

И.Н. Дубина

МАТЕМАТИКО- СТАТИСТИЧЕСКИЕ МЕТОДЫ

В ЭМПИРИЧЕСКИХ
СОЦИАЛЬНО-
ЭКОНОМИЧЕСКИХ
ИССЛЕДОВАНИЯХ



И.Н.Дубина

МАТЕМАТИКО- СТАТИСТИЧЕСКИЕ МЕТОДЫ В ЭМПИРИЧЕСКИХ СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

Допущено
учебно-методическим объединением
по классическому университетскому образованию
в качестве учебного пособия по дисциплине
федерального компонента “Эконометрика”
для студентов высших учебных заведений,
обучающихся по направлению 080100 “Экономика”



**МОСКВА
“ФИНАНСЫ И СТАТИСТИКА”**



**ИЗДАТЕЛЬСКИЙ ДОМ
“ИНФРА-М”**

2010

РЕЦЕНЗЕНТЫ:

Кафедра информационных систем в экономике
Алтайского государственного университета
(заведующий кафедрой – О.П. Мамченко,
доктор экономических наук, профессор);

П.М. Симонов,
доктор физико-математических наук, доцент,
заместитель декана по научной работе
экономического факультета
Пермского государственного университета

Дубина И.Н.

Д79

Математико-статистические методы в эмпирических социально-экономических исследованиях: учеб. пособие / И.Н. Дубина. – М.: Финансы и статистика; ИНФРА-М. – 2010. – 416 с.: ил.

ISBN 978-5-279-03107-8 (Финансы и статистика)

ISBN 978-5-16-003910-7(ИНФРА-М)

Рассмотрены основы теории социально-экономических измерений, представлены подходы к проектированию измерительных шкал, к проверке и обоснованию надежности результатов измерения, исследовательских гипотез и метода прогнозирования. Цель пособия – формирование, развитие и закрепление навыков использования специальных методов и прикладного программного обеспечения (математико-статистических пакетов SPSS, STATISTICA и WINSTEPS) для математически корректного анализа результатов количественных исследований в экономических и социальных областях.

Для студентов, магистрантов и аспирантов экономических вузов. Может быть полезно широкому кругу исследователей и практиков, использующих методы теории измерений и статистического анализа.

Д $\frac{1602090000 - 039}{010(01) - 2010}$ 34 – 2009

УДК 330.43я(075.8)
ББК 65.49в631.8я73

ISBN 978-5-279-03107-8
ISBN 978-5-16-003910-7

© Дубина И.Н., 2010
© Издательство «Финансы и статистика», 2010

ОГЛАВЛЕНИЕ

Предисловие	7
Индикативное задание	10
Часть первая. ПРОЕКТИРОВАНИЕ И ОЦЕНКА КАЧЕСТВА СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ИЗМЕРЕНИЙ	
Глава 1. Измерение и типы измерительных шкал	13
1.1. Измерение и измерительная шкала	13
1.2. Типы измерительных шкал	16
1.3. Базовые процедуры обработки данных и проблема выбора измерительной шкалы	23
Глава 2. Оценки и измерения	26
2.1. Различия между измерением и оценкой	26
2.2. Шкала Лайкерта	27
2.3. Возможности преобразования данных из неметрических шкал в метрические	33
2.4. Модель Раша	36
2.5. Реализация модели Раша в программе WINSTEPS	47
2.6. Метод QIPM: пример компромисса между точностью и простотой	54
Глава 3. Формирование и анализ выборочной совокупности	61
3.1. Особенности выборочного метода исследования	61
3.2. Методы формирования выборки	64
3.3. Анализ точности результатов выборочного исследования: доверительный интервал	68
3.4. Использование статистических пакетов для определения базовых характеристик выборки	79
3.5. Определение необходимого объема выборки	88

Глава 4. Оценка качества измерений и измерительных инструментов в социально-экономических исследованиях	101
4.1. Показатели качества измерений	101
4.2. Валидность как характеристика измерений.....	103
4.3. Подходы к оценке надежности измерений.....	107
4.4. Использование статистических пакетов для оценки надежности измерений	123
4.5. Оценка согласованности измерений на основе модели Раша	135
4.6. Оценка качества измерительных инструментов	140

Часть вторая. АНАЛИЗ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ И СТАТИСТИЧЕСКИЕ ВЫВОДЫ

Глава 5. Основные принципы проверки статистических гипотез	148
5.1. Статистическая достоверность связи и статистическая гипотеза	148
5.2. Нулевая и альтернативная гипотезы	151
5.3. Виды статистических ошибок и уровни статистической значимости	155
5.4. Общая процедура проверки статистических гипотез	162
5.5. Статистические методы и критерии проверки гипотез	164
5.6. Статистическая значимость и практическая важность	166
Глава 6. Параметрические методы проверки статистических гипотез	170
6.1. Применение z -критерия	170
6.2. Применение t -критерия Стьюдента для независимых выборок	174
6.3. Применение t -критерия Стьюдента для связанных выборок.....	183

Глава 7. Дисперсионный анализ	188
7.1. Особенности применения дисперсионного анализа	188
7.2. Однофакторный дисперсионный анализ	190
7.3. Критерии множественных сравнений	196
7.4. Реализация метода однофакторного дисперсионного анализа в статистических пакетах	200
7.5. Многофакторный дисперсионный анализ	210
7.6. Дисперсионный анализ с повторными измерениями	215
Глава 8. Непараметрические методы проверки статистических гипотез	226
8.1. Критерий χ^2 Пирсона	226
8.2. Тест Мак-Немара	235
8.3. Критерий Манна–Уитни	241
8.4. Тест Колмогорова–Смирнова	245
8.5. Критерий знаков и T -критерий Уилкоксона	248
8.6. Критерии Краскела–Уоллиса и χ^2 Фридмана	251
Часть третья. КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ И СТАТИСТИЧЕСКОЕ ПРОГНОЗИРОВАНИЕ	
Глава 9. Меры связи и их проверка на значимость	256
9.1. Виды связей исследуемых признаков	256
9.2. Сущность и задачи корреляционного анализа	260
9.3. Меры связи для переменных в метрических шкалах	264
9.4. Меры связи для переменных в неметрических шкалах	268
Глава 10. Регрессионный анализ	279
10.1. Модели парной и множественной регрессии	279
10.2. Проверка обоснованности регрессионной модели	288
10.3. Построение регрессионных моделей в программе Excel	297
10.4. Построение регрессионных моделей в пакете SPSS	304

Глава 11. Моделирование и прогнозирование на основе анализа временных рядов	341
11.1. Основные характеристики и общие модели временных рядов	341
11.2. Простейшие модели временных рядов и автокорреляционный анализ.....	346
11.3. Методы сглаживания временных рядов.....	356
11.4. Регрессионные и авторегрессионные модели временных рядов	378
Библиографический список	393
Приложения	398

ПРЕДИСЛОВИЕ

Получить и обработать необходимую информацию, используемую в теории и практике социально-экономических и гуманитарных наук, можно лишь на основе *грамотного* применения статистических подходов и методов. И хотя статистика является обязательной дисциплиной для экономических, социологических и многих других вузовских специальностей, студенты, изучающие этот курс, имеют часто лишь самое общее представление о возможностях статистического анализа. В частности, в рамках базовых курсов по математике и статистике, как правило, не рассматриваются (или рассматриваются на самом вводимом уровне) специальные математико-статистические методы и программное обеспечение для проектирования экспериментальных исследований и обработки данных. Для сравнения: многие американские и европейские университеты предлагают специальные курсы по основам теории измерений в гуманитарной и социальной сферах, методам проектирования измерительных шкал, количественным методам исследования бизнеса и социально-экономических процессов.

У студентов российской высшей школы часто отсутствуют необходимые знания и навыки социально-экономических измерений, проектирования исследований и статистической обработки данных. Результаты, представляемые в некоторых дипломных и даже диссертационных работах, а также публикациях, часто являются не только малообоснованными, но и полностью лишены смысла с математической точки зрения.

Методы и инструменты, используемые российскими молодыми исследователями, часто не соответствуют современным научным стандартам, и поэтому они плохо «вписываются» в мировой научный дискурс, а работы российских ученых почти не представлены в международных нетехнических научных журналах.

Среди исследователей, к сожалению, распространено убеждение, что статистические методы нужны только для обработки результатов наблюдений. В данном учебном пособии акцент

сделан на применении математико-статистических методов на этапах планирования эксперимента и оценки качества полученных исследовательских результатов. Именно в этих областях у многих студентов и молодых исследователей, даже имеющих солидную математико-статистическую подготовку, отсутствуют необходимые знания.

Учебное пособие посвящено вопросам применения математико-статистических методов в эмпирических исследованиях, т.е. таких исследованиях, которые используют количественные данные, полученные опытным путем — в результате экспериментов, наблюдений, измерений и представленные в числовом виде. Речь идет не только о теоретических исследованиях, а в первую очередь — о практических исследованиях, проводимых для решения реальных задач в областях маркетинга, управления, прогнозирования и оптимизации производства товаров, оценки качества товаров и услуг и многих других сферах деятельности.

Цель данного пособия — сформулировать основы статистической грамотности в тех областях, которые часто остаются за пределами базовых вузовских курсов по математике, статистике и эконометрике, а также многих учебников и учебных пособий. Некоторые темы, рассматриваемые в пособии, связанные с оценкой качества социально-экономических измерений, почти не отражены в учебной литературе. Особое внимание уделяется математическим основам проектирования количественных социально-экономических измерений: подробно рассматриваются виды измерительных шкал, методы калибровки и конвертирования данных из неметрических в метрические шкалы (в том числе с использованием модели Раша), а также подходы и методы для оценки точности и надежности социально-экономических измерений (с применением коэффициента альфа Кронбаха, коэффициентов Гутмана и других критериев и показателей). Более обстоятельно, чем в других учебниках и учебных пособиях, освещаются вопросы статистической проверки исследовательских гипотез, разбираются особенности использования статистических критериев и методов для такой проверки и получения корректных (с математической точки зрения) результатов. Заметное место в пособии занимает обсуждение вопросов корректного и адекватного использования методов корреляционно-регрессионного анализа и анализа временных рядов для прогнозирования социально-экономических процессов.

Большое внимание уделяется работе с программными продуктами для обработки и анализа исследовательских данных — WINSTEPS, STATISTICA, SPSS. Помимо специальных статистических пакетов в книге приводится описание возможности использования основных статистических функций Excel, которые могут быть применены для решения относительно простых задач статистического анализа.

В пособии дано много примеров и решений для иллюстрации «внутренних» механизмов работы тех или иных процедур, применяемых для анализа исследовательских результатов. Это необходимо, поскольку работа с современными статистическими пакетами, реализующими данные процедуры, и, самое главное, корректная интерпретация полученных результатов требуют ясного понимания теоретических основ применяемых процедур. В частности, нужно четко понимать, что означают те или иные показатели, на каких принципах основано их получение, с какими предпосылками и ограничениями они связаны. После рассмотрения всех этих моментов с иллюстрацией на сравнительно простых примерах приводится описание того, как аналогичную, но более сложную задачу можно решить с помощью того или иного программного продукта.

Автор старался не перегружать изложение материала абстрактными формулами, теориями и выводами в целях большей доступности, однако в то же время стремился избегать тривиализации статистических подходов и выводов.

Учебный материал подается структурированно, включает большое количество заданий, ко многим задачам и упражнениям даны комментарии, поэтому материал можно изучать самостоятельно. Многие упражнения, приведенные в пособии, основаны на реальном исследовательском материале.

Автор надеется, что это пособие поможет читателям повысить уровень исследовательской квалификации, получить знания и навыки по проектированию измерительных шкал для проведения социально-экономических исследований, по оценке точности и надежности социально-экономических измерений, использованию математико-статистических методов обработки и анализа результатов научных исследований, а также будет способствовать более качественному решению практических задач в реальных областях экономики и бизнеса.

ИНДИКАТИВНОЕ ЗАДАНИЕ

Возможно, читатель недоумевает: почему непосредственно за предисловием предлагается учебное задание (ведь учебное пособие, как правило, начинается с изложения материала по некоторой теме, после чего следуют задания). Это объясняется тем, что данное задание является ознакомительным и в определенном смысле индикативным (англ. *indicative*, т.е. указывающий, показывающий).

Читателю предлагается найти (самостоятельно или по согласованию с преподавателем) любую статью, связанную с использованием математико-статистического аппарата для обработки данных эмпирических исследований в социальной или экономической сфере (желательно в области профильной специальности и (или) исследовательских интересов читателя). Такую статью можно найти, например, в журналах «Экономическая социология», «Экономика и математические методы», «Социология: методология, методы, математические модели», «Социологические исследования» и других специальных периодических изданиях.

В подразделе Библиографический список этого пособия указаны адреса интернет-сайтов, на которых представлены содержание выпусков этих журналов, аннотации и тексты многих статей. Исследования с использованием статистических методов и моделей также содержатся на сайтах «Статистические методы», «Образовательный портал Экономика, Социология, Менеджмент: Методы сбора и анализа социологических данных» и других сайтах, адреса которых указаны в подразделе Библиографический список.

Выбрав статью и ознакомившись с ее текстом, попытайтесь ответить на приведенные ниже вопросы и выполнить простые задания. Возможно, в тексте нет прямых ответов на вопросы, которые вам предлагаются. В таком случае попробуйте восстановить логику исследователя по косвенным признакам и *предположите* возможные ответы.

Для выполнения этого задания не нужно обращаться к каким-либо учебным материалам. Основная цель задания – самоопределение исходного уровня понимания текста, описывающего использование подходов статистического анализа. Поэтому данное задание и названо индикативным. Автор советует читателю сохранить статью, а также записать и сохранить ответы на предлагаемые вопросы, а после освоения всего материала пособия вернуться к этой же статье и вновь ответить на вопросы данного задания. Затем сравнить эти ответы с ответами, которые вы дали ранее и проанализировать их. Изменились ли ваши ответы и ваше понимание статьи после работы с книгой?

Разбираться с чужим текстом гораздо сложнее, чем со своим, тем более восстанавливать логику авторов статьи и находить ответы на вопросы. Поэтому автор советует читателю при ответах на вопросы пытаться аргументировать ответы, а не ограничиваться ответами «да», «нет», «не знаю». Неважно, «правильным» будет обоснование или нет, главное на этом этапе – попытка понимания и осмысления подходов и методов для решения задач, рассматриваемых в статье.

Один из главных вопросов индикативного задания – обоснование репрезентативности выборки. Эта проблема, несмотря на кажущуюся простоту, весьма важна. Дело в том, что далеко не все авторы эмпирических социально-экономических исследований уделяют внимание обоснованию репрезентативности используемых в исследовании выборочных совокупностей. Следовательно, результаты таких исследований могут быть поставлены под сомнение¹.

Вопросы и задания по статье

• Основано ли исследование, представленное в статье, на выборке из генеральной совокупности?

Если да, то

а) каков размер выборки?

б) какова репрезентативность выборки?

в) каким образом формировалась выборка?

г) можно ли обобщить полученные результаты на генеральную совокупность?

¹ Вопросу репрезентативности и ее обоснованию посвящена в значительной степени гл. 2 данного учебного пособия.

• Какие переменные использовались? Какие из них являются зависимыми, а какие независимыми?

• Какие измерительные шкалы применяются в данном исследовании?

• Какие методы и подходы к обработке данных используются авторами?

• Какие статистические гипотезы проверяются и какими методами?

• Укажите несколько (до 10) терминов, встречающихся в статье и непонятных для вас (если такие есть).

• Оцените, насколько понятна вам терминология, используемая в статье:

полностью понятна

в основном понятна

отчасти понятна

плохо понятна

совсем не понятна

• Оцените, насколько вам понятны используемые в статье математические символы и формулы:

полностью понятны

в основном понятны

отчасти понятны

плохо понятны

совсем не понятны

• Оцените, насколько вам понятны основные результаты и выводы исследования, представленного в статье?

полностью понятны

в основном понятны

отчасти понятны

плохо понятны

совсем не понятны

• Оцените, насколько вы знакомы со статистическими подходами и методами, которые описаны и/или используются в статье?

я их постоянно использую

хорошо знаком (а)

отчасти знаком (а)

плохо знаком (а)

совсем не знаком (а)

Часть первая

Проектирование и оценка качества социально- экономических измерений

Если что-либо существует,
оно может быть измерено.

Э. Торндайк

Точно так же, как все искусства
стремятся к музыке, так и все науки
стремятся к измерениям.

Дж. Сантаяна

Не все то, что может быть сосчитано,
может быть измерено,
и не все то, что может быть измерено,
может быть сосчитано.

А. Эйнштейн

Глава 1

Измерение и типы измерительных шкал

1.1. Измерение и измерительная шкала

Высказывания, вынесенные в качестве эпитафии к части первой данного пособия, в определенной степени отражают проблему измерения и его роль в научном исследовании. Измерение, понимаемое прежде всего как численное выражение некоторого свойства, традиционно рассматривалось в качестве необходимого и важнейшего компонента научного исследования. Более того, измеримость часто выступала критерием «научности» вообще: если мы не можем выделить свойства объектов, измерить их признаки и установить на этой основе причинно-следствен-

ные связи, то мы не можем говорить о научности и объективности исследования. Причем этот принцип распространялся не только на естественно-научные, но и на социальные и гуманитарные области исследования. Иллюстрацией к такому подходу могут служить высказывания Э. Торндайка и Дж. Сантаяны, приведенные в эпиграфе. В то же время в современной науке сформировалось осознание сложности измерительных процедур и несводимости их к простому численному представлению свойств исследуемых объектов. Это понимание выразил А. Эйнштейн фразой, также представленной в эпиграфе.

В широком смысле слова любое исследование – это процесс сбора и анализа информации для формулировки и проверки идей и выводов. Статистические подходы применимы для исследований, в которых используются *данные*, т.е. информация, представляемая в числовом виде. При всей тривиальности этих положений следует понимать, что недостаточно просто собрать данные и никакие данные не говорят сами за себя.

Более того, число обладает некой «магической силой» для исследователя: число как бы «само по себе» *a priori* означает и даже подчеркивает объективность исследования, а иногда и направляет порядок действия исследователя. Ведь всем знакомы элементарные арифметические операции, и у начинающих исследователей часто возникает соблазн получить некий результат, «что-нибудь сложив, разделив или умножив».

Иллюстрацией к вышеизложенному является история, за которой, возможно, стоит и реальный случай. Некий начинающий маркетинголог, оценивая привлекательность населенных пунктов для строительства горнолыжных курортов в регионе, использовал три критерия для расчета «интегрального показателя привлекательности». В результате «исследования» был выявлен «лучший» пункт с максимальным рейтингом.

Население, чел.	3012
Высота над уровнем моря, м	1753
Дата основания, год	1925
Итого	6690

Комментировать эти результаты нет необходимости – их абсурдность очевидна. Но подобные анекдоты, как это ни печально, встречаются и в реальной исследовательской практике.

Профессор Форман приводит в своей книге [43] следующий поучительный случай. В 1990 г. в США был опубликован рейтинг мест отдыха пенсионеров, в котором по семи критериям был ранжирован 151 город, и в результате суммирования получен некий итоговый показатель. На основе примененной методики Лас-Вегас оказался на 105-м месте. В 1994 г. была использована иная процедура сбора и обработки данных. В результате Лас-Вегас оказался первым в рейтинге. Безусловно, причина столь радикальных перемен — не в изменении условий отдыха в Лас-Вегасе, а в изменении методологии измерений!

