

$$R = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\overline{b_x^2} \overline{b_y^2}}}; E(x) = \sum x p$$



$$\chi^2 = \sum \frac{(m - m')^2}{m'}$$

И. И. Елисеева, В. О. Рукавишников

ЛОГИКА ПРИКЛАДНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

★ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА ДЛЯ ЭКОНОМИСТОВ ★

И. И. Елисеева, В. О. Рукавишников

ЛОГИКА ПРИКЛАДНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

МОСКВА «ФИНАНСЫ И СТАТИСТИКА» 1982

ББК 22.172

Е51

Редколлегия серии
«Математическая статистика
для экономистов»:

А. Я. Боярский, Н. К. Дружинин, А. М. Дубров, Ю. Н. Тюрин

Ирина Ильинична Елисеева
Владимир Олегович Рукавишников

ЛОГИКА ПРИКЛАДНОГО
СТАТИСТИЧЕСКОГО АНАЛИЗА

Рецензенты *В. П. Трофимов, Г. Г. Пирогов*
Зав. редакцией *Р. А. Казьмина*
Редактор *К. С. Исаева*
Мл. редакторы *Н. Е. Константинова, Е. С. Уварова, В. Л. Долгова*
Корректоры *Г. А. Башарина, Л. Г. Захарко*
Техн. редакторы *К. К. Букалова, Г. А. Полякова*
Худож. редактор *М. К. Гуров*

ИБ № 950

Сдано в набор 14.06.82. Подписано в печать 18.10.82. А12897
Формат 84×108¹/₃₂. Бум. кн.-журн. Гарнитура «Литературная».
Печать высокая. П. л. 6 Усл. п. л. 10,08 Усл. кр.-отт. 10,29 Уч.-изд. л. 10,00
Тираж 7000 экз. Заказ 1010. Цена 1 р. 30 к.
Издательство «Финансы и статистика», Москва, ул Чернышевского, 7.
Московская типография № 4 Союзполиграфпрома
при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли
129041 г. Москва, Б. Переяславская, 46

Елисеева И. И., Рукавишников В. О.

Е51 **Логика прикладного статистического анализа. —**
М.: Финансы и статистика, 1982. — 192 с., ил. —
(Мат. статистика для экономистов).
1 р. 30 к.

Показаны логические принципы статистического описания сложных объектов и процессов. Значительное внимание уделено построению модельных схем, формализующих социально-экономические явления. Описывается применение многомерных математико-статистических моделей для задач экономического анализа на предприятиях и изучения структуры социально-экономических объектов.
Для статистиков, экономистов, социологов, демографов.

Е 1702060000—181 22—82
010(01)—82

ББК 22.172
517.8

© Издательство «Финансы и статистика», 1982

ПРЕДИСЛОВИЕ

Одной из характерных черт современного состояния социально-экономических исследований является широкое применение в них методов математической статистики. При этом важен учет как природы изучаемого явления, так и математико-статистических особенностей используемого метода. Недооценка той или иной стороны этого положения, как правило, приводит к неудачам.

Несколько слов о структуре книги.

Логика статистического метода определяется спецификой объекта исследования, природой изучаемых связей, целями, задачами и гипотезами исследования. Этим вопросам посвящена первая глава книги.

Во второй главе обсуждаются методологические и прикладные аспекты проблемы описания многомерных статистических объектов. Рассмотрение этих вопросов ведется на примерах регрессионного анализа, факторного анализа и анализа главных компонент — сегодня едва ли не наиболее распространенных на практике методов анализа статистических данных. Подробно обсуждаются проблемы сравнения результатов факторного анализа, полученных либо на различных массивах данных, либо различными методами, тесно связанные с более общими проблемами обобщения результатов статистических исследований. При рассмотрении различных аспектов применения регрессионного анализа выделяются вопросы отбора существенных переменных, сравнения моделей и построения регрессионных моделей в условиях коррелированности входных переменных. Обсуждение последнего вопроса связывается с использованием в регрессии метода главных компонент.

