



БИБЛИОТЕЧКА  
ИНОСТРАННЫХ  
КНИГ ДЛЯ  
ЭКОНОМИСТОВ  
И СТАТИСТИКОВ

А. ЭРЕНБЕРГ

# АНАЛИЗ И ИНТЕРПРЕТАЦИЯ СТАТИСТИЧЕСКИХ ДАННЫХ



**БИБЛИОТЕЧКА  
ИНОСТРАННЫХ КНИГ  
ДЛЯ ЭКОНОМИСТОВ  
И СТАТИСТИКОВ**

# DATA REDUCTION

A. S. C. Ehrenberg

London Graduate School  
of Business Studies

A Wiley-Interscience Publication

John Wiley & Sons  
London·New York·Sydney·Toronto

# АНАЛИЗ И ИНТЕРПРЕТАЦИЯ СТАТИСТИЧЕСКИХ ДАННЫХ

Перевод с английского Б. И. КЛИМЕНКО  
Под редакцией и с предисловием А. А. РЫВКИНА

## БИБЛИОТЕЧКА ИНОСТРАННЫХ КНИГ ДЛЯ ЭКОНОМИСТОВ И СТАТИСТИКОВ

Издательство «Финансы и статистика» выпускает на русском языке серию книг иностранных авторов, рассчитанных на специалистов, нуждающихся в пополнении своих математических и статистических знаний. Задача серии — познакомить советского читателя с методами, применяемыми за рубежом в экономическом анализе и различных хозяйственных расчетах. В серию включаются также работы по общим вопросам статистики.

Вышли из печати книги:

1. М. Броуди. О статистическом рассуждении. 1968.
2. А. Бернштейн. Справочник статистических решений. 1968.
3. У. Дж. Рейхман. Применение статистики. 1969.
4. Х. Крыньский. Математика для экономистов. 1970.
5. С. Дайменд. Мир вероятностей. 1970.
6. А. Хьютсон. Дисперсионный анализ. 1971.
7. С. Лизер. Эконометрические методы и задачи. 1971.
8. Эм. Борель, Р. Дельтейль, Р. Юрон. Вероятности, ошибки. 1972.
9. Статистические методы исследования корреляций в экономике. 1972.
10. Л. Столерю. Равновесие и экономический рост. 1974.
11. Я. Окунь. Факторный анализ. 1974.
12. С. Сирль, У. Госман. Матричная алгебра в экономике. 1974.
13. Е. Гренъ. Статистические игры и их применение. 1975.
14. Д. Тёрнер. Вероятность, статистика и исследование операций. 1976.
15. Э. Кейн. Экономическая статистика и эконометрия. Вып. 1. 1977.
16. Э. Кейн. Экономическая статистика и эконометрия. Вып. 2. 1977.
17. Э. Колкот. Проверка значимости. 1978.
18. Г. Дэвид. Метод парных сравнений. 1978.
19. М. Г. Кенуй. Быстрые статистические вычисления. 1979.
20. Дж. Вайнберг, Дж. Шумекер. Статистика. 1979.
21. Н. Хастингс, Дж. Пикок. Справочник по статистическим распределениям. 1980.
22. А. Гильберт. Как работать с матрицами. 1981.
23. М. Кендэл. Временные ряды. 1981.
24. Ю. Кюн. Описательная и индуктивная статистика. 1981.

Подготавливаются к изданию:

Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике.

Кимбл Г. Как правильно пользоваться статистикой.

### РЕДКОЛЛЕГИЯ СЕРИИ

В. М. Иванова, В. А. Колемаев, Г. Г. Пирогов,  
А. А. Рывкин, Е. М. Четыркин, Р. М. Энгов.

9  $\frac{10805^1-142}{010(01)-81}$  43—81(C) 0702000000 1702060000

<sup>1</sup> Второй индекс 10803.

## ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

При первом знакомстве книга А. Эренберга «Анализ и интерпретация статистических данных» воспринимается как одно из многочисленных пособий и руководств по статистике, излагающих традиционный материал на простых примерах, которые доступны читателю, не обладающему специальной подготовкой. В первых главах автор ведет неторопливую и обстоятельную беседу с читателем, как бы заботясь лишь о естественности изложения и о подготовке базы для следующих, технически более сложных разделов. Даже тот, кому трудно пользоваться методами численного анализа статистических данных из-за предубеждения к математическим выкладкам, обнаружит, что ему предлагают приемы, органично вытекающие из содержательного и целенаправленного изучения конкретной ситуации, почувствует, как исчезает психологический барьер, отделяющий его от мира, говорящего на языке современной статистики. Преодолеть такой барьер можно лишь в процессе активного усвоения материала, постепенно приобретая на практике необходимые навыки. В большинстве других пособий и руководств предполагается, что читатель уже владеет соответствующей ситуацией, и основной упор делается на систематическое изложение теоретического курса. В действительности, однако, подобное предположение далеко не всегда оправдано. Нередко людям свойственно считать себя неспособными к математике из-за того, что им не удалось связать специфику формального мышления с логикой практической деятельности, с задачами, возникающими в их повседневной работе. И, конечно, один из лучших способов привить читателю вкус к анализу обобщающих статистических показателей типа средней и дисперсии — сопоставление различных приемов наиболее компактного и вместе с тем весьма информативного представления данных в виде таблиц и диаграмм. Умение правильно составить таблицу или построить диаграмму, сделать их наглядными и выразительными важно и само по себе. Насыщенный ими материал обладает гораздо большей информативностью, в силу чего расширяется круг потребителей, которые легче и быстрее воспримут заключенное в нем содержание.

В последующих главах автор постепенно усложняет рассматриваемую задачу, вводит новые аспекты анализа и знакомит читателя с соответствующими понятиями теории вероятностей и математической статистики. Главное, к чему А. Эренберг стремится

на протяжении всей книги, — избежать абстрактного теоретизирования, сделать каждое новое понятие оправданным логикой анализа и потребностями прикладного исследования, не только обрисовать возможные применения аппарата, но и показать, что недооценка содержательной стороны дела и излишнее доверие к формальным выкладкам могут привести к ложным выводам.

Книга А. Эренберга, как видно из названия, охватывает не только применение собственно математической статистики, но и более широкую область — *анализ данных*. Параллельное использование этих понятий стало настолько привычным, что почти утерянной оказалась способность их различать. На деле они далеко не тождественны, ибо первое относится к методу, а второе охватывает как метод, так и методологию исследования, включая анализ самих методов, указание их места в процессе решения практических проблем. Книг, в которых ставится и успешно решается задача помочь читателю в конкретной работе с данными, практически нет. В большинстве случаев авторы ограничиваются набором иллюстраций к технике вычислений, либо вовсе оставляя в стороне вопросы методологии, либо пытаясь взглянуть на реальный мир сквозь призму теории, лежащей в основе излагаемых ими методов. А. Эренберг подходит к оценке разработанных статистических методов с позиций исследователя, имеющего определенную практическую цель, когда эти методы существенны лишь постольку, поскольку они приближают его к решению стоящей перед ним содержательной проблемы.

Методы статистики оказываются, безусловно, полезными лишь в тех определенных ситуациях, для которых они подробно разработаны. Сама проверка выполнения всех исходных гипотез, обеспечивающих их применение, может быть проведена далеко не всегда и еще гораздо реже завершается получением утвердительного ответа, особенно если мы располагаем немногими наблюдениями интересующего нас явления. Приступая к анализу конкретных данных, исследователь может располагать различным числом наблюдений и более или менее развитой теорией изучаемых процессов. Наиболее благоприятная ситуация складывается, когда теория способна предложить ему содержательные гипотезы, а данных достаточно, чтобы убедиться в их достоверности или с уверенностью отвергнуть сделанные предположения. В таком случае потребность в привлечении статистических методов может возникнуть, когда проверяется весьма сложная гипотеза, состоящая в объяснении совместного поведения многочисленных факторов, которые объединены в систему моделью, содержащей ряд внутренних связей между соответствующими переменными. А. Эренберг не рассматривает такую возможность и потому мы также оставим ее в стороне. В менее сложных случаях исследователь, располагая достаточным количеством исходных данных, получит интересующий его вывод, применяя лишь статистические понятия, необходимые для измерения основных характеристик имеющейся в его распоряжении совокупности.