Основное искусство¹ статистического анализа заключается в получении и интерпретации данных. Получая данные, исследователь приписывает изучаемому объекту или событию числа по некоторому правилу, т.е. он так или иначе использует ту или иную процедуру измерения. Объектами исследования могут быть города, страны, машины, детали, люди, явления и т.д. Предметом изучения являются свойства этих объектов или событий. Например, в качестве свойства человека (группы людей, города, страны) может выступать богатство. Исследователь фиксирует выраженность интересующих его свойств изучаемых объектов с помощью обнаружения некоторых признаков (например, количества денег). Степень выраженности свойств может (но не обязательно) характеризоваться числом. Например, богатство определяется как наличие или отсутствие денег (это может кодироваться, например, числами 0 и 1), а также их количество.

Один из распространенных, хотя и не бесспорных, подходов к пониманию измерения можно сформулировать так: *измерение* — это кодирование и соотнесение степени выраженности признаков эмпирических объектов или событий с помощью чисел в соответствии с определенными правилами (измерительными шкалами). Очень важно понимать, что эти правила измерения определяют как результат измерения, так и исследовательский вывод (сравните результаты в примере о Лас-Вегасе).

Измерительная шкала определяет тип получаемых данных и множество операций, которые можно осуществлять с этими данными.

¹ Здесь слово «искусство» употреблено в смысле греч. *techné* — мастерство, т.е. совокупность навыков и приемов в какой-либо сфере.

Измерительная шкала – это математическая характеристика переменной, определяемая процессом измерения и типом измеряемой переменной, являющаяся основным критерием для выбора соответствующего математико-статистического метода. Применяемая измерительная шкала – основной элемент при обосновании возможных методов обработки данных и интерпретации результатов. Прежде чем использовать ту или иную математическую операцию (пусть даже из области элементарной арифметики, не говоря уже о сложных статистических методах), очень важно понимать, позволяет ли выбранная шкала осуществлять такую операцию, проведено ли измерение таким способом, для которого возможно *корректное* использование данной операции.

Вопрос о типе используемой шкалы особо важен в социально-экономических и гуманитарных исследованиях, поскольку многие измерения в этих сферах являются косвенными. Непосредственно мы фиксируем некоторые признаки (время решения задачи, ответы на вопросы, оценки респондентов и т.д.) и на этой основе делаем выводы о некотором скрытом (*латентном*) свойстве, недоступном прямому измерению (например, коммуникативность, креативность, подготовленность по предмету, факторы организационного климата и т.п.).

Определение шкалы измерений является первым этапом любого эмпирического исследования. К сожалению, многие исследователи проходят данный этап «автоматически», совсем не задумываясь об этом или руководствуясь лишь «здравым смыслом». Именно поэтому гл. 1 и 2 этого пособия посвящены измерительным шкалам.

1.2. Типы измерительных шкал

Выделяют три основных атрибута измерительных шкал, наличие или отсутствие которых определяет принадлежность шкалы к той или иной категории.

Упорядоченность шкалы означает, что одна позиция шкалы, определяемая числом и соответствующая выраженности измеряемого свойства, больше, меньше или равна другой позиции.

Интервальность шкалы означает, что интервалы между позициями шкалы равны между собой.

Нулевая точка (или точка отсчета) шкалы означает, что набор чисел, соответствующих выраженности измеряемого при-

знака, имеет точку отсчета, обозначаемую 0, которая соответствует полному отсутствию измеряемого свойства.

Наиболее общей классификацией измерительных шкал является разделение на *неметрические шкалы* (в которых единица измерения отсутствует) и *метрические шкалы* (в которых может быть установлена единица измерения, например килограмм, градус, рубль и т.д.). Метрические шкалы являются более *мощными*, т.е. они лучше дифференцируют измеряемые объекты (отражают больше информации о различии измеряемых объектов).

К неметрическим шкалам относят номинативную и порядковую шкалы. В *номинативной шкале (nominal scale)* отсутствуют все главные атрибуты измерительных шкал, а именно упорядоченность, интервальность, нулевая точка. Для обозначения такой шкалы также используются термины *шкала наименований* и *номинальная шкала*.

Номинативная шкала используется для классификации или идентификации объектов (группировки по классам, каждому из которых приписывается число). Объекты группируются по классам таким образом, чтобы внутри класса они были идентичны по измеряемому свойству. Это самая простая шкала из тех, что могут рассматриваться как измерительные, хотя фактически она не ассоциируется с измерением и не связана с понятиями «величина» и «число». Она используется только с целью отличить один объект от другого.

Примером свойства, измеряемого в такой шкале, может являться пол. Эта переменная имеет две категории (наименьшее количество категорий, которое может иметь измеряемая переменная). Это свойство может быть измерено отнесением человека (объекта) в соответствующую категорию при подсчете количества мужчин и женщин. Категория «мужчины» может, например, кодироваться цифрой 1, а категория «женщины» — цифрой 2. В результате можно сделать заключение о том, к какой категории отнесено больше объектов, а к какой меньше. Однако на основе этих данных нельзя произвести логические или математические операции над самими категориями, т.е. для такой шкалы нельзя, например, сказать, что 1 меньше, чем 2, или что $2 - 1 = 1$. Другими примерами свойств, измеряемых в номинативной шкале, являются семейное положение, национальность, религиозная принадлежность, место рождения.

Категории переменных, измеряемых в номинативной шкале, должны быть взаимоисключающими, чтобы не было неопределенности относительно классификации исследуемых объектов. Эти категории должны также быть исчерпывающими, т.е. каждому возможному значению должна однозначно соответствовать определенная категория (хотя бы категория *другие* или *прочие*). Категории данных в номинативной шкале должны включать сопоставимые экземпляры.

Категории измеряемой переменной фактически не являются численными, и их можно сравнивать друг с другом только на основе количества наблюдений, относящихся к ним. При использовании номинативных шкал цифры, кодирующие соответствующие категории, являются лишь метками (именами) и не несут в себе численного значения. Как мы показали на примере, к данным, представленным в такой шкале, арифметические операции не применимы: их нельзя складывать, вычитать, умножать или делить. Однако разработаны специальные процедуры обработки и анализа данных в номинативных шкалах (например, критерий χ^2 для статистической проверки гипотез). Они основаны на определении частоты (подсчете количества появлений переменной в различных категориях).

В *порядковой шкале (rating scale)* присутствует упорядоченность, но отсутствуют атрибуты интервальности и нулевой точки. Для обозначения такой шкалы также используются термины *ранговая шкала* и *шкала рангов*. Порядковая шкала наиболее распространена в социальных и гуманитарных исследованиях.

Результатом измерений в порядковой шкале является упорядочение объектов. Шкала ранжирует объекты, приписывает им числа в зависимости от выраженности измеряемого свойства по некоторому признаку (в порядке убывания или возрастания). В отличие от номинативной шкалы здесь можно не просто определить, что один объект отличен от другого, но и то, что один объект больше или меньше другого. Следовательно, шкала показывает, больше или меньше выражено свойство (измеряемая величина), но не насколько больше или насколько меньше оно выражено, а тем более во сколько раз больше или меньше.

Приведем несколько примеров измерений в порядковой шкале.

Пример 1.1. Социально-экономический статус определяется в следующих категориях: 1) высший класс; 2) средний класс; 3) низший класс. Эти категории ранжируются, например, по признаку уровня дохода. Понятно, что в данной шкале значение 1 больше, чем 2, но непонятно насколько. То же самое можно сказать о должностном статусе, например: специалист, ведущий специалист, главный специалист, заместитель начальника отдела, начальник отдела и т.д.

Пример 1.2. Четверым гонщикам присвоены «места» в соответствии с тем, кто раньше достиг финиша: $A - 1$; $B - 2$; $C - 3$; $D - 4$.

Основываясь только на этих данных, можно судить лишь о том, кто приехал раньше, а кто позже. Нельзя сказать, *насколько* раньше приехал один из гонщиков, чем другой. Более того, совершенно очевидно, что разрыв между гонщиками, например, A и B меньше, чем B и D , хотя $2 - 1 = 1$ и $4 - 2 = 2$. Гонщик A мог, например, достичь финиша на 20 мин. раньше гонщика B , от которого гонщик D отстал всего на 1 мин.

Пример 1.3. Выступления спортсменов-фигуристов, как правило, оцениваются по шкале от 1 до 10. При этом оценки, например, от 3 до 5 встречаются гораздо реже, чем оценки от 7 до 9, а наиболее значимыми оценками являются оценки от 9 до 10. Очевидно, что интервал между оценками 9,9 и 10,0 для судей и для спортсменов гораздо больше, чем интервал между оценками 4,9 и 5,0.

Пример 1.4. В школе и вузе используется 5-балльная система оценки знаний (1, 2, 3, 4, 5). Можем ли мы ответить, *насколько* 5 «лучше», чем 4? Настолько же, насколько 4 «лучше», чем 3? Возможно, для одного студента разница между 5 и 4 невелика, а разница между 4 и 3 значительна. Возможно, эта разница изменяется от курса к курсу. Если считать, что оценки учеников или студентов отражают их знания, то в случае применения к этой шкале действий арифметики получается, что знания отличника равны сумме знаний двоечника и троечника ($5 = 2 + 3$), а между знаниями отличника и троечника такая же разница, как между знаниями «ударника» и двоечника ($5 - 3 = 4 - 2$).

Следовательно, мы не можем сказать, что в такой шкале сохраняется равенство интервалов между пунктами. Сложение, как и большинство других математических операций, предполагает, что интервалы между значениями равны. Если интервалы не равны, то $2 + 2$ может означать 5 или 55. Поэтому широко распространенная процедура усреднения оценок (баллов), т.е. вычисления среднего арифметического, — это операция, лишенная смысла с математической точки зрения.

Таким образом, возможности статистического анализа для данных в порядковых шкалах ограничены. Строго говоря, упорядочение — это единственная математическая операция, применимая к порядковой шкале. Использование многих операций с порядковыми данными (например, вычисление среднего арифметического) математически некорректно, но широко распространено на практике. Например, в рейтинге мест отдыха американских пенсионеров (см. пример в подразд. 1.1) усреднение баллов порядковой шкалы представляло собой математически бессмысленную процедуру.

В то же время шкала может вполне корректно использоваться в экспериментальных исследованиях, но для этого необходимо применять адекватные методы обработки данных, например конвертирование порядковых шкал в метрические шкалы или использовать непараметрические методы для обработки данных. (Подробно методы обработки результатов измерений в порядковой шкале будут рассмотрены ниже.)

К метрическим шкалам относят интервальную шкалу и шкалу отношений. В *интервальной шкале (interval scale)* присутствуют упорядоченность и интервальность, но нет нулевой точки. В этой шкале исследуемому объекту присваивается число единиц измерения, пропорциональное выраженности измеряемого свойства. Соответствующие интервалы разных участков шкалы имеют одно и то же значение, поэтому измерения в интервальной шкале допускают не только классификацию и ранжирование, но и точное определение различий между категориями.

Примеры интервальной шкалы: фиксация времени и даты, температурные шкалы Цельсия и Фаренгейта. Например, для известных шкал измерения времени можно сказать, что интервал между 3 и 6 ч равен интервалу между 4 и 7 ч. Но можно ли сказать, что 6 ч в два раза больше, чем 3 ч? Очевидно, нет. Аналогично, если сегодня уличный термометр фиксирует температуру -10°C , а вчера было -20°C , мы можем сказать, что сегодня теплее на 10° . Но можно ли сказать, что сегодня теплее в 2 раза? Это кажется нелепым. Однако сказать, что температура в 300°C отличается от температуры в 200°C в два раза больше, чем температура в 150°C отличается от температуры в 100°C , кажется вполне логичным.

Подобные ответы связаны с тем, что интервальные шкалы не имеют нулевой точки отсчета. Точнее, выбор нулевой точки

в интервальной шкале условен (произволен). Имея данные, представленные в интервальной шкале, мы можем судить о том, *насколько* больше или *насколько* меньше выражено измеряемое свойство, но не о том, *во сколько раз* больше или меньше. Для интервальных шкал характерна произвольность выбора нулевой точки, т.е. 0 не соответствует полному отсутствию измеряемого свойства, поэтому измерения не соответствуют абсолютному количеству измеряемого свойства. Например, нулевое значение температуры по шкале Цельсия не соответствует полному отсутствию тепла. Для интервальных шкал мы можем корректно использовать большинство математических операций, поэтому если, например, имеются данные в порядковой шкале, то с точки зрения математической корректности целесообразно конвертировать их в интервальную шкалу, используя специальные процедуры, которые будут рассмотрены далее. Но на практике иногда возникают сложности как при преобразовании исходных данных в метрические шкалы, так и при интерпретации и анализе полученных результатов.

В *шкале отношений (ratio scale)* присутствуют все атрибуты измерительных шкал: упорядоченность, интервальность, нулевая точка. Для обозначения такой шкалы также используется термин *абсолютная шкала*, что подчеркивает абсолютный характер нулевой точки.

Шкала отношений позволяет оценивать, во сколько раз свойство одного объекта больше или меньше аналогичного свойства другого объекта, принимаемого за эталон, единицу. Эта шкала характеризуется всеми атрибутами интервальной шкалы и, кроме того, имеет фиксированную нулевую точку (0), которая не является условной, так как она соответствует полному отсутствию измеряемого свойства. Например, переменная *количество сотрудников* имеет *фактическое* начало отсчета, так как нулевое значение соответствует отсутствию сотрудников вообще. Аналогично, нулевое значение может соответствовать отсутствию образования, дохода, детей, количеству лет в браке и т.п. Другими примерами измерения в абсолютной шкале являются определение физических характеристик (веса, длины, высоты, площади, температуры по температурной шкале Кельвина), запись возраста респондента, денежных потоков, численности населения, количества сотрудников, временных периодов (например, времени выполнения задания) и т.п.

В силу фиксированности нулевой точки при сравнении исследуемых объектов можно судить не только о том, насколько больше или меньше выражено свойство, но и во сколько раз больше или меньше оно выражено. Для абсолютной шкалы возможно корректное использование любых математико-статистических операций и процедур обработки данных. Впрочем, стоит отметить, что существование нулевой точки редко является критичным для статистического анализа. Например, выполнение большинства известных параметрических методов (работающих с метрическими шкалами) для проверки статистических гипотез предполагает, что данные представлены в интервальной шкале.

Рассмотренная классификация измерительных шкал не является единственной. Некоторые специалисты (например, Дж. Хили) выделяют лишь три категории измерительных шкал, не проводя существенных различий между интервальной и относительной шкалами и объединяя эти шкалы в одну [29]. Авторы некоторых учебных и специальных изданий, напротив, выделяют большее количество шкал. Например, профессор А.И. Орлов выделяет шесть типов шкал (шкалу наименований, порядковую шкалу, шкалу интервалов, шкалу отношений, шкалу разностей, абсолютную шкалу) [21].

В качестве примеров конкретных шкал, применяемых, например, в маркетинговых исследованиях, можно назвать следующие.

Шкала Лайкерта (Likert scale) используется при определении представлений, мнений и отношений к товарам, торговым маркам, магазинам и т.п. Шкала состоит из набора утверждений, по которым респондент выражает свое отношение («полностью согласен», «согласен», «затрудняюсь ответить», «не согласен», «полностью не согласен»).

Шкала Осгуда (Osgood scale) применяется при определении различий в отношении к товарам, магазинам и иному путем указания респондентом мнения по поводу различных утверждений, касающихся предмета опроса в категориях «сложный – простой», «удобный – неудобный», «дорогой – дешевый». Шкала может содержать пять или семь позиций по мере возрастания (убывания) признака.

Ценностная шкала Рокича (Rokeach value scale) применяется для оценки поведения потребителей на основе измерения ценностей. Респондентам предлагается ранжировать по степени важности список из наборов задач, целей, а также наиболее предпоч-

тительных для них форм поведения в отношении приобретения ими товаров.

1.3. Базовые процедуры обработки данных и проблема выбора измерительной шкалы

Базовыми и наиболее распространенными процедурами обработки данных практически любого экспериментального исследования являются процедуры получения агрегированных и комплексных оценок. В эксперименте, основанном на выборочном исследовании, измеряются показатели отдельных элементов выборки. Набор таких измерений по каждому показателю может быть значительным (десятки, сотни, тысячи). Для анализа показателя по всей выборке необходимо *агрегировать* измерения, т.е. получить *агрегированные*¹ (коллективные, групповые, обобщенные) оценки. Получение агрегированных оценок должно выполняться корректно. Для величин, измеренных в метрических шкалах (интервальной и относительной), агрегирование, как правило, осуществляется путем вычисления среднего арифметического по группе. Для метрических шкал эта простая процедура вполне корректна. Наиболее распространенная ошибка (с точки зрения математической корректности) — использование арифметических действий для агрегирования данных в неметрических шкалах (чаще всего порядковых), например вычисление среднего балла успеваемости по отдельному предмету по студенческой группе. (Более подробно этот вопрос разбирается на примере шкалы Лайкерта в гл. 2.)

На практике также довольно часто встречаются ситуации, когда необходимо синтезировать отдельные показатели в некоторый интегрированный (*комплексный*) показатель. Например, показатель инновационности компании интегрируется на основе нескольких показателей, характеризующих ее организационный климат. Другим, более распространенным, примером является усреднение баллов успеваемости студента (или студенческой группы) по нескольким дисциплинам.

Очень часто, как и в случае с агрегированными оценками, исследователи используют арифметические действия для интегрирования показателей, не задумываясь о содержательной и математической строгости применяемых процедур.

¹ Лат. *aggregare* — присоединять.

Таким образом, при выборе методов статистической обработки полученных данных в первую очередь следует обращать внимание на тип измерительной шкалы. Идеальным относительно корректности использования математико-статистического аппарата является выбор только тех методов, которые могут применяться для обработки данных, представленных в соответствующих шкалах. Однако на практике многие математические операции используются для данных, измеренных в таких шкалах, для которых эти операции в принципе не применимы. В частности, такие основные статистические показатели, как среднее арифметическое, стандартное отклонение или коэффициент корреляции являются настолько привычными для представления результатов, что исследователь редко задумывается о корректности использования этих показателей для данных, полученных в разных шкалах.

Дифференциация номинативной шкалы от порядковой, а тем более от метрической шкалы, обычно не вызывает трудностей. Гораздо сложнее определить различие между другими шкалами (особенно порядковой и интервальной). Весьма распространенной является ошибка, когда исследователь интерпретирует данные так, как если бы они были измерены по шкале, допускающей более сложные операции, чем позволяет эта шкала в действительности. Наиболее часто данные порядковой шкалы интерпретируются как измерения в интервальной шкале, поскольку статистические процедуры для этой шкалы являются более информативными. (Далее будут предложены соответствующие примеры и задания для анализа.) Это, безусловно, создает большие трудности для анализа и интерпретации результатов исследования с точки зрения обоснования их корректности, поэтому исследователь должен проявлять осторожность при интерпретации и оценке результатов.

В гл. 2 мы рассмотрим подходы к обоснованию выбора шкал и конвертации данных из неметрических в метрические шкалы.

