



**Орлов**

**Александр Иванович**

доктор экон. наук, доктор техн. наук, канд. физ.-мат. наук,  
профессор, зав. лаб. экономико-математических методов в  
контроллинге МГТУ им. Н.Э. Баумана

УДК 303.4; 519.2

## КОНТРОЛЛИНГ СТАТИСТИЧЕСКИХ МЕТОДОВ

*Автор в течение полувека консультировал научных работников различных специальностей, рецензировал их статьи и книги, оппонировал диссертации. Были выявлены разнообразные недостатки при проведении исследований и публикации их результатов, которые мешают их восприятию, а в ряде случаев ставят под сомнение адекватность выводов. Поэтому автор пришел к выводу о целесообразности выработки, обсуждения и распространения системы основных требований к разработке и применению статистических методов анализа данных, к их описанию в публикациях, диссертациях и т.п. Настоящая статья посвящена первоначальному рассмотрению ряда формулировок таких требований. Автор исходит из современной парадигмы прикладной статистики, сменившей примитивную парадигму XIX в. и устаревшую парадигму середины XX в.*

**Ключевые слова:** контроллинг, статистические методы, анализ данных, вероятностно-статистическая модель, прикладная статистика, парадигма, непараметрика, статистика нечисловых данных, теория измерений, регрессионный анализ, нейросетевые методы.

---

**Orlov Alexander**, Doctor of Economics, Doctor of Technical Sciences, PhD in Physical and Mathematical Sciences, Professor, Head of the laboratory of economic and mathematical methods in controlling, BMSTU

## CONTROLLING STATISTICAL METHODS

*The author has been advising researchers of various specialties for half a century, reviewing their articles and books, and opposing dissertations. Various shortcomings have been identified when conducting research and publishing their results, which interfere with their perception, and in some cases cast doubt on the adequacy of the conclusions. Therefore, the author concluded that it is advisable to develop, discuss and disseminate a system of basic requirements for the development and application of statistical methods of data analysis, for their description in publications, dissertations, etc. This article is devoted to the initial consideration of a number of such requirements formulations. The author proceeds from the modern paradigm of applied statistics, which replaced the primitive paradigm of the XIX century and the outdated paradigm of the mid-XX century.*

**Keywords:** controlling, statistical methods, data analysis, probabilistic-statistical model, applied statistics, paradigm, nonparametrics, statistics of non-numeric data, measurement theory, regression analysis, neural network methods.

---

### Введение

В современном контроллинге много различных направлений. Начиная со статьи [3] автор развивает контроллинг организационно-экономических методов. Контроллинг в этой области – это разработка процедур управления соответствием, используемых и вновь создаваемых (внедряемых) организационно-экономических методов, поставленным задачам. Статистические методы и, прежде всего, прикладная статистика – важная часть организационно-экономических методов. В настоящей статье, развивая положения доклада [11], рассматриваются проблемы контроллинга статистических методов.

Представляется полезным выработать, обсудить и распространить основные требования к разработке и применению статистических методов анализа данных, к их описанию в публикациях, диссертациях и т.п. Почему такая работа понадобилась? Казалось бы, имеется много книг, статей, и Интернет-ресурсов, им и надо следовать. Однако зачастую не удается извлечь из литературных источников конкретные рекомендации по проведению собственных работ и подготовке их к публикации. К тому же приходится констатировать, что в публикациях зачастую имеются ошибки, десятилетиями кочующие из одной публикации в другую. Одна из таких ошибок проанализирована в статье [4].

Постоянное консультирование в течение полувека научных работников различных специальностей, рецензирование их статей и книг, оппонирование диссертаций дало возможность автору познакомиться с сотнями конкретных исследований по разработке и применению статистических методов. Критический анализ накопленного материала позволил разработать общий подход к проведению таких исследований и ряд частных методов, отраженных в наших учебниках. Кроме того, выявлены разнообразные недостатки при проведении исследований и публикации их результатов, которые мешают адекватному восприятию, а в ряде случаев ставят под сомнение обоснованность выводов. Этим обосновано убеждение автора в том, что целесообразно сформулировать и обсудить естественные требования к методам обработки данных и представлению результатов статистического анализа конкретных данных.

С целью «стандартизации математических орудий» (в терминологии Н. Бурбаки [1, с.253])

представляется целесообразным развернуть работу по сертификации статистических методов и соответствующих пакетов программ, а также учебных курсов и материалов, правил подготовки к публикации результатов теоретических и практических исследований.

