Реферат

**Современные эконометрические методы**

**О развитии эконометрических методов**

Современное состояние в эконометрике, как и в других областях, определяется прошлым. Кратко рассмотрим историю эконометрики и прикладной статистики, начав с их практической пользы.

**Что дает прикладная статистика народному хозяйству?** Так называлась статья [1], в которой приводились многочисленные примеры успешного использования методов эконометрики и прикладной математической статистики при решении практических задач. Обширный перечень примеров приведен в предыдущих главах настоящей книги. Его можно продолжать практически безгранично. Так, в любом номере журнала "Заводская лаборатория" есть работы, в которых те или иные методы эконометрики и прикладной статистики применяются для решения прикладных технико-экономических задач.

Поэтому бесспорно совершенно, что методы эконометрики и прикладной статистики успешно применяются в различных отраслях народного хозяйства, практически во всех областях науки. Согласно докладу [2], в 1988 г. затраты в СССР на статистический анализ данных оценивались в 2 миллиарда рублей ежегодно.

Большая практическая значимость эконометрики и прикладной статистики, особенно в экономике, менеджменте, технических исследованиях и разработках, оправдывает целесообразность развития их методологии, в которых эти области научной и прикладной деятельности рассматривалась бы как целое, "с высоты птичьего полета". Чтобы иметь возможность обсуждения тенденций развития эконометрики и статистических методов в XXI веке, необходимо хотя бы кратко рассмотреть их историю.

**Об истории эконометрики и прикладной статистики.** Типовые примеры раннего этапа применения статистических методов описаны в Ветхом Завете (см., например, Книгу Чисел). С математической точки зрения они сводились к подсчетам числа попаданий значений наблюдаемых признаков в определенные градации. В дальнейшем результаты стали представлять в виде таблиц и диаграмм, как это и сейчас делают Госкомстат РФ (Российское статистическое агентство). Надо признать, что по сравнению с Ветхим Заветом есть прогресс - в Библии не было таблиц. Однако нет продвижения по сравнению с работами российских статистиков конца девятнадцатого - начала двадцатого века (типовой монографией тех времен можно считать книгу [3], которая в настоящее время ещё легко доступна).

Сразу после возникновения теории вероятностей (Паскаль, Ферма, 17 век) вероятностные модели стали использоваться при обработке статистических данных. Например, изучалась частота рождения мальчиков и девочек, было установлено отличие вероятности рождения мальчика от 0.5, анализировались причины того, что в парижских приютах эта вероятность не та, что в самом Париже, и т.д. Имеется достаточно много публикаций по истории теории вероятностей, однако в некоторых из них имеются неточные утверждения, что заставило одного из крупнейших ученых ХХ в. академика Украинской АН Б.В. Гнеденко включить в очередное издание своего курса [4] главу по истории математики случайного.

В 1794 г. (по другим данным - в 1795 г.) К. Гаусс разработал метод наименьших квадратов, один из наиболее популярных ныне статистических методов (см. главу 5 выше), и применил его при расчете орбиты астероида Церера - для борьбы с ошибками астрономических наблюдений. В Х1Х веке заметный вклад в развитие практической статистики внес бельгиец А. Кетле, на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей, таких, как доля самоубийств среди всех смертей. Интересно, что основные идеи статистического приемочного контроля и сертификации продукции обсуждались академиком М.В. Остроградским и применялись в российской армии ещё в середине Х1Х в.. Статистические методы управления качеством, сертификации и классификации продукции и сейчас весьма актуальны (см. главу 13 выше).

Современный этап развития прикладной статистики можно отсчитывать с 1900 г., когда англичанин К. Пирсон основан журнал "Biometrika". Первая треть ХХ в. прошла под знаком параметрической статистики. Изучались методы, основанные на анализе данных из параметрических семейств распределений, описываемых кривыми из т.н. семейства Пирсона. Наиболее популярным было нормальное (гауссово) распределение. Для проверки гипотез использовались критерии Пирсона, Стьюдента, Фишера. Были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

Разработанную в первой трети ХХ в. теорию называем параметрической статистикой, поскольку ее основной объект изучения - это выборки из распределений, описываемых одним или небольшим числом параметров. Наиболее общим является семейство кривых Пирсона, задаваемых четырьмя параметрами. Как правило, нельзя указать каких-либо веских причин, по которым конкретное распределение результатов наблюдений должно входить в то или иное параметрическое семейство (подробнее см. начало главы 4). Исключения хорошо известны: если вероятностная модель предусматривает суммирование независимых случайных величин, то сумму естественно описывать нормальным распределением; если же в модели рассматривается произведение таких величин, то итог, видимо, приближается логарифмически нормальным распределением, и т.д. Однако в подавляющем большинстве реальных ситуаций подобных моделей нет, и приближение реального распределения с помощью кривых из семейства Пирсона или его подсемейств - чисто формальная операция.

Именно из таких соображений критиковал параметрическую статистику академик АН СССР С.Н. Бернштейн в 1927 г. в своем докладе на Всероссийском съезде математиков [5]. Однако эта теория, к сожалению, до сих пор остается основой преподавания статистических методов и продолжает использоваться основной массой прикладников, остающихся далекими от новых веяний в статистике. Почему так происходит? Чтобы попытаться ответить на этот вопрос, обратимся к одной из статистических наук - наукометрии, в которой статистическими методами анализируется развитие научных исследований.

**Наукометрия прикладной статистики.** Проведенный несколько лет назад анализ прикладной статистики как области научно-практической деятельности (в рамках движения за создание Всесоюзной статистической ассоциации, учрежденной в 1990 г.) показал, в частности, что актуальными для специалистов в настоящее время являются не менее чем 100 тысяч публикаций (подробнее см. статьи [6,7]). Реально же каждый из них знаком с существенно меньшим количеством книг и статей. Так, в наиболее солидном и обширном из научных изданий в области эконометрики и прикладной статистики - трехтомнике Кендалла и Стьюарта [8-10] всего около 2 тысяч литературных ссылок. При всей очевидности соображений о многократном дублировании ценных идей в различных публикациях приходится признать, что каждый специалист по эконометрике и прикладной статистике владеет лишь небольшой частью накопленных в этой области знаний. Не удивительно, что приходится постоянно сталкиваться с игнорированием или повторением ранее полученных результатов, с уходом в тупиковые (с точки зрения практики) направления исследований, с беспомощностью при обращении к реальным данным, и т.д. Все это - одно из проявлений адапционного механизма торможения развития науки, которая оказывается не в состоянии даже осмыслить ранее полученные результаты. Об этом печальном явлении еще более 30 лет назад писали В.В.Налимов и другие науковеды (см., например, [11]).

Традиционный предрассудок состоит в том, что каждый новый результат, полученный исследователем - это кирпич, вложенный в непрерывно растущее здание науки, который непременно будет проанализирован и использован научным сообществом. Реальная ситуация - совсем иная. Как известно, большинство книг в центральных библиотеках никто никогда не читал. Так что с новым результатом, скорее всего, познакомятся лишь несколько человек, да и то поверхностно, а использовать его будут, в лучшем случае, сам автор в дальнейших работах и его ученики.

Основа профессиональных знаний экономиста, менеджера, исследователя и инженера закладывается в период обучения. Затем они пополняются в том узком направлении, в котором работает специалист. Следующий этап - их тиражирование новому поколению. В результате вузовские учебники отстоят от современного развития на десятки лет. Так, учебники по математической статистике, по нашей экспертной оценке, в основном соответствуют 40-60-м годам ХХ в. А потому тем же годам соответствует по своему научному и методологическому уровню большинство вновь публикуемых исследований и тем более - прикладных работ. Одновременно приходится признать, что результаты, которым не повезло, поскольку они не вошли в учебники, независимо от их научной и (или) прикладной ценности почти все забываются.

Активно продолжается развитие тупиковых направлений. Психологически это понятно. Приведу пример из своего опыта. В свое время по заказу Госстандарта я разработал методы оценки параметров гамма-распределения (см. государственный стандарт [12]. Поэтому мне близки и интересны работы по оцениванию параметров по выборкам из распределений, принадлежащих тем или иным параметрическим семействам, понятия функции максимального правдоподобия, эффективности оценок, использование неравенства Рао-Крамера и т.д. К сожалению, я знаю, что это - тупиковая ветвь, поскольку реальные данные не подчиняются каким-либо параметрическим семействам, надо применять иные статистические методы, о которых речь пойдет ниже. Понятно, что специалистам по параметрической статистике, потратившим многие годы на совершенствование в своей области, психологически трудно согласиться с подобным утверждением. В том числе и мне было трудно перейти на другую позицию, отраженную в настоящей книге и исходящую из потребностей прикладных работ.

**Точки роста**

Отечественная литература по эконометрике и прикладной статистике столь же необозрима, как и мировая. Только в секции "Математические методы исследования" журнала "Заводская лаборатория" с 1960-х годов опубликовано более 1000 статей. Не будем даже пытаться перечислять коллективы исследователей или основные монографии в этой области. Отметим только одно издание. По нашему мнению, наилучшей отечественной книгой по прикладной статистике является сборник статистических таблиц Л.Н. Большева и Н.В.Смирнова [13] с подробными комментариями, играющими роль сжатого учебника и справочника.

Основная цель настоящей главы - выделить и обсудить "точки роста" эконометрики и прикладной статистики, те их направления, которые представляются перспективными в будущем, в XXI веке, но пока в большинстве учебных изданий отодвинуты на задний план традиционными постановками.