Задания для самостоятельного выполнения

1. Назовите главные различия номинативной, порядковой, интервальной и абсолютной измерительных шкал. Как эти различия влияют на выбор математических методов, которые могут быть использованы для обработки данных?

2. Назовите шкалу, которая может быть применена для измерения следующих признаков. Ответ обоснуйте.

- Регионы, в которых расположены офисы компании;
- количество сотрудников;
- доход компании;
- затраты на осуществление рекламной кампании;
- профессиональный статус (должность) в анкете;
- профессиональный статус (должность) как мера продвижения по службе;
- телефонный номер;
- время решения задачи;
- профессия;
- номера домов;
- сортность продукции;
- образование (начальное, среднее, высшее, ученая степень);
- образование (количество классов школы);
- возраст;
- год рождения;
- место рождения;
- успеваемость (средний балл);
- профилирующая дисциплина (специальность).

3. Ответьте, в какой шкале могут быть измерены характеристики следующих объектов: 1) стиральный порошок; 2) сотрудники фирмы; 3) деятельность компании:

а) вес, объем, цвет, запах, цена, способность удалять загрязнения, воздействие на цвет стираемых вещей, скорость отстирывания, воздействие на элементы стиральной машины и т. д.;

б) возраст, пол, цвет волос, семейный статус, мотивация, ответственность, квалификация, производительность труда и т. д.;

в) торговый оборот, суммарные затраты, сезонность продаж, наиболее успешно продаваемые товары и др.

Предложите не менее одной «новой» характеристики (не включенной в список) по каждому объекту и обоснуйте выбор шкалы для ее измерения.

4. Представьте данные о заработной плате каждого из сотрудников компании с помощью номинативной, порядковой, интервальной и абсолютной шкал. Проиллюстрируйте на конкретных примерах потерю информации при переходе от более мощных к менее мощным шкалам (абсолютная – интервальная – порядковая – номинативная).

Пояснение. Заработная плата, выраженная в «классическом» денежном выражении, представляется в шкале отношений. Для представления заработной платы в порядковой шкале можно, например, ранжировать список сотрудников в зависимости от оплаты их труда. Для представления в номинативной шкале можно, допустим, выделить сотрудников с повременной и фиксированной оплатой и т.д. Для представления заработной платы в интервальной шкале можно предложить шкалу с равенством интервалов, но условной нулевой точкой, т.е. такой, например, когда нельзя точно сказать, сколько получает конкретный сотрудник в денежном выражении, но можно судить о его заработной плате относительно «среднего уровня». Предложите собственный способ измерения заработной платы в разных шкалах.

5. Вы собираетесь подготовить и издать книгу «100 лучших предприятий города N». Предложите характеристики, которые вы будете измерять с указанием шкал. Будете ли вы использовать какой-либо комплексный показатель? Обоснуйте ответ.

6. Вам необходимо разработать анкету для опроса студентов, на основе которого можно было бы оценить качество преподаваемых курсов. Что именно вы бы оценивали? Какие вопросы могли быть включены в анкету? Какие шкалы вы бы использовали?

7. Маркетолог авиакомпании разрабатывает исследование для оценки мнения пассажиров. Какие вопросы могут быть включены в анкету? Предложите вопросы, отклики на которые могут быть измерены по каждой из четырех типов шкал.

8. Разрабатывается исследование для оценки мнения студентов о высшем учебном заведении. Какие вопросы могут быть включены в анкету? Предложите вопросы, ответы на которые могут быть измерены по каждой из четырех типов шкал.

Глава 2

Оценки и измерения

2.1. Различия между измерением и оценкой

В гл. 1 мы предложили достаточно широкое определение измерения как операции соотнесения признаков эмпирических объектов с помощью чисел в соответствии с правилами, кото-

рые определяются измерительными шкалами. Это определение не является единственным и бесспорным. Существует разделяемый многими исследователями подход, согласно которому измерение следует отличать от оценки – другого способа количественной характеристики величин. Различие измерения и оценки основано на используемой измерительной модели (включающей измерительную шкалу) и получаемых результатах.

Оценка дает приближенное значение величины или параметра, причем в отличие от измерения нет однозначного соответствия между величиной и ее количественным выражением в определенных единицах. Считается, что по исходным данным, представленным в неметрических шкалах, получаются *оценки* свойств исследуемого объекта, тогда как *измерения* могут быть получены при использовании либо метрических шкал, либо определенных процедур преобразования исходных данных в метрические шкалы.

Принято считать, что измерение в отличие от оценки характеризуется научной обоснованностью, корректностью использования математико-статистического аппарата, большей точностью, надежностью и объективностью, стандартизованностью и универсальностью, меньшей чувствительностью к смене наборов исходных данных и т.д. Хотя четкое и однозначное различие между оценками и измерениями провести достаточно сложно, в данной главе с целью демонстрации такого различия подробно рассмотрены два подхода к измерению: «измерение» на основе шкалы Лайкерта и измерение с использованием модели Раша. Можно считать, что первый подход продуцирует оценки, а второй – именно измерения.

2.2. Шкала Лайкерта

Шкала Лайкерта (Likert scale) впервые была предложена в 1932 г. американским специалистом в областях организационной психологии и управления Ренсисом Лайкертом (Rensis Likert)¹. С тех пор эта шкала широко применяется для измерения отношений респондентов к предмету исследования.

Данный подход предполагает, что респондент выражает свое согласие или несогласие с каждым суждением (из предложенного

¹ Вариантом произношения фамилии является Ликерт.

набора) по шкале оценок (из трех-семи или более пунктов), а его место на итоговой шкале определяется суммой оценок каждого отдельного суждения. Лайкерт обосновал предположение о самом существовании измеряемой однородной латентной переменной и показал, что его шкала приемлема для ее измерения.

Хотя «классическая» шкала Лайкерта включает пять пунктов, сходные шкалы с другим количеством позиций шкалы (3, 4, 6, 7,...) также называют шкалами Лайкерта; чаще всего используется 5- или 7-позиционная шкала. Шкала может содержать или не содержать «нейтральную точку» (при нечетном числе позиций). Частным случаем шкалы Лайкерта является дихотомическая шкала (две позиции).

Оценки по шкале Лайкерта могут быть выражены в категориях согласия, частоты, важности, качества и т.п., например:

(5) Полностью согласен – (4) Согласен – (3) Нейтральное отношение – (2) Не согласен – (1) Полностью не согласен;

(5) Всегда – (4) Очень часто – (3) Иногда – (2) Редко – (1) Никогда;

(4) Очень важно – (3) Важно – (2) В некоторой степени важно – (1) Неважно;

(5) Очень хорошо – (4) Хорошо – (3) Приемлемо – (2) Плохо – (1) Очень плохо.

Позиции шкалы Лайкерта обычно кодируются числами от 1 до 5 (для 5-позиционной шкалы), от 1 до 6 (для 6-позиционной шкалы) и т.д., хотя кодирование может включать отрицательные и нулевое значения (например, для нейтральной позиции), что, впрочем, используется нечасто. Примеры кодировки позиций шкалы Лайкерта приведены на рис. 2.1.

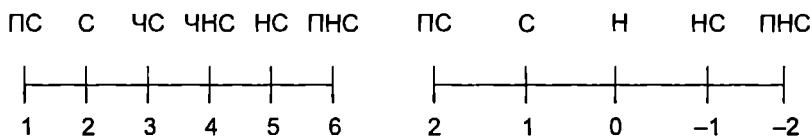


Рис. 2.1. Примеры кодировки позиций шкалы Лайкерта:

ПС – полностью согласен; С – согласен; ЧС – частично согласен;

ЧНС – частично не согласен; НС – не согласен;

ПНС – полностью не согласен;

Н – нейтральное отношение

Направление кодирования зависит от формулировки вопроса. Для позитивно сформулированных утверждений (например, «Творческие идеи в данной организации поощряются») позитивно выраженное отношение (например, «Согласен») кодируется, как правило, большим числом (например, 5), а негативно выраженное («Не согласен») – меньшим. При негативной формулировке (например, «Мой руководитель игнорирует мои предложения» или «Преподаватель дает слишком много заданий на дом») позитивно выраженное отношение может кодироваться меньшим числом, а негативно выраженное – большим.

В случае, если анкета содержит как прямые, так и обратные вопросы или утверждения (например, «Мой руководитель поддерживает мои инициативы» и «Мой руководитель не поддерживает мои инициативы»), отклики на обратные утверждения перекодируются по формуле

$$r = (\min - s) + \max,$$

где r – отклик на прямые утверждения (вопросы);
 s – отклик на обратные утверждения (вопросы);
 \min, \max – минимальное и максимальные значения шкалы.

Агрегированная оценка составит, например $(4 + 3 + 5)/3 = 4$. Средний «словесный» результат (в данном случае «Согласен») сохранится при любом линейном преобразовании численных значений пунктов шкалы. На первый взгляд никаких изъянов в таком подходе к агрегированию нет. Выше мы уже обсуждали, что агрегирование оценок на основе вычисления среднего арифметического корректно лишь для метрических шкал. Может ли шкала Лайкерта рассматриваться как метрическая? Другими словами, могут ли измерения в шкале Лайкерта по крайней мере интерпретироваться как интервальные?

В шкале Лайкерта очень наглядно проявляется дилемма между простотой измерительных процедур и математической строгостью. В литературе встречаются и аргументируются различные точки зрения на возможности использования шкалы Лайкерта. Эта шкала, как показал еще ее автор, является порядковой (т.е. неметрической) шкалой, однако многие исследователи считают возможным использовать ее как интервальную шкалу и применять процедуры для обработки данных, приемлемые для метрических шкал, так как, по их мнению, это не приводит к серьез-

ным ошибкам в результатах. Основными аргументами в пользу такого утверждения являются следующие *предсуждения*:

- интервалы между позициями шкалы Лайкерта в отдельном пункте (вопросе или утверждении) анкеты приблизительно одинаковы;

- относительные интервалы между позициями шкалы Лайкерта изменяются незначительно при переходе от вопроса к вопросу (от утверждения к утверждению).

Если эти два условия можно считать выполненными, то к шкале Лайкерта могут быть применены процедуры, приемлемые для интервальных шкал.

Посмотрим, что означает предпосылка о равенстве интервалов в 5-позиционной шкале Лайкерта (рис. 2.2).

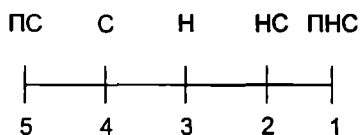


Рис. 2.2. Пример 5-позиционной шкалы Лайкерта:

ПС – полностью согласен; С – согласен; Н – нейтральное отношение; НС – не согласен; ПНС – полностью не согласен

Равенство межпозиционных интервалов должно означать следующее:

$$\begin{aligned} \text{ПС} - \text{С} &= \text{С} - \text{Н} = \text{Н} - \text{НС} = \text{НС} - \text{ПНС} = \\ &= 5 - 4 = 4 - 3 = 3 - 2 = 2 - 1 = 1. \end{aligned}$$

Пусть мы имеем в качестве пункта анкеты следующее утверждение: «Члены моего коллектива открыты новым идеям». Для респондента А, который в целом согласен с этим утверждением, интервалы между позициями шкалы могут быть такими, как, например, на рис. 2.3, а.

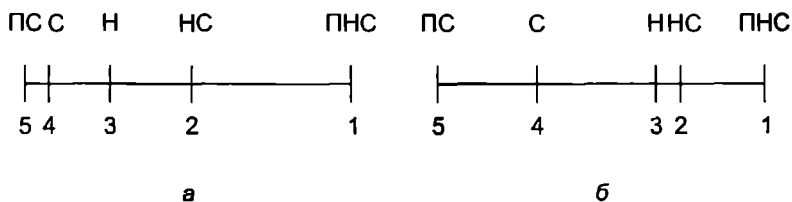


Рис. 2.3. Пример неравенства интервалов в шкале Лайкерта:

а – случай согласия респондентов; б – случай несогласия респондентов

В этом случае респондент *A* скорее выберет одну из позиций – ПС или С, нежели позицию Н, а тем более ПНС, т.е. в этом случае говорить о равенстве интервалов (даже приближительном) не приходится. Предположим, что некий респондент *B* скорее не согласен с предложенным утверждением. Для него интервалы между позициями шкалы могут быть такими, как на рис. 2.3,б.

Интервал между позициями Н и ПНС для респондента *B* намного меньше, чем интервалы между другими позициями. Очевидно, что вычисление арифметического среднего для данного случая некорректно.

Аналогично складывается ситуация и с получением комплексных оценок в шкале Лайкерта. Предположим, что измеряемым показателем является «Поддержка творческой инициативы сотрудников руководством компании». Анкета включает два пункта: «Мой руководитель открыт для новых идей» и «Мой руководитель поддерживает мои новые предложения». Очевидно, что эти пункты имеют разный «вес». Поэтому отклик «Согласен» при ответе на первый пункт может не соответствовать позиции «Согласен» при отклике на второй пункт (одно дело – просто воспринимать новые идеи, другое – их поддерживать). Возможно, «вес» позиции «Согласен» (4) в первом пункте соответствует нейтральной позиции (3) во втором пункте (рис. 2.4).

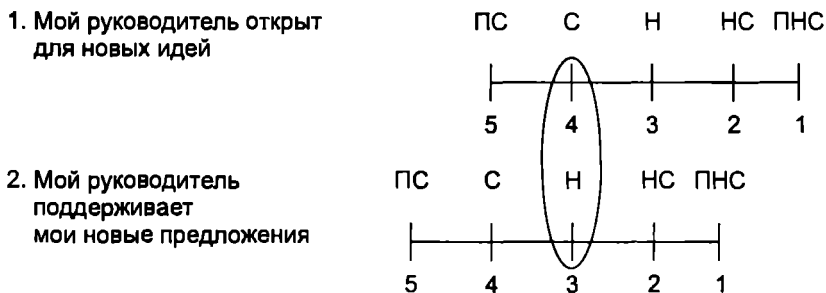


Рис. 2.4. Пример неэквивалентности значений позиций шкалы Лайкерта для разных пунктов анкеты

Еще более рельефно неэквивалентность значений позиций шкалы при переходе от пункта к пункту при получении комплексных оценок в шкале Лайкерта демонстрирует пример, приведенный в книге Т. Бонда и К. Фокс [35].

Приведем еще один пример из этой книги.

Оценивается отношение школьников к работе за компьютером. В анкете используются несколько пунктов, в том числе: «Я настолько боюсь компьютера, что стараюсь избегать пользоваться им» и «Я опасюсь, что сделаю ошибку при работе за компьютером». Школьник, выбирающий ответ «Полностью согласен» по второму пункту, может выбрать ответ «Не согласен» по первому, и комплексная оценка, получаемая традиционным способом, будет равна семи, хотя очевидно, что «вклад» пунктов в общую оценку не равноценен. Школьник, выбравший ответы «Полностью согласен» по первому пункту и «Не согласен» по второму, получит те же семь баллов, что кажется нелогичным. Более реалистичную оценку можно было бы получить, «выравнивая вес» этих пунктов, например, следующим образом: 2 в первом пункте равно 5 во втором (рис. 2.5).

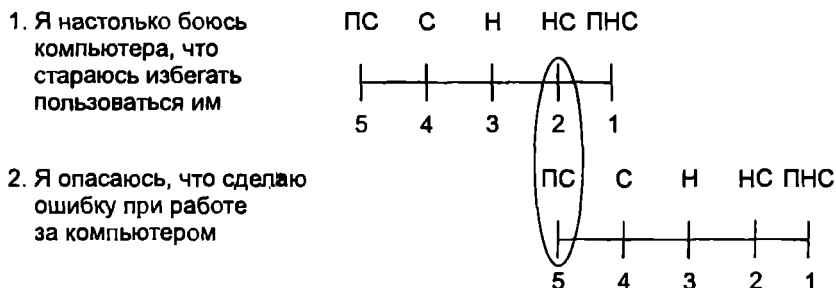


Рис. 2.5. Пример «выравнивания» пунктов при получении комплексной оценки в шкале Лайкерта

Таким образом, в реальной исследовательской практике достаточно сложно обеспечить выполнение двух основных условий (равенство межпозиционных интервалов шкалы и эквивалентность позиций шкалы при переходе от пункта к пункту) для того, чтобы шкалу Лайкерта можно было бы считать интервальной. Даже для приблизительного выполнения этих условий «веса» пунктов анкеты должны быть примерно равны, а отношения респондентов примерно одинаковы. Если обосновать это не удастся, для обработки данных должны использоваться методы, применимые для порядковых шкал, либо использоваться процедуры преобразования данных в метрические шкалы.

2.3. Возможности преобразования данных из неметрических шкал в метрические

При использовании статистических методов требуется обосновать правомерность применения выбранной шкалы, поэтому исследователь стремится найти свидетельства того, что его шкала может рассматриваться как метрическая. Почти повсеместно используемые в исследовательской практике подходы к получению агрегированных и комплексных оценок по шкале Лайкерта путем вычисления арифметического среднего работают с практической точки зрения, но лишены смысла с точки зрения математической, поскольку эта шкала не интервальная. С другой стороны, для получения корректных результатов исследования необходимо обеспечить корректность получаемых измерений и применяемых процедур обработки. В то же время шкала Лайкерта является простым и удобным инструментом для сбора данных. Возможен ли компромисс между практическим удобством и математической строгостью? Возможно ли значимое обоснование данных в шкале Лайкерта как данных интервальной шкалы? Возможно ли некое преобразование данных из шкалы Лайкерта в метрическую шкалу?

Одним из самых «сильных» аргументов в пользу метрического характера шкалы является распределение измерений по нормальному закону¹. Исходным принципом здесь служит предположение о том, что измеряемое свойство распределено в генеральной совокупности в соответствии с нормальным законом. Следовательно, измерение данного свойства также должно обеспечивать нормальное распределение. Если это так, то шкала может считаться по крайней мере интервальной [18]. Если это не так, то измеряемое свойство отражается в не метрической шкале, а, например, в порядковой. Таким образом, основной подход к разработке и анализу измерительной шкалы заключается в обеспечении и проверке нормального распределения полученных измерений. Если выборочное распределение не отличается от нормального, то, по мнению некоторых специалистов,

¹ При дальнейшем изложении предполагается, что читатели знакомы с нормальным законом распределения. При необходимости пополнить свои знания в этой области следует обратиться к любому вузовскому учебнику по теории вероятностей и математической статистике.

измеряемое свойство отражается в интервальной шкале. Однако такой подход к обоснованию метрического характера измерений принимается не всеми исследователями. Например, А.И. Орлов в одной из своих работ показывает, что в большинстве случаев распределения результатов измерения существенно отличаются от нормальных [23].

Причиной отклонения формы выборочного распределения признака от нормального вида чаще всего является особенность процедуры измерения: используемая шкала может обладать неравномерной чувствительностью к измеряемому свойству в разных частях диапазона его изменчивости. Для проверки нормальности распределения данных используются различные процедуры, позволяющие выяснить, отличается ли от нормального выборочное распределение измеренной переменной. Необходимость такого сравнения особенно актуальна, когда возникают сомнения, в какой шкале представлен признак – порядковой или интервальной. Это важно, поскольку от этого зависит возможность использования методов для дальнейшего анализа. Метрические шкалы позволяют использовать гораздо более широкий диапазон методов, чем порядковые, поэтому и исследовательские выводы могут быть более глубокими и информативными. При значительном отклонении эмпирического распределения от нормального следует отказаться от предположения о том, что признак измерен в метрической шкале. Однако остается вопрос о *степени значительности* отклонения распределений измерения от нормального закона.