Очевидно, стандартизация полезна только тогда, когда она проводится квалифицированными специалистами, в противном случае вместо пользы имеем вред. Примером является печальная судьба многообразия стандартов по статистическим методам управления качеством, большую часть которых пришлось отменить из-за ошибок разработчиков. Эта ситуация со стандартизацией подробно проанализирована в [10]. Очевидно, что проект нормативного документа должен подвергаться тщательному обсуждению на основе анализа высококвалифицированными специалистами. Однако такие специалисты предпочитают заниматься собственными исследованиями.

Настоящая статья посвящена требованиям к методам обработки данных и представлению результатов статистического анализа конкретных данных. Автор исходит из современной парадигмы прикладной статистики, о которой необходимо сказать несколько слов.

### О смене парадигм прикладной статистики

Статистические методы анализа данных широко применяются исследователями в различных областях науки. Центром этого инструментария является прикладная статистика, т.е. наука о том, как обрабатывать данные. Применение методов прикладной статистики в той или иной конкретной области деятельности порождает соответствующие науки. Например, применение в экономике и управлении (эконометрика), в биологии (биометрия), в технических исследованиях (технометрика), в химии (хеометрика), в медицине (доказательная медицина), в науковедении и управлении наукой (наукометрия) и т.д.

Обсудим смену парадигм прикладной статистики. Под парадигмой понимаем принятую наиболее квалифицированным ядром исследователей модель адекватной деятельности в той или иной области науки. Обсудим изменение с течением времени основ общепринятой специалистами модели действий в области прикладной статистики и методов анализа данных, более широко – в области математических методов исследования.

Рассмотрим три реально используемых в настоящее время парадигмы – примитивную, устаревшую, современную. Примитивная соответствует взглядам XIX и начала XX в., устаревшая – середине XX в., современная – XXI в. Поясним на примере действий современных исследователей, придерживающихся той или иной парадигмы.

Исходя из примитивной парадигмы, наивные (плохо знакомые с современной прикладной статистикой) авторы применяют широко известные расчетные формулы классического критерия Стьюдента для проверки статистической гипотезы о равенстве 0 математического ожидания без какого-либо обоснования и верят в то, что действуют правильно.

Согласно устаревшей парадигме, в начале исследования принимают (обычно без какого-либо обоснования, тем более строгого), что результаты измерений имеют нормальное распределение, затем применяют классический критерий Стьюдента (в предположениях нормальности распределения результатов измерений (наблюдений, испытаний, анализов, опытов) – это обосновано).

Согласно современной парадигме для проверки рассматриваемой гипотезы следует использовать непараметрические методы (основанные на центральной предельной теореме), поскольку хорошо известно, что распределения реальных данных, как правило, не являются нормальными. Бесспорно, что обоснованность статистических выводов возрастает при переходе от примитивной парадигмы к устаревшей и далее к современной. Несмотря на прогресс в развитии прикладной статистики, в настоящее время в практике научной работы в различных областях используются все три парадигмы. Обсудим, как это влияет на качество результатов исследовательской деятельности, как следствие, на качество научных публикаций.

Констатируем, что примитивная парадигма – это парадигма поваренной книги. Придерживающиеся этой парадигмы лица без осмысления следуют составленным кем-то рецептам. Применение распространенных программных продуктов без осмысления применяемых методов может провоцировать такие расчеты. Однако довольно часто итоговые выводы оказываются полезными с позиций прикладной области. Но иногда они могут быть и грубо ошибочными. Об опасности бездумного примене-

ния программных продуктов предупреждал еще проф. В.В. Налимов [2], выдающийся исследователь в области статистических методов.

