УДК 519.2:338 (075.8)
ББК 60.6я73
О-66

Р е ц е н з е н т ы:
заместитель директора Института проблем управления РАН
д-р техн. наук, проф., чл.-кор. РАН *Д.А. Новиков*;
заведующий кафедрой
«Системы управления экономическими объектами»
Московского авиационного института
(Государственного технического университета)
д-р экон. наук, проф. *В.Д. Калачанов*

Орлов А. И.

О-66 Организационно-экономическое моделирование : учебник : в 3 ч. / А. И. Орлов. — М. : Изд-во МГТУ им. Н. Э. Баумана, 2012.

ISBN 978-5-7038-3276-9

Ч. 3 : Статистические методы анализа данных. — 623, [1] с. : ил.

ISBN 978-5-7038-3566-1

Изложены современные методы анализа статистических данных. Рассмотрены начала выборочных исследований и основные задачи описания данных, оценивания и проверки гипотез, статистические методы анализа числовых данных, многомерный статистический анализ и статистические методы анализа динамики. Приведены основные понятия теории статистического моделирования на примерах моделей экономики и управления, медицины, социологии, демографии, истории, электротехники.

Материал учебника соответствует курсам лекций, которые автор читает в МГТУ им. Н.Э. Баумана.

Для студентов и преподавателей вузов, слушателей институтов повышения квалификации, структур второго образования и программ МВА, инженеров различных специальностей, менеджеров, экономистов, социологов, научных и практических работников, чья деятельность связана с анализом данных.

УДК 519.2:338 (075.8)
ББК 60.6я73

ISBN 978-5-7038-3566-1 (ч. 3)
ISBN 978-5-7038-3276-9

Оглавление

Предисловие

Статистические методы анализа данных активно применяются в технических исследованиях, экономике, теории и практике управления (менеджменте), социологии, медицине, геологии, истории и т. д. С результатами наблюдений (измерений, испытаний, опытов), а также с их анализом имеют дело специалисты всех отраслей практической деятельности во многих областях теоретических исследований.

Учебник позволяет овладеть современными статистическими методами на уровне, достаточном для использования этих методов в научной и практической деятельности. Он входит в серию книг по организационно-экономическому моделированию и высоким статистическим технологиям. В нем рассмотрены современные методы анализа данных, соответствующие последним научным достижениям отечественной вероятностно-статистической школы. Издание подготовлено в рамках инновационной образовательной программы «Менеджмент высоких технологий» МГТУ им. Н.Э. Баумана. Следует отметить, что нет оснований противопоставлять субъективные экспертные данные объективным результатам измерений (наблюдений, испытаний, анализов, опытов), поскольку для их описания и анализа используют одни и те же вероятностно-статистические методы и модели.

Книга предназначена для студентов и аспирантов различных специальностей, прежде всего технических, управленческих и экономических («Менеджмент высоких технологий», «Менеджмент организации»), слушателей институтов повышения квалификации, структур послевузовского (в том числе второго) образования, в частности программ MBA (Master of Business Administration — мастер делового администрирования), преподавателей вузов, сотрудников

научно-исследовательских организаций и подразделений. Учебник может быть полезен при изучении следующих дисциплин: «Организационно-экономическое моделирование», «Статистические методы», «Прикладная статистика», «Эконометрика», «Методы анализа данных», «Многомерный статистический анализ», «Общая теория статистики», «Планирование эксперимента», «Биометрика», «Теория принятия решений», «Управленческие решения», «Экономико-математическое моделирование», «Математические методы прогнозирования», «Прогнозирование и технико-экономическое планирование», «Хемометрия», «Математические методы в экономике», «Маркетинговые исследования», «Математические методы оценки», «Математические методы в социологии», «Математические методы в геологии» и т. п.

Введение

Статистические методы анализа данных применяют во многих областях деятельности человека. Целесообразно выделить три вида научной и прикладной деятельности в области применения статистических методов анализа данных (по степени специфичности методов, сопряженной с погруженностью в конкретные проблемы):

1) разработка и исследование методов общего назначения без учета специфики области применения;

2) разработка и исследование статистических моделей реальных явлений и процессов в соответствии с потребностями той или иной конкретной области применения;

3) использование статистических методов и моделей для статистического анализа конкретных данных.

По мере движения от первого вида к третьему область применения конкретного статистического метода сужается, при этом возрастает его роль для анализа определенной ситуации. Если первому виду деятельности соответствуют научные результаты, значимость которых оценивается по общенаучным критериям, то для третьего вида деятельности основным является успешное решение конкретных задач той или иной области применения (техники и технологии, экономики, социологии, медицины и др.). Второй вид деятельности занимает промежуточное положение. Это связано с тем, что, с одной стороны, теоретическое изучение свойств статистических методов и моделей, предназначенных для определенной области применения, может быть весьма сложным и математизированным, а с другой — результаты представляют интерес лишь для некоторой группы специалистов. Можно утверждать, что второй вид деятельности нацелен на решение типовых задач конкретной области применения.

Прикладная статистика. Статистические методы анализа данных, относящиеся к первому виду деятельности, обычно называют методами прикладной статистики. Таким образом, прикладная статистика — наука о том, как обрабатывать данные произвольной природы, без учета специфики конкретной области применения [1].

Математические основы прикладной статистики и статистических методов анализа данных — теория вероятностей и математическая статистика. Курс математической статистики состоит в основном из доказательств теорем, тогда как курс прикладной статистики представляет собой методологию анализа данных и алгоритмы расчетов, теоремы приводятся только для обоснования этих алгоритмов, доказательства, как правило, опускаются.

Прикладная статистика — одна из статистических наук, не относящаяся к математике, это методическая дисциплина, являющаяся центром, идейным ядром статистики. К прикладной статистике относятся задачи описания данных, оценивания и проверки гипотез.

Описание вида данных и при необходимости механизма их порождения — начало любого статистического исследования. Для описания данных применяют детерминированные и вероятностно-статистические методы. С помощью детерминированных методов можно проанализировать только данные, находящиеся в распоряжении исследователя. Например, получены таблицы, рассчитанные органами официальной государственной статистики, на основе представленных предприятиями и организациями статистических отчетов. Применить имеющиеся результаты к более широкой (генеральной) совокупности, использовать их для прогнозирования и управления можно лишь на основе вероятностно-статистического моделирования. В связи с этим в прикладную статистику часто входят методы, опирающиеся на теорию вероятностей.

Нецелесообразно противопоставлять детерминированные и вероятностно-статистические методы. Их можно

рассматривать как последовательные этапы статистического анализа. На первом этапе необходимо проанализировать имеющиеся данные, представить их в удобном для восприятия виде с помощью таблиц и диаграмм. Второй этап — изучение статистических данных на основе тех или иных вероятностно-статистических моделей. Возможность более глубокого изучения реального явления или процесса обеспечивается разработкой адекватной математической модели.

В простейшем случае статистические данные — это значения некоторого признака, свойственного изучаемым объектам. Значения признака могут быть количественными или качественными (представляют собой указание на категорию, к которой может принадлежать объект). При измерении по нескольким количественным или качественным признакам в качестве статистических данных об объекте получают вектор, который можно рассматривать как новый вид данных. В таком случае выборка состоит из набора векторов. Если часть координат вектора — числа, а часть — качественные (категоризованные) данные, то речь идет о векторе разнотипных данных.

