

Методы **В.М.Симчера** многомерного анализа статистических данных

ТЕМА 1 Типология методов многомерного анализа данных

ТЕМА 2 Методы многомерного корреляционного изучения данных

ТЕМА 3 Методы многомерной обработки данных

ТЕМА 4 Методы многомерного моделирования данных

ТЕМА 5 Методы многомерного сопоставления данных



В.М.Симчера

Методы многомерного анализа статистических данных

Рекомендовано
Учебно-методическим объединением
по образованию в области финансов,
учета и мировой экономики
в качестве учебного пособия
для студентов, обучающихся по специальностям
“Финансы и кредит”, “Бухгалтерский учет,
анализ и аудит”, “Мировая экономика”,
“Налоги и налогообложение”



**МОСКВА
“ФИНАНСЫ И СТАТИСТИКА”
2008**

УДК 519.237(075.8)

ББК 22.172я73

С37

РЕЦЕНЗЕНТЫ:

Кафедра статистики

Финансовой академии при Правительстве РФ
(заведующий кафедрой – профессор В.Н. Салин);

Б.И. Исаков,

доктор экономических наук, профессор

*На 1-й странице обложки – репродукция картины
«Демографические волны» М.А. Королева, председателя
Межгосударственного статистического комитета СНГ,
из частной коллекции В.Л. Соколина, руководителя
Федеральной службы государственной статистики.
Публикуется с разрешения М.А. Королева и В.Л. Соколина*

Симчера В.М.

С37 Методы многомерного анализа статистических данных:
учеб. пособие. – М.: Финансы и статистика, 2008. – 400 с.: ил.
ISBN 978-5-279-03184-9

В данном пособии в отличие от аналогичных изданий процедуры применения сложных методов многомерного анализа рассматриваются по шагам на конкретных, а не на абстрактных примерах (их представлено в книге более 150). Каждая из пяти тем пособия включает вводные замечания, типологию решаемых задач, задачи для самостоятельных занятий, тесты и вопросы для самоконтроля. Примеры решения большеразмерных задач приведены с применением пакетов электронной обработки данных, в частности пакетов SAS, Statistica, Stadia, Statgraphics.

Для студентов, аспирантов и специалистов, занимающихся статистическими вычислениями повышенной сложности.

С $\frac{1602090000 - 125}{010(01) - 2008}$ 50 – 2007

УДК 519.237(075.8)

ББК 22.172я73

ISBN 978-5-279-03184-9

© Симчера В.М., 2008

© Издательство «Финансы и статистика»,
2008

Предисловие

Методов многомерного анализа данных много, но они разрозненные и, как правило, несводимые в единое целое. Многообразие этих методов обусловлено объективным многообразием изучаемых явлений, которые данные методы призваны отображать и измерять. Ценность их определяется тем, насколько каждый из них и все они адекватны изучаемым предметам, полно и достоверно выявляют и объясняют скрытые причинно-следственные связи признаков, которые не могут быть установлены и предъявлены с помощью плоских одномерных расчетов и примитивных цифровых иллюстраций.

Распространенное пренебрежение этими методами, игнорирование их в условиях принятия сложных управленческих решений чревато большими упущенными выгодами и потерями материальных, трудовых и финансовых ресурсов, которые вне реализации оптимальных схем используются некомплексно, а следовательно, неэффективно. Отсюда вытекает объективная необходимость изучения методов многомерного анализа данных и, несмотря на сложные процедуры их приложения, столь же объективная целесообразность их широкого практического применения.

В отличие от простых одномерных методов, оперирующих ограниченными и, как правило, однородными наборами объектов наблюдения и очевидными взаимосвязями между их признаками, многомерные методы имеют дело с неограниченными и разрозненными наборами наблюдаемых объектов и неочевидными и, как правило, многообразными и по преимуществу разнонаправленными взаимосвязями между их признаками.

Фундаментальное отличие состоит в том, что само множество наблюдаемых объектов и признаков, как и гипотезы и закономерности распределения и изменения их значений в пространстве и времени, здесь неизвестны и не даны, а должны быть найдены, выступая каждый раз не только целью определения исходных условий, но и сущностью самого исследования.

