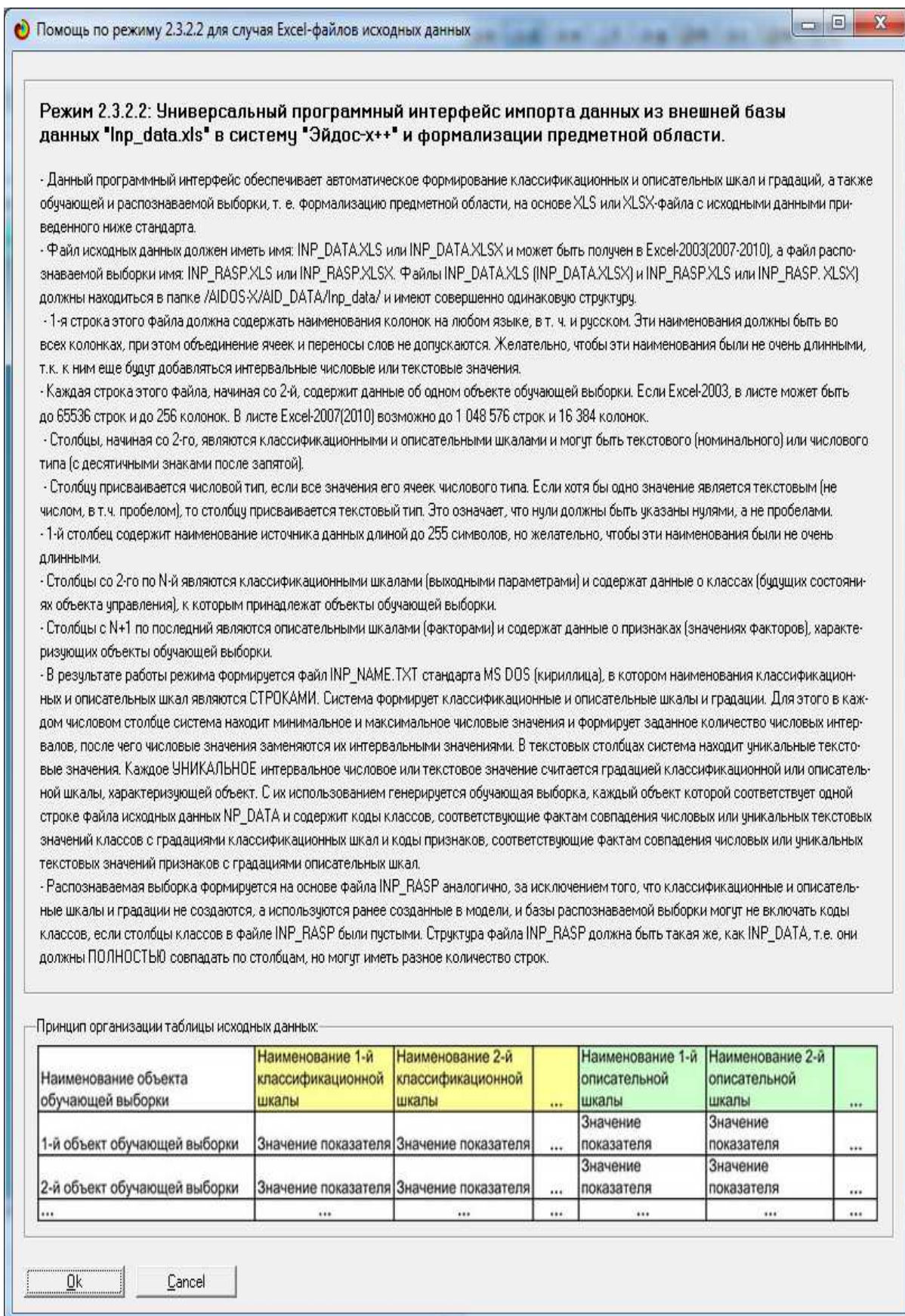


Рисунок 4 – Описание стандарта базы данных, непосредственно воспринимаемой одним из программных интерфейсов системы «Эйдос» с внешними базами данных

Исходный текст этой программы приведен ниже:

```
*****
FUNCTION Main()

LOCAL  GetList[0], GetOptions, nColor, oMessageBox, oMenuWords, oDlg

DC_IconDefault(1000)

SET DECIMALS TO 15
SET DATE GERMAN
SET ESCAPE On

SET COLLATION TO SYSTEM // Руссификация
*SET COLLATION TO ASCII // Руссификация

PUBLIC aSay[30], Mess97, Mess98, Mess99 // Массив сообщений отображаемых стадий исполнения (до 30 на экране)
PUBLIC Time_progress, Wsego, oProgress, lOk
PUBLIC nEvery := 100 // Количество корректировок прогресс-бар

*****
g = 0
s = 0
mRegim = 1
@g , 0 DCGROUP oGroup1 CAPTION 'Задайте вариант использования программы:' SIZE 78.0, 7.5
@++s, 2 DCRADIO mRegim VALUE 1 PROMPT 'Загрузка символического ряда из файла:' PARENT oGroup1
@++s, 2 DCRADIO mRegim VALUE 2 PROMPT 'Расчет арифметической прогрессии' PARENT oGroup1
@++s, 2 DCRADIO mRegim VALUE 3 PROMPT 'Расчет геометрической прогрессии' PARENT oGroup1
@++s, 2 DCRADIO mRegim VALUE 4 PROMPT 'Расчет ряда Фибоначчи' PARENT oGroup1
@++s, 2 DCRADIO mRegim VALUE 5 PROMPT 'Расчет ряда случайных чисел:' PARENT oGroup1

P1 = 45
P2 = 60

// Загрузка символического ряда из файла

s = 1
cFile = 'Inp_data.txt'

nElement = 1
mUpper = .T.
@0.7,43 DCGROUP oGroup2 CAPTION '' SIZE 33, 5.5
@ 1+0.1, 4.5 DCSAY "Имя файла:" EDITPROTECT {||.NOT.mRegim=1} HIDE {||.NOT.mRegim=1} PARENT oGroup1
@ 1,15 DCSAY "" GET cFile PICTURE "XXXXXXXXXXXX" EDITPROTECT {||.NOT.mRegim=1} HIDE {||.NOT.mRegim=1} PARENT oGroup2
@ 2, 2 DCRADIO nElement VALUE 1 PROMPT 'Элементы-слова (числа)' EDITPROTECT {||.NOT.mRegim=1} HIDE {||.NOT.mRegim=1} PARENT oGroup2
@ 3, 2 DCRADIO nElement VALUE 2 PROMPT 'Элементы-символы (цифры)' EDITPROTECT {||.NOT.mRegim=1} HIDE {||.NOT.mRegim=1} PARENT oGroup2
@ 4, 2 DCCHECKBOX mUpper PROMPT 'Перевести в заглавные' EDITPROTECT {||.NOT.mRegim=1} HIDE {||.NOT.mRegim=1} PARENT oGroup2

P1 = 35
```

P2 = 61

// Расчет арифметической прогрессии

s = 2

N1 = 1

N2 = N1+99

D = 1

```
@ s+0.2, P1 DCSAY "Номер начального элемента ряда:" EDITPROTECT { || .NOT.mRegim=2 } HIDE { || .NOT.mRegim=2 } PARENT oGroup1
@ s , P2 DCSAY "" GET N1 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=2 } HIDE { || .NOT.mRegim=2 } PARENT oGroup1
@ ++s+0.2, P1 DCSAY "Номер конечного элемента ряда:" EDITPROTECT { || .NOT.mRegim=2 } HIDE { || .NOT.mRegim=2 } PARENT oGroup1
@ s , P2 DCSAY "" GET N2 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=2 } HIDE { || .NOT.mRegim=2 } PARENT oGroup1
@ ++s+0.2, P1 DCSAY "Шаг прогрессии:" EDITPROTECT { || .NOT.mRegim=2 } HIDE { || .NOT.mRegim=2 } PARENT oGroup1
@ s , P2 DCSAY "" GET D PICTURE "#####" EDITPROTECT { || .NOT.mRegim=2 } HIDE { || .NOT.mRegim=2 } PARENT oGroup1
```

// Расчет геометрической прогрессии

s = 3

N1 = 1

N2 = N1+99

Q = 1.1

```
@ s+0.2, P1 DCSAY "Номер начального элемента ряда:" EDITPROTECT { || .NOT.mRegim=3 } HIDE { || .NOT.mRegim=3 } PARENT oGroup1
@ s , P2 DCSAY "" GET N1 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=3 } HIDE { || .NOT.mRegim=3 } PARENT oGroup1
@ ++s+0.2, P1 DCSAY "Номер конечного элемента ряда:" EDITPROTECT { || .NOT.mRegim=3 } HIDE { || .NOT.mRegim=3 } PARENT oGroup1
@ s , P2 DCSAY "" GET N2 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=3 } HIDE { || .NOT.mRegim=3 } PARENT oGroup1
@ ++s+0.2, P1 DCSAY "Знаменатель прогрессии:" EDITPROTECT { || .NOT.mRegim=3 } HIDE { || .NOT.mRegim=3 } PARENT oGroup1
@ s , P2 DCSAY "" GET Q PICTURE "###.#####" EDITPROTECT { || .NOT.mRegim=3 } HIDE { || .NOT.mRegim=3 } PARENT oGroup1
```

// Расчет ряда Фибоначчи

s = 4

N1 = 1

N2 = N1+99

```
@ s+0.2, P1 DCSAY "Номер начального элемента ряда:" EDITPROTECT { || .NOT.mRegim=4 } HIDE { || .NOT.mRegim=4 } PARENT oGroup1
@ s , P2 DCSAY "" GET N1 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=4 } HIDE { || .NOT.mRegim=4 } PARENT oGroup1
@ ++s+0.2, P1 DCSAY "Номер конечного элемента ряда:" EDITPROTECT { || .NOT.mRegim=4 } HIDE { || .NOT.mRegim=4 } PARENT oGroup1
@ s , P2 DCSAY "" GET N2 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=4 } HIDE { || .NOT.mRegim=4 } PARENT oGroup1
```

// Расчет ряда случайных чисел с равномерным распределением

s = 5

R1 = 100

R2 = 1

```
@ s+0.2, P1 DCSAY "Количество элементов ряда:" EDITPROTECT { || .NOT.mRegim=5 } HIDE { || .NOT.mRegim=5 } PARENT oGroup1
@ s , P2 DCSAY "" GET R1 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=5 } HIDE { || .NOT.mRegim=5 } PARENT oGroup1
@ ++s+0.2, P1 DCSAY "Число разрядов в элементе:" EDITPROTECT { || .NOT.mRegim=5 } HIDE { || .NOT.mRegim=5 } PARENT oGroup1
@ s , P2 DCSAY "" GET R2 PICTURE "#####" EDITPROTECT { || .NOT.mRegim=5 } HIDE { || .NOT.mRegim=5 } PARENT oGroup1
```

P1 = 35

P2 = 61

s = 8.0

mGroup = 2

```
@ s+0.2,P1 DCSAY "Количество слов (чисел) в группе:"
```



```

@ s      ,P2 DCSAY "" GET mGroup PICTURE "#####"

DCGETOPTIONS TABSTOP
DCREAD GUI ;
  TO lExit ;
  FIT ;
  OPTIONS GetOptions ;
  ADDBUTTONS;
  MODAL ;
  TITLE '(С) Луценко Е.В. АСК-анализ символьных и числовых рядов'

*****
IF lExit
  ** Button Ok
ELSE
  QUIT
ENDIF
*****

*****
*****

T_Mess1 = "Начало: "+TIME()           // Начало
Sec_1   = (DOY(DATE())-1)*86400+SECONDS()

IF mRegim = 5 // Расчет ряда случайных чисел (с равномерным распределением)
  N1 = 1
  N2 = R1
ENDIF

nMax = N2 - N1 + 1
Mess = 'АСК-анализ рядов. Генерация ряда'
@ 4,5 DCPROGRESS oProgress SIZE 70,1.1 MAXCOUNT nMax COLOR GRA_CLR_CYAN PERCENT EVERY 100
DCREAD GUI TITLE Mess PARENT @oDialog FIT EXIT
oDialog:show()
nTime = 0
DC_GetProgress(oProgress,0,nMax)

***** формирование текстовой переменной с символами *****

mInpData := "" // Текстовая переменная для загрузки текстового файла

DO CASE
  CASE mRegim = 1 // Загрузка символьного ряда из файла:

    IF .NOT. FILE(cFile)
      Mess = 'В текущей папке нет файла: "'
      Mess = STRTRAN(Mess, "#", cFile)
      LB_Warning(Mess)
    
```

```

CLOSE ALL
RETURN NIL
ELSE
mInpData = CharOne(' ', FILESTR(cFile)) // Загрузка cFile
IF mUpper // Перевести в заглавные
    mInpData = UPPER(mInpData)
ENDIF
IF nElement = 2 // Элементы - символы (цифры)
    mInpData2 = ""
    FOR j=1 TO LEN(mInpData)
        mInpData2 = mInpData2 + SUBSTR(mInpData, j, 1) + " "
    NEXT
    mInpData = CharOne(' ', mInpData2) // Удалить подряд идущие пробелы
ENDIF
mOptions = 'Загрузка символического ряда из файла: "#". Количество слов (чисел) в группе: @"'
mOptions = STRTRAN(mOptions, "#", cFile)
mOptions = STRTRAN(mOptions, "@", ALLTRIM(STR(mGroup)))
ENDIF

CASE mRegim = 2 // Расчет арифметической прогрессии

FOR n = N1 TO N2
    Xn = ROUND(N1+D*(n-1), 0)
    mInpData = mInpData + ALLTRIM(STR(Xn)) + " " // Текстовая переменная для загрузки текстового файла
    DC_GetProgress(oProgress, ++nTime, nMax)
NEXT
mOptions = 'Расчет элементов арифметической прогрессии от: "#" до "@" с шагом "D".'
mOptions = STRTRAN(mOptions, "#", ALLTRIM(STR(N1)))
mOptions = STRTRAN(mOptions, "@", ALLTRIM(STR(N2)))
mOptions = STRTRAN(mOptions, "D", ALLTRIM(STR(D)))

CASE mRegim = 3 // Расчет геометрической прогрессии

FOR n = N1 TO N2
    Xn = ROUND(N1*Q^(n-1), 0)
    mInpData = mInpData + ALLTRIM(STR(Xn)) + " " // Текстовая переменная для загрузки текстового файла
    DC_GetProgress(oProgress, ++nTime, nMax)
NEXT
mOptions = 'Расчет элементов геометрической прогрессии от: "#" до "@" со знаменателем "Q".'
mOptions = STRTRAN(mOptions, "#", ALLTRIM(STR(N1)))
mOptions = STRTRAN(mOptions, "@", ALLTRIM(STR(N2)))
mOptions = STRTRAN(mOptions, "Q", ALLTRIM(STR(Q)))

CASE mRegim = 4 // Расчет ряда Фибоначчи

FOR n = N1 TO N2
    SQRT5 = SQRT(5)
    Xn = 1/SQRT5*((1+SQRT5)/2)^n-1/SQRT5*((1-SQRT5)/2)^n
    Xn = ROUND(Xn, 0)
    mInpData = mInpData + ALLTRIM(STR(Xn)) + " " // Текстовая переменная для загрузки текстового файла

```

```

        DC_GetProgress(oProgress, ++nTime, nMax)
    NEXT
    mOptions = 'Расчет элементов ряда Фибоначчи от: "#" до "@".'
    mOptions = STRTRAN(mOptions, "#", ALLTRIM(STR(N1)))
    mOptions = STRTRAN(mOptions, "@", ALLTRIM(STR(N2)))

CASE mRegim = 5 // Расчет ряда случайных чисел (с равномерным рапределением)

    N1 = 1
    N2 = R1

    FOR j = N1 TO N2
        Xn = SUBSTR(ALLTRIM(STR(RANDOM())),1,R2) // Генерация 5-разрядного псевдослучайного числа, преобразование го в текстовую
        форму и получение старшего разряда
        mInpData = mInpData + ALLTRIM(Xn) + " " // Текстовая переменная для загрузки текстового файла
        DC_GetProgress(oProgress, ++nTime, nMax)
    NEXT
    mOptions = 'Расчет # элементов ряда $-разрядных случайных чисел (с равномерным рапределением).'
    mOptions = STRTRAN(mOptions, "#", ALLTRIM(STR(R1)))
    mOptions = STRTRAN(mOptions, "$", ALLTRIM(STR(R2)))

ENDCASE

STRFILE(mOptions, 'Options.txt')
STRFILE(mInpData, 'Inp_data.txt')

*MsgBox('STOP')
DC_GetProgress(oProgress,nMax,nMax)
oDialog:Destroy()

***** Формирование БД Inp_data.dbf на основе текстовой переменной ****

**** Создание БД Inp_data.dbf
CLOSE ALL
CrLf = CHR(13)+CHR(10) // Конец строки (записи)
mInpName := "" // TXT-переменная с наименованиями полей

aStructure := { { "ObjName", "C", 250, 0 }, ;
                { "Futur" , "C", 250, 0 }, ;
                { "Retro" , "C", 250, 0 } }
DbCreate( "Inp_data.dbf", aStructure )

mInpName = mInpName + "Futur" + CrLf + "Retro" + CrLf
STRFILE(mInpName, "Inp_name.txt")

CLOSE ALL
USE Inp_data EXCLUSIVE NEW

IF NUMTOKEN(mInpData," ") >= mGroup + 1

```

```

nMax = NUMTOKEN(mInpData," ") - mGroup - 1

Mess = 'ACK-анализ рядов. Формирование БД "Inp_data.dbf"'
@ 4,5 DCPROGRESS oProgress2 SIZE 70,1.1 MAXCOUNT nMax COLOR GRA_CLR_CYAN PERCENT EVERY 100
DCREAD GUI TITLE Mess PARENT @oDialog2 FIT EXIT
oDialog2:show()
nTime = 0
DC_GetProgress(oProgress,0,nMax)

*** Начало цикла по словам *****

FOR t=1 TO NUMTOKEN(mInpData," ") - mGroup - 1 // Цикл по текущей дате

    mWordR = ""
    FOR j=1 TO mGroup // Пропущенная группа
        mWordR = mWordR + TOKEN(mInpData," ",t+j-1) + " "
    NEXT

    mWordF = ""
    FOR j=1 TO mGroup // Следующая группа
        mWordF = mWordF + TOKEN(mInpData," ",t+mGroup+j-1) + " "
    NEXT

    APPEND BLANK

    FIELDPUT(1, ALLTRIM(STR(t)))
    FIELDPUT(2, mWordF)
    FIELDPUT(3, mWordR)

    DC_GetProgress(oProgress2, ++nTime, nMax)

NEXT

* MsgBox('STOP')
DC_GetProgress(oProgress2,nMax,nMax)
oDialog2:Destroy()

ENDIF

CLOSE ALL

***** Прошло секунд с начала процесса
Sec_2 = (DOY(DATE())-1)*86400+SECONDS() - Sec_1
Sec_2 = (DOY(DATE())-1)*86400+SECONDS() - Sec_1
ch2 = INT(Sec_2/3600) && Часы
mm2 = INT(Sec_2/60)-ch2*60 && Минуты
cc2 = Sec_2-ch2*3600-mm2*60 && Секунды
Mess = 'Процесс создания БД "Inp_data.dbf" и "Inp_name.txt" завершился успешно! Время исполнения # секунд!'
Mess = STRTRAN(Mess,"#",STRTRAN(STR(cc2,2)," ","0"))

```

```

    LB_Warning(Mess, '(C) Луценко Е.В. АСК-анализ символьных и числовых рядов')
RETURN NIL
*****
FUNCTION LB_Warning( message, ctitle )

    LOCAL aMsg := {}
    DEFAULT cTitle TO ''
    IF valtype(message) # 'A'
        aadd(aMsg,message)
    ELSE
        aMsg := message
    ENDIF
    IF LEN(ALLTRIM(ctitle)) > 0
        DC_MsgBox( ,,aMsg,ctitle)
    ELSE
        DC_MsgBox( ,,aMsg,'Универсальная когнитивная аналитическая система "Эйдос-X++"')
    ENDIF

RETURN NIL

```

* Примечание: красным цветом выделена строка, в которой задается тип генератора псевдослучайных **чисел**

Это сделано с целью облегчить программистам ее реализацию на других языках программирования, если у них возникнет такое желание.

7.3.5.2. ХАРАКТЕРИСТИКА ИСХОДНЫХ ДАННЫХ

В исследовании, описанном ниже в данной статье, авторами исследовались числовые псевдослучайные последовательности из одnorазрядных чисел с различной длиной последовательности, используемой в качестве обучающей выборки: 10, 20, 100, 1000, 2000, 3000, 5000, 7000, 10000, 20000, 30000 чисел. Эти последовательности программа записывает в виде DOS-TXT-файла: c:\Aidos-X\AID_DATA\Inp_data\100\Inp_data.txt в папку, из которой система «Эйдос» берет внешние исходные данные.

7.3.5.3. ЗАВИСИМОСТЬ ДОСТОВЕРНОСТИ МОДЕЛИ ОТ ОБЪЕМОВ ИСХОДНЫХ ДАННЫХ

При создании моделей была исследована зависимость последующей группы из двух одnorазрядных псевдослучайных чисел на предыдущей. Вероятности верной идентификации и неидентификации пар псевдослучайных чисел и значения асимптотического информационного критерия качества шума в различных моделях, созданных на основе 10, 20, 100, 1000, 2000, 3000, 5000, 7000, 10000, 20000, 30000 чисел приведены в таблице 1 на рисунках 5 и 6:

Таблица 1 – Вероятности верной идентификации и неидентификации в моделях, созданных на основе различных объемов выборки

Объем выборки	Вероятность верной идентификации %	Вероятность верной неидентификации %	Асимптотический информационный критерий качества шума	
			% от теоретически максимально возможного	бит
10	100,00000	100,00000	15,656	0,99255
20	100,00000	91,28151	13,004	0,82445
100	100,00000	93,82898	11,343	0,71914
1000	99,29789	62,53239	8,049	0,51030
2000	82,82424	62,24528	7,409	0,46971
3000	79,31265	58,13248	7,113	0,45098
5000	73,86432	56,75676	6,663	0,42243
7000	72,67400	54,79248	6,119	0,38793
10000	68,23047	54,70407	5,994	0,38001
20000	63,56453	52,35414	5,232	0,33168
30000	60,07601	53,29280	5,130	0,32525

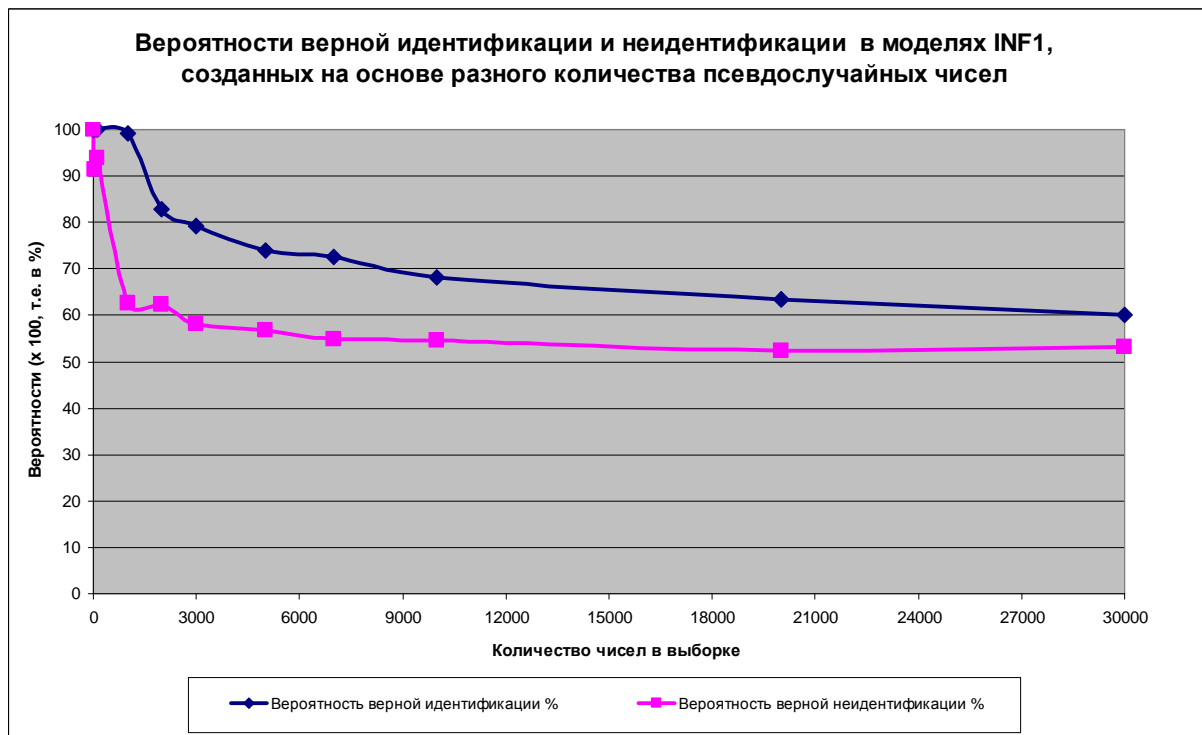


Рисунок 5 – Вероятности верной идентификации и неидентификации в моделях INF1 [15], созданных на основе разного количества псевдослучайных чисел

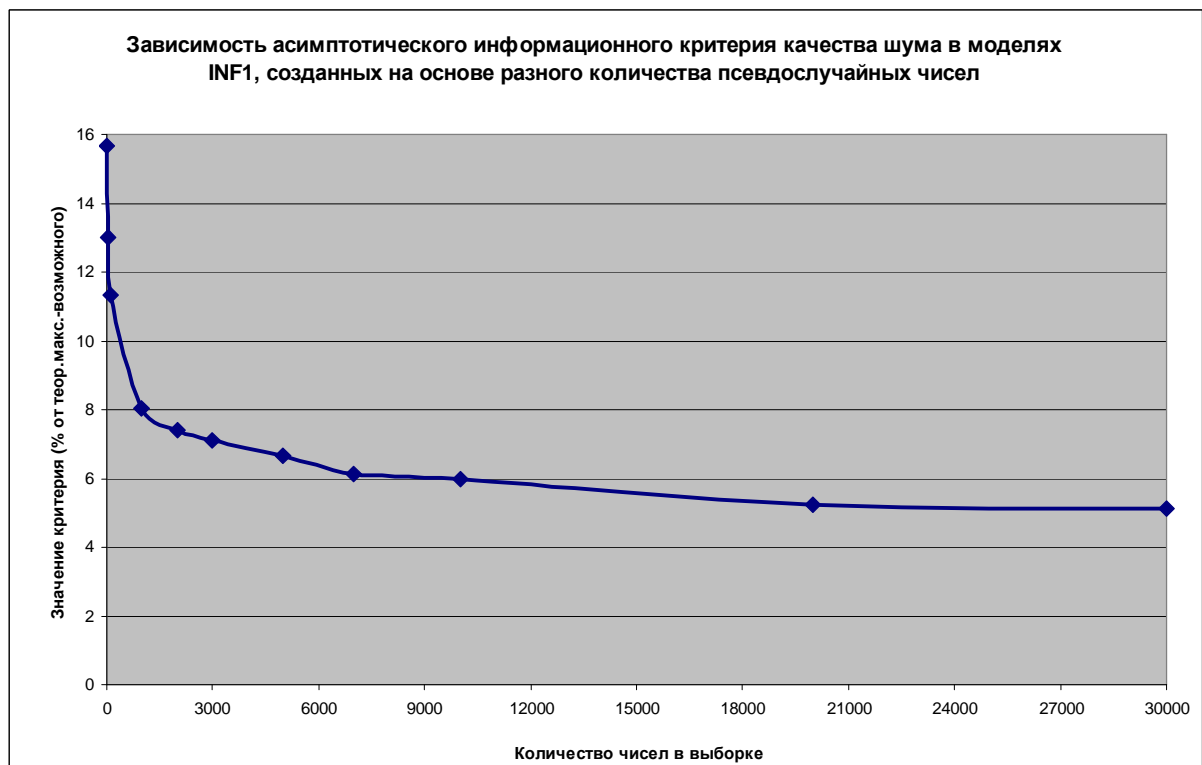


Рисунок 6 – Зависимость асимптотического информационного критерия качества шума в моделях INF1 [15], созданных на основе разного количества псевдослучайных чисел

Из таблицы 1 и рисунков 5 и 6 мы видим, что:

– модели, созданные на основе сравнительно небольшого количества псевдослучайных чисел (до 1000), имеют очень высокую достоверность идентификации пары последующих чисел по паре предшествующих, близкую к 100%;

– при увеличении объема выборки от 1000 до 10000 чисел достоверность сначала быстро, а затем все медленнее и медленнее снижается, т.е. асимптотически сходится к некоторому значению (пределу);

– при объемах выборки от 10000 до 30000 чисел достоверность модели стабилизируется и практически не меняется, асимптотически приближаясь к некоторому предельному значению.

На основе этих результатов можно сделать следующие **выводы**:

1. Системе «Эйдос» успешно удается выявить закономерности взаимосвязи между предыдущей и последующей парой псевдослучайных чисел. Это означает, что **качество шума, генерируемого стандартным генератором псевдослучайных чисел языка программирования xVbase++ (RANDOM()), можно считать довольно низким.**

2. Когда чисел менее 1000, то выявление закономерностей взаимосвязи между предыдущей и последующей парой псевдослучайных чисел для системы «Эйдос» является тривиальной (элементарной) задачей.

3. Но и для моделей, созданных на основе значительно большего количества псевдослучайных чисел: 10000, 20000 и 30000, тоже совершенно очевидно, что полученные результаты были бы невозможны, если бы последовательность чисел была действительно случайной, т.е. шум был качественным.

На основании того факта, что при объемах выборки от 10000 до 30000 чисел достоверность модели практически не меняется, асимптотически приближаясь к некоторому предельному значению, можно сделать вывод о том, предложенный асимптотический информационный критерий качества шума действительно работает. То есть можно считать, что значение этого критерия для выборки 10000 уже достаточно хорошо отражает качество шума и при дальнейшем увеличении объема выборки меняется не существенно.

Отметим, что для чистого шума количественное значение этого критерия должно быть равно 0 и чем ближе критерий к этому значению при таких объемах выборки, при увеличении которых этот критерий уже существенно не меняется, тем ближе к чистому шуму сигнал, на основе которого создана модель.