Устаревшая парадигма – это парадигма середины XX в. В ней застыли взгляды начала XX в., когда были получены первые результаты новой отрасли науки – математической статистики. Согласно устаревшей парадигме элементы выборки рассматриваются как независимые случайные величины, распределения которых входят в то или иное параметрическое семейство распределений: нормальных, логистических, экспоненциальных, Вейбулла-Гнеденко, Коши, Лапласа, гамма-распределений, бета-распределений и др. Все эти семейства входят в четырехпараметрическое семейство распределений, введенное основателем математической статистики К. Пирсоном в начале XX в. С целью упорядочения результатов измерений (наблюдений, анализов, испытаний, опытов, обследований) он и его последователи приняли рабочую гипотезу, что распределения реальных данных всегда совпадают с какими-то элементами его четырехпараметрического семейства. Затем началось развитие теории параметрической математической статистики, в которой задачи оценивания и проверки гипотез решались для выборок из тех или иных параметрических семейств. Был получен ряд замечательных математических моделей и результатов, например, связанных с методом максимального правдоподобия, критериями Стьюдента, Пирсона (хи-квадрат), Фишера, неравенством Рао-Крамера и др. Многомерное нормальное распределение оказалось весьма полезным для развития регрессионного и дискриминантного анализов. Видимо, потому, что плотность такого распределения в точке  $Y$  является квадратичной формой от координат  $Y$ , а алгоритмы регрессионного и дискриминантного анализов соответствуют преобразованиям этой квадратичной формы при линейной замене координат.

Параметрической математической статистике посвящено основное содержание распространенных и в настоящее время вузовских учебников по математической статистике. В отличие от примитивной парадигмы, имеется строгая математическая теория, позволяющая на основе гипотезы, что распределения элементов выборки входят в то или иное параметрическое семейство, получать расчетные алгоритмы и на их основе – полезные практические рекомендации. Однако

у этой математико-статистической теории есть принципиальный недостаток – распределения реальных данных, как правило, не являются нормальными и вообще не входят в четырехпараметрическое семейство Пирсона. Это утверждение строго обосновано [12].

В прикладных работах иногда пытаются проверить нормальность или, например, экспоненциальность реальных данных. Зачастую отклонить гипотезу нормальности не удастся. Но это нельзя рассматривать как окончательное подтверждение гипотезы о нормальности распределения рассматриваемых данных, поскольку для тех же данных обычно не удастся отклонить и гипотезу о том, что распределение данных соответствует другому популярному распределению. Причина этого внешне парадоксального явления очевидна – недостаточный (малый) объем выборки. Например, известно, что для того, чтобы выяснить, какому распределению соответствуют анализируемые данные – нормальному или логистическому, необходимо не менее 2500 наблюдений [12]. Реальные объемы выборок в социально-экономических исследованиях, как правило, значительно меньше. Констатируем, что на основе параметрической математической статистики осуществлены попытки решения многих прикладных задач в конкретных областях исследования. Но в ряде случаев получены ошибочные выводы, хотя доля таких случаев заметно меньше, чем при опоре на примитивную парадигму.

На смену параметрической статистике пришла непараметрическая, основанная на принципиально иной модели порождения данных. В ней, в отличие от параметрической, элементы выборки с числовыми значениями предполагаются имеющими произвольную функцию распределения (во многих случаях добавляют еще условие непрерывности).

Развитие непараметрической статистики к настоящему времени достигло такого уровня, что ее методами можно решать столь же обширный круг задач анализа данных, что и параметрическими. Преимущество непараметрической статистики по сравнению с параметрической состоит в том, что нет необходимости принимать необоснованные предположения о виде функции распределения.

У непараметрической статистики есть и недостатки. Один из них порожден тем, что реальные статистические данные довольно часто содержат

совпадения. Дело в том, что если функция распределения элементов выборки непрерывна, как это принято в непараметрической статистике, то вероятность совпадения двух или более элементов выборки равна 0. Одной из причин возникновения обсуждаемого противоречия является то, что свойства прагматических чисел, используемых для записи результатов измерений (наблюдений, испытаний, опытов, анализов, обследований), отличаются от свойств математических чисел (например, прагматические числа записываются с помощью конечного числа цифр, а почти все действительные числа требуют в теории бесконечного ряда цифр). Предложенные автором подходы к анализу совпадений при применении непараметрических статистических методов позволяют частично снять рассматриваемое противоречие [6].

#### Современная парадигма прикладной статистики и системная нечеткая интервальная математика

Современная парадигма прикладной статистики основана на применении методов непараметрической и нечисловой статистики [5]. В настоящее время теоретические исследования по прикладной статистике проводятся в основном в соответствии с современной парадигмой [7].

Наши работы, посвященные современной парадигме прикладной статистики, послужили основой для создания нового перспективного направления теоретической и вычислительной математики – системной нечеткой интервальной математики. Ее основная идея – это переход от классических действительных чисел как основы математики к прагматическим числам с конечным количеством градаций, к нечетким и интервальным числам. Ключевой публикацией является монография 2014 г., вызвавшая значительный интерес у научной общественности [14]. Ее продолжением является монография [15], подготовленная на основе работ авторов 2014–2021 гг.