Одним элементом выборки, т. е. одним измерением, может быть функция в целом (электрокардиограмма больного, амплитуда биений вала двигателя, временной ряд, описывающий динамику показателей хозяйственной деятельности определенной фирмы) и другие математические объекты (бинарные отношения). Так, при опросах экспертов часто используют упорядочения (ранжировки) объектов экспертизы — образцов продукции, инвестиционных проектов, вариантов управленческих решений.

Итак, математическая природа элементов выборки в разных задачах прикладной статистики может быть различной. Однако можно выделить два класса статистических данных — числовые и нечисловые данные. Соответственно прикладную статистику подразделяют на числовую и нечисловую статистику.

Числовые статистические данные — числа, векторы, функции. Их можно складывать, умножать на коэффициенты, поэтому в числовой статистике большое значение имеют разнообразные суммы. В качестве математического аппарата анализа сумм случайных элементов выборки используют классические законы больших чисел и центральные предельные теоремы.

Нечисловые статистические данные — категоризованные данные, векторы разнотипных признаков, бинарные отношения, множества, нечеткие множества и др. Их нельзя складывать и умножать на коэффициенты. Эти данные представляют собой элементы нечисловых математических пространств (множеств). Математический аппарат анализа нечисловых статистических данных основан на использовании расстояний между элементами (мер близости, показателей различия, псевдометрик) в таких пространствах. С помощью расстояний определяют эмпирические и теоретические средние величины, доказывают законы больших чисел, строят непараметрические оценки плотности распределения вероятностей, решают задачи диагностики и кластерного анализа и т. д. [2].

В прикладных исследованиях используют различные виды статистических данных, что связано, в частности, со способами их получения. Например, если испытания некоторых технических устройств продолжают до определенного момента времени, то получают так называемые цензурированные данные, состоящие из набора чисел — продолжительности работы ряда устройств до отказа — и информации о том, что остальные устройства продолжали работать в момент окончания испытания. Цензурированные данные часто применяют при оценке и контроле надежности технических устройств.

Основные области применения прикладной статистики в зависимости от вида статистических данных приведены далее (модели порождения цензурированных данных входят в состав каждой из рассматриваемых областей):

Статистические данные	Область применения прикладной статистики
Числа	Статистика (случайных) величин
Конечномерные векторы	Многомерный статистический анализ
Функции	Статистика случайных процессов и временных рядов
Объекты нечисловой природы	Статистика нечисловых данных (статистика объектов нечисловой природы)

Вероятностно-статистическое моделирование. При применении статистических методов в конкретных областях знаний и отраслях народного хозяйства получаем научно-практические дисциплины «Статистические методы в промышленности», «Статистические методы в медицине» и др. С этой точки зрения эконометрика представляет собой дисциплину «Статистические методы в экономике» [3]. Перечисленные дисциплины обычно основаны на вероятностно-статистических моделях, сформированных в соответствии с особенностями области применения.

Основная часть настоящей книги посвящена статистическим методам и вероятностно-статистическому моделированию в технико-экономических исследованиях (логистике, управлении качеством, электротехнике), в экономике и управлении (налогообложении, маркетинге), в демографии, истории, медицине и социологии.

При выборе вероятностно-статистических моделей автор во многом исходил из имеющегося у него опыта решения конкретных прикладных задач, а также старался не повторять уже известный в литературе материал. В связи с этим в издании не рассмотрены вопросы надежности и безопасности технических устройств и технологий, теории массового обслуживания, сложные системы эконометрических уравнений.

Статистический анализ конкретных данных. Применение статистических методов и моделей для статистиче-

ского анализа конкретных данных тесно связано с проблемами соответствующей области применения. Результаты третьего вида научной и прикладной деятельности находятся на стыке дисциплин (являются междисциплинарными). Эти результаты можно рассматривать как примеры практического применения статистических методов, что и сделано в настоящем учебнике. Но не меньше оснований относить их к конкретной области применения.

Примеры практического применения статистических методов включены во все главы книги. При выборе примеров предпочтение отдавалось исследованиям, в которых автор принимал непосредственное участие. Однако описание примеров было адаптировано для использования в учебном процессе. Заказчики прикладных исследований получают отчеты, в которых проблемы соответствующих областей применения рассмотрены подробнее [4].

Высокие статистические технологии. Термин «высокие технологии», популярный в современной научно-технической литературе, используют для обозначения наиболее передовых технологий, основанных на последних достижениях научно-технического прогресса [5]. Такие технологии, существующие и в технологиях статистического анализа данных, подробно изучены в настоящем учебнике.

Слово «высокие» означает, что статистическая технология опирается на современные достижения статистической теории и практики. Другими словами, математическая основа технологии получена сравнительно недавно в рамках научной дисциплины; алгоритмы расчетов разработаны и обоснованы в соответствии с нею.

Слово «статистические» подробно объясняется в данной работе. С точки зрения автора, статистические данные представляют собой результаты измерений (наблюдений, испытаний, анализов, опытов), а статистические технологии — технологии анализа статистических данных.

Наконец, сравнительно редко используемый применительно к статистике термин «технологии». Статистический анализ данных включает в себя процедуры и алгоритмы,

выполняемые последовательно, параллельно или по более сложной схеме. Можно выделить следующие этапы применения статистических технологий:

- планирование статистического исследования;
- организация сбора необходимых статистических данных по оптимальной или рациональной программе (планирование выборки, создание организационной структуры и подбор команды статистиков, подготовка кадров, которые будут заниматься сбором данных, а также контролеров данных и т. п.);
- непосредственный сбор данных и их фиксация на тех или иных носителях (с контролем качества сбора и отбраковкой ошибочных данных по соображениям, связанным с конкретной областью применения);
- первичное описание данных (расчет различных параметров выборки, характеристик, функций распределения, непараметрических оценок плотности, построение гистограмм, корреляционных полей, различных таблиц и диаграмм и т. д.);
- оценивание числовых или нечисловых характеристик, а также параметров распределений (например, непараметрическое интервальное оценивание коэффициента вариации или восстановление зависимости между откликом и факторами, т. е. оценивание функции);
- проверка статистических гипотез (иногда их цепочек — после проверки предыдущей гипотезы принимают решение о проверке последующей гипотезы);
- применение различных алгоритмов многомерного статистического анализа, алгоритмов диагностики и построения классификаций, статистики нечисловых и интервальных данных, анализа временных рядов и др.;
- проверка устойчивости полученных оценок и выводов относительно допустимых отклонений исходных данных и предпосылок используемых вероятностно-статистических моделей, в частности изучение свойств оценок методом размножения выборок;

• применение полученных статистических результатов в прикладных целях (для диагностики конкретных материалов, построения прогнозов, выбора инвестиционного проекта из предложенных вариантов, нахождения оптимального режима проведения технологического процесса, подведения итогов испытаний образцов технических устройств и др.);

• составление итоговых отчетов для тех, кто не является специалистами в статистических методах анализа данных, в том числе для руководства — лиц, принимающих решения.