На практике используются два основных способа проверки нормальности распределения данных:

- визуально-графический способ (нормальность оценивается по различным диаграммам и графикам частот встречаемости признака);
- статистические критерии нормальности (Колмогорова–Смирнова, Шапиро–Уилка др.).

Процедуры проверки нормальности распределения включены в статистические пакеты SPSS, STATISTICA и др. (подразд. 3.4).

Если исходные данные не проходят проверку на соответствие нормальному распределению, можно их нормализовать, т.е. привести к нормальному виду. Наиболее простым и распространенным нормализующим преобразованием исходных данных является приведение данных к так называемому стандарт-

ному z -распределению со средним, равным 0, и стандартным отклонением, равным 1:

$$z = \frac{r - m}{s},$$

где r – исходные данные;
 m – среднее значение по признаку;
 s – стандартное отклонение по признаку.

Допустим, мы имеем по некоторому измеряемому признаку 10 измерений, полученных, например, по шкале Лайкерта (табл. 2.1).

Таблица 2.1

Исходные данные в шкале Лайкерта

№	1	2	3	4	5	6	7	8	9	10
r	3	4	4	2	3	4	5	2	3	4

Среднее значение по имеющимся исходным данным составляет 3,4; стандартное отклонение – 0,966. Этого достаточно для нормализации данных (табл. 2.2).

Таблица 2.2

Нормализованные данные (z -шкала)

№	1	2	3	4	5	6	7	8	9	10
r	-0,41	0,62	0,62	-1,45	-0,41	0,62	1,66	-1,45	-0,41	0,62

Среднее значение по нормализованным данным равно 0; стандартное отклонение – 1. В пакете SPSS при вычислении базовых статистик (*Analyze, Descriptive Statistics, Descriptive*) можно выбрать опцию перевода исходных данных в нормализованную z -шкалу (*Save standardized values as variables*).

В эмпирических исследованиях широко используется также *стандартная T-шкала*, связанная с z -шкалой простым преобразованием:

$$T = 10z + 50.$$

T -распределение имеет среднее, равное 50, и стандартное отклонение, равное 10.

Описанные подходы к нормализации данных имеют принципиальный недостаток с точки зрения математической строгости.

Они изначально используют процедуры, неприменимые для порядковых шкал (определение среднего значения и стандартного отклонения) по причинам, которые подробно обсуждались выше. Однако исследователи «забывают» о такой «мелочи», и обычно считается, что нормализованные данные, полученные в порядковых шкалах, можно рассматривать как интервальные со всеми вытекающими последствиями по применению методов обработки данных.

Еще одним распространенным подходом к преобразованию порядковых шкал в интервальные является замена интервалов на их средние значения. Предположим, что имеются данные о среднемесячном доходе, представленные в порядковой шкале с пятью позициями:

- 1 – до 4000 руб.;
- 2 – от 4000 до 8000 руб.;
- 3 – от 8000 до 16000 руб.;
- 4 – от 16000 до 24000 руб.;
- 5 – свыше 24000 руб.

В таком случае мы можем перейти, например, к шкале, имеющей свойства абсолютной: 2000, 6000, 12000, 20000, 30000. Очевидно, что подобный подход имеет существенные ограничения. Он применим лишь в случаях, когда содержательный смысл измеряемых величин (денежные величины, временные периоды, размеры, количество определенных товаров и т.п.) позволяет осуществить такое преобразование, т.е. когда изначально имеются признаки, которые могли бы быть измерены в метрических шкалах, но по каким-то причинам (например, этическим или в целях упрощения анкеты и процесса обработки данных) они были представлены в порядковой шкале. Тем не менее подобный подход позволяет корректно использовать методы обработки данных, которые неприменимы к порядковым шкалам. Однако точность такого преобразования и полученных результатов существенно зависит от выбора позиций порядковой шкалы.

2.4. Модель Раша

Процедуру более корректного преобразования данных из шкалы Лайкерта в интервальную шкалу можно осуществить на основе *модели Раша* (*Rasch model*), предложенной датским мате-

матиком Г. Рашем еще в конце 1950-х гг., хотя понимание и признание среди исследователей и специалистов эта модель получила лишь в 1970 – 1980-х гг. Первоначально модель Раша была разработана для оценки способностей, знаний, умений, подготовленности и т.п. (*abilities*), позднее ее также начали использовать для оценки отношений (*attitudes*) и восприятий (*perceptions*). В настоящее время модель Раша наиболее активно применяется в образовательно-педагогической практике (в инструментах тестирования), но имеет серьезные перспективы для социально-экономических и гуманитарных исследований.

Эта модель была разработана в связи с обнаружившимися проблемами представления данных в порядковых шкалах. Во-первых, эти шкалы не обеспечивают линейности (пропорциональности) в представлении данных, а во-вторых, получаемые данные субъективны и ситуативны в том смысле, что они зависят от конкретных респондентов и конкретных вопросов. Изменение набора респондентов и/или набора вопросов в анкете может привести к другим результатам. Г. Раш предложил сравнительно простую модель для «объективного» измерения, решающую эти две проблемы.

В контексте измерения способностей ключевая идея модели Раша может быть сформулирована следующим образом: вероятность правильного ответа на вопрос или задание теста (P) является функцией способностей (знаний) тестируемого и зависит от способностей (знаний) тестируемого (B) и сложности вопроса (D):

$$P(B) = \frac{e^{B-D}}{1 + e^{B-D}}.$$

В контексте измерения отношений (восприятий, оценок) эта идея может быть интерпретирована следующим образом. Вероятность выбора (позитивного отклика) определенной позиции пункта анкеты (*endorsement*) зависит от отношения (*attitude*) респондента к пункту (вопросу или утверждению) анкеты и от степени соответствия (*agreeability*) этого пункта и позиции «действительному положению вещей».

Эта базовая идея может быть проиллюстрирована с помощью *характеристической кривой* заданий теста (или пунктов анкеты). Пусть в тесте Задание 2 сложнее Задания 1. Тогда при

той же подготовленности испытуемого (или при тех же самых знаниях) вероятность правильного ответа на Задание 2 будет меньше, чем вероятность правильного ответа на Задание 1.

Аналогично в случае измерения отношений, если в анкете Пункт 2 объективно сложнее для выбора респондента, чем Пункт 1, то при одном и том же отношении респондента к предмету исследования вероятность выбора Пункта 2 будет меньше, чем вероятность выбора Пункта 1 (рис. 2.6).

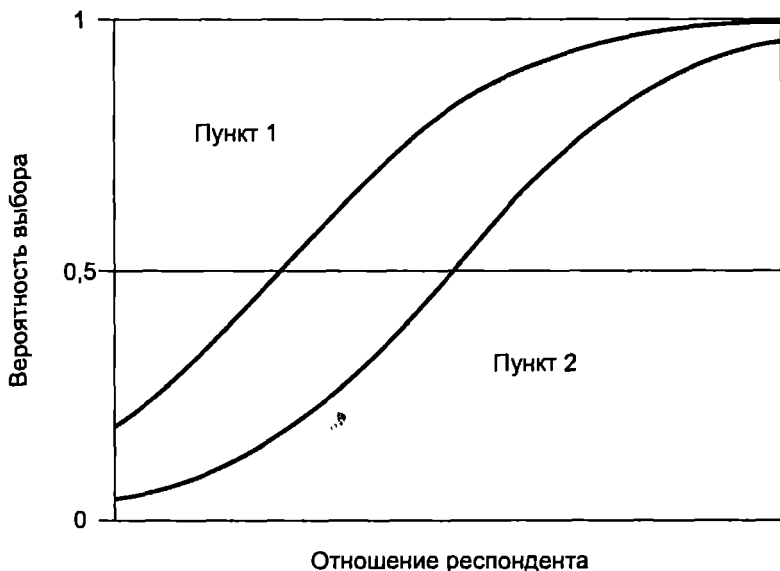


Рис. 2.6. Характеристическая кривая отклика респондента

Единицей шкалы Раша является «логит» (d):

$$d = \ln\left(\frac{P}{1-P}\right),$$

где P – вероятность правильного ответа на вопрос или (в контексте измерения отношений) вероятность выбора пункта (позиции) анкеты.

Чем больше вероятность того, что испытуемый правильно ответит на вопрос, тем большее значение по шкале Раша имеет ответ. На практике при получении данных по шкале Раша вместо

вероятностей используют частоты откликов, т.е. чем чаще тестируемый отвечает правильно, тем выше его показатель по шкале Раша.

Если подставить в эту формулу (модели Раша) значения вероятности успешного (правильного) ответа на вопрос, то получим

$$d = B - D.$$

Этот результат показывает, что измерения по шкале Раша зависят только от способности испытуемого и сложности вопроса, т.е. мы получаем измерение, независимое от выборки испытуемых или респондентов (*person free measurement*) и набора вопросов теста или пунктов анкеты (*item free calibration*). Это одно из важнейших свойств шкалы Раша наряду с интервальностью. Независимость измерения от теста и респондентов (в пределах ошибки измерения) позволяет специалистам в области теории измерений называть шкалу Раша «объективной» измерительной шкалой и сравнивать со шкалами, используемыми в точных и естественных науках, например в физике.

Для определения сложности задания теста применяется следующая формула:

$$d = \ln\left(\frac{1 - P}{P}\right).$$

Следовательно, чем проще задание, тем меньшее значение оно имеет по шкале Раша.

Некоторую сложность может вызвать интерпретация данных по шкале Раша. Значение 0 по шкале Раша отражает ситуацию, когда знания (способности) тестируемого соответствуют сложности задачи (вопроса), а вероятность правильного ответа на вопрос составляет 50%. При уменьшении сложности вопроса на один логит вероятность правильного ответа возрастает приблизительно до 75%. При увеличении сложности вопроса на один логит вероятность правильного ответа уменьшается примерно до 25%.

Иначе значение 0 можно интерпретировать следующим образом: 50% тестируемых правильно отвечают на вопрос, 50% — неправильно. Значение $-2,94$ соответствует «простому» вопросу, на который 95% тестируемых отвечают правильно и лишь 5% — неправильно. Значение $+2,94$, напротив, соответствует «сложному» вопросу, на который лишь 5% тестируемых отвечают правильно (рис. 2.7).

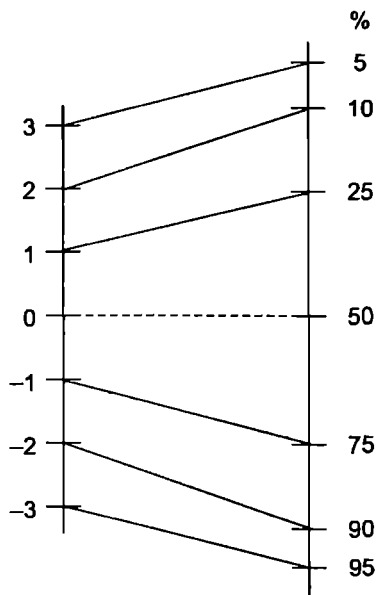


Рис. 2.7. Интерпретация шкалы Раша

При измерении отношений значения шкалы Раша можно интерпретировать аналогично. Например, при оценке организационного климата для творчества и инноваций в компании значение 0 соответствует «нейтральному» климату (50% положительных и 50% отрицательных откликов), значение $-2,94$ соответствует 95% положительных и 5% отрицательных откликов и т.д. Иначе значение $-2,94$ может быть интерпретировано как показатель организационного климата, в котором 95% положительных факторов для творчества и инноваций и лишь 5% отрицательных факторов. Значение $2,94$ соответствует резко отрицательному («холодному») климату для творчества и инноваций. В «классической» интерпретации значений шкалы Раша чаще встречаемый признак (с большим количеством баллов) считается более «легким» и имеет меньшее значение. Для более «естественной» интерпретации результатов измерения отношений (как в рассмотренном случае оценки организационного климата) значения шкалы Раша для пунктов анкеты следует умножить на -1 . Этот пример показывает, что результат интерпретации зависит от предмета исследования, формулировки вопросов или утверж-

дений (в «негативной» или «позитивной» форме), а также от прямого или обратного кодирования позиций в исходной шкале.

Модель Раша обладает рядом преимуществ по сравнению с порядковыми шкалами, в частности шкалой Лайкерта:

- модель обеспечивает калибровку пунктов анкеты в соответствии с их «весом» («сложностью»);

- «вес» пунктов анкеты может быть сравнительно просто вычислен на основе частотности откликов;

- единица измерения шкалы Раша, представляющая собой результат преобразования исходных данных, имеет стандартизованный характер, что дает возможность объективно сравнивать результаты, полученные с помощью разных инструментов и по разным выборкам;

- данные представляются в интервальной шкале;

- определяется статистика соответствия (*fit statistics*) каждого пункта анкеты и откликов каждого респондента измерительной модели, что является чрезвычайно важным инструментом оценки точности результатов измерений, существенно дополняющим классические методы (гл. 4 данного учебного пособия);

- пункты анкеты, которые каким-либо респондентом оставлены без ответа, не представляют трудностей при использовании модели Раша, т.е. пропуски заданий (пунктов анкеты) не критичны для обработки данных;

- измерения по модели Раша не зависят от выборки (*sample free*) и набора пунктов анкеты (*item free*), т.е. теоретически они не зависят от распределения индивидуальных откликов и набора вопросов. Это позволяет получать объективированные результаты, что, во-первых, положительно отражается на качестве исследовательских результатов, а во-вторых, отвечает принципу объективности как одного из критериев научного исследования. По этой причине измерения, основанные на модели Раша, называют «объективными измерениями» (*objective measurement*) и даже «фундаментальными измерениями» (*fundamental measurement*).

Разработаны специальные программные продукты для конвертирования исходных данных в шкалу Раша и выполнения всех необходимых расчетов (программы WINSTEPS, RUMM, BiLog, MultiLog и др.). Процедуры, реализованные в этих программах, достаточно сложны и громоздки. Для того чтобы проиллюстрировать общий подход к реализации подобных процедур, рассмотрим простейший алгоритм преобразования исходных данных в шкалу Раша.

Пример 2.1. В дихотомической шкале измерение представлено двумя возможными состояниями (1 и 0). Примером такой шкалы может являться некий тест на знание учебного предмета, в котором испытуемому требуется выбрать правильный ответ из двух вариантов по каждому вопросу: правильный ответ кодируется 1, неправильный – 0.

Предположим, что в тест включено семь заданий (вопросов): 1, 2, ..., 7. Допустим, что на вопросы (задания) ответили семь испытуемых, которых обозначим латинскими буквами: *A*, *B*, ..., *G*. Исходные данные представлены в табл. 2.3.

Таблица 2.3

Исходные данные для конвертирования в шкалу Раша

Испытуемый	Номер задания						
	1	2	3	4	5	6	7
<i>A</i>	1	0	1	1	0	1	0
<i>B</i>	0	1	0	1	1	0	0
<i>C</i>	0	1	1	1	1	1	0
<i>D</i>	0	1	1	1	1	1	1
<i>E</i>	1	0	1	0	1	1	0
<i>F</i>	0	0	1	1	0	1	1
<i>G</i>	1	1	1	1	1	0	0

Для конвертирования исходных данных в шкалу Раша необходимо определить общую сумму баллов (S_1), набранных испытуемыми, и общую сумму баллов по каждому пункту теста (S_2). Далее определяется максимальное значение, которое *мог бы* набрать испытуемый ($Max1$): $Max1 = 1 \cdot 7 = 7$. В контексте тестового исследования (измерения способностей, знаний, подготовленности и т.д.) для дихотомической шкалы это значение соответствует количеству вопросов, на каждый из которых может быть дан правильный ответ (7). Отношение S_1 к $Max1$ может быть интерпретировано (при большом количестве пунктов теста) как частота (или вероятность) правильного ответа (P_1). Это значение определяется для каждого испытуемого. В итоге для каждого испытуемого определяется его балл по шкале Раша $D_1 = \ln(P_1/(1 - P_1))$. Например, при проверке знаний или подготовленности тестируемого более высокие значения D_1 соответствуют лучшей подготовленности, и наоборот.

Аналогично вычисляется максимальное количество баллов, которое может быть приписано каждому заданию теста (Max_2). В данном случае количество пунктов равно 7, максимальное значение по каждому заданию равно 1, поэтому $Max_2 = 1 \cdot 7 = 7$. Далее для каждого пункта вычисляется отношение $P_2 = S_2 / Max_2$ и значение $D_2 = \ln((1 - P_2) / P_2)$. Обратите внимание на то, что в последней формуле числитель и знаменатель поменялись местами (что эквивалентно смене знака перед логарифмом). В классическом варианте модели Раша, которая разрабатывалась в контексте именно тестовых исследований, это было сделано для более «естественной» интерпретации результатов. В таком случае более высокие значения D_2 соответствуют большей сложности тестового задания (вопроса), и наоборот. Следовательно, чем сложнее вопрос, тем меньше вероятность правильного ответа на него и тем меньшее количество испытуемых отвечает на него правильно. В табл. 2.4 приведены результаты конвертирования исходных данных в шкалу Раша.

Подобный подход можно распространить на случай измерения отношений. Дихотомические шкалы нередко используются в социологических, маркетинговых и иных опросах, например:

«Я планирую приобрести ноутбук (... название) в течение ближайших 6 мес. Да Нет».

Если в таком случае положительный отклик респондента («Да») будет кодироваться как 1, а отрицательный («Нет») как 0, то большие значения для респондента по шкале Раша будут соответствовать его более позитивному (благоприятному) отношению к исследуемому явлению (объекту). В связи с этим рассмотренный подход к обработке данных в дихотомической шкале может быть перенесен и в сферу социально-экономических исследований. В то же время в контексте измерения отношений логичнее получать более высокие значения по шкале Раша для пунктов, по которым большим количеством респондентов выражено благоприятное отношение к предмету исследования. В такой ситуации более логичным представляется использовать для оценки пунктов (индикаторов) «прямую» формулу перевода суммарных откликов в логиты, а именно $D_2 = \ln(P_2 / (1 - P_2))$. В частности, такой подход целесообразен при оценке организационно-психологических условий для работы сотрудников организации.

Результаты конвертирования исходных данных

Испытуемый	Задания теста							S _i	P _i	D _i
	1	2	3	4	5	6	7			
A	1	0	1	1	0	1	0	4	0,57	0,29
B	0	1	0	1	1	0	0	3	0,43	-0,29
C	0	1	1	1	1	1	0	5	0,71	0,92
D	0	1	1	1	1	1	1	6	0,86	1,79
E	1	0	1	0	1	1	0	4	0,57	0,29
F	0	0	1	1	0	1	1	4	0,57	0,29
G	1	1	1	1	1	0	0	5	0,71	0,92
S ₂	3	4	6	6	5	5	2			
P ₂	0,43	0,57	0,86	0,86	0,71	0,71	0,29			
D ₂	0,29	-0,29	-1,79	-1,79	-0,92	-0,92	0,92			

Такой подход к конвертированию данных может быть расширен и на многопозиционные порядковые шкалы, в частности на шкалу Лайкерта. В отличие от дихотомической шкалы эта шкала имеет не один пороговый переход «правильно—неправильно», а несколько, например переходы между позициями «всегда» и «часто», «часто» и «иногда», «иногда» и «никогда». Переход от одной позиции к другой может рассматриваться как более или менее «сложный» для респондента, что отражает степень соответствия (*agreeability*) позиции действительному положению вещей.