Приходится констатировать, что в настоящее время значительная доля прикладных работ осуществляется в традициях устаревшей или даже примитивной парадигм. Такие работы нецелесообразно огульно отрицать. Они могут приносить пользу в конкретных областях. Однако бесспорно, что переход на современную парадигму прикладной статистики повысит научный уровень исследований, а также позволит полу-

чить важные результаты в конкретных областях. К сожалению, многие исследователи, связанные с анализом данных, в том числе разработчики программных продуктов по этой тематике, недостаточно знакомы с непараметрической и нечисловой статистикой. Необходимо шире распространять информацию о современной парадигме прикладной статистики.

Опора на подходы и результаты непараметрической и нечисловой статистики — одно из основных требований к статистическим методам анализа данных. Раскроем это утверждение.

### Роль вероятностно-статистических моделей данных

Первый этап при разработке и применении методов прикладной статистики — это выбор и обоснование вероятностно-статистических моделей данных. При описании, применении и обсуждении тех или иных процедур анализа статистических данных обычно сосредотачивают внимание на расчетных формулах. Однако алгоритмы расчетов основаны на вероятностно-статистических моделях порождения изучаемых данных. С этих моделей и надо начинать, и при проведении исследования, и при его описании.

Например, в работах по прикладной статистике наивные авторы под выборкой обычно понимают конечную последовательность чисел. Квалифицированные исследователи в большинстве случаев используют наиболее распространенную модель выборки, согласно которой результаты измерений рассматриваются как конечная последовательность реализаций независимых одинаково распределенных случайных величин, моделирующих результаты измерений (наблюдений, испытаний, опытов, анализов, обследований) [12].

Если общая функция распределения этих случайных величин является произвольной, то необходимо обратиться к методам непараметрической статистики. Для реальных данных совпадения результатов встречаются достаточно часто. Следовательно, в таких случаях наблюдаются отклонения от непараметрической модели с непрерывными функциями распределения. Как уже отмечалось выше, модель анализа совпадений при расчете значений непараметрических ранговых статистик предложена в работе [6]. Статистика интервальных данных как составная часть нечисловой статистики была создана для обработки округленных данных и данных с совпадениями [12].

Отметим устойчивость предубеждений. Например, в до сих пор продолжающейся развиваться параметрической статистике пропагандируется использование метода максимального правдоподобия, хотя одношаговые оценки имеют столь же хорошие свойства, что и оценки максимального правдоподобия. В ряде случаев система уравнений максимального правдоподобия не имеет явного решения в виде конечных расчетных формул, и соответствующие оценки рекомендуется находить теми и или иными итерационными методами. Их сходимость, как правило, не изучают, хотя есть примеры, в которых отсутствие сходимости продемонстрировано. Между тем одношаговые оценки вычисляются по конечным формулам, без всяких итераций [12].

Заметна склонность теоретиков в области математической и прикладной статистики к использованию многомерных нормальных распределений. Именно для таких распределений найдены явные формулы для различных характеристик в многомерном статистическом анализе, прежде всего в регрессионном. По авторской экспертной оценке, причина в том, что таким теоретикам удается использовать хорошо развитую в линейной алгебре теорию квадратичных форм.

Давно установлено, что распределения почти всех реальных данных не являются нормальными (гауссовскими) [12]. Выдвигают теоретические аргументы в обоснование использования нормального распределения. Так, утверждают, что зависимость значения случайной величины от многих факторов влечет нормальность. Иногда увеличивают обоснованность такого суждения, добавляя, что факторы являются независимыми и сравнимыми по величине случайными величинами. Однако близость к нормальному распределению можно ожидать лишь при справедливости аддитивной модели порождения данных, когда факторы складываются (это утверждение вытекает из Центральной предельной теоремы теории вероятностей). Если же случайная величина формируется путем перемножения (мультипликативная модель порождения данных), то ее распределение является (в асимптотике) логарифмически нормальным. Если же справедлива модель «самого слабого» звена (или «самого сильного» — рекорда), т.е. значение случайной величины равно крайнему члену вариационного ряда значений факторов (соответственно, мини-

муму или максимуму), то имеем в пределе распределение Вейбулла-Гнеденко.