Возможны и иные этапы применения статистических технологий. Квалифицированное и результативное применение статистических методов — отнюдь не проверка одной отдельно взятой статистической гипотезы или оценка параметров одного заданного распределения из фиксированного семейства. Подобного рода операции представляют собой отдельные кирпичики, из которых состоит статистическая технология.

Процедура статистического анализа данных — информационный технологический процесс (информационная технология). В настоящее время было бы несерьезно говорить об автоматизации всего процесса статистического анализа данных, поскольку существует много нерешенных проблем, вызывающих дискуссии среди статистиков.

Опишем опыт внедрения высоких статистических технологий. Организованный в 1989 г. Институт высоких статистических технологий и эконометрики (ИВСТЭ) в настоящее время действует на базе кафедры ИБМ-2 «Экономика и организация производства» Московского государственного технического университета им. Н.Э. Баумана. Институт занимается развитием, изучением и внедрением высоких статистических технологий. Основной интерес представляет применение высоких статистических технологий для анализа конкретных экономических данных. Наиболее перспективно использование высоких статистических технологий для поддержки принятия управленческих решений прежде

всего в таком новом для России современном направлении экономической науки и практики, как контроллинг.

Вначале ИВСТЭ действовал как Всесоюзный центр статистических методов и информатики Центрального правления Всесоюзного экономического общества. В 1990—1992 гг. было выполнено более 100 хоздоговорных работ, в том числе для НИЦ по безопасности атомной энергетики, ВНИИ нефтепереработки, ПО «Пластик», ФГУП «ЦНИИ черной металлургии им. И.П. Бардина», НИИ стали, ВНИИ эластомерных материалов и изделий, НИИ прикладной химии, ЦНИИ химии и механики, НПО «Орион», ВНИИ экономических проблем развития науки и техники, ПО «Уралмаш», «АвтоВАЗ», МИИТ и др. В институте

- разрабатывались эконометрические методы анализа нечисловых данных, а также процедуры расчета и прогнозирования индекса инфляции и валового внутреннего продукта (ВВП);
- развивалась методология построения и использования математических моделей процессов налогообложения (для Министерства налогов и сборов РФ), методология оценки рисков реализации инновационных проектов высшей школы (для Министерства промышленности, науки и технологий РФ);
- оценивалось влияние различных факторов на формирование налогооблагаемой базы ряда налогов (для Министерства финансов РФ);
- прорабатывались перспективы применения современных статистических и экспертных методов для анализа данных о научном потенциале (для Министерства промышленности, науки и технологий РФ);
- разрабатывалось методологическое, программное и информационное обеспечение анализа рисков химико-технологических объектов (для Международного научно-технического центра), методы использования экспертных оценок в задачах экологического страхования (совместно с Институтом проблем рынка РАН);

• проводились маркетинговые исследования (в частности, для Institute for Market Research GfK MR, Промрадтехбанка, для фирмы, торгующей растворимым кофе);

• прогнозировалось социально-экономическое развитие России методом сценариев;

• проводились работы по экономико-математическому моделированию развития малых предприятий и созданию систем информационной поддержки принятия решений.

В 2010—2012 гг. ИВСТЭ совместно с Группой компаний «Волга — Днепр» и Ульяновским государственным университетом активно участвует в проекте «Разработка математического аппарата, программного и информационного обеспечения автоматизированной системы прогнозирования и предотвращения авиационных происшествий при организации и производстве воздушных перевозок».

Программное обеспечение статистических методов. Как правило, статистическую обработку данных проводят с помощью соответствующих программных продуктов. В учебник не были включены ссылки на программные продукты по следующим причинам: быстрое обновление программных продуктов; каждый программный продукт обладает определенными достоинствами и недостатками, в связи с чем крайне трудно обосновать, какой из программных продуктов следует предпочесть [6].

С течением времени различие между математической и прикладной статистикой усиливается. Это проявляется в том, что большинство методов, входящих в статистические пакеты программ (например, в Statgraphics и SPSS или в Statistica), даже не упоминаются в учебниках по математической статистике. В результате этого специалист по математической статистике оказывается зачастую беспомощным при обработке реальных данных, а программные продукты по статистических методам применяют лица без необходимой теоретической подготовки [7].

По оценкам экспертов, распространенные статистические программные продукты обычно соответствуют уров-

ню научных исследований 1960—1970-х гг. В них отсутствует большинство статистических методов, включенных в современные учебники [1—3].

Перспективы развития статистических методов. Теория статистических методов нацелена на решение реальных задач, поэтому в ней постоянно возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится с помощью математических средств, т. е. путем доказательства теорем. При разработке и применении статистических методов важна методологическая составляющая (как именно сформулировать задачи, какие предположения принять для дальнейшего математического изучения), а также современные информационные технологии, в частности компьютерный эксперимент.

Актуальной является задача анализа истории статистических методов для выявления тенденций их развития, применения тенденций для прогнозирования и планирования исследований в области статистических методов.

Ситуация с внедрением современных статистических методов на отечественных предприятиях и в организациях различных отраслей народного хозяйства внушает оптимизм. Продолжают развиваться структуры, в которых требуются статистические методы, — подразделения качества, надежности, управления персоналом, центральные заводские лаборатории и др. В последние годы получили распространение службы контроллинга, маркетинга и сбыта, логистики, сертификации, прогнозирования и планирования, инноваций и инвестиций, управления рисками, экологии, использующие различные статистические методы (в частности, методы экспертных оценок). Рассмотренные в учебнике методы необходимы органам государственного и муниципального управления, организациям силовых ведомств, транспорта и связи, медицины, образования, агропромышленного комплекса, научным и практическим работникам всех областей деятельности.

Основные этапы становления статистических методов. В качестве примера первого применения статистических методов можно привести Библию, Ветхий Завет. Там описана процедура и даны результаты переписи военнообязанных. Само слово «статистика» происходит от латинского слова *status* — состояние дел. Вначале под статистикой понимали описание экономического и политического состояния государства или его части. Например, к 1792 г. относится следующее определение: статистика описывает состояние государства в настоящее время или в некоторый известный момент в прошлом. И сейчас деятельность государственных статистических служб достаточно хорошо соответствует этому определению.

Однако постепенно термин «статистика» стал использоваться более широко. Так, Наполеон Бонапарт под этим термином понимал «бюджет вещей». 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», т. е. статистика уже не связывается ни с государствоведением, ни с социально-экономическими проблемами.

В 1954 г. академик 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» [8].

Под «статистикой» также часто понимают набор количественных данных о некотором явлении или процессе. Специалисты в области статистических методов «статистикой» называют функцию результатов наблюдений, используемую для оценивания характеристик и параметров распределений и проверки гипотез.

После возникновения теории вероятностей как науки (Паскаль, Ферма, XVII в.) вероятностные модели стали использоваться при обработке статистических данных. В 1794 г. К. Гаусс разработал метод наименьших квадратов (гл. 6), один из наиболее популярных статистических методов, и применил его при расчете орбиты астероида Церера. В XIX в. заметный вклад в развитие практической статистики внес А. Кетле (1791—1874), на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей.

Параметрическая статистика. С 1900 г. были изучены методы, основанные на анализе данных из параметрических семейств распределений. Наиболее распространенным 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. Были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

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

С математической точки зрения параметрическая статистика позволяет получить теоретические схемы, на основе которых построена теория. Профессионалам следует обра-

тить внимание на теорию достаточных статистик, неравенство Рао — Крамера, теорию оптимального оценивания и др.