Например, первый пороговый переход моделируется как вероятность выбора 1 («всегда») вместо 0 («часто») по каждому пункту анкеты:

$$P_{i1}(x = 1 / B, D_i, F_1) = \frac{e^{(B - [D_i + F_1])}}{1 + e^{(B - [D_i + F_1])}},$$

где P_{i1} — вероятность выбора респондентом позиции «всегда» (1) вместо «часто» (0) по i -му пункту анкеты;

D_i — «сложность» пункта анкеты;

F_1 — «сложность» первого порогового перехода.

Пример 2.2. Предположим, что в анкету включено семь пунктов (1, 2, ..., 7), представляющих собой утверждение, по которому респон-

дент выражает степень своего согласия/несогласия. Для оценки откликов используется 5-позиционная шкала Лайкерта с позициями «Полностью не согласен» (ПНС), «Не согласен» (НС), «Нейтральное отношение» (Н), «Согласен» (С), «Полностью согласен» (ПС). Позиции шкалы кодируются от 1 до 5. Пусть на вопросы анкеты ответили 40 респондентов. Исходные данные (частоты откликов по каждому пункту) представлены в табл. 2.5.

Таблица 2.5

Отклики респондентов на пункты анкеты

Пункт анкеты	ПНС (1)	НС (2)	Н (3)	С (4)	ПС (5)	S	P	D	D_1
1	5	8	11	10	6	124	0,62	-0,49	0,49
2	9	7	10	9	5	114	0,57	-0,28	0,28
3	9	8	9	10	4	112	0,56	-0,24	0,24
4	7	9	9	8	7	119	0,595	-0,38	0,38
5	3	4	11	12	10	142	0,71	-0,90	0,90
6	5	9	9	9	8	126	0,63	-0,53	0,53
7	12	13	8	4	3	93	0,465	0,14	-0,14

Из таблицы видно, что при ответе на первый пункт анкеты 5 респондентов выбрали ПНС, 8 респондентов – НС, 11 респондентов – Н, 10 респондентов – С, 6 респондентов – ПС и т.д.

В наиболее простом варианте алгоритм перевода этих данных в шкалу Раша следующий. Рассчитывается балл соответствующего пункта (*item counts*) как сумма произведений количества респондентов, выбравших соответствующую категорию, и «балл» этой категории (S). Например, для первого пункта: $5 \cdot 1 + 8 \cdot 2 + 11 \cdot 3 + 10 \cdot 4 + 6 \cdot 5 = 124$. Максимальный балл для каждого пункта составляет: $40 \cdot 5 = 200$. Тогда «вероятность» выбора первого пункта составит: $P = 124/200 = 0,62$. Значение этого пункта по шкале Раша (в классическом варианте) определяется как $D = \ln((1 - P)/P) = -0,49$. Аналогично вычисляются агрегированные оценки для всех пунктов анкеты.

Здесь мы вновь сталкиваемся с вопросом об интерпретации результатов. В данном случае пункты, оцененные большим баллом, имеют меньшее значение, что полностью соответствует классической («тестовой») интерпретации измерений на основе

модели Раша. В ситуациях измерения латентных переменных через фиксацию откликов респондентов, если большей степени измеряемого свойства соответствует большее значение в порядковой шкале, более удобным и понятным является использование формулы

$$D = \ln(P / (1 - P))$$

для преобразования данных в шкалу Раша. Результаты измерения будут одинаковыми по абсолютной величине, но отличаться знаком (значения D_1 в табл. 2.5).

Те пункты, по которым не получено отклика от респондентов, не учитываются при подсчете максимально возможного балла для каждого пункта. Например, если по первому пункту один из респондентов не выразил своего мнения, то значение максимального балла для этого пункта составляет не $40 - 5 = 200$, а $39 - 5 = 195$. В остальном процедура конвертирования данных не отличается от описанной выше.

В этом примере мы определяли лишь агрегированные оценки для пунктов анкеты и не рассчитывали значения для каждого респондента и для каждой конкретной позиции каждого пункта. Более сложный подход³к преобразованию данных в шкалу Раша позволяет учесть и эти показатели, что необходимо, в частности, для анализа надежности откликов респондентов и пунктов шкалы (реализовано в программе WINSTEPS).

На основе модели Раша были разработаны другие измерительные модели, включающие большее количество параметров. Например, известна двухпараметрическая модель, в которой параметр a соответствует кривизне характеристической кривой, т.е. дифференцирующей способности тестового задания:

$$P_i(B) = \frac{1}{1 + e^{-a(B-D)}}.$$

Трехпараметрическая модель, предложенная А. Бирнбаумом, включает еще один дополнительный параметр (часто называемый «параметром угадывания»), определяющий точку пересечения характеристической кривой с осью ординат:

$$P_i(B) = c_j + (1 - c_j) \frac{1}{1 + e^{-a(B-D)}}.$$

О преимуществах и недостатках двух- и трехпараметрических моделей в сравнении с классической моделью Раша до сих пор спорят специалисты в области теории измерений (см., например, [1; 2; 19; 45; 55; 65; 67]).

2.5. Реализация модели Раша в программе WINSTEPS

Рассмотрим процедуру обработки исходных данных на основе модели Раша с помощью программы WINSTEPS (демонстрационная версия этой программы доступна на сайте www.winsteps.com). В этой программе используются более сложные алгоритмы, чем те, что были проиллюстрированы выше на простых примерах. Реализованные алгоритмы основаны на итерационных процедурах для подбора и оценки параметров модели Раша [55].

Пример 2.3. Рассмотрим тест, в котором данные были представлены в дихотомической шкале (см. табл. 2.4). Исходный файл для работы в программе WINSTEPS имеет достаточно сложный формат, поэтому мы покажем, как этот файл может быть сформирован на основе данных, предварительно представленных в программе SPSS, с помощью конвертера, включенного в программу WINSTEPS. Исходные данные могут быть перенесены в SPSS из MS Excel или MS Word обычным копированием.

Структура данных должна быть следующей: столбцы массива соответствуют испытуемым, а строки — заданиям. Предположим, что исходные данные перенесены в программу SPSS и переменные, соответствующие заданиям, обозначены как *item1*, *item2*, ..., *item7*. Файл сохранен под именем *Rasch-binary.sav*. После запуска программы WINSTEPS необходимо на вопрос *Would you like help setting up your analysis?* ответить *No*. Затем из пункта меню SPSS (или SPSS/SAS) выбрать опцию *Select SPSS file and variables*. В диалоговом окне выбрать сохраненный ранее файл (*Rasch-binary.sav*). В открывшемся окне программы WordPad следует скопировать имена переменных, которые расположены ниже строки *Variable Format Label* (в конце файла), и вставить их после строки *Item Response Variables*.

В результате редактируемый файл в программе WordPad содержит следующий фрагмент.


```

; Number of Cases: 7
; Number of SPSS Variables: 7
! Person Label Variables. (Do not delete this line)
! Item Response Variables. (Do not delete this line)
ITEM1 ; F4.0
ITEM2 ; F4.0
ITEM3 ; F4.0
ITEM4 ; F4.0
ITEM5 ; F4.0
ITEM6 ; F4.0
ITEM7 ; F4.0
!Other SPSS Variables (ignored)
;Variable Format Label
ITEM1 ; F4.0
ITEM2 ; F4.0
ITEM3 ; F4.0

```

После сохранения файла, не закрывая его, нужно перейти в программу WINSTEPS и из пункта меню SPSS (SPSS/SAS) выбрать опцию *Construct WINSTEPS file from SPSS file*. В появившемся окне *Permanent Winsteps control file* задать имя файла с расширением .txt, например Rasch-binary.txt. В открывшемся окне программы WordPad необходимо проверить значения параметров NI (количество заданий NI – 7), ITEM1 (номер столбца, с которого начинаются числовые данные после строки END NAMES: ITEM1 = 1), CODES (коды переменной в исходных данных, CODES = «01»). В итоге подготовленный файл с исходными данными содержит следующий фрагмент.

```

ITEM1 = 1 ; Starting column of item responses
NI = 7 ; Number of items
; SPSS Cases processed = 7
; datum: 0 count: 18
; datum: 1 count: 31
XWIDE = 1 ; this matches the biggest data value observed
CODES = «01» ; matches the data
NAME1 = 9 ; Starting column for person label in data record
; Person Label variables: in label: in line
NAMLEN = 0 ; Length of person label
&END ; Item labels follow:

```

```
ITEM1 ; Item 1
ITEM2 ; Item 2
ITEM3 ; Item 3
ITEM4 ; Item 4
ITEM5 ; Item 5
ITEM6 Item 6
ITEM7 Item 7
END NAMES
1011010 1
0101100 2
0111110 3
0111111 4
1010110 5
0011011 6
1111100 7
```

После сохранения файла программу WordPad можно закрыть. Далее необходимо перейти в программу WINSTEPS, щелкнуть по клавише *Enter*, в меню *File* выбрать опцию *Open File* и открыть сохраненный файл (*Rasch-binary.txt*). В командной строке программы WINSTEPS на запрос *Report output file name* указать имя файла, который будет содержать отчет о рабочем сеансе в WINSTEPS, например *output.txt* (если такой отчет не нужен, нажать *Enter*). На запрос *Extra specifications* щелкнуть по клавише *Enter*. Ввод исходных данных в программе WINSTEPS завершен. На монитор и в файл отчета выводится общая статистика по обработке исходных данных (ее описание мы опускаем).

Результаты конвертирования исходных данных в шкалу Раша по всем испытуемым приводятся в отчете PERSON STATISTICS (табл. 2.6), который формируется выбором опции *PERSON: measure* из меню *Output Tables*.

В столбце ENTRY NUMBER указаны номера испытуемых. Результаты измерений (столбец MEASURE) приведены по убыванию (в контексте тестирования это означает, что подготовленность (знания) испытуемого 4 по всем заданиям теста на 1,03 логита лучше, чем, например, испытуемого 3, и на 2,49 логита лучше, чем испытуемого 2). В столбце RAW SCORE приведены суммарные баллы испытуемых по исходным данным, в столбце COUNT – количество заданий, по которым проводился расчет, в столбце MODEL S.E. – ошибка измерения на основе модели

Отчет программы WINSTEPS по респондентам

PERSON STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEA CORR.	PER- SON
					MNSQ	ZSTD	MNSQ	ZSTD		
4	6	7	2.13	1.14	1.00	.2	.59	.40	.40	4
3	5	7	1.10	.92	.51	-1.2	.40	.84	.84	3
7	5	7	1.10	.92	.99	.1	.92	.43	.43	7
1	4	7	.34	.85	.97	.0	.86	.50	.50	1
5	4	7	.34	.85	1.22	.7	1.33	.24	.24	5
6	4	7	.34	.85	1.17	.6	1.18	.31	.31	6
2	3	7	-.36	.85	1.15	.5	1.04	.33	.33	2
MEAN	4.4	7.0	.71	.91	1.00	.1	.90			
S.D.	.9	.0	.74	.10	.22	.6	.30			

Раша, в столбце PERSON указываются имена (метки) испытуемых (в данном случае имена испытуемых не задавались, поэтому повторяется их номер). Содержание других столбцов отчета будет рассмотрено в гл. 4. В строках MEAN и S.D. указаны средние значения и стандартные отклонения по соответствующим значениям в столбцах.

Результаты измерений по заданиям теста приводятся в отчете ITEM STATISTICS (табл. 2.7), который формируется выбором опции *ITEM: measure* из меню *Output Tables*.

В этом отчете результаты измерений (MEASURE) также приводятся по убыванию: от более сложных — к менее сложным заданиям (например, задание 7 на 0,7 логита сложнее задания 1 и т.д.). В столбце ENTRY NUMBER указаны номера заданий; RAW SCORE — суммарные баллы по каждому заданию; COUNT — количество испытуемых, по которым проводился расчет сложности; MODEL S.E. — ошибка измерения сложности заданий по модели Раша; ITEM — имена (метки) заданий. В строках MEAN и S.D. указаны средние значения и стандартные отклонения по соответствующим значениям в столбцах.

При необходимости можно вывести результаты без упорядочения результатов измерения, а в соответствии с номером респондента или пункта анкеты (*PERSON: entry* или *ITEM: entry*), а также в соответствии с показателями статистики соответствия полученных измерений модели Раша (подразд. 4.4).

Как видим, результаты измерений по абсолютным значениям отличаются от тех, что были получены при расчете «вручную» по простейшему алгоритму (см. табл. 2.5), хотя при ранжировании по величине они будут располагаться в одном и том же порядке. Это обстоятельство связано с тем, что в программе WINSTEPS применяются более точные алгоритмы. В алгоритмах, которые реализованы в программе WINSTEPS и других специальных программных продуктах, использующих модель Раша, учитывается много дополнительных факторов, например, делается поправка на размер выборки, распределение и разброс данных, учитываются случайные выбросы, рассчитываются ошибки калибровки данных и измерений и т.д. [55].

Аналогично могут быть обработаны исследовательские данные, представленные в других шкалах. В гл. 4 будет рассмотрен пример обработки данных по оценке организационного климата с помощью программы WINSTEPS и показано, как оценивается качество полученных измерений.

Отчет программы WINSTEPS по пунктам анкеты

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEA CORR.	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD		
7	2	7	1.71	.88	.89	-.1	.90	.0	.44	ITEM7
1	3	7	1.01	.81	1.39	1.3	1.39	1.2	-.14	ITEM1
2	4	7	.38	.81	.92	-.2	.87	-.3	.44	ITEM2
5	5	7	-.31	.87	1.01	.1	.85	-.1	.32	ITEM5
6	5	7	-.31	.87	.97	.1	.99	.2	.29	ITEM6
3	6	7	-1.24	1.10	.73	-.2	.47	-.3	.59	ITEM3
4	6	7	-1.24	1.10	1.03	.3	.84	.2	.21	ITEM4
MEAN	4.4	7.0	.00	.92	.99	.2	.90	.1		
S.D.	1.4	.0	1.03	.12	.18	.5	.25	.5		

Модель Раша трактуется как модель «объективных измерений», которые не зависят (или зависят в значительно меньшей степени, чем оценочные результаты по данным в неметрических шкалах) от респондентов и измерительных инструментов. Эта модель также обеспечивает возможность корректного использования очень широкого круга статистических методов. Эти обстоятельства позволяют считать результаты, получаемые на основе применения модели Раша, более надежными, чем результаты, полученные по исходным данным, например в шкале Лайкерта.

Поэтому когда результаты, полученные по данным в порядковой шкале, согласуются с результатами по данным, конвертированным в шкалу Раша, это можно считать дополнительным подтверждением надежности полученных результатов, используемого исследовательского инструмента, подхода, гипотезы и т.д. Как правило, при хорошо спланированном и проведенном эксперименте, с использованием обоснованных и надежных инструментов (гл. 4), а также при правильном использовании процедуры конвертирования исходных данных в измерения по шкале Раша, результаты, полученные по исходным данным (оценки) и по данным, конвертированным в шкалу Раша (измерения), в целом хорошо согласуются¹. Однако конвертация данных позволяет выявить менее заметные различия и связи, провести более глубокий и математически корректный анализ данных, сделать более обоснованные выводы.

Вместе с тем не нужно забывать, что модель Раша – это прежде всего *математическая модель*, поэтому не следует руководствоваться только формальными соображениями при принятии тех или иных выводов. Всегда нужно помнить о содержательном наполнении полученных результатов и руководствоваться также содержательными критериями, в том числе здравым смыслом и практической важностью полученных результатов (об этом пойдет речь во второй и третьей частях данного учебного пособия).

Как и любая модель, модель Раша – это инструмент для уточнения и детализации содержательных выводов исследования.

¹ В то же время следует отметить, что тема согласованности результатов традиционных моделей измерения и модели Раша представлена в литературе довольно слабо.

2.6. Метод QIPM: пример компромисса между точностью и простотой

Обсуждая виды шкал, и в частности шкалу Лайкерта, мы уже указывали, что исследователи часто интерпретируют данные в порядковых шкалах как интервальные и используют различные методы обработки данных в ущерб строгости и корректности их применения. Очень наглядно эта ситуация проявляется в использовании подхода к выявлению приоритетных факторов, определяющих деятельность организации: QIPM (*Quality Improvement Priority Matrix* – матрица приоритетов улучшения качества). QIPM – это простой и наглядный метод оценки важности (*importance*) различных факторов, определяющих деятельность организации, и степени их осуществления-реализации (*performance*) [40]. Вариантами такого подхода является оценка факторов по параметрам «важность» и «удовлетворенность» (*satisfaction*) клиентов компании или «влияние на бизнес» (*business value impact*) и «простота реализации» (*ease of implementation*). Оцененные факторы (после агрегирования индивидуальных оценок) отображаются на графике в соответствующих осях, например «важность–реализованность». Приоритетными, т.е. требующими мер по изменению и совершенствованию, считаются факторы с высокой важностью, но низкой реализованностью (факторы, которые попадают в юго-восточный квадрант).

Если данные оцениваются в других категориях, то приоритетность может определяться попаданием в другой квадрант. Например, для оценки соотношения «влияние на бизнес–простота реализации» в качестве приоритетных будут рассматриваться факторы, попадающие в северо-восточный квадрант.

Такой подход не требует ни много времени и ресурсов для сбора данных, ни использования сложного аналитического инструментария для их обработки и анализа. В то же время метод нагляден и удобен для принятия решений, он широко используется в различных государственных и коммерческих организациях США и Европы. В частности, автору книги известны примеры применения такого подхода компанией IBM, Министерством финансов США, Университетом Дж. Вашингтона (США). Работы в области менеджмента качества с использованием QIPM были удостоены престижной премии Malcolm Baldrige National Quality Award в 1995 г. [38].

Однако этот «классический» подход к выделению приоритетных факторов является довольно поверхностным и не обеспечивает возможности проведения серьезного анализа ситуации в организации и выделения приоритетов ее развития. Определенные приоритетных факторов на основе их «попадания» в юго-восточный квадрант не обеспечивает разделения приоритетности с достаточной точностью, поскольку часто от 1/4 до 1/2 всех факторов оказываются в этом квадранте. Еще одна сложность использования такого подхода заключается в «пограничном эффекте». Например, фактор с высокой важностью (близкой к 9) и реализованностью немногим больше средней (например, 5,01) попадет в северо-восточный («успешный») квадрант. И напротив, фактор с важностью немногим больше 5 (например, 5,01) и реализованностью немногим меньше средней (например, 4,99) окажется в юго-восточном («приоритетном») квадранте.