Предметом наблюдения и изучения здесь выступают как параметрические, так и непараметрические (качественные поряд-

ковые или ранговые) связи, детерминированные и недетерминированные (стохастические), массовые и робастные, реальные и ложные, наблюдаемые и ненаблюдаемые (латентные) связи, т.е. все виды и формы связей, недоступные для простого наблюдения и изучения. При этом разнородность наблюдаемых объектов и многообразие признаков, характеризующих их, неочевидность и разнонаправленность взаимосвязей между ними определяются многомерной природой наблюдаемых явлений, формирующих сложное матричное множество пересекающихся неоднородных объектов и комплексных признаков, выявление и изучение которых невозможны с помощью простых одномерных методов. В результате возникает объективная необходимость обращения к методам многомерного анализа данных, успех в применении которых определяется знанием природы изучаемых объектов, их размерности и многообразных форм многомерных взаимосвязей*.

* Конечно, при наличии однородных совокупностей наблюдаемых явлений и однонаправленных процессов их изменения в пространстве и времени, аппроксимируемых линейными функциями, вопрос выбора методов анализа их связей, оценки сходимости и проверки достоверности получаемых результатов не возникает.

Определять характер и значение факторов методами функционального, дисперсионного, индексного, корреляционного или многомерного анализа при наличии линейных и однонаправленных связей – это безразлично. Все эти методы обеспечивают получение относительно одинаково сходимых и относительно правдоподобных результатов. В этих случаях предпочтение всегда и везде, при прочих равных условиях, отдадут методу, с помощью которого искомый результат находится кратчайшим путем и с минимальными затратами сил и средств.

Принципиальная разница появляется в случаях наличия неоднородных, нелинейных и разнонаправленных данных и их связей, которых в природе и обществе подавляющее большинство. Вот тогда приходится обращаться к методам компонентного и факторного анализа, поскольку с помощью всех остальных методов эти данные и связи попросту не аппроксимируются.

Методы одно- и многомерного факторного анализа двух показателей и более в обиходе обычно путают и представляют как равноценные и взаимозаменяемые, что не соответствует действительности.

Методы одномерного анализа, под которыми понимают методы определения общего прироста за счет отдельных факторов (в том числе и большего числа чем один-два), в частности методы функционально-стоимостного (нормативного), индексного, дисперсионного и корреляционного анализа, аппроксимируют, как правило, один-единственный тип нормаль-

Несмотря на актуальность и важность, многомерные методы в современной литературе (не говоря о работах прошлых времен) представлены чрезвычайно усложненно. К сожалению, прозрачно объединить в целое и извлечь нечто рациональное из того, что к настоящему времени опубликовано, очень сложно. А студентам сегодня преподаются в качестве многомерных методов скорее наборы некоторых усложненных счетных процедур, не имеющих, как правило, ни начала, ни конца, но не логически связанные и практически понятые и применимые алгоритмы. Подобные материи в нормальных дидактических средах понять нельзя, их можно только запомнить. Это объясняется во многом, конечно, не отсутствием желания излагать сложное прозрачно и просто, а сложностью и многомерностью самого предмета анализа, объективной невозможностью его простого представления, о чем речь шла ранее.

Нельзя объять необъятное, нельзя алгоритм, требующий выполнения нескольких десятков и даже сотен действий, заменить алгоритмом из двух-трех действий. Попытка такой замены – это всегда упрощенчество и, следовательно, разрушение сути самого дела, а не упрощение, удача или гениальный ход, как это иногда может показаться легковерным. И поэтому не случайно, а скорее закономерно, что все (или почти все) современные, и тем более старые, традиционные учебные пособия по статистике оперируют, как правило, одномерными методами сбора, обработки и анализа данных, упрощая многомерные как чрезвычайно сложные, которые в рамках учебных пособий физически просто невозможно представить.