Использование модели на основе семейства нормальных распределений можно сравнить с поиском под ярким фонарем потерянных в темных кустах ключей. Очевидно, под фонарем искать легче. Можно продемонстрировать активность. Однако нецелесообразно надеяться на благоприятный исход поисков ключей.

Из сказанного вытекает следующее требование к статистическим методам обработки данных: если по каким-либо причинам исследователь желает применить параметрическое семейство распределений, его использование должно быть тщательно обосновано путем проверки статистической гипотезы согласия как с рассматриваемым семейством, так и с альтернативными семействами.

### Роль вероятностно-статистических моделей в многомерном статистическом анализе

Начнем с регрессионного анализа. Используют несколько основных типов регрессионных моделей. Обсудим простейшую постановку: одна независимая переменная и одна зависимая. Кратко охарактеризуем основные используемые модели.

Часто применяют модели метода наименьших квадратов с детерминированной независимой переменной и параметрической зависимостью (линейной, квадратической и т.п.). Естественно принять, что распределение отклонений произвольно (т.е. рассматривают непараметрическую модель). Вывод предельных распределений оценок параметров и регрессионной зависимости основан на Центральной предельной теореме и теореме о линеаризации [12].

Принципиально иной тип моделей основан на выборке случайных векторов. В большинстве случаев зависимость является параметрической, ее параметры оценивают по выборочным данным. Естественно принять, что распределение двумерного вектора произвольно. Об оценке дисперсии независимой переменной (в отличие от зависимой) можно говорить только в модели на основе выборки случайных векторов, равно как и о коэффициенте детерминации как критерии качества модели, в противном случае возможны принципиальные ошибки [8].

Другой тип моделей регрессионного анализа, основанный на выборке случайных векторов —

это непараметрическая регрессия, в которой как зависимость, так и отклонения от нее являются непараметрическими. Зависимость (как условное среднее) оценивается с помощью непараметрических оценок плотности распределения случайного вектора.

Еще один вариант — модель, в которой тренд линейен, а периодическая и случайная составляющие и отклонения от них являются непараметрическими. Он является промежуточным между двумя только что рассмотренными.

В моделях следующего типа малые погрешности имеются как в значениях зависимой переменной, так и в значениях независимой переменной. Значения переменных естественно описывать интервалами. В прошлом столетии этот раздел прикладной статистики, посвященный моделям указанного типа, назывался конъюнктным анализом, сейчас он входит в статистику интервальных данных [12].

Возможно дальнейшее развитие приведенной выше классификации моделей регрессионного анализа. Так, обычно принимают, что погрешности (ошибки, невязки) — это независимые одинаково распределенные случайные величины. Можно отказаться как от требования одинаковой распределенности, так и от требования независимости. Если средняя квадратическая погрешность пропорциональна измеряемой величине, то приходим к необходимости минимизации не суммы квадратов разностей значений зависимой переменной и функции от независимой переменной, а другого критерия оптимизации. А именно, в квадрат возводятся частные от деления указанных величин на значения функции от значений независимой переменной. Другими словами, в методе наименьших квадратов надо заменить абсолютные отклонения относительными.

Вместо суммы квадратов отклонений можно использовать другие постановки задачи оптимизации, например, минимизировать сумму модулей отклонений (метод наименьших модулей) или максимальное (по модулю) отклонение (метод минимакса).

К регрессионному анализу примыкают задачи сглаживания временных рядов и статистики случайных процессов, в которых отклонения от функции времени зависимы (в отличие от регрессионного анализа, в котором такие отклонения — независимые случайные величины). Другими словами, при моделировании

временных рядов вполне естественно отказаться от требования независимости погрешностей. Более того, поскольку зависимость между значениями случайно функции от времени, как правило, убывает при увеличении расстояния между моментами измерения, то независимость погрешностей можно постулировать лишь тогда, когда моменты измерений значительно отличаются друг от друга.

Можно описывать погрешности не случайными величинами, а нечеткими числами, частным случаем которых являются интервалы, о которых уже шла речь выше. Автор не пытается описать все различные постановки регрессионного анализа. Для этого нужны монографии. Однако проведенный выше краткий анализ многообразия моделей регрессионного анализа приводит к выводу, что не существует какой-либо единой «стандартной модели» [9]. Следовательно, при решении и описании задачи восстановления зависимости необходимо начинать с выбора и обоснования той или иной вероятностно-статистической модели порождения данных.