Из вышесказанного следуют такие формулировки асимптотического информационного критерия близости сигнала к шуму:

– **сигнал тем ближе к шуму, чем быстрее при неограниченном увеличении числа отсчетов стремится к нулю количество информации в значениях одних его элементов о значениях других элементов;**

– **для шума количество информации в одних его элементах о значениях других асимптотически стремится к нулю при неограниченном увеличении количества элементов.**

Полученные закономерности можно считать примерами действия закона больших чисел (в его содержательной интерпретации; математические формулировки еще предстоит получить).

7.3.5.4. SWOT-АНАЛИЗ ВЛИЯНИЯ ПРЕДШЕСТВУЮЩИХ ПАР ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ НА ПОСЛЕДУЮЩИЕ

В созданной модели INF1 отражено, какое количество информации содержится в предшествующей паре псевдослучайных чисел о последующей. Это количество информации может быть положительным (если говорит о том, что произойдет), и отрица-

тельным (если говорит о том, чего не произойдет), также больше или меньше по модулю (чем больше модуль – тем сильнее влияние).

Вся эта информация отражена в SWOT-матрице и SWOT-диаграмме, которые являются стандартными выходными формами системы “Эйдос” (рисунки 7, 8) [16].

На инвертированных SWOT-матрицах и SWOT-диаграммах (предложены автором в работе [16]), мы видим, какие последующие пары псевдослучайных чисел обуславливает предыдущая пара 1_1 (рисунки 9 и 10):

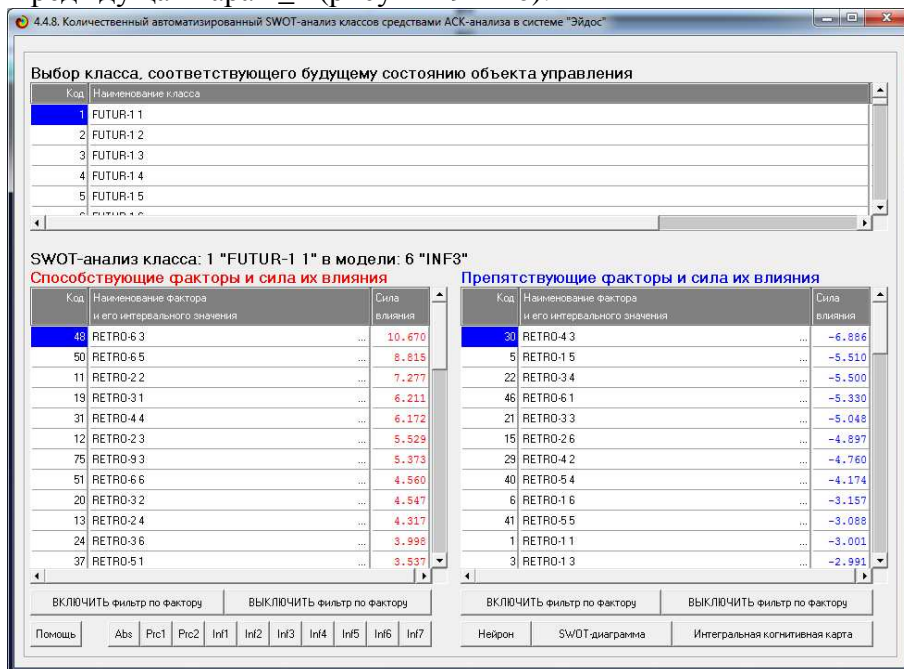


Рисунок 7 – Пример SWOT-матрицы, показывающей зависимости между предыдущими парами чисел и последующей парой 1_1 в модели INF1 [15], созданной основе 30000 псевдослучайных чисел

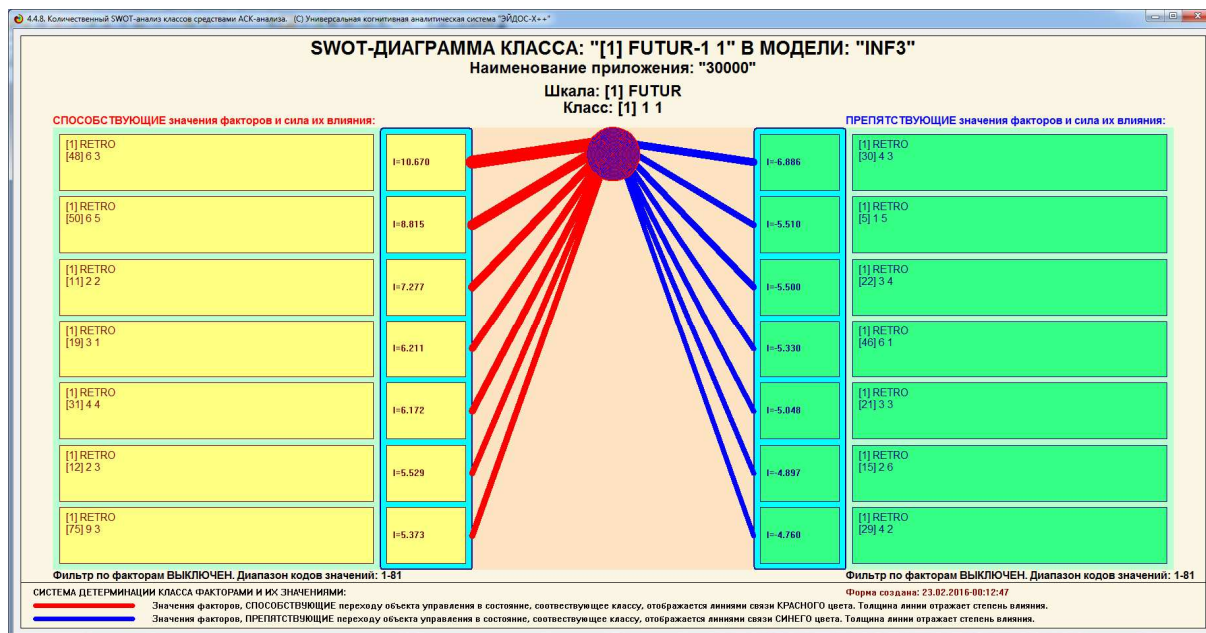


Рисунок 8 – Пример SWOT-диаграммы, показывающей зависимости между предыдущими парами чисел и последующей парой 1_1 в модели INF1 [15], созданной основе 30000 псевдослучайных чисел

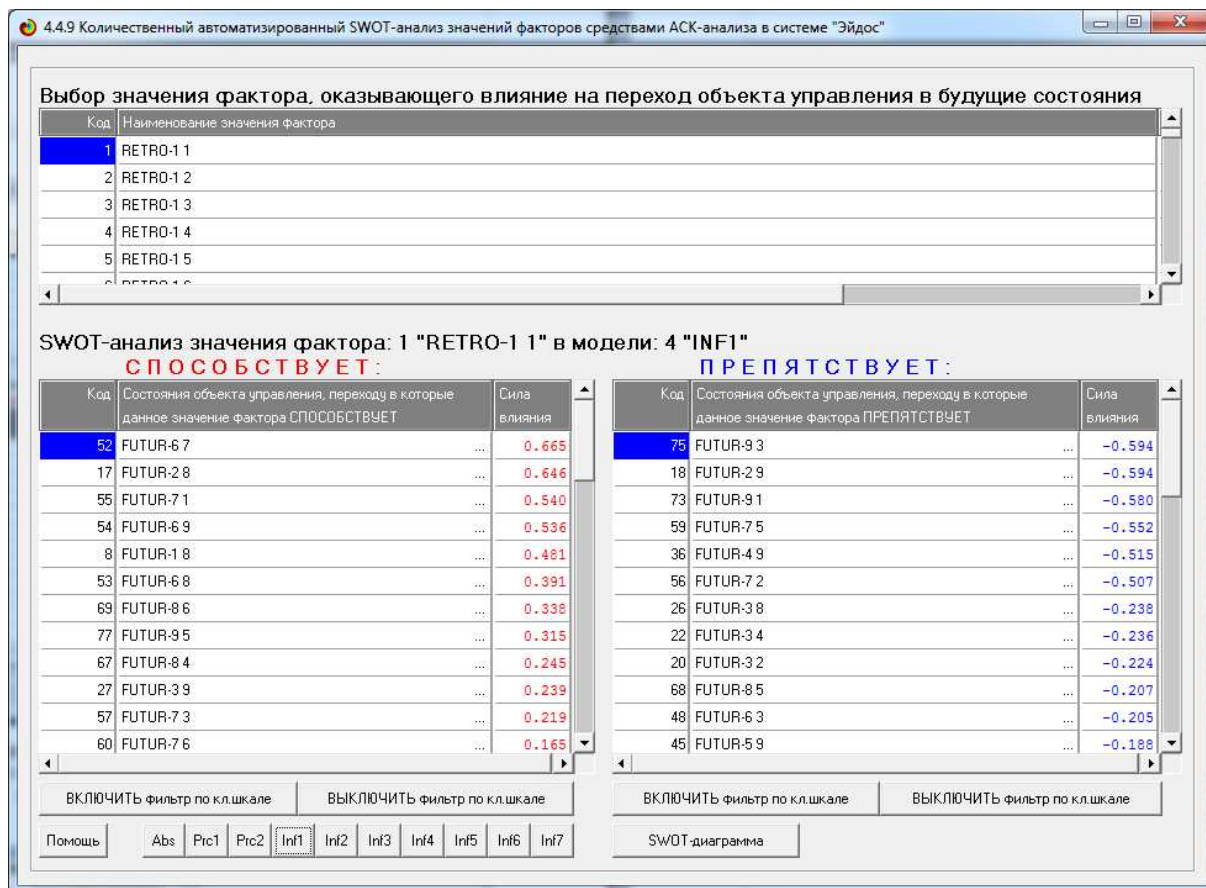


Рисунок 9 – Пример SWOT-матрицы, показывающей зависимости между предыдущей парой чисел 1_1 и последующими в модели INF1 [15], созданной основе 30000 псевдослучайных чисел

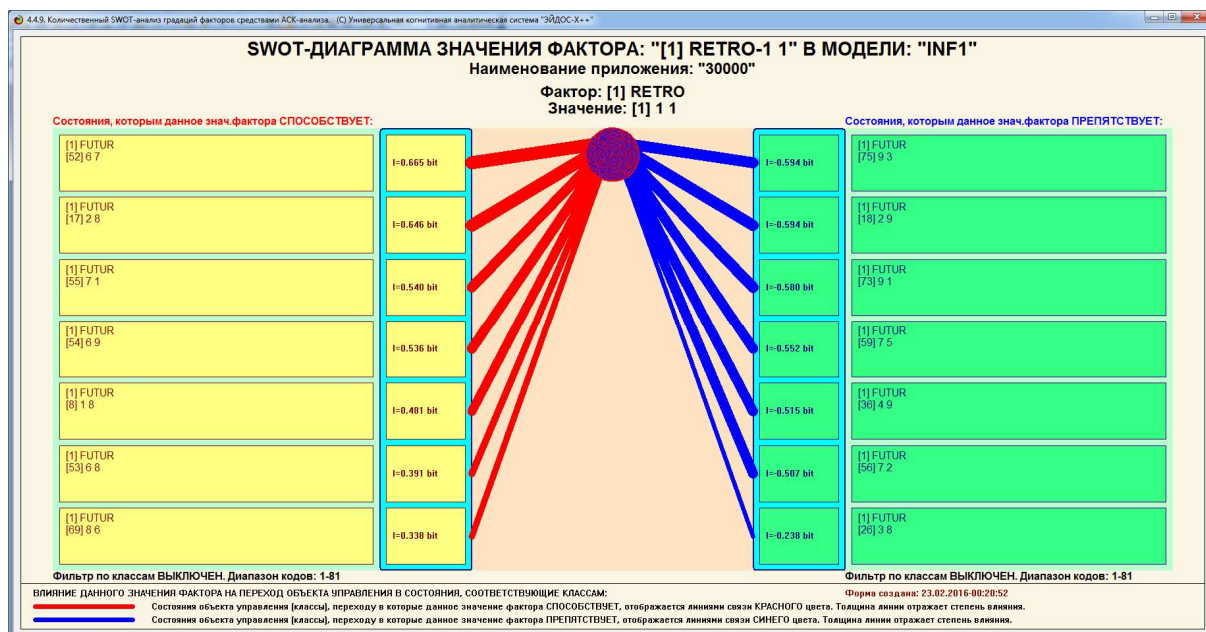


Рисунок 10 – Пример SWOT-диаграммы, показывающей зависимости между предыдущей парой чисел 1_1 и последующими в модели INF1 [15], созданной основе 30000 псевдослучайных чисел

Из рисунков 7, 8 и 9, 10 хорошо видно, что система детерминации будущих пар псевдослучайных чисел, выявленная системой «Эйдос», весьма мало напоминает случайную. При использовании других датчиков псевдослучайных чисел картина может быть иной, но это предмет дальнейших исследований.

7.3.5.5. НАГЛЯДНОЕ ГРАФИЧЕСКОЕ ОТОБРАЖЕНИЕ ЗАКОНОМЕРНОСТЕЙ В СОЗДАННЫХ МОДЕЛЯХ В ФОРМЕ КОГНИТИВНЫХ ФУНКЦИЙ

Когнитивные функции – это наглядное графическое отображение матриц различных моделей, создаваемых системой «Эйдос» и перечисленных на рисунке (2) [1, 3].

Когнитивные функции являются обобщением понятия функции, которое более пригодно для адекватного отражения причинно-следственных зависимостей в реальной области, т.к. они отражают количество информации содержится в значении аргумента о значении функции.

Определены понятия нередуцированных, частично и полностью редуцированных прямых и обратных, позитивных и негативных когнитивных функций и метод формирования редуцированных когнитивных функций, являющийся обобщением известного взвешенного метода наименьших квадратов на основе учета в качестве весов наблюдений количества информации в значениях аргумента о значениях функции [1, 3]¹¹⁹.

Человек обладает естественной высокоразвитой способностью обнаруживать в изображениях закономерности, и в некоторых случаях эта способность на много превосходит аналогичные возможности программного обеспечения и компьютеров. Поэтому когнитивные функции могут быть очень полезны для решения задач, решаемых в данной статье.

В качестве примера рассмотрим две когнитивные функции: матрицы абсолютных частот ABS и матрицы информативностей INF1 для модели созданной на основе 30000 псевдослучайных чисел (рисунки 11 и 12):

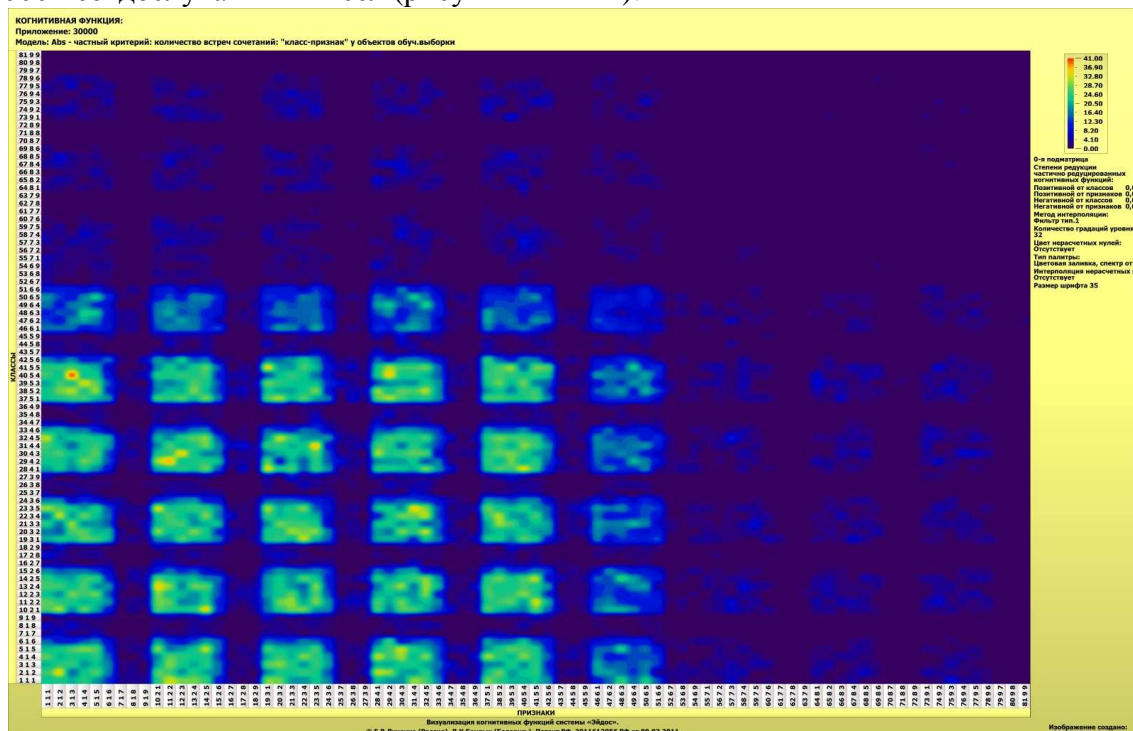


Рисунок 11 – Когнитивная функция матрицы абсолютных частот ABS для модели, созданной на основе 30000 псевдослучайных чисел

¹¹⁹ Подборка публикаций по когнитивным функциям: <http://www.twirpx.com/file/775236/>

В лабораторных работах, встроенных в систему систему «Эйдос-X++», уже есть работа вычислительного типа 2.01: «Исследование случайной семантической информационной модели при различных объемах выборки» [38]¹²⁰ (рисунок 13):

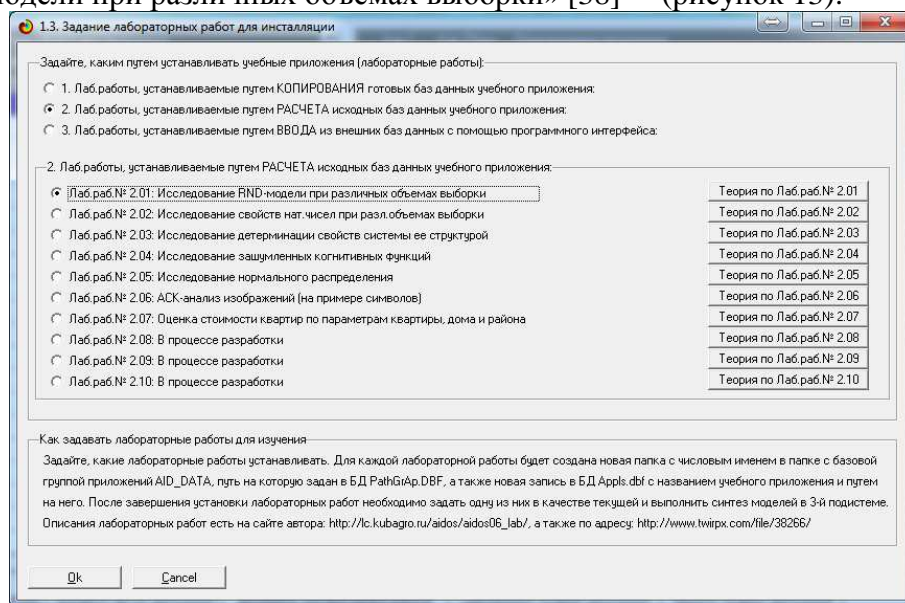


Рисунок 13. Экранная форма системы «Эйдос-X++», обеспечивающая установку лабораторной работы по исследованию псевдослучайных моделей

На основе материалов данной статьи может быть реализована еще одна лабораторная работа (2.08) по дисциплинам, связанным с интеллектуальными технологиями, представлением знаний и системами искусственного интеллекта, а также в других областях [38].

В работах, приведенных в списке литературы [26-40], приведены примеры применения сходных подходов к анализу текстов, последовательности миллиона десятичных знаков числа π и др.

7.3.7. Перспективы

В качестве перспективы продолжения намеченного в данном разделе направления исследований авторы планируют:

- усовершенствовать описанную выше программу генерации исходных данных, которое обеспечит использование для генерации псевдослучайной последовательности различные алгоритмы;

- усовершенствовать описанную выше программу генерации исходных данных, которое обеспечит графическую визуализацию зависимости последующих значений элементов ряда от предыдущих;

- интегрировать описанную выше программу генерации исходных данных в состав системы «Эйдос» как один из видов программного интерфейса с внешними данными и лабораторную работу вычислительного типа (2.01, см. рисунок 13);

- провести численные исследования и сравнения качества шума, получаемого с помощью различных алгоритмов, а также с помощью различных архиваторов и методов шифрования;

- разработать в системе «Эйдос» выходную форму со значениями предложенного в данном разделе асимптотического информационного критерия качества шума для всех создаваемых в системе моделей;

¹²⁰ http://lc.kubagro.ru/aidos/aidos06_lab/lab_10.htm

– применить предельные теоремы теории вероятностей и математической статистики для изучения асимптотических свойств предложенного информационного критерия качества шума.

ЗАКЛЮЧЕНИЕ

Подведем итоги. Монография продолжает две наши предыдущие книги "Системная нечеткая интервальная математика" (Орлов А.И., Луценко Е.В., 2014) и "Перспективные математические и инструментальные методы контроллинга" (Орлов А.И., Луценко Е.В., Лойко В.И., 2015). Первая из них была посвящена разработке новой области математики, во второй дана панорама основанных на первой книге новых перспективных математических и инструментальных методов контроллинга. Более подробным исследованиям посвящена настоящая третья книга.

Более точно, она посвящена примерам таких исследований. Так, в главе 1 методология применения организационно-экономического моделирования при решении задач контроллинга затронута лишь в той мере, которая необходима для адекватного материала остальных глав. Глава 2, с одной стороны, впервые дает анализ развития статистических методов в нашей стране, с другой стороны, должна быть продолжена в серии продвинутых историко-статистических исследований. В главе 3 рассмотрены лишь отдельные разделы прикладной статистики. Ничего не сказано, например, об алгоритмах классификации. Очевидно, впрочем, что достаточно подробное описание прикладной статистики возможно лишь в многотомном энциклопедическом издании. Глава 4 посвящена ракетно-космической промышленности, хотя организационно-экономическое обеспечение контроллинга, инноваций и менеджмента в любой отрасли промышленности, области деятельности заслуживает не менее подробного описания. Различные применения автоматизированного системно-когнитивного анализа (АСК-анализа) в главе 5 (при решении ряда экономических задач) и главе 6 (при оценке результатов научной и преподавательской деятельности) - лишь примеры эффективного применения этого метода. Целесообразно дальнейшее развитие подхода на основе АСК-анализа и теории информации для решения задач прикладной статистики (глава 7).

Авторы надеются, что настоящая монография, несмотря на некоторую ее фрагментарность, будет интересной и полезной для студентов, аспирантов, преподавателей и научных работников, интересующихся современными интеллектуальными математическими и информационными технологиями организационно-экономического моделирования и перспективами их применения в исследованиях в области контроллинга, инноватики и менеджмента, а также в учебном процессе.

ЛИТЕРАТУРА

Литература к главе 1

1. Орлов А.И. Научная школа кафедры «Экономика и организация производства» в области эконометрики // Четвёртые Чарновские Чтения. Сборник трудов. Материалы IV международной научной конференции по организации производства. Москва, 5-6 декабря 2014 г. – М.: НП «Объединение контроллеров», 2014. – С.326 - 337.
2. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.
3. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотипное. – М.: Наука, 1969. – 512 с.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики / 3-е изд.- М.: Наука, 1983. - 416 с. (1-е изд. – 1965).
5. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.
6. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
7. Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. – 64 с.
8. Орлов А.И. О перестройке статистической науки и её применений // Вестник статистики. 1990. №1. С.65 – 71.
9. Орлов А.И. Создана единая статистическая ассоциация // Вестник Академии наук СССР. 1991. №7. С.152-153.
10. Колобов А.А., Омельченко И.Н., Орлов А.И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. – М.: Экзамен, 2008. – 621 с.
11. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. – М.: КноРус, 2010. – 192 с.
12. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Том 78. №1, часть I. С.87-93.
13. Орлов А.И. Основные черты новой парадигмы математической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 90. С. 45-71.
14. Орлов А.И. Эконометрика: Учебное пособие для вузов. — М.: Экзамен, 2002.— 576 с.
15. Орлов А.И. Эконометрика: Учебник для вузов. Издание второе, исправленное и дополненное. — М.: Экзамен, 2003. — 576 с.
16. Орлов А.И. Эконометрика: Учебник для вузов. Издание третье, исправленное и дополненное. — М.: Экзамен, 2004. — 576 с.
17. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. – Ростов-на-Дону: Феникс, 2009. - 572 с.
18. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
19. Орлов А.И. Организационно-экономическое моделирование: учебник : в 3 ч. Ч.2. Экспертные оценки. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. – 486 с.
20. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
21. Орлов А.И. Принятие решений. Теория и методы разработки управленческих решений. — М.: ИКЦ «МарТ»; Ростов н/д: Издательский центр «МарТ», 2005. — 496 с.
22. Орлов А.И. Теория принятия решений. – М.: Экзамен, 2004. – 576 с.
23. Орлов А.И. Оптимальные методы в экономике и управлении: Учеб. пособие по курсу «Организационно-экономическое моделирование». – М.: Изд-во МГТУ им. Н.Э. Баумана, 2007. – 44 с.
24. Орлов А.И. Организационно-экономическое моделирование: теория принятия решений. Учеб. для вузов. – М. : КноРус, 2011. – 568 с.
25. Орлов А.И., Федосеев В.Н. Менеджмент в техносфере. Учебное пособие для вузов. – М.: Академия, 2003. – 384 с.
26. Орлов А.И. Менеджмент: организационно-экономическое моделирование. Учебное пособие для вузов. – Ростов-на-Дону: Феникс, 2009. - 475 с.
27. Файоль А. Общее и промышленное управление. – Л.-М.: Центральный институт труда, 1924. Переиздание: Контроллинг. 1992. Вып. 2. 151 с.
28. Муравьева В.С., Орлов А.И. Организационно-экономические проблемы прогнозирования на промышленном предприятии // Управление большими системами. Выпуск 17. М.: ИПУ РАН, 2007. С.143-158.

29. Муравьева В.С. Организационно-экономические методы прогнозирования на промышленных предприятиях // Седьмой всероссийский симпозиум «Стратегическое планирование и развитие предприятий». Тезисы докладов и сообщений. Секция 2. Модели и методы разработки стратегии предприятия. - М.: ЦЭМИ РАН, 2006. С. 174-176.
30. Орлов А.И. Оценка инфляции по независимой информации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 108. С. 259–287.
31. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. №7. С.64-66.
32. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.
33. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.
34. Муравьева В.С., Орлов А.И. Непараметрическое оценивание точки пересечения регрессионных прямых // Заводская лаборатория. Диагностика материалов. 2008. Т.74. №1. С.63-68.
35. Муравьева В.С. Точка встречи: асимптотическое распределение уровня качества и временного лага // Заводская лаборатория. Диагностика материалов. 2008. Т.74. №3. С.70-73.
36. Муравьева В.С. Разработка организационно-экономических инструментов и методов управления промышленными предприятиями на основе эконометрического прогнозирования. Автореферат дисс. канд. экон. наук. - М.: МГТУ им. Н.Э. Баумана, 2011. - 16 с.
37. Лындина М.И., Орлов А.И. Методы прогнозирования для ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 196–221.
38. Фалько С.Г., Орлов А.И. «Шесть сигм» как подход к совершенствованию бизнеса // Контроллинг. 2004. №4(12). С.42-46.
39. Орлов А.И. «Шесть сигм» - новая система внедрения математических методов исследования // Заводская лаборатория. Диагностика материалов. 2006. Т.72. №5. С.50-53.
40. Орлов А.И. Проблемы внедрения математических и инструментальных методов контроллинга // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 107. С. 1017 – 1048.
41. Орлов А.И. Устойчивые экономико-математические методы и модели. – Saarbrücken (Germany), Lambert Academic Publishing, 2011. – 436 с.
42. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
43. Орлов А.И. Неформальная информационная экономика будущего // Неформальные институты в современной экономике России: Материалы Третьих Друкеровских чтений.- М.: Доброе слово: ИПУ РАН, 2007. – С.72-87.
44. Орлов А.И. Неформальная информационная экономика будущего – новая организационно-экономическая теория // Стратегическое планирование и развитие предприятий. Секция 4 / Материалы Девятого всероссийского симпозиума.– М.: ЦЭМИ РАН, 2008. – С.123-124.
45. Орлов А.И. Экономико-математические методы в контроллинге и неформальная информационная экономика будущего // Формування ринкової економіки: Зб. наук. праць. Спец. вип., присвяч. Міжнар. наук.-практ. конф. «Контролінг у бізнесі: теорія і практика». – К.: КНЕУ, 2008. – С.43-50.
46. Орлов А.И. Основные идеи неформальной информационной экономики будущего – новой организационно-экономической теории // Четвертая международная конференция по проблемам управления: Сборник трудов. – М.: ИПУ РАН, 2009. – С.672 - 686.
47. Орлов А.И. Основные идеи солидарной информационной экономики - базовой организационно-экономической теории // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 112. С. 52 – 77.
48. Орлов А.И. Солидарная информационная экономика - экономика без рынка и денег // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 388 – 418.
49. Орлов А.И. Эконометрическая поддержка контроллинга // Контроллинг. - 2002. - №1. - С.42-53.
50. Орлов А.И. Контроллинг организационно-экономических методов // Контроллинг. 2008. №4 (28). С.12-18.
51. Куликова С.Ю., Муравьева В.С., Орлов А.И. Организационно-экономическое моделирование в контроллинге // Контроллинг. 2009. № 5 (33). С. 42-47.

52. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. 33.

53. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф. С.Г. Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с.

Литература к разделу 2.1

1. Никитина Е.П., Фрейдлина В.Д., Ярхо А.В. Коллекция определений термина «статистика». – М.: МГУ им. М.В. Ломоносова, 1972. – 46 с.

2. Ленин В.И. Развитие капитализма в России. Процесс образования внутреннего рынка для крупной промышленности. – М.: Политиздат, 1986. – XII, 610 с.

3. Гнеденко Б.В. Очерк по истории теории вероятностей. – М.: УРСС, 2001. – 88 с.

4. Клейн Ф. Лекции о развитии математики в XIX столетии. Часть I. – М.-Л.: Объединенное научно-техническое издательство НКТП СССР, 1937. – 432 с.

5. Плошко Б.Г., Елисеева И.И. История статистики: Учебное пособие. – М.: Финансы и статистика, 1990. – 295 с.

6. Бернштейн С.Н. Современное состояние теории вероятностей и ее приложений. – В сб.: Труды Всероссийского съезда математиков в Москве 27 апреля – 4 мая 1927 г. – М. – Л.: ГИЗ, 1928. – С.50 – 63.

7. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов // Заводская лаборатория. 1992. Т.58. № 1. – С.67 – 74.

8. Орлов А.И. О перестройке статистической науки и её применений // Вестник статистики. 1990. № 1. – С.65 – 71.

9. Кендалл М., Стьюарт А. Теория распределений. – М.: Наука, 1966. – 566 с.

10. Кендалл М., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. – 899 с.

11. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. – М.: Наука, 1976. – 736 с.

12. Налимов В.В., Мульченко З.М. Наукометрия. Изучение развития науки как информационного процесса. – М.: Наука, 1969. – 192 с.

13. ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения. – М.: Изд-во стандартов. 1984. – 53 с.

14. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1965 (1-е изд.), 1968 (2-е изд.), 1983 (3-е изд.).

15. Смирнов Н.В. О приближении плотностей распределения случайных величин // Ученые записки МГПИ им. В.П.Потемкина. 1951. Т.XVI. Вып.3. – С. 69 – 96.

16. Орлов А.И. Эконометрика. Учебник для вузов. Изд. 3-е, исправленное и дополненное. – М.: Экзамен, 2004. – 576 с.

17. Орлов А.И. О развитии прикладной статистики. – В сб.: Современные проблемы кибернетики (прикладная статистика). – М.: Знание, 1981. – С.3-14.

18. Тутубалин В.Н. Границы применимости (вероятностно-статистические методы и их возможности). – М.: Знание, 1977. – 64 с.

19. Орлов А.И. Сертификация и статистические методы // Заводская лаборатория. 1997. Т.63. № 3. – С.55 – 62.

20. Орлов А.И. Что дает прикладная статистика народному хозяйству? // Вестник статистики. 1986. №8. – С.52 – 56.

21. Орлов А.И., Орлова Л.А. Применение эконометрических методов при решении задач контроллинга // Контроллинг. 2003. №4(8). – С.50 – 54.

22. Комаров Д.М., Орлов А.И. Роль методологических исследований в разработке методоориентированных экспертных систем (на примере оптимизационных и статистических методов). – В сб.: Вопросы применения экспертных систем. – Минск: Центросистем, 1988. – С.151 – 160.

23. The teaching of statistics / Studies in mathematical education. Vol.7. – Paris, UNESCO, 1991. – 258 pp.

24. Kotz S., Smith K. The Hausdorff Space and Applied Statistics: A View from USSR (Котц С., Смит К. Пространство Хаусдорфа и прикладная статистика: точка зрения ученых СССР). – The American Statistician. November 1988. Vol. 42. № 4. – P.241 – 44.

25. Кудлаев Э.М., Орлов А.И. Вероятностно-статистические методы исследования в работах А.Н. Колмогорова // Заводская лаборатория. 2003. Т.69. № 5. – С.55 – 61.

26. Корнев В.П. Видные деятели отечественной статистики. 1686 – 1990. Биографический словарь. – М.: Финансы и статистика, 1993. – 200 с.

27. Чехов А.П. Остров Сахалин / Сочинения. Тома 14 –15. – М.: Наука, 1978. – 928 с.

28. Орлов А.И. Основные черты новой парадигмы математической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С.188-214. – IDA [article ID]: 0901306013. – Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/13.pdf>.
29. Орлов А.И. О развитии статистики объектов нечисловой природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №09(093). С. 273 – 309. – IDA [article ID]: 0931309019. – Режим доступа: <http://ej.kubagro.ru/2013/09/pdf/19.pdf>
30. Орлов А.И. Математические методы теории классификации / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 423 – 459. – IDA [article ID]: 0951401023. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/23.pdf>
31. Орлов А.И. Основные идеи статистики интервальных данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №10(094). С. 867 – 892. – IDA [article ID]: 0941310060. – Режим доступа: <http://ej.kubagro.ru/2013/10/pdf/60.pdf>
32. Орлов А.И., Луценко Е.В. О развитии системной нечеткой интервальной математики // Философия математики: актуальные проблемы. Математика и реальность. Тезисы Третьей всероссийской научной конференции; 27-28 сентября 2013 г. / Редкол.: Бажанов В.А. и др. – Москва, Центр стратегической конъюнктуры, 2013. – С.190–193.
33. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>
34. Луценко Е.В., Орлов А.И. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/07.pdf>

Литература к разделу 2.2

1. Орлов А.И. Основные этапы становления статистических методов // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 1209 – 1233. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/86.pdf>.
2. Кудлаев Э.М., Орлов А.И. Вероятностно-статистические методы исследования в работах А.Н.Колмогорова // Заводская лаборатория. Диагностика материалов. 2003. Т.69. № 5. С.55-61.
3. Колмогоров в воспоминаниях / Под ред. А.Н.Ширяева. – М.: Физматлит, 1993. - 736 с.
4. Колмогоров А.Н. Избранные труды: Математика и механика. - М.: Наука, 1985. - 470 с.
5. Колмогоров А.Н. Теория вероятностей и математическая статистика. – М.: Наука, 1986. - 535 с.
6. Колмогоров А.Н. Теория информации и теория алгоритмов. – М.: Наука, 1987. - 304 с.
7. Колмогоров А.Н. Основные понятия теории вероятностей. – М.-Л.: ОНТИ, 1936. - 80 с. (3-е издание. – М.: Фазис, 1998. - 144 с. – Серия «Стохастика», вып.1.)
8. Орлов А.И. Эконометрика. Учебник для вузов. Изд. 3-е, исправленное и дополненное. - М.: Экзамен, 2004. – 576 с.
9. Орлов А.И. Прикладная статистика. Учебник. - М.: Экзамен, 2006. - 672 с.
10. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1965 (1-е изд.), 1968 (2-е изд.), 1983 (3-е изд.).
11. Тюрин Ю.Н. Линейная модель в многомерной непараметрической статистике. - В сб.: Многомерный статистический анализ в социально-экономических исследованиях. - М.: Наука, 1974. С.7-24.
12. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>
13. Орлов А.И. О критериях Колмогорова и Смирнова // Заводская лаборатория. 1995. Т.61. № 7. С.59-61.
14. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория.1985. Т.51. №1. С.60-62.

15. Тюрин Ю.Н. Исследования по непараметрической статистике (непараметрические методы и линейная модель). Автореф. дисс. докт. физ.-мат. наук. - М.: МГУ, 1985. - 33 с.
16. Орлов А.И. Методы поиска наиболее информативных множеств признаков в регрессионном анализе // Заводская лаборатория. 1995. Т.61. №1. С.56-58.
17. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
18. Сердобольский В.И., Орлов А.И. Статистический анализ при большом числе параметров. – В сб.: Программно-алгоритмическое обеспечение прикладного многомерного статистического анализа / Тезисы докладов III Всесоюзной школы-семинара. - М.: ЦЭМИ АН СССР, 1987. С.151-160.
19. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
20. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. 1991. Т.57. №7. С.64-66.
21. Орлов А.И. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 189 – 218. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/13.pdf>
22. Орлов А.И. Непараметрический метод наименьших квадратов с периодической составляющей // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №1. С.65-75.
23. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Т.1: Пер. с англ. – М.: Мир, 1984. - 528 с.
24. Гнеденко Б.В. Курс теории вероятностей: Учебник. 7-е изд., исправл. - М.: Эдиториал УРСС, 2001. - 320 с.
25. Ширяев А.Н. Статистический последовательный анализ. – М.: Наука, 1976. - 240 с.
26. Круглов В.М., Королев В.Ю. Предельные теоремы для случайных сумм. – М.: МГУ, 1990. - 188 с.
27. Gnedenko B.V., Korolev V.Yu. Random summation: limit theorems and applications. – CRC Press, Boca Raton, Fl., 1996. - 268 pp.
28. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979. - 296 с.
29. Остроградский М.В. Об одном вопросе, касающемся вероятностей / Полное собрание трудов. Т.3. – Киев: Издательство Академии наук УССР, 1961. – С.215-237.
30. Колмогоров А.Н. Статистический приемочный контроль при допустимом числе дефектных изделий, равном нулю. – Л.: Знание, 1951. - 24 с.
31. Гнеденко Б.В. Математика и контроль качества продукции. – М.: Знание, 1978. - 64 с.
32. Беляев Ю.К. Вероятностные методы выборочного контроля. – М.: Наука, 1975. - 407 с.
33. Лумельский Я.П. Статистические оценки результатов контроля качества. – М.: Изд-во стандартов, 1979. - 200 с.
34. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов // Заводская лаборатория. 1992. Т.58. № 1. С.67-74.
35. Орлов А.И. Сертификация и статистические методы // Заводская лаборатория. 1997. Т.63. № 3. С.55-62.
36. Кас М., Kiefer J., Wolfowitz J. On test of normality and other tests of goodness of fit based on distance methods / Ann. Math. Statist. 1955. V.26. No.2. P.189-211.
37. Рунион Р. Справочник по непараметрической статистике. - М.: Финансы и статистика, 1982. - 198 с.
38. Ивченко Г.И., Медведев Ю.И. Математическая статистика. – М.: Высшая школа, 1984. - 248 с.
39. Лумельский Я.П. К вопросу сравнения несмещенных и других оценок. – В сб.: Прикладная статистика. – М.: Наука, 1983. С.316-319.
40. Невзоров В.Б. Рекорды. Математическая теория. - М.: Фазис, 2000. -ХП+244 с.
41. Журбенко И.Г., Кудлаев Э.М. О выявлении эффекта воздействия в рандомизированных экспериментах // Успехи математических наук. 1984. Т.39. Вып.1. С.3-38.
42. Булинский А.В., Колмогоров А.Н. Линейные выборочные оценки сумм / Теория вероятностей и ее применения. 1979. Т.24. № 2. С. 241-251.
43. Кудлаев Э.М. Рандомизированная процедура оценки айсберговой угрозы, связанная с добычей нефти и газа на морском арктическом шельфе // Заводская лаборатория. Диагностика материалов. 2009. Т.75. № 1. С.75-77.

Литература к разделу 2.3

1. Орлов А.И. Основные этапы становления статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 1209 – 1233. – IDA [article ID]: 0971403086. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/86.pdf>
2. Орлов А.И. Вероятностно-статистические методы в работах А.Н. Колмогорова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №04(098). С. 158 – 180. – IDA [article ID]: 0981404011. – Режим доступа: <http://ej.kubagro.ru/2014/04/pdf/11.pdf>
3. Орлов А.И. Математические методы исследования в работах Бориса Владимировича Гнеденко // Заводская лаборатория. Диагностика материалов. 2007. Т.73. №7. С.66-72.
4. Гнеденко Б.В. Курс теории вероятностей: Учебник. 7-е изд., исправл. - М.: Эдиториал УРСС, 2001. - 320 с.
5. Гнеденко Б.В., Колмогоров А.Н. Предельные распределения для сумм независимых случайных величин. - М.-Л.: ГТТИ, 1949. 264 с.
6. Гнеденко Б.В., Королюк В.С., Ющенко Е.Л. Элементы программирования (2-е изд.). - М.: Физматгиз, 1963. 348 с.
7. Добровольская Н.К. Борис Владимирович Гнеденко // Киевские математики-педагоги. - Киев: Изд-во «Вища школа», 1979. С.37-60.
8. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
9. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
10. Орлов А.И. О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №09(093). С. 273 – 309. – IDA [article ID]: 0931309019. – Режим доступа: <http://ej.kubagro.ru/2013/09/pdf/19.pdf>
11. Гнеденко Б.В. Предельные теоремы для максимального члена вариационного ряда // Доклады АН СССР. 1941. Т.32. С.7-9.
12. Кудлаев Э.М. Оценивание параметров распределения Вейбулла-Гнеденко // Техническая кибернетика. 1986. № 6. С.5-18.
13. Гнеденко Б.В., Кудлаев Э.М. О случайных величинах, обусловленных суммами независимых случайных величин // Вестник МГУ им. М.В. Ломоносова. Сер. мат. и мех. 1995. Вып.1. С.23-31.
14. Боев Г.П., Виноградов Ю.К., Гнеденко Б.В. Методика составления эмпирических зависимостей и номограмм в текстильном деле. - М.: Гизлегпром, 1936. - 128 с.
15. Гнеденко Б.В. К теории счетчиков Гейгер-Мюллера // Журнал экспериментальной и теоретической физики. 1941. Т. 11. Вып. 1. С. 101-106.
16. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. - М.: Наука, 1966. - 301 с.
17. Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию массового обслуживания (изд. б-е). - М.: Наука, 1964. - 146 с.
18. Гнеденко Б.В. Даниелян Э.А., Димитров Б.Н. и др. Приоритетные системы обслуживания. - М.: МГУ, 1973. - 447 с.
19. Гнеденко Б.В. Каждому специалисту нужно знать математическую статистику // Вестник высшей школы. 1961. № 12. С. 29-30.
20. Орлов А.И., Орловский И.В. Оценка остаточного члена порядка n^{-2} для функции распределения двухвыборочной статистики Смирнова. – В сб.: Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. – Пермь: Изд-во Пермского государственного университета, 1978. С.100-109.
21. Колмогоров А.Н. О работах Б.В. Гнеденко по теории вероятностей // Теория вероятностей и ее применения. 1962. Т.7. №2. С.323-329.
22. Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности. - М.: Наука, 1965. - 524 с.
23. Гнеденко Б.В. Математика и контроль качества продукции. - М.: Знание, 1978. - 64 с.
24. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов // Заводская лаборатория. 1992. Т.58. № 1. С.67-74.
25. Орлов А.И. Сертификация и статистические методы // Заводская лаборатория. 1997. Т.63. № 3. С.55-62.
26. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979. - 296 с.

27. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
28. Гнеденко Б.В. Введение в специальность математика. - М., Наука, 1991. - 340 с.
29. Гнеденко Б.В. Очерки истории математики в России. - М.: ГТТИ, 1946. - 247 с.
30. Гнеденко Б.В. О некоторых задачах истории математики // Труды третьего Всесоюзного математического съезда. (Москва, июнь-июль 1956). Т. II. Краткое содержание обзорных и секционных докладов. М.: Изд-во АН СССР, 1956. С.100-101.
31. Гнеденко Б.В. Очерк по истории теории вероятностей. - М.: УРСС, 2001. - 88 с.
32. Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию вероятностей. - М.: ГТТИ, 1946. - 128 с.
33. Орлов А.И., Розенталь А.Л. Про хоккей и теорию вероятностей (беседа с Б.В.Гнеденко) // Пионер. 1971. №10. С.72-73.
34. Орлов А.И. Как поймать случай? (Беседа с Б.В.Гнеденко) // Пионер. 1971. №12. С.68-70.
35. Гнеденко Б.В. О математике. - М.: Эдиториал УРСС. 2000. - 208 с.
36. Гнеденко Б.В., Орлов А.И. Роль математических методов исследования в кардинальном ускорении научно-технического прогресса // Заводская лаборатория. 1988. Т.54. № 1. С.1-4.
37. Гнеденко Б.В., Орлов А.И. О применении современных статистических методов в управлении качеством продукции // Надежность и контроль качества. 1990. №3. С.62-62.
38. Гнеденко Б.В. Математическая статистика - мощное орудие в работе заводской лаборатории // Заводская лаборатория. 1961. Т.27. № 10. С. 1251-1253.

Литература к разделу 2.4

1. Орлов А.И. Основные этапы становления статистических методов / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). – IDA [article ID]: 0971401086. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/86.pdf>
2. Орлов А.И. Вероятностно-статистические методы в работах А.Н. Колмогорова / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №04(098). С. 158 – 180. – IDA [article ID]: 0981404011. – Режим доступа: <http://ej.kubagro.ru/2014/04/pdf/11.pdf>
3. Орлов А.И. Вероятностно-статистические методы в работах Б.В. Гнеденко / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 31 – 52. – IDA [article ID]: 1001406002. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/02.pdf>
4. Колмогоров А.Н. Теория вероятностей и математическая статистика. – М.: Наука, 1986. - 535 с.
5. Смирнов Н.В. Теория вероятностей и математическая статистика. Избранные труды. – М.: Наука, 1970. – 289 с.
6. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1965 (1-е изд.), 1968 (2-е изд.), 1983 (3-е изд.).
7. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотипное. – М.: Наука, 1969. – 512 с.
8. Смирнов Н.В., Белугин Д.А. Теория вероятностей и математическая статистика в приложении к геодезии. – М.: Недра, 1969. – 380 с.
9. Смирнов Н.В. О приближении плотностей распределения случайных величин // Ученые записки МГПИ им. В.П.Потемкина. 1951. Т. XVI. Вып.3. С. 69-96.
10. Орлов А.И. Прикладная статистика. Учебник. - М.: Экзамен, 2006. - 672 с.
11. Орлов А.И. О критериях Колмогорова и Смирнова // Заводская лаборатория. 1995. Т.61. № 7. С.59-61.
12. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – IDA [article ID]: 0971403047. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>
13. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979. - 296 с.
14. Орлов А.И. Скорость сходимости распределения статистики Мизеса – Смирнова // Теория вероятностей и ее применения. 1974. Т.19. №4. С.766-786.
15. Орлов А.И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №11. С.66-70.

16. Орлов А.И., Миронова Н.Г., Фомин В.Н., Черномордик О.М. Методика. Проверка однородности двух выборок параметров продукции при оценке ее технического уровня и качества. - М.: ВНИИ-Стандартизации, 1987. - 116 с.
17. Большев Л.Н. Избранные труды. Теория вероятностей и математическая статистика. - М.: Наука, 1987. - 286 с.
18. Орлов А.И. Статистические пакеты – инструменты исследователя // Заводская лаборатория. 2008. Т.74. №5. С.76-78.
19. Орлов А.И. О развитии прикладной статистики // Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С.3-14.
20. Налимов В.В. Канатоходец. Воспоминания. — М.: Издательская группа «Прогресс», 1994. — 456 с.
21. Налимов В.В. Применение математической статистики при анализе вещества.- М.: Физматгиз, 1960. - 430 с.
22. Налимов В.В., Чернова Н.К. Статистические методы планирования экстремальных экспериментов. - М. Физматгиз, 1965. - 340 с.
23. Налимов В.В. Теория эксперимента. - М.: Наука, 1971. - 207 с.
24. Налимов В.В., Голикова Т.И. Логические основания планирования эксперимента. Изд. 2-е, переработанное и расширенное. - М.: Металлургия, 1981. - 151 с.
25. Маркова Е.В., Никитина Е.П. Математическая теория эксперимента: история, развитие, будущее // Заводская лаборатория. 2002. Т.68. № 1. С.112-118.
26. Математическая теория планирования эксперимента / Под ред. С.М. Ермакова. - М.: Наука, 1983. - 392 с.
27. Налимов В.В., Мульченко З.М. Наукометрия. Изучение развития науки как информационного процесса. - М.: Наука, 1969. - 192 с.
28. Налимов В.В. Вероятностная модель языка. Изд. 2-е, расширенное. - М.; Наука, 1979. - 303 с.
29. Орлов А.И. Математика нечеткости. - Журнал «Наука и жизнь». 1982. No.7. с.60-67.
30. Налимов В.В. Спонтанность сознания: Вероятностная теория смыслов и смысловая архитектура личности. - М.: Изд-во «Прометей» МГПИ им. В.И. Ленина, 1989. - 288 с.
31. Налимов В.В. В поисках иных смыслов. - М.: Издательская группа «Прогресс», 1993. - 280 с.
32. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
33. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). - Краснодар, КубГАУ. 2014. - 600 с.
34. Kotz S., Smith K. The Hausdorff Space and Applied Statistics: A View from USSR (Котц С., Смит К. Пространство Хаусдорфа и прикладная статистика: точка зрения ученых СССР). - The American Statistician. November 1988. Vol. 42. № 4. P. 241-244.
35. Kotz S. Statistical Terminology - Russian Vs. English - in the Light of the Development of Statistics in the USSR // The American Statistician, 1965. Vol. 19, № 3, P.16-22.
36. Kotz S. Statistics in the USSR // Survey, 1965. Vol. 57, October, P.132-141.
37. Плошко Б.Г., Елисеева И.И. История статистики: Учебное пособие. - М.: Финансы и статистика. 1990. - 295 с.
38. Прикладная статистика / Ученые записки по статистике. Т.45. - М.: Наука, 1983. - 383 с.
39. Тимофеев К. Что же такое прикладная статистика? // Вестник статистики. 1985. № 10. С.66-67.
40. Орлов А.И. Что дает прикладная статистика народному хозяйству? // Вестник статистики. 1986. № 8. С.52 - 56.
41. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.
42. Шеремет Н. О так называемой прикладной статистике // Вестник статистики. 1987. № 2. С.67-71.
43. Weinberg J.H., Schumaker J. Statistics: An Intuitive Approach (2-nd ed.). - Belmont, CA: Brooks-Cole. 1969.
44. Мандель И. Теория статистики и прикладная статистика // Вестник статистики. 1987. № 7. С.76-79.
45. Цонев В. К дискуссии по вопросу: что же такое прикладная статистика // Вестник статистики. 1988. № 2. С.67-68.
46. Маркович М. Хроника и информация // Вестник статистики. 1986. № 11. С.62-64.
47. Статистика и перестройка: Ученые записки по статистике. Т.55. - М.: Наука, 1991. - 280 с.
48. Орлов А.И. Создана единая статистическая ассоциация // Вестник Академии наук СССР. 1991. № 7. С.152-153.

49. Орлов А.И. Всесоюзная статистическая ассоциация - гарантия успешного внедрения современных статистических методов // Надежность и контроль качества. 1991. № 6. С.54-55.
50. Орлов А.И. О перестройке статистической науки и её применений. / Вестник статистики. 1990. № 1. С.65 - 71.
51. Устав Всесоюзной статистической ассоциации (ВСА). 1-й Пленум Правления ВСА // Вестник статистики. 1991. № 2. С.71-76.
52. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Т. 78. №1. Часть I. С.87-93.
53. Орлов А.И. О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №09(093). С. 273 – 309. – IDA [article ID]: 0931309019. – Режим доступа: <http://ej.kubagro.ru/2013/09/pdf/19.pdf>
54. Орлов А.И. Основные идеи статистики интервальных данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №10(094). С. 867 – 892. – IDA [article ID]: 0941310060. – Режим доступа: <http://ej.kubagro.ru/2013/10/pdf/60.pdf>
55. Орлов А.И. Основные черты новой парадигмы математической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С. 187 – 213. – IDA [article ID]: 0901306013. – Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/13.pdf>
56. Вероятность и математическая статистика. Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Изд-во «Большая Российская Энциклопедия», 1999. – 910 с.

Литература к разделу 2.5

1. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 672 с.
2. Орлов А.И. Эконометрика. Изд. 3-е, переработанное и дополненное. - М.: Экзамен, 2004. – 576 с.
3. Орлов А.И. Теория принятия решений.– М.: Экзамен, 2006. – 576 с.
4. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.
5. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотипное. – М.: Наука, 1969. – 512 с.
6. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики / 3-е изд.- М.: Наука, 1983. - 416 с. (1-е изд. – 1965).
7. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.
8. Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. – 64 с.
9. Орлов А.И. О перестройке статистической науки и её применений. – Журнал «Вестник статистики». 1990. №.1. С.65 – 71.
10. Орлов А.И. Непараметрическая и прикладная статистика в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 197–226.
11. Материалы республиканской научно-практической конференции "Статистика и ее применения - 2015". Под редакцией профессора А.А. Абдушукурова. - Ташкент: НУУз, 2015. - 500 с.
12. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. № 3. С. 52-60.
13. Орлов А.И. Точки роста статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 136–162.
14. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф. С.Г. Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с.
15. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.
16. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.
17. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
18. Орлов А.И. Устойчивые экономико-математические методы и модели. - Saarbrücken (Germany), Lambert Academic Publishing, 2011. - 436 с.
19. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.

20. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
21. Орлов А.И. О реальных возможностях бутстрепа как статистического метода // Заводская лаборатория. Диагностика материалов. 1987. Т.53. №10. С.82-85.
22. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 163 – 195.
23. Орлов А.И. Основные идеи статистики интервальных данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 94. С. 55-70.
24. Орлов А.И. Статистика интервальных данных (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №3. С. 61 - 69.
25. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
26. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.
27. Орлов А.И. О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 41-50.
28. Орлов А.И. Теория нечетких множеств – часть теории вероятностей // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 92. С. 51-60.
29. Орлов А.И. Экспертные оценки // Заводская лаборатория. Диагностика материалов. 1996. Т.62. №1. С.54-60.
30. Орлов А.И. Организационно-экономическое моделирование: учеб. Ч.2. Экспертные оценки. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. - 486 с.
31. Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1-11.
32. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 89. С. 175-200.
33. Орлов А.И. Оценки плотности распределения вероятностей в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 15-32.
34. Орлов А.И. Предельные теоремы для ядерных оценок плотности в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 108. С. 316 – 333.
35. Орлов А.И. Предельная теория непараметрических статистик // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 31-52.
36. Орлов А.И. Многообразие объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 102. С. 32 – 63.
37. Орлов А.И. Вероятностные модели порождения нечисловых данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 105. С. 39–66.
38. Орлов А.И. Теория люсианов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 275 – 304.
39. Орлов А.И. Расстояния в пространствах статистических данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета, 2014. № 101. С. 227 – 252.
40. Орлов А.И. Математические методы теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 95. С. 23 – 45.
41. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 33–49.
42. Орлов А.И. Базовые результаты математической теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 219–239.

43. Орлов А.И. Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера-Уэлча вместо критерия Стьюдента // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 197–218.
44. Загоруйко Н.Г., Орлов А.И. Некоторые нерешенные математические задачи прикладной статистики // Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С.53-63.
45. Орлов А.И. Некоторые нерешенные вопросы в области математических методов исследования // Заводская лаборатория. Диагностика материалов. 2002. Т.68. №3. С. 52-56.
46. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
47. Орлов А.И. Высокие статистические технологии // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №11. С. 55-60.
48. Орлов А.И. О высоких статистических технологиях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 105. С. 14–38.

Литература к разделу 3.1

1. Орлов А.И. Что дает прикладная статистика народному хозяйству? // Вестник статистики. 1986. № 8. С.52 – 56.
2. Горский В.Г. Планирование промышленных экспериментов. - М.: «Металлургия», 1978. - 112 с.
3. Гуда А.Н. Модели, методы и средства анализа данных в затрудненных условиях. Автореф. дисс. докт. технич. наук. - Таганрог: Таганрогский государственный радиотехнический университет, 1997. 38 с.
4. Налимов В.В. Применение математической статистики при анализе вещества. - М.: Физматгиз, 1960. - 430 с.
5. Налимов В.В., Чернова Н.Л. Статистические методы планирования экстремальных экспериментов. - М.: Физматгиз, 1965. - 340 с.
6. Налимов В.В. Канатоходец. Воспоминания. - М.: Издательская группа "Прогресс", 1994. - 456 с.
7. Гнеденко Б.В., Орлов А.И. Роль математических методов исследования в кардинальном ускорении научно-технического прогресса // Заводская лаборатория. Диагностика материалов. 1988. Т.54. №1. С.1-4.
8. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов. // Заводская лаборатория. Диагностика материалов. 1992. Т.58. №1. С.67-74.
9. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №3. С. 52-60.
10. Горский В.Г., Орлов А.И. Математические методы исследования: итоги и перспективы // Заводская лаборатория. Диагностика материалов. 2002. Т.68. №1. С.108-112.
11. Комаров Д.М., Орлов А.И. Роль методологических исследований в разработке методоориентированных экспертных систем (на примере оптимизационных и статистических методов) // Вопросы применения экспертных систем. - Минск: Центросистем, 1988. С.151-160.
12. Ленин В.И. Развитие капитализма в России. Процесс образования внутреннего рынка для крупной промышленности. - М.: Политиздат, 1986. - XII, 610 с.
13. Гнеденко Б.В. Курс теории вероятностей: Учебник. - Изд. 6-е, перераб. и доп. - М.: Наука, Гл. ред. физ.-мат. лит., 1988. - 448 с.
14. Гнеденко Б.В. Очерк по истории теории вероятностей. – М.: Едиториал УРСС, 2001. – 88 с.
15. Клейн Ф. Лекции о развитии математики в XIX столетии. Часть I. - М. -Л.: Объединенное научно-техническое издательство НКТП СССР, 1937. - 432 с.
16. Плошко Б.Г., Елисеева И.И. История статистики: Учеб. пособие. - М.: Финансы и статистика. 1990. - 295 с.
17. Остроградский М.В. Об одном вопросе, касающемся вероятностей / Полное собрание трудов. Т.3. – Киев: Издательство Академии наук УССР, 1961. – С.215 – 237.
18. Гнеденко Б.В. Математическая статистика и контроль качества. - М.: Знание, 1976. - 64 с.
19. Орлов А.И. Сертификация и статистические методы (обобщающая статья). // Заводская лаборатория. Диагностика материалов. 1997. Т.63. №3. С. 55-62.
20. Орлов А.И. Основные проблемы контроллинга качества // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 111. С. 20-52.
21. Бернштейн С.Н. Современное состояние теории вероятностей и ее приложений // Труды Всероссийского съезда математиков в Москве 27 апреля - 4 мая 1927 г. - М.-Л.: ГИЗ, 1928. - С.50-63.
22. Орлов А.И. Создана единая статистическая ассоциация // Вестник Академии наук СССР. 1991. № 7. С.152-153.