При описании современного этапа развития эконометрических и статистических методов целесообразно выделить пять актуальных направлений, в которых развивается современная прикладная статистика, т.е. пять "точек роста": непараметрика (т.е. непараметрическая статистика), робастность, бутстреп, статистика интервальных данных, статистика нечисловых данных (в несколько иной терминологии - статистика объектов нечисловой природы). Обсудим их.

**Непараметрическая статистика** (см. также главу 4). В первой трети ХХ в., одновременно с параметрической статистикой, в работах Спирмена и Кендалла появились первые непараметрические методы, основанные на коэффициентах ранговой корреляции, носящих ныне имена этих статистиков. Но непараметрика, не делающая нереалистических предположений о том, что функции распределения результатов наблюдений принадлежат тем или иным параметрическим семействам распределений, стала заметной частью статистики лишь со второй трети ХХ века. В 30-е годы появились работы А.Н.Колмогорова и Н.В.Смирнова, предложивших и изучивших статистические критерии, носящие в настоящее время их имена. Эти критерии основаны на использовании так называемого эмпирического процесса. (Как известно, эмпирический процесс – это разность между эмпирической и теоретической функциями распределения, умноженная на квадратный корень из объема выборки.) В работе А.Н.Колмогорова 1933 г. изучено предельное распределение супремума модуля эмпирического процесса, называемого сейчас критерием Колмогорова. Затем Н.В. Смирнов исследовал супремум и инфимум эмпирического процесса, а также интеграл (по теоретической функции распределения) квадрата эмпирического процесса.

Следует отметить, что встречающееся иногда в литературе словосочетание "критерий Колмогорова-Смирнова" некорректно, поскольку эти два статистика никогда не печатались вместе и не изучали один и тот же критерий схожими методами. Корректно сочетание "критерий типа Колмогорова-Смирнова", применяемое для обозначения критериев, основанных на использовании супремума функций от эмпирического процесса.

После второй мировой войны развитие непараметрической статистики пошло быстрыми темпами. Большую роль сыграли работы Ф. Вилкоксона и его школы. К настоящему времени с помощью непараметрических методов можно решать практически тот же круг статистических задач, что и с помощью параметрических. Однако для обеспечения широкого внедрения непараметрических методов необходимо провести еще целый комплекс теоретических и пилотных (т.е. пробных) прикладных работ. Все большую роль играют непараметрические оценки плотности, непараметрические методы регрессии и распознавания образов (дискриминантного анализа). В нашей стране непараметрические методы получили достаточно большую известность после выхода в 1965 г. первого издания упомянутого выше сборника статистических таблиц Л.Н. Большева и Н.В.Смирнова [13], содержащего подробные таблицы для основных непараметрических критериев.

Тем не менее параметрические методы всё еще популярнее непараметрических, особенно среди тех прикладников, кто слабо знаком со статистическими методами. Неоднократно публиковались (см. начало гл.4) экспериментальные данные, свидетельствующие о том, что распределения реально наблюдаемых случайных величин, в частности, ошибок измерения, в подавляющем большинстве случаев отличны от нормальных (гауссовских). Тем не менее теоретики продолжают строить и изучать статистические модели, основанные на гауссовости, а практики - применять подобные методы и модели. Другими словами, "ищут под фонарем, а не там, где потеряли".

**Устойчивость статистических процедур (робастность)** (см. также главу 10). Если в параметрических постановках на данных накладываются слишком жесткие требования - их функции распределения должны принадлежать определенному параметрическому семейству, то в непараметрических, наоборот, излишне слабые - требуется лишь, чтобы функции распределения были непрерывны. При этом игнорируется априорная информация о том, каков "примерный вид" распределения. Априори можно ожидать, что учет этого "примерного вида" улучшит показатели качества статистических процедур. Развитием этой идеи является теория устойчивости (робастности) статистических процедур, в которой предполагается, что распределение исходных данных мало отличается от некоторого параметрического семейства. За рубежом эту теорию разрабатывали П.Хубер, Ф.Хампель и многие другие. Из монографий на русском языке, трактующих о робастности и устойчивости статистических процедур, самой ранней и наиболее общей была книга [14], следующей - монография [15]. Частными случаями реализации идеи робастности (устойчивости) статистических процедур являются статистика объектов нечисловой природы (см. главу 8) и статистика интервальных данных (см. главу 3)..

Имеется большое разнообразие моделей робастности 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. Однако эта модель представляется "тупиковой", поскольку в большинстве случаев большие выбросы либо невозможны из-за ограниченности шкалы прибора либо интервала изменения измеряемой величины, либо от них можно избавиться, применяя лишь статистики, построенные по центральной части вариационного ряда. Кроме того, в подобных моделях обычно считается известной частота засорения, что в сочетании со сказанным выше делает их малопригодными для практического использования.

Более перспективным представляется, например, модель малых отклонений распределений, в которой расстояние между распределением каждого элемента выборки и базовым распределением не превосходит заданной малой величины, и модель статистики интервальных данных.

**Бутстреп (размножение выборок)** (см. также главу 11). Другое из упомянутых выше направлений - бутстреп - связано с интенсивным использованием возможностей вычислительной техники. Основная идея состоит в том, чтобы теоретическое исследование заменить вычислительным экспериментом. Вместо описания выборки распределением из параметрического семейства строим большое число "похожих" выборок, т.е. "размножаем" выборку. Затем вместо оценивания характеристик (и параметров) и проверки гипотез на основе свойств теоретического распределения решаем эти задачи вычислительным методом, рассчитывая интересующие нас статистики по каждой из "похожих" выборок и анализируя полученные при этом распределения. Например, вместо того, чтобы теоретическим путем находить распределение статистики, доверительные интервалы и другие характеристики, моделируют большое число выборок, похожих на исходную, затем рассчитывают соответствующие значения интересующей исследователя статистики и изучают их эмпирическое распределение. Квантили этого распределения задают доверительные интервалы, и т.д.

Термин "бутстреп" мгновенно получил широкую известность после первой же статьи Б.Эфрона 1979 г. по этой тематике. Он сразу же стал обсуждаться в массе публикаций, в том числе и научно-популярных. В "Заводской лаборатории" № 10 за 1987 г. была помещена подборка статей по бутстрепу. На русском языке выпущен сборник статей Б. Эфрона [16]. Основная идея бутстрепа по Б. Эфрону состоит в том, что методом Монте-Карло (статистических испытаний) многократно извлекаются выборки из эмпирического распределения. Эти выборки, естественно, являются вариантами исходной, напоминают ее.

Сама по себе идея "размножения выборок" была известна гораздо раньше. Одна из статей Б. Эфрона в сборнике [16] называется так: "Бутстреп-методы: новый взгляд на метод складного ножа". Упомянутый "метод складного ножа" (jackknife) предложен М. Кенуем еще в 1949 г., за 30 лет до появления статьи Б.Эфрона. "Размножение выборок" при этом осуществляется путем исключения одного наблюдения. Таким образом для выборки объема *n* получаем *n* "похожих" на нее выборок объема *(n - 1)* каждая. Если же исключать по 2 наблюдения, то число "похожих" выборок возрастает до *n (n - 1) / 2* объема *(n - 2)* каждая.

Преимущества и недостатки бутстрепа как статистического метода обсуждались в главе 11 выше. Там же приводится информация о ряде аналогичных методов. Необходимо подчеркнуть, что бутстреп по Эфрону - лишь один из вариантов методов "размножения выборки" (resampling), и, на наш взгляд, не самый удачный. Метод "складного ножа" представляется более полезным. На его основе можно сформулировать следующую простую практическую рекомендацию.

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

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

**Статистика интервальных данных** (см. также главу 9). Перспективное и быстро развивающееся направление последних лет - прикладная математическая статистика интервальных данных. Речь идет о развитии методов математической статистики в ситуации, когда статистические данные - не числа, а интервалы, в частности, порожденные наложением ошибок измерения на значения случайных величин.

Статистика интервальных данных идейно связана с интервальной математикой, в которой в роли чисел выступают интервалы. Это направление математики является дальнейшим развитием всем известных правил приближенных вычислений, посвященных выражению погрешностей суммы, разности, произведения, частного через погрешности тех чисел, над которыми осуществляются перечисленные операции. К настоящему времени удалось решить, в частности, ряд задач теории интервальных дифференциальных уравнений, в которых коэффициенты, начальные условия и решения описываются с помощью интервалов.

Одна из ведущих научных школ в области статистики интервальных данных - это школа проф. А.П. Вощинина, активно работающая с конца 70-х годов. В частности, изучены проблемы регрессионного анализа, планирования эксперимента, сравнения альтернатив и принятия решений в условиях интервальной неопределенности.

Рассмотрим другое направление в статистике интервальных данных, которое также представляется перспективным. В нем развиваются асимптотические методы статистического анализа интервальных данных при больших объемах выборок и малых погрешностях измерений. В отличие от классической математической статистики, сначала устремляется к бесконечности объем выборки и только потом - уменьшаются до нуля погрешности. В частности, с помощью такой асимптотики были сформулированы правила выбора метода оценивания параметров гамма-распределения в ГОСТ 11.011-83 [12].