Возможна другая крайность: исследователь впервые приступает к анализу принципиально новой проблемы. У него нет ни теоретической базы, ни достаточного количества данных. Статистические методы, как бы они ни были совершенны, способны при этом сыграть лишь второстепенную роль, будучи подспорьем (и то весьма условным) в том анализе, который предстоит провести для накопления систематизированной информации и последующего создания теории.

Итак, остаются лишь две промежуточные ситуации: богатая информационная база при слабых теоретических обобщениях или развитая теория при малом числе достаточно однородных наблюдений. При первой из них статистические методы помогут осуществить структуризацию данных и будут полезны при выборе нуждающихся в проверке теоретических гипотез. При второй — потребность в помощи со стороны этих методов еще большая, но возможности их применения существенно ограничены. Здесь возникает дополнительная сложность, поскольку развитость теории, лишенной мощного эмпирического основания, обычно связана с обилием умозрительных построений, не всегда поддающихся непосредственному сопоставлению с действительностью. Никаких определенных выводов с помощью статистических методов сделать при этом не удастся, они лишь участвуют в процессе конструирования новых умозрительных теорий, доставляя авторам более или менее эффектные аргументы для защиты своей точки зрения.

Таким образом, мощность многих статистических методов возрастает по мере того, как увеличивается возможность без них обойтись: в любой сколько-нибудь сложной обстановке их роль становится вспомогательной, как, впрочем, и любого формального инструментария, а на первое место выдвигается содержательный анализ изучаемой проблемы. Примерно так выглядит общеметодологическая позиция А. Эренберга, которую он стремится обосновать на страницах своей книги. И если с точки зрения используемого формального аппарата эта книга доступна широкому кругу читателей, то ее методологическая основа далеко не тривиальна и не все выводы автора бесспорны.

Основное внимание А. Эренберг уделяет изучению ситуации, где отсутствует развитая теория и перед исследователем стоит задача обнаружения связей между наблюдаемыми показателями, имеющих характер общих закономерностей.

Пытаясь объяснить свое отношение к поиску закономерностей в социальных, экономических и биологических науках, А. Эренберг неоднократно обращается к сопоставлению с законами физики. «Очень часто встречаются попытки, — пишет он в параграфе 5.4, — сравнения законов в физике и в социальных науках. Физические законы обычно считают точными, хорошо обоснованными теоретически и позволяющими получать простые и достаточно надежные предсказания. В отличие от них соотношения в общественных и биологических науках обычно трактуют как прибли-

зительные, имеющие лишь статистический характер, не обладающие хорошим теоретическим фундаментом, а потому полученные с их помощью прогнозы могут быть лишь довольно неопределенными и приблизительными.

При подобных сопоставлениях обычно сравнивают хорошо обоснованные физические законы с некоторыми соотношениями, обнаруженными в социологии, экономике или биологии в результате проведения одного или двух изолированных исследований. Однако соотношения, подобные закону Бойля, базируются на тысячах экспериментов. Долгие десятилетия прошли, прежде чем в результате обобщения и теоретического объяснения эмпирических результатов мог быть сформулирован физический закон.

Но даже и в наше время законы физики и химии не являются абсолютно точными. Они существенно упрощают реальную действительность». А. Эренберг считает, что физические законы успешно применяются не потому, что они точные, а потому, что они исследованы при широком многообразии условий протекания описываемых ими процессов.

Закономерности в общественных науках часто пытаются обнаружить на основе единичных или однократных наблюдений, в силу чего полученные соотношения в значительной степени отражают специфические условия протекания конкретного процесса, а не общую закономерность. Выход А. Эренберг видит в расширении эмпирической базы социально-экономических и биологических исследований, в том, чтобы ограничиться поиском и верификацией ненаправленных связей и не претендовать на отражение причинно-следственных зависимостей.

Поскольку даже законы физики, имеющие надежное эмпирическое обоснование, не являются абсолютно точными и существенно упрощают реальность, нельзя требовать такой точности и от закономерностей в социальных науках и биологии. Достаточно, если эти закономерности хорошо отражают существо связей. Отсюда вытекает и позиция А. Эренберга по отношению к методам статистической обработки данных. Метод наименьших квадратов и многие другие методы, использующие оптимизацию, представляют собой попытку достичь предельной точности в процессе статистической обработки данных, в то время как сами исходные показатели и гипотезы о характере связей такой точностью не обладают.