23. Орлов А.И. Первый Всемирный конгресс Общества математической статистики и теории вероятностей им. Бернулли // Надежность и контроль качества. 1987. № 6. С. 54-59.
24. Орлов А.И. О перестройке статистической науки и её применений // Вестник статистики. 1990. № 1. С.65 - 71.
25. Кендалл М., Стьюарт А. Теория распределений. - М.: Наука, 1966. 566 с.
26. Кендалл М., Стьюарт А. Статистические выводы и связи. - М.: Наука, 1973. 899 с.
27. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976. 736 с.
28. Налимов В.В., Мульченко З.М. Наукометрия. Изучение развития науки как информационного процесса. - М.: Наука, 1969. 192 с.
29. ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения. - М.: Изд-во стандартов. 1984. - 53 с.
30. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1965 (1-е изд.), 1968 (2-е изд.), 1983 (3-е изд.).
31. Орлов А.И. Точки роста статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 136–162.
32. Орлов А.И. Пути развития статистических методов: непараметрика, робастность, бутстреп и реалистическая статистика // Надежность и контроль качества. 1991. № 8. С. 3-8.
33. Орлов А.И. О критериях Колмогорова и Смирнова // Заводская лаборатория. Диагностика материалов. 1995. Т.61. №7. С.59-61.
34. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 97. С. 32-45.
35. Холлендер М., Вулф Д. Непараметрические методы статистики. - М.: Финансы и статистика, 1985. - 518 с.
36. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.
37. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.
38. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. №7. С.64-66.
39. Хьюбер П. Робастность в статистике. - М.: Мир, 1984. - 304 с.
40. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. - М.: Мир, 1989. - 512 с.
41. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
42. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания: Статистическая обработка неоднородных совокупностей. - М.: Статистика, 1980. - 208 с.
43. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.
44. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
45. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 163 – 195.
46. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 114. С. 27 – 41.
47. Efron B. Bootstrap Methods: Another Look at the Jackknife // Ann. Statist. 1979. V.7. № 1. P.1-26.
48. Диаконис П., Эфрон Б. Статистические методы с интенсивным использованием ЭВМ // В мире науки. 1983. № 7. С.60-73.
49. Подборка статей по бутстрепу / Заводская лаборатория. 1987. Т.53. № 10. С.76-99.
50. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. - М.: Финансы и статистика, 1988. 263 с.
51. Орлов А.И. О реальных возможностях бутстреп как статистического метода // Заводская лаборатория. Диагностика материалов. 1987. Т.53. №10. С.82-85.
52. Дискуссия по анализу интервальных данных / Заводская лаборатория. Диагностика материалов. 1990. Т.56. № 7. С.75-95.

53. Сборник трудов Международной конференции по интервальным и стохастическим методам в науке и технике. Т.т. 1,2. - М.: МЭИ, 1992.
54. Шокин Ю.И. Интервальный анализ. - Новосибирск: Наука, 1981. 284 с.
55. Вошинин А.П. Метод оптимизации объектов по интервальным моделям целевой функции. - М.: МЭИ, 1987. 109 с.
56. Вошинин А.П., Сотиров Г.Р. Оптимизация в условиях неопределенности. - М.: МЭИ - София: Техника, 1989. 224 с.
57. Вошинин А.П., Акматбеков Р.А. Оптимизация по регрессионным моделям и планирование эксперимента. - Бишкек: Илим, 1991. 164 с.
58. Вошинин А.П., Скибицкий Н.В. Интервальный подход к выражению неопределенности измерений и калибровке цифровых измерительных систем // Заводская лаборатория. Диагностика материалов. 2017. Т.73. № 11. С. 66-71.
59. Скибицкий Н.В., Севальнев Н.В. Интервальные модели в задачах оптимального управления с дифференциальными связями // Заводская лаборатория. Диагностика материалов. 2015. Т.81. № 11. С. 73-80.
60. Орлов А.И. Комментарий IV к статье А.П. Вошинина, А.Ф. Бочкова, Г.Р. Сотирова «Метод анализа данных при интервальной нестатистической ошибке» // Заводская лаборатория. Диагностика материалов. 1990. Т.56. №7. С.86-89.
61. Orlov A.I. Interval statistics // Interval Computations. 1992. № 1(3). Pp. 44-52.
62. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
63. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 574 с.
64. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
65. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.
66. Орлов А.И. Организационно-экономическое моделирование: учебник : в 3 ч. Ч.2. Экспертные оценки. М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. 486 с.
67. Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1-11.
68. Суппес П., Зинес Дж. Основы теории измерений // Психологические измерения. -М., Мир,1967. С. 9-110.
69. Пфанцагль И. Теория измерений. - М.: Мир, 1976. 166 с.
70. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. - М.: Мир, 1976. 168 с.
71. Дэвид Г. Метод парных сравнений. - М.: Статистика, 1978. 144 с.
72. Матерон Ж. Случайные множества и интегральная геометрия. - М.: Мир, 1978. 318 с.
73. Терехина А.Ю. Анализ данных методами многомерного шкалирования. - М.: Наука, 1986. 168 с.
74. Перекрест В.Т. Нелинейный типологический анализ социально-экономической информации: Математические и вычислительные методы. - Л.: Наука, 1983. 176 с.
75. Кемени Дж., Снелл Дж. Кибернетическое моделирование: Некоторые приложения. - М.: Советское радио, 1972. 192 с.
76. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1981. - 80 с.
77. Орлов А.И. Статистика объектов нечисловой природы и экспертные оценки // Экспертные оценки. Вопросы кибернетики. Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1979. С.17-33.
78. Орлов А.И. Статистика объектов нечисловой природы (Обзор) // Заводская лаборатория. Диагностика материалов. 1990. Т.56. №3. С.76-83.
79. Орлов А.И. Тридцать лет статистики объектов нечисловой природы (обзор) // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №5. С.55-64.
80. Орлов А.И. О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 41-50.
81. Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. - 220 с.
82. Рыданова Г.В. Некоторые вопросы статистического анализа случайных бинарных векторов. Автореф. дисс. канд. физ.-мат. наук. - М.: МГУ, 1988. 16 с.
83. Орлов А.И. Асимптотическое поведение статистик интегрального типа // Доклады АН СССР. 1974. Т.219. № 4. С. 808-811.

84. Орлов А.И. Асимптотическое поведение статистик интегрального типа // Вероятностные процессы и их приложения. - М.: МИЭМ, 1989. С.118-123.
85. Орлов А.И. Задачи оптимизации и нечеткие переменные. - М.: Знание, 1980. - 64 с.
86. Орлов А.И. Теория нечетких множеств – часть теории вероятностей // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 92. С. 51-60.
87. Орлов А.И. Расстояния в пространствах статистических данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета, 2014. № 101. С. 227 – 252.
88. Орлов А.И. Заметки по теории классификации // Социология: методология, методы, математические модели. 1991. № 2. С. 28-50.
89. Орлов А.И. Математические методы теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 95. С. 23 – 45.
90. Орлов А.И. Базовые результаты математической теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 219–239.
91. Титма М.Х., Тоодинг Л.М. Математические методы в арсенале социолога // Социологические исследования. 1986. № 4. С. 123-128.
92. Загоруйко Н.Г., Орлов А.И. Некоторые нерешенные математические задачи прикладной статистики // Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. С. 53-63.

Литература к разделу 3.2

1. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Большая Российская энциклопедия, 1999. – 910с.
2. Гнеденко Б.В. Курс теории вероятностей: Учебник. 7-е изд., исправл. - М.: Эдиториал УРСС, 2001. – 320 с.
3. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 556 – 586. – IDA [article ID]: 0891305038. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/38.pdf>
4. Орлов А. И. О средних величинах // Управление большими системами. Выпуск 46. М.: ИПУ РАН, 2013. С.88-117.
5. Рао С.Р. Линейные статистические методы и их применения. – М.: Наука, 1968. – 548 с.
6. Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики / А.И. Орлов, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>
7. Орлов А.И. О развитии статистики объектов нечисловой природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №09(093). С. 273 – 309. – IDA [article ID]: 0931309019. – Режим доступа: <http://ej.kubagro.ru/2013/09/pdf/19.pdf>
8. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
9. Прохоров Ю. В. Сходимость случайных процессов и предельные теоремы теории вероятностей // Теория вероятностей и ее применения. 1956. Т. 1, № 2. С. 177–238.
10. Келли Дж. Общая топология. - М.: Наука, 1968. - 384 с.
11. Орлов А.И. Асимптотическое поведение статистик интегрального типа // Доклады АН СССР. 1974. Т.219. № 4. С. 808-811.
12. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
13. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1964.- 576 с.
14. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, Омега-квадрат и ошибки при их применении / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – IDA [article ID]: 0971403047. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>
15. Биллингсли П. Сходимость вероятностных мер. – М.: Наука, 1977. - 352 с.

16. Орлов А.И. Предельная теория непараметрических статистик / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 226 – 244. – IDA [article ID]: 1001406011. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/11.pdf>

Литература к разделу 3.3

1. Орлов А.И. Основные этапы становления статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 97. С. 73–85.
2. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №3. С. 52-60.
3. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
4. Орлов А.И. Состояние и перспективы развития прикладной и теоретической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 115. С. 202–226.
5. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. – Ростов-на-Дону: Феникс, 2009. - 572 с.
6. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.
7. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.
8. Налимов В.В. Применение математической статистики при анализе вещества. - М.: ГИФМЛ, 1960. - 430 с.
9. Clancey V.J. Statistical methods in chemical analyses // Nature. 1947. V.159. № 4036. P.339-340.
10. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. - Л.: Энергоатомиздат, 1985. - 248 с.
11. Новицкий П.В. Основы информационной теории измерительных устройств. -Л.: энергия, 1968. - 248 с.
12. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. №7. С.64-66.
13. Боровков А.А. Теория вероятностей. - М.: Наука, 1976. - 352 с.
14. Петров В.В. Суммы независимых случайных величин. - М.: Наука, 1972. - 416 с.
15. Золотарев В.М. Современная теория суммирования независимых случайных величин. - М.: Наука, 1986. - 416 с.
16. Егорова Л.А., Харитонов Ю.С., Соколовская Л.В. О применении непараметрического X-критерия Ван-дер-Вардена при статистической обработке результатов наблюдений // Заводская лаборатория. Диагностика материалов. 1976. Т.42. №10. С. 1237-1239.
17. Орлов А.И. Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера-Уэлча вместо критерия Стьюдента // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 197–218.
18. Артемьев Б.Г., Голубов С.М. Справочное пособие для работников метрологических служб.- М.: Изд-во стандартов, 1982. - 280 с.
19. Орлов А.И. Устойчивость в социально-экономических моделях. — М.: Наука, 1979. — 296 с.
20. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.
21. Орлов А.И. Устойчивые экономико-математические методы и модели. Saarbrücken (Germany), Lambert Academic Publishing, 2011. 436 с.
22. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
23. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. - 416 с.
24. Орлов А.И. Неустойчивость параметрических методов отбраковки резко выделяющихся наблюдений. // Заводская лаборатория. Диагностика материалов. 1992. Т.58. №7. С.40-42.
25. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.

26. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.

Литература к разделу 3.4

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.
2. Опыт применения ЭВМ в социологических исследованиях. - М.: Институт социологических исследований АН СССР, Советская социологическая ассоциация, 1977. - 158 с.
3. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
4. Орлов А.И. Общий взгляд на статистику объектов нечисловой природы. - В сб.: Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. - С.58-92.
5. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
6. Сэндидж Ч., Фрайбургер В., Ротцолл К. Реклама: теория и практика: Пер. с англ. - М.: Прогресс, 1989. - 630 с.
7. Ядов В.А. Стратегии и методы качественного анализа данных. - Журнал «Социология: методология, методы, математические модели». 1991. №1. С.14-31.

Литература к разделу 3.5

1. Аффифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. - М.: Мир, 1982. - 488 с.
2. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.
3. Боровков А.А. Математическая статистика. - М.: Наука, 1984. - 472 с.
4. Ван-дер-Варден Б.Л. Математическая статистика. - М.: ИЛ, 1960. - 434 с.
5. Гаек Я., Шидак З. Теория ранговых критериев / Пер. с англ. - М.: Наука, 1971. - 376 с.
6. Гублер Е.В., Генкин А.А. Применение критериев непараметрической статистики в медико-биологических исследованиях. - Л.: Медицина, 1973. - 144 с.
7. Ивченко Г.И., Медведев Ю.И. Математическая статистика. - М.: Высшая школа, 1984. - 248 с.
8. Камень Ю.Э., Камень Я.Э., Орлов А.И. Реальные и номинальные уровни значимости в задачах проверки статистических гипотез. - Журнал «Заводская лаборатория». 1986. Т.52. № 12. С.55-57.
9. Карякин Р.Н., Орлов А.И., Адамов С.Ю. Вероятностная теория высших гармоник помех, создаваемых электровозами. - В сб.: Прикладной многомерный статистический анализ. Ученые записки по статистике, т.33. - М.: Наука, 1978. С.376-380.
10. Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. - М.: Наука, 1973. - 900 с.
11. Кокс Д.Р., Хинкли Д.В. Теоретическая статистика. - М.: Мир, 1978. - 560 с.
12. Крамер Г. Математические методы статистики / Пер. с англ. / 2-е изд. - М.: Мир, 1975. - 648 с.
13. Леман Э. Проверка статистических гипотез. - М.: Наука, 1979. - 408 с.
14. Методика. Проверка однородности двух выборок параметров продукции при оценке ее технического уровня и качества. - М.: ВНИИ стандартизации, 1987. - 116 с.
15. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
16. Новиков Д.А. Статистические методы в педагогических исследованиях (типовые случаи). - М.: МЗ-Пресс, 2004. - 67 с.
17. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
18. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов. - Журнал «Заводская лаборатория». 1992. Т.58. №1. С.67-74.
19. Орлов А.И. Высокие статистические технологии. - Журнал «Заводская лаборатория». 2003. Т.69. №11. С.55-60.
20. Орлов А.И. Эконометрика. Из. 3-е, перераб. и дополн. - М.: Экзамен, 2004. - 576 с.
21. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
22. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотипное. - М.: Наука, 1969. - 512 с.
23. Смолянский М.Л. Таблицы неопределенных интегралов. - М.: ГИФМЛ, 1961. - 108 с.
24. Холлендер М., Вульф Д. Непараметрические методы статистики. - М.: Финансы и статистика, 1983. - 518 с.

Литература к разделу 3.6

1. Практикум по эконометрике: Учеб. пособие / И.И. Елисева, С.В. Курышева, Н.М. Гордеенко и др.; Под ред. И.И. Елисеевой. - М.: Финансы и статистика. 2001. - 192 с.

2. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. – Ростов-на-Дону: Феникс, 2009. - 572 с.
3. Крюкова Е.М. Применение методов организационно-экономического прогнозирования в отрасли лова черных металлов // Заводская лаборатория. Диагностика материалов. - 2008. - Т.74. - №7. - С.67 – 72.
4. Орлов А.И. Непараметрический метод наименьших квадратов: учет сезонности // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып. XXI. – Пермь: Перм. ун-т, 2008. – С.135-148.
5. Орлов А.И. Непараметрический метод наименьших квадратов с периодической составляющей: условия применимости // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып. XXII. – Пермь: Перм. ун-т, 2010. – С.96-108.
6. Орлов А.И. Устойчивые экономико-математические методы и модели. Разработка и развитие устойчивых экономико-математических методов и моделей для модернизации управления предприятиями. – Saarbrücken: LAP, 2011. – 436 с.
7. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
8. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. - 2012. - Том 78. - №1, часть I. - С.87-93.
9. Орлов А.И. Основные черты новой парадигмы математической статистики / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С.188-214. – IDA [article ID]: 0901306013. – Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/13.pdf>, 1,688 у.п.л.
10. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(89). – С. 554 – 584. IDA [article ID]: 0891304038. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/38.pdf>, 1,938 у.п.л., импакт-фактор РИНЦ=0,577
11. Орлов А.И. Организационно-экономическое моделирование при решении задач управления хозяйственными единицами / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №03(87). – С. 679–705. – Режим доступа: <http://ej.kubagro.ru/2013/03/pdf/53.pdf>
12. Орлов А.И. Метод оценивания длины периода и периодической составляющей сигнала // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. – Пермь: Изд-во Пермского государственного университета, 1999. – С.38-49.
13. Орлов А.И. Проблемы методологии государственной политики и управления в неформальной информационной экономике будущего / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(88). – С. 592 – 618. – IDA [article ID]: 0881304041. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/41.pdf>

Литература к разделу 4.1

1. Орлов А.И. Об организационно-экономическом обеспечении решения задач управления космической деятельностью // XL Академические чтения по космонавтике, посвященные памяти академика С.П. Королёва и других выдающихся отечественных ученых - пионеров освоения космического пространства : сборник тезисов / Российская академия наук, Государственная корпорация по космической деятельности "РОСКОСМОС", Комиссия РАН по разработке научного наследия пионеров освоения космического пространства, Московский государственный университет им. Н.Э. Баумана. - М.: МГТУ им. Н.Э. Баумана, 2015. - С. 138-139.
2. Новиков А.М., Новиков Д.А. Методология. - М.: СИНТЕГ, 2007. - 668 с.
3. Орлов А.И. Организационно-экономическое обеспечение инновационной деятельности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 96. С. 605–632.
4. Проектирование интегрированных производственно- корпоративных структур: эффективность, организация, управление / Колобов А.А., Омельченко И.Н., Орлов А.И. и др. / Под ред. А.А. Колобова, А.И. Орлова. Научное издание. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2006. — 728 с.

5. Колобов А. А., Омельченко И. Н., Орлов А. И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. Учебник для вузов. — М.: Экзамен, 2008. — 621 с.
6. Орлов А.И. Менеджмент: организационно-экономическое моделирование. Учебное пособие для вузов. – Ростов-на-Дону: Феникс, 2009. - 475 с.
7. Загонова Н.С., Орлов А.И. Эконометрическая поддержка контроллинга инноваций. Нечеткий выбор // Российское предпринимательство. 2004. №4. С.54-57.
8. Орлов А.И. Инновационная деятельность: организационно-экономическое обеспечение и Интернет-аукционы. - Проблемы информационной экономики. Выпуск VII. Стратегия инновационного развития российской экономики: Сб. научных трудов / Под ред. Р.М. Нижегородцева. - М.: Книжный дом «ЛИБРОКОМ», 2008. – С.325-345.
9. Орлов А.И. О подходах к разработке организационно-экономического обеспечения решения задач управления в аэрокосмической отрасли // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 884-896.
10. Орлов А.И. Организационно-экономическое обеспечение инновационной деятельности в ракетно-космической отрасли // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 102. С. 112–143.
11. Волков В.А., Орлов А.И. О разработке организационно-экономического обеспечения решения задач управления космической деятельностью // К.Э. Циолковский и стратегия развития космонавтики: Материалы XLIX научных чтений памяти К.Э. Циолковского. 2014. – Калуга, 2014. - С.251-252.
12. Орлов А.И. О методологических основах разработки организационно-экономического обеспечения решения задач управления космической деятельностью // Актуальные проблемы космонавтики: Труды XXXIX академических чтений по космонавтике, посвященных памяти академика С.П. Королева и других выдающихся отечественных ученых-пионеров освоения космического пространства. Москва, 27 – 30 января 2015 г. - М.: МГТУ им. Н.Э. Баумана, 2015. - С.153-155.
13. Орлов А. И., Пугач О. В. Подходы к общей теории риска // Управление большими системами. Выпуск 40. М.: ИПУ РАН, 2012. С.49-82.
14. Орлов А.И. Многообразии рисков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 111. С. 53-80.
15. Орлов А.И. Современное состояние контроллинга рисков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 98. С. 933-942.
16. Орлов А.И. Аддитивно-мультипликативная модель оценки рисков при создании ракетно-космической техники // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 102. С. 78–111.
17. Орлов А.И., Цисарский А.Д. Особенности оценки рисков при создании ракетно-космической техники // Национальные интересы: приоритеты и безопасность. 2013. – №43(232). – С.37 – 46.
18. Орлов А.И., Цисарский А.Д. Аддитивно-мультипликативная модель оценки рисков и ее применение при разработке инновационно-инвестиционных проектов создания ракетно-космической техники // Проблемы управления безопасностью сложных систем: Труды XXI Международной конференции. Москва, декабрь 2013 г. / Под ред. Н.И. Архиповой, В.В. Кульбы. М.: РГГУ, 2013. С.394-398.
19. Орлов А.И. Цисарский А.Д. Особенности оценки рисков при реализации инновационно-инвестиционных проектов в космической отрасли // Актуальные проблемы российской космонавтики. Материалы XXXVIII Академических чтений по космонавтике. Москва, январь 2014 г. / Под общей редакцией А.К. Медведевой. М.: Комиссия РАН по разработке научного наследия пионеров освоения космического пространства, 2014. С.210-210.
20. Волков В.А., Орлов А.И. Организационно-экономические подходы к оценке реализуемости инновационно-инвестиционных проектов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 97. С. 1074-1086.
21. Волков В.А., Орлов А.И. Организационно-экономические подходы к оценке реализуемости проектов по созданию ракетно-космической техники // Экономический анализ: теория и практика. 2014. № 11 (362). С.41–47.
22. Волков В.А., Баев Г.О., Орлов А.И., Фалько С.Г. Требования и оценка реализуемости проектов создания изделий ракетно-космической техники // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 939-958.
23. Волков В.А. Методы оценки и управления реализуемостью проектов по созданию ракетно-космической техники. Автореферат диссертации на соискание ученой степени кандидата экономических наук. - М.: МГТУ им. Н.Э. Баумана, 2015. - 15 с.
24. Контроллинг / А.М. Карминский, С.Г. Фалько, А.А. Жевага, Н.Ю. Иванова; под ред. А.М. Карминского, С.Г. Фалько. – 3-е изд., дораб. – М.: ИД «ФОРУМ»: ИНФРА-М, 2013. – 336 с.

25. Фалько С.Г. Контроллинг для руководителей и специалистов.- М.: Финансы и статистика, 2008. – 272 с.
26. Фалько С.Г., Иванова Н.Ю. Управление нововведениями на высокотехнологичных предприятиях. – М.: МГТУ им. Н.Э. Баумана, 2007. – 256 с.
27. Орлов А.И. Основные проблемы контроллинга качества // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 111. С. 20-52.
28. Орлов А.И. Контроллинг организационно-экономических методов // Контроллинг. – 2008. – №4 (28). – С.12-18.
29. Орлов А.И. Новая область контроллинга – контроллинг организационно-экономических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 1126-1137.
30. Орлов А.И. Наука как объект управления // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 1243 – 1273.
31. Орлов А.И. О строительстве науки в отдельно взятой стране // Электронный журнал «Biocosmology – neo-Aristotelism». – 2014, Summer. – Vol.4 – No. 3. – Pp. 203 – 223. [Электронный ресурс]. Режим доступа: <https://sites.google.com/site/biocosmologyneoaristotelism/home/1> (дата обращения 27.02.2016).
32. Чеботарев П.Ю. Оценка ученых: пейзаж перед битвой // Наукометрия и экспертиза в управлении наукой. – М.: ИПУ РАН, 2013. – С. 506–537.
33. Гринченко С.Н. Является ли мировая наука «организмом»? // Биокосмология – нео-Аристотелизм. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. Vol. 4. № 1–2 (Winter / Spring 2014). – P. 115–122.
34. Хруцкий К.С. Триади́ческий биокосмологический подход к вопросам развития науки России // Биокосмология – нео-Аристотелизм. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. → Vol.3. № 3 (Summer 2013). – P. 375–390.
35. Управление большими системами / Сборник трудов. Специальный выпуск 44. Наукометрия и экспертиза в управлении наукой / Под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева. М.: ИПУ РАН, 2013. – 568 с.
36. Наукометрия и экспертиза в управлении наукой: сборник статей / Под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева. М.: ИПУ РАН, 2013. – 572 с.
37. Орлов А.И. О ключевых показателях эффективности научной деятельности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 111. С. 81-112.
38. Мухин В.В., Орлов А.И. О контроллинге научной деятельности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 1222-1237.
39. Мухин В.В., Орлов А.И. Совершенствование организационных структур и контроллинг персонала на предприятиях типа "Научно-исследовательский институт" ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 109. С. 265–296.
40. Орлов А.И. Последствия принятия решений для научно-технического и экономического развития // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 355–387.
41. Лындина М.И., Орлов А.И. Математическая теория рейтингов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 114. С. 1 – 26.
42. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 33–49.
43. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
44. Лындина М.И., Орлов А.И. Методы прогнозирования для ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 196–221.
45. Орлов А.И. Развитие статистических и экспертных методов прогнозирования в аэрокосмической отрасли // Актуальные проблемы российской космонавтики. Материалы XXXVIII Академических

чтений по космонавтике. Москва, январь 2014 г. / Под общей редакцией А.К. Медведевой. М.: Комиссия РАН по разработке научного наследия пионеров освоения космического пространства, 2014. С.209-209.

46. Орлов А.И. Принятие решений и экспертные оценки в авиации и ракетно-космической промышленности // Теория активных систем: Труды международной научно-практической конференции (17-19 ноября 2014 г., Москва, Россия). Общая редакция – В.Н. Бурков, Д.А. Новиков. – М.: ИПУ РАН, 2014. [Электронный ресурс]. Режим доступа: <http://www.mtas.ru/upload/library/tas2014/S2-PDF/2-10.pdf>

47. Орлов А.И. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 91. С. 133-162.

48. Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1-11.

49. Орлов А.И. Анализ экспертных упорядочений // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 112. С. 21–51.

50. Орлов А.И. Основные идеи солидарной информационной экономики - базовой организационно-экономической теории // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 112. С. 52 – 77.

51. Орлов А.И. Солидарная информационная экономика - экономика без рынка и денег // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 388 – 418.

52. Орлов А.И. Оценка погрешностей характеристик финансовых потоков инвестиционных проектов в ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. – № 109. С. 238–264.

53. Орлов А.И. Основные идеи статистики интервальных данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 94. С. 55-70.

54. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

55. Орлов А.И. Оценка инфляции по независимой информации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 108. С. 259–287.

Литература к разделу 4.2

1. Качество и успешность государственных политик и управления / Якунин В.И., Сулакшин С.С., Багдасарян В.Э. и др. – М.: Научный эксперт, 2012. – 488 с.

2. Кара-Мурза С.Г. Кризисное обществоведение. Часть первая. Курс лекций. – М.: Научный эксперт, 2011. – 464 с.;

3. Кара-Мурза С.Г. Кризисное обществоведение. Часть вторая. Курс лекций. – М.: Научный эксперт, 2012. – 384 с.

4. Якунин В.И., Багдасарян В.Э., Сулакшин С.С. Новые технологии борьбы с российской государственностью. – М.: Научный эксперт, 2009. – 424 с.

5. Колобов А.А., Омельченко И.Н., Орлов А.И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. – М.: Издательство «Экзамен», 2008. – 621 с.

6. Единство материального и духовного [Электронный ресурс]. URL: http://www.cemi.rssi.ru/news/science/index.php?ELEMENT_ID=8850 (дата обращения 28.07.2015).

7. Солидарная информационная экономика [Электронный ресурс]. Режим доступа: <http://forum.orlovs.pp.ru/viewtopic.php?f=2&t=570> (дата обращения 28.07.2015).

8. Публикации по солидарной информационной экономике [Электронный ресурс]. Режим доступа: <http://forum.orlovs.pp.ru/viewtopic.php?f=2&t=951> (дата обращения 28.07.2015).

9. Орлов А.И. Менеджмент: организационно-экономическое моделирование.— Ростов-на-Дону: Феникс, 2009. — 475 с.

10. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 576 с.

11. Орлов А.И. Организационно-экономическое моделирование: теория принятия решений. — М.: КноРус, 2015. — 568 с.

12. Орлов А.И. Организационно-экономическое моделирование : учебник. Ч.2. Экспертные оценки. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. — 486 с.

13. Сетевая экспертиза / Под ред. чл.-к. РАН Д.А. Новикова, проф. А.Н. Райкова. – М.: Этвес, 2019. – 168 с.

14. Друкер П.Ф. Новые реальности в правительстве и политике, в экономике и бизнесе, в обществе и мировоззрении: Пер. с англ. - М.: Бук Чембэр Интернэшнл, 1994. - 380 с.
15. Стиглиц Дж. Экономисты виноваты в кризисе, но есть шанс исправить дело // Газета «Ведомости». Аналитика. 20.08.2010. [Электронный ресурс]. Режим доступа: <http://www.ippnou.ru/article.php?idarticle=008769> (дата обращения 28.07.2015).
16. Кропоткин П.А. Взаимная помощь среди животных и людей как двигатель прогресса. Изд.2, доп. - М.: Книжный дом «ЛИБРОКОМ», 2011. - 280 с.
17. Аристотель. Экономика. Книги I – III // Вестник древней истории, 1969, № 3. Перевод с древнегреческого и латыни Г. А. Тароняна. [Электронный ресурс]. Режим доступа: <http://ancientrome.ru/antlitr/aristot/index.htm> (дата обращения 28.07.2015).
18. Орлов А.И. Организационно-экономическое моделирование, эконометрика и статистика в техническом университете // Вестник МГТУ им. Н.Э. Баумана. Сер. «Естественные науки». 2012. №1. С. 106-118.
19. Орлов А.И. Организационно-экономическое моделирование при решении задач управления хозяйственными единицами // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 87. С. 637–664.
20. Экономика предприятия: учебник / А.П. Аксенов, И.Э. Берзинь, Н.Ю. Иванова и др. ; под ред. С.Г. Фалько. — М. : КНОРУС, 2011. — 352 с.
21. Хруцкий К.С., Карпов А.В. Новгородское вече - к вопросу об изменении методологии изучения ключевого для Российской истории феномена: альтернатива биокосмологического подхода // *Biocosmology – neo-Aristotelism. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism*. Vol. 4. No. 4 (Autumn 2014). Pp. 502-540.
22. Ефремов И.А. Туманность Андромеды. - М.: Эксмо, 2009. - 768 с.
23. Агапова И.И. История экономической мысли. - М.: ЭКМОС, 1998. - 248 с.
24. Бэкон Ф. Сочинения в двух томах. Т. 2. - М.: Мысль, 1972. – 582 с.
25. Форд Г. Моя жизнь. Мои достижения. - М.: Попурри, 2009. - 352 с.
26. Лившиц В.Н. Системный анализ рыночного реформирования нестационарной экономики России: 1992–2013. - М.: ЛЕНАНД, 2013. - 640 с.
27. Кейнс Дж.М. Экономические возможности наших внуков (1931) // Вопросы экономики. 2009. № 6. С. 60–69.
28. Орлов А.И. Как нам реорганизовать хозяйство России? // Национальные интересы: приоритеты и безопасность. 2015. № 19 (304). С. 51-60.
29. Глушков В.М. Макроэкономические модели и принципы построения ОГАС. - М.: Статистика, 1975. - 160 с.
30. Бир Ст. Мозг фирмы. - М.: Радио и связь, 1993. - 416 с.
31. Великое кольцо. Журнал инновационного отдела при МГК СКМ РФ. Январь 2007. №1. С. 1 - 24. [Электронный ресурс]. Режим доступа: <http://rusprogressivelib.files.wordpress.com/2009/11/vk1.pdf> (дата обращения 28.07.2015).
32. Герасимов И. Интернет, Open Source и Открытое сетевое общество. 2008. [Электронный ресурс]. Режим доступа: <http://www.interface.ru/home.asp?artId=9835> (дата обращения 28.07.2015).
33. Теория управления организационными системами. [Электронный ресурс]. Режим доступа: <http://www.mtas.ru/> (дата обращения 28.07.2015).
34. Cockshott W. Paul and Cottrell Allin F. Information and Economics: A Critique of Hayek. November 1996. [Электронный ресурс]. Режим доступа: http://ricardo.ecn.wfu.edu/~cottrell/socialism_book/hayek_critique.pdf (дата обращения 28.07.2015).
35. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
36. Орлов А.И. Устойчивые экономико-математические методы и модели. Разработка и развитие устойчивых экономико-математических методов и моделей для модернизации управления предприятиями. – Saarbrücken: Lambert Academic Publishing, 2011. - 436 с.
37. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
38. Кокшотт П. Расчёт в натуральной форме: от Нейрата до Канторовича. 15 мая 2008. Перевод С. Маркова под ред. С. Голикова и Д. Левыкина [Электронный ресурс]. Режим доступа: http://ricardo.ecn.wfu.edu/~cottrell/socialism_book/hayek_critique.pdf (дата обращения 28.07.2015).
39. Файоль А. Общее и промышленное управление. – М.: Контролинг, 1992. – 111 с.
40. Фалько С.Г. Контролинг для руководителей и специалистов. - М.: Финансы и статистика, 2008. - 272 с.