В рамках рассматриваемого научного направления, разработана общая схема исследования, включающая расчет нотны (максимально возможного отклонения статистики, вызванного интервальностью исходных данных) и рационального объема выборки (превышение которого не дает существенного повышения точности оценивания). Она применена к оцениванию математического ожидания, дисперсии, коэффициента вариации, параметров гамма-распределения и характеристик аддитивных статистик, при проверке гипотез о параметрах нормального распределения, в т.ч. с помощью критерия Стьюдента, а также гипотезы однородности с помощью критерия Смирнова. Разработаны подходы к рассмотрению интервальных данных в основных постановках регрессионного, дискриминантного и кластерного анализов. В частности, изучено влияние погрешностей измерений и наблюдений на свойства алгоритмов регрессионного анализа, разработаны способы расчета нотн и рациональных объемов выборок, введены и исследованы новые понятия многомерных и асимптотических нотн, доказаны соответствующие предельные теоремы. Начата разработка интервального дискриминантного анализа, в частности, рассмотрено влияние интервальности данных на введенный в главе 5 показатель качества классификации. Изучено асимптотическое поведение оценок метода моментов и оценок максимального правдоподобия (а также более общих - оценок минимального контраста), проведено асимптотическое сравнение этих методов в случае интервальных данных. Найдены общие условия, при которых, в отличие от классической математической статистики, метод моментов дает более точные оценки, чем метод максимального правдоподобия.

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

**Статистика объектов нечисловой природы как часть прикладной статистики.** Согласно общепринятой в настоящее время классификации статистических методов прикладная статистика делится на следующие четыре области:

статистика (числовых) случайных величин (см. главу 4),

многомерный статистический анализ (см. главу 5),

статистика временных рядов и случайных процессов (см. главу 6),

статистика объектов нечисловой природы (см. главу 8),.

Первые три из этих областей являются классическими. Они были хорошо известны еще в первой половине ХХ в. Остановимся на четвертой, сравнительно недавно вошедшей в массовое сознание специалистов. Ее именуют также статистикой нечисловых данных или попросту нечисловой статистикой. Анализ динамики развития эконометрики и прикладной статистики приводит к выводу, что в XXI в. она станет центральной областью прикладной статистики, поскольку содержит наиболее общие подходы и результаты.

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

Примерами объектов нечисловой природы являются (подробнее см. главу 8):

значения качественных признаков, т.е. результаты кодировки объектов с помощью заданного перечня категорий (градаций);

упорядочения (ранжировки) экспертами образцов продукции (при оценке её технического уровня и конкурентоспособности)) или заявок на проведение научных работ (при проведении конкурсов на выделение грантов);

классификации, т.е. разбиения объектов на группы сходных между собой (кластеры);

толерантности, т.е. бинарные отношения, описывающие сходство объектов между собой, например, сходства тематики научных работ, оцениваемого экспертами с целью рационального формирования экспертных советов внутри определенной области науки;

результаты парных сравнений или контроля качества продукции по альтернативному признаку ("годен" - "брак"), т.е. последовательности из 0 и 1;

множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга;

слова, предложения, тексты;

вектора, координаты которых - совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности (т.н. форма № 1-наука) или заполненная компьютеризированная история болезни, в которой часть признаков носит качественный характер, а часть - количественный;

ответы на вопросы экспертной, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возможно, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты; и т.д.

Интервальные данные (см. выше) тоже можно рассматривать как пример объектов нечисловой природы, а именно, как частный случай нечетких множеств.

С начала 70-х годов под влиянием запросов прикладных исследований в социально-экономических, технических, медицинских науках в России активно развивается статистика объектов нечисловой природы, известная также как статистика нечисловых данных или нечисловая статистика. В создании этой сравнительно новой области эконометрики и прикладной математической статистики приоритет принадлежит российским ученым.

Большую роль сыграл основанный в 1973 г. научный семинар "Экспертные оценки и анализ данных". В 60-е годы советское научное сообщество стало интересоваться методами экспертных оценок (об их истории и современном состоянии см. главу 12). Как следствие, началось знакомство с конкретными математизированными теориями, связанными с этими методами. Речь идет о репрезентативной теории измерений, ставшей известной в нашей стране по статье П.Суппеса и Дж.Зинеса в сборнике [17] и книге И.Пфанцагля [18], о теории нечеткости, современный этап которой начался с работ Л.А.Заде [19], теории парных сравнений, описанной в монографии Г.Дэвида [20]. К этому кругу идей примыкают теория случайных множеств (см., например, книгу Ж.Матерона [21]) и методы многомерного шкалирования (описаны, в частности, в монографиях А.Ю.Терехиной [22] и В.Т.Перекреста [23]). Но наибольшее влияние оказали идеи Дж.Кемени, который аксиоматически ввел расстояние между ранжировками (теперь оно именуется в литературе расстоянием Кемени) и предложил использовать в качестве средней величины решение оптимизационной задачи (теперь - медиана Кемени). Его скромная книжка [24], написанная в соавторстве с Дж.Снеллом, породила большой поток исследований.

В течение 70-х годов на основе запросов теории экспертных оценок (а также социологии, экономики, техники и медицины) развивались конкретные направления статистики объектов нечисловой природы. Были установлены связи между конкретными видами таких объектов, разработаны для них вероятностные модели (см. главу 8). Научные итоги этого периода подведены в монографиях [14,25,26]).

Следующий этап - выделение статистики объектов нечисловой природы в качестве самостоятельного направления в эконометрике и прикладной статистике, ядром которого являются методы статистического анализа данных произвольной природы. Программа развития этого нового научного направления впервые была сформулирована в статье [27]. Реализация этой программы была осуществлена в 80-е годы. Для работ этого периода характерна сосредоточенность на внутренних проблемах нечисловой статистики. Ссылки на конкретные монографии, сборники, статьи и иные публикации нескольких десятков авторов приведены в главе 8. Отметим лишь сборник научных статей [28], полностью посвященный нечисловой статистике.

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

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

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

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

Первой обсудим проблему определения средних величин. В рамках репрезентативной теории измерений удается указать вид средних величин, соответствующих тем или иным шкалам измерения (см. главу 3). В классической математической статистике средние величины вводят с помощью операций сложения (выборочное среднее арифметическое, математическое ожидание) или упорядочения (выборочная и теоретическая медианы). В пространствах произвольной природы средние значения нельзя определить с помощью операций сложения или упорядочения. Теоретические и эмпирические средние приходится вводить как решения экстремальных задач. Для теоретического среднего это - задача минимизации математического ожидания (в классическом смысле) расстояния от случайного элемента со значениями в рассматриваемом пространстве до фиксированной точки этого пространства (минимизируется указанная функция от этой точки). Для эмпирического среднего математическое ожидание берется по эмпирическому распределению, т.е. берется сумма расстояний от некоторой точки до элементов выборки и затем минимизируется по этой точке. При этом как эмпирическое, так и теоретическое средние как решения экстремальных задач могут быть не единственным элементом пространства, а состоять из множества таких элементов, которое может оказаться и пустым. Тем не менее удалось сформулировать и доказать законы больших чисел для средних величин, определенных указанным образом, т.е. установить сходимость эмпирических средних к теоретическим .

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

В статистике в пространствах произвольной природы большую роль играют непараметрические оценки плотности, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализов. В нечисловой статистике предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в частности, доказана их состоятельность, изучена скорость сходимости и установлен примечательный факт совпадения наилучшей скорости сходимости в произвольном случае с той, которая имеет быть в классической теории для числовых случайных величин.

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

Для проверки гипотез могут быть использованы статистики интегрального типа, в частности, типа омега-квадрат. Любопытно, что предельная теория таких статистик, построенная первоначально в классической постановке [29], приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида [30], поскольку при этом удалось провести рассуждения, опираясь на базовые математические соотношения, а не на те частные (с общей точки зрения), что были связаны с конечномерным пространством.

Представляют практический интерес результаты, связанные с конкретными областями статистики объектов нечисловой природы, в частности, со статистикой нечетких множеств, развитой в книге [31], и со статистикой случайных множеств [14] (следует отметить, что теория нечетких множеств в определенном смысле сводится к теории случайных множеств [14,31]), с непараметрической теорией парных сравнений, с аксиоматическим введением метрик в конкретных пространствах объектов нечисловой природы [28], и с рядом других конкретных постановок (см. главу 8).

Для анализа нечисловых, в частности, экспертных данных весьма важны методы классификации. С другой стороны, наиболее естественно ставить и решать задачи классификации, основанные на использовании расстояний или показателей различия, в рамках статистики объектов нечисловой природы. Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа). Современное состояние дискриминантного и кластерного анализа с точки зрения статистики объектов нечисловой природы отражено в главе 5.

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

**Другие точки роста.** Выше рассмотрены пять "точек роста" эконометрики и прикладной статистики. Разумеется, они не исчерпывают все многообразие фронта научных исследований в рассматриваемых областях. Кроме того, в настоящей главе почти не затронуты разнообразные применения эконометрических и статистических методов в конкретных прикладных исследованиях и разработках. Много интересных проблем есть в планировании экспериментов, особенно кинетических (см., например, статью [31]), при анализе проблем надежности, в новых статистических методах управления качеством продукции (см. главу 13), в том числе в связи с идеями Г. Тагути, при анализе рисков (см. главу 14), в вопросах экологии и безопасности и др.

В течение последних более чем 60 лет в России наблюдается огромный разрыв между государственной статистикой и научным сообществом специалистов по статистическим методам (подробнее об этом см. статью [7]). В учебнике по истории статистики [32] даже не упоминаются имена членов-корреспондентов АН СССР Н.В.Смирнова и Л.Н. Большева! А ведь они – единственные представители именно математической статистики как таковой в Академии наук в ХХ в. (еще ряд членов Академии наук имели математическую статистику среди своих интересов, но Н.В. Смирнов и Л.Н. Большев занимались практически только ею). Поэтому нет ничего удивительного в том, что тенденции развития современной эконометрики и прикладной математической статистики столь же мало обсуждаются отечественными авторами, как и ее история.