В третьей главе всесторонне рассматривается сравнительно малоизвестный отечественным специалистам подход к построению объяснительных статистических моделей в форме систем структурных уравнений множественной регрессии. В тех случаях, когда предполагается причинная интерпретация таких систем, они часто называются

причинными моделями. Вопросы методологии и методики разработки структурных и причинных моделей в нашей литературе до сих пор систематически изложены не были. Эти вопросы рассматриваются в книге во взаимосвязи с общими проблемами методологии статистического объяснения, эмпирического обоснования гипотез и построения теории. Значительное внимание, как и во второй главе, уделено здесь вопросам сравнения моделей. Кратко излагается схема одного из возможных подходов к разработке причинных моделей систем с целенаправленным поведением на основе нечисловой информации, который может быть полезен при построении прогнозных моделей развития сложных социально-экономических систем.

Ввиду распространенности категоризованных переменных в описаниях социально-экономических объектов, четвертая глава книги посвящена методам анализа таблиц сопряженности. Для анализа таких таблиц в последние годы был создан ряд новых методов построения моделей, описывающих и объясняющих связи переменных. В главе дается обзор существующих подходов, среди которых выделяется метод логлинейного анализа. Показываются возможности еще одного приема анализа таблиц сопряженности, пока не получившего должного распространения в исследовательской практике, но заслуживающего внимания: использование канонических меток — ортогонального преобразования переменных — и канонических корреляций, позволяющих получить хорошую аппроксимацию структуры исходных данных, расклассифицированных в таблице сопряженности.

Таким образом, в этой книге рассматриваются в основном вопросы методологии, методики и организации прикладного статистического исследования с использованием многомерных методов изучения взаимосвязей. Такое сужение проблематики логики прикладного статистического анализа было обусловлено невозможностью в рамках книги небольшого объема подробно и всесторонне осветить все многообразие аспектов данной темы.

Доступность ЭВМ и внешняя простота решения задач анализа данных с помощью стандартных программных средств привели к тому, что зачастую на практике многие из многомерных методов используются неправильно, что в свою очередь порождает ошибочные решения и выводы научно-практического характера, а тем самым дискредитируется и сама идея активного применения современных

методов прикладной статистики в комплексных системных исследованиях социально-экономических явлений. Поэтому в своей работе мы стремились показать не столько «технологии» разработки статистических моделей, сколько трудности, о которых должен знать каждый исследователь, решивший заняться этим увлекательным делом. А трудности, как показывает опыт, даже при разработке простых моделей заставляют по-иному взглянуть на многие этапы исследования, внешне, казалось бы, связанные, скорее, с теоретической проработкой проблемы, нежели со статистическим анализом. Так, может обнаружиться, что неточности измерительных инструментов излишне велики или вообще едва ли могут быть оценены, что переменные сильно коррелированы между собой или что нет достаточного числа переменных — непосредственно измеряемых индикаторов искомого скрытого качества, что градации шкал качественных признаков, используемых в таблицах сопряженности, описаны не строго и этим осложняют изучение взаимосвязей признаков, и т. д. Построение простой модели превращается в самостоятельное, сложное и комплексное исследование, открывающее новые аспекты проблемы, и в конечном итоге — новые горизонты познания сущности изучаемого явления.

Рукавишниковым В. О. написаны § 1.1, главы 2 и 3, § 4.4, Елисеевой И. И. — § 1.2, 4.1, 4.2, 4.3.

Глава 1. ОБЩИЕ ВОПРОСЫ СТАТИСТИЧЕСКОГО АНАЛИЗА В ПРИКЛАДНЫХ СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

1.1. СТАТИСТИЧЕСКИЙ МЕТОД ИССЛЕДОВАНИЯ

Статистический метод применяется в различных областях науки и практики. Иногда для краткости его обозначают просто термином «статистика». Следует, однако, различать употребление этого термина для обозначения статистики как метода и статистики как науки, ибо отсутствие ясного осознания этих различий в условиях исторически сложившегося неоднозначного его использования служит основой для дискуссии о содержании и предмете статистики¹.