Иллюстрировать с помощью таких приемов что-то, конечно, можно и нужно, доказать и объяснить практически ничего нельзя. Столетиями, начиная с Я.Бернулли (1654–1705), пробовали, ничего не вышло. Не выйдет и впредь, когда доказывать и объяснять приходится куда более сложные виртуальные

распределений и строятся в предположении существования и воспроизводства линейных связей, тогда как методы многомерного факторного анализа аппроксимируют много различных типов распределения и строятся в предположении существования и моделирования нелинейных связей. Количество сопрягаемых факторов при этом не имеет значения; и в том и в другом случае их может быть сколь угодно много.

В зарубежной литературе методы факторного анализа называются и представляются как собственно методы факторного анализа (Factors Analysis), а методы одномерного анализа – как One Way или Two Way Analysis.

вещи и нейронные связи, о существовании которых прежде понятия не имели.

Почему? Объяснение простое. При изучении и применении одномерных статистических методов ограничиваются обычно заранее известными, как правило, упрощенными предпосылками, допущениями и гипотезами, не требующими предварительных доказательств и проверок. Здесь повсеместно доминируют здравый смысл и принцип правдоподобия, ходячие истины типа «все сущее – разумно, все разумное – сущее», «все очевидное – верно, все верное – очевидно», а все неочевидное – неверно, и оно, конечно же, только от болезненного воображения или дьявола.

Словом, дело обычно представлялось (представляется и теперь) так, что основу основ, фундамент статистики как науки составляют эмпирические наблюдения за окружающим нас миром, упрощенные детерминированные законы его познания, не требующие доказательств индуктивные обобщения, а все остальное (стохастические процессы, индетерминизм, случайность, вероятность, риски, методы проверки статистических гипотез, ненаблюдаемые факты и т.д.) – это упражнения досужего ума.

В связи с этим в прошлом почти все ограничивалось имитацией и во многом воображаемыми оценками. Не отвечая своему призванию и назначению, статистика была вынуждена адаптировать факты, представлять их в ожидаемом виде, а не как соответствующие духу и букве праведной жизни. И так обстояло дело (продолжает обстоять и теперь) отнюдь не в одной России. Так оно в принципе обстоит и в других странах, включая страны так называемой развитой демократии и экономики, откуда, в сущности, как раз и происходит практика формального представления и субъективной интерпретации данных.

Причина заключается в том, что изучением неочевидных событий и процессов, законами распределения скрытых, латентных и, как правило, ненаблюдаемых явлений, сложными гипотезами и законами, скажем, многомерного, робастного или биномиального распределений, распределениями Пуассона, Байеса, Вальраса, Парето, Коши, Вейбула и т.д., изучением случайности как сложной закономерности, словом, изучением реального мира традиционная статистика мало интересовалась и ограниченно не занималась. Не особо занимается она этим по тем же соображениям сложности предмета сколько-нибудь значимо и теперь.

В результате за рамками традиционных методов оставались (и остаются) все (или почти все) приемы и процедуры статистического эксперимента, эконометрики, актуарных вычислений, электронной обработки данных, прикладной статистики, важнейшие гипотезы статистических распределений, критерии и приемы их проверки, независимые, нелинейные и непараметрические связи и т.д. Отсюда общая ограниченность традиционной статистики, убывающий коэффициент ее полезного действия, недовольство ею и игнорирование ее фундаментальных ценностей.

Между тем окружающий нас мир — это мир сложных, многомерных и противоречивых процессов, фактов и событий, адекватное познание и одномерное объяснение которого невозможно (и никогда не было и не будет возможно) с помощью простых приемов, на уровне здравого смысла, эмпирически, одним лишь методом проб и ошибок. Преодолевать ограниченность современной традиционной статистики вне апелляции и практического применения методов многомерного анализа данных далее невозможно.

Конечно, концептуальное понимание отличий многомерных от одномерных методов сбора, обработки, анализа, моделирования и сопоставления данных необходимо, но недостаточно в учебном процессе. Здесь дополнительно требуются знания профессиональных различий, без которых осмысленно решать и предметно объяснять решение прикладных задач проблематично. В чем состоят эти отличия?