Предложен усовершенствованный подход использования метода QIPM, который позволяет автоматизировать обработку данных и повысить точность получаемых результатов [10; 40]. Для анализа полученных данных для каждого фактора рассчитываются средние арифметические значения «важности» (I) и «реализованности» (P), а также индекс приоритетности $IPR = I/P$. Чем выше этот индекс, тем более приоритетным является соответствующий фактор. На основе индекса приоритетности осуществлялась кластеризация факторов, которая включала следующие этапы:

1) выбор необходимого количества кластеров (с увеличением количества кластеров точность оценки их приоритетности изменяется);

2) расчет «интервала приоритетности» Δ , определяющего «попадание» фактора в тот или иной кластер:

$$\Delta = \frac{(IPR \max - IPR \min)}{\text{Количество кластеров}};$$

3) построение матрицы QIPM и выделение факторов, попадающих в разные кластеры (например, цветом).

Выполнение этих операций может быть относительно просто автоматизировано. Например, в [10] приведено описание программного модуля, включающего набор макросов, разработанных на основе Visual Basic for Application для MS Excel. Итогом

работы модуля является построение диаграммы, представляющей собой матрицу QIPM, где точки, соответствующие оцениваемым факторам, окрашены в различные цвета в зависимости от номера кластера, к которому они относятся в соответствии со значением коэффициента приоритетности. Программа позволяет выбирать нужное количество кластеров, при этом рассчитывается количество точек (факторов) в кластере, коэффициент корреляции «важность—реализованность» для всех факторов и факторов внутри каждого кластера (коэффициент корреляции внутри кластера обычно оказывается выше, чем «общий» коэффициент корреляции).

Но, говоря строго, такой подход не совсем корректен с математической точки зрения, поскольку агрегированные оценки факторов, позиционируемые на графике, рассчитываются как среднее арифметическое индивидуальных оценок, отражаемых в порядковой шкале. Использование метода могло быть более корректным в случае преобразования данных в интервальную шкалу, например с помощью модели Раша. Однако практика показывает, что такое преобразование часто не изменяет значительным образом результаты (для такого подхода), поэтому исследователи нередко жертвуют математической корректностью и точностью измерений в пользу простоты и удобства использования метода.

В качестве примера приведем результаты использования метода QIPM для мониторинга и совершенствования маркетинговой деятельности компании, занимающейся комплексным оснащением объектов торговли, общественного питания и пищевых производств. Для сбора исходных данных была разработана анкета, включающая набор факторов, степень важности и качество реализации которых оценивались клиентами компании по шкале от 1 до 5. Анкета включала пункты по таким темам, как ассортиментная политика, качество обслуживания, конкурентные преимущества, ценовая политика, работа сервисной службы, эффективность рекламы и др. (всего 29 факторов). После обработки была построена матрица QIPM и выделены три кластера приоритетности факторов (рис. 2.8).

Для проверки сопоставимости результатов исходные данные из шкалы Лайкерта конвертировались в шкалу Раша и тоже представлялись в матрице QIPM для визуального сравнения расположения факторов в матрице.

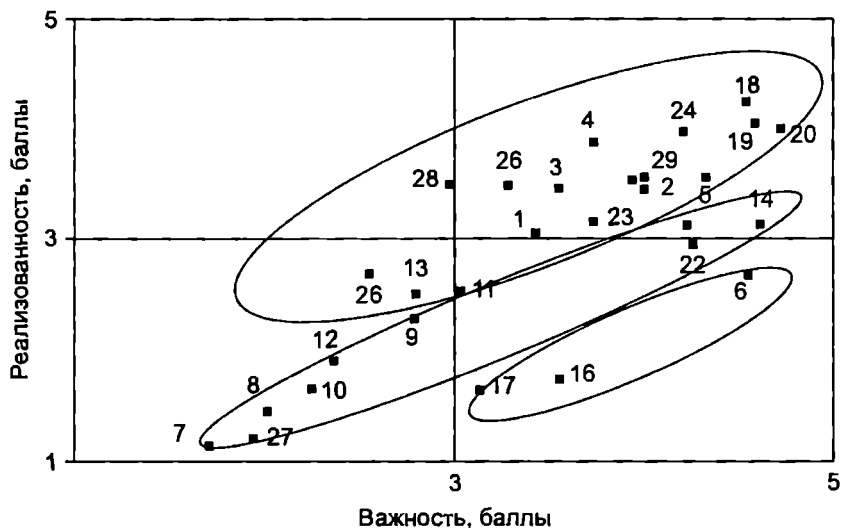


Рис. 2.8. Данные в шкале Лайкерта на диаграмме QIPM

Результаты оценки важности и реализованности в двух шкалах дополнительно сравнивались после ранжирования факторов соответственно по важности и реализованности.

Сопоставление результатов показывает почти полную их идентичность. Поэтому в задачах такого рода, где высокая точность не требуется, нет необходимости использовать сложный математико-статистический инструментарий для измерений, а достаточно ограничиться оценками, что чаще всего и происходит на практике.

В качестве материала для выполнения задания, связанного с использованием QIPM, читателю предлагается выборка из данных, полученных автором данного учебного пособия совместно с профессором кафедры теории менеджмента Университета Дж. Вашингтона (США) С. Ампли (табл. П.10). Метод QIPM использовался для мониторинга деятельности кафедры с целью определения приоритетных направлений ее развития. Преподаватели кафедры оценивали в категориях «важность – реализованность» такие факторы, как заработная плата, компьютерные лаборатории, финансовая поддержка научных исследований, расписание занятий, библиотечные ресурсы, помощь в подготовке заявок на получение грантов, зарубежные командировки,

уровень знания английского языка студентами, функционирование веб-сайтов кафедры и университета и др. (всего оценивалось 52 фактора). Для оценки факторов по критериям «важность — реализованность» по методу QIPM использовалась порядковая шкала с позициями от 1 до 9. Предложенные подходы для анализа и интерпретации данных, полученных с помощью QIPM, и результаты применения этих подходов были представлены и обсуждены на годовом собрании преподавателей и сотрудников кафедры теории менеджмента Университета Дж. Вашингтона. По итогам обсуждения был разработан проект плана стратегического развития кафедры и сформулированы предложения по использованию разработанных методов на других кафедрах университета.

Задания для самостоятельного выполнения

1. Выберите из задания 2 к гл. 1 признак, который может быть измерен в шкале Лайкерта. На примере выбранного свойства покажите преимущества и недостатки использования этой измерительной шкалы.

Цель задания — освоение методики конвертирования данных, а не получение содержательного вывода по конкретному исследованию. Поэтому содержательный смысл вопросов (показателей) не приводится, а используются формальные обозначения показателей: «пункт1», «пункт2» и т.д. Рекомендуется выполнить задание в среде Excel с использованием функций СРЗНАЧ (вычисление среднего значения) и стандартных графических инструментов.

2. Выполните упражнения по преобразованию данных (выборка из данных по оценке организационного климата двух компаний (табл. П.9). Данные представлены в шкале Лайкерта.

2а. Рассматривая каждый пункт анкеты в качестве индикатора, характеризующего организационный климат, на одном графике по исходным данным постройте и сравните профили организационного климата двух компаний (профиль климата — ломаная линия, состоящая из отрезков, соединяющих точки, координатами которой являются идентификационный номер измеряемого показателя (на оси абсцисс) и измеренное значение этого показателя (на оси ординат)). Использовать исходные данные в шкале Лайкерта.

2б. Конвертируйте исходные данные в стандартные *z*- и *T*-шкалы. По конвертированным данным постройте профили организационного климата двух компаний и сравните их. Объясните сложность использования *z*- и *T*-шкал для данного случая. (При конвертировании используйте стандартные функции программы Excel СРЗНАЧ и СТАНДОТКЛОН.)

2в. Конвертируйте исходные данные в шкалу Раша с помощью программы Excel по алгоритму, представленному в подразд. 2.4. По конвертированным данным постройте профили организационного климата двух компаний и сравните их с результатами выполнения пункта 2а данной задачи. Предложите возможные объяснения изменения результатов, сделайте выводы.

При преобразовании данных полезным может оказаться использование функции Excel СЧЕТЕСЛИ (диапазон поиска; значение) для расчета частот (количества респондентов, поставивших тот или иной балл).

2г. Конвертируйте исходные данные в шкалу Раша с помощью программы WINSTEPS. По полученным данным вновь постройте профили организационного климата двух компаний и сравните их с результатами пунктов 2а, 2б и 2в.

2д. На одном графике совместите организационные профили первой компании, полученные по трем наборам данных: 1) по исходным данным в шкале Лайкерта; 2) по данным в шкале Раша, полученным при выполнении пункта 2в; 3) по данным в шкале Раша, полученным при выполнении пункта 2г. Проследите за изменением результатов, сделайте выводы. Насколько существенны, по вашему мнению, изменения в измерениях показателей организационного климата при использовании разных шкал? Обоснуйте ответ.

3. Примените QIPM по выборке (табл. П10), содержащей оценки 33 факторов 11 респондентами в категориях «важность» (*I*) и «реализованность» (*P*).

Цель задания – освоение методики QIPM и конвертирования данных, а не получение содержательного вывода по конкретному исследованию. Поэтому в задании используются формальные обозначения факторов: «фактор1», «фактор2» и т.д.

3а. Агрегируйте оценки по каждой из категорий в двух шкалах: порядковой шкале и шкале Раша. Учитывайте, что от некоторых респондентов не получены отклики по ряду пунктов.

Результаты представьте на трех графиках в осях «Важность» (*I*) и «Реализованность» (*P*). На первом графике должны быть представлены результаты, полученные по исходным данным; на втором – после преобразования данных в шкалу Раша с помощью программы Excel по алгоритму, представленному в подразд. 2.4; на третьем – после преобразования данных в шкалу Раша с помощью программы WINSTEPS (см. задание 2г). Сравните распределение полученных результатов на диаграмме QIPM и сделайте выводы.

3б. Результаты представьте в виде трех таблиц с четырьмя столбцами: «Фактор», «I», «P» и «Квадрант». В каждой из этих таблиц представьте один из трех наборов данных, полученных при выполнении задания 3а.

Пояснение. Квадранты диаграммы QIPM можно обозначить как «юго-восточный» (правый нижний) – 1, «северо-восточный» (правый верхний) – 2, «юго-западный» (левый верхний) – 3, «северо-западный» (левый нижний) – 4. «Попадание» в квадрант можно определить, используя логическую функцию Excel ЕСЛИ, вычисляя с ее помощью положение фактора относительно средних значений «важности» и «реализованности».

Сравните полученные результаты с точки зрения приоритетности факторов. Изменилась ли приоритетность факторов при переходе к другой шкале? (Для оценки изменения приоритетности факторы можно ранжировать по характеристикам (*I* и *P*) и проследить изменение рангов при изменении шкалы.) Сделайте выводы.

3в. По данным в шкале Лайкерта определите индекс приоритетности каждого фактора. Ранжируйте факторы по приоритетности. Сравните результаты с результатами, полученными при выполнении заданий 3а и 3б. Объясните сложности, связанные с определением индекса приоритетности по данным, конвертированным в шкалу Раша.

3г. На основе алгоритма, описанного в подразд. 2.6, проведите кластеризацию факторов по приоритетности (по данным в шкале Лайкерта). Количество кластеров: а) 3, б) 4. Сравните результаты с результатами, полученными при выполнении пунктов 3.1–3.3 и сделайте выводы.

Формирование и анализ выборочной совокупности

3.1. Особенности выборочного метода исследования

Основная проблема, часто возникающая при проведении исследований, – это ограниченность временных, финансовых, технических, человеческих и прочих ресурсов, имеющихся у исследователя. Все множество (полный набор) элементов, относительно которого осуществляется заключение по результатам измерения, называется *генеральной совокупностью*, или *популяцией* (population). Соответственно объектами измерения являются элементы генеральной совокупности. Исследование *всех* элементов генеральной совокупности называется *сплошным исследованием*.

Например, необходимо оценить политические предпочтения граждан накануне выборов. Наиболее точные результаты были бы получены после опроса всех потенциальных избирателей (сплошное исследование). Но это невозможно в стране, электорат которой составляют несколько десятков миллионов человек. Даже для гораздо меньших групп (студенты университета, жители городского района и т.д.) сбор данных по всем членам группы чаще всего затруднен или невозможен. У исследователя практически всегда недостаточно времени и ресурсов для того, чтобы измерить характеристику каждого объекта генеральной совокупности.

Для решения этой проблемы исследователь формирует выборку. *Выборка (sample)* – это ограниченная по численности группа объектов (например, респондентов), отбираемая из генеральной совокупности для изучения ее свойств. Цель формирования выборки – получение достоверных выводов о свойствах исследуемой генеральной совокупности. Исследование объектов выборки называется *выборочным исследованием*. После анализа результатов выборочного исследования проверяется ряд гипотез для обоснования возможности переноса этих результатов на всю генеральную совокупность (рис. 3.1).

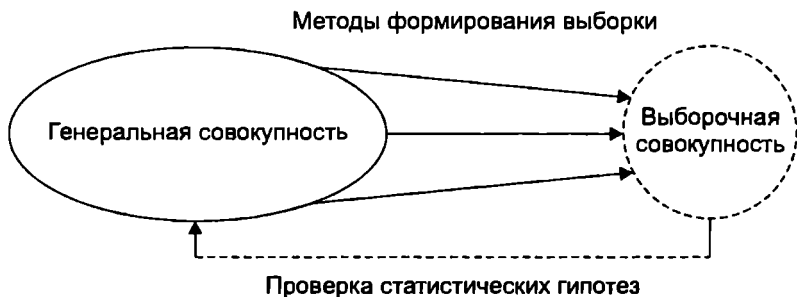


Рис. 3.1. Иллюстрация принципа выборочного исследования

Выборочное исследование имеет ряд преимуществ по сравнению со сплошным. Во-первых, это существенное снижение затрат, во-вторых, сокращение времени на получение данных и их обработку. Часто результаты выборочного исследования рассматриваются как менее точные и надежные, чем результаты сплошного исследования. Однако это не совсем верное суждение. При использовании корректных подходов выборочное исследование обеспечивает очень точное представление о генеральной совокупности. Более того, выборочное исследование подчас обеспечивает лучшее качество измерений, чем исследование всей генеральной совокупности по причине более качественного сбора данных и их анализа. Например, оно может обеспечить лучшее интервьюирование, наблюдение, обработку и другое, чем в случае измерения всех элементов генеральной совокупности. В некоторых ситуациях обследование всех элементов популяции вообще невозможно, например, если исследование предполагает разрушение или порчу объекта (контроль качества продукции и др.).

Выборочное исследование не имеет смысла, если *все* элементы генеральной совокупности различны (дисперсия измеряемых параметров значительна). Например, если мы выясняем мнение по определенной теме, а *все* представители генеральной совокупности имеют *разное* мнение, нам необходимо будет опросить всех представителей, т.е. провести сплошное исследование.

При формировании выборки нужен такой отбор элементов (наблюдений), чтобы сформированная выборка была репрезентативной (представительной) по отношению к генеральной совокупности, из которой она была извлечена. **Репрезентатив-**

ность выборки — способность выборки представлять изучаемые явления достаточно полно с точки зрения их изменчивости в генеральной совокупности. Эта характеристика показывает, можно ли распространить результаты, полученные на данной выборке, на всю генеральную совокупность, из которой взята эта выборка. Репрезентативная выборка должна воспроизводить все важные характеристики изучаемой генеральной совокупности. Например, если мы определяем среднюю заработную плату работников отрасли и генеральная совокупность на 60% состоит из женщин и на 40% — из мужчин, то и выборка из нее должна отражать те же пропорции. По выражению Дж. Хили, репрезентативная выборка — это генеральная совокупность в миниатюре. Для формулирования и обоснования статистических выводов репрезентативность выборок играет решающую роль: если выборки не являются репрезентативными, то обобщение выводов, полученных на основе выборки, на генеральную совокупность становится неприемлемым.

Репрезентативность характеризует соответствие показателей, полученных в результате выборочного наблюдения, показателям, свойственным для всей генеральной совокупности. Расхождение между указанными показателями представляет собой *ошибку репрезентативности*, которая может быть *случайной* (сбалансированной) или *систематической* (несбалансированной, направленной). Примером систематической ошибки репрезентативности выборки является прогнозирование президентских выборов в США в 1936 г., когда одним из главных кандидатов на этот пост был Ф. Рузвельт, претендовавший на второй срок. Для исследования была сформирована очень большая выборка (свыше 2 млн респондентов), но с систематической ошибкой — проводилось изучение мнений верхнего и среднего классов, тогда как Рузвельта в основном поддерживал низший класс, который и обеспечил его блестящую победу на выборах. Причинами систематической ошибки могут быть также время, день или место проведения измерений, метод формирования выборки и др.

Репрезентативность — это неформализуемая характеристика. Однозначно и точно сказать, основываясь на четких показателях, что данная выборка является репрезентативной, невозможно. Но о репрезентативности выборки, т.е. о том, насколько результаты выборочного исследования соответствуют (или могут соответствовать) результатам сплошного исследования, можно

судить по таким критериям, как способ формирования выборки, ее объем и вариативность (изменчивость) данных. Эти критерии будут подробно рассмотрены ниже.

3.2. Методы формирования выборки

Известно несколько методов формирования выборок из генеральных совокупностей. В наиболее общем виде формируемые выборки можно разделить на две группы: *вероятностные* и *условно-вероятностные*. Вероятностная выборка часто называется также *случайной* выборкой. Однако термин «вероятностная выборка» все же предпочтительнее, так как в обиходе под случайным часто подразумевается то, что происходит «по случайному стечению обстоятельств», «наугад» и т.п. Но на самом деле вероятностные выборки формируются с помощью таких подходов, которые предусматривают тщательное следование алгоритму, не оставляя место бессистемности и «случайности» в обыденном понимании этого слова.

Условно-вероятностные выборки формируются «по желанию», «по присутствию», «по доступности» и другим аналогичным принципам и критериям. Очевидно, что лишь к вероятностным выборкам можно в полной мере применить методы получения статистических выводов о генеральной совокупности. В то же время условно-вероятностные выборки также весьма часто используются в исследовательской практике, особенно применительно к небольшим группам. В частности, среди исследователей, работающих в вузах, распространена практика использования студентов в качестве респондентов для «отладки» исследовательских подходов и инструментов. При этом, как правило, опрашиваются студенты в тех студенческих группах, с которыми и работает преподаватель-исследователь. Такие «удобные» для исследователя выборки, безусловно, полезны для определенных целей (например, для предварительного формулирования исследовательских гипотез и тестирования различных форм опроса перед тем, как начать выполнение более сложного и дорогого проекта). Сформировать условно-вероятностные выборки обычно достаточно просто и недорого. Однако чаще всего такие выборки нерепрезентативны, и переносить результаты, полученные по этим выборкам, на другие группы нельзя. Следовательно, нельзя обосновывать статистические выводы и ги-

потезы только на выборках, не являющихся вероятностными, даже при наличии очень веских доказательств в пользу обосновываемых гипотез.

Методы формирования вероятностных выборок основаны на принципе равновероятностного отбора (*equal probability of selection method*): вероятность попадания в выборку должна быть одинакова для любого элемента (объекта) исследуемой генеральной совокупности. Вероятность того, что выборка, сформированная по такому принципу, будет репрезентативной, очень высока, однако гарантировать это все же невозможно.

Для такого формирования выборки необходимо иметь:

а) список всех элементов (объектов) исследуемой генеральной совокупности;

б) систему отбора элементов из этого списка, которая бы обеспечивала равную вероятность включения в выборку любого элемента из списка.