Проведенное А. Эренбергом сопоставление физических законов и социальных или биологических закономерностей является недостаточно полным, а указанная им ограниченность многих статистических методов вряд ли дает основание для того скептического к ним отношения, которое проявляется в некоторых разделах книги.

По крайней мере необходимо иметь в виду, что общественные науки изучают объект, способный к целенаправленным действиям, что само по себе ограничивает области, где могут быть обнаружены закономерности достаточно общего характера.

Если точку зрения А. Эренберга относить не только к изучению склонностей индивидуальных потребителей того или иного товара (а именно такие примеры преобладают в книге), но и к исследованию закономерностей, проявляющихся на уровне всего народного хозяйства, то придется обратить внимание и на ряд других, гораздо более существенных методологических различий между методами, применяемыми в общественных и естественных науках. В частности, в физике весьма велика роль статистических закономерностей, которые становятся все более детерминированными при переходе на макроуровень. Так, стохастическое поведение молекул определяет температуру состоящего из них тела и т. п. В экономике и в социальных науках проблема перехода с микроуровня на макроуровень гораздо менее разработана. До сих пор основные макропоказатели в экономике формируются в виде суммы абсолютных величин соответствующих микропоказателей, например валовой национальный продукт в статистике капиталистических государств представляет собой агрегат, полученный в виде суммы продаж отдельными производителями товаров и услуг, за вычетом расходов на приобретение товаров, необходимых для обеспечения производственных нужд предприятия. Такой показатель, безусловно, имеет смысл, но, как правило, не тождественный смыслу отдельных образующих его слагаемых. Например, в физике от энергии движения молекул зависит количество соударений и, следовательно, температура тела, но при этом суммарная энергия молекул вовсе не будет равна энергии тела. Точно так же ВВП скорее характеризует некоторую, сравнимую с предыдущим периодом, хозяйственную активность, интенсивность взаимодействия элементов системы, а не ее продуктивность.

Вопрос о сквозных измерениях в экономике, охватывающих как микро-, так и макроуровень, вообще далеко не прост и вряд ли может быть решен на основе статистических подходов, поскольку получение сопоставимых и интерпретируемых данных возможно лишь на основе определенного встроенного в систему механизма их формирования, который в свою очередь оказывает воздействие на получаемый результат.

А. Эренберг постоянно подчеркивает свое более чем сдержанное отношение к методу наименьших квадратов, к построенным на его основе уравнениям регрессии, к факторному и компонентному анализу. В его позиции в этих вопросах имеет место явное смещение акцентов. Критиковать тот или иной метод можно лишь за отсутствие строгости положенного в его основу алгоритма. Если же все необходимые предпосылки указаны, а вывод логически безупречен, то остается одно из двух: либо рассуждать о силе метода, анализируя выполнимость предпосылок в тех или иных конкретных условиях, либо предъявить претензии к практикам, неоправданно выбирающим сферу и способы его применения.

А. Эренберг прибегает и к тем, и к другим аргументам. Но

он также пытается противопоставить одному методу другой. В частности, он утверждает, что в случае, когда мы располагаем различными группами наблюдений, каждая из которых соответствует специфическим условиям эксперимента, эффективнее проводить линию регрессии не с помощью метода наименьших квадратов, а используя значения групповых средних. Как только рассуждения об ограниченности любого конкретного подхода сменяются предложением определенных альтернатив, обсуждение смещается в область, где помимо общих методологических соображений становятся возможными точные сопоставления.

Надо сказать, что вряд ли кому-либо из исследователей удастся описать такую достаточно общую ситуацию, в которой предлагаемый А. Эренбергом метод окажется лучше метода наименьших квадратов или одной из его модификаций. Если мы располагаем несколькими группами наблюдений и у нас есть основания считать, что принадлежность элементов к разным группам существенно влияет на характер связи между переменными, то вряд ли мы станем применять метод наименьших квадратов сразу ко всем наблюдениям, не проведя предварительного анализа. Мы установим число наблюдений в каждой из групп, примем во внимание особенности получения данных о каждой из них, содержание стоящей перед нами задачи и с учетом всех этих обстоятельств выберем более подходящую модификацию метода наименьших квадратов.