Параметрическую статистику часто критикуют, так как нельзя указать каких-либо веских причин, по которым распределение результатов конкретных наблюдений непременно должно входить в параметрическое семейство [9].

Статистические методы в России. В России были получены многие фундаментальные результаты прикладной статистики [10]. Первое статистико-экономическое обозрение России составлено И.К. Кирилловым (1689—1737), обер-секретарем Сената, под названием «Цветущее состояние Всероссийского государства» [11]. Научный труд по вопросам организации учета населения в России «Рассуждение о ревизии поголовной и касающемся до оной» написан в 1747 г. В.Н. Татищевым (1686—1750), известным государственным деятелем. Он одним из первых применял анкеты для сбора статистических данных. Большой вклад в теорию и практику отечественной статистики внес М.В. Ломоносов (1711—1765).

Огромное значение имеют работы А.Н. Колмогорова (1903—1987), которые дали первоначальный толчок к развитию ряда направлений прикладной статистики, а также Н.В. Смирнова (1900—1966) и Л.Н. Большева (1922—1978) [10]. До сих пор для специалистов важны работы А.Н. Колмогорова по аксиоматическому подходу к теории вероятностей, по критерию согласия эмпирического распределения с теоретическим распределением, по свойствам медианы как оценки центра распределения, по эффекту «вздувания» коэффициента корреляции, по теории средних величин, по статистической теории кристаллизации металлов, по методу наименьших квадратов, по свойствам сумм случайного числа случайных слагаемых, по статистическому контролю, по несмещенным оценкам, по аксиоматическому получению логарифмически нормального закона распределения при дроблении, по методам обнаружения различий при экспериментах типа погодных [12].

Идеи А.Н. Колмогорова продолжил развивать его ученик Б.В. Гнеденко (1912—1995), занимавшийся предельными теоремами теории вероятностей, математической статистикой, теорией надежности, статистическими методами управления качеством, теорией массового обслуживания [13]. По его мнению, важнейшими аспектами востребованности теории и успешного применения ее на практике являются:

- наличие в теории богатого набора математических моделей, отражающих разнообразные явления предметной области;
- наличие в предметной области специалистов, способных понять математические модели и превратить их в «руководящие указания» на производстве;
- наличие литературы самого разного уровня, отражающей достижения теории и практику ее применения;
- возможность прямого контакта между создателями теории и специалистами предметной области для взаимной корректировки задач теории и методов ее приложения в предметной области.

Статистические методы применял В.В. Налимов (1910—1997) — создатель и руководитель нескольких новых научных направлений: метрологии количественного анализа, химической кибернетики, математической теории эксперимента и наукометрии. Он также 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.

Наряду с перечисленными исследователями следует отметить А.Я. Хинчина, С.Н. Бернштейна, Е.Е. Слуцкого, В.С. Немчинова, В.И. Романовского, К. Круга, А.А. Любищева, А.П. Вощинина и др. В 1990 г. была образована Всесоюзная статистическая ассоциация (ВСА), объединившая статистиков всех направлений — специалистов по при-

кладной и математической статистике, по надежности (в основном представителей оборонно-промышленного комплекса), преподавателей экономико-статистических дисциплин, работников официальной государственной статистики. Ведущую роль в создании ВСА сыграл Всесоюзный центр статистических методов и информатики [15]. Отметим выпуск энциклопедии «Вероятность и математическая статистика» [16], содержащей полезную информацию для специалистов по статистическим методам.

Работы по прикладной статистике 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. Отечественные работы по статистическим методам в основном публикуются в журнале «Заводская лаборатория» в разделе «Математические методы исследования», созданном в 1961 г. В нем за 50 лет помещено около 1 000 статей по различным направлениям прикладной статистики, прежде всего по статистическому анализу числовых величин, по статистике нечисловых данных, по многомерному статистическому анализу, по планированию эксперимента, по опыту применения статистических методов при решении конкретных прикладных задач.

Точки роста. Выделим пять актуальных направлений (точек роста), в которых развивается современная прикладная статистика: непараметрическая статистика (непараметрика), устойчивость статистических процедур (робастность), бутстреп (размножение выборок), статистика интервальных данных, статистика нечисловых данных (в другой терминологии — статистика объектов нечисловой природы, нечисловая статистика).

1. 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 непарамет-

рическими. В первой трети XX в. в работах Ч. Спирмена (1863—1945) и М. Кендалла (1907—1983) были описаны первые методы непараметрической статистики, основанные на коэффициентах ранговой корреляции. Но непараметрическая статистика, не содержащая нереалистических предположений о принадлежности функции распределения результатов наблюдений тем или иным параметрическим семействам распределений, стала заметной частью статистики лишь со второй трети XX в. В 1930-е гг. появились работы А.Н. Колмогорова и Н.В. Смирнова, предложивших и изучивших статистические критерии. После Второй мировой войны развитие непараметрической статистики пошло быстрыми темпами. Большой вклад в развитие статистики внес Ф. Вилкоксон (1892—1965). К настоящему времени с помощью методов непараметрической статистики можно решать практически те же статистические задачи, что и с помощью методов параметрической статистики. Важную роль играют непараметрические оценки плотности, непараметрические методы регрессии и распознавания образов (дискриминантного анализа).

2. Устойчивость статистических процедур (робастность). Если в параметрических постановках на вероятностные модели статистических данных накладываются слишком жесткие требования (их функции распределения должны принадлежать определенному параметрическому семейству), то в непараметрических постановках — излишне слабые требования (функции распределения должны быть непрерывны). При этом игнорируется априорная информация о «примерном виде» распределения. Априори можно ожидать, что учет «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 распределения некоторого параметрического семейства. За рубежом эту теорию разрабатывали П. Хубер, Ф. Хампель и др.

Частными случаями реализации идеи устойчивости статистических процедур являются статистика объектов нечисловой природы и статистика интервальных данных.

Существует много моделей устойчивости в зависимости от того, какие именно отклонения от заданного параметрического семейства допускаются. Среди теоретиков наиболее популярной оказалась модель выбросов, в которой исходная выборка «засоряется» малым числом выбросов, имеющих принципиально иное распределение. Более перспективна модель малых отклонений распределений, в которой расстояние между распределением каждого элемента выборки и базовым распределением не превосходит заданного минимального значения, и модель статистики интервальных данных [2, 17].

***3. Бутстреп* (*размножение выборок*)**. Бутстреп связан с интенсивным использованием возможностей компьютеров. Основная идея заключается в замене теоретического исследования вычислительным экспериментом. Например, вместо описания выборки распределением из параметрического семейства формируется большое число «похожих» выборок, т. е. осуществляется размножение выборки. Далее на основе свойств теоретического распределения с помощью вычислительного метода решаются задачи, рассчитываются интересующие статистики по каждой из «похожих» выборок и анализируются полученные распределения. Квантили этого распределения задают доверительные интервалы и т. д.