**О некоторых нерешенных вопросах эконометрики и прикладной статистики**

За последние 30 лет выявился целый ряд нерешенных вопросов эконометрики и прикладной статистики , как чисто научных, так и научно-организационных. Обсудим пять из них:

влияние отклонений от традиционных предпосылок (вероятностно-статистических моделей) на свойства эконометрических и статистических процедур;

оправданность использования асимптотических теоретических результатов эконометрики и прикладной математической статистики при конечных объемах выборок;

формулировки и обоснования правил выбора одного из многих критериев для проверки конкретной гипотезы;

конкретные способы организации теоретических работ в области эконометрики и прикладной математической статистики;

организация и проведение прикладных работ с использованием методов эконометрики и прикладной математической статистики.

Настоящий раздел отнюдь не претендует на решение перечисленных вопросов. Его цель гораздо скромнее - обратить внимание на существование ряда нерешенных вопросов в надежде, что коллективными усилиями удастся продвинуться в их решении.

**Влияние отклонений от традиционных предпосылок.** В вероятностной теории статистических методов выборка обычно моделируется как конечная последовательность независимых одинаково распределенных случайных величин или векторов. Часто предполагается, что эти величины (вектора) имеют нормальное распределение.

На основе сформулированных классических предпосылок построено огромное здание классической математической статистики с большим числом теорем. Оно за последние 100 лет обросло горой учебников и программных продуктов.

Однако при внимательном взгляде совершенно ясна нереалистичность классических предпосылок. Независимость результатов измерений обычно принимается "из общих предположений", между тем во многих случаях очевидна их коррелированность [33]. Одинаковая распределенность также вызывает сомнения из-за изменения во времени свойств измеряемых образцов, средств измерения и психофизического состояния специалистов, проводящих измерения (наблюдения, испытания, анализы, опыты). Даже обоснованность самой возможности применения вероятностных моделей также часто вызывает сомнения, например, при моделировании уникальных измерений (теорию вероятностей обычно привлекают при изучении массовых явлений). И уж совсем редко распределения результатов измерений можно считать нормальными (см. главу 4).

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

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

**Пример 2.** Для проверки однородности двух независимых выборок (на самом деле - для проверки равенства математических ожиданий) обычно рекомендуют использовать двухвыборочный критерий Стьюдента. Что будет при отклонении от нормальности распределений, из которых взяты выборки? Если объемы выборок равны или если дисперсии результатов наблюдений в выборках совпадают, то в асимптотике (когда объемы выборок безгранично возрастают) классический метод является корректным. Если же объемы выборок существенно отличаются и их дисперсии различны, то двухвыборочную статистику Стьюдента применять нельзя. Поскольку проверка равенства дисперсий - более сложная задача, чем проверка равенства математических ожиданий, то для выборок разного объема использовать двухвыборочную статистику Стьюдента не следует, лучше применять критерий Крамера- Уэлча, как это подробно обосновано в главе 4.

**Пример 3.** В задаче отбраковки (исключения) резко выделяющихся наблюдений (выбросов) расчетные методы, основанные на нормальности, являются крайне неустойчивыми по отношению к отклонениям от нормальности, что полностью лишает эти методы научной обоснованности (подробнее см. главу 4).

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

Итак, имеется **необходимость изучения свойств расчетных методов классической математической статистики, опирающихся на предположение нормальности, в ситуациях, когда это предположение не выполнено.** Аппаратом для такого изучения наряду с методом Монте-Карло (статистических испытаний) могут послужить предельные теоремы теории вероятностей (и опирающиеся на них асимптотические методы математической статистики), прежде всего ЦПТ, поскольку интересующие нас расчетные методы обычно используют разнообразные суммы.

Пока подобное изучение не проведено, остается неясной научная ценность, например, применения факторного анализа к векторам из переменных, принимающих небольшое число градаций и к тому же измеренных в порядковой шкале. Этот пример показывает важность еще одного направления исследований - изучения свойств алгоритмов, предназначенных для анализа числовых данных, в случаях, когда данные измерены в шкалах, отличных от абсолютной, в частности, в порядковой шкале. Подробнее это направление рассмотрено в главе 3.

Из большого числа возможных постановок, относящихся к изучению влияния отклонений от традиционных предпосылок, укажем лишь на то, что реальные данные имеют небольшое число значащих цифр (обычно от 2 до 5), в то время как в классической математической статистике используются непрерывные случайные величины, для которых вероятность получения подобного результата наблюдения равна 0. Действительно, вероятность того, что хотя бы один элемент выборки из распределения с непрерывной функцией распределение попадет в заданное счетное множество, в частности, в множество рациональных чисел, равна 0 (согласно классическим свойствам вероятностной меры). Событиями, имеющими вероятность 0, принято пренебрегать. Следовательно, с точки зрения классической математической статистики любыми реальными данными нужно пренебречь! Выходов из этого парадокса несколько. Один из них - бурно развивающаяся в настоящее время статистика интервальных данных (см. главу 9), другой - использование классических поправок Шеппарда для сгруппированных данных [34,35]. Здесь еще много работы. Так, даже для такого широко используемого статистического показателя, как коэффициент корреляции, поправки на группировку (поправки Шеппарда) были получены сравнительно недавно - лишь в 1980 г. [35].

Почему на первый план выдвинуто изучение классических алгоритмов, а не построение новых, специально предназначенных для работы в условиях отклонения от классических предпосылок? Во-первых, потому, что классические алгоритмы в настоящее время наиболее распространены (благодаря сложившейся системе образования как прикладников, так и математиков). Во-вторых, более новые подходы зачастую методологически уязвимы. Так, известная робастная модель засорения Тьюки-Хубера (см. главу 10) нацелена на борьбу с большими выбросами, которые зачастую физически невозможны из-за ограниченности интервала возможных значений измеряемой характеристики, в котором работает конкретное средство измерения. Следовательно, модель Тьюки-Хубера имеет скорее теоретическое значение, чем практическое. Сказанное, конечно, не означает, что следует прекратить разработку, изучение и внедрение непараметрических и устойчивых методов, выделенных выше как "точки роста" современных эконометрики и прикладной статистики.

**Использование асимптотических результатов при конечных объемах выборок.** Как отмечено выше, изучение классических алгоритмов во многих случаях может быть проведено с помощью асимптотических методов математической статистики, в частности, с помощью ЦПТ и методов наследования сходимости [14, п.2.4]. Отрыв классической математической статистики от нужд прикладных исследований проявился, в частности, в том, что в распространенных монографиях недостает математического аппарата, необходимого, в частности, для изучения двухвыборочных статистик. Суть в том, что переходить к пределу приходится не по одному параметру, а по двум – объемам двух выборок. Пришлось разработать соответствующую теорию – теорию наследования сходимости, изложенную в монографии [14, п.2.4].

Однако применять результаты подобного изучения придется при конечных объемах выборок. Возникает целый букет проблем, связанных с таким переходом. Часть из них обсуждалась в статье [37] в связи с изучением свойств статистик, построенных по выборкам из конкретных распределений.

Однако при обсуждении влияния отклонений от исходных предположений на свойства статистических процедур возникают дополнительные проблемы. Какие отклонения считать типичными? Ориентироваться ли на наиболее "вредные" отклонения, в наибольшей степени искажающие свойства алгоритмов, или же сосредоточить внимание на "типичных" отклонениях?

При первом подходе получаем гарантированный результат, но "цена" этого результата может быть излишне высокой. В качестве примера укажем на универсальное неравенство Берри-Эссеена для погрешности в ЦПТ [38,39]. Совершенно справедливо подчеркивает академик РАН А.А. Боровков [39, с,172], что "скорость сходимости в реальных задачах, как правило, оказывается лучше."

При втором подходе возникает вопрос, какие отклонения считать "типичными". Попытаться ответить на этот вопрос можно, анализируя большие массивы реальных данных. Вполне естественно, что ответы различных исследовательских групп будут различаться.

Одна из ложных идей - использование при анализе возможных отклонений только какого-либо конкретного параметрического семейства – распределений Вейбулла-Гнеденко, трехпараметрического семейства гамма - распределений и др. Как уже отмечалось выше, еще в 1927 г. акад. АН СССР С.Н. Бернштейн обсуждал методологическую ошибку, состоящую в сведении всех эмпирических распределений к четырехпараметрическому семейству Пирсона [5]. Однако и до сих пор параметрические методы статистики весьма популярны, особенно среди прикладников, и вина за это заблуждение лежит прежде всего на преподавателях статистических методов.

**Выбор одного из многих критериев для проверки конкретной гипотезы.** Во многих случаях для решения конкретной практической задачи разработано много методов, и специалист по математическим методам исследования стоит перед проблемой: какой из них предложить прикладнику для анализа конкретных данных?

В качестве примера рассмотрим задачу проверки однородности двух независимых выборок. Как известно [13], для ее решения можно предложить массу критериев: Стьюдента, Крамера-Уэлча, Лорда, хи - квадрат, Вилкоксона (Манна-Уитни), Ван – дер - Вардена, Сэвиджа, Н.В.Смирнова, типа омега-квадрат (Лемана-Розенблатта), Г.В. Мартынова и др. Какой выбрать?

Естественным образом приходит в голову идея "голосования": провести проверку по многим критериям, а затем принять решение "по большинству голосов". С точки зрения статистической теории такая процедура приводит попросту к построению еще одного критерия, который априори ничем не лучше прежних (но и не хуже), но более труден для изучения. С другой стороны, если совпадают решения по всем рассмотренным статистическим критериям, исходящим из различных принципов, то в соответствии с концепцией устойчивости, развитой в монографии [14], это повышает доверие к полученному общему решению.