Статистический метод — это одно из средств познания объективной действительности, выступающее в качестве опосредствующего звена в познавательном процессе между исследователем и изучаемыми им объектами. Любая конкретная наука, использующая статистический метод, не лишается при этом ни своего предмета, ни специфических приемов исследования. Метод, о котором идет речь, дает возможность математически выразить найденные в процессе исследования конкретные зависимости и отношения, а в целом ряде случаев помогает находить их.

Таким образом, статистический метод в качестве орудия познания выступает прежде всего как способ описания изучаемых конкретной наукой явлений. Способ описания предполагает и систему объяснения исследуемой области явлений. Статистический метод как способ объяснения требует, чтобы соответствующая область науки или практики располагала определенным концептуальным аппаратом (сис-

¹См., например, материалы дискуссии на страницах журнала «Вестник статистики» в 1975—1978 гг., а также [28, с. 272—295].

темой понятий), с использованием которого производятся систематизация, анализ и обработка статистических данных. При этом статистическими данными называются сведения о некотором числе объектов в какой-либо более или менее обширной совокупности, обладающей теми или иными признаками, способными изменять свое количественное значение или качественное состояние при переходе от одного объекта — единицы совокупности к другому.

В статистике термины «признак» и «переменная», используемые часто как синонимы, определяются как величина, которая может принимать любое значение из некоторого заданного множества значений. Система признаков-переменных и область определения каждого из них задаются исследователем на основе теории изучаемого явления. Различают также переменные и случайные переменные (случайные величины). Случайная переменная — это величина, которая в совокупности наблюдений, выполненных в повторяющемся комплексе условий, может принимать различные значения из определенного множества, причем заранее неизвестно, какие именно. Свойства случайной величины будут полностью определены, если установлен закон (функция) ее распределения, устанавливающий связь между возможными значениями случайной величины и соответствующими им вероятностями или относительными частотами их наблюдения в заданной ситуации. Наконец, переменные могут быть качественными (нечисловыми) и количественными (дискретными или непрерывными).

Статистический метод обычно определяют как комплекс приемов и принципов, согласно которым производятся сбор, анализ, сравнение и интерпретация статистических данных для получения научных и практических выводов. Математическое содержание этого комплекса образует математическая статистика — самостоятельная математическая наука, отрасль прикладной математики. В то же время статистический метод исследования имеет и нематематическую сторону, к которой следует отнести вопросы методологии, методики и логики организации и проведения статистического исследования, включая вопросы планирования выборочных обследований, организации обработки статистических данных различного типа средствами современной электронно-вычислительной техники, формы представления результатов расчетов и т. п., к нематематической стороне статистического метода относятся также методологические и методические проблемы постановки задачи содержа-

тельной интерпретации итогов математико-статистического анализа, обеспечения достоверности вывода. Таким образом, статистический метод сегодня — это определенная стратегия научного поиска, которая использует математический аппарат и математические концепции в рамках системного подхода к изучаемым явлениям или объектам.

Конечная цель каждого конкретного (или, как принято говорить, прикладного) экономического, социологического и любого другого исследования, базирующегося на статистических данных, — установление статистической закономерности, которая при определенных условиях может быть интерпретирована как закон. Примерами таких законов в сфере технических наук являются законы Бойля — Мариотта и Гей-Люссака, определяющие свойства газов, закон Ома в электротехнике и т. д. Подобные законы не универсальны, поскольку имеются причины, ограничивающие сферу их действия. Они феноменологичны, поскольку справедливы на уровне совокупностей, рассматриваемых без учета структуры и особенностей поведения отдельных единиц, из которых состоят эти совокупности (так, например, названные выше законы сформулированы без учета молекулярной или атомной структуры вещества). Они статистичны, поскольку описывают закономерности, отчетливо проявляющиеся только при осреднении большого числа экспериментальных данных. Математическое описание каждой из названных выше эмпирических закономерностей является в то же время ее моделью, справедливой лишь при условии выполнения определенных допущений и предположений.

Термин «модель», использованный нами, как известно, чрезвычайно многозначен. В контексте рассматриваемых в этой книге вопросов мы будем понимать под моделью любое формальное описание связей между определенными символами.