Предположим, что по выборке делаются какие-либо статистические выводы. Насколько эти выводы устойчивы? Если есть другие (контрольные) выборки, описывающие это же явление, можно применить к ним ту же статистическую процедуру и сравнить результаты. Если таких выборок не существует, то следует их построить искусственно. Выбирается исходная выборка и исключается один элемент. Имеется похожая выборка, взятая из того же распределения, только с объемом на единицу меньше. Затем воз-

вращается этот элемент выборки и исключается другой. Получается вторая похожая выборка. Поступая так со всеми элементами исходной выборки, имеем число выборок, похожих на исходную выборку, равное ее объему. Остается обработать выборки тем же способом, что и исходную выборку, и изучить устойчивость получаемых выводов — разброс оценок параметров, частоты принятия или отклонения гипотез и т. д.

Есть много способов развития идеи размножения выборок. Первый вариант — построение по исходной выборке эмпирической функции распределения, а затем переход каким-либо образом от кусочно-постоянной функции к непрерывной функции распределения, например соединение точек $\left(x(i);\ i/n\right),\ i=1,\ 2,\ ...,\ n,$ отрезками прямых. Второй вариант перехода к непрерывному распределению — построение непараметрической оценки плотности. После этого рекомендуется брать размноженные выборки из этого распределения (являющегося состоятельной оценкой исходного распределения), непрерывность защитит от совпадений элементов в этих выборках.

Третий 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 варианте соединяем вместе идеи устойчивости и бутстрепа. Поскольку всегда имеются погрешности измерения, то реальные данные — это не числа, а интервалы (результат измерения плюс-минус погрешность).

4. Статистика интервальных данных. Перспективное и быстро развивающееся направление последних лет — статистика интервальных данных, в которой рассматриваются асимптотические методы статистического анализа интервальных данных при больших объемах выборок и малых погрешностях измерений [2].

В рамках данного научного направления:

- разработана общая схема исследования, включающая в себя расчет нотны (максимально возможного отклонения статистики, вызванного интервальностью исходных данных) и рационального объема выборки (превышение этого объема не дает существенного повышения точности оценивания);
- оценены математическое ожидание, дисперсия, коэффициент вариации, параметры гамма-распределения и характеристики аддитивных статистик;
- осуществлена проверка гипотез о параметрах нормального распределения, в том числе 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;
- изучено влияние погрешностей измерений и наблюдений на свойства алгоритмов регрессионного анализа,
- введены и исследованы новые понятия многомерных и асимптотических нотн и доказаны соответствующие предельные теоремы;
- разработан интервальный дискриминантный анализ, в частности, установлено влияние интервальности данных на показатель качества классификации;
- изучено асимптотическое поведение оценок метода моментов и оценок максимального правдоподобия, а также более общих оценок минимального контраста и проведено асимптотическое сравнение точности указанных выше методов в случае интервальных данных;
- найдены условия, при которых в отличие от классической математической статистики метод моментов дает более точные оценки, чем метод максимального правдоподобия.

В области асимптотической статистики интервальных данных российская наука имеет мировой приоритет. Во все

виды статистического программного обеспечения включают алгоритмы интервальной статистики, «параллельные» обычно используемым алгоритмам прикладной математической статистики. Это позволяет в явном виде учесть наличие погрешностей результатов наблюдений.

5. Статистика нечисловой природы. Анализ динамики развития прикладной статистики приводит к выводу, что в XXI в. статистика нечисловой природы станет центральной областью прикладной статистики, поскольку содержит наиболее общие подходы и результаты.

Исходный объект прикладной математической статистики — выборка. В вероятностной теории статистики выборка представляет собой совокупность независимых одинаково распределенных случайных элементов. Какова природа этих элементов? В классической математической статистике элементы выборки — числа, в многомерном статистическом анализе — векторы, в нечисловой статистике элементы выборки — объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы принадлежат пространствам, не имеющим линейной (векторной) структуры [2].

С начала 1970-х гг. под влиянием запросов прикладных исследований в социально-экономических, технических, медицинских науках в России активно развивается статистика объектов нечисловой природы. В создании этой сравнительно новой области прикладной математической статистики приоритет принадлежит российским ученым. Большое значение для развития нечисловой статистики имели запросы теории и практики экспертных оценок [18].

Учебник состоит из трех частей (12 глав). В части I (главы 1—4) рассмотрены проблемы организации выборочных исследований на примере двух конкретных маркетинговых опросов, модели случайных выборок, в том числе гипергеометрическая и биномиальная, методы доверительного оценивания доли и проверки однородности двух биномиальных выборок, модели порождения данных, методы

их описания с помощью таблиц и диаграмм, выборочных характеристик и эмпирической функции распределения, непараметрических оценок плотности (в пространствах произвольной природы). Показано, что распределение результатов наблюдений (испытаний, измерений, анализов, опытов), как правило, отличается от нормального распределения. Большое внимание уделено непараметрическим методам анализа статистических данных, методам оценивания параметров и характеристик. Разработаны и изучены одношаговые оценки для замены устаревших оценок максимального правдоподобия. Исследована асимптотика решений экстремальных статистических задач и устойчивость (робастность) статистических процедур. Оценивание для сгруппированных данных построено на основе формулы Эйлера — Маклорена и поправок Шеппарда. Для проверки гипотез разработан метод моментов, реализованный на примере гипотезы согласия с гамма-распределением. Продемонстрирована крайняя неустойчивость параметрических методов отбраковки выбросов, приводящая к выводу о невозможности их научно обоснованного использования. Сформулирована предельная теория непараметрических критериев, опирающаяся на метод приближения ступенчатыми функциями. Разработан метод проверки гипотез по совокупности малых выборок для применения в асимптотике растущей размерности, когда число неизвестных параметров увеличивается вместе с объемом данных. Рассмотрена проблема множественных проверок статистических гипотез, актуальная при разработке высоких статистических технологий анализа данных.

В части II (главы 5—7) приведены конкретные статистические методы анализа данных различных типов. Разобраны методы точечного и доверительного непараметрического оценивания основных характеристик распределения (математического ожидания, медианы, дисперсии, среднего квадратического отклонения, коэффициента вариации), методы проверки однородности характеристик двух незави-

симых выборок, обоснована необходимость использования непараметрического критерия Крамера — Уэлча вместо статистики критерия Стьюдента. Изучены свойства двухвыборочного критерия Вилкоксона, обосновано применение состоятельных критериев проверки однородности независимых выборок. Разработаны методы проверки однородности связанных выборок, в том числе на основе критериев проверки гипотезы симметрии. Перечислены основные постановки многомерного статистического анализа. Рассмотрены линейный (Пирсона) и непараметрические (Спирмена, Кендалла) коэффициенты парной корреляции. Изложена задача восстановления линейной зависимости между двумя переменными на основе непараметрического метода наименьших квадратов, а также основы линейного регрессионного анализа, теории индексов, в том числе индексов потребительских цен, статистические методы анализа динамики, в том числе методы анализа и прогнозирования временны́х рядов и системы эконометрических уравнений. Включены оригинальные подходы к оцениванию периода и периодической составляющей сигналов.