Для отбора списочных элементов может использоваться процедура генерирования случайных чисел, включенная во многие статистические и математические программные продукты. В программе Excel такая процедура может быть осуществлена выбором функции СЛЧИС (в англоязычной версии – RAND). Эта функция возвращает равномерно распределенное случайное число в диапазоне от 0 до 1. Для того чтобы получить случайное вещественное число в интервале между числами a и b , можно использовать следующую формулу:

$$\text{СЛЧИС()}*(b - a) + a.$$

Пример 3.1. Если, например, элементами генеральной совокупности являются студенты факультета в количестве 2000 чел. и всем им присвоены соответствующие идентификационные номера (от 1 до 2000), то формирование вероятностной выборки может быть сделано с помощью функции СЛЧИС()*1999 + 1.

Полученный результат сопоставляется с номером из списка, и таким образом отбирается соответствующий студент в выборку. Когда формируемая выборка достигает желаемого объема, процедура завершается.

Выборка может формироваться на основе повторного или бесповторного вероятностного отбора. Сущность *повторного отбора* состоит в том, что каждый отобранный элемент (например, идентификационный номер) «возвращается» в генеральную совокупность, т.е. существует вероятность, что он снова попа-

дет в выборку. Но на практике такой подход осуществляется редко. Обычно выборочные исследования проводятся по схеме *бесповторного отбора*, когда повторное попадание одного и того же элемента генеральной совокупности в выборку исключено (например, повторяющийся идентификационный номер игнорируется и тем самым попадает в выборку только один раз). Если объем формируемой выборки намного меньше объема генеральной совокупности, то тогда вероятность извлечения одного и того же элемента дважды весьма мала, поэтому игнорирование повторений не приводит к искажению наших заключений о генеральной совокупности.

Для большого списка элементов и при отсутствии компьютерных средств, с помощью которых можно осуществить указанную процедуру, такой подход трудно реализуем. С целью сэкономить время и другие ресурсы в этой ситуации исследователи часто формируют *систематическую вероятностную выборку*. При этом в выборку случайно отбирается только первый элемент, а затем каждый i -й элемент списка генеральной совокупности. Например, если есть список из 10000 элементов генеральной совокупности и необходима выборка объемом 200, нужно случайным образом отобрать один элемент списка, а затем взять элементы списка, отстоящие через каждые $10000/200$, т.е. 50 номеров от него (вверх и вниз по списку).

При формировании систематических выборок нет строгого соответствия принципу равновероятностного отбора, так как, после того как отобран первый элемент, для остальных элементов вероятность отбора уже не является одинаковой, как это было ранее. Происходящий при этом рост вероятности ошибки очень незначителен, но лишь при условии, что список, из которого отбираются элементы, является сам по себе случайным (хорошо перемешанным) или по крайней мере нециклическим (непериодическим) по отношению к измеряемым признакам объекта.

Следующий метод формирования вероятностной выборки — это получение *стратифицированной (типической, районированной)* выборки. Этот метод гарантирует репрезентативность выборки в *отношении выбранных признаков объекта*. Список элементов генеральной совокупности разбивается на подписки в соответствии с категориями нужного признака, и затем формируется выборка на основе этих подписков. И если из каждого подписка извлекается количество элементов, пропорциональ-

ное доле соответствующей категории в генеральной совокупности, то полученная выборка будет репрезентативной для данной генеральной совокупности в отношении выбранного признака.

Пример 3.2. Необходимо сформировать выборку объемом 50 студентов, в которой был бы пропорционально представлен каждый из 10 факультетов университета. Если, допустим, на одном факультете обучаются 30% студентов университета, а на другом – лишь 2% всех студентов, то ранее описанные процедуры формирования выборок дадут выборку с очень малым (или даже с нулевым) количеством студентов второго факультета. Если генеральная совокупность будет предварительно разбита на подписки (по факультетам), то тогда можно отобрать ровно 5 студентов соответствующего факультета.

Если мы проводим стратифицированный отбор по *принципу пропорциональной численности единиц типических групп*, то объем выборки определяется по формуле

$$n_i = n \frac{N_i}{N},$$

где n_i – объем выборки из типической группы;

n – общий объем выборки;

N_i – объем типической группы;

N – объем генеральной совокупности.

В этом случае объем типической выборки от первого факультета составит 15 студентов, от второго – 1 студент.

Применение метода формирования стратифицированных выборок ограничено главным образом тем, что очень часто точный состав генеральной совокупности неизвестен. Если у нас нет никакой информации о генеральной совокупности, то мы не сможем построить схему стратифицирования и определить, сколько элементов следует отобрать из каждого подписка.

Впрочем, эта трудность осложняет применение *всех* описанных процедур для формирования выборок. Во многих случаях список элементов генеральной совокупности отсутствует или получить его крайне сложно (не существует полного списка граждан страны или даже области, эти списки устарели и т.д.). Еще более сложно изучить (измерить, оценить) все объекты (элементы списка), отобранные в выборку, например, часть отобранных респондентов могут отказаться участвовать в опросе.

В случае неполноты списков может формироваться *групповая (кластерная) выборка*, которая предполагает отбор *групп* элементов (кластеров), а не отдельных элементов. Такими кластерами могут являться, например, административные районы города. Из них можно сформировать выборку. Кластеры должны быть примерно одного объема, так как (это будет показано во второй части данного учебного пособия) большинство тестов для проверки гипотез работают при предположении примерно равного (по крайней мере однопорядкового) объема выборок.

Затем внутри каждого кластера (района, попавшего в выборку) формируется подвыборка по микрорайонам (избирательным округам, кварталам и т.п.). Подобная процедура может быть продолжена и дальше, вплоть до жителей отдельных домов. Если все подвыборки являются вероятностными, то в результате будет получена выборка, которая с большой вероятностью окажется репрезентативной для всего города (причем она будет получена без использования списка всех жителей города).

Групповая выборка менее надежна из-за нескольких этапов отбора, каждый из которых может вносить ошибку, но если использование других методов формирования выборки невозможно, то формирование групповой выборки может оказаться единственной альтернативой.

3.3. Анализ точности результатов выборочного исследования: доверительный интервал

Довольно распространенным является мнение о том, что увеличение точности результатов выборочного исследования пропорционально увеличению числа элементов выборки, что не совсем справедливо. Более 90% ошибок измерений имеют «невыборочные» источники, и лишь 10% ошибок появляются в результате неправильного формирования выборки [39]. Мнение об обязательном повышении точности с ростом численности выборки основывается на трех заблуждениях, к сожалению, разделяемых многими исследователями.

Заблуждение 1. Чем больше выборка, тем она репрезентативнее.

То, что это не всегда так, демонстрирует пример с прогнозированием исхода президентских выборов в США в 1936 г.: выборка объемом свыше 2 млн респондентов не обеспечила репрезентативности, прогноз оказался не верен. Выборка, состоящая

из 100 элементов, может обеспечить лучшую репрезентативность, чем выборка, состоящая из 10000 элементов.

Заблуждение 2. Выборка должна состоять как минимум из 10% элементов генеральной совокупности.

На самом деле *абсолютный* объем выборки – фактор гораздо более важный, чем объем выборки *относительно* объема генеральной совокупности. Например, выборка объемом 100 элементов, извлеченная из 1000 элементов генеральной совокупности, может обеспечить примерно такую же точность оценки, как выборка объемом 1000 элементов, отобранных из 100 млн элементов, формирующих генеральную совокупность. Так, при опросах общественного мнения о возможном исходе выборов президента США обычно формируется выборка, насчитывающая «всего» 1500–2000 респондентов, тогда как весь электорат США насчитывает около 150 млн чел. Однако при формировании выборки из большой популяции возникает проблема отбора «соответствующих» элементов.

Заблуждение 3. Заранее невозможно ответить на вопрос о необходимой и достаточной численности выборки.

Это отчасти так. Действительно, требуемый объем выборки можно определить после того, как проведен анализ хотя бы нескольких случайным образом предварительно отобранных элементов генеральной совокупности. Но это может (и должно) быть сделано на стадии предварительного тестирования, что является обязательным этапом любого серьезного исследования. Кроме того, сделать заключение о необходимом объеме выборки можно по результатам предыдущих исследований.

Следует запомнить, что необходимый объем выборки представляет собой функцию вариации измеряемых параметров генеральной совокупности и точности оценки этих параметров, требуемой исследователем. Например, в зависимости от требуемой точности для одной и той же совокупности может быть достаточно 50, 500 или 5000 элементов выборки. И напротив, заданную точность могут обеспечить 100, 1000 или 10000 элементов в зависимости от того, насколько велика дисперсия полученных измерений.

Общим правилом при формировании выборки является следующее: чем больше дисперсия оцениваемых параметров генеральной совокупности, тем больший объем выборки требуется

для того, чтобы обеспечить требуемую точность. Например, если мы выясняем мнение по определенной теме, когда все представители популяции (элементы генеральной совокупности) имеют предположительно одинаковое мнение, нам достаточно опросить одного представителя. Если возможны два мнения, нам необходима выборка по меньшей мере из двух элементов (но при этом мы должны позаботиться о том, чтобы в выборку попали носители данных различных мнений, и нам понадобится большая выборка, чтобы определить частотность встречаемости этих мнений). Если возможны 10 точек зрения, нужна еще большая выборка, и т.д.

С практической точки зрения главными факторами, определяющими объем выборки, являются желаемая точность оценки и бюджет исследования. По поводу бюджета особые комментарии не требуются. Если, скажем, в бюджете исследователя имеется 1000 руб., а телефонный опрос каждого респондента требует в среднем 10 руб., то максимальный объем выборки – 100 элементов. Если нужна выборка большего объема, то необходимо выбирать другие, более дешевые формы опроса. Более интересным в контексте тематики этого учебного пособия и данной главы в частности является вопрос об объеме выборки с позиций точности оценки измеряемых параметров.

Оценка параметра может быть *точечной* или *интервальной*. Примером точечной оценки является результат гипотетического выборочного исследования, установившего, что 61% опрошенных пользователей телефона марки «Brand Name» остались довольны покупкой, или 73% граждан, участвовавших в опросе общественного мнения, поддерживают новые инициативы правительства. Как мы уже говорили, результаты выборочного исследования нельзя напрямую переносить на генеральную совокупность, т.е. мы не можем сказать, что *именно* 61% *всех* пользователей данной марки телефонов ими довольны, или что *точно* 73% *всех* граждан страны поддерживают новый закон. Очевидно, что здесь должен быть некоторый допуск, или интервал, в пределах которого можно формулировать результат, например: от 45 до 75% пользователей телефона остались довольны сделанной покупкой и т.п. Последний подход, безусловно, более оправдан при выборочном исследовании.

Точность интервальной оценки параметра, измеряемого при выборочном исследовании, определяется двумя показателями:

а) интервалом, в котором ожидается обнаружить оцениваемый параметр;

б) вероятностью обнаружения этого параметра в данном интервале.

Эти два показателя объединяет понятие *доверительного интервала*. Процедура определения доверительного интервала основана на *центральной предельной теореме* — одной из главных теорем теории вероятностей и статистики. Согласно этой теореме распределение средних значений выборок, извлекаемых из одной и той же совокупности, соответствует нормальному распределению. Более того, когда выборки становятся достаточно большими, то выборочные средние подчиняются нормальному закону, даже если исходная переменная не является нормальной. Среднее значение всех выборочных средних равно среднему значению генеральной совокупности (μ), а стандартное отклонение выборочных средних (σ_x) определяется по следующей формуле:

$$\sigma_x = \frac{\sigma}{\sqrt{n}},$$

где n — объем выборки;

σ — стандартное отклонение¹ по генеральной совокупности (корень квадратный из дисперсии):

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}},$$

N — объем генеральной совокупности.

Величина σ_x называется также *стандартной ошибкой среднего* (*standard error of the mean*)². Вычисление стандартной ошибки среднего основывается на предположении о нормальности измеряемой переменной. Если это предположение не выполнено, то оценка может оказаться неверной, особенно для малых выборок.

¹ Используется также термин *среднее квадратическое отклонение*.

² Термин был впервые введен Юлом (Yule) в 1897 г. Используется также термин *средняя ошибка выборки*.

Естественным образом возникает вопрос о том, какой объем выборки может считаться «достаточно большим». Известно эмпирическое правило, согласно которому принимается, что если объем выборки (n) равен 100 или более, то применима центральная предельная теорема, и допущение о нормальности распределения всех возможных выборочных средних может быть принято. Показано, что при увеличении объема выборки до 100 и более качество оценки стандартной ошибки среднего улучшается и без предположения нормальности выборки. Если же n меньше 100, то нужно иметь доказательства нормальности распределения генеральной совокупности, и только в этом случае можно полагать, что распределение, которому подчиняются выборочные статистики, является нормальным. Если таких доказательств нет, необходимо проверять нормальность выборочного распределения.

Поскольку в большинстве случаев значение стандартного отклонения по генеральной совокупности (σ) неизвестно, его заменяют выборочным стандартным отклонением (s). Выборочное стандартное отклонение (стандартное отклонение по выборке) определяется как

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}},$$

где \bar{x} – среднее значение изучаемого признака по выборочным наблюдениям.

В математической статистике доказано, что стандартные отклонения по генеральной и выборочной совокупностям связаны соотношением

$$\sigma = s \sqrt{\frac{n}{n-1}}.$$

Таким образом, стандартная ошибка среднего рассчитывается как

$$\sigma_x = \frac{s}{\sqrt{n}}$$

(предполагается, что выборка формируется в результате случайного повторного отбора). Отсюда следует, что стандартное отклонение по выборке определяет интервал попадания среднего по всей генеральной совокупности. Стандартная ошибка среднего зависит от стандартного отклонения по выборке и ее объема. Например, если стандартное отклонение по выборке $s = 0,74$ и $n = 10$, то стандартное отклонение выборочных средних равно

$$\sigma_x = \frac{0,74}{\sqrt{10}} = 0,234.$$

Если стандартное отклонение по выборке уменьшается в *два* раза, то оцениваемое изменение измеряемого параметра по генеральной совокупности также уменьшается в *два* раза:

$$\sigma_x = \frac{0,37}{\sqrt{10}} = 0,117.$$

Предположим, что по результатам выборочного исследования мы получаем стандартную ошибку среднего:

$$\sigma_x = \frac{0,8}{\sqrt{25}} = 0,16.$$

При увеличении количества респондентов в *четыре* раза, при том же самом значении стандартного отклонения по выборке, мы можем обеспечить увеличение точности лишь в *два* раза:

$$\sigma_x = \frac{0,8}{100} = 0,08.$$

При бесповторном случайном отборе численность генеральной совокупности в ходе формирования выборки сокращается. Поэтому при такой схеме формирования выборки стандартное отклонение выборочных средних рассчитывается как

$$\sigma_x = \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}.$$

Очевидно, что для применения этой формулы (и этого механизма отбора соответственно) нам должна быть известна численность генеральной совокупности N .

Для нормального распределения существует универсальное соотношение между относительной частотой встречаемости в генеральной совокупности значений x , средним значением (μ) и стандартным отклонением (σ). Это соотношение называется *законом нормального распределения*:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Это соотношение удобно представить для *стандартного нормального распределения*¹ (или *z-распределения*) в виде

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Любое нормальное распределение может быть сведено к *z-распределению* с помощью простого преобразования:

$$z = \frac{x - \mu}{\sigma}.$$

Последняя формула называется *стандартным z-преобразованием*, переводящим измерения в *стандартную z-шкалу*. В результате такого преобразования значения z выражаются в единицах стандартного отклонения от среднего.

Отсюда следует, для любого z можно однозначно определить площадь под кривой любого нормального распределения вне зависимости от величины среднего значения и стандартного отклонения. Так, для $z = 1$ около 68,27% всех значений признака располагаются в пределах одного стандартного отклонения по обе стороны от среднего значения при любом нормальном распределении (рис. 3.2).

В пределах трех стандартных отклонений ($z = 3$) умещается почти вся генеральная совокупность, а именно 99,73%.

¹ Стандартное нормальное распределение имеет среднее значение, равное 0, и стандартное отклонение, равное 1. Поэтому для обозначения стандартного нормального распределения также используется термин *единичное нормальное распределение*.

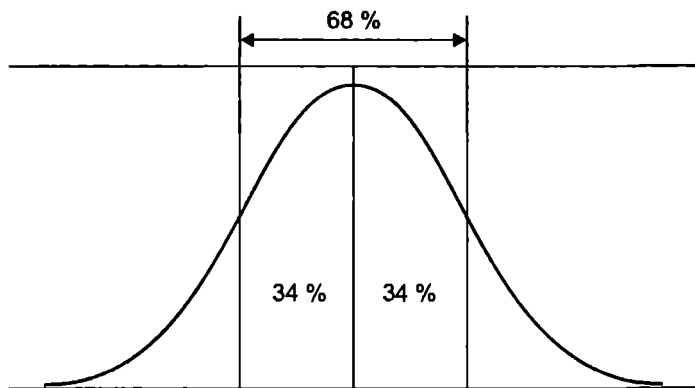


Рис. 3.2. Площадь под кривой нормального распределения ($z = 1$)

Важным практическим следствием этого свойства является возможность оценить точность определения изучаемого параметра на основе выборочного исследования. Так, с вероятностью $P = 0,6827$ значение изучаемого параметра по генеральной совокупности, который оценивается по элементам выборки, будет попадать в интервал $\bar{x} \pm \sigma_x$.

Для $z = 2$ значение вероятности составит 0,9544, т.е. в 95,44% случаев значение исследуемого параметра будет попадать в интервал $\bar{x} \pm 2\sigma_x$. Для $z = 3$ значение вероятности составит 0,9973, т.е. в 99,73% случаев значение параметра будет лежать в интервале $\bar{x} \pm 3\sigma_x$.

Значения z для других значений вероятности можно определить из табл. П.1 приложения 1 или с помощью функции Excel НОРМСТОБР, возвращающей значение z при заданной вероятности.

Если известно среднее арифметическое значение по выборке (\bar{x}) и выборочное стандартное отклонение (s), легко определить стандартную ошибку среднего σ_x . Используя соответствующую статистическую таблицу и задавая необходимые значения вероятности (требуемый уровень статистической значимости¹

¹ Подробно это понятие разбирается в гл. 5.

$p = 1 - P$), можно определить значение z , которое соответствует заданному значению вероятности попадания среднего значения параметра по генеральной совокупности в интервал $\Delta = \bar{x} \pm z\sigma_x$.

Величина Δ называется *доверительным интервалом* (*confidence interval*), а величина $\delta = \pm z\sigma_x$ называется *предельной ошибкой среднего* (или *предельной ошибкой выборки*). Доверительный интервал фактически характеризует *точность оценки* измеряемой величины. Таким образом, для оценки точности выборочных измерений достаточно определить среднее значение и стандартное отклонение по выборке, а также задать уровень значимости.

Очевидно, что с увеличением значения z возрастает вероятность попадания среднего в доверительный интервал Δ , но при этом диапазон оценки становится неопределеннее и размытее, что уменьшает точность оценки¹. Поэтому не следует стремиться задавать очень большое значение вероятности. Вполне достаточным является 90%-ный или 95%-ный уровень значимости ($p < 0,05$). Поскольку стандартное отклонение средних значительно меньше стандартного отклонения индивидуальных откликов, приемлемым считается даже 68%-ный доверительный интервал [53].