Распространено, особенно среди математиков, ложное и вредное мнение о необходимости поиска оптимальных методов, решений и т.д. Дело в том, что оптимальность обычно исчезает при отклонении от исходных предпосылок. Так, среднее арифметическое в качестве оценки математического ожидания является оптимальной оценкой только тогда, когда исходное распределение - нормальное (см., например, монографию [40]), в то время как состоятельной оценкой - всегда, лишь бы математическое ожидание существовало. С другой стороны, для любого произвольно взятого метода оценивания или проверки гипотез обычно можно так сформулировать понятие оптимальности, чтобы рассматриваемый метод стал оптимальным – с этой специально выбранной точки зрения. Возьмем, например, выборочную медиану как оценку математического ожидания. Она, разумеется, оптимальна, хотя и в другом смысле, чем среднее арифметическое (оптимальное для нормального распределения). А именно, для распределения Лапласа выборочная медиана является оценкой максимального правдоподобия, а потому оптимальной - в том смысле, в каком оптимальной является любая оценка максимального правдоподобия. Соответствующее понятие оптимальности требует аккуратных формулировок, оно строго изложено в монографии [41]. Как известно, оценки максимального правдоподобия удобны при теоретических рассмотрениях, а при анализе конкретных экономических, технических и иных данных следует применять одношаговые оценки (см. об этом статью [42]).

Критерии однородности были проанализированы в монографии проф. Я.Ю. Никитина [43]. Естественных подходов к сравнению критериев несколько - на основе асимптотической относительной эффективности по Бахадуру, Ходжесу - Леману, Питмену. И выяснилось, что каждый критерий является оптимальным при соответствующей альтернативе или подходящем распределении на множестве альтернатив. При этом математические выкладки обычно используют альтернативу сдвига, сравнительно редко встречающуюся в практике анализа реальных статистических данных (в связи с критерием Вилкоксона эта альтернатива обсуждалась в главе 4). Итог печален - блестящая математическая техника, продемонстрированная в монографии [43], не позволяет дать рекомендации для выбора критерия проверки однородности при анализе реальных данных. Другими словами, с точки зрения работы прикладника, т.е. анализа конкретных данных, монография [43] бесполезна. Блестящее владение математикой и огромное трудолюбие, продемонстрированные автором этой монографии, увы, ничего не принесли практике.

Конечно, каждый практически работающий статистик так или иначе решает для себя проблему выбора статистического критерия. На основе ряда методологических соображений в главе 4 мы остановили свой выбор на состоятельном против любой альтернативы критерии типа омега-квадрат (Лемана-Розенблатта). Однако остается чувство неудовлетворенности в связи с недостаточной теоретической обоснованностью этого выбора.

**Организация теоретических работ в области эконометрики и прикладной статистики.** Выше продемонстрирована необходимость большой теоретической работы по развитию нацеленных на практическое использование математических методов исследования. В статье [6] 1992 г. обоснован вывод о необходимости создания сети научно-исследовательских организаций, которая выполняла бы такую работу. Как известно, количество научных работников к настоящему времени сократилось по крайней мере в 3 раза по сравнению с началом 1990-х годов, так что на осуществление в ближайшие годы сформулированной в [6] научно-организационной программы надеяться не приходится.

Приходится с сожалением констатировать, что в рамках научной специальности "теория вероятностей и математическая статистика" наблюдается четко выраженное игнорирование проблем статистического анализа реальных данных и уход в глубь узкоматематических исследований, которые ничего не могут дать практике. Причины этого явления, типичного для математических дисциплин, обсуждались выше. Поэтому нет оснований ожидать, что при "естественном ходе событий" будут получены существенные продвижения в рассмотренных выше нерешенных проблемах эконометрики и прикладной математической статистики.

Помочь может выделение государственными структурами системы грантов, направленных на поддержку работ в области нерешенных эконометрики и прикладной математической статистики. Принципиальным шагом явилось бы выделение эконометрики и прикладной математической статистики как самостоятельных научных направлений, отличных как от чисто математических дисциплин типа "теории вероятностей и математической статистики", так и от, например, ветви экономической теории, известной в официальных кругах под названием "статистика".

**О прикладных работах с использованием методов прикладной статистики.** Проблемы организации теоретических работ в области эконометрики и прикладной математической статистики лишь в перспективе важны для практической работы. Как правило, те, кто обрабатывает реальные данные, недостаточно знакомы с теоретическими основами алгоритмов и тем более не следят за событиями "на переднем крае" обсуждаемой научно-методической дисциплины. Это вполне естественно, поскольку основная специальность у таких специалистов - иная.

Несколько огрубляя, можно сказать, что реально используется только то, что имеется в учебниках и справочниках, в широко распространенных программных продуктах, а научные публикации с точки зрения прикладника представляют собой "информационный шум". Ситуация усугубляется традиционным ненормальным положением в отечественной статистике [7], наличием ошибок во многих изданиях.

К сожалению, учебная и научная литература на русском языке (как, впрочем, и на иных языках) по эконометрике и прикладной статистике в целом далека от совершенства, переполнена устаревшими методологическими подходами и прямыми ошибками. До сих пор наилучшим изданием остаются "Таблицы математической статистики" Л.Н. Большева и Н.В.Смирнова [13], созданные в 60-х годах.

Хотя студенты почти всех специальностей изучают в конце курса высшей математики раздел "теория вероятностей и математическая статистика", реально они знакомятся лишь с некоторыми основными понятиями и результатами, которых явно не достаточно для практической работы. С некоторыми математическими методами исследования студенты встречаются в специальных курсах (например, таких, как "Прогнозирование и технико-экономическое планирование", "Технико-экономический анализ", "Контроль качества продукции", "Маркетинг", "Контроллинг", "Математические методы прогнозирования" и др.), однако изложение в большинстве случаев носит весьма сокращенный и рецептурный характер. В результате подавляющую часть специалистов по эконометрике, прикладной математической статистике и их применению следует считать самоучками.

Поэтому большое значение имеет введение в технических вузах курса "Прикладная математическая статистика", а на экономических факультетах таких вузов – курса «Эконометрика», поскольку эконометрика – это, как известно, статистический анализ конкретных экономических данных (см. главу 1). Это естественно делать, например, в рамках подпрограммы "Технологии подготовки кадров для национальной технологической базы" федеральной целевой программы "Национальная технологическая база". Естественно, что курсы "Прикладная математическая статистика" и «Эконометрика» должны быть обеспечены соответствующими учебниками и учебными пособиями, методическими материалами и обучающими компьютерными системами.

Только через систему образования можно поднять уровень массового применения эконометрики и прикладной статистики и сократить отставание от "переднего края" теории. А это отставание в настоящее время составляет не менее 20 (но и не более 100) лет.

**Высокие статистические технологии и эконометрика**

В настоящем пункте подробно обсуждается ранее введенное понятие "высокие статистические технологии". Рассматриваются причины широкого распространения устаревших и частично ошибочных "низких" статистических технологий. Показано, что из всех путей повышения качества прикладных статистических исследований наиболее эффективным является расширение обучения "высоким статистическим технологиям", в том числе под именем эконометрики. Описан опыт работы Института высоких статистических технологий и эконометрики МГТУ им. Н.Э. Баумана.

Термин "высокие технологии" популярен в современной научно-технической литературе. Он используется для обозначения наиболее передовых технологий, опирающихся на последние достижения научно-технического прогресса. Есть такие технологии и среди технологий статистического анализа данных - как в любой интенсивно развивающейся научно-практической области.

Примеры высоких статистических технологий и входящих в них алгоритмов анализа данных, подробный анализ современного состояния и перспектив развития даны выше при обсуждении “точек роста” эконометрики как научно-практической дисциплины. В качестве "высоких статистических технологий" были выделены технологии непараметрического анализа данных; устойчивые (робастные) технологии; технологии, основанные на размножении выборок, на использовании достижений статистики нечисловых данных и статистики интервальных данных.

**Термин "высокие статистические технологии".** Обсудим пока не вполне привычный термин "высокие статистические технологии". Каждое из трех слов несет свою смысловую нагрузку.

"Высокие", как и в других областях, означает, что статистическая технология опирается на современные достижения статистической теории и практики, в частности, теории вероятностей и прикладной математической статистики. При этом "опирается на современные научные достижения" означает, во-первых, что математическая основа технологии получена сравнительно недавно в рамках соответствующей научной дисциплины, во-вторых, что алгоритмы расчетов разработаны и обоснованы в соответствии в нею (а не являются т.н. "эвристическими"). Со временем, если новые подходы и результаты не заставляют пересмотреть оценку применимости и возможностей технологии, заменить ее на более современную, "высокие статистические технологии" переходят в "классические статистические технологии", такие, как метод наименьших квадратов. Итак, высокие статистические технологии - плоды недавних серьезных научных исследований. Здесь два ключевых понятия - "молодость" технологии (во всяком случае, не старше 50 лет, а лучше - не старше 10 или 30 лет) и опора на "высокую науку".

Термин "статистические" привычен, но разъяснить его нелегко. Во всяком случае, к деятельности Государственного комитета РФ по статистике высокие статистические технологии отношения не имеют. Как известно, сотрудники проф. В.В. Налимова собрали более 200 определений термина "статистика" [44]. Полемика вокруг терминологии иногда принимает весьма острые формы (см., например, редакционные замечания к статье [1], написанные в стиле известных высказываний о генетике и кибернетике конца 1940-х годов). Современное представление о терминологии в области теории вероятностей и математической статистики отражено в Приложении 1 к настоящей книге, подготовленном в противовес распространенным ошибкам и неточностям в этой области. В частности, с точки зрения эконометрики статистические данные – это результаты измерений, наблюдений, испытаний, анализов, опытов, а "статистические технологии" - это технологии анализа статистических данных.