При применении одномерных методов наблюдаемые объекты (их множество может быть сколь угодно большим; минимальное множество, соответствующее требованиям малой выборки, во всех случаях должно составлять 30 ед.; в более продвинутых случаях — 100 ед.) характеризуются одним-единственным наблюдаемым признаком (например, наблюдаемые предприятия — численностью занятых), при двумерном пространстве — соответственно двумя признаками (например, численностью занятых и производительностью труда), а при применении многомерных методов и, следовательно, при работе в многомерных пространствах — тремя признаками и более (например, численностью занятых, производительностью труда и фондовооруженностью, фондами, выручкой, фондоотдачей и прибылью). Это пример семимерной модели анализа связей).

В изложенной нами постановке задачи одномерный статистический анализ представляет частный случай многомерного,

его отправную точку, иллюстративный стенд. И в этом понимании он всегда необходим и важен.

Целью традиционных методов является по преимуществу декларация фактов, иллюстрация конкретных случаев, фотографическая характеристика наблюдаемых объектов и их признаков, а не их представление как носителей и образцов, моделирующих и объясняющих законы поведения целых явлений. И в этом случае методы одномерного анализа имеют право на существование, но только на существование в отнюдь не универсальном масштабе и значении. Цель многомерных методов — выявление именно этих законов путем установления характера распределения и тесноты связей между многими (обычно тремя и больше) не только очевидными, но и скрытыми признаками, позволяющими полно и детально изучать и объяснять наблюдаемые процессы, что в принципе невозможно сделать, оперируя традиционными одномерными методами.

Еще более фундаментальное отличие касается теоретических допущений и гипотез, аппроксимирующих признаки и связи ненаблюдаемых явлений.

Практически все задачи одномерного анализа ставятся и решаются в предположении того, что в природе существует так называемый гауссовский закон распределения данных. Под этот закон подгоняются или подстраиваются многие многомерные статистические данные и связи. То же самое происходит, когда речь идет о решении некоторого класса специфических многомерных задач, эмпирическое распределение данных в которых сходится или хорошо согласуется с гауссовским распределением.

Других типов распределений (нелинейных, непараметрических, робастных, нейронных и т.д.) эмпирическая статистика не знает или почти не знает. Какой выход в этом случае? Законы одних распределений подменяются другими, вводится сушая чехарда или, что еще хуже, вопрос о гипотезе попросту игнорируется.

Конечно, это крайний случай. Специалисты, разумеется, обычно обращаются к так называемым комбинаторным методам или искусственно подгоняют наблюдаемые данные, отсекая те массивы данных, которые не укладываются в их схемы, квалифицируя такого рода данные как ненормальные, случайные или вырожденные. Это иногда дает неплохие результаты, но в большинстве случаев находится далеко от конструктивного решения. Комбинаторика, как и подмена так называемых ненормальных распределений и нелинейных связей нормальными

распределениями и линейными связями, в сущности, не дает ничего: многомерный анализ остается лишенным ясных теоретических оснований, а получаемые результаты — необходимого содержательного смысла*.

* Допустим, наблюдаемые семь факторных признаков-показателей, детерминирующих один результативный, например рост ВВП, соответствуют семи типам разных распределений. Предположим, эмпирические данные о численности занятых соответствуют теоретическим значениям нормального распределения Гаусса, производительность труда — биномиального, фонды — распределения Парето, фондовооруженность — распределения Вальраса, фондоотдача — показательного распределения, издержки — распределения Пуассона, прибыль — распределения Байеса, а распределение ВВП на разных отрезках пространства и времени робастно, асимметрично и разнонаправленно.

Как в этом случае установить общий закон их распределения и, следовательно, общую их связь с законом распределения ВВП? Можно ли вывести закон распределения ВВП как гибрид из семи других представленных типов распределений, которым следуют приведенные факторы ВВП? Без доказательств очевидно, что нельзя, любая попытка синтеза данных разнотипных распределений будет примитивной фикцией.