В случае, когда выборка состоит из менее 100 элементов или когда нет достаточных оснований считать выборочное распределение нормальным, для определения доверительного интервала рекомендуется использовать другое теоретическое распределение – t -распределение Стьюдента. Это распределение подробно будет рассмотрено во второй части данного учебного пособия, а пока скажем, что в этом случае процедура определения доверительного интервала аналогична случаю больших выборок, но вместо значения z используется значение t -критерия Стьюдента. Это значение зависит от объема выборки и задаваемого уровня вероятности. Значение t -критерия можно определить из табл. П.2 или с помощью функции Excel СТЬЮДРАСПОБР($1 - P$; $n - 1$).

Пример 3.3. Группа студентов университета разрабатывает проект создания «Университетского ланч-клуба» (гибрида студенческой столовой и фешенебельного ресторана). Студенты решают провести мониторинг ряда параметров, на основе чего и принять решение об организации работы предполагаемого заведения.

¹ Чем менее определенным является прогноз, тем с большей вероятностью он осуществится.

Пусть в качестве измеряемого параметра выступает параметр «Предпочтительное время обеда». Проведено предварительное выборочное измерение (объем выборки – 10 элементов, выборки формировались в результате случайного повторного отбора). Респондентам предлагалось выбрать время обеда из пяти предложенных вариантов: 11.00, 11.30, 12.00, 12.30, 13.00. Отклики респондентов кодировались как 1, 2, 3, 4, 5 соответственно выбранному времени. Поскольку интервалы между позициями выбранной шкалы равны и соответствуют 30 мин., используемую шкалу можно рассматривать как интервальную. Результаты предварительного выборочного исследования представлены в табл. 3.1.

Таблица 3.1

Данные для определения доверительного интервала

Респондент	Отклик
1	3
2	2
3	3
4	2
5	5
6	1
7	4
8	3
9	4
10	3
\bar{x} (среднее значение)	3,0
s (стандартное отклонение)	1,155

Определяя стандартную ошибку среднего по характеристикам выборки, получим

$$\sigma_x = \frac{1,155}{\sqrt{10}} = 0,365.$$

Значение стандартного отклонения переводится во временной диапазон исходя из дизайна анкеты: интервал в 1 пункт соответствует 30 мин. Поэтому значение 0,5 соответствует $0,5 \cdot 30 = 15$ мин., а 0,365 соответствует $0,365 \cdot 30 = 11$ мин. Таким образом, по результатам выборочного исследования мы можем оценить интервал изменения среднего по генеральной совокупности. Для $z = 1$ он составляет $3,0 \pm 0,365$, т.е. 12.00 ± 11 мин. (рис. 3.3).

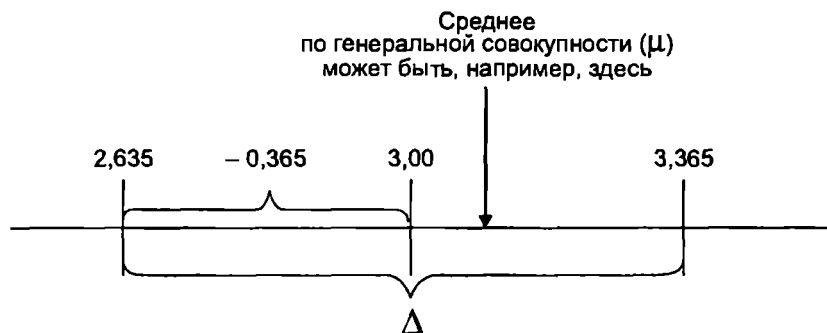


Рис. 3.3. Доверительный интервал

При этом вероятность попадания среднего по генеральной совокупности в этот интервал равна приблизительно 68%, так как мы принимаем $z = 1$. Аналогично с вероятностью 95% по характеристикам выборки можно утверждать, что интервал изменения среднего по генеральной совокупности $\Delta = 3,0 \pm 0,365 \cdot 1,96 = 3,0 \pm 0,715$ (12,00 \pm 21,5 мин.).

Однако, как отмечалось ранее, такой подход к определению доверительного интервала не совсем корректен, так как объем выборки мал ($n < 100$). Поэтому в данном случае необходимо использовать значение t -критерия. Для $n = 10$ и $P = 0,95$ (95%) $t = 2,262$ (см. табл. П.2). Соответственно $\Delta = 3,0 \pm 0,365 \cdot 2,262 = 3,0 \pm 0,826$ (12,00 \pm 25 мин.).

В ряде случаев подобные задачи необходимо решать по *сгруппированным* данным, когда указывается, сколько раз в выборке встречается определенное значение изучаемого признака. В таком случае среднее значение признака рассчитывается не как *среднее арифметическое* (сумма значений признака по элементам выборки, деленная на количество элементов), а как *среднее арифметическое взвешенное*¹. Число одинаковых значений признака в ряду распределения называется *частотой* и обычно обозначается f .

¹ Разные подходы к определению среднего (в том числе среднего арифметического, среднего арифметического взвешенного, среднего гармонического, среднего геометрического и др.) подробнее описаны в учебной литературе по математической статистике, например [20].

Среднее арифметическое взвешенное определяется как

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i},$$

а стандартное отклонение взвешенное – как

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}.$$

С примерами сгруппированных данных читатель познакомится при решении задач, содержащихся в первой части данного учебного пособия.

3.4. Использование статистических пакетов для определения базовых характеристик выборки

Многие статистические методы и процедуры, рассматриваемые в данном учебном пособии, реализованы в пакетах SPSS и STATISTICA. Параллельное рассмотрение возможностей применения обоих пакетов для решения исследовательских задач не связано с тем, что эти пакеты являются взаимоисключающими или, наоборот, взаимодополняющими. Область их функционального пересечения очень широка, и большинство процедур, реализованных в одном пакете, содержится и в другом. Однако каждый из этих пакетов имеет особенности, которые делают решение одних задач более удобным с помощью одного пакета, других – с помощью другого. Например, выходные формы SPSS более удобны для анализа и интеграции в исследовательские отчеты, а пакет STATISTICA по некоторым видам задач (например, анализ временных рядов) предоставляет возможность сделать более развернутый анализ. В ряде случаев в этих пакетах используются различные варианты реализации некоторых статистических методов. Например, метод контрастов, реализованный в SPSS, применяет статистику t -критерия Стьюдента, а аналогичный метод, реализованный в пакете STATISTICA, – статистику F -критерия Фишера.

При изложении материала, связанного с использованием статистических пакетов, несколько большее внимание будет уделено пакету SPSS, поскольку, по мнению автора данного учебного

пособия, этот пакет более простой для освоения и во многих случаях более удобный для работы.

В этом подразделе мы опишем основные процедуры работы с этими пакетами для определения таких базовых (элементарных) статистик, как среднее значение, стандартное отклонение, доверительный интервал и др. Цель этого подраздела – объяснить принципы работы со статистическими пакетами SPSS и STATISTICA, поэтому читатель, умеющий работать с этими программами, может пропустить данный подраздел без ущерба для понимания дальнейшего материала.

Описание процедуры расчета элементарных статистик представим общими сведениями о структуре данных и переменных, с которыми работает пакет SPSS. Это сделано для того, чтобы облегчить процесс освоения пакета SPSS читателям, которые не имеют даже первоначального опыта работы с этим статистическим пакетом. В пакете STATISTICA правила и структура ввода и представления данных аналогичны, поэтому мы ограничимся рассмотрением вопросов о подготовке данных в среде SPSS.

Исходные данные для обработки могут быть введены в пакет SPSS несколькими способами: путем набора данных в таблице SPSS, копированием данных из других приложений, например Excel (используются стандартные команды **Copy–Paste**), открытием файла с данными, конвертированием данных из файлов других форматов и др. Затем рекомендуется определить переменные в соответствии со структурой массива данных и их особенностями. Это можно осуществить, перейдя с вкладки *Data View* на вкладку *Variable View* (или двойным щелчком левой кнопкой мыши на заголовке выбранного столбца). Далее в соответствующих полях можно задать или изменить имя переменной, ее тип, шкалу измерения, формат представления и метку.

По умолчанию каждая переменная имеет имя, состоящее из букв *var* и пяти цифр (00001, 00002 и т. д.). При изменении имени необходимо соблюдать следующие правила:

- имя должно начинаться с буквы;
- имя не может содержать больше восьми символов;
- в нерусифицированных версиях пакета SPSS не рекомендуется использовать кириллицу (по крайней мере при определении переменных), хотя во многих случаях это не вызывает проблем;
- допускается применение цифр и некоторых специальных символов («_», «.», «\$» и др.);

- не допускается использование пробелов;
- в имени не различаются прописные и строчные буквы.

По умолчанию новая переменная является числовой. Тип и формат переменной можно изменить, щелкнув по правому краю поля *Type*. Здесь можно изменить количество цифр после десятичного знака, использовать в качестве десятичного знака точку или запятую, задать параметры переменной в денежном и других форматах и т.д.

Для улучшения идентификации переменных можно и полезно ввести метку (*label*) для используемых переменных. Метки применяются при формировании различных выходных форм и играют роль краткого содержательного описания. Например, в качестве метки можно задать расширенное название изучаемого признака (фактора), а также математическое или логическое выражение, которое применяется для вычисления переменной. При анализе данных метка оказывается очень полезной, так как она позволяет лучше ориентироваться в файле данных.

При сборе данных редко удается избежать отсутствия ряда показателей (измерений) по некоторым наблюдениям (объектам). При использовании многих статистических процедур, прежде всего связанных с анализом одной переменной, это не приводит к сколько-нибудь серьезным осложнениям: просто уменьшается число используемых наблюдений. Однако применение некоторых операций, например вычисление коэффициента корреляции, создает проблемы, так как отсутствие хотя бы одного значения в наблюдении исключает его из дальнейшего анализа. Для обозначения отсутствующих данных в пакете SPSS используется специальный символ (,). Пользователь пакета может определить и другой код для пропущенных данных. Исследователь также может сам указать, что те или иные данные следует считать пропущенными. Например, все значения, лежащие в определенном интервале, указанном пользователем, будут считаться пропущенными (*Range of missing values*). Можно задать фиксированные значения переменной (до трех значений), которые будут считаться пропущенными (*Discrete missing values*), а также комбинацию интервала и фиксированного значения (*Range plus one discrete missing value*). Все эти варианты станут доступны в диалоговом окне *Missing Values*, если щелкнуть по правому краю поля *Missing*. По умолчанию установлена опция *No missing values*, т.е. все значения переменной считаются допустимыми для обработки.

Иногда исследователь предпочитает ограничить анализ имеющихся данных определенным подмножеством, с которым и проводятся желаемые статистические процедуры. Для этого можно воспользоваться командой **Select Cases** из основного меню *Data*. В диалоговом окне *Select Cases* можно определить условия отбора наблюдений и задать несколько режимов отбора:

- *All cases* – все элементы выборки используются для анализа (установлено по умолчанию);

- *Random sample of cases* – из имеющихся данных случайным образом формируется подвыборка (задается определенный процент элементов выборки либо фиксированное число элементов);

- *Based on time or case range* – подмножество данных определяется указанием диапазона изменений номеров наблюдений (элементов исходной выборки);

- *Use filter variable* – подмножество данных формируется на основе задаваемой маски (фильтра). Маска может представлять собой отдельную переменную, включающую такое же количество элементов, что и исходный набор наблюдений. В подмножество данных отбираются те наблюдения, для которых элемент в фильтре имеет ненулевое значение;

- *If condition is satisfied* – отбираются элементы, удовлетворяющие некоторому условию, которое задается логическим выражением (с помощью элементов на панели «калькулятора» в диалоговом окне *Select Cases*: *If* и/или из списка функций *Functions* в этом же окне). Если условие истинно, то наблюдение выбирается, в противном случае – отбрасывается.

Предположим, что когда мы опрашиваем обучающихся о предпочитаемом времени обеда, мы отмечаем также их статус. Пусть это будет переменная *status*, измеряемая в номинативной шкале и имеющая три значения: 1 – студент, 2 – магистрант, 3 – аспирант. Допустим, мы хотим ограничить анализ группой, содержащей только студентов и магистрантов. В этом случае можно задать в диалоговом окне *Select Cases*: *If* следующее логическое выражение: $(status = 1) | (status = 2)$. Знак «|» означает логическое «или». Точно так же можно формулировать условия, используя две и более переменные. Например, мы отмечаем место проживания студентов (переменная *living*) как 1 – общежитие вблизи университета; 2 – в центре города; 3 – на окраине города; 4 – за городом. Допустим, нас по-прежнему интересуют

студенты и магистранты, но только те, кто проживает в общежитии или в центре города. В этом случае мы можем задать следующее условие: $((\text{status} = 1) | (\text{status} = 2)) \& ((\text{living} = 1) | (\text{living} = 2))$. Знак «&» означает логическое «и». Аналогичное условие для данного случая можно задать следующим образом: $(\sim \text{status} = 3) \& ((\text{living} = 1) | (\text{living} = 2))$. Знак «~» означает логическое «не».

Элементы исходной выборки, не попавшие в подмножество, не используются в дальнейших вычислениях, но сохраняются в исходном файле данных. Необработываемые наблюдения легко отличить, так как, во-первых, соответствующий им номер строки перечеркнут, а во-вторых, к массиву данных добавляется новая переменная-фильтр со значениями *Selected* (1) и *Not Selected* (0) соответственно для элементов, как вошедших, так и не вошедших в подмножество. Данная операция действует только в пределах текущего сеанса работы. При повторной загрузке файла вновь становится доступной вся исходная выборка. Переменная-фильтр при этом сохраняется, но она оказывается неактивна (для ее активизации можно использовать режим *Use filter variable*). В случае необходимости можно удалить все элементы, не попавшие в анализируемую подгруппу, выделив в поле *Unselected Cases Are* режим *Deleted*. Однако это делать не рекомендуется, так как в процессе анализа приходится часто изменять или переконфигурировать исходный набор данных.

Теперь рассмотрим процедуру вычисления элементарных статистик в пакете SPSS и проверку нормальности выборочных данных на ранее приведенном примере. Переменные обозначены как *person* (идентификатор респондента) и *time* (время обеда), переменным присвоены метки соответственно *Respondents* и *Lunch time*. В основном меню *Analyze* выбирается опция *Descriptive Statistics*, затем опция *Explore*. В поле *Dependent List* переносится переменная *Lunch time* (рис. 3.4).

Далее в диалоговом окне *Statistics* (переход осуществляется щелчком по соответствующей кнопке в окне *Explore*) устанавливается уровень статистической значимости в процентах (по умолчанию 95%). В окне *Plots* устанавливаются режимы графического представления исходных данных (например, полигон частот) и проверки распределения данных на нормальность. Для проверки нормальности в этом окне выбираем режим (устанавливаем флажок) *Normality plots with tests*.

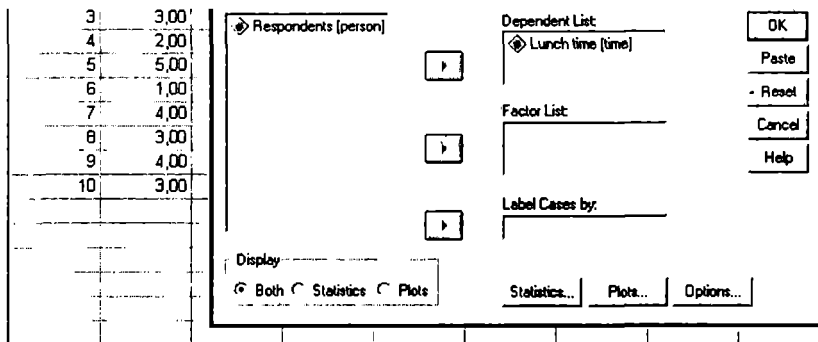


Рис. 3.4. Диалоговое окно *Explore*

После щелчка по кнопке *Continue* в окне *Plots*, а затем по кнопке *OK* в окне *Explore* формируется отчет *Descriptives* (табл. 3.2 в котором представлены среднее значение по выборке (3,00 стандартное отклонение (1,155), стандартная ошибка средних (0,365), нижняя и верхняя границы доверительного интервала ($3,0 - 0,826 = 2,174$ и $3,0 + 0,826 = 3,826$), минимальное (1,0) максимальное (5,0) значения элементов выборки, их размах (4,0

Таблица 3

Описательные статистики в отчете *Descriptives*

			Statistic	Std. Error
Lunch time	Mean		3,0000	,36515
	95% Confidence Interval for Mean	Lower Bound	2,1740	
		Upper Bound	3,8260	
	Median		3,0000	
	Variance		1,333	
	Std. Deviation		1,15470	
	Minimum		1,00	
	Maximum		5,00	
	Range		4,00	

О нормальности распределения можно судить по графику *Normal Q-Q Plot* (рис. 3.5). В случае нормального распределения выборочных данных все точки должны лежать на одной прямой или рядом с ней. Чем больше отклонение точек от прямой линии, тем меньше выборочное распределение соответствует нормальному. В данном случае отклонение от нормального можно считать незначительным.

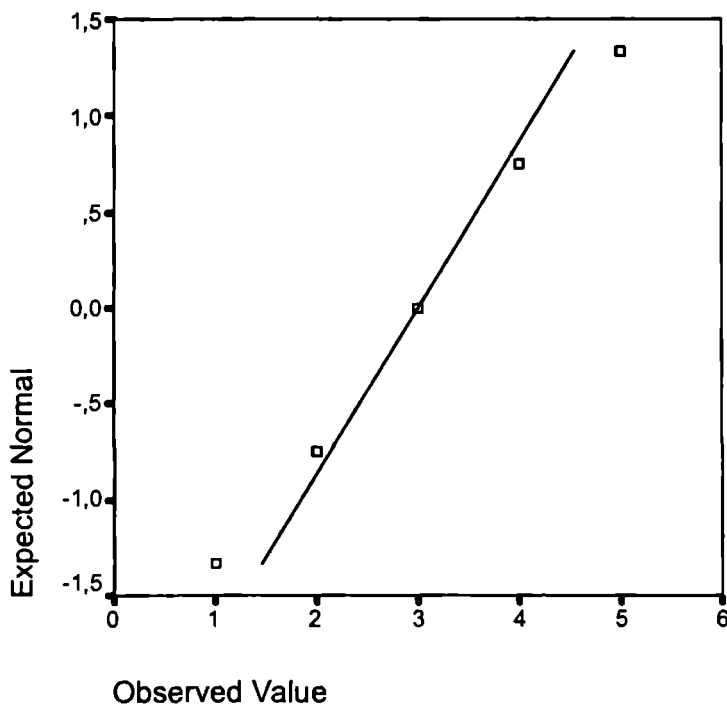


Рис. 3.5. График для оценки нормальности распределения

О нормальности распределения также можно судить с помощью специальных тестов. В выходной форме пакета SPSS приводятся результаты проверки выборочного распределения на нормальность двумя методами: Колмогорова–Смирнова (Kolmogorov–Smirnov test) и Шапиро–Уилка (Shapiro–Wilk test). Последний тест считается более гибким и универсальным. В табл. 3.3 приведена статистика по этим методам.

Статистические критерии для оценки нормальности распределения

	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Lunch time	,200	10	,200	,953	10	,703

Если уровень значимости, отражаемый в столбце Sig. (