Известно, что с развитием научного познания явлений определенной предметной области теории феноменологического типа уступают место теориям, в которых раскрывается конкретный механизм происходящих процессов. Соответственно изменяются цели и задачи статистического метода исследования. В первом случае целью статистического исследования является описание внешнего проявления взаимосвязей измеряемых свойств изучаемых объектов, адекватно и, как правило, однозначно, отображаемых в системе признаков-переменных, входящих в статистическое

описание. Во втором случае исследуются гипотезы о внутреннем механизме взаимосвязей, т. е. изучаются так называемые структурные модели. При этом в моделях могут присутствовать как непосредственно наблюдаемые переменные, так и ненаблюдаемые характеристики, адекватное отображение которых в форме математической модели, равно как и указание способа их косвенного измерения, представляет порой сложную теоретическую и методическую задачу. Ненаблюдаемые непосредственно свойства объектов — это характеристики, само существование которых оправдано лишь в контексте определенной теории.

Принятая математико-статистическая модель в этом, как и в любом другом случае, в свою очередь во многом определяет математическую форму теории и границы применимости выводов.

Изложенное выше представление о статистическом методе основано в основном на опыте его применения в технических, естественных, медико-биологических, социологических и экономических исследованиях¹. Выявление и осмысление общих черт статистического метода, применяемого к объектам разной природы, привело к появлению термина «прикладная статистика» как наименования сложившейся специфической области знания, прикладной дисциплины, исследующей вопросы применения математической статистики к реальным данным. Поэтому нельзя не упомянуть

¹Все сказанное ранее относилось в первую очередь к области применения статистических методов для анализа данных наблюдений. Проникновение в структуру данных и составляет цель статистического метода в этом случае. Модели, разрабатываемые исследователями, можно назвать *аналитическими*; выражения, описывающие взаимосвязи зависимых и независимых величин, содержатся в них в явном виде. К этому классу моделей следует отнести модели распределения для описания вероятностных свойств исследуемых признаков (например, с помощью стандартных теоретических статистических распределений из семейства распределений Джонсона или Пирсона), регрессионные модели и многомерные факторные модели, модели планирования эксперимента и анализа временных рядов и т. д. К статистическим (в широком смысле) моделям относят и *имитационные* модели, описанные с помощью последовательности определенных математических операций, таких, как решение систем дифференциальных уравнений с использованием случайных или псевдослучайных чисел, неоднократное преобразование матриц переходных вероятностей в цепях Маркова и т. д. Особенности статистической методологии для этого класса моделей нами не рассматриваются; читателям, интересующимся этими вопросами, рекомендуем обратиться к литературе по методам системного моделирования и анализа стохастических процессов.

о том, что термины «статистический метод», «статистическое исследование» широко используют также в рамках самостоятельной области знания — социально-экономической статистики, где в силу сложившейся традиции в их содержание часто вкладывается смысл, несколько отличающийся от представленного выше. В рамках социально-экономической статистики были созданы специфические методология и методы, не имеющие очевидной математико-статистической природы, — индексный анализ, типологические группировки и т. д. В последние годы общепринятый характер приобретает точка зрения о том, что успех в решении стоящих перед статистической практикой и социально-экономической статистикой как наукой задач во многом зависит от рационального сочетания в обработке, обобщении и анализе данных методологии и приемов как собственно социально-экономической статистики, так и прикладной статистики. Дело в том, что пропагандировавшееся длительное время мнение о существовании принципиальной разницы между математической статистикой, приложениями статистического метода в сфере естественных и технических наук и социально-экономической статистикой привело к тому, что в практической деятельности отечественных статистиков и экономистов мало применялись математико-статистические методы исследования.