Тем не менее существуют весьма важные аспекты проблемы, благодаря которым претензии А. Эренберга становятся достаточно обоснованными. Так, метод наименьших квадратов, особенно если учесть многочисленность его применяемых модификаций, недостаточно удобен в процессе проведения самого исследования. Он оказывается слишком точным и громоздким, чтобы исследователь, не обладающий высокой квалификацией, мог постоянно ощущать связь между получаемыми результатами и характером изменений, происходящих при переходе от одной группы к другой. Работу, состоящую в поиске формы соотношения между переменными, в сопоставлении воздействия на общий результат каждой из рассматриваемых групп наблюдений, безусловно, удобнее вести, опираясь на более наглядные и менее точные методы. Однако при оформлении окончательного результата целесообразно привлечь более сильные методы.

А. Эренберг критикует главным образом методы статистического анализа, которые в той или иной форме опираются на поиск «наилучшего», с точки зрения определенного критерия, решения задачи, т. е. являются оптимизационными. При этом он, безусловно, прав, когда обращает внимание на тот факт, что поиски «наилучшего» решения для конкретной совокупности наблюдений могут вовсе не приблизить нас к получению достаточно общей закономерности.

Каждый, кто строил регрессионные модели, знает, насколько чувствительными оказываются их коэффициенты к добавлению

новых данных, если объем имеющейся информации невелик. Перед исследователем возникает целая серия сложных проблем, решение которых в большинстве случаев ему не под силу.

Во-первых, как относиться к возможности невыполнения гипотез, лежащих в основе того или иного метода? Ведь каждый метод работает в предположении, что эти гипотезы реализованы. Но даже если полученные отклонения от рассчитанной с помощью метода теоретической линии оказываются именно такими, как предполагалось вначале, то что из этого факта следует? Ведь может случиться, что свойства остатков были обеспечены при «наилучшей» подгонке линии регрессии к исходным данным за счет искажения самой модели. (Кстати, когда это происходит, то непременно наблюдается неустойчивость получаемых оценок при изменении совокупности принимаемых во внимание наблюдений.) Лишь столкнувшись с неинтерпретируемостью параметров, мы начинаем задумываться над тем, что могло привести к такому результату. Если же параметры укладываются в приемлемый для нас диапазон, то мы склонны воспринимать полученный результат как вполне объективный.

Во-вторых, насколько мы можем быть уверены в содержательной репрезентативности анализируемой ситуации? Разве наше знание реального процесса настолько полно, что позволяет рассчитывать на нереальность существования неучтенного, но важного для общего описания фактора? Ведь оптимальное решение может быть полезно только при условии, что нам точно известны все обстоятельства анализируемого или контролируемого процесса. Когда это не так, оптимизация является, с методологической точки зрения, исключительно рискованной. Отрицательный эффект от воздействия неучтенных факторов легко может перевесить выигрыш за счет выбора «наилучшего» решения. Более того, когда отсутствует своего рода резерв прочности, неучет существенного фактора приводит, как правило, к нереализуемым результатам.

Приведенных аргументов, по-видимому, достаточно, чтобы у читателя возник естественный в таком случае вопрос: как же быть практике, какие методы применять, а от применения каких следует воздержаться? Общего ответа на этот вопрос существовать не может. Каждая конкретная задача предъявляет свои требования к методам ее решения, в том числе и к статистическим методам, если в их применении появляется необходимость. С уверенностью можно сказать лишь одно: формальное применение статистических методов, не опирающееся на глубокое понимание существа проблемы и теоретических предпосылок, которые лежат в основе этих методов, только случайно может привести к осмысленным результатам. Чтобы с большими основаниями можно было рассчитывать на успех, необходимо внимательно относиться к методологическим аспектам статистического исследования. Книга А. Эренберга, безусловно, поможет в овладении необходимыми для этого навыками.