41. Орлов А.И. Неформальная информационная экономика будущего // Неформальные институты в современной экономике России: Материалы Третьих Друкеровских чтений. - М.: Доброе слово: ИПУ РАН, 2007. – С.72-87.
42. Орлов А.И. Сети экспертов в неформальной информационной экономике будущего // Теория активных систем / Труды международной научно-практической конференции (17-19 ноября 2009 г., Москва, Россия). Том I. Общая редакция – В.Н. Бурков, Д.А. Новиков. – М.: ИПУ РАН, 2009. – С. 279 - 287.
43. Орлов А.И. Институциональные проблемы неформальной информационной экономики будущего // Информационная экономика: институциональные проблемы: Материалы Девярых Друкеровских чтений / Под ред. Р.М. Нижегородцева. – М.: Доброе слово, 2009. - С.179-184.
44. Орлов А.И. Неформальная информационная экономика будущего – базовая организационно-экономическая теория // Вестник Южно-Российского государственного технического университета (Новочеркасского политехнического института). Серия «Социально-экономические науки». 2010. No.2. С .55-67.
45. Орлов А.И. Основные идеи неформальной информационной экономики будущего // ЭТАП: Экономическая Теория, Анализ, Практика. 2010, № 1. С.89-105.
46. Орлов А.И. Подходы неформальной информационной экономики будущего к управлению развитием крупномасштабных систем // IDO science. 2010. №2. С. 17-26.
47. Орлов А.И. Аристотель и неформальная информационная экономика будущего // Biocosmology – neo-Aristotelism. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. Vol.2. No.3. Summer, 2012. Pp. 150-164.
48. Орлов А.И. Основные идеи солидарной информационной экономики – новой базовой организационно-экономической теории // Научный эксперт. 2013. № 1 - 2. С. 69-81.
49. Orlov A. I. Functionalist-Organic Information Economy – the Organizational-Economic Theory of Innovation Development // Biocosmology – neo-Aristotelism. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. → Vol.3. No.1 (Winter 2013). – Pp. 52-59.
50. Орлов А.И. Проблемы методологии государственной политики и управления в неформальной информационной экономике будущего // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 88. С. 653-679
51. Орлов А.И. Солидарная информационная экономика – инструмент реализации национальных интересов // Национальные интересы: приоритеты и безопасность. – 2013. – № 33 (222). – С.2–10.
52. Орлов А.И. Солидарная информационная экономика как экономическая составляющая государственной идеологии России // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 98. С. 1149-1160
53. Орлов А.И. Солидарная информационная экономика – составная часть теории управления социально-экономическими системами // XII Всероссийское совещание по проблемам управления ВСПУ-2014, Москва, 16-19 июня 2014 г.: Труды. [Электронный ресурс] М.: Институт проблем управления им. В.А. Трапезникова РАН, 2014, 9616 с. Электрон. текстовые дан. (1074 файл: 537 МБ), 1 электрон. опт. диск (DVD-ROM), ISBN 978-5-91450-151-5. Номер государственной регистрации: 0321401153. – С.5774-5784.
54. Орлов А.И. Экономическая составляющая государственной идеологии России - солидарная информационная экономика // Государственная идеология и современная Россия. Материалы Всероссийской научно-общественной конференции. Москва, 28 марта 2014 г. - М.: Наука и политика, 2014. 280 с. – С.183 – 193.
55. Орлов А.И. Солидарная информационная экономика — организационно-экономическая теория инновационного развития России. Инженерный журнал: наука и инновации. 2014. № 2 (26). С. 13. URL: <http://engjournal.ru/catalog/indust/hidden/1207.html>
56. Орлов А.И. Солидарная информационная экономика – составная часть теории управления крупномасштабными социально-экономическими системами // Управление развитием крупномасштабных систем MLSD'2014. Сборник научных трудов. Под общей редакцией С.Н. Васильева, А.Д. Цвиркуна. Москва, ИПУ РАН, 2014. С. 67-76.

Литература к разделу 4.3

1. Бурков В.Н., Новиков Д.А. Как управлять проектами. - М.: Синтег, 1997. - 188 с.
2. Новиков Д.А., Матвеев А.А., Цветков А. В. Модели и методы управления портфелями проектов. - М.: ПМСОФТ, 2005. - 206 с.
3. Новиков Д.А., Нижегородцев Р.М., Гонтарева И.В. Управление проектами. - М.: Либроком, 2009. — 384 с.
4. ГОСТ Р 54869—2011. Проектный менеджмент. Требования к управлению проектом.

5. ГОСТ Р 54870—2011. Проектный менеджмент. Требования к управлению портфелем проектов.
6. ГОСТ Р 54871—2011. Проектный менеджмент. Требования к управлению программой.
7. Орлов А.И. Менеджмент: организационно-экономическое моделирование. – Ростов-на-Дону: Феникс, 2009. – 475 с.
8. Орлов А.И. Инновационная деятельность: организационно-экономическое обеспечение и Интернет-аукционы. - Проблемы информационной экономики. Выпуск VII. Стратегия инновационного развития российской экономики: Сб. научных трудов / Под ред. Р.М. Нижегородцева. - М.: Книжный дом «ЛИБРОКОМ», 2008. – С.325-345.
9. Виленский П.Л., Смоляк С.А., Лившиц В.Н. Оценка эффективности инвестиционных проектов: теория и практика. 4-е изд., перераб. и доп. – М.: Дело, 2008. – 1104 с.
10. Орлов А.И. Новая парадигма математических методов экономики // Экономический анализ: теория и практика. – 2013. – № 36 (339). – С.25–30.
11. Бендиков М.А. Оценка реализуемости инновационного проекта // Менеджмент в России и за рубежом. 2001. №2. С. 27-43.
12. Хрусталёв О.Е., Хрусталёв Ю.Е. Инструментальные методы оценки реализуемости наукоемкого инвестиционного проекта // Экономический анализ: теория и практика. 2011. № 27(234). С.8-18.
13. Неволин И.В., Хрусталёв О.Е., Хрусталёв Ю.Е. Методология оценки финансовой значимости и реализуемости инновационных проектов создания интеллектуальной продукции // Финансовая аналитика: проблемы и решения. 2013. №11(149). С.39-45.
14. Хрусталёв Е.Ю., Хрусталёв О.Е. Финансовая устойчивость наукоемкого предприятия как фактор оценки реализуемости инновационного проекта // Национальные интересы: приоритеты и безопасность. 2013. №33(222). С.16-23.
15. Проектирование интегрированных производственно- корпоративных структур: эффективность, организация, управление / Колобов А.А., Омельченко И.Н., Орлов А.И. и др. / Под ред. А.А. Колобова, А.И. Орлова. Научное издание. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2006. — 728 с.
16. Колобов А. А., Омельченко И. Н., Орлов А. И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. Учебник для вузов. — М.: Экзамен, 2008. — 621 с.
17. Хрусталёв Е.Ю., Соколов Н.А., Хрусталёв О.Е. Концепция оценки и управления риском при реализации инновационных проектов создания интеллектуальной продукции // Экономический анализ: теория и практика. 2013. № 44(347). С.2-13.
18. Хрусталёв Е.Ю., Стрельникова И.А. Институциональный метод повышения реализуемости наукоемких инвестиционных проектов // Экономический анализ: теория и практика. 2011. № 3(210). С.2-8.
19. Орлов А.И., Цисарский А.Д. Особенности оценки рисков при создании ракетно-космической техники // Национальные интересы: приоритеты и безопасность. – 2013. – №43(232). – С.37 – 46.
20. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
21. Орлов А.И. Организационно-экономическое моделирование: учеб. Ч.2. Экспертные оценки. М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. 486 с.
22. Орлов А.И. Организационно-экономическое моделирование: теория принятия решений. М.: КноРус, 2011. 568 с.
23. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 574 с.
24. Цисарский А.Д. Повышение эффективности реализации проектов по созданию перспективных образцов ракетно-космической техники на основе концепции Requirements Engineering // Национальные интересы: приоритеты и безопасность. 2013. №31 (220). С.25-29.
25. Вторые Чарновские чтения. Сборник трудов. Материалы II международной научной конференции по организации производства. Москва, 7 – 8 декабря 2012 г. – М.: НП «Объединение контроллеров», 2013. – 183 с.
26. Луценко Е.В., Орлов А.И. О развитии системной нечеткой интервальной математики // Философия математики: актуальные проблемы. Математика и реальность. Тезисы Третьей всероссийской научной конференции; 27-28 сентября 2013 г. / Редкол.: Бажанов В.А. и др. – Москва, Центр стратегической конъюнктуры, 2013. – С.190–193.
27. Луценко Е.В., Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>.

Литература к разделу 4.4

1. Орлов А.И. Менеджмент: организационно-экономическое моделирование. – Ростов-на-Дону: Феникс, 2009. – 475 с.
2. Цисарский А.Д. Повышение эффективности реализации проектов по созданию перспективных образцов ракетно-космической техники на основе концепции Requirements Engineering // Национальные интересы: приоритеты и безопасность. 2013. №31 (220). С.25-29.
3. Ebert C., Dumke R. Software Measurement.- Heidelberg: Springer Verlag, 2007.
4. IAG Business Analysis Benchmark (103 Unternehmen, durchschnittlich 3 Mio USA\$ Projektumfang). –New Castle DE: IAG Consulting, 2008.
5. Standish Group: Solutions for Enterprise Project and Portfolio Management. - West Yarmouth, USA: The Standish Group International, 2009.
6. Фалько С.Г., Цисарский А.Д., Баев Г.О. Управление себестоимостью и прогнозирование цен создания РКТ // Контроллинг, 2013, № 1(47).
7. Орлов А.И., Цисарский А.Д. Особенности оценки рисков при создании ракетно-космической техники // Национальные интересы: приоритеты и безопасность. – 2013. – №43(232). – С.37 – 46.
8. Волков В.А., Орлов А.И. Организационно-экономические подходы к оценке реализуемости проектов по созданию ракетно-космической техники // Экономический анализ: теория и практика. – 2014. – № 11 (362). – С.41–47.
9. Волков В.А., Орлов А.И. Организационно-экономические подходы к оценке реализуемости инновационно-инвестиционных проектов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 181 – 202. – IDA [article ID]: 0971403013. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/13.pdf>
10. Баев Г.О. Использование метода целевых издержек для управления себестоимостью ракетно-космической техники // Инженерный журнал: наука и инновации, 2014, вып. 1. URL: <http://engjournal.ru/catalog/indust/hidden/1171.html>
11. Бриль А. Р. Функционально-стоимостный анализ в экономических расчётах. — Л.: Изд-во ЛГУ, 1989. — 148 с.

Литература к разделу 4.5

1. Орлов А.И. Теория принятия решений. – М.: Экзамен, 2006. – 576 с.
2. Орлов А.И., Пугач О.В. Подходы к общей теории риска // Управление большими системами. Вып. 40. 2012. С.49-82.
3. Орлов А.И. Современное состояние контроллинга рисков / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №04(098). С. 32 – 64. – IDA [article ID]: 0981404003. – Режим доступа: <http://ej.kubagro.ru/2014/04/pdf/03.pdf>
4. Орлов А.И. Менеджмент: организационно-экономическое моделирование. – Ростов-на-Дону: Феникс, 2009. - 475 с.
5. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.
6. Орлов А.И., Рухлинский В.М., Шаров В.Д. Экономическая оценка рисков при управлении безопасностью полетов // Материалы I Международной конференции «Стратегическое управление и контроллинг в некоммерческих и публичных организациях: фонды, университеты, муниципалитеты, ассоциации и партнерства»: выпуск №1 / Под научн. ред. С.Л. Байдакова и С.Г. Фалько. – М.: НП «ОК», 2011. – С.108-114.
7. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. – М.: КноРус, 2010. – 192 с.
8. Орлов А.И. Организационно-экономическое моделирование. Ч.2. Экспертные оценки. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. – 486 с.
9. Пугач О.В. Математические методы оценки рисков // Заводская лаборатория. Диагностика материалов. 2013. Т.79. №7. С.64–69.
10. Орлов А.И., Цисарский А.Д. Особенности оценки рисков при создании ракетно-космической техники // Национальные интересы: приоритеты и безопасность. – 2013. – №43(232). – С.37 – 46.
11. Орлов А.И., Семенов П.М., Жихарев В.Н., В.А. Цупин В.А. Методология оценки рисков реализации инновационных проектов // Управление большими системами. Материалы Международной научно-практической конференция (22-26 сентября 1997 г., Москва, Россия). - М.: СИНТЕГ, 1997. – С. 109 – 109.

12. Вологжанина С.А., Орлов А.И. Об одном подходе к оценке рисков для малых предприятий (на примере выполнения инновационных проектов в ВУЗах). - Подготовка специалистов в области малого бизнеса в высшей школе. Сборник научных статей. - М.: Изд-во ООО "ЭЛИКС +", 2001. - С.40-53.
13. Орлов А.И. Эконометрика. - М.: Экзамен, 2004. - 576 с.
14. Хрусталев С.А., Орлов А.И., Шаров В.Д. Оценка эффективности управленческих решений в автоматизированной системе прогнозирования и предотвращения авиационных происшествий // Известия Самарского научного центра Российской академии наук. 2012. Том 14. №4(2). С.535-539.
15. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.
16. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2014. - №06(100). С. 1 - 30. - IDA [article ID]: 1001406001. - Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/01.pdf>

Литература к разделу 4.6

1. Муравьева В.С., Орлов А.И. Организационно-экономические проблемы прогнозирования на промышленном предприятии/ Управление большими системами. Выпуск 17. М.: ИПУ РАН, 2007. С.143-158.
2. Орлов А.И. Менеджмент: организационно-экономическое моделирование. - Ростов-на-Дону: Феникс, 2009. - 475 с.
3. Орлов А.И. Эконометрика. 3-е изд. — М.: Экзамен, 2004. — 576 с.
4. Карминский А.М., Фалько С.Г., Жевага А.А., Иванова Н.Ю. Контроллинг: учебник/ под ред. А.М. Карминского, С.Г. Фалько. 3-е изд., дораб. - М.: ИД «ФОРУМ»: ИНФРА-М, 2013. - 336 с.
5. Мухин В.В. О контроллинге научной деятельности / В.В. Мухин, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2014. - №06 (100). С. 256 - 275. - IDA [article ID]: 1001406013. - Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/13.pdf>
6. Орлов А.И. Эконометрическая поддержка контроллинга // Контроллинг. 2002. №1. С.42-53.
7. Колобов А.А., Омельченко И.Н., Орлов А.И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. — М.: Экзамен, 2008. — 621 с.
8. Глущенко В.В. Менеджмент: системные основы: 2-е изд., доп. и испр. - Железнодорожный, Моск. обл.: ТОО НПЦ «Крылья», 1998. - 224 с.
9. Орлов А.И. Организационно-экономическое моделирование: в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
10. Орлов А.И. Статистические методы прогнозирования // Малая российская энциклопедия прогнозистики. - М.: Институт экономических стратегий, 2007. - С.148-153.
11. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
12. Неуймин Я.Г. Модели в науке и технике. История, теория, практика. - Л.: Наука, 1984. - 190 с.
13. Орлов А.И. О реальных возможностях бутстрапа как статистического метода // Заводская лаборатория. - 1987. - Т.53. - №10. - С. 82-85.
14. Орлов А.И. Непараметрический метод наименьших квадратов с периодической составляющей // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №1. С.65-75.
15. Орлов А.И. Организационно-экономическое моделирование: в 3 ч. Ч.2. Экспертные оценки. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. - 486 с.
16. Орлов А.И. Сценарии социально-экономического развития России до 2007 г. // Обозреватель-Observer. 1999. №10 (117). С.47-50.
17. Модели и методы анализа и синтеза сценариев развития социально-экономических систем / Под ред. В.В. Шульца, В.В. Кульбы; Центр исследования проблем безопасности РАН, Ин-т проблем управления им. В.А. Трапезникова РАН. В двух книгах. Кн. 1. - М.: Наука, 2012. - 304 с.
18. Модели и методы анализа и синтеза сценариев развития социально-экономических систем / Под ред. В.В. Шульца, В.В. Кульбы; Центр исследования проблем безопасности РАН, Ин-т проблем управления им. В.А. Трапезникова РАН. В двух книгах. Кн. 2. - М.: Наука, 2012. - 358 с.
19. Науман Э. Принять решение - но как?: Пер. с нем. - М.: Мир, 1987. - 198 с.
20. Хрусталев С.А., Орлов А.И., Шаров В.Д. Математические методы оценки эффективности управленческих решений // Заводская лаборатория. Диагностика материалов. 2013. Т.79. №11. С. 67-72.

21. Орлов А.И., Савинов Ю.Г., Богданов А.Ю. Экспертные технологии и их применение при оценивании вероятностей редких событий // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №3. С.63-69.
22. Муравьева В.С. Разработка организационно-экономических инструментов и методов управления промышленными предприятиями на основе эконометрического прогнозирования. Автореферат диссертации на соискание ученой степени кандидата экономических наук. – М.: МГТУ им. Н.Э. Баумана, 2011. – 16 с.
23. Сетевая экспертиза / Под ред. Д.А. Новикова, А.Н. Райкова. – М.: Эгвес, 2010. – 168 с.
24. Райков А.Н. Ситуационная комната для поддержки корпоративных решений // Открытые системы. 1999. № 07-08. Режим доступа: <http://www.osp.ru/os/1999/07-08/179889/>
25. Соколов А.В. Форсайт: взгляд в будущее // Форсайт. 2007. № 1 (1). С. 8-15.
26. Федеральный закон Российской Федерации от 28 июня 2014 г. N 172-ФЗ «О стратегическом планировании в Российской Федерации».

Литература к разделу 4.7

1. Математические модели в экономике. Расчет индекса инфляции / Орлов А.И., Балашов В.В., Куроптев О.В., Канакова Е.М., Рафальская А.С. - М.: Изд-во Московского государственного института электроники и математики (технического ун-та), 1994. - 32 с.
2. Орлов А.И. Как использовать индекс инфляции? // Международная газета «Наука и технология в России». 1995. № 9-10 (15-16). С.16-17.
3. Орлов А.И. Нас ограбили на триллион долларов (беседу вел В.С. Кожемяко) // Газета «Правда». 1996. 13 марта. № 38 (27684). С.1-1.
4. Орлов А.И. Насколько понизился наш уровень жизни? // Диалог. 1996. № 4. С.43-43.
5. Орлов А.И., Иванова И.Г., Точенная Н.С. Инфляция: вчера, сегодня, завтра // Международная газета «Наука и технология в России». 1996. № 1 (17). С.9-9.
6. Орлов А.И. Какова цена «реформ»? // Газета «Правда». 1997. 22-29 августа. № 32 (27803). С.2-2.
7. Расчет, прогнозирование и применение индексов инфляции на основе независимой информации / Орлов А.И., Жихарев В.Н., Цупин В.А., Васюкевич В.А., Балашов В.В., Иванова И.Г., Канакова Е.М., Куроптев О.В., Рафальская А.С. // Управление большими системами. Материалы Международной научно-практической конференции (22-26 сентября 1997 г., Москва, Россия). – М.: СИНТЕГ, 1997. С.81-81.
8. Орлов А.И. Экономическое положение населения России на пороге XXI века // Россия на пороге XXI века (методологический аспект изучения современных процессов). Тезисы научно-методической конференции. - М.: МГИЭМ (ту), 1997. С.48-49.
9. Динамика цен и уровень жизни / Орлов А.И., Жихарев В.Н., Цупин В.А., Иванова И.Г. // Россия сегодня: общество, культура, государство, человек. Тезисы докладов Межвузовской научно-теоретической конференции. – М.: МГИЭМ (ту), 1998. С. 108-109.
10. Орлов А.И., Жихарев В.Н., Цупин В.А. Анализ динамики цен на продовольственные товары в Москве и Московской области // Научные труды Рижского института мировой экономики. Вып.2. - Рига: РИМЭ, 1998. С.19-25.
11. Как оценивать уровень жизни? (На примере московского региона) / Орлов А.И., Жихарев В.Н., Цупин В.А., Балашов В.В. // Обозреватель-Observer. 1999. № 5 (112). С. 80-83.
12. Федосеев В.Н., Орлов А.И. За что нас покупают (состояние рыночной мотивации труда в России) // Российское предпринимательство. 2000. № 6. С.10-19.
13. Орлов А.И. Сколько в России богачей? // Газета «Правда». 2001. 18 января. № 6 (28269). С.1 - 1.
14. Орлов А.И. Сколько богатых в России? // Газета «Дуэль». 2002. 25 июня. №26 (271). С.4-4.
15. Орлов А.И. Эконометрика: Учебник для вузов. — М.: Экзамен, 2002 (1-е изд.), 2003 (2-е изд.), 2004 (3-е изд.). — 576 с.
16. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах экономики и управления / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №04(098). С. 105 – 125. – IDA [article ID]: 0981404008. – Режим доступа: <http://ej.kubagro.ru/2014/04/pdf/08.pdf>
17. Орлов А.И., Орлова Л.А. Гуляй, Россия от рубля... и ниже. Интервальная оценка инфляции по независимой информации // Российское предпринимательство. 2004. № 10. С.44-49.
18. Муравьева В.С. Разработка организационно-экономических инструментов и методов управления промышленными предприятиями на основе эконометрического прогнозирования. Автореф. дисс. ... кандидата экономических наук. - М.: МГТУ им. Н.Э. Баумана, 2011. - 16 с.
19. Орлов А.И. Цена рубля советского и рубля антисоветского // Газета «Правда». 2008. 29 февраля. № 22.

20. Орлов А.И. Организационно-экономическое моделирование процессов управления промышленными предприятиями в условиях рисков инфляции. - Стратегическое планирование и развитие предприятий. Секция 4 / Материалы Девятого всероссийского симпозиума. Москва, 15-16 апреля 2008 г. Под ред. чл.-корр. РАН Г.Б. Клейнера. – М.: ЦЭМИ РАН, 2008. – С.124–126.
21. Куликова С.Ю., Муравьева В.С., Орлов А.И. Контроллинг уровня потребительских цен и прожиточного минимума // Материалы II Международной научно-практической конференции по контроллингу. / Под науч. ред. С.Г. Фалько. – М.: НП «Объединение контроллеров», 2012. - С. 37 – 47.
22. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. – Ростов-на-Дону: Феникс, 2009. - 572 с.
23. Федоров А.Г. Исследование динамики потребительских цен [Электронный ресурс] Молодежный научно-технический вестник, № 08, август 2013. URL: <http://sntbul.bmstu.ru/doc/605856.html>
24. Гаврилова В. Д., Евдокимов А. А. Исследование динамики потребительских цен [Электронный ресурс] Молодежный научно-технический вестник, № 01, январь 2014. URL: <http://sntbul.bmstu.ru/doc/687936.html>
25. Добротворский Н., Седов А. Курс холодильника к кошельку: живем, как в 1985 году! / Газета «Комсомольская правда». 2003. 10 декабря.
26. Орлов А.И., Орлова Л.А. Тенденции макроэкономического развития России // Экономика XXI века. 2002. № 12. С. 27 – 37.
27. Орлов А.И., Орлова Л.А. Тенденции макроэкономического развития России // Современное управление. 2003. № 7. С. 7 - 16.
28. Орлов А.И. Теория принятия решений. – М.: Экзамен, 2006. – 576 с.
29. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
30. Орлов А.И. Размытые цены. Нечисловая экономика и управление инвестиционным процессом // Российское предпринимательство. 2001. № 12. С.103-108.
31. Проскурина О.Ю. Анализ результатов измерения индекса инфляции в конце 2006 г. [Электронный ресурс] Доклады Научный семинар Лаборатории экономико-математических методов в контроллинге МГТУ им. Н.Э. Баумана за I полугодие 2007 года. URL: <http://ibm.bmstu.ru/nil/nsem0701.html>
32. Панфилова Ю., Угодников К. Как вы считаете? // Итоги, 2005, 14 ноября, №46 (492).
33. Рост цен в России [Электронный ресурс] URL: <http://forum.orlovs.pp.ru/viewtopic.php?f=2&t=660&p=7770#p7770>
34. Ложь официальной статистики [Электронный ресурс] URL: <http://forum.orlovs.pp.ru/viewtopic.php?f=2&t=1051&p=7708#p7708>
35. Потребительская корзина и прожиточный минимум по Оршански [Электронный ресурс] URL: <http://forum.orlovs.pp.ru/viewtopic.php?f=2&t=1453>
36. Национальное счетоводство / Под ред. Г.Д. Кулагиной. - М.: Финансы и статистика, 1997. - 448 с.

Литература к разделу 4.8

1. Орлов А.И. Последствия принятия решений для научно-технического и экономического развития // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 355 – 387.
2. Цисарский А.Д. Кадровое обеспечение процессов управления инновационной деятельностью на предприятиях ракетно-космической отрасли // Инновации в менеджменте. 2014. № 2 (2). С. 72 - 79.
3. Профессиональные стандарты. [Электронный ресурс]. URL: <http://base.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=157436;fld=134;dst=1000000001,0;rnd=0.10866944314321314> (дата обращения 30.12.2015)).
4. Мухин Ю.И. "Антиаполлон". Лунная афера США. - М.: Яуза, Эксмо, 2005. - 432 с.
5. Орлов А.И. Основные идеи солидарной информационной экономики - базовой организационно-экономической теории // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 112. С. 52 – 77.
6. Орлов А.И. Солидарная информационная экономика - экономика без рынка и денег // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 388 – 418.
7. Альтшуллер Г.С. Найти идею: Введение в ТРИЗ — теорию решения изобретательских задач. - М.: Альпина Паблишер, 2014. - 319 с.
8. Пойа Д. Как решать задачу. Изд. 2-е. - М.: Учпедгиз, 1961. - 208 с.
9. Пойа Д. Математическое открытие. Решение задач: основные понятия, изучение и преподавание. - М.: Наука, 1970. - 452 с.

10. Пойа Д. Математика и правдоподобные рассуждения. Изд. 2-е, испр.- М.: Наука, 1975. - 464 с.
11. Орлов А.А. Тайны и секреты компьютера. - М.: «Горячая Линия – Телеком», 2002. – 604 с.
12. Юзбекова И. Российский искусственный интеллект обманул жюри теста Тьюринга // РБК - РИА "Росбизнесконсалтинг" 09.06.2014 [Электронный ресурс]. URL: [http://top.rbc.ru/society/09/06/2014/929457.shtml#xtor=AL-\[internal traffic\]-\[rbc.ru\]-\[main body\]-\[item 11\]](http://top.rbc.ru/society/09/06/2014/929457.shtml#xtor=AL-[internal traffic]-[rbc.ru]-[main body]-[item 11]) (дата обращения 03.01.2016)/
13. Орлов А.И. Миф XX века: искусственный интеллект // Подводная лодка, 2003. № 11. С.102-103.
14. Орлов А.И. Искусственный интеллект или мощный калькулятор? // Магия ПК. 2003. № 3(59). С.42-45.
15. Орлов А.А. Компьютеры будущего. [Электронный ресурс] URL: <http://orlovs.pp.ru/comp.php> (дата обращения 01.09.2015).
16. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 163 – 195.
17. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 114. С. 27 – 41.
18. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф. С.Г. Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с. В РИНЦ: <http://elibrary.ru/item.asp?id=23209923>
19. Лындина М.И., Орлов А.И. Методы прогнозирования для ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 196–221.

Литература к разделу 4.9

1. Орлов А.И. Последствия принятия решений для научно-технического и экономического развития // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 355 – 387.
2. Цисарский А.Д. Кадровое обеспечение процессов управления инновационной деятельностью на предприятиях ракетно-космической отрасли // Инновации в менеджменте. 2014. № 2 (2). С. 72 - 79.
3. Мухин В.В., Орлов А.И. О контроллинге научной деятельности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 1222-1237.
4. Мухин В.В., Орлов А.И. Совершенствование организационных структур и контроллинг персонала на предприятиях типа "Научно-исследовательский институт" ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 109. С. 265–296.
5. Орлов А.И. Прогноз развития информационно-коммуникационных технологий // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 116. С. 435–461. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/31.pdf>
6. Указ Президента РФ №597 от 7 мая 2012 г. «О мероприятиях по реализации государственной социальной политики» [Электронный ресурс]. Режим доступа: <http://www.rosmintrud.ru/docs/president/ukaz/37> (дата обращения 17.02.2016).
7. Постановление Правительства РФ №23 от 22 января 2013 г. «О Правилах разработки, утверждения и применения профессиональных стандартов» [Электронный ресурс]. Режим доступа: <http://www.rosmintrud.ru/docs/government/103> (дата обращения 17.02.2016).
8. Федеральный закон N 236-ФЗ от 03.12.2012 г. «О внесении изменений в Трудовой Кодекс Российской Федерации и Статью 1 Федерального закона "О техническом регулировании"» [Электронный ресурс]. Режим доступа: <http://base.consultant.ru/cons/CGI/online.cgi?req=doc;base=LAW;n=138556> (дата обращения 17.02.2016).
9. Профессиональные стандарты [Электронный ресурс]. Режим доступа: <http://www.sut.ru/home/hide/professionalnye-standarty#ps1> (дата обращения 18.02.2016).
10. Профессиональные стандарты. [Электронный ресурс]. Режим доступа: <http://base.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=157436;fld=134;dst=1000000001,0;rnd=0.10866944314321314> (дата обращения 17.02.2016).
11. Орлов А.И. Новая парадигма разработки и преподавания организационно-экономического моделирования, эконометрики и статистики в техническом университете // Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С.131-144.