Часть III (главы 8—12) посвящена вероятностно-статистическому моделированию в различных областях применения [19]. Рассмотрены основные понятия теории статистического моделирования; демографические модели; статистические модели движения товарных потоков в процессе работы склада (модели логистики); статистическое моделирование исторических процессов, позволившее существенно уточнить хронологию древнего мира и средневековья; вероятностно-статистическое моделирование помех, создаваемых электровозами. Описан подход к моделированию взаимовлияний факторов методом Жихарева — Орлова — Кольцова, на основе которого разработана система моделей налогообложения и проанализированы макроэкономические балансовые соотношения. Изучена эконометрическая база метода — моделирование и анализ многомерных временны́х рядов. Рассмотрены комплекс

статистических методов управления качеством, в том числе методы обнаружения разладки с помощью контрольных карт, весьма актуальные не только для организации производства, но и в менеджменте; статистическое моделирование в экспертных исследованиях. Приведены примеры процедур экспертных оценок, выделены основные стадии экспертного опроса. В качестве примера применения общенаучной теории измерений получены правила выбора вида средних величин в зависимости от типов шкал, в которых измерены ответы экспертов. Показано использование методов средних арифметических и медиан баллов в сочетании с процедурами согласования кластеризованных ранжировок. Рассмотрены математические методы анализа экспертных оценок, в частности расстояние Кемени и медиана Кемени, в пространствах бинарных отношений; медико-статистические технологии в научных медицинских исследованиях; проблемы внедрения высоких статистических технологий. Проанализировано развитие статистического инструментария отечественных социологов за последние 30 лет, изложены перспективы применения люсианов, асимптотика квантования и выбор числа градаций в социологических анкетах.

Автор настоящего учебника более 40 лет постоянно занимается статистическими методами. В издание включены теоретические и практические результаты, полученные им в 1970-х гг. и в последние годы. Литературные ссылки помогут углубленно изучить материал. В части 1 учебника помещена краткая информация о деятельности автора как научного работника и преподавателя, о ранее выпущенных им монографиях, учебниках, учебных пособиях.

В отличие от учебной литературы по математическим дисциплинам, в настоящей книге практически отсутствуют доказательства. Однако в нескольких случаях они приведены.

Автор благодарен сотрудникам редакции Издательства МГТУ им. Н.Э. Баумана, членам редколлегии и секции «Математические методы исследования» журнала «Завод-

ская лаборатория», всему коллективу кафедры ИБМ-2 «Экономика и организация производства» МГТУ им. Н.Э. Баумана и заведующему кафедрой профессору С.Г. Фалько за постоянную поддержку проектов по разработке и внедрению организационно-экономических, эконометрических и статистических курсов, членам Ученого совета, поддержавшим инициативу о введении статистических методов в учебный процесс, декану факультета «Инженерный бизнес и менеджмент» профессору И.Н. Омельченко за совместные научные исследования, рецензентам — заведующему кафедрой «Системы управления экономическими объектами» Московского авиационного института В.Д. Калачанову и заместителю директора Института проблем управления РАН Д.А. Новикову.

Автор благодарен за помощь в написании гл. 11 сыну А.А. Орлову и жене Л.А. Орловой.

С текущей научной информацией по теории и практике статистических методов анализа данных можно ознакомиться на сайте «Высокие статистические технологии» http://orlovs.pp.ru, а также на странице «Лаборатория экономико-математических методов в контроллинге» http://www.ibm.bmstu.ru/nil/lab.html (сайт научно-учебного комплекса «Инженерный бизнес и менеджмент» МГТУ им. Н.Э. Баумана). Достаточно большой объем информации содержит еженедельник «Эконометрика».

Читатели могут сообщать свои вопросы и замечания по адресу Издательства или непосредственно автору по электронной почте E-mail: prof-orlov@mail.ru.

Литература

1. *Орлов А.И.* Прикладная статистика. М.: Экзамен, 2006. 671 с.

2. *Орлов А.И.* Организационно-экономическое моделирование: В 3 ч. Ч.1: Нечисловая статистика. М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. 541 с.

3. *Орлов А.И.* Эконометрика. Ростов н/Д: Феникс, 2009. 572 с.

4. *Математическое* моделирование процессов налогообложения (подходы к проблеме)/ А.И. Орлов, М.А. Кастосов, Н.Ю. Иванова и др. М.: Изд-во ЦЭО Минобразования РФ, 1997. 232 с.

5. *Орлов А.И.* Высокие статистические технологии // Заводская лаборатория. 2003. Т. 69. № 11. С. 55–60.

6. *Орлов А.И.* Математическое обеспечение сертификации: сравнительный анализ диалоговых систем по статистическому контролю // Заводская лаборатория. 1996. Т. 62. № 7. С. 46—49.

7. *Орлов А.И.* Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория. 1985. Т. 51. №1. С. 60—62.

8. *Никитина Е.П., Фрейдлина В.Д., Ярхо А.В.* Коллекция определений термина «статистика». М.: МГУ, 1972. 46 с.

9. *Орлов А.И.* О развитии прикладной статистики // Современные проблемы кибернетики (прикладная статистика). М.: Знание, 1981. С. 3—14.

10. *Большев Л.Н.*, *Смирнов Н.В.* Таблицы математической статистики. М.: Наука, 1983. 416 с.

11. *Плошко Б.Г., Елисеева И.И.* История статистики. М.: Финансы и статистика. 1990. 295 с.

12. *Кудлаев Э.М., Орлов А.И.* Вероятностно-статистические методы исследования в работах А.Н. Колмогорова // Заводская лаборатория. 2003. Т. 69. № 5. С. 55—61.

13. *Орлов А.И.* Математические методы исследования в работах Бориса Владимировича Гнеденко // Заводская лаборатория. 2007. Т. 73. №7. С. 66—72.

14. *Смирнов Н.В.* Теория вероятностей и математическая статистика: Избранные труды. М.: Наука, 1970. 289 с.

15. *Kotz S., Smith K.* The Hausdorff Space and Applied Statistics: A View from USSR // The American Statistician. November 1988. Vol. 42. No 4. P. 241—244.

16. *Вероятность* и математическая статистика. Энциклопедия / Под ред. Ю.В. Прохорова. М.: Большая Российская Энциклопедия, 1999. 910 с.

17. *Орлов А.И.* Устойчивость в социально-экономических моделях. М.: Наука, 1979. 296 с.

18. *Орлов А.И.* Организационно-экономическое моделирование: В 3 ч. Ч. 2: Экспертные оценки. М.: Изд-во МГТУ им. Н.Э. Баумана. 2011. 486 с.

19. *Неуймин Я.Г.* Модели в науке и технике. История, теория, практика. Л.: Наука, 1984. 190 с.

Часть I

Основные постановки задач анализа данных

Глава 1. Выборочные исследования

Термин «выборочные исследования» применяют, когда невозможно изучить все элементы представляющей интерес совокупности. Приходится знакомиться с частью совокупности, т. е. с выборкой, а затем с помощью вероятностно-статистических методов и моделей переносить выводы, сделанные при рассмотрении выборки, на совокупность в целом. Выборочные исследования включают в себя способы получения и анализа статистических данных, поэтому составляют важный раздел статистических методов, эконометрики и прикладной статистики [1].