Можно ли построить на основе частично робастных, асимметричных, разнонаправленно меняющихся и по большей части хаотических данных тренд, аппроксимирующий какой-либо из известных или конструирующийся какой-либо новый, неизвестный тип распределений? В принципе можно, но проблематично. С аналогичной задачей сталкиваются практически всегда, аппроксимируя эмпирические данные по любому наблюдаемому показателю, но всего лишь в единичных случаях, при наличии предельных массивов данных, когда, опираясь на закон больших чисел, удается констатировать приемлемую сходимость эмпирических данных с теоретическими их значениями. В подавляющем большинстве статистических экспериментов сходимость обычно плохая или вовсе отсутствует, что указывает на неудачную аппроксимацию данных с помощью избранного типа распределений и требует либо увеличения круга наблюдаемых данных, либо поиска нового типа и, следовательно, проведения повторного эксперимента. Именно путем такой, подчас многократной итерации нахождение тренда в сложных случаях становится возможным.

В таких ситуациях, располагая общим трендом и опираясь на метод дедукции, можно шаг за шагом решать задачу нахождения частных трендов (путем построения евклидовых пространств и определения минимальных парных расстояний между наблюдаемыми частными и общим трендами). Впрочем, при подобных допущениях методом индукции также решается задача нахождения неизвестного общего тренда на основе известных частных трендов.

Конечно, в данном случае речь идет всего лишь об идее решения задачи, но идее конструктивной, заслуживающей экспериментальной проверки, идее, которая открывает перспективы построения единой теории многомерных распределений.

Понятно, что при наличии разных типов распределений и задач апеллировать к одним и тем же процедурам их идентификации и решения в принципе невозможно. На встречающиеся имитации здесь вообще не следует обращать внимание.

Как действовать? Разбивать сложные многомерные пространства на части, вводить их типы, соответствующие типам существующих теоретических допущений и распределений, т.е. решать задачу отдельно по каждому признаку или однородной группе признаков, применяя разные методы. Словом, классические многомерные данные и задачи модифицировать и переводить в форму специфических данных и задач, соответствующих известным и доступным методам их обработки и решения.

В сущности, это означает переход от методов многомерного решения соответствующих классов задач к методам одномерного. Понятно, что, если подобные задачи существуют в природе, их разбиение может быть признано приемлемым и полезным, если нет – каждый раз оно будет профанацией и пустой тратой времени. Нельзя ведь, к примеру, матрицу 1000×1000 позиций разбить на пять или десять меньших и при этом найти полноценное решение искомых задач, в частности коэффициенты полных затрат или объективно обусловленные издержки и цены. Именно поэтому, к сожалению, приходится признавать факт существования огромной массы нерешаемых статистических задач и апеллировать к поиску новых теорий и статистической идентификации новых информационных пространств.

По той же причине приходится накладывать своего рода ограничение на существующее множество методов многомерного анализа данных и в обиходе обращаться только к тем из них, которые адекватно аппроксимируют известные законы и гипотезы распределения данных и нашли широкое применение в статистической работе.

Среди этих методов, кроме методов корреляционного изучения связей, особое значение имеют методы компонентного, факторного, дискриминантного и кластерного анализа, моделирования и сопоставления данных, которые представляют предмет настоящего учебного пособия.

К решению задач именно этого рода с помощью представленных в пособии методов как раз и призывает автор настоящей работы. И не только призывает, но и методично, шаг за шагом показывает, как это следует делать, что заслуживает одобрения и, на мой взгляд, может вызвать неподдельный интерес и принести пользу как студентам, так и многим исследователям, а также всем тем, кто занимается применением и развитием фундаментальных методов современной прикладной статистики.

Академик Н.П. Федоренко

Типология методов многомерного анализа данных

1.1. Вводные замечания

Условием успеха любого исследования является сходимость теоретических допущений наблюдаемым фактам, и наоборот. И неважно, идет ли речь об индуктивном или дедуктивном исследовании, детерминированных или индетерминированных фактах и их причинно-следственных связях, — условие каждый раз остается неизменным.

Сходимость теории и фактов, их адекватность или неадекватность устанавливаются путем идентификации теоретических и эмпирических распределений, существующих в природе и известных науке.