### Теория измерений как основа построения вероятностно-статистических моделей

Согласно современным воззрениям, при проведении статистического анализа данных необходимо исходить из теории измерений [12]. Согласно этой теории, первый шаг при анализе данных — это выявление шкал, в которых они измерены. Основное требование: применяемые статистические методы должны соответствовать шкалам, в которых измерены данные.

Приведем пример. Статистические выводы, основанные на расчете средних величин, должны быть инвариантны относительно допустимых преобразований шкал измерения статистических данных. Доказано, что для данных, измеренных в порядковой шкале, в качестве средних величин можно использовать только конечное число функций от результатов измерений, а именно члены вариационного ряда. При нечетном объеме выборки — медиану, а при четном — левую медиану или правую медиану. Применение, например, среднего арифметического или среднего геометрического недопустимо. Как следствие, поскольку широко используемые в прикладных исследованиях ранги или баллы, как правило, измерены в порядковой шкале, рассчитывать для них среднее арифмети-

ческое нельзя. В частности, согласно современной прикладной статистике нельзя оценивать успеваемость учащихся по среднему баллу экзаменационных оценок.

Основное требование: статистические выводы, основанные на расчете тех или иных статистик (функций от результатов измерений), должны быть инвариантны относительно допустимых преобразований шкал измерения данных. Следовательно, перед исследователями в области теории прикладной статистики возникает перво-степенная задача: для каждой используемой ими шкалы выяснить, какими алгоритмами анализа данных из рассматриваемого ими семейства алгоритмов можно пользоваться в этой шкале. Выше кратко описаны выводы относительно использования семейства средних по Коши.

Важна и обратная задача — для определенного алгоритма анализа данных выяснить, в какой шкале можно им пользоваться. Установлено, что коэффициент линейной парной корреляции Пирсона соответствует шкале интервалов, в то время как непараметрические ранговые коэффициенты корреляции Спирмена и Кендалла нацелены на изучение взаимосвязи порядковых переменных.

На основе теории измерений кратко обсудим довольно широко известный метод анализа иерархий. Исходные данные в этом методе — это результаты парных сравнений, они измерены в порядковых шкалах. А результаты расчетов по методу анализа иерархий выражены в шкале интервалов, как утверждают энтузиасты этого метода. С точки зрения теории измерений такое недопустимо. Результаты расчетов (статистические выводы) должны быть измерены в той же шкале, что и исходные данные. Следовательно, с точки зрения теории измерений методом анализа иерархий пользоваться не следует. Автор рекомендует применять адекватные метода анализа экспертных оценок, в частности, методы средних арифметических рангов, медиан рангов, согласования кластеризованных ранжировок [13].

### Обучающие выборки в задачах диагностики и нейросети

Естественно распространить разрабатываемые требования на смежную (близкородственную) область — нейросетевую обработку данных. Учитывая значительное взаимопроникновение вероятностно-статистических и нейросетевых методов, это представляется весьма полезным.

Рассмотрим в качестве базового примера соотношение прикладной статистики и нейросетевой обработки данных в области математической теории классификации. В этой теории выделяют три раздела:

- построение классификаций;
- изучение классификаций;
- применение классификаций [12].

Если изучение классификаций обычно рассматривают как часть статистики нечисловых данных, то две другие области имеют в литературе самые разные названия. Синонимы понятия «построение классификаций», по мнению автора, таковы: кластер-анализ, распознавание образов без учителя, типология, таксономия, группировка, классификация без учителя, дихотомия. Аналогично синонимами термина «применение классификаций» являются: методы дискриминации (дискриминантный анализ), математические методы диагностики, распознавание образов с учителем, автоматическая классификация с учителем, статистическая классификация.

Здесь под «учителем» понимают способы построения правил принятия решений на основе обучающих выборок. Предполагается, что для каждого из классов имеется обучающая выборка, т.е. выборка элементов из этого класса. На основе обучающих выборок строится правило принятия решений о том, к какому классу отнести вновь поступающий объект.

Когда говорят об алгоритмах «без учителя», то это значит, что речь идет о построении классификации на основе анализа данных единой обучающей выборки, для элементов которой не указано, к какому классу этот элемент относится. Алгоритмы «без учителя» основаны на тех или иных мерах близости между элементами (показателях различия).