1.1. Организация выборочных исследований

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

Оценка функции спроса. Функция спроса часто встречается в учебниках по экономической теории, при этом обычно не рассказывается, как она получена. Однако оценить ее по эмпирическим данным не так уж трудно. Например, можно выяснить ожидаемый спрос с помощью простого приема: узнать у потенциальных потребителей, какую максимальную цену они готовы заплатить за определенный товар. Пусть выборка состоит из 20 опрошенных потребителей, которые назвали следующие максимально допустимые для них цены, руб.: 40; 25; 30; 50; 35; 20; 50; 32; 15; 40; 20; 40; 45; 30; 50; 25; 35; 20; 35; 40.

Упорядочим приведенные значения в порядке возрастания и сведем их в табл. 1.1. В первом столбце указаны номера различных значений, названных потребителями, во

Глава 2. Описание данных

Выделяют три основные области статистических методов обработки результатов наблюдений — описание данных, оценивание (характеристик и параметров распределений, регрессионных зависимостей и др.) и проверка статистических гипотез.

Величины, используемые при описании данных, применяют на дальнейших этапах статистического анализа — оценивании и проверке гипотез, а также при решении задач, возникающих при применении вероятностно-статистических методов принятия решений, например при статистическом контроле качества продукции и статистическом регулировании технологических процессов.

2.1. Модели порождения данных

Статистические данные — результаты наблюдений (измерений, испытаний, опытов, анализов). Функции результатов наблюдений, используемые, в частности, для оценки параметров распределений и (или) проверки статистических гипотез, называют статистиками[*]. Если в вероятностной модели результаты наблюдений рассматривают как случайные величины (случайные элементы), то статистики как функции случайных величин (элементов) являются случайными величинами (элементами). Статистики — выборочные аналоги характеристик случайных величин (математического ожидания, медианы, дисперсии, моментов и

[*] С точки зрения математиков, речь идет об измеримых функциях.

Глава 3. Оценивание

При применении статистических методов необходимо оценивать параметры распределений, функции распределения и их плотности, зависимости между переменными и другие составляющие организационно-экономических моделей. Часто используют модели на основе параметрических семейств распределений, в которых следует оценить значение параметра распределения. Методы статистического оценивания определяются применяемой моделью. В гл. 3 рассмотрено оценивание параметров, а также оценивание путем решения экстремальных статистических задач, к которым сводятся многие постановки прикладной статистики, и робастные (устойчивые) методы оценивания, в том числе по сгруппированным данным.

3.1. Методы оценивания параметров

Некоторые статистические методы основаны на параметрических моделях. Термин «параметрический» означает следующее: вероятностно-статистическая модель полностью описывается конечномерным вектором фиксированной размерности. Причем размерность не зависит от объема выборки. Далее приведены примеры методов оценивания, используемых в параметрических моделях.

Рассмотрим выборку $x_1, x_2, \ldots, x_n$ из распределения с плотностью $f(x; \theta_0)$, где $f(x; \theta_0)$ — элемент параметрического семейства плотностей распределения вероятностей $\{f(x; \theta), \theta \in \Theta\}$; Θ — k-мерное пространство параметров, являющееся подмножеством евклидова пространства $\boldsymbol{R}^k$; конкретное значение параметра θ_0 неизвестно. Обычно в

Глава 4. Проверка гипотез

Глава посвящена одному из основных разделов статистических методов — избранным задачам проверки статистических гипотез. Разработан метод моментов, реализованный на примере гипотезы согласия с гамма-распределением. Продемонстрирована крайняя неустойчивость параметрических методов отбраковки выбросов, приводящая к выводу о невозможности их научно обоснованного использования. Изложена проблема множественных проверок статистических гипотез, актуальная при разработке высоких статистических технологий анализа данных.

4.1. Метод моментов проверки гипотез

Как уже было отмечено в гл. 3, к методу моментов относят все статистические процедуры, основанные на использовании выборочных моментов и их функций. В непараметрической статистике на основе выборочных моментов проводится точечное и интервальное оценивание таких характеристик распределения, как математическое ожидание, дисперсия, среднее квадратическое отклонение, коэффициент вариации (см. гл. 5). Для проверки гипотез в непараметрической статистике также используется метод моментов. Пример — критерий Крамера — Уэлча для проверки равенства математических ожиданий по двум независимым выборкам.

В практике применения статистических методов (согласно классическим схемам параметрической статистики) довольно часто возникает необходимость проверки гипотезы о принадлежности функции распределения результатов

Часть II

Конкретные статистические методы

Глава 5. Статистические методы анализа числовых выборок

Рассмотрено несколько типовых задач анализа числовых данных, часто встречающихся при применении статистических методов в различных областях научных исследований и отраслях народного хозяйства. В настоящей главе выборка моделируется как совокупность независимых одинаково распределенных числовых случайных величин с произвольной функцией распределения.

5.1. Оценивание основных характеристик распределения

Существенная часть алгоритмов статистического анализа данных исходит из предположения о нормальности распределения результатов наблюдений. Как уже было отмечено (см. гл. 2), распределения погрешностей физических измерений, как правило, отличны от нормальных распределений. Вследствие наличия отклонений от нормальности свойства алгоритмов могут в одних случаях изменяться сравнительно слабо, как при проверке гипотезы однородности математических ожиданий для выборок равного объема. В других случаях изменения свойств таковы, что алгоритмы из научных переходят в эвристические. Например, свойства алгоритмов отбраковки выбросов (резко выделяющихся наблюдений) крайне неустойчивы по отношению к отклонениям от нормальности: если зафиксировать правило отбраковки, то крайне неустойчив уровень значимости, а если зафиксировать уровень значимости, то крайне

Глава 6. Многомерный статистический анализ

В многомерном статистическом анализе выборка состоит из элементов многомерного пространства. Отсюда и название этого раздела статистических методов. Из многих задач многомерного статистического анализа рассмотрены основные задачи корреляции, восстановления зависимости, индексы.

6.1. Коэффициенты корреляции

Термин «корреляция» означает связь. В области статистических методов этот термин обычно используется в словосочетании «коэффициенты корреляции». Рассмотрим линейные и непараметрические парные коэффициенты корреляции как способы измерения связи двух случайных переменных.

Исходные данные — набор случайных векторов $(x_i, y_i) = (x_i(\omega), y_i(\omega))$, $i = 1, 2, \ldots, n$. ***Выборочным коэффициентом корреляции, или выборочным линейным парным коэффициентом корреляции Пирсона,*** называется число

$$r_n = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}.$$

При $r_n = 1$ $y_i = ax_i + b$, причем $a > 0$. Если $r_n = -1$, то $y_i = ax_i + b$, причем $a < 0$. Таким образом, близость коэф-

Глава 7. Статистические методы анализа динамики

Анализ динамики — это анализ временных рядов. Под временными рядами понимают детерминированные или случайные функции времени. Время предполагается дискретным, в противном случае говорят о случайных процессах, а не о временных рядах.

7.1. Методы анализа и прогнозирования временных рядов

Модели стационарных и нестационарных временных рядов. Пусть $t = 0, \pm 1, \pm 2, \pm 3, \ldots$ Сначала рассмотрим временной ряд $X(t)$, который принимает числовые значения (цена на хлеб или курс обмена доллара на рубли). Обычно в поведении временного ряда выявляют две основные тенденции — тренд и периодические колебания.