12. Орлов А.И. Новая парадигма прикладной статистики // Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С.206-217.
13. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Т. 78. №1, часть I. С.87-93.
14. Орлов А.И. Организационно-экономическое моделирование, эконометрика и статистика в техническом университете // Вестник МГТУ им. Н.Э. Баумана. Сер. «Естественные науки». 2012. №1. С. 106-118.
15. Орлов А.И. Новая парадигма математической статистики // Материалы республиканской научно-практической конференции «Статистика и её применения – 2012». Под редакцией профессора А.А. Абдушукурова. – Ташкент: НУУз, 2012. – С.21-36.
16. Орлов А.И. Новая парадигма организационно-экономического моделирования, эконометрики и статистики // Вторые Чарновские Чтения. Сборник тезисов. Материалы II международной научной конференции по организации производства. Москва, 7 – 8 декабря 2012 г. – М.: НП «Объединение контролеров», 2012. – С. 116-120.
17. Орлов А.И. Новая парадигма математических методов экономики // Экономический анализ: теория и практика. – 2013. – № 36 (339). – С.25–30.
18. Орлов А.И. Организационно-экономическое моделирование, эконометрика и статистика при решении задач экономики и организации производства // Инженерный журнал: наука и инновации, 2014, вып. 1. URL: <http://engjournal.ru/catalog/indust/hidden/1198.html>
19. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах экономики и управления / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №04(098). С. 105 – 125. – IDA [article ID]: 0981404008. – Режим доступа: <http://ej.kubagro.ru/2014/04/pdf/08.pdf> , 1,312 у.п.л., импакт-фактор РИНЦ=0,346
20. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах управления // Труды X Международной конференции «Идентификация систем и задачи управления» SICPRO '15. Москва, 26-29 января 2015 г. М.: Институт проблем управления им. В.А. Трапезникова, 2015. 1484 с. [Электронный ресурс]. - Электронные текстовые дан. (121 файл, 68,5 Мб). - М.: ИРУ РАН, 2015. - 1 электронно-оптический диск (CD-ROM). - Системные требования: Pentium 4. Acrobat PReader 6.0 и выше. - Загл. с экрана. - ISBN 978-5-91450-162-1. - С.34 - 42.
21. Орлов А.И. Новая парадигма математических методов исследования // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №.7 С. 5-5.
22. Орлов А.И. Основные черты новой парадигмы математической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 90. С. 45-71.
23. Орлов А.И. Высокие статистические технологии // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №11. С.55-60.
24. Орлов А.И. О высоких статистических технологиях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 105. С. 14 – 38.
25. Орлов А.И. Эконометрика: Учебник для вузов. — М.: Экзамен, 2002 (1-е изд.), 2003 (2-е изд.), 2004 (3-е изд.). — 576 с.
26. Орлов А.И. Организационно-экономическое моделирование: теория принятия решений. Учеб. для вузов. М. : КноРус, 2011. 568 с.
27. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
28. Лындина М.И., Орлов А.И. Методы прогнозирования для ракетно-космической промышленности // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 196–221.
29. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
30. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.
31. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
32. Орлов А.И. Устойчивые экономико-математические методы и модели. Saarbrücken (Germany), Lambert Academic Publishing, 2011. 436 с.
33. Орлов А.И. Сценарии социально-экономического развития России до 2007 г. // Обозреватель-Observer. 1999. № 10 (117). С.47-50.

34. Орлов А.И. Сценарии социально-экономического развития России на период до 2007 г. и в XXI в. // Экономика XXI века. 2000. № 8. С. 3-22. [Электронный ресурс]. Режим доступа: <http://orlovs.pp.ru/econ.php#e1p10> (дата обращения 18.02.2016).
35. Орлов А.И. Сценарии социально-экономического развития России в XXI в. // Обозреватель-Observer. 2000. № 10-11. С. 82-82.
36. Орлов А.И. Грядущая смута 2012 года // Вестник Академии Прогнозирования (Исследований Будущего). № 12. 2004 : Труды Академии прогнозирования. Выпуск № 9. 2004. - С.42 - 45.
37. Проектирование интегрированных производственно- корпоративных структур: эффективность, организация, управление / Колобов А.А., Омельченко И.Н., Орлов А.И. и др. / Под ред. А.А. Колобова, А.И. Орлова. Научное издание. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2006. — 728 с.
38. Колобов А.А., Омельченко И.Н., Орлов А.И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. Учебник для вузов. — М.: Экзамен, 2008. — 621 с.
39. Орлов А.И. Менеджмент: организационно-экономическое моделирование.— Ростов-на-Дону: Феникс, 2009. — 475 с.
40. Орлов А.И. Аристотель и неформальная информационная экономика будущего // *Biocosmology – neo-Aristotelism. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. Vol.2. No.3. Summer, 2012*. Pp. 150-164.
41. Orlov A. I. Functional-Organic Information Economy – the Organizational-Economic Theory of Innovation Development // *Biocosmology – neo-Aristotelism. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. – Vol.3. No.1 (Winter 2013)*. – Pp. 52-59.
42. Орлов А.И. Проблемы методологии государственной политики и управления в неформальной информационной экономике будущего // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 88. С. 653-679.
43. Орлов А.И. Солидарная информационная экономика как экономическая составляющая государственной идеологии России // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 98. С. 1149-1160.
44. Орлов А.И. Основные идеи солидарной информационной экономики - базовой организационно-экономической теории // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 112. С. 52 – 77.
45. Орлов А.И. Солидарная информационная экономика - экономика без рынка и денег // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 113. С. 388 – 418.
46. Орлов А.И. Организационно-экономическое моделирование: учебник : в 3 ч. Ч.2. Экспертные оценки. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. — 486 с.
47. Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1-11.
48. Орлов А.И. Сети экспертов в неформальной информационной экономике будущего // Теория активных систем / Труды международной научно-практической конференции (17-19 ноября 2009 г., Москва, Россия). Том I. Общая редакция – В.Н. Бурков, Д.А. Новиков. – М.: ИПУ РАН, 2009. – С. 279 - 287. [Электронный ресурс]. Режим доступа: <http://elibrary.ru/item.asp?id=22014052> (дата обращения 18.02.2016).
49. Beer Stafford. World in torment. 1992. [Электронный ресурс]. Режим доступа: http://www.ototsky.mgn.ru/it/papers/world_in_torment.pdf (дата обращения 18.02.2016).
50. Хруцкий К.С., Карпов А.В. Новгородское вече - к вопросу об изменении методологии изучения ключевого для Российской истории феномена: альтернатива биокосмологического подхода // *Biocosmology – neo-Aristotelism. Bilingual Electronic Journal of Universalizing Scientific and Philosophical Research based upon the Original Aristotelian Cosmological Organicism. Vol. 4. No. 4 (Autumn 2014)*. Pp. 502-540.
51. Носовский Г.В., Фоменко А.Т. Как было на самом деле. Реконструкция подлинной истории. - М.: АСТ, 2015. - 767 с. [Электронный ресурс]. Режим доступа: <http://chronologia.org/xpron7/index.html> (дата обращения 18.02.2016).
52. Хруцкий К.С., Смирнова О.А. Вопросы нравственного хозяйства в свете русской философско-религиозной традиции на рубеже XIX-XX столетий: с позиций Биокосмологии // *Biocosmology – neo-Aristotelism, Vol.4, No.3 (Summer 2014)*, с. 224–271. [Электронный ресурс]. Режим доступа: <http://en.biocosmology.ru/electronic-journal-biocosmology---neo-aristotelism> (дата обращения 18.02.2016).
53. Глушков В.М. Макроэкономические модели и принципы построения ОГАС. - М.: Статистика, 1975. - 160 с.

54. Бир Ст. Мозг фирмы. - М.: Радио и связь, 1993. - 416 с.
55. Орлов А.И. Принятие решений. Теория и методы разработки управленческих решений. Учебное пособие для вузов. - М.: ИКЦ «МарТ»; Ростов н/Д: Издательский центр «МарТ», 2005. - 496 с.
56. Орлов А.И. Теория принятия решений. Учебник для вузов. — М.: Экзамен, 2006. — 576 с.
57. Доклад о мировом развитии 2016 «Цифровые дивиденды». Обзор. - Вашингтон: Всемирный банк, 2016 [Электронный ресурс]. Режим доступа: <http://www.inesnet.ru/wp-content/uploads/2016/01/World0developm0I0dividends0overview.pdf> (дата обращения 18.02.2016).
58. Винер Н. Кибернетика и общество / Пер. с англ. Е.Г. Панфилова; Общ. ред. и предисловие Э.Я. Кольмана. - М.: Издательство иностранной литературы, 1958. - 199 с.

Литература к разделу 5.1

1. Ткачев А.Н. Гуманистическая экономика и цели региональной администрации / А.Н. Ткачев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(006). С. 214 – 227. – IDA [article ID]: 0060404018. – Режим доступа: <http://ej.kubagro.ru/2004/04/pdf/18.pdf>, 0,875 у.п.л.
2. Ткачев А.Н. Качество жизни населения, как интегральный критерий оценки эффективности деятельности региональной администрации / А.Н. Ткачев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №02(004). С. 171 – 185. – IDA [article ID]: 0040402014. – Режим доступа: <http://ej.kubagro.ru/2004/02/pdf/14.pdf>, 0,938 у.п.л.
3. Ткачев А.Н. Формальная постановка задачи и синтез многоуровневой модели влияния инвестиций на экономическую составляющую качества жизни / А.Н. Ткачев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(006). С. 185 – 213. – IDA [article ID]: 0060404017. – Режим доступа: <http://ej.kubagro.ru/2004/04/pdf/17.pdf>, 1,812 у.п.л.
4. Ткачев А.Н. Исследование многоуровневой семантической информационной модели влияния инвестиций на уровень качества жизни населения региона / А.Н. Ткачев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(006). С. 228 – 267. – IDA [article ID]: 0060404019. – Режим доступа: <http://ej.kubagro.ru/2004/04/pdf/19.pdf>, 2,5 у.п.л.
5. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с.
6. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
7. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с.
8. Симанков В.С., Луценко Е.В., Лаптев В.Н. Системный анализ в адаптивном управлении: Монография (научное издание). /Под науч. ред. В.С.Симанкова. – Краснодар: ИСТЭК КубГТУ, 2001. – 258с.
9. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
10. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с.
11. Луценко Е.В., Лойко В.И., Семантические информационные модели управления агропромышленным комплексом. Монография (научное издание). – Краснодар: КубГАУ. 2005. – 480 с.
12. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп.– Краснодар: КубГАУ, 2006. – 615 с.
13. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с.
14. Наприев И.Л., Луценко Е.В., Чистилин А.Н. Образ-Я и стилевые особенности деятельности сотрудников органов внутренних дел в экстремальных условиях. Монография (научное издание). – Краснодар: КубГАУ. 2008. – 262 с.

15. Луценко Е. В., Лойко В.И., Великанова Л.О. Прогнозирование и принятие решений в растениеводстве с применением технологий искусственного интеллекта: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 257 с.
16. Трунев А.П., Луценко Е.В. Астросоциотипология: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 264 с.
17. Луценко Е.В., Коржаков В.Е., Лаптев В.Н. Теоретические основы и технология применения системно-когнитивного анализа в автоматизированных системах обработки информации и управления (АСОИУ) (на примере АСУ вузом): Под науч. ред.д.э.н., проф. Е.В.Луценко. Монография (научное издание). – Майкоп: АГУ. 2009. – 536 с.
18. Луценко Е.В., Коржаков В.Е., Ермоленко В.В. Интеллектуальные системы в контроллинге и менеджменте средних и малых фирм: Под науч. ред. д.э.н., проф. Е.В.Луценко. Монография (научное издание). – Майкоп: АГУ. 2011. – 392 с.
19. Наприев И.Л., Луценко Е.В. Образ-я и стилевые особенности личности в экстремальных условиях: Монография (научное издание). – Saarbrucken, Germany: LAP Lambert Academic Publishing GmbH & Co. KG., 2012. – 262 с. Номер проекта: 39475, ISBN: 978-3-8473-3424-8
20. Трунев А.П., Луценко Е.В. Автоматизированный системно-когнитивный анализ влияния факторов космической среды на ноосферу, магнитосферу и литосферу Земли: Под науч. ред. д.т.н., проф. В.И.Лойко. Монография (научное издание). – Краснодар, КубГАУ. 2012. – 480 с. ISBN 978-5-94672-519-4
21. Трубилин А.И., Барановская Т.П., Лойко В.И., Луценко Е.В. Модели и методы управления экономикой АПК региона. Монография (научное издание). – Краснодар: КубГАУ. 2012. – 528 с. ISBN 978-5-94672-584-2
22. Горпинченко К.Н., Луценко Е.В. Прогнозирование и принятие решений по выбору агротехнологий в зерновом производстве с применением методов искусственного интеллекта (на примере СК-анализа). Монография (научное издание). – Краснодар, КубГАУ. 2013. – 168 с. ISBN 978-5-94672-644-3
23. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0
24. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0
25. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф.С.Г.Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с. ISBN 978-5-94672-923-9
26. Страница проф.Е.В.Луценко на сайте Научного журнала КубГАУ: <http://ej.kubagro.ru/a/viewaut.asp?id=11>
27. Страница сайта проф.Е.В.Луценко, посвященная АСК-анализу и системе «Эйдос»: <http://lc.kubagro.ru/aidos/index.htm>
28. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
29. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
30. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.
31. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (На-

учный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.

32. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (алгоритм и программная реализация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1371 – 1421. – IDA [article ID]: 1041410100. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/100.pdf>, 3,188 у.п.л.

33. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (математические аспекты) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 814 – 845. – IDA [article ID]: 1051501050. – Режим доступа: <http://ej.kubagro.ru/2015/01/pdf/50.pdf>, 2 у.п.л.

34. Луценко Е.В. Решение задач статистики методами теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №02(106). С. 1 – 47. – IDA [article ID]: 1061502001. – Режим доступа: <http://ej.kubagro.ru/2015/02/pdf/01.pdf>, 2,938 у.п.л.

35. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

36. Орлов А.И. Проблемы управления экологической безопасностью. Итоги двадцати лет научных исследований и преподавания. – Saarbrücken: Palmarium Academic Publishing. 2012. – 344 с. [Электронный ресурс]. – Режим доступа: <http://www.bmstu.ru/ps/~orlov/fileman/ls/Орлов%20А.И.%20Проблемы%20управления%20экологической%20безопасностью>

Литература к разделу 5.2.1.

1. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

2. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0

3. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

4. Макаревич О.А. Управление агропромышленным холдингом с применением технологий искусственного интеллекта: Монография (научное издание). – М: "Финансы и статистика", 2009. – 215 с.: ил. – Режим доступа: <http://elibrary.ru/item.asp?id=19916607>

5. Луценко Е.В. Системно-когнитивный подход к построению многоуровневой семантической информационной модели управления агропромышленным холдингом / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 194 – 214. – Шифр Информрегистра: 0420800012\0095, IDA [article ID]: 0410807011. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/11.pdf>, 1,312 у.п.л.

6. Луценко Е.В. Исследование характеристик исходных данных по агропромышленному холдингу и разработка программного интерфейса их объединения и стандартизации (формализация предметной области) / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 215 – 246. – Шифр Информрегистра: 0420800012\0094, IDA [article ID]: 0410807012. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/12.pdf>, 2 у.п.л.

7. Луценко Е.В. Синтез и верификация двухуровневой семантической информационной модели агропромышленного холдинга / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(042). С. 1 – 15. – Шифр Информрегистра: 0420800012\0120, IDA [article ID]: 0420808001. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/01.pdf>, 0,938 у.п.л.
8. Луценко Е.В. Решение задач прогнозирования и поддержки принятия решений (управления) для агропромышленного холдинга на основе его двухуровневой семантической информационной модели / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(042). С. 16 – 34. – Шифр Информрегистра: 0420800012\0119, IDA [article ID]: 0420808002. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/02.pdf>, 1,188 у.п.л.
9. Луценко Е.В. Исследование двухуровневой семантической информационной модели агропромышленного холдинга / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(042). С. 35 – 75. – Шифр Информрегистра: 0420800012\0118, IDA [article ID]: 0420808003. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/03.pdf>, 2,562 у.п.л.
10. Луценко Е.В. Методология применения системно-когнитивного анализа для синтеза многоуровневой семантической информационной модели агропромышленного холдинга и решения на ее основе задач прогнозирования, поддержки принятия управленческих решений и научных исследований / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(045). С. 11 – 29. – Шифр Информрегистра: 0420900012\0006, IDA [article ID]: 0450901002. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/02.pdf>, 1,188 у.п.л.
11. Луценко Е.В. Автоматизированные технологии управления знаниями в агропромышленном холдинге / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №08(052). С. 98 – 109. – Шифр Информрегистра: 0420900012\0088, IDA [article ID]: 0520908007. – Режим доступа: <http://ej.kubagro.ru/2009/08/pdf/07.pdf>, 0,75 у.п.л.
12. Луценко Е.В. Управление агропромышленным холдингом на основе когнитивных функций связи результатов работы холдинга и характеристик его предприятий / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 248 – 260. – Шифр Информрегистра: 0420900012\0111, IDA [article ID]: 0540910015. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/15.pdf>, 0,812 у.п.л.
13. Макаревич О.А. Применение технологий искусственного интеллекта для прогнозирования и управления в агропромышленном холдинге / О.А. Макаревич, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №05(059). С. 149 – 157. – Шифр Информрегистра: 0421000012\0093, IDA [article ID]: 0591005010. – Режим доступа: <http://ej.kubagro.ru/2010/05/pdf/10.pdf>, 0,562 у.п.л.
14. Автоматизированный системно-когнитивный анализ и его применение для управления социально-экономическими системами в АПК / Е.В. Луценко, В.И. Лойко, О.А. Макаревич, Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №04(078). С. 654 – 698. – IDA [article ID]: 0781204055. – Режим доступа: <http://ej.kubagro.ru/2012/04/pdf/55.pdf>, 2,812 у.п.л.
15. Лойко В.И. Потокные модели управления эффективностью инвестиций в агропромышленных объединениях / В.И. Лойко, Т.П. Барановская, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 615 – 631. – IDA [article ID]: 0831209043. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/43.pdf>, 1,062 у.п.л.
16. Лойко В.И. Инвестиционно-ресурсное управление сельскохозяйственным производством / В.И. Лойко, Т.П. Барановская, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 582 – 614. – IDA [article ID]: 0831209042. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/42.pdf>, 2,062 у.п.л.

17. Луценко Е.В. Интеллектуальные модели инвестиционного управления АПК / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 540 – 581. – IDA [article ID]: 0831209041. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/41.pdf>, 2,625 у.п.л.

18. Луценко Е.В. Концептуальные основы управления экономической устойчивостью перерабатывающего комплекса региона с применением технологий искусственного интеллекта / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №03(087). С. 739 – 748. – IDA [article ID]: 0871303057. – Режим доступа: <http://ej.kubagro.ru/2013/03/pdf/57.pdf>, 0,625 у.п.л.

19. Луценко Е.В. Синтез, верификация и исследование на устойчивость системно-когнитивной модели перерабатывающего комплекса региона / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 305 – 333. – IDA [article ID]: 1011407016. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/16.pdf>, 1,812 у.п.л.

Литература к разделу 5.2.2

1. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

2. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0

3. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

4. Луценко Е.В. Системно-когнитивные модели прогнозирования развития многоотраслевой агропромышленной корпорации. Часть I. Когнитивная структуризация и формализация предметной области / Е.В. Луценко, В.И. Лойко, Т.П. Барановская, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №09(113). – IDA [article ID]: 1131509097. – Режим доступа: <http://ej.kubagro.ru/2015/09/pdf/97.pdf>, 1,125 у.п.л.

5. Макаревич О.А. Управление агропромышленным холдингом с применением технологий искусственного интеллекта: Монография (научное издание). – М: "Финансы и статистика", 2009. – 215 с.: ил. – Режим доступа: <http://elibrary.ru/item.asp?id=19916607>

6. Луценко Е.В. Системно-когнитивный подход к построению многоуровневой семантической информационной модели управления агропромышленным холдингом / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 194 – 214. – Шифр Информрегистра: 0420800012\0095, IDA [article ID]: 0410807011. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/11.pdf>, 1,312 у.п.л.

7. Луценко Е.В. Исследование характеристик исходных данных по агропромышленному холдингу и разработка программного интерфейса их объединения и стандартизации (формализация предметной области) / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 215 – 246. – Шифр Информрегистра: 0420800012\0094, IDA [article ID]: 0410807012. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/12.pdf>, 2 у.п.л.

8. Луценко Е.В. Синтез и верификация двухуровневой семантической информационной модели агропромышленного холдинга / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(042). С. 1 – 15. – Шифр Информрегистра: 0420800012\0120, IDA [article ID]: 0420808001. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/01.pdf>, 0,938 у.п.л.

9. Луценко Е.В. Решение задач прогнозирования и поддержки принятия решений (управления) для агропромышленного холдинга на основе его двухуровневой семантической информационной модели / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(042). С. 16 – 34. – Шифр Информрегистра: 0420800012\0119, IDA [article ID]: 0420808002. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/02.pdf>, 1,188 у.п.л.
10. Луценко Е.В. Исследование двухуровневой семантической информационной модели агропромышленного холдинга / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(042). С. 35 – 75. – Шифр Информрегистра: 0420800012\0118, IDA [article ID]: 0420808003. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/03.pdf>, 2,562 у.п.л.
11. Луценко Е.В. Методология применения системно-когнитивного анализа для синтеза многоуровневой семантической информационной модели агропромышленного холдинга и решения на ее основе задач прогнозирования, поддержки принятия управленческих решений и научных исследований / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(045). С. 11 – 29. – Шифр Информрегистра: 0420900012\0006, IDA [article ID]: 0450901002. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/02.pdf>, 1,188 у.п.л.
12. Луценко Е.В. Автоматизированные технологии управления знаниями в агропромышленном холдинге / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №08(052). С. 98 – 109. – Шифр Информрегистра: 0420900012\0088, IDA [article ID]: 0520908007. – Режим доступа: <http://ej.kubagro.ru/2009/08/pdf/07.pdf>, 0,75 у.п.л.
13. Луценко Е.В. Управление агропромышленным холдингом на основе когнитивных функций связи результатов работы холдинга и характеристик его предприятий / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 248 – 260. – Шифр Информрегистра: 0420900012\0111, IDA [article ID]: 0540910015. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/15.pdf>, 0,812 у.п.л.
14. Макаревич О.А. Применение технологий искусственного интеллекта для прогнозирования и управления в агропромышленном холдинге / О.А. Макаревич, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №05(059). С. 149 – 157. – Шифр Информрегистра: 0421000012\0093, IDA [article ID]: 0591005010. – Режим доступа: <http://ej.kubagro.ru/2010/05/pdf/10.pdf>, 0,562 у.п.л.
15. Автоматизированный системно-когнитивный анализ и его применение для управления социально-экономическими системами в АПК / Е.В. Луценко, В.И. Лойко, О.А. Макаревич, Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №04(078). С. 654 – 698. – IDA [article ID]: 0781204055. – Режим доступа: <http://ej.kubagro.ru/2012/04/pdf/55.pdf>, 2,812 у.п.л.
16. Лойко В.И. Поточковые модели управления эффективностью инвестиций в агропромышленных объединениях / В.И. Лойко, Т.П. Барановская, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 615 – 631. – IDA [article ID]: 0831209043. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/43.pdf>, 1,062 у.п.л.
17. Лойко В.И. Инвестиционно-ресурсное управление сельскохозяйственным производством / В.И. Лойко, Т.П. Барановская, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 582 – 614. – IDA [article ID]: 0831209042. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/42.pdf>, 2,062 у.п.л.
18. Луценко Е.В. Интеллектуальные модели инвестиционного управления АПК / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 540 – 581. – IDA [article ID]: 0831209041. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/41.pdf>, 2,625 у.п.л.

19. Луценко Е.В. Концептуальные основы управления экономической устойчивостью перерабатывающего комплекса региона с применением технологий искусственного интеллекта / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №03(087). С. 739 – 748. – IDA [article ID]: 0871303057. – Режим доступа: <http://ej.kubagro.ru/2013/03/pdf/57.pdf>, 0,625 у.п.л.

20. Луценко Е.В. Синтез, верификация и исследование на устойчивость системно-когнитивной модели перерабатывающего комплекса региона / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 305 – 333. – IDA [article ID]: 1011407016. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/16.pdf>, 1,812 у.п.л.

Литература к разделу 5.3

1. Луценко Е.В. Концептуальные основы управления экономической устойчивостью перерабатывающего комплекса региона с применением технологий искусственного интеллекта / Е.В. Луценко, В.И. Лойко, Т.П. Барановская // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №03(087). С. 739 – 748. – IDA [article ID]: 0871303057. – Режим доступа: <http://ej.kubagro.ru/2013/03/pdf/57.pdf>, 0,625 у.п.л.

2. Луценко Е.В. Применение СК-анализа и системы «Эйдос» для синтеза когнитивной матричной передаточной функции сложного объекта управления на основе эмпирических данных / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №01(75). С. 681 – 714. – Шифр Информрегистра: 0421200012\0008. – Режим доступа: <http://ej.kubagro.ru/2012/01/pdf/53.pdf>, 2,125 у.п.л.

3. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (когнитивная структуризация и формализация предметной области) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(51). С. 1 – 37. – Шифр Информрегистра: 0420900012\0073. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/01.pdf>, 2,312 у.п.л.

4. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (синтез и верификация семантической информационной модели) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(51). С. 38 – 46. – Шифр Информрегистра: 0420900012\0072. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/02.pdf>, 0,562 у.п.л.

5. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (прогнозирование, принятие решений и исследование предметной области) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(51). С. 47 – 82. – Шифр Информрегистра: 0420900012\0071. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/03.pdf>, 2,25 у.п.л.

6. Луценко Е.В. Прогнозирование урожайности подсолнечника по Краснодарскому краю с применением системно-когнитивного анализа (Часть 2-я: Формальная постановка задачи и преобразование исходных данных в информацию, а ее в знания) / Е.В. Луценко, Н.О. Познышева // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №10(84). С. 384 – 409. – Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/31.pdf>, 1,625 у.п.л.

7. Луценко Е.В. Прогнозирование урожайности подсолнечника по Краснодарскому краю с применением системно-когнитивного анализа (Часть 3-я: Решение задач прогнозирования и исследования предметной области) / Е.В. Луценко, Н.О. Познышева // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №10(84). С. 410 – 435. – Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/32.pdf>, 1,625 у.п.л.

8. Лойко В.И. Структура 1 производственной системы с вертикальной интеграцией / В.И. Лойко, В.В. Крохмаль // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). С. 224 – 239. – Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/07.pdf>, 1 у.п.л.

9. Крохмаль В.В. Структура 2 производственной системы с вертикальной интеграцией / В.В. Крохмаль, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). С. 240 – 254. – Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/08.pdf>, 0,938 у.п.л.
10. Луценко Е.В., Лойко В.И., Семантические информационные модели управления агропромышленным комплексом. Монография (научное издание). – Краснодар: КубГАУ. 2005. – 480 с.
11. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.
12. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.
13. Луценко Е.В. Разработка без программирования и применение в адаптивном режиме методик риэлтерской экспресс-оценки по методу аналогий (сравнительных продаж) в системно-когнитивном анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №10(094). С. 507 – 564. – IDA [article ID]: 0941310036. – Режим доступа: <http://ej.kubagro.ru/2013/10/pdf/36.pdf>, 3,625 у.п.л.
14. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с.
15. Луценко Е.В. Системно-когнитивный подход к построению многоуровневой семантической информационной модели управления агропромышленным холдингом / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 194 – 214. – Шифр Информрегистра: 0420800012\0095, IDA [article ID]: 0410807011. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/11.pdf>, 1,312 у.п.л.
16. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(063). С. 1 – 23. – Шифр Информрегистра: 0421000012\0233, IDA [article ID]: 0631009001. – Режим доступа: <http://ej.kubagro.ru/2010/09/pdf/01.pdf>, 1,438 у.п.л.
17. Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики / А.И. Орлов, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>, 3,375 у.п.л.

Литература к разделу 6.1

1. Раннев Г.Г. P224 Интеллектуальные средства измерений : учебник для студ. высш. учеб. заведений / Г. Г. Раннев. — М. : Издательский центр «Академия», 2010. — 272 с. ISBN 978-5-7695-6469-7. http://www.academia-moscow.ru/ftp_share/books/fragments/fragment_13431.pdf
2. ГОСТ Р 8.673-2009: Государственная система обеспечения единства измерений. Датчики интеллектуальные и системы измерительные интеллектуальные. Основные термины и определения. http://standartgost.ru/g/ГОСТ_Р_8.673-2009
3. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. <http://elibrary.ru/item.asp?id=18632909>
4. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. <http://elibrary.ru/item.asp?id=18271217>

5. Луценко Е.В. Формирование субъективных (виртуальных) моделей физической и социальной реальности сознанием человека и неоправданное приращение им онтологического статуса (гипостазирование) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №09(113). С. 1 – 32. – IDA [article ID]: 1131509001. – Режим доступа: <http://ej.kubagro.ru/2015/09/pdf/01.pdf>, 2 у.п.л.
6. Луценко Е.В. Принципы и перспективы корректной содержательной интерпретации субъективных (виртуальных) моделей физической и социальной реальности, формируемых сознанием человека / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №01(115). С. 22 – 75. – IDA [article ID]: 1151601003. – Режим доступа: <http://ej.kubagro.ru/2016/01/pdf/03.pdf>, 3,375 у.п.л.
7. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. количественная гидроэкология: методы системной идентификации. –Тольятти: ИЭВБ РАН, 2003. – 463 с. ISBN 5-93424-109-5. <http://www.biometrica.tomsk.ru/konstan.htm> <http://www.twirpx.com/file/339044/>
8. Трунев А.П., Луценко Е.В. Автоматизированный системно-когнитивный анализ влияния факторов космической среды на ноосферу, магнитосферу и литосферу Земли: Под науч. ред. д.т.н., проф. В.И.Лойко. Монография (научное издание). – Краснодар, КубГАУ. 2012. – 480 с. ISBN 978-5-94672-519-4. <http://elibrary.ru/item.asp?id=21683737>
9. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
10. Орлов А.И. Теория измерений как часть методов анализа данных: размышления над переводом статьи П.Ф. Веллемана и Л. Уилкинсона // Социология: методология, методы, математическое моделирование. 2012. № 35. С. 155-174.
11. Хаббард Дуглас У. Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе / Дуглас У. Хаббард / [Пер. с англ. Е. Пестеревой]. — М.: ЗАО «Олимп–Бизнес», 2009. — 320 с.: ил. ISBN 978-5-9693-0163-4 (рус.). <http://www.twirpx.com/file/1546361/>
12. Клир Дж. Системология. Автоматизация решения системных задач. Перевод с английского М. А. Зуева, под редакцией А. И. Горлина. М.: Радио и связь, 1990. –534 с. <http://www.twirpx.com/file/486296/>
13. Демидов Я.П.. Системное измерение экономических процессов и явлений. – Казань: Изд-во МОиН РТ, 2011. – 268с.
14. Луценко Е.В. Подчиняются ли социально-экономические явления каким-то аналогам или обобщениям принципа относительности Галилея и Эйнштейна и выполняются ли для них теорема Нётер и законы сохранения? / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 219 – 254. – IDA [article ID]: 0911307014. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/14.pdf>, 2,25 у.п.л.
15. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>
16. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф.С.Г.Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с. ISBN 978-5-94672-923-9. <http://elibrary.ru/item.asp?id=23209923>
17. Луценко Е.В. Существование, несуществование и изменение как эмерджентные свойства систем. / Квантовая Магия, том 5, вып. 1, стр. 1215-1239, 2008. [Электронный ресурс]. – Режим доступа: <http://quantmagic.narod.ru/volumes/VOL512008/p1215.html>
18. Луценко Е.В. Исследование влияния подсистем различных уровней иерархии на эмерджентные свойства системы в целом с применением АСК-анализа и интеллектуальной системы "Эйдос" (микроструктура системы как фактор управления ее макросвойствами) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №01(075). С. 638 – 680. – Шифр Информрегистр: 0421200012/0025, IDA [article ID]: 0751201052. – Режим доступа: <http://ej.kubagro.ru/2012/01/pdf/52.pdf>, 2,688 у.п.л.
19. Чередищенко Н.А., Луценко Е.В., Бандык Д.К., Трунев А.П. Прогнозирование землетрясений на основе астрономических данных с применением АСК-анализа на примере большого калифорнийского

разлома Сан-Андреас / Н.А. Чердниченко, Е.В. Луценко, Д.К. Бандык, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 1322 – 1377. – IDA [article ID]: 0911307093. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/93.pdf>, 3,5 у.п.л.