Сходимость на уровне необходимых и достаточных условий, а не абстрактных идеалов проверяется по известным критериям статистического согласия (или в случае их отсутствия) — по критериям правдоподобия и здравого смысла. Именно так, как много столетий назад, измышления и теории проверяются фактами и практикой, а факты и практика — теорией и измышлениями. Ничего другого человек за это время не смог или не успел придумать.

И там, где теория подтверждается фактами, а факты — теорией, возможен и проводится статистический эксперимент, способный дать (и обычно дающий) весомые результаты, имеющие ясный смысл и первостепенное научное и практическое значение, а там, где не подтверждается, подобный эксперимент невозможен, его инициирование бессмысленно, а результаты фальшивые.

Приходится не без сожаления констатировать, что первый случай (и скорее не столько в общественных, сколько в естественных науках) по разным причинам, а главное, по причине лимита самих знаний и ресурсов их реализации, ограничен и

Тема 4. Методы многомерного моделирования данных	206
4.1. Вводные замечания	206
4.2. Методические указания по изучению темы	207
4.2.1. Методы градиентного анализа	207
4.2.2. Методы линейного программирования	212
4.2.3. Методы межотраслевого анализа	216
4.2.4. Методы технико-экономических обоснований	224
4.3. Примеры решения типовых задач	228
4.3.1. Примеры градиентного анализа	228
4.3.2. Примеры линейного программирования	235
4.3.3. Примеры межотраслевого анализа	237
4.3.4. Примеры технико-экономического обоснования	245
4.4. Задачи для самостоятельных занятий	285
4.4.1. Задачи градиентного анализа	285
4.4.2. Задачи линейного программирования	288
Рекомендуемая литература	290
Тема 5. Методы многомерного сопоставления данных	292
5.1. Вводные замечания	292
5.2. Методы определения однородных типов сопоставляемых региональных образований	292
5.3. Методические рекомендации по определению однородности и тесноты взаимосвязи сопоставляемых региональных показателей	321
5.4. Методы определения многомерных рейтингов деловой активности, предпринимательской уверенности и инвестиционной привлекательности регионов России	326
5.5. Методы многомерного прогнозирования региональных темпов экономического роста	351
Вопросы для самоконтроля	367
Рекомендуемая литература	369
Русско-английский словарь терминов по многомерному анализу статистических данных	372
Публикации В.М. Симчеры по методам многомерного анализа статистических данных	392

Contents

Foreword	3
Topic 1. Typology of multivariate data analysis methods	11
1.1. Introductory remarks.....	11
1.2. Methods of multivariate either empirical and practical data analysis	14
Recommended literature for further studying of topic	29
Topic 2. Methods of multivariate correlation data analysis	30
2.1. Introductory remarks.....	30
2.2. Methodological directions on topic studying	35
2.3. Examples of routine problems solving	52
2.4. Exercises	73
Tests and questions.....	77
Recommended literature for further studying of topic	97
Topic 3. Methods of multivariate data processing	98
3.1. Introductory remarks.....	98
3.2. Methodological directions on topic studying	101
3.2.1. Methods of component analysis	104
3.2.2. Methods of factorial analysis	113
3.2.3. Methods of cluster analysis	116
3.2.4. Methods of discriminant analysis	122
3.3. Examples of routine problems solving	125
3.3.1. Examples on component analysis	125
3.3.2. Examples on factorial analysis	131
3.3.3. Examples on cluster analysis.....	134
3.3.4. Examples on discriminant analysis	146
3.4. Exercises	185
3.4.1. Problems on component analysis	185
3.4.2. Problems on factorial analysis	188
3.4.3. Problems on cluster analysis	190
3.4.4. Problems on discriminant analysis.....	192
3.5. Variants of computer control problems	193
3.5.1. Variants of main control problems	195
3.5.2. Variants of additional control problems	198
3.5.3. Variants of audit problems	199
Tests and questions.....	200
A topics for classes	202
Recommended literature for further studying of topic	204