Под ***трендом*** понимают зависимость $X(t)$ от времени линейного, квадратичного или другого типа, которую выявляют тем или иным способом эмпирического сглаживания (например, экспоненциального) либо модельно-расчетным путем, в частности с помощью метода наименьших квадратов. Другими словами, тренд — очищенная от случайностей основная тенденция временного ряда.

Временной ряд обычно колеблется вокруг тренда, причем отклонения от тренда часто обнаруживают правильность. В основном это связано с естественной или назначенной периодичностью — сезонной или недельной, месячной или квартальной (например, в соответствии с

Часть III

Вероятностно-статистическое моделирование

Глава 8. Основы вероятностно-статистического моделирования

Рассмотрена устойчивость статистических выводов, разработана общая схема устойчивости и предложен принцип уравнивания погрешностей. Приведены примеры вероятностно-статистического моделирования в демографии, логистике, истории и электротехнике.

8.1. Основные понятия теории вероятностно-статистического моделирования

Модель (обобщенная модель) — создаваемый в целях получения и (или) хранения информации специфический объект (в форме мысленного образа, описания знаковыми средствами либо материальной системы), который отражает свойства, характеристики и связи объекта-оригинала произвольной природы, существенные для задачи, решаемой субъектом [1]. Модели часто описываются словами или формулами, алгоритмами и иными математическими средствами.

Математические модели. Как правило, при более тщательном анализе явления или процесса словесных моделей недостаточно. Необходимо применение сложных математических моделей. Так, при принятии решений в менеджменте производственных систем используются модели:

- технологических процессов (модели контроля и управления);

Глава 9. Статистические модели динамики

Рассмотрен метод компьютерно-статистического моделирования (далее метод ЖОК*) для оценки результатов влияния описывающих ситуацию факторов на итоговые показатели и друг на друга. Такой метод позволяет получать выводы, полезные для управления различными структурами на микро- и макроуровнях — от бригад и предприятий до государства в целом. В методе используется модель многомерного временного ряда, у которой коэффициенты непосредственного влияния факторов друг на друга и начальные условия задаются экспертами.

9.1. Метод компьютерно-статистического моделирования результатов взаимовлияний факторов

Опишем основные составляющие компьютерно-статистического метода и результаты его практического применения.

1. Определение экспертным путем списка факторов, которые необходимо учитывать при анализе конкретной ситуации. В качестве примера рассмотрим типовое промыш-

* Метод ЖОК получил название по первым буквам фамилий основных разработчиков — В.Н. Жихарева, А.И. Орлова, В.Г. Кольцова. Опыт практического применения этого метода описан в работах [6, 7]. Метод ЖОК развивает идеи когнитивного подхода при решении слабоструктурированных задач, разработанного в Институте проблем управления РАН [8, 9], но на основе иного математического обеспечения.

Глава 10. Статистические модели управления качеством

Одна из наиболее важных областей применения статистических методов — обеспечение качества, основанное на применении статистического моделирования. Статистическим методам управления качеством и посвящена настоящая глава. Приведены общие сведения о месте статистических методов в принятии решений при управлении качеством и сертификации продукции. Рассмотрен статистический контроль качества и продемонстрирована его высокая экономическая эффективность.

10.1. Основы статистического контроля качества

Статистические методы сертификации в России. Методы статистики — именно то средство, которое необходимо изучить, чтобы внедрить управление качеством. Они — наиболее важная составная часть комплексной системы всеобщего управления качеством на фирме. В японских корпорациях все, начиная от председателя совета директоров и до рядового рабочего в цехе, обязаны знать хотя бы основы статистических методов [1].

Сертификация — официальная гарантия поставки производителем продукции, удовлетворяющей установленным требованиям. Поставщики и продавцы должны иметь сертификаты качества на предлагаемые ими товары и услуги. Маркетинг включает в себя работы по сертификации.

Существует несколько уровней сертификации. Говоря о сертификации продукции, могут иметь в виду качество ее

Глава 11. Статистические модели в медицине

Рассмотрена организация клинико-статистических исследований и экспериментов, приведены примеры применения статистических методов в научных медицинских исследованиях.

11.1. Клинико-статистические исследования

Под ***клинико-статистическими исследованиями*** понимают специально организованный сбор и анализ медицинских данных о течении заболеваний у пациентов, о динамике объективных и субъективных показателей их состояния, о реакции на те или иные лечебные воздействия. Исследуются одна или более групп лиц (больных или здоровых), выводы делаются по группам в целом, а не по каждому конкретному пациенту. Цель исследований — перенести выводы, сделанные для выборки, на генеральную совокупность, т. е. клинико-статистическое исследование ориентировано на получение полезных рекомендаций, касающихся тех пациентов, которые попадут в поле зрения врачей после окончания исследования. Таким образом, имеется потенциальное противоречие интересов практикующего врача и научного работника, проводящего клинико-статистическое исследование. Первый заинтересован оказать наилучшую возможную помощь каждому пациенту, а второй разрабатывает рекомендации для будущих больных.

Сбор данных и карта больного. Информация о каждом отдельном пациенте обычно содержится в его истории

Глава 12. Статистические методы в социологии

Социология (от лат. *societas* — общество и греч. *logos* — учение) — наука об обществе как целостной системе и об отдельных социальных институтах, процессах, социальных группах и общностях, отношениях личности и общества, закономерностях массового поведения людей. Среди общественных отношений важное место занимают экономические отношения.

Проанализирована динамика развития статистического инструментария социологов, рассмотрено применение теории люсианов для анализа дихотомических данных, методов анализа сгруппированных данных и принципа уравнивания погрешностей, сделаны полезные выводы для теории управления запасами. Методы социометрии изложены применительно к управлению малыми группами людей.

12.1. Развитие статистического инструментария социологов

Принципиальный прорыв в развитии статистического инструментария произошел в СССР в 1970-е гг. Именно тогда в арсенале отечественных социологов появились теория измерений и теория нечетких множеств, математические методы классификации и многомерное шкалирование, непараметрическая статистика и статистика нечисловых данных. В дальнейшие десятилетия шло естественное развитие научного аппарата. К сожалению, нельзя утверждать, что в последние годы темпы этого развития усилились. Постепенно

Учебное издание

Орлов Александр Иванович

ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Часть 3

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Редактор *А.С. Водчиц*
Технический редактор *Э.А. Кулакова*
Корректор *О.В. Калашникова*
Художник *Н.Г. Столярова*
Компьютерная графика *В.А. Филатовой*
Компьютерная верстка *И.А. Марковой*

Оригинал-макет подготовлен
в Издательстве МГТУ им. Н.Э. Баумана.

Санитарно-эпидемиологическое заключение
№ 77.99.60.953.Д.003961.04.08 от 22.04.2008 г.

Подписано в печать 23.05.12. Формат 84×108 1/32.
Усл. печ. л. 32,76. Тираж 500 экз. Заказ

Издательство МГТУ им. Н.Э. Баумана.
105005, Москва, 2-я Бауманская ул., д. 5, стр. 1.
E-mail: press@bmstu.ru
http://www.baumanpress.ru

Отпечатано в типографии МГТУ им. Н.Э. Баумана.
105005, Москва, 2-я Бауманская, 5, стр. 1.
E-mail: baumanprint@gmail.com

ISBN 978-5-7038-3566-1