В настоящее время «нейросети» – весьма популярный термин. Речь идет о различных математических моделях (а также разработанных на их основе алгоритмах, их программной или аппаратной реализации), построенных по аналогии с сетями нервных клеток живого организма. Первые такие модели были разработаны в середине XX в. при изучении процессов, протекающих в мозге человека. Была сделана попытка смоделировать эти процессы (на уровне знаний того времени). В настоящее время известно, что человеческий мозг работает иначе, чем предполагают энтузиасты нейросетей.

При внимательном анализе основных идей нейросетевых методов становится очевидным,

что эти модели предназначены прежде всего для решения задач классификации на основе анализа обучающих выборок. Другими словами, классические задачи теории классификации решаются с помощью нейросетей не так, как ранее в прикладной статистике.

Теория математической статистики позволяет сравнивать алгоритмы классификации по качеству. Для задач диагностики целесообразно проводить сравнение на основе прогностической силы алгоритма [12]. Оказывается, что нейросетевые алгоритмы, как правило, не являются оптимальными. Например, в теории классификации доказано, что для отнесения вновь поступающего объекта в один из двух классов, заданных обучающими выборками, (асимптотически) оптимальным является решающее правило, основанное на непараметрических оценках плотностей распределений вероятностей, соответствующих классам [12]. Нейросетевые методы не могут дать лучшего результата, чем это решающее правило. К сожалению, частое упоминание нейросетей в современной литературе приводит к забвению оптимальных методов и алгоритмов (это естественно в силу ограниченных возможностей человеческого мозга по восприятию и осмыслению информации), что, естественно, снижает эффективность технологических решений искусственного интеллекта.

Таким образом, приходим к выводу, что нейросети, методы распознавания образов, и, например, генетические алгоритмы – это другие названия ряда давно разрабатываемых разделов прикладной статистики (статистических методов анализа данных) [12]. Усилиями журналистов и публицистов, не очень разбирающихся в идеях и научных результатах прикладной статистики, новая терминология оказалась в центре внимания научной общественности.

### Выводы

В настоящей статье автором обоснована необходимость разработки системы требований к статистическим моделям и методам при их создании, применении и преподавании, в том числе при их описании в научных и методических публикациях.

Подчеркивается, что прежде всего должна быть представлена и обоснована вероятностно-статистическая модель порождения данных. Полезный анализ иерархической структуры

понятия «модель» и потенциальных источников ошибок при построении, изучении, применении и преподавании вероятностно-статистических моделей реальных данных представлен в статье [16]. Автором приведены краткие формулировки ряда требований к статистическим методам, проанализированных выше.

Поскольку практически все распределения реальных данных ненормальны, предпочтения следует отдавать непараметрическим постановкам. Возможность применения параметрических семейств распределений должна быть тщательно

обоснована. В соответствии с теорией проверки статистических гипотез должны быть указаны не только нулевая гипотеза, но и альтернативная, только тогда можно обсуждать мощность критерия.

Необходимо изучение устойчивости выводов, получаемых на основе организационно-экономической модели, относительно допустимых изменений исходных данных и предпосылок модели [12]. В частности, статистические выводы должны быть инвариантны относительно допустимых преобразований шкал.

### Литература:

1. Бурбаки Н. Очерки по истории математики. – М.: Изд-во Иностранной литературы. 1963. – 292 с.
2. Налимов В.В. Теория эксперимента. – М.: Наука, 1971. – 208 с.
3. Орлов А.И. Контроллинг организационно-экономических методов // Контроллинг. 2008. № 4 (28). С. 12-18.
4. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, Омега-квадрат и ошибки при их применении // Научный журнал КубГАУ. 2014. № 97. С. 647-675.
5. Орлов А.И. О новой парадигме математических методов исследования // Научный журнал КубГАУ. 2016. № 122. С. 807-832.
6. Орлов А.И. Модель анализа совпадений при расчете непараметрических ранговых статистик // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №11. С. 66-72.
7. Орлов А.И. Развитие математических методов исследования (2006-2015 гг.) // Заводская лаборатория. Диагностика материалов. 2017. Т. 83. №1. Ч. 1. С. 78-86.
8. Орлов А.И. Ошибки при использовании коэффициентов корреляции и детерминации // Заводская лаборатория. Диагностика материалов. 2018. Т. 84. № 3. С. 68-72.
9. Орлов А.И. Многообразие моделей регрессионного анализа (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2018. Т. 84. № 5. С. 63-73.
10. Орлов А.И. Эконометрика: учебное пособие. – М.: Саратов: Интернет-Университет Информационных Технологий (ИНТУ-ИТ), Ай Пи Ар Медиа, 2020. – 676 с.
11. Орлов А.И. Контроллинг и статистические методы // Контроллинг в экономике, организации производства и управлении: сборник научных трудов X Международного конгресса по контроллингу (Ярославль, 22 октября 2021 г.) / Под научной редакцией д.э.н., профессора С.Г. Фалько / НП «Объединение контроллеров». – М.: НП «Объединение контроллеров», 2021. С. 65-74.
12. Орлов А.И. Прикладной статистический анализ. – М.: Ай Пи Ар Медиа, 2022. – 812 с.
13. Орлов А.И. Искусственный интеллект: экспертные оценки. – М.: Ай Пи Ар Медиа, 2022. – 436 с.
14. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.
15. Орлов А.И., Луценко Е.В. Анализ данных, информации и знаний в системной нечеткой интервальной математике: научная монография. – Краснодар: КубГАУ, 2022. – 405 с.
16. Савельев О.Ю. Модель: иерархия понятия и потенциальный источник ошибок // Инновации в менеджменте. 2021. № 28. С. 54-58.