Некоторые представители традиционной школы социально-экономической статистики критически относятся к возможности использования математики в своей научной или практической работе, утверждая, что математика имеет дело с абстрактными величинами, а статистика — с количествами, являющимися отображениями свойств совокупности объектов определенного качества. Действительно, предметом математики являются не объекты или явления реального мира, а абстрактные объекты — формальные, знаковые модели. Поэтому вполне обоснована характеристика математики как своеобразного формального языка. Следовательно, решение вопроса о принципиальной приложимости или неприложимости математики в той или иной сфере науки сводится к выяснению возможности выразить на математическом языке систему понятий и представлений данной предметной области. Это положение в полной мере применимо и к рассмотрению возможностей математической статистики в изучении социально-экономических явлений. Поэтому вопрос о взаимоотношении математической и социально-экономической статистики следует ставить как задачу рас-

ширения сферы применения и содержания языка прикладной статистики.

Многие проблемы на этом пути еще остаются нерешенными. К ним относится проблема модификации или адаптации существующих математико-статистических методов, методологии и методики их применения с учетом свойств социально-экономических данных, так как условия и предположения, использованные математиками при разработке этих методов и выполняющиеся с разумной строгостью в других приложениях, для социально-экономических данных чаще нарушаются, чем выполняются. Отмеченная ограниченная возможность переноса математических приемов исследования порождает методологические ошибки двоякого рода: отрицание приложимости сложившегося математико-статистического аппарата в социально-экономических исследованиях и представление, будто один и тот же метод без каких-либо модификаций аппарата или методики применения способен равно удовлетворительно описывать явления различной природы. Поэтому наряду с модификацией существующих методов требуется разработка нового математического аппарата, методологии и методики анализа, ориентированных на социально-экономическую информацию, и прежде всего на нечисловые данные. Непосредственно к названным выше примыкают проблемы интерпретации и устойчивости статистических выводов в условиях возможных колебаний исходных данных и нарушения предпосылок, использованных при разработке моделей и т. д.

Методы статистического анализа и моделирования (в том числе и рассмотренные нами в этой книге) образуют теоретическую основу организации всего процесса прикладного статистического исследования. Вместе с тем эта теоретическая основа является лишь частью методологии статистического исследования. Вопросы организации и управления процессом сбора и анализа эмпирических данных также входят в статистическую методологию и в значительной степени определяют всю логику прикладного статистического анализа. Можно выделить два подхода к организации процесса анализа данных, сложившиеся на практике в зависимости от развитости представлений исследователей о природе изучаемых ими явлений или объектов и, следовательно, от уровня знаний о структуре анализируемых данных, о взаимосвязях переменных и т. п.:

1. Структура данных заранее известна; соответственно заранее определяются методы их описания.

2. Структура данных точно неизвестна; соответственно заранее не определен набор и последовательность применения статистических приемов анализа, условия перехода от одного метода (уровня анализа) к другому. Формирование «траектории анализа» происходит под управлением исследователя. Повышение степени автоматизации процесса анализа достигается за счет организации на ЭВМ диалогового (интерактивного) режима обработки с переходом с одного уровня анализа на другой в зависимости от результатов расчетов и степени достижения поставленных целей.

Важно подчеркнуть, что весь процесс статистического анализа подчиняется и регулируется исходной формулировкой цели и задач исследования. Хотя до сих пор мы не уделяли особого внимания вопросу об использовании ЭВМ, применение ЭВМ в той или иной форме подразумевалось во всех случаях. Мы сознательно не акцентировали внимания читателя на этой стороне организации прикладного статистического исследования, чтобы не возникло впечатления, что современные многомерные статистические методы всегда требуют применения больших ЭВМ. Тем не менее есть очень веские аргументы в пользу использования ЭВМ в исследованиях многомерных статистических объектов: а) большая скорость вычислений; б) возможность оперирования очень большими массивами данных и неоднократного обращения к ним; в) возможности применения нескольких методов анализа одновременно (или практически одновременно) и гибкой параллельной разработки нескольких альтернативных моделей; г) наличие в современных ВЦ пакетов стандартных программ статистических методов, в том числе и тех, о которых идет речь в этой книге.

Все эти достоинства современных средств вычислительной техники в сочетании с богатым арсеналом методов статистического анализа данных различной природы делают статистический метод чрезвычайно мощным орудием познания в умелых руках исследователей.

1.2. ОСОБЕННОСТИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ОБЪЕКТОВ И ЗАДАЧИ ПРИКЛАДНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

Совокупность социально-экономических явлений — это всегда реализация определенной закономерности в конкретных условиях места и времени, представленная целым рядом единичных явлений.

Единица любой совокупности социально-экономических явлений обладает множеством различных свойств или, иначе говоря, признаков. Признаки, непосредственно отражающие качество явления, позволяющие уяснить его сущность — так называемые существенные признаки, — образуют систему, в которой один признак взаимосвязан с другими. Это порождает многомерность описания социально-экономических объектов. Вариация таких признаков всегда ограничена единым качеством явления. Скажем, себестоимость единицы одного и того же вида продукции по совокупности однородных предприятий не может принимать сколь угодно большие и сколь угодно малые значения, ее изменение будет лежать в довольно узких пределах, определяемых примерно одинаковой трудоемкостью производства продукции, уровнем развития техники, однородным характером организации труда и т. д. Очевидна связанность себестоимости с другими существенными признаками: уровнем производительности труда на предприятии, фондооснащенностью и фондоотдачей, квалификацией работников и др. Изменение одной из этих характеристик необходимо приводит к изменению других. Измерение отдельных признаков позволяет построить интегральную характеристику явления, оценка связи признаков — выявить связь различных сторон явления.

Вопрос об определении единицы совокупности тесно связан с вопросом об уровне исследования природы явления. Можно изучать, например, производительность труда на уровне отрасли, отдельного предприятия, цеха, бригады, наконец, на уровне отдельного работника. В каждом случае единица совокупности будет особой. Уровень исследования определяет круг выдвигаемых задач, на каждом уровне в качестве единицы наблюдения выступает то, в чем проявляется изучаемая закономерность, где можно проследить ее действие.

Особенности условий формирования социально-экономических явлений, неодинаковость темпов их развития определяют наличие в реальных совокупностях особых типов элементов (классов явлений). Однако даже при достаточно полном описании типов не всегда можно точно решить, к какому из них принадлежит объект. По-видимому, в достаточно общем виде совокупность социально-экономических объектов можно представить как статистический класс или коллектив, включающий ядро и окружающие его явления. Ядро — концентрированное выражение специфи-

ческих свойств класса. Окружающие ядро явления — это явления переходного качества, принадлежность которых к данному классу может быть гарантирована лишь с определенной вероятностью. Такие явления образуют полосу размыва между классами.

Соотношение между ядром и окружающими его явлениями в разных классах будет, конечно, различным. Это зависит от устойчивости класса, длительности его существования, взаимодействия с другими классами той же совокупности, с другими совокупностями.

Социально-экономические явления выделяются своей динамичностью. Причина этого в особой природе социально-экономических явлений и процессов, в центре которых человек — сложный социальный организм с субъективными установками, способностью активно воздействовать на окружающий мир. Изменение социальных явлений имеет направленность (тенденцию) и происходит достаточно интенсивно.

Сравним, например, динамику производительности труда и динамику какого-либо биологического вида. Эволюция вида совершается тысячи лет, тогда как качественные сдвиги в производительности общественного труда могут происходить неоднократно на глазах одного поколения.

Чтобы понять логику социально-экономических приложений статистического метода, полезно представить совокупность социально-экономических объектов в виде упорядоченного набора данных с параметрами N, D, T , где N — численность единиц совокупности ($i = 1, \dots, N$); D — число признаков, фиксируемых по каждой единице (j -й признак, принадлежащий i -й единице, будем обозначать x_{ij}); T — моменты времени, в которые измеряются признаки x_{ijt} . Если совокупность изучается в статике, то все данные можно представить в виде таблицы, в которой строка несет всю информацию о единичном объекте, столбец — о признаке. Размеры такой таблицы часто оказываются достаточно большими: число строк может колебаться от нескольких десятков до сотен тысяч, а число столбцов — от одного-двух до нескольких сотен. Основные статистические методы обработки социально-экономической информации исходят именно из этой формы организации представления исходных данных.