20. Лопатина Л.М. Концептуальная постановка задачи: "Прогнозирование количественных и качественных результатов выращивания заданной культуры в заданной точке" / Л.М. Лопатина, И.А. Драгавцева, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №05(007). С. 86 – 100. – IDA [article ID]: 0070405008. – Режим доступа: <http://ej.kubagro.ru/2004/05/pdf/08.pdf>, 0,938 у.п.л.

21. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.

22. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012(0197), IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.

23. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 328 – 356. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 у.п.л.

24. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408 с. <http://www.twirpx.com/file/124076/>

25. Орлов А.И. Точки роста статистических методов / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 136 – 162. – IDA [article ID]: 1031409011. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/11.pdf>, 1,688 у.п.л.

26. Луценко Е.В. Решение задач статистики методами теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №02(106). С. 1 – 47. – IDA [article ID]: 1061502001. – Режим доступа: <http://ej.kubagro.ru/2015/02/pdf/01.pdf>, 2,938 у.п.л.

27. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.

28. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.

29. Луценко Е.В. Коэффициент эмерджентности классических и квантовых статистических систем / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С. 214 – 235. – IDA [article ID]: 0901306014. – Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/14.pdf>, 1,375 у.п.л.

30. Луценко Е.В. СК-анализ и система "Эйдос" в свете философии Платона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(045). С. 91 – 100. – Шифр Информрегистра: 0420900012(0010), IDA [article ID]: 0450901008. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/08.pdf>, 0,625 у.п.л.

31. Пойа Д. Математика и правдоподобные рассуждения. 2 изд. – М., Наука, 1975 – 464 с. <http://ilib.mccme.ru/djvu/polya/rassuzhdenija.htm>
32. *Hartley R.V.L.* Transmission of information. — Bell System Technical Journal - 7. — 1928. — С. 535-63. перевод: Хартли Р. Передача информации / Хартли Р. // Теория информации и ее приложения / Под ред. А.А. Харкевича. – М.: Физматгиз, 1959. – С. 5 – 35. http://www.dotrose.com/etext/90_Miscellaneous/transmission_of_information_1928b.pdf,
33. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 117 – 193. – Шифр Информрегистра: 0420800012\0091, IDA [article ID]: 0410807010. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/10.pdf>, 4,812 у.п.л.
34. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с. <http://elibrary.ru/item.asp?id=18630282>
35. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с. <http://elibrary.ru/item.asp?id=21745340>
36. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с. <http://elibrary.ru/item.asp?id=18828433>
37. Симанков В.С., Луценко Е.В., Лаптев В.Н. Системный анализ в адаптивном управлении: Монография (научное издание). /Под науч. ред. В.С.Симанкова. – Краснодар: ИСТЭК КубГТУ, 2001. – 258с. <http://elibrary.ru/item.asp?id=21747625>
38. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с. <http://elibrary.ru/item.asp?id=18632737>
39. Луценко Е.В., Лойко В.И., Семантические информационные модели управления агропромышленным комплексом. Монография (научное издание). – Краснодар: КубГАУ. 2005. – 480 с. <http://elibrary.ru/item.asp?id=21720635>
40. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп.– Краснодар: КубГАУ, 2006. – 615 с. <http://elibrary.ru/item.asp?id=18632602>
41. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с. <http://elibrary.ru/item.asp?id=21683721>
42. Наприев И.Л., Луценко Е.В., Чистилин А.Н. Образ-Я и стилевые особенности деятельности сотрудников органов внутренних дел в экстремальных условиях. Монография (научное издание). – Краснодар: КубГАУ. 2008. – 262 с. <http://elibrary.ru/item.asp?id=21683724>
43. Луценко Е. В., Лойко В.И., Великанова Л.О. Прогнозирование и принятие решений в растениеводстве с применением технологий искусственного интеллекта: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 257 с. <http://elibrary.ru/item.asp?id=21683725>
44. Трунев А.П., Луценко Е.В. Астросоциотипология: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 264 с. http://lc.kubagro.ru/aidos/aidos08_TL/Monography-TL.rar
45. Луценко Е.В., Коржаков В.Е., Лаптев В.Н. Теоретические основы и технология применения системно-когнитивного анализа в автоматизированных системах обработки информации и управления (АСОИУ) (на примере АСУ вузом): Под науч. ред. д.э.н., проф. Е.В.Луценко. Монография (научное издание). – Майкоп: АГУ. 2009. – 536 с. <http://elibrary.ru/item.asp?id=18633313>
46. Луценко Е.В., Коржаков В.Е., Ермоленко В.В. Интеллектуальные системы в контроллинге и менеджменте средних и малых фирм: Под науч. ред. д.э.н., проф. Е.В.Луценко. Монография (научное издание). – Майкоп: АГУ. 2011. – 392 с. <http://elibrary.ru/item.asp?id=21683734>
47. Наприев И.Л., Луценко Е.В. Образ-Я и стилевые особенности личности в экстремальных условиях: Монография (научное издание). – Saarbrücken, Germany: LAP Lambert Academic Publishing GmbH & Co. KG., 2012. – 262 с. Номер проекта: 39475, ISBN: 978-3-8473-3424-8.
48. Трубилин А.И., Барановская Т.П., Лойко В.И., Луценко Е.В. Модели и методы управления экономикой АПК региона. Монография (научное издание). – Краснодар: КубГАУ. 2012. – 528 с. ISBN 978-5-94672-584-2. <http://elibrary.ru/item.asp?id=21683702>
49. Горпинченко К.Н., Луценко Е.В. Прогнозирование и принятие решений по выбору агротехнологий в зерновом производстве с применением методов искусственного интеллекта (на примере СК-

анализа). Монография (научное издание). – Краснодар, КубГАУ. 2013. – 168 с. ISBN 978-5-94672-644-3. <http://elibrary.ru/item.asp?id=20213254>

50. Луценко Е.В. Решение задач ампелографии с применением АСК-анализа изображений листьев по их внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык, Л.П. Трошин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №08(112). С. 862 – 910. – IDA [article ID]: 1121508064. – Режим доступа: <http://ej.kubagro.ru/2015/08/pdf/64.pdf>, 3,062 у.п.л.

51. Луценко Е.В. Системно-когнитивный анализ и система "Эйдос" и их применение для построения интеллектуальных измерительных систем // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74. <http://elibrary.ru/contents.asp?issueid=1267409> <http://elibrary.ru/item.asp?id=21538328>

52. Луценко Е.В. Математический метод АСК-анализа - системная теория информации // Заводская лаборатория. Диагностика материалов. 2015. Т. 81. № 5. С. 61 - 71. <http://www.zldm.ru/content/article.php?ID=1985>

53. Луценко Е.В. АСК-анализ влияния экологических факторов на качество жизни населения региона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №06(110). С. 1 – 37. – IDA [article ID]: 1101506001. – Режим доступа: <http://ej.kubagro.ru/2015/06/pdf/01.pdf>, 2,312 у.п.л.

54. Луценко Е.В. Реализация психологических, педагогических и профориентационных тестов и супертестов без программирования в среде интеллектуальной системы «Эйдос-Х++» (На примере теста: «Анализ особенностей индивидуального стиля педагогической деятельности») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 1057 – 1085. – IDA [article ID]: 0881304076. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/76.pdf>, 1,812 у.п.л.

55. Луценко Е.В. Реализация тестов и супертестов для ветеринарной и медицинской диагностики в среде системы искусственного интеллекта «Эйдос-Х++» без программирования / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 167 – 207. – IDA [article ID]: 0891305014. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/14.pdf>, 2,562 у.п.л.

56. Луценко Е.В. Прогнозирование длительности послеоперационного восстановительного периода методом сердечно-дыхательного синхронизма (СДС) с применением АСК-анализа (часть 1) / Е.В. Луценко, Е.В. Сергеева // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №10(064). С. 142 – 178. – Шифр Информрегистра: 0421000012\0280, IDA [article ID]: 0641010014. – Режим доступа: <http://ej.kubagro.ru/2010/10/pdf/14.pdf>, 2,312 у.п.л.

57. Луценко Е.В. Прогнозирование длительности послеоперационного восстановительного периода методом сердечно-дыхательного синхронизма (СДС) с применением АСК-анализа (часть 2) / Е.В. Луценко, Е.В. Сергеева // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №10(064). С. 179 – 203. – Шифр Информрегистра: 0421000012\0279, IDA [article ID]: 0641010015. – Режим доступа: <http://ej.kubagro.ru/2010/10/pdf/15.pdf>, 1,562 у.п.л.

58. Чердниченко Н.А. Моделирование смещения полюса Земли и алгоритм прогнозирования его динамики с применением АСК-анализа / Н.А. Чердниченко, Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №05(099). С. 149 – 188. – IDA [article ID]: 0991405010. – Режим доступа: <http://ej.kubagro.ru/2014/05/pdf/10.pdf>, 2,5 у.п.л.

59. Чердниченко Н.А. Прогнозирование глобальных климатических аномалий Эль-Ниньо и Ла-Нинья с применением системы искусственного интеллекта Aidos-X / Н.А. Чердниченко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 128 – 160. – IDA [article ID]: 1051501007. – Режим доступа: <http://ej.kubagro.ru/2015/01/pdf/07.pdf>, 2,062 у.п.л.

60. Чердниченко Н.А. Моделирование и прогноз динамики глобальных климатических аномалий типа Эль-Ниньо и Ла-Нинья / Н.А. Чердниченко, А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный жур-

нал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №06(110). С. 1545 – 1577. – IDA [article ID]: 1101506102. – Режим доступа: <http://ej.kubagro.ru/2015/06/pdf/102.pdf>, 2,062 у.п.л.

61. Silvano Martelo, Paolo Toth. Knapsack problems. – Wiley, 1990. – 306 с.

62. Луценко Е.В. Количественное измерение сходства-различия клонов винограда по контурам листьев с применением АСК-анализа и системы «Эйдос» / Луценко Е.В., Трошин Л.П. // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). – IDA [article ID]: 1161602077. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/77.pdf>, 1,313 у.п.л.

Литература к разделу 6.2

1. Ричард Броди. Психические вирусы. Методическое пособие для слушателей курса. «Современные психотехнологии». Москва, 2002, 192 стр. [http://yandex.ru/yandsearch?text= Ричард%20Броди.%20ПСИХИЧЕСКИЕ%20ВИРУСЫ](http://yandex.ru/yandsearch?text=Ричард%20Броди.%20ПСИХИЧЕСКИЕ%20ВИРУСЫ)

2. Кара-Мурза С.Г. Манипуляция сознанием. — М.: Изд-во: Эксмо, 2005. — 832 с. ISBN 5-699-08331-6/ [http://socioline.ru/files/5/52/kara-murza s. - manipulyaciya_soznaniem_politicheskii_bestseller_-_2005.pdf](http://socioline.ru/files/5/52/kara-murza_s.-manipulyaciya_soznaniem_politicheskii_bestseller_-_2005.pdf) Луценко Е.В. Тотальная ложь как стратегическое информационное оружие общества периода глобализации и дополненной реальности (применим ли в современном обществе принцип наблюдаемости как критерий реальности) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1411 – 1429. – IDA [article ID]: 1011407091. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/91.pdf>, 1,188 у.п.л.

3. Луценко Е.В. Синтез и верификация многокритериальной системно-когнитивной модели университетского рейтинга Гардиан и ее применение для сопоставимой оценки эффективности российских вузов с учетом направления подготовки / Луценко Е.В. // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №03(107). – IDA [article ID]: 1071503001. – Режим доступа: <http://ej.kubagro.ru/2015/03/pdf/01.pdf>, 3,875 у.п.л.

4. Луценко Е.В. Виртуализация общества как основной информационный аспект глобализации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №01(009). С. 6 – 43. – IDA [article ID]: 0090501002. – Режим доступа: <http://ej.kubagro.ru/2005/01/pdf/02.pdf>, 2,375 у.п.л.

5. Вяткин В.Б. Групповой плагиат: от студента до министра. - Троицкий вариант — Наука - <http://trv-science.ru> - [Электронный ресурс]. Адрес доступа: <http://trv-science.ru/2011/11/08/gruppovoj-plagiat-ot-studenta-do-ministra/> или: <http://trv-science.ru/2011/11/08/gruppovoj-plagiat-ot-studenta-do-ministra/print/>

6. Луценко Е.В. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 498 – 544. – IDA [article ID]: 1031409032. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/32.pdf>, 2,938 у.п.л.

7. Луценко Е.В. Современное состояние и перспективы развития Политематического сетевого электронного научного журнала Кубанского государственного аграрного университета / Е.В. Луценко, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 146 – 176. – IDA [article ID]: 1001406008. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/08.pdf>, 1,938 у.п.л.

8. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.

9. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013.

– №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

Литература к разделу 6.3.

1. Луценко Е.В. Хиршамания при оценке результатов научной деятельности, ее негативные последствия и попытка их преодоления с применением многокритериального подхода и теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №04(108). С. 1 – 29. – IDA [article ID]: 1081504001. – Режим доступа: <http://ej.kubagro.ru/2015/04/pdf/01.pdf>, 1,812 у.п.л.
2. Чеботарев П. Ю., “Наукометрия: как с её помощью лечить, а не калечить?”, УБС, 44 (2013), 14–31. <http://onr-russia.ru/sites/default/files/zatravka.pdf>
3. Орлов А.И. О некоторых методологически ошибочных методах анализа и оценки результатов научной деятельности // Россия: тенденции и перспективы развития. Ежегодник. Вып. 8. / РАН. ИНИОН. Отд. науч. сотрудничества и междунар. связей; Отв. ред. Ю.С. Пивоваров. – М., 2013а. – Ч. 2. – С.528–533.
4. Орлов А.И. Два типа методологических ошибок при управлении научной деятельностью // Управление большими системами / Сборник трудов. Специальный выпуск 44. Наукометрия и экспертиза в управлении наукой / [под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева]. – 2013б. – С.32–54. [Электронный ресурс]. Режим доступа: http://ubs.mtas.ru/archive/search_resul...n_id=19050 (дата обращения 30.07.2014).
5. Орлов А.И. Наукометрия и управление научной деятельностью // Управление большими системами / Сборник трудов. Специальный выпуск 44. Наукометрия и экспертиза в управлении наукой / [под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева]. – 2013в. – С. 538–568. [Электронный ресурс]. Режим доступа: http://ubs.mtas.ru/archive/search_resul...n_id=19078 (дата обращения 30.07.2014).
6. Орлов А.И. О ключевых показателях эффективности научной деятельности / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №07(111). С. 81 – 112. – IDA [article ID]: 1111507006. – Режим доступа: <http://ej.kubagro.ru/2015/07/pdf/06.pdf>, 2 у.п.л.
7. Орлов А.И. Примеры методологических ошибок при управлении научной деятельностью // Проблемы наукометрии: состояние и перспективы развития. Международная конференция. – М.: Ин-т проблем развития науки РАН, 2013е. – С.107 – 109.
8. Орлов А.И. Критерии выбора показателей эффективности научной деятельности // Контроллинг. – 2013. – №3(49). – С.72–78.
9. Форум: <http://forum.orlovs.pp.ru/viewtopic.php?p=8357>
10. Бернал Дж. Наука и общество. М.: Изд-во иностр. лит., 1953. Режим доступа URL: <http://www.twirpx.com/file/498382/>
11. Прайс Д. Малая наука, большая наука // Наука о науке, М.: Изд-во «Прогресс», 1966. Режим доступа URL: <http://www.garfield.library.upenn.edu/essays/v10p072y1987.pdf>; <http://lis.sagepub.com/content/35/2/115.abstract>
12. Добров Г.М. Наука о науке. Киев: Наукова Думка, 1989, 302 с.
13. Налимов В.В., Мульченко З.М. Наукометрия. М.: Наука, 1969, 192 с. Режим доступа URL: <http://www.e-reading.by/divureader.php/113843/Nalimov - Naukometriya.html>
14. Хайтун С.Д. Наукометрия. Состояние и перспективы. М.: Наука, 1983, 344 с. Режим доступа URL: <http://librarun.org/book/12517/1>
15. Бедный Б.И., Миронос А.А., Сорокин Ю.М., Сулейманов Е.В. Наука и научная деятельность: организация, технологии, информационное обеспечение / Под ред. проф. Б.И. Бедного. - Нижний Новгород: Изд-во ННГУ, 2013. - 228 с.
16. Мирский Э.М. Междисциплинарные исследования и дисциплинарная организация науки, М. Наука, 1980, 304 с.
17. Пельц Д., Энрюс Ф. Ученые в организациях. Оптимальные условия для исследований и разработок. М.: Прогресс, 1973, 469 с. Режим доступа URL: http://publ.lib.ru/ARCHIVES/P/PEL'C_Donat/d/Pelc_D..html
18. Гарфилд Ю. Можно ли выявлять и оценивать научные достижения и научную продуктивность? // Вестник АН СССР, 1982. - № 7. - С. 42-50. Режим доступа URL: <http://www.prometeus.nsc.ru/science/citation/garfield.ssi>
19. Управление большими системами / Сборник трудов. Специальный выпуск 44 - Наукометрия и экспертиза в управлении наукой / [под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева]. М.: ИПУ РАН, 2013. - 568 с. Режим доступа URL: http://ubs.mtas.ru/archive/index.php?SECTION_ID=685

20. Игра в цифры, или как теперь оценивают труд учёного (сборник статей о библиометрике). М.: МЦНМО, 2011. 72 с. Режим доступа URL: <http://www.mccme.ru/free-books/bibliometric.pdf>; или <http://www.twirpx.com/file/753485/>
21. Бедный Б.И., Сорокин Ю.М. О показателях научного цитирования и их применении // Высшее образование в России. 2012. № 3. С. 17-28. Режим доступа URL: <http://vovr.ru/upload/bednyi-sorokin%203-12.pdf>
22. Мотрошилова Н.В. Реальные факторы научно-исследовательского труда и измерения цитирования // Управление большими системами. - 2013. - № 44 - С. 453-475. Режим доступа URL: http://ubs.mtas.ru/archive/index.php7SECTION_ID=685
23. Орлов А.И. Наука как объект управления / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1244 – 1274. – IDA [article ID]: 1011407082. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/82.pdf>, 1,938 у.п.л.
24. Юревич М. А. Методические проблемы оценки результативности исследователя // Альманах “Наука. Инновации. Образование”. 2014, вып. 16. Режим доступа URL: <http://riep.ru>; Обсуждение статьи М.А. Юревича “Методические проблемы оценки результативности исследователя” - там же.
25. Алескерев Ф.Т., Писляков В.В., Субочев А.Н., Чистяков А.Г. Построение рейтингов журналов по менеджменту с помощью методов теории коллективного выбора: препринт WP7/2011/04. Нац. иссл. ун-т «Высшая школа экономики». М.: Изд. дом Высшей школы экономики, 2011. - 44 с. Режим доступа URL: http://www.hse.ru/data/2011/06/29/1216101480/WP7_2011_04_final.pdf
26. Бедный Б.И., Миронос А.А., Серова Т.В. Продуктивность исследовательской работы обучающихся (наукометрические оценки) // Высшее образование в России. 2006. - №7. - С. 20-36. Режим доступа URL: <http://www.phd.unn.ru/files/2014/04/008.pdf>
27. Солошенко Н.С., Кириллова О.В. Отражение российских журналов в БД Science Citation Index и SCOPUS // Educational Technology & Society. 2006. V.9. No.3. P.313320.
28. Теста Д. Процесс отбора журналов в Thomson Reuters. Режим доступа URL: http://thomsonreuters.com/content/science/pdf/ssr/journal_selection_essay-russian.pdf
http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/
29. Кириллова О. В. Подготовка российских журналов для зарубежной аналитической базы данных Scopus. Рекомендации и комментарии. Режим доступа - URL: <http://elsevierscience.ru/info/add-journal-to-scopus/>
30. Цыганов А. В. Краткое описание наукометрических показателей, основанных на цитируемости // Управление большими системами / Сборник трудов. Специальный выпуск 44 - Наукометрия и экспертиза в управлении наукой / [под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева]. М.: ИПУ РАН, 2013, с. 248- 261.
31. Publish or Perish. Режим доступа URL: <http://www.harzing.com/pop.htm#metrics> (дата обращения 26.08.2015.)
32. Штовба С.Д., Штовба Е.В. Индекс цитирования, учитывающий скрытую диффузию научных знаний // Научно-техническая информация. Сер. 1 «Организация и методика информационной работы». - 2013. - №7. - С. 28-31. Режим доступа URL: <http://shtovba.vk.vntu.edu.ua/file/6ad63e809551b1e63ab2b9e21f9190e2.pdf>
33. Гринченко С. Н. Имеет ли решение задача перманентной оценки вклада ученого в науку? // Управление большими системами / Сборник трудов. Специальный выпуск 44 - Наукометрия и экспертиза в управлении наукой / [под ред. Д.А. Новикова, А.И. Орлова, П.Ю. Чеботарева]. М.: ИПУ РАН, 2013, с. 280 - 291. Режим доступа URL: http://ubs.mtas.ru/archive/index.php?SECTION_ID=685
34. Михайлов О. В. Цитируемость ученого: важнейший ли это критерий качества его научной деятельности? // Informetrics.ru. Электронный журнал. Статья № 1079. Режим доступа URL: <http://www.informetrics.ru/articles/sn.php?id=56> (дата обращения: 26.08.2015).
35. Орлов А.И. Методологические ошибки ведут к неправильным управленческим решениям // Управление большими системами. Вып. 27. - М.: ИПУ РАН, 2009. - С. 59-65.
36. Эпштейн В. Л. О контрпродуктивности использования наукометрического показателя результативности научной деятельности для будущего России // Проблемы управления. - 2007. - №3. - С. 70-72. Режим доступа - URL: <http://cvberleninka.ru/article/n/o-kontrproduktivnosti-ispolzovaniya-naukometricheskogo-pokazatelva-rezultativnosti-nauchnov-devatelnosti-dlya-buduschego-rossii>
37. Муравьев А.А. К вопросу о классификации российских журналов по экономике и смежным дисциплинам // Научные доклады. - 2012. -Т.14 (R). - С. 1- 60. Режим доступа: http://www.gsom.spbu.ru/files/upload/niim/publishing/2012/wp_muravyev.pdf

38. Силина А.Ю., Васильева В.Д., Дербишер В.Е., Гермашев И.В. Систематизация наукометрических показателей эффективности научной деятельности // Информационные технологии. - 2009. - №6. - С. 53-56.
39. Международный союз математиков предостерегает от неправильного использования статистики цитирований // Полит.ру / Наука. - 16 июня 2008. Режим доступа - URL: <http://www.polit.ru/news/2008/06/16/mathunion/> (дата обращения: 08.01.2013).
40. Солошенко Н.С., Кириллова О.В. Отражение российских журналов в БД Science Citation Index и SCOPUS // Educational Technology & Society. 2006. - V.9. - No.3. - P.313-320.
41. Hirsch J. E. An index to quantify an individual's scientific research output // Proceedings of the National Academy of Sciences. 2005. - Vol. 102. - No. 46. - P. 16569-16572. Режим доступа - URL: <http://www.pnas.org/content/102/46/16569.full>
42. www.elibrary.ru/defaultx.asp - научная электронная библиотека
43. <http://school-collection.edu.ru/> - федеральное хранилище Единая коллекция цифровых образовательных ресурсов
44. www.diss.rsl.ru - электронная библиотека диссертаций
45. <http://www.edu.ru/> - федеральный портал Российское образование
46. <http://www.igumo.ru/> - интернет-портал Института гуманитарного образования и информационных технологий
47. www.edu.ru - сайт Министерства образования РФ
48. <http://riep.ru> - сайт Российского научно-исследовательского института экономики, политики и права в научно-технической сфере (РИЭПП)
49. www.humanities.edu.ru - сайт «Гуманитарное образование»
50. www.edu.ru - федеральный портал «Российское образование»
51. <http://www.eduhmao.ru/info/1/4382/> - информационно-просветительский портал
52. <http://www.iqlib.ru> - электронная библиотека образовательных и просветительских изданий
53. <http://www.integro.ru> - Центр Системных Исследований «Интегро»

Литература к разделу 6.4

1. Хагуров Т.А., Остапенко А.А. Реформа образования глазами учителей и преподавателей: опыт социологического исследования / Ин-т социологии РАН; Рос. акад. социал. наук, Краснодар. регион. отделение. – М.-Краснодар: Парабеллум, 2013. – 107 с. http://ost101.narod.ru/2013_Khagurov_Ostapenko_Reforma_obrazovaniya.pdf
2. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
3. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
4. Луценко Е.В. Реализация психологических, педагогических и профориентационных тестов и супертестов без программирования в среде интеллектуальной системы «Эйдос-Х++» (На примере теста: «Анализ особенностей индивидуального стиля педагогической деятельности») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 1057 – 1085. – IDA [article ID]: 0881304076. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/76.pdf>, 1,812 у.п.л.
5. Луценко Е.В. Реализация тестов и супертестов для ветеринарной и медицинской диагностики в среде системы искусственного интеллекта «Эйдос-Х++» без программирования / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 167 – 207. – IDA [article ID]: 0891305014. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/14.pdf>, 2,562 у.п.л.
6. Луценко Е.В. Разработка без программирования и применение в адаптивном режиме методик ризлтерской экспресс-оценки по методу аналогий (сравнительных продаж) в системно-когнитивном ана-

лизе и интеллектуальной системе «Эйдос» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №10(094). С. 507 – 564. – IDA [article ID]: 0941310036. – Режим доступа: <http://ej.kubagro.ru/2013/10/pdf/36.pdf>, 3,625 у.п.л.

7. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.

8. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.

9. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/07.pdf>, 3,875 у.п.л.

10. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (алгоритм и программная реализация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1371 – 1421. – IDA [article ID]: 1041410100. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/100.pdf>, 3,188 у.п.л.

11. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (математические аспекты) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 814 – 845. – IDA [article ID]: 1051501050. – Режим доступа: <http://ej.kubagro.ru/2015/01/pdf/50.pdf>, 2 у.п.л.

12. Луценко Е.В. Решение задач статистики методами теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №02(106). С. 1 – 47. – IDA [article ID]: 1061502001. – Режим доступа: <http://ej.kubagro.ru/2015/02/pdf/01.pdf>, 2,938 у.п.л.

13. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

14. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.

15. Луценко Е.В. Системно-когнитивный анализ и система "Эйдос" и их применение для построения интеллектуальных измерительных систем // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74.

<http://elibrary.ru/contents.asp?issueid=1267409> <http://elibrary.ru/item.asp?id=21538328>

Литература к разделу 7.1

1. Орлов А.И. Эконометрика. Учебник. М.: Издательство "Экзамен", 2002. - 576с. [Электронный ресурс]. – Режим доступа: <http://orlovs.pp.ru/econ/economet.zip>

2. Орлов А.И. Точки роста статистических методов / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал

КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 136 – 162. – IDA [article ID]: 1031409011. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/11.pdf>, 1,688 у.п.л.

3. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 163 – 195. – IDA [article ID]: 1031409012. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/12.pdf>, 2,062 у.п.л.

4. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408 с.

5. Хаббард Дуглас У. Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе. Олимп-Бизнес. 2009. – 320 с. ISBN 978-5-9693-0163-4.

6. Вяткин В.Б. Синергетическая теория информации. Часть 1. Синергетический подход к определению количества информации / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №10(044). С. 174 – 197. – Шифр Информрегистра: 0420800012\0137, IDA [article ID]: 0440810012. – Режим доступа: <http://ej.kubagro.ru/2008/10/pdf/12.pdf>, 1,5 у.п.л.

7. Вяткин В.Б. Синергетическая теория информации Часть 2. Отражение дискретных систем в плоскости признаков их описания / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(045). С. 154 – 183. – Шифр Информрегистра: 0420900012\0001, IDA [article ID]: 0450901012. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/12.pdf>, 1,875 у.п.л.

8. Вяткин В.Б. Синергетическая теория информации Часть 3. Информационные функции и энтропия Больцмана / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №02(046). С. 165 – 174. – Шифр Информрегистра: 0420900012\0015, IDA [article ID]: 0460902011. – Режим доступа: <http://ej.kubagro.ru/2009/02/pdf/11.pdf>, 0,625 у.п.л.

9. Вяткин В.Б. Хаос и порядок дискретных систем в свете синергетической теории информации / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №03(047). С. 96 – 129. – Шифр Информрегистра: 0420900012\0027, IDA [article ID]: 0470903008. – Режим доступа: <http://ej.kubagro.ru/2009/03/pdf/08.pdf>, 2,125 у.п.л.