### References:

1. Burbaki N. Oчерki po istorii matematiki. – M.: Izd-vo Inostrannoj literatury. 1963. – 292 s.
2. Nalimov V.V. Teorija jeksperimenta. – M.: Nauka, 1971. – 208 s.
3. Orlov A.I. Kontrolling organizacionno-jekonomicheskikh metodov // Kontrolling. 2008. № 4 (28). S. 12-18.
4. Orlov A.I. Neparametricheskie kriterii soglasija Kolmogorova, Smirnova, Omega-kvadrat i oshibki pri ih primenenii // Nauchnyj zhurnal KubGAU. 2014. № 97. S. 647-675.
5. Orlov A.I. O novoj paradigme matematicheskikh metodov issledovaniya // Nauchnyj zhurnal KubGAU. 2016. № 122. S. 807-832.
6. Orlov A.I. Model' analiza sovpadenij pri raschete neparametricheskikh rangovykh statistik // Zavodskaja laboratorija. Diagnostika materialov. 2017. T.83. №11. S. 66-72.
7. Orlov A.I. Razvitie matematicheskikh metodov issledovaniya (2006-2015 gg.) // Zavodskaja laboratorija. Diagnostika materialov. 2017. T. 83. №1. Ch. 1. S. 78-86.
8. Orlov A.I. Oshibki pri ispol'zovanii koeficientov korrelyacii i determinacii // Zavodskaja laboratorija. Diagnostika materialov. 2018. T. 84. № 3. S. 68-72.

9. Orlov A.I. Mnogoobrazie modelej regressionnogo analiza (obobshhajushhaja stat'ja) / Zavodskaja laboratorija. Diagnostika materialov. 2018. T. 84. № 5. S. 63-73.
  10. Orlov A.I. Jekonometrika: uchebnoe posobie. – M.: Saratov: Internet-Universitet Informacionnyh Tehnologij (INTUIT), Aj Pi Ar Media, 2020. – 676 c.
  11. Orlov A.I. Kontrolling i statisticheskie metody // Kontrolling v jekonomike, organizacii proizvodstva i upravlenii: sbornik nauchnyh trudov X Mezhdunarodnogo kongressa po kontrollingu (Jaroslavl', 22 oktjabrja 2021 g.) / Pod nauchnoj redakciej d.je.n., professora S.G. Fal'ko / NP «Ob#edinenie kontrollerov». – M.: NP «Ob#edinenie kontrollerov», 2021. S. 65-74.
  12. Orlov A.I. Prikladnoj statisticheskij analiz. – M.: Aj Pi Ar Media, 2022. – 812 c.
  13. Orlov A.I. Iskusstvennyj intellekt: jekspertnye ocenki. – M.: Aj Pi Ar Media, 2022. – 436 c.
  14. Orlov A.I., Lucenko E.V. Sistemnaja nechetkaja interval'naja matematika. Monografija (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s.
  15. Orlov A.I., Lucenko E.V. Analiz dannyh, informacii i znaniy v sistemnoj nechetkoj interval'noj matematike: nauchnaja monografija. – Krasnodar: KubGAU, 2022. – 405 s.
  16. Savel'ev O.Ju. Model': ierarhija ponjatija i potencial'nyj istochnik oshibok // Innovacii v menedzhmente. 2021. № 28. S. 54-58.
-