Наконец, редко используемый применительно к статистике термин "технологии". Статистический анализ данных, как правило, включает в себя целый ряд процедур и алгоритмов, выполняемых последовательно, параллельно или по более сложной схеме. В частности, можно выделить следующие этапы:

- планирование статистического исследования;

- организация сбора необходимых статистических данных по оптимальной или хотя бы рациональной программе (планирование выборки, создание организационной структуры и подбор команды эконометриков или 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 предметной области);

- первичное описание данных (расчет различных выборочных характеристик, функций распределения, непараметрических оценок плотности, построение гистограмм, корреляционных полей, различных таблиц и диаграмм и т.д.),

- оценивание тех или иных числовых или нечисловых характеристик и 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 принимается решение о проверке той или иной последующей гипотезы),

- более углубленное изучение, т.е. применение различных алгоритмов многомерного статистического анализа, алгоритмов диагностики и построения классификации, 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, допустимых преобразований шкал измерения, в частности, изучение свойств оценок методом размножения выборок;

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

- составление итоговых отчетов, в частности, предназначенных для тех, кто не является специалистами в эконометрических и статистических методах анализа данных, в том числе для руководства - "лиц, принимающих решения".

Возможны и иные структуризации статистических технологий. Важно подчеркнуть, что квалифицированное и результативное применение статистических методов - это отнюдь не проверка одной отдельно взятой статистической гипотезы или оценка параметров одного заданного распределения из фиксированного семейства. Подобного рода операции - только отдельные кирпичики, из которых складывается здание статистической технологии. Между тем учебники и монографии по статистике обычно рассказывают об отдельных кирпичиках, но не обсуждают проблемы их организации в технологию, предназначенную для прикладного использования.

Итак, процедура эконометрического или статистического анализа данных – это информационный технологический процесс, другими словами, та или иная информационная технология. Статистическая информация подвергается разнообразным операциям (последовательно, параллельно или по более сложным схемам). В настоящее время об автоматизации всего процесса статистического анализа данных говорить было бы несерьезно, поскольку имеется слишком много нерешенных проблем, вызывающих дискуссии среди статистиков. "Экспертные системы" в области статистического анализа данных пока не стали рабочим инструментом статистиков. Ясно, что и не могли стать. Можно сказать и жестче - это пока научная фантастика или даже вредная утопия.

В литературе статистические технологии рассматриваются явно недостаточно. В частности, обычно все внимание сосредотачивается на том или ином элементе технологической цепочки, а переход от одного элемента к другому остается в тени. Между тем проблема "стыковки" статистических алгоритмов, как известно, требует специального рассмотрения, поскольку в результате использования предыдущего алгоритма зачастую нарушаются условия применимости последующего. В частности, результаты наблюдений могут перестать быть независимыми, может измениться их распределение и т.п. (см. обсуждение этой проблемы в статье [45]).

Например, при проверке статистических гипотез большое значение имеют такие хорошо известные характеристики статистических критериев, как уровень значимости и мощность. Методы их расчета и использования при проверке одной гипотезы обычно хорошо известны. Если же сначала проверяется одна гипотеза, а потом с учетом результатов ее проверки - вторая, то итоговая процедура, которую также можно рассматривать как проверку некоторой (более сложной) статистической гипотезы, имеет характеристики (уровень значимости и мощность), которые, как правило, нельзя просто выразить через характеристики двух составляющих гипотез, а потому они обычно неизвестны. В результате итоговую процедуру нельзя рассматривать как научно обоснованную, она относится к эвристическим алгоритмам. Конечно, после соответствующего изучения, например, методом Монте-Карло, она может войти в число научно обоснованных процедур прикладной статистики.

**Почему живучи "низкие статистические технологии"?** "Высоким статистическим технологиям" противостоят, естественно, "низкие статистические технологии". Это те технологии, которые не соответствуют современному уровню науки и техники. Обычно они одновременно и устарели, и не адекватны сути решаемых эконометрических и статистических задач.

Примеры таких технологий неоднократно критически рассматривались на страницах различных изданий. В главе 4 рассматривались примеры неправильного использования критерия Вилкоксона для проверки совпадения теоретических медиан или функций распределения двух выборок. Можно также вспомнить критику использования классических процентных точек критериев Колмогорова и омега-квадрат в ситуациях, когда параметры оцениваются по выборке и эти оценки подставляются в "теоретическую" функцию распределения [46]. Приходилось констатировать широкое распространение таких порочных технологий и конкретных алгоритмов, в том числе в государственных и международных стандартах (перечень ошибочных стандартов дан в статье [47]), учебниках и распространенных пособиях. Тиражирование ошибок происходит обычно в процессе обучения в вузах или путем самообразования при использовании недоброкачественной литературы.

На первый взгляд вызывает удивление устойчивость "низких статистических технологий", их постоянное возрождение во все новых статьях, монографиях, учебниках. Поэтому, как ни странно, наиболее "долгоживущими" оказываются не работы, посвященные новым научным результатам, а публикации, разоблачающие ошибки, типа статьи [46]. Прошло больше 15 лет с момента ее публикации, но она по-прежнему актуальна, поскольку ошибочное применение критериев Колмогорова и омега-квадрат по-прежнему распространено.

Целесообразно рассмотреть здесь по крайней мере три обстоятельства, которые определяют эту устойчивость ошибок.

Во-первых, прочно закрепившаяся традиция. Учебники по т.н. «Общей теории статистики», написанные экономистами (поскольку учебная дисциплина "статистика" официально относится к экономике), если беспристрастно проанализировать их содержание, состоят в основном из введения в прикладную статистику, изложенного в стиле «низких статистических технологий», на уровне 1950-х годов. К "низкой" прикладной статистике добавлена некоторая информация о деятельности органов Госкомстата РФ. Примерно таково же положение со статистическими методами в медицине - одни и те же "низкие статистические технологии" переписываются из книги в книгу. Кратко говоря, «профессора-невежды порождают новых невежд» [7]. Так мы писали в 1990 г., но никто из указанных невежд даже не поинтересовался, какие ошибки имеются в виду. Новое поколение, обучившись ошибочным алгоритмам, их использует, а с течением времени и достижением должностей, ученых званий и степеней– пишет новые учебники со старыми ошибками.

Руководство Госкомстата РФ, воспользовавшись катаклизмами начала 1990-х годов, сделало вид, что ему неизвестно о создании в 1990 г. Всесоюзной статистической ассоциации и секции статистических методов в ее составе. Госкомстат РФ по-прежнему закрыт от "высоких статистических технологий" и работает на уровне позапрошлого века. Защита стала надежнее, поскольку в соответствии с современным стилем аппаратной работы на письма и обращения можно не отвечать.

Второе обстоятельство связано с большими трудностями при оценке экономической эффективности применения статистических методов вообще и при оценке вреда от применения ошибочных методов в частности. (А без такой оценки как докажешь, что "высокие статистические технологии" лучше "низких"?) Некоторые соображения по первому из этих вопросов приведены в статье [1], содержащей оценки экономической эффективности ряда работ по применению статистических методов. При оценке вреда от применения ошибочных методов приходится учитывать, что общий успех в конкретной инженерной или научной работе вполне мог быть достигнут вопреки их применению, за счет "запаса прочности" других составляющих общей работы. Например, преимущество одного технологического приема над другим можно продемонстрировать как с помощью критерия Крамера-Уэлча проверки равенства математических ожиданий (что правильно), так и с помощью двухвыборочного критерия Стьюдента (что, вообще говоря, неверно, т.к. обычно не выполняются условия применимости этого критерия - нет ни нормальности распределения, ни равенства дисперсий). Кроме того, приходится выдерживать натиск невежд, защищающих свои ошибочные работы, например, государственные стандарты. Вместо исправления ошибок применяются самые разные приемы бюрократической борьбы с теми, кто разоблачает ошибки (подробнее см. статью [47]).

Третье существенное обстоятельство – трудности со знакомством с высокими статистическими технологиями. В течение последних 10 лет только журнал "Заводская лаборатория" систематически предоставлял такие возможности. К сожалению, поток современных отечественных и переводных статистических книг, выпускавшихся ранее, в частности, издательством “Финансы и статистика”, практически превратился в узкий ручеек… Возможно, более существенным является влияние естественной задержки во времени между созданием "новых статистических технологий" и написанием полноценной и объемной учебной и методической литературы. Она должна позволять знакомиться с новой методологией, новыми методами, теоремами, алгоритмами, технологиями не по кратким оригинальным статьям, а при обычном обучении в высшей школе.

**Как ускорить внедрение "высоких статистических технологий"?** Таким образом, весь арсенал используемых эконометрических и статистических методов можно распределить по трем потокам:

* высокие статистические технологии;
* классические статистические технологии,
* низкие статистические технологии.

Основная современная проблема статистических технологий состоит в обеспечении того, чтобы в конкретных эконометрических и статистических исследованиях использовались только технологии первых двух типов. При этом под классическими статистическими технологиями понимаем технологии почтенного возраста, сохранившие свое значение для современной статистической практики. Таковы метод наименьших квадратов, статистики Колмогорова, Смирнова, омега-квадрат, непараметрические коэффициенты корреляции Спирмена и Кендалла и многие другие статистические процедуры.

Каковы возможные пути решения основной современной проблемы в области статистических технологий?