Прочитать эту книгу полезно как тому, кто начинает изучать методы статистического анализа данных, так и специалисту, имеющему опыт самостоятельной работы. Первый, несомненно, оценит простоту изложения, которая достигается благодаря максимально комфортной для читателя психологической атмосфере, активизирующей его интуитивный и практический опыт и тем самым способствующей более быстрому и прочному восприятию теоретических понятий и идей. Второй обнаружит в книге много нетривиальных соображений методологического характера, что должно побудить его еще раз проанализировать свое отношение к результатам, получаемым с помощью популярных статистических методов при решении практических задач. Преподаватель вуза найдет в книге А. Эренберга много полезных методических приемов и примеров, позволяющих оживить и обогатить лекционный материал. Некоторая дискуссионность изложения дает ему возможность не только продемонстрировать студентам разнообразие точек зрения на роль статистических методов, но и показать, что существенные проблемы могут найти решение лишь благодаря умелому сочетанию содержательного и формального подходов к анализу конкретных данных и интерпретации полученных при этом выводов.

*А. А. РЫВКИН*

## ПРЕДИСЛОВИЕ

Эта книга представляет собой начальный курс по обработке и интерпретации статистических данных. Она предназначена всем, кто в своей повседневной работе сталкивается со всевозможной статистической информацией.

Многие стремятся больше узнать о статистике, о том, как обрабатываются данные и как сформировать суждение об анализе, проведенном другими людьми, какая техника обработки предпочтительнее в тех или иных условиях. Статистика необходима администраторам и предпринимателям, экономистам и инженерно-техническим работникам, а также ученым, преподавателям и студентам, специализирующимся в области общественных и естественных наук.

Читатель вправе задать вопрос: чем предлагаемая книга отличается от множества уже имеющихся книг по статистике? Во-первых, в ней явно рассматриваются приемы и методы, позволяющие автору избежать ситуации, когда, приводя необозримые таблицы чисел, он тем самым фактически расписывается в собственной беспомощности: «Я не понимаю этих данных, может быть, *вы* сможете понять их». С самого начала акцент должен быть сделан на сведения первоначальных обширных массивов чисел к краткому, но несущему в себе значительную информацию резюме, и каждому исследователю надлежит стремиться именно к такой подаче результатов своей работы. К сожалению, большинство книг свидетельствует о том, что подобное стремление со стороны авторов явно недостаточно.

Во-вторых, книга проливает свет на сущность статистических процедур. В ней не просто на манер повареной книги излагается техника обработки данных. Объяснить, *почему* мы поступаем так или иначе, — вот чему уделяется наибольшее внимание.

В-третьих, хотя в книге излагаются главным образом основы курса статистики, в ней все же используется достаточно сложная техника, иногда необходимая для получения наиболее существенных выводов. Подобная техника применяется специалистами-статистиками, но и рядовой читатель также должен иметь общее представление о соответствующих принципах обработки статистической информации. Это поможет ему как при использовании уже имеющихся результатов, так и при самостоятельной обработке новых данных.

Многие традиционно изучаемые в статистике понятия, такие, как частотные распределения, вероятность, проверка значимости, описываются и обсуждаются лишь в заключительной части книги, поскольку они имеют менее широкую область применения. Наконец, в книге последовательно проводится мысль о том, что полученные результаты могут оказаться полезными только в той степени, в которой они позволяют нам делать обобщения. Таким образом, обобщение эмпирических данных — главная тема книги.

Цель книги — научить читателя считать, выявлять связи, скрытые в числовых данных, получать обобщенные характеристики, удобные для интерпретации, использования и представления широкой аудитории. *Необходимо заставить данные говорить.*

Однако научиться считать — не значит стать математиком. Математика в книге сведена к минимуму. Читатель только должен уметь обращаться с уравнениями типа  $y = 2x + 5$  и иногда с некоторыми простыми алгебраическими выражениями, логарифмами или квадратными корнями. Подлинная роль математики проявляется на *более поздних* стадиях анализа, когда математика помогает обобщить многочисленные и разнообразные результаты. В переходе от рассмотрения простых соотношений к построенным более общим моделям и теориям отражается естественный ход процесса познания. Мы будем сталкиваться с этим процессом в различных разделах книги.

Чтобы на основе имеющихся данных получить обобщающие результаты, необходимо рассмотреть эти данные в той конкретной форме, в которой они были получены. Поэтому числовые примеры играют в книге исключительно важную роль. Некоторые из них взяты из моей повседневной работы и связаны с теми обычными скучными проблемами, с которыми читатель, возможно, также встречается в своей практике.