10. Вяткин В.Б. Информационно-синергетический анализ электронных систем атомов химических элементов. Часть 1. Структурная организация электронных систем в плоскости подболочек / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №04(048). С. 24 – 44. – Шифр Информрегистра: 0420900012\0036, IDA [article ID]: 0480904003. – Режим доступа: <http://ej.kubagro.ru/2009/04/pdf/03.pdf>, 1,312 у.п.л.

11. Вяткин В.Б. Синергетическая теория информации. Часть 4. Квантовые аспекты отражения конечных множеств / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(069). С. 45 – 59. – Шифр Информрегистра: 0421100012\0169, IDA [article ID]: 0691105006. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/06.pdf>, 0,938 у.п.л.

12. Вяткин В.Б. Синергетическая теория информации: пояснения и терминологические замечания / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №06(080). С. 557 – 592. – IDA [article ID]: 0801206046. – Режим доступа: <http://ej.kubagro.ru/2012/06/pdf/46.pdf>, 2,25 у.п.л.

13. Вяткин В.Б. Орбитальная система распределения электронов в атоме и структура периодической системы элементов / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 1460 – 1493. – IDA [article ID]: 0891305100. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/100.pdf>, 2,125 у.п.л.

14. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (алгоритм и программная реализация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1371 – 1421. – IDA [article ID]: 1041410100. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/100.pdf>, 3,188 у.п.л.

15. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (математические аспекты) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 832 – 863. – IDA [article ID]: 1051501050. – Режим доступа: <http://ej.kubagro.ru/2015/01/pdf/50.pdf>, 2 у.п.л.
16. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.
17. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. [Электронный ресурс]. – Режим доступа: http://lc.kubagro.ru/aidos/aidos14_OL/index.htm
18. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. [Электронный ресурс]. – Режим доступа: http://lc.kubagro.ru/aidos/aidos14_L3/index.htm
19. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
20. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
21. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.¹²¹
22. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(011). С. 181 – 199. – IDA [article ID]: 0110503019. – Режим доступа: <http://ej.kubagro.ru/2005/03/pdf/19.pdf>, 1,188 у.п.л.
23. Луценко Е.В. Системно-когнитивный анализ функций и восстановление их значений по признакам аргумента на основе априорной информации (интеллектуальные технологии интерполяции, экстраполяции, прогнозирования и принятия решений по картографическим базам данных) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 130 – 154. – Шифр Информрегистра: 0420900012\0066, IDA [article ID]: 0510907006. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/06.pdf>, 1,562 у.п.л.
24. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(063). С. 1 – 23. – Шифр Информрегистра: 0421000012\0233, IDA [article ID]: 0631009001. – Режим доступа: <http://ej.kubagro.ru/2010/09/pdf/01.pdf>, 1,438 у.п.л.
25. Трунев А.П. Автоматизированный системно-когнитивный анализ влияния тел Солнечной системы на движение полюса Земли и визуализация причинно-следственных зависимостей в виде когнитивных функций / А.П. Трунев, Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №01(065). С. 232 – 258. – Шифр Информрегистра:

¹²¹ Для удобства читателей эта и другие работы автора размещены на личном сайте: <http://lc.kubagro.ru/>

0421100012\0002, IDA [article ID]: 0651101020. – Режим доступа: <http://ej.kubagro.ru/2011/01/pdf/20.pdf>, 1,688 у.п.л.

26. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.

27. Луценко Е.В. Развитие интеллектуальной системы «Эйдос-астра», снимающее ограничения на размерность баз знаний и разрешение когнитивных функций / Е.В. Луценко, А.П. Трунев, Е.А. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(069). С. 353 – 377. – Шифр Информрегистра: 0421100012\0159, IDA [article ID]: 0691105031. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/31.pdf>, 1,562 у.п.л.

28. Луценко Е.В. Применение СК-анализа и системы «Эйдос» для синтеза когнитивной матричной передаточной функции сложного объекта управления на основе эмпирических данных / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №01(075). С. 681 – 714. – Шифр Информрегистра: 0421200012\0008, IDA [article ID]: 0751201053. – Режим доступа: <http://ej.kubagro.ru/2012/01/pdf/53.pdf>, 2,125 у.п.л.

29. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/07.pdf>, 3,875 у.п.л.

30. Луценко Е.В. Системно-когнитивный анализ и система «Эйдос» и их применение для построения интеллектуальных измерительных систем // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74.

31. Луценко Е.В. Бандык Д.К. Подсистема визуализации когнитивных (каузальных) функций системы «Эйдос». // Е.В. Луценко (Россия), Д.К. Бандык (Белоруссия). Пат. № 2011612056 РФ. Заяв. № 2011610347 РФ 20.01.2011. Оpubл. от 09.03.2011. – Режим доступа: <http://lc.kubagro.ru/aidos/2011612056.jpg>

32. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.

33. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012\0197, IDA [article ID]: 0701106018. – Режим доступа:

34. Луценко Е.В. Системно-когнитивный анализ и система "Эйдос" и их применение для построения интеллектуальных измерительных систем. // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74. <http://elibrary.ru/item.asp?id=21538328> <http://elibrary.ru/contents.asp?issueid=1267409>

35. Луценко Е.В. Коэффициент эмерджентности классических и квантовых статистических систем / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С. 214 – 235. – IDA [article ID]: 0901306014. – Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/14.pdf>, 1,375 у.п.л.

36. Трунев А.П. Гравитационные волны и коэффициент эмерджентности классических и квантовых систем / А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 1343 – 1366. – IDA [article ID]: 0971403092. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/92.pdf>, 1,5 у.п.л.

37. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания об-

разов "Эйдос-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с. [Электронный ресурс]. - Режим доступа: <http://lc.kubagro.ru/aidos/aidos96/index.htm>

38. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2014. - №07(101). С. 1367 – 1409. - IDA [article ID]: 1011407090. - Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

Литература к разделу 7.2.1

1. Орлов А.И. Точки роста статистических методов / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2014. - №09(103). С. 136 – 162. - IDA [article ID]: 1031409011. - Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/11.pdf>, 1,688 у.п.л.

2. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2014. - №09(103). С. 163 – 195. - IDA [article ID]: 1031409012. - Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/12.pdf>, 2,062 у.п.л.

3. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2013. - №08(092). С. 859 – 883. - IDA [article ID]: 0921308058. - Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

4. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2013. - №04(088). С. 340 – 359. - IDA [article ID]: 0881304022. - Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.

5. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2009. - №10(054). С. 48 – 77. - Шифр Информрегистра: 0420900012\0110, IDA [article ID]: 0540910004. - Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.

6. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2012. - №09(083). С. 328 – 356. - IDA [article ID]: 0831209025. - Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 у.п.л.

7. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). - Краснодар: КубГАУ. 2002. - 605 с.¹²²

8. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2005. - №03(011). С. 181 – 199. - IDA [article ID]: 0110503019. - Режим доступа: <http://ej.kubagro.ru/2005/03/pdf/19.pdf>, 1,188 у.п.л.

9. Луценко Е.В. Системно-когнитивный анализ функций и восстановление их значений по признакам аргумента на основе априорной информации (интеллектуальные технологии интерполяции, экстраполяции, прогнозирования и принятия решений по картографическим базам данных) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2009. - №07(051). С. 130 – 154. - Шифр Информрегистра: 0420900012\0066, IDA [article ID]: 0510907006. - Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/06.pdf>, 1,562 у.п.л.

¹²² Для удобства читателей эта и другие работы автора размещены на личном сайте: <http://lc.kubagro.ru/>

10. Луценко Е.В. Управление агропромышленным холдингом на основе когнитивных функций связи результатов работы холдинга и характеристик его предприятий / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 248 – 260. – Шифр Информрегистра: 0420900012\0111, IDA [article ID]: 0540910015. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/15.pdf>, 0,812 у.п.л.
11. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(063). С. 1 – 23. – Шифр Информрегистра: 0421000012\0233, IDA [article ID]: 0631009001. – Режим доступа: <http://ej.kubagro.ru/2010/09/pdf/01.pdf>, 1,438 у.п.л.
12. Трунев А.П. Автоматизированный системно-когнитивный анализ влияния тел Солнечной системы на движение полюса Земли и визуализация причинно-следственных зависимостей в виде когнитивных функций / А.П. Трунев, Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №01(065). С. 232 – 258. – Шифр Информрегистра: 0421100012\0002, IDA [article ID]: 0651101020. – Режим доступа: <http://ej.kubagro.ru/2011/01/pdf/20.pdf>, 1,688 у.п.л.
13. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.
14. Луценко Е.В. Развитие интеллектуальной системы «Эйдос-астра», снимающее ограничения на размерность баз знаний и разрешение когнитивных функций / Е.В. Луценко, А.П. Трунев, Е.А. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(069). С. 353 – 377. – Шифр Информрегистра: 0421100012\0159, IDA [article ID]: 0691105031. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/31.pdf>, 1,562 у.п.л.
15. Луценко Е.В. Применение СК-анализа и системы «Эйдос» для синтеза когнитивной матричной передаточной функции сложного объекта управления на основе эмпирических данных / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №01(075). С. 681 – 714. – Шифр Информрегистра: 0421200012\0008, IDA [article ID]: 0751201053. – Режим доступа: <http://ej.kubagro.ru/2012/01/pdf/53.pdf>, 2,125 у.п.л.
16. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/07.pdf>, 3,875 у.п.л.
17. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. – Режим доступа: http://lc.kubagro.ru/aidos/aidos14_OL/index.htm
18. Луценко Е.В. Системно-когнитивный анализ и система «Эйдос» и их применение для построения интеллектуальных измерительных систем // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74.
19. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408 с.
20. Хаббард Дуглас У. Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе. Олимп-Бизнес. 2009. –320 с. ISBN 978-5-9693-0163-4.
21. Вяткин В.Б. Синергетическая теория информации. Часть 1. Синергетический подход к определению количества информации / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №10(044). С. 174 – 197. – Шифр Информрегистра: 0420800012\0137, IDA [article ID]: 0440810012. – Режим доступа: <http://ej.kubagro.ru/2008/10/pdf/12.pdf>, 1,5 у.п.л.
22. Вяткин В.Б. Синергетическая теория информации Часть 2. Отражение дискретных систем в плоскости признаков их описания / В.Б. Вяткин // Политематический сетевой электронный научный

журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(045). С. 154 – 183. – Шифр Информрегистра: 0420900012\0001, IDA [article ID]: 0450901012. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/12.pdf>, 1,875 у.п.л.

23. Вяткин В.Б. Синергетическая теория информации Часть 3. Информационные функции и энтропия Больцмана / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №02(046). С. 165 – 174. – Шифр Информрегистра: 0420900012\0015, IDA [article ID]: 0460902011. – Режим доступа: <http://ej.kubagro.ru/2009/02/pdf/11.pdf>, 0,625 у.п.л.

24. Вяткин В.Б. Хаос и порядок дискретных систем в свете синергетической теории информации / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №03(047). С. 96 – 129. – Шифр Информрегистра: 0420900012\0027, IDA [article ID]: 0470903008. – Режим доступа: <http://ej.kubagro.ru/2009/03/pdf/08.pdf>, 2,125 у.п.л.

25. Вяткин В.Б. Информационно-синергетический анализ электронных систем атомов химических элементов. Часть 1. Структурная организация электронных систем в плоскости подоболочек / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №04(048). С. 24 – 44. – Шифр Информрегистра: 0420900012\0036, IDA [article ID]: 0480904003. – Режим доступа: <http://ej.kubagro.ru/2009/04/pdf/03.pdf>, 1,312 у.п.л.

26. Вяткин В.Б. Синергетическая теория информации. Часть 4. Квантовые аспекты отражения конечных множеств / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(069). С. 45 – 59. – Шифр Информрегистра: 0421100012\0169, IDA [article ID]: 0691105006. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/06.pdf>, 0,938 у.п.л.

27. Вяткин В.Б. Синергетическая теория информации: пояснения и терминологические замечания / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №06(080). С. 557 – 592. – IDA [article ID]: 0801206046. – Режим доступа: <http://ej.kubagro.ru/2012/06/pdf/46.pdf>, 2,25 у.п.л.

28. Вяткин В.Б. Орбитальная система распределения электронов в атоме и структура периодической системы элементов / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 1460 – 1493. – IDA [article ID]: 0891305100. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/100.pdf>, 2,125 у.п.л.

29. Спиридонова О.Н. Эконометрическое моделирование стоимости квартир в г. Москва, район Замоскворечье // Материалы VI Международной студенческой электронной научной конференции «Студенческий научный форум» URL: <http://www.scienceforum.ru/2014/489/626>, <http://www.scienceforum.ru/2014/pdf/1236.pdf>, (дата обращения: 30.12.2014).

30. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.

31. Луценко Е.В. Бандык Д.К. Подсистема визуализации когнитивных (каузальных) функций системы «Эйдос». // Е.В. Луценко (Россия), Д.К. Бандык (Белоруссия). Пат. № 2011612056 РФ. Заяв. № 2011610347 РФ 20.01.2011. Оpubл. от 09.03.2011. – Режим доступа: <http://lc.kubagro.ru/aidos/2011612056.jpg>

32. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. - 456 с. - Режим доступа: <http://www.twirpx.com/file/95399/>

33. Кендалл М., Стюарт А. Многомерный статистический анализ и временные ряды, М.: Наука, Т. 3, 1976. - 736 с.. – Режим доступа: <http://www.twirpx.com/file/21817/>

34. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.

35. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012(0197, IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.
36. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (алгоритм и программная реализация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1410 – 1460. – IDA [article ID]: 1041410100. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/100.pdf>, 3,188 у.п.л.
37. Дорохов Е. В. Применение адаптивных, ARIMA и ARCH методов при прогнозировании краткосрочной динамики российского фондового рынка. [Электронный ресурс]. – Режим доступа: http://finbiz.spb.ru/download/3_2007_doroh.pdf
38. Замков О. О., Толстопятенко А. В., Черемных Ю. Н. *Математические методы в экономике: Учебник* /Московский государственный университет. - М.: ДИС, 1997. - 368 с.
39. Носач В.В. Решение задач аппроксимации с помощью персональных компьютеров / М.: Микап, 1994. - 382 с.
40. Копяев Б.В. В методе наименьших квадратов надо заменить абсолютные отклонения относительными // Заводская лаборатория. 2012. № 7. С.76-76.
41. Луценко Е.В. Системно-когнитивный анализ и система "Эйдос" и их применение для построения интеллектуальных измерительных систем. // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74. <http://elibrary.ru/item.asp?id=21538328> <http://elibrary.ru/contents.asp?issueid=1267409>
42. Орлов А.И. Эконометрика. Учебник для вузов. - М.: Экзамен, 2002, 2003 (изд. 2-е, переработанное и дополненное), 2004 (изд. 3-е, переработанное и дополненное). - 576 с. <http://ibm.bmstu.ru/nil/biblio.html#books-13-econ>).

Литература к разделу 7.2.2

1. Орлов А.И. Точки роста статистических методов / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 136 – 162. – IDA [article ID]: 1031409011. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/11.pdf>, 1,688 у.п.л.
2. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 163 – 195. – IDA [article ID]: 1031409012. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/12.pdf>, 2,062 у.п.л.
3. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
4. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
5. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 48 – 77. – Шифр Информрегистра: 0420900012(0110, IDA [article ID]: 0540910004. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.
6. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. –

№09(083). С. 328 – 356. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 у.п.л.

7. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.¹²³

8. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(011). С. 181 – 199. – IDA [article ID]: 0110503019. – Режим доступа: <http://ej.kubagro.ru/2005/03/pdf/19.pdf>, 1,188 у.п.л.

9. Луценко Е.В. Системно-когнитивный анализ функций и восстановление их значений по признакам аргумента на основе априорной информации (интеллектуальные технологии интерполяции, экстраполяции, прогнозирования и принятия решений по картографическим базам данных) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 130 – 154. – Шифр Информрегистра: 0420900012\0066, IDA [article ID]: 0510907006. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/06.pdf>, 1,562 у.п.л.

10. Луценко Е.В. Управление агропромышленным холдингом на основе когнитивных функций связи результатов работы холдинга и характеристик его предприятий / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 248 – 260. – Шифр Информрегистра: 0420900012\0111, IDA [article ID]: 0540910015. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/15.pdf>, 0,812 у.п.л.

11. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(063). С. 1 – 23. – Шифр Информрегистра: 0421000012\0233, IDA [article ID]: 0631009001. – Режим доступа: <http://ej.kubagro.ru/2010/09/pdf/01.pdf>, 1,438 у.п.л.

12. Трунев А.П. Автоматизированный системно-когнитивный анализ влияния тел Солнечной системы на движение полюса Земли и визуализация причинно-следственных зависимостей в виде когнитивных функций / А.П. Трунев, Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №01(065). С. 232 – 258. – Шифр Информрегистра: 0421100012\0002, IDA [article ID]: 0651101020. – Режим доступа: <http://ej.kubagro.ru/2011/01/pdf/20.pdf>, 1,688 у.п.л.

13. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.

14. Луценко Е.В. Развитие интеллектуальной системы «Эйдос-астра», снимающее ограничения на размерность баз знаний и разрешение когнитивных функций / Е.В. Луценко, А.П. Трунев, Е.А. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(069). С. 353 – 377. – Шифр Информрегистра: 0421100012\0159, IDA [article ID]: 0691105031. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/31.pdf>, 1,562 у.п.л.

15. Луценко Е.В. Применение СК-анализа и системы «Эйдос» для синтеза когнитивной матричной передаточной функции сложного объекта управления на основе эмпирических данных / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №01(075). С. 681 – 714. – Шифр Информрегистра: 0421200012\0008, IDA [article ID]: 0751201053. – Режим доступа: <http://ej.kubagro.ru/2012/01/pdf/53.pdf>, 2,125 у.п.л.

16. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике /

¹²³ Для удобства читателей эта и другие работы автора размещены на личном сайте: <http://lc.kubagro.ru/>

Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/07.pdf>, 3,875 у.п.л.

17. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. – Режим доступа: http://lc.kubagro.ru/aidos/aidos14_OL/index.htm

18. Луценко Е.В. Системно-когнитивный анализ и система «Эйдос» и их применение для построения интеллектуальных измерительных систем // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №5. С.64-74.

19. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408 с.

20. Хаббард Дуглас У. Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе. Олимп-Бизнес. 2009. –320 с. ISBN 978-5-9693-0163-4.

21. Вяткин В.Б. Синергетическая теория информации. Часть 1. Синергетический подход к определению количества информации / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №10(044). С. 174 – 197. – Шифр Информрегистра: 0420800012\0137, IDA [article ID]: 0440810012. – Режим доступа: <http://ej.kubagro.ru/2008/10/pdf/12.pdf>, 1,5 у.п.л.

22. Вяткин В.Б. Синергетическая теория информации Часть 2. Отражение дискретных систем в плоскости признаков их описания / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(045). С. 154 – 183. – Шифр Информрегистра: 0420900012\0001, IDA [article ID]: 0450901012. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/12.pdf>, 1,875 у.п.л.

23. Вяткин В.Б. Синергетическая теория информации Часть 3. Информационные функции и энтропия Больцмана / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №02(046). С. 165 – 174. – Шифр Информрегистра: 0420900012\0015, IDA [article ID]: 0460902011. – Режим доступа: <http://ej.kubagro.ru/2009/02/pdf/11.pdf>, 0,625 у.п.л.

24. Вяткин В.Б. Хаос и порядок дискретных систем в свете синергетической теории информации / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №03(047). С. 96 – 129. – Шифр Информрегистра: 0420900012\0027, IDA [article ID]: 0470903008. – Режим доступа: <http://ej.kubagro.ru/2009/03/pdf/08.pdf>, 2,125 у.п.л.

25. Вяткин В.Б. Информационно-синергетический анализ электронных систем атомов химических элементов. Часть 1. Структурная организация электронных систем в плоскости подоболочек / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №04(048). С. 24 – 44. – Шифр Информрегистра: 0420900012\0036, IDA [article ID]: 0480904003. – Режим доступа: <http://ej.kubagro.ru/2009/04/pdf/03.pdf>, 1,312 у.п.л.

26. Вяткин В.Б. Синергетическая теория информации. Часть 4. Квантовые аспекты отражения конечных множеств / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(069). С. 45 – 59. – Шифр Информрегистра: 0421100012\0169, IDA [article ID]: 0691105006. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/06.pdf>, 0,938 у.п.л.

27. Вяткин В.Б. Синергетическая теория информации: пояснения и терминологические замечания / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №06(080). С. 557 – 592. – IDA [article ID]: 0801206046. – Режим доступа: <http://ej.kubagro.ru/2012/06/pdf/46.pdf>, 2,25 у.п.л.

28. Вяткин В.Б. Орбитальная система распределения электронов в атоме и структура периодической системы элементов / В.Б. Вяткин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 1460 – 1493. – IDA [article ID]: 0891305100. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/100.pdf>, 2,125 у.п.л.

29. Спиридонова О.Н. Эконометрическое моделирование стоимости квартир в г. Москва, район Замоскворечье // Материалы VI Международной студенческой электронной научной конференции «Студенческий научный форум» URL: <http://www.scienceforum.ru/2014/489/626>, <http://www.scienceforum.ru/2014/pdf/1236.pdf>, (дата обращения: 30.12.2014).

30. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012/0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.
31. Луценко Е.В. Бандык Д.К. Подсистема визуализации когнитивных (каузальных) функций системы «Эйдос». // Е.В. Луценко (Россия), Д.К. Бандык (Белоруссия). Пат. № 2011612056 РФ. Заяв. № 2011610347 РФ 20.01.2011. Оpubл. от 09.03.2011. – Режим доступа: <http://lc.kubagro.ru/aidos/2011612056.jpg>
32. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. – 456 с. – Режим доступа: <http://www.twirpx.com/file/95399/>
33. Кендалл М., Стюарт А. Многомерный статистический анализ и временные ряды, М.: Наука, Т. 3, 1976. – 736 с. – Режим доступа: <http://www.twirpx.com/file/21817/>
34. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.
35. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012/0197, IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.

Литература к разделу 7.3

1. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>
2. Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики / А.И. Орлов, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>, 3,375 у.п.л.
3. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: <http://ej.kubagro.ru/2014/01/pdf/07.pdf>, 3,875 у.п.л.
4. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (алгоритм и программная реализация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1371 – 1421. – IDA [article ID]: 1041410100. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/100.pdf>, 3,188 у.п.л.
5. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (математические аспекты) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 814 – 845. – IDA [article ID]: 1051501050. – Режим доступа: <http://ej.kubagro.ru/2015/01/pdf/50.pdf>, 2 у.п.л.
6. Луценко Е.В. Решение задач статистики методами теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №02(106). С. 1 – 47. – IDA [article ID]: 1061502001. – Режим доступа: <http://ej.kubagro.ru/2015/02/pdf/01.pdf>, 2,938 у.п.л.
7. Лебедько Е.Г. Математические основы передачи информации. Ч.1, 2: учеб. пособие для вузов.- СПб: СПбГУИТМО, 2005.- 91 с. <http://www.twirpx.com/file/967816/>

8. Лебедько Е.Г. Математические основы передачи информации. Ч.3, 4: учеб. пособие для вузов.- СПб: СПбГУИТМО, 2009.- 120 с. <http://www.twirpx.com/file/421244/>
9. Лебедько Е.Г. Математические основы передачи информации. Ч.5: учеб. пособие для вузов.- СПб: СПбГУИТМО, 2010.- 93 с. <http://www.twirpx.com/file/367753/>
10. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №10(114). С. 27 – 41. – IDA [article ID]: 1141510002. – Режим доступа: <http://ej.kubagro.ru/2015/10/pdf/02.pdf>, 0,938 у.п.л.
11. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012\0197, IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.
12. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
13. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.
14. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине / Луценко Е.В. // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,750 у.п.л.
15. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
16. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.
17. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. <http://elibrary.ru/item.asp?id=18632909>
18. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. <http://elibrary.ru/item.asp?id=18271217>
19. Стабин И.П., Моисеева В.С. Автоматизированный системный анализ.- М.: Машиностроение, 1984. -309 с.
20. Перегудов Ф. И., Тарасенко Ф. П.. Основы системного анализа. Томск Изд-во науч.-техн. лит. 1997. 389с.
21. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ. М.: Высшая школа, 1989. - 320 с.,
22. Симанков В.С. Автоматизация системных исследований в альтернативной энергетике. Диссерт. на соиск. уч. ст. докт, техн. наук. По спец.: 05.13.01. <http://tekhnosfera.com/avtomatizatsiya-sistemnyh-issledovaniy-v-alternativnoy-energetike>

23. Клир Дж. Системология. Автоматизация решения системных задач. Москва: Радио и связь. 1990. 538 с. <http://www.twirpx.com/file/486296/>
24. Лефевр В.А. Конфликтующие структуры. Издание второе, переработанное и дополненное. — М.: Изд-во «Советское радио», 1973. — 158 с. с ил.
25. Хаббард Дуглас У. Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе / Дуглас У. Хаббард / [Пер. с англ. Е. Пестеревой]. — М.: ЗАО «Олимп–Бизнес», 2009. — 320 с.: ил. ISBN 978-5-9693-0163-4 (рус.). <http://www.twirpx.com/file/1546361/>
26. Луценко Е.В. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2004. — №01(003). С. 312 – 340. — IDA [article ID]: 0030401013. — Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/13.pdf>, 1,812 у.п.л.
27. Луценко Е.В. Коэффициент эмерджентности классических и квантовых статистических систем / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2013. — №06(090). С. 214 – 235. — IDA [article ID]: 0901306014. — Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/14.pdf>, 1,375 у.п.л.
28. Луценко Е.В. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2003. — №02(002). С. 146 – 164. — IDA [article ID]: 0020302013. — Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/13.pdf>, 1,188 у.п.л.
29. Луценко Е.В. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2004. — №03(005). С. 44 – 64. — IDA [article ID]: 0050403003. — Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/03.pdf>, 1,312 у.п.л.
30. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (когнитивная структуризация и формализация предметной области) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2009. — №07(051). С. 1 – 37. — Шифр Информрегистра: 0420900012\0073, IDA [article ID]: 0510907001. — Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/01.pdf>, 2,312 у.п.л.
31. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (синтез и верификация семантической информационной модели) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2009. — №07(051). С. 38 – 46. — Шифр Информрегистра: 0420900012\0072, IDA [article ID]: 0510907002. — Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/02.pdf>, 0,562 у.п.л.
32. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (прогнозирование, принятие решений и исследование предметной области) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2009. — №07(051). С. 47 – 82. — Шифр Информрегистра: 0420900012\0071, IDA [article ID]: 0510907003. — Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/03.pdf>, 2,25 у.п.л.
33. Луценко Е.В. Исследование символьных и цифровых рядов методами теории информации и АСК-анализа (на примере числа Пи с одним миллионом знаков после запятой) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2014. — №05(099). С. 319 – 355. — IDA [article ID]: 0991405022. — Режим доступа: <http://ej.kubagro.ru/2014/05/pdf/22.pdf>, 2,312 у.п.л.
34. Луценко Е.В. Применение теории информации и АСК-анализа для экспериментальных исследований в теории чисел / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: КубГАУ, 2014. — №03(097). С. 676 – 717. — IDA [article ID]: 0971403048. — Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/48.pdf>, 2,625 у.п.л.
35. Луценко Е.В. АСК-анализ проблематики статей Научного журнала КубГАУ в динамике / Е.В. Луценко, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — Краснодар: Куб-

- ГАУ, 2014. – №06(100). С. 109 – 145. – IDA [article ID]: 1001406007. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/07.pdf>, 2,312 у.п.л.
36. Луценко Е.В. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 498 – 544. – IDA [article ID]: 1031409032. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/32.pdf>, 2,938 у.п.л.
37. Луценко Е.В. Существование, несуществование и изменение как эмерджентные свойства систем. / Квантовая Магия, том 5, вып. 1, стр. 1215-1239, 2008. [Электронный ресурс]. – Режим доступа: <http://quantmagic.narod.ru/volumes/VOL512008/p1215.html>
38. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с. – Режим доступа: <http://elibrary.ru/item.asp?id=21683721>
39. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 117 – 193. – Шифр Информрегистра: 0420800012\0091, IDA [article ID]: 0410807010. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/10.pdf>, 4,812 у.п.л.
40. Луценко Е.В. Виртуализация общества как основной информационный аспект глобализации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №01(009). С. 6 – 43. – IDA [article ID]: 0090501002. – Режим доступа: <http://ej.kubagro.ru/2005/01/pdf/02.pdf>, 2,375 у.п.л.
41. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
42. Орлов А.И. Состояние и перспективы развития прикладной и теоретической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 115. С. 202–226.
43. Эйнштейн А. Собрание научных трудов в 4 томах. Том 1. Работы по теории относительности 1905-1920. Под редакцией И.Е. Тамма, Я.А. Смородинского, Б.Г. Кузнецова. - М.: Наука, 1965. - С. 613-616.
44. Колмогоров А.Н. Теория информации и теория алгоритмов. - М.: Наука, 1987. - 304 с.
45. Kolmogorov A.N. On tables of random numbers // Sankhya. Indian J. Statist. Ser.A. 1963. V.25. N 4. P. 369-376.
46. Колмогоров А.Н. Три подхода к определению понятия "количество информации" // Проблемы передачи информации. 1965. Т.1. №1. С. 3-11.
47. Колмогоров А.Н. К логическим основам теории информации и теории вероятностей // Проблемы передачи информации. 1969. Т.5. №3. С. 3-7.
48. Колмогоров А.Н. Комбинаторные основания теории информации и исчисления вероятностей // Успехи математических наук. 1983. Т.38. Вып.4. С. 27-36.
49. Орлов А.И. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 91. С. 133-162.
50. Орлов А.И. Непараметрические оценки циклов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 115. С. 183 – 201.
51. Новиков Д.А. Системный анализ и Systems analysis // Инновации в менеджменте. 2015. №3(5). С.50-56.

Научное издание

**Орлов Александр Иванович,
Луценко Евгений Вениаминович,
Лойко Валерий Иванович**

**ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ,
МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ
ОБЕСПЕЧЕНИЕ КОНТРОЛЛИНГА,
ИННОВАЦИЙ И МЕНЕДЖМЕНТА**

Монография

В авторской редакции
Компьютерная верстка – Е. В. Луценко

Подписано в печать 24.09.2016. Формат 60 × 84 ¹/₈.
Усл. печ. л. – 70,0. Уч.-изд. л. – 41,0
Тираж 45 экз. Заказ № 570.

Типография Кубанского государственного аграрного университета.
350044, г. Краснодар, ул. Калинина, 13