Бороться с конкретными невеждами - дело почти безнадежное. Отстаивая свое положение и должности, они либо нагло игнорируют информацию о своих ошибках, как это делают авторы ряда учебников по "Общей теории статистики", либо с помощью различных бюрократических приемов уходят и от ответственности, и от исправления ошибок по существу (как это было со стандартами по статистическим методам - см. статью [6]). Третий вариант - признание и исправление ошибок - встречается, увы, редко. Но встречается.

Конечно, необходима демонстрация квалифицированного применения высоких статистических технологий. В 1960-70-х годах этим занималась лаборатория акад. А.Н. Колмогорова в МГУ им. М.В. Ломоносова. Секция "Математические методы исследования" журнала 2Заводская лаборатория" опубликовала за последние 40 лет более 1000 статей в стиле "высоких статистических технологий". В настоящее время действует Институт высоких статистических технологий и эконометрики МГТУ им. Н.Э.Баумана. Есть, конечно, целый ряд других научных коллективов, работающих на уровне "высоких статистических технологий".

Но самое основное - обучение. Какие бы новые научные результаты ни были получены, если они остаются неизвестными студентам, то новое поколение исследователей и инженеров вынуждено осваивать их по одиночке, а то и переоткрывать. Т.е. практически новые научные результаты почти исчезают, едва появившись. Как уже от меячалось, избыток публикаций превратился в тормоз развития. По нашим данным, к настоящему времени по статистическим технологиям опубликовано не менее миллиона статей и книг, из них не менее 100 тысяч являются актуальными для современного специалиста. Реальное число публикаций, которые способен освоить исследователь, по нашей оценке, не превышает 2-3 тысяч. Во всяком случае, в наиболее "толстом" (на русском языке) трехтомнике по статистике М. Дж. Кендалла и А. Стьюарта [8-10] приведено около 2 тысяч литературных ссылок. Итак, каждый исследователь знаком не более чем с 2-3% актуальных литературных источников. Поскольку существенная часть публикаций заражена "низкими статистическими технологиями", то исследователь самоучка имеет мало шансов выйти на уровень "высоких статистических технологий". Одновременно приходится констатировать, что масса полезных результатов погребена в изданиях прошлых десятилетий и имеет мало шансов встать в ряды "высоких статистических технологий" без специально организованных усилий современных специалистов.

Итак, еще и еще раз: основное - обучение. Несколько огрубляя, можно сказать: что то, что попало в учебные курсы и соответствующие учебные пособия - то сохраняется, что не попало - то пропадает. Подробнее об обучении - несколько позже. Сейчас - об упомянутом выше Институте высоких статистических технологий и эконометрики МГТУ им. Н.Э.Баумана.

**Институт высоких статистических технологий и эконометрики.** Организованный нами в 1989 г. Институт высоких статистических технологий и эконометрики (ИВСТЭ) действует на базе кафедры ИБМ-2 “Экономика и организация производства” Московского государственного технического университета им. Н.Э.Баумана. Институт на хоздоговорных и госбюджетных началах занимается развитием, изучением и внедрением эконометрики и "высоких статистических технологий", т.е. наиболее современных технологий анализа экономических, технических, социологических, медицинских данных, ориентированных на использование в условиях современного производства и экономики. Основной интерес представляют применения "высоких статистических технологий" для анализа конкретных экономических данных, т.е. в эконометрике. Наиболее перспективным представляется применение "высоких статистических технологий" для поддержки принятия управленческих решений, прежде всего в таком новом (для России) современном направлении экономической науки и практики, как контроллинг (см., например, монографию [48]).

Вначале Институт 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 НИЦентра по безопасности атомной энергетики, ВНИИ нефтепереработки, ПО “Пластик”, ЦНИИ черной металлургии им. 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, Казахского политехнического института, Донецкого государственного университета и многих других.

Затем Институт в качестве Лаборатории эконометрических исследований разрабатывал эконометрические методы анализа нечисловых данных, а также процедуры расчета и прогнозирования индекса инфляции и валового внутреннего продукта. Институт высоких статистических технологий и эконометрики развивал методологию построения и использования математических моделей процессов налогообложения (для Министерства налогов и сборов РФ), методологию оценки рисков реализации инновационных проектов высшей школы (для Министерства промышленности, науки и технологий РФ). Институт оценивал влияние различных факторов на формирование налогооблагаемой базы ряда налогов (для Минфина РФ), прорабатывал перспективы применения современных статистических и экспертных методов для анализа данных о научном потенциале (для Министерства промышленности, науки и технологий РФ). Важное направление связано с эколого-экономической тематикой - разработка методологического, программного и информационного обеспечения анализа рисков химико-технологических объектов (для Международного научно-технического центра), методов использования экспертных оценок в задачах экологического страхования (совместно с Институтом проблем рынка РАН). Институт проводил маркетинговые исследования (в частности, для Institute for Market Research GfK MR, Промрадтехбанка, фирм, торгующих растворимым кофе, программным обеспечением, оказывающих образовательные услуги). Интерес вызывали работы Института по прогнозированию социально-экономического развития России методом сценариев, по экономико-математическому моделированию развития малых предприятий и созданию современных систем информационной поддержки принятия решений для таких организаций.

Институт ведет фундаментальные исследования в области высоких статистических технологий и эконометрики, в частности, в рамках НИЧ МГТУ им. Н.Э. Баумана и Российского фонда фундаментальных исследований. Информация об Институте представлена на сайте в ИНТЕРНЕТе (http://antorlov.nm.ru, зеркала http://antorlov.euro.ru, http://www.newtech.ru/~orlov), который в 2000 г. посетили более 10000 пользователей. Институтом издается еженедельная компьютерная газета «Эконометрика» (около 1000 подписчиков). Архив выпусков газеты "Эконометрика" можно рассматривать как хрестоматию по различным разделам эконометрики, а также по высоким статистическим технологиям.

**Зачем нужны высокие статистические технологии, разве недостаточно обычных статистических методов?** Это вполне естественный вопрос. Мы считаем и доказываем своими теоретическими и прикладными работами, что совершенно недостаточно. Так, многие данные в информационных системах имеют нечисловой характер, например, являются словами или принимают значения из конечных множеств. Нечисловой характер имеют и упорядочения, которые дают эксперты или менеджеры, например, выбирая главную цель, следующую по важности и т.д. Значит, нужна статистика нечисловых данных. Она построена (см. главу 8). Далее, многие величины известны не абсолютно точно, а с некоторой погрешностью - от и до. Другими словами, исходные данные - не числа, а интервалы. Нужна статистика интервальных данных. Она развита (см. главу 9). В монографии [48] по контроллингу на с.138 хорошо сказано: "Нечеткая логика - мощный элегантный инструмент современной науки, который на Западе (и на Востоке - в Японии, Китае - А.О.) можно встретить в десятках изделий - от бытовых видеокамер до систем управления сооружениями, - у нас до самого последнего времени был практически неизвестен". Напомним, первая монография российского автора по теории нечеткости была выпущена в 1980 г. [49]. Ни статистики нечисловых данных, ни статистики интервальных данных, ни статистики нечетких данных нет и не могло быть в классической статистике. Все это - высокие статистические технологии. Они разработаны за последние 10-30-50 лет. А обычные вузовские курсы по общей теории статистики и по математической статистике разбирают научные результаты, полученные в первой половине ХХ века.

Важная часть эконометрики - применение высоких статистических технологий к анализу конкретных экономических данных, что зачастую требует дополнительной теоретической работы по доработке статистических технологий применительно к конкретной ситуации. Большое значение имеют конкретные эконометрические модели, например, модели экспертных оценок (глава 12) или экономики качества (глава 13). И конечно, такие конкретные применения, как расчет и прогнозирование индекса инфляции (глава 7).. Сейчас уже многим ясно, что годовой бухгалтерский баланс предприятия может быть использован для оценки его финансово-хозяйственной деятельности только с привлечением данных об инфляции.

Термин "эконометрика" пока мало известен в России. А между тем в мировой науке эконометрика занимает достойное место. Напомним, что Нобелевские премии по экономике получили эконометрики Ян Тильберген, Рагнар Фриш, Лоуренс Клейн, Трюгве Хаавельмо. В 2000 г. к ним добавились еще двое Джеймс Хекман и Дэниель Мак-Фадден. Выпускается ряд научных журналов, полностью посвященных эконометрике, в том числе: Journal of Econometrics (Швеция), Econometric Reviews (США), Econometrica (США), Sankhya (Indian Journal of Statistics. Ser.D. Quantitative Economics. Индия), Publications Econometriques (Франция).

Применение эконометрики дает заметный экономический эффект. Например, в США - не менее 20 миллиардов долларов ежегодно только в области статистического контроля качества. А что у нас? Повторим, что в секции "Математические методы исследования" журнала "Заводская лаборатория" за последние 40 лет напечатано более 1000 статей по высоким статистическим технологиям и их применениям. Однако в нашей стране по ряду причин эконометрика не была сформирована как самостоятельное направление научной и практической деятельности, в отличие, например, от Польши, не говоря уже об англосаксонских странах. В результате специалистов - эконометриков у нас на порядок меньше, чем в США и Великобритании (Американская статистическая ассоциация включает более 20000 членов).

**Преподавание высоких статистических технологий и эконометрики.** Приходится с сожалением констатировать, что в России практически отсутствует подготовка специалистов по высоким статистическим технологиям. В курсах по теории вероятностей и математической статистике обычно даются лишь классические основы этих дисциплин, разработанные в первой половине ХХ в., а преподаватели свою научную деятельность предпочитают посвящать доказательству теорем, имеющих лишь внутриматематический интерес, а не высоким статистическим технологиям.