*Преобразование данных* — это радикальный подход к статистическим методам и проблемам. Цель книги заключается также в том, чтобы сделать понятным способ мышления, принятый статистиками и учеными. Основной принцип при анализе данных состоит в необходимости добиваться результатов, пригодных для использования в последующей работе. Нет ничего проще, чем посмотреть, согласуются ли новые результаты со старыми (например, по-прежнему ли выполняется соотношение  $y = 2x + 5$ ?). Подобное использование имеющихся результатов ведет к эмпирическим обобщениям, т. е. к получению результатов, которые будут иметь место при широком диапазоне изменения условий наблюдения. Именно так обычно протекает процесс научного познания.

### Структура книги

Книга состоит из пяти частей. В первой и второй частях рассматриваются способы анализа различных совокупностей данных на базе одной переменной и соотношений между переменными.

В третьей части обсуждаются обычные статистические процедуры, позволяющие описать нерегулярные изменения признака на множестве значений некоторой ограниченной совокупности наблюдений. Здесь же дается практическая оценка методов корреляционного и регрессионного анализа, а также вводится понятие стохастической модели.

В четвертой части мы имеем дело с выборками и процедурой статистического вывода на основе выборочных данных. (Преподаватели, желающие использовать книгу в соответствии с обычной последовательностью изложения курса статистики, могут начать с каких-либо глав из третьей или четвертой части.)

В пятой части сравниваются различные подходы к сбору статистических данных и обсуждается связь между описанием наблюдаемого явления и его объяснением.

Каждая глава заканчивается упражнениями. Содержащиеся в них вопросы подробно обсуждаются. Эти упражнения помогут читателю усвоить основные положения курса, которые одним могли показаться слишком элементарными, а другим слишком сложными, а также сделают изложение основного текста главы более полным и подробным, что особенно важно при первом чтении.

Лондон  
Январь 1974 г.

*А. ЭРЕНБЕРГ*

Данные часто бывают представлены в таком виде, что на их основе нельзя немедленно сделать какие-либо определенные выводы. В этом случае читатель может либо оставить их без внимания, либо проанализировать самостоятельно, либо обратиться к автору с просьбой сформулировать его точку зрения. В последних двух случаях полезно знать, что можно сделать для облегчения понимания информации, содержащейся в представленных данных.

В гл. 1 настоящей книги описаны основные процедуры, позволяющие лучше усвоить смысл табличных данных. В гл. 2 мы будем иметь дело с ситуацией, когда уже накоплен некоторый опыт работы с аналогичными показателями.

После выявления особенностей данных полученные результаты необходимо обобщить. И здесь, конечно, более эффективны обобщающие показатели, нежели просто большие массивы чисел. Соответственно в гл. 3 обсуждаются границы использования таблиц и графиков.

Численные результаты не могут интерпретироваться изолированно, сами по себе. Их необходимо сопоставить с уже имеющимися образцами и теоретическими понятиями. Этому посвящается гл. 4.

# I ГЛАВА

## СРЕДНИЕ И СПОСОБ ПРЕДСТАВЛЕНИЯ ДАННЫХ

Вначале мы рассмотрим таблицу необработанных данных. В этой же главе мы обсудим способы, с помощью которых можно лучше понять содержащуюся в них информацию.

Таблица 1.1. Данные по четырем районам для восьми трехмесячных периодов

|   | 1969—1970 гг. |       |        |       |       |       |       |       |
|---|---------------|-------|--------|-------|-------|-------|-------|-------|
|   | 13—15         | 16—18 | 19—21  | 22—24 | 25—27 | 28—30 | 31—33 | 34—36 |
| А | 97,63         | 92,24 | 100,90 | 90,39 | 95,69 | 94,44 | 91,13 | 97,81 |
| Б | 48,29         | 42,31 | 49,98  | 39,09 | 46,38 | 49,74 | 41,74 | 37,39 |
| В | 75,23         | 75,16 | 100,11 | 74,23 | 74,23 | 76,97 | 71,66 | 76,47 |
| Г | 49,69         | 57,21 | 80,19  | 51,09 | 52,88 | 49,41 | 59,32 | 52,56 |