Такое представление данных приводит по меньшей мере к двум типам задач: выявлению сходства между объектами — строками таблицы (задача классификации) и анализу

взаимосвязей признаков. В конечном итоге целью в обоих случаях является свертывание информации. Возможность короткого описания исходного массива обуславливает некая скрытая закономерность, и само получение свертки данных означает познание этой объективной закономерности.

Свертка столбцов таблицы, или редукция признакового пространства, имеет объективную основу — связанность признаков. Это позволяет говорить о необходимости выявления скрытых факторов, проявляющихся внешне в устойчивых сочетаниях значений непосредственно измеряемых признаков (подробнее речь об этом пойдет в гл. 2).

Информативность статистических данных многократно возрастает, если минимизация описания включает как выделение классов, так и редукцию их характеристик. При решении задачи классификации важную роль играет нахождение небольшого числа наиболее информативных признаков описания. Это оказывается возможным благодаря взаимосвязи признаков и дублированию ими информации, неинформативности тех признаков, которые слабо варьируют и т. д. Сокращение числа переменных и повышение их информативности можно осуществить путем агрегирования ряда признаков. Тот или иной вариант конкретизации постановки этой задачи, определяющий выбор меры информативности и класса допустимых преобразований признаков, приводит к тому или иному конкретному методу снижения размерности (факторному анализу, экстремальной группировке параметров и т. д.).

Задача измерения связей тесно связана с задачей изучения динамики явлений. Анализ рядов динамики предполагает выделение качественно отличных периодов и их обобщенную характеристику; выявление тенденции (тренда); разложение факторов на основные (эволюционные), периодические и случайные; прогнозирование. Анализ временных рядов — чрезвычайно сложная проблема, характеризующаяся специфическими постановкой задачи исследования и используемыми методами — в настоящей работе не рассматривается.

Разграничивая задачи статистического исследования, можно разграничить и средства, используемые для их решения, однако это разделение весьма относительно, как относительно и грани переходов от одной задачи к другой.

Исходным условием и основной гарантией достоверности и сопоставимости первичной информации является правильное описание изучаемой совокупности, что в свою очередь

предполагает определение образующих ее единиц, продолжительности периода времени, достаточного для выявления изучаемых процессов, и, наконец, местоположения единиц совокупности. В ряде случаев время получения данных также может влиять на их достоверность.

Определение объекта исследования не так просто и очевидно, как это может показаться. Например, при изучении образа жизни населения города встает вопрос — включать или не включать в обследование лиц, живущих в общежитиях, гостиницах; при анализе работы грузового автотранспорта — учитывать ли все те хозяйства, где есть хотя бы один грузовой автомобиль или только крупные автохозяйства и т. д.

Сбор сведений об изучаемой совокупности может производиться с разной степенью полноты охвата единиц. Сплошной охват единиц не является обязательным условием, так как и при несплошном учете можно обеспечить достоверность статистических выводов, соблюдая правила выборочного наблюдения.

Проектирование выборочного наблюдения, а также оценка его результатов предполагают заранее известную величину допустимой погрешности выборочного показателя и соответственно знание того, какая ошибка является недопустимой при замене сплошного учета несплошным. Допустимый размер погрешности устанавливают исходя из целей исследований и особенностей изучаемой закономерности. Довольно часто при этом учитывается эффект действия погрешности на принятие решений. Например, по материалам выборочного учета населения города определяется средний размер жилой площади, приходящейся на 1 человека. Возникает вопрос: какая при этом может быть допущена погрешность? Если предел возможной ошибки выборки (с принятой доверительной вероятностью) получился равным $1,7 \text{ м}^2$, можно ли использовать выборочные данные? Известно, что собранные материалы будут использоваться не просто для характеристики обеспеченности населения жилой площадью, а в связи с изучением семейной структуры. В этом случае нужно иметь представление, когда начинает проявляться эффект повышения жилищной обеспеченности на образование новых семей. Можно полагать, что на уровень брачности влияет изолированность жилища, наличие отдельной комнаты, в переводе на размеры жилой площади можно сказать, что возможность образования новой брачной пары предполагает по крайней мере дополнительно