В настоящее время появилась надежда на эконометрику. В России начинают развертываться эконометрические исследования и преподавание эконометрики, в том числе не только Институтом высоких статистических технологий и эконометрики. Преподавание этой дисциплины ведется в Московском государственном университете экономики, статистики и информатики (МЭСИ), на экономическом факультете МГУ им. М.В. Ломоносова, в Высшей школе экономики и еще в нескольких экономических учебных заведениях. Среди технических вузов факультет "Инженерный бизнес и менеджмент" МГТУ им. Н.Э.Баумана имеет в настоящее время приоритет в преподавания эконометрики.

Мы полагаем, что экономисты, менеджеры и инженеры, прежде всего специалисты по контроллингу [48], должны быть вооружены современными средствами информационной поддержки, в том числе высокими статистическими технологиями и эконометрикой. Очевидно, преподавание должно идти впереди практического применения. Ведь как применять то, чего не знаешь?

Один раз - в 1990-1992 гг. мы уже обожглись на недооценке необходимости предварительной подготовки тех, для кого предназначены современные компьютерные средства. Наш коллектив (Всесоюзный центр статистических методов и информатики Центрального правления Всесоюзного экономического общества)разработал систему диалоговых программных систем обеспечения качества продукции. Их созданием руководили ведущие специалисты страны. Но распространение программных продуктов шло на 1-2 порядка медленнее, чем ожидалось. Причина стала ясна не сразу. Как оказалось, работники предприятий просто не понимали возможностей разработанных систем, не знали, какие задачи можно решать с их помощью, какой экономический эффект они дадут. А не понимали и не знали потому, что в вузах никто их не учил статистическим методам управления качеством. Без такого систематического обучения нельзя обойтись - сложные концепции "на пальцах" за пять минут не объяснишь.

Есть и противоположный пример - положительный. В середине 1980-х годов в советской средней школе ввели новый предмет "Информатика". И сейчас молодое поколение превосходно владеет компьютерами, мгновенно осваивая быстро появляющиеся новинки, и этим заметно отличается от тех, кому за 30-40 лет. Если бы удалось ввести в средней школе курс вероятности и статистики - а такой курс есть в Японии и США, Швейцарии, Кении и Ботсване, почти во всех странах (см. подготовленный ЮНЕСКО сборник докладов [50]) - то ситуация могла бы быть резко улучшена. Надо, конечно, добиться, чтобы такой курс был построен на высоких статистических технологиях, а не на низких. Другими словами, он должен отражать современные достижения, а не концепции пятидесятилетней или столетней давности.

Необходимо активизировать деятельность Российской ассоциации статистических методов. Но не стоит ограничиваться только внутренними проблемами сообщества специалистов по статистическим методам. Например, в созданном в России профессиональном экономическом обществе - Ассоциации контроллеров России - необходимо, на наш взгляд, выделить направление, посвященное применению высоких статистических технологий и эконометрики в контроллинге, а также учесть необходимость обучения основам этого направления при формировании мощной образовательной базы контроллинга.

 **Литература**

1. Орлов А.И. Что дает прикладная статистика народному хозяйству?/ Вестник статистики. 1986. № 8. С.52 - 56
2. Комаров Д;М., Орлов А.И. Роль методологических исследований в разработке методоориентированных экспертных систем (на примере оптимизационных и статистических методов) - В сб.: Вопросы применения экспертных систем. - Минск: Центросистем, 1988. С.151-160.
3. Ленин В.И. Развитие капитализма в России. Процесс образования внутреннего рынка для крупной промышленности. - М.: Политиздат, 1986. - XII+610 с.
4. Гнеденко Б.В. Курс теории вероятностей: Учебник. - Изд. 6-е, переработанное и дополненное. - М.: Наука, Гл. ред. физ. - мат. лит., 1988. - 448 с.
5. Бернштейн С.Н. Современное состояние теории вероятностей и ее приложений. - В сб.: Труды Всероссийского съезда математиков в Москве 27 апреля - 4 мая 1927 г. - М.-Л.: ГИЗ, 1928. С.50-63.
6. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов. / Заводская лаборатория. 1992. Т.58. № 1. С.67-74.
7. Орлов А.И. О перестройке статистической науки и её применений. / Вестник статистики. 1990. № 1. С.65 - 71.
8. Кендалл М., Стьюарт А. Теория распределений. - М.: Наука, 1966. - 566 с.
9. Кендалл М., Стьюарт А. Статистические выводы и связи. - М.: Наука, 1973. - 899 с.
10. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976. - 736 с.
11. Налимов В.В., Мульченко З.М. Наукометрия. Изучение развития науки как информационного процесса. - М.: Наука, 1969. - 192 с.
12. ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения. - М.: Изд-во стандартов. 1984. - 53 с.
13. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1965 (1-е изд.), 1968 (2-е изд.), 1983 (3-е изд.).
14. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука,1979. - 296 с.
15. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания: Статистическая обработка неоднородных совокупностей. - М;: Статистика, 1980. - 208 с.
16. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. - М.: Финансы и статистика, 1988. - 263 с.
17. Суппес П., Зинес Дж. Основы теории измерений. - В сб.: Психологические измерения. -М: Мир,1967. С. 9-110.
18. Пфанцагль И. Теория измерений. - М.: Мир, 1976. - 166 с.
19. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. - М.: Мир, 1976. - 168 с.
20. Дэвид Г. Метод парных сравнений. - М.: Статистика, 1978. - 144 с.
21. Матерон Ж. Случайные множества и интегральная геометрия. - М.: Мир, 1978. - 318 с.
22. Терехина А.Ю. Анализ данных методами многомерного шкалирования. - М.: Наука, 1986. - 168 с.
23. Перекрест В.Т. Нелинейный типологический анализ социально-экономической информации: Математические и вычислительные методы. - Л.: Наука, 1983. - 176 с.
24. Кемени Дж., Снелл Дж. Кибернетическое моделирование: Некоторые приложения. - М.: Советское радио, 1972. - 192 с.
25. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1981. - 80 с.
26. Литвак Б.Г. Экспертная информация: Методы получения и анализа. - М.: Радио и связь, 1982. - 184 с.
27. Орлов А.И. Статистика объектов нечисловой природы и экспертные оценки. - В сб.: Экспертные оценки. Вопросы кибернетики. Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1979. С.17-33.
28. Анализ нечисловой информации в социологических исследованиях. / Под ред. В.Г. Андреенкова, А.И.Орлова, Ю.Н. Толстовой. - М.: Наука, 1985. - 220 с.
29. Орлов А.И. Асимптотическое поведение статистик интегрального типа. / Доклады АН СССР. 1974. Т.219. № 4. С.808-811.
30. Орлов А.И. Асимптотическое поведение статистик интегрального типа. - В сб.: Вероятностные процессы и их приложения. Межвузовский сборник. - М.: МИЭМ, 1989. С.118-123.
31. Горский В.Г. Современные статистические методы обработки и планирования экспериментов в химической технологии. - В сб.: Инженерно-химическая наука для передовых технологий. Международная школа повышения квалификации Труды третьей сессии. 26-30 мая 1997, Казань, Россия / Под ред. В.А. Махлина. - М.: Научно-исследовательский физико-химический институт им. Карпова, 1997. С.261-293.
32. Плошко Б.Г., Елисеева И.И. История статистики: Учебное пособие. - М.: Финансы и статистика. 1990. - 295 с.
33. Эльясберг П.Е. Измерительная информация. Сколько ее нужно, как ее обрабатывать? - М.: Наука, 1983. - 208 с.
34. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.
35. Орлов А.И., Орловский И.В. О поправках на группировку. - В сб.: Прикладной многомерный статистический анализ. - М.: Наука, 1978. - С.339-342.
36. Орлов А.И. Поправка на группировку для коэффициента корреляции. / Экономика и математические методы. - 1980. - Т.XVI. - №4. - С.800-801.
37. Орлов А.И. Методы оценки близости допредельных и предельных распределений статистик. / Заводская лаборатория. - 1998. - Т.64. - № 5. - С.64-67.
38. Феллер В. Введение в теорию вероятностей и ее приложения. Т.2. - М.: Мир, 1984. - 751 с.
39. Боровков А.А. Теория вероятностей. - М.: Наука, 1976. - 352 с.
40. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.
41. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. - М.: Наука, 1979. - 528 с.
42. Орлов А.И. О нецелесообразности использования итеративных процедур нахождения оценок максимального правдоподобия. / "Заводская лаборатория", 1986. Т.52. No.5. С.67-69.
43. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
44. Никитина Е.П., Фрейдлина В.Д., Ярхо А.В. Коллекция определений термина "статистика" / Межфакультетская лаборатория статистических методов. Вып.37. - М.: Изд-во Московского государственного университета им. М.В. Ломоносова, 1972. - 46 с.
45. Орлов А.И. Проблема множественных проверок статистических гипотез. / Заводская лаборатория. 1996. Т.62. No.5. С.51-54.
46. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат. / Заводская лаборатория. - 1985. - Т.51. - No.1. - С.60-62.
47. Орлов А.И. Сертификация и статистические методы. / Заводская лаборатория. 1997. Т.63. No.З. С.55-62.
48. Контроллинг в бизнесе. Методологические и практические основы построения контроллинга в организациях / А.М. Карминский, Н.И. Оленев, А.Г. Примак, С.Г.Фалько. - М.: Финансы и статистика, 1998. - 256 с.
49. Орлов А. И. Задачи оптимизации и нечеткие переменные. - М.: Знание, 1980.- 64 с.
50. The teaching of statistics / Studies in mathematics education. Vol.7. - Paris, UNESCO, 1989. - 258 pp.