Topic 4. Methods of multivariate data modeling	206
4.1. Introductory remarks	206
4.2. Methodological directions on topic studying	207
4.2.1. Methods of gradient analysis	207
4.2.2. Methods of linear programming	212
4.2.3. Methods of inter-branch analysis	216
4.2.4. Methods of feasibility studies	224
4.3. Examples on problems solving	228
4.3.1. Examples on gradient analysis	228
4.3.2. Examples on linear programming	235
4.3.3. Examples on inter-branch analysis	237
4.3.4. Examples on feasibility studies	245
4.4. Exercises	285
4.4.1. Problems on gradient analysis	285
4.4.2. Problems on linear programming	288
Recommended literature for further studying of topic	290
Topic 5. Methods of multivariate data comparisons	292
5.1. Introductory remarks	292
5.2. Determination methods of uniform types of regional formations being compared	292
5.3. Methodological recommendations on determination of homogeneity and closeness relations of regional indicators being compared	321
5.4. Determination methods of multivariate ratings (by the example of rating determination of business activity, business confidence and investments attractiveness of Russian regions)	326
5.5. Methods of multivariate forecasts on economic growth rate	351
Questions	367
Recommended literature	369
Russian-English Glossary of widely used terms on methods of multivariate analysis of statistical data	372
Others books by Professor Vasily M. Simchera on methods of multivariate analysis of statistical data	392

Methods of multivariate analysis of statistical data

Vassily M. Simchera

Moscow, Financy i Statistica Publishers, 2007, 400 p.

Unlike simple methods which operate with limited and as a rule uniform sets of observations as well with as obvious relationships between their characteristics, the multiple methods deal with unlimited and as a rule separate and not obvious sets of observations as well as with multi-directional and diverse relationships between their characteristics.

The subjects of observation and study in this book are as parametric as non-parametric (qualitative, ordinal and rank) relations, deterministic and non-deterministic (stochastic), mass and robust, true and false, observable and non-observable (latent) phenomena, i.e. all kinds and types of statistical combinations and relations which are inaccessible for direct survey and study.

At the same time the diversity of objects being observed and multiplicity of their properties, unobviousness and multi-direction of relations between them are defined by multi-dimensional nature of phenomena being observed which form the complex matrix multitude of intercrossed multi-dimensional objects and features reflecting their relations which one cannot explore and study using simple methods.

This book contains five topics referring to methods of multivariate data analysis representing the subject of scrutiny and application of modern statistics successful application of which became possible because of knowledge of subjects nature being studied, their dimensions and manifold types of multivariate relations.

There are methodological instructions for studying each topic, typology of tasks being solved and examples of solving typical tasks as well as exercises for practice presented in brief.

The solving big dimensional problems is made with appliance of modern packages of electronic data processing in particular Statistica, Stadia and Statgrafhics.

This book is intended for students, post-graduates and experts engaged in application of multi dimensional methods for solving big dimensional applied statistical tasks problems.

Учебное издание

Симчера Василий Михайлович

**МЕТОДЫ МНОГОМЕРНОГО АНАЛИЗА
СТАТИСТИЧЕСКИХ ДАННЫХ**

Заведующая редакцией *Л.А. Табакова*
Ведущий редактор *Н.А. Кузнецова*
Младший редактор *Н.А. Федорова*
Художественный редактор *Ю.И. Артюхов*
Технический редактор *Т.С. Маринина*
Корректоры *Т.М. Васильева, Г.Д. Кузнецова*
Компьютерный набор *И.В. Витте,*
Е.Ф. Тимохиной, О.В. Фортунатовой
Компьютерная верстка *И.В. Витте, Е.Ф. Тимохиной*
Оформление художника *Н.М. Биксентеева*

ИБ № 5081

Подписано в печать 12.10.2007. Формат 60x88/16
Гарнитура «Таймс». Печать офсетная
Усл.п.л. 24,5. Уч.-изд. л. 22,7
Тираж 2000 экз. Заказ «С» 125

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефоны: (495) 625-35-02, 625-47-08
Факс (495) 625-09-57
E-mail: mail@finstat.ru <http://www.finstat.ru>

ООО «Великолукская городская типография»
182100, Псковская область, г. Великие Луки,
ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: zakaz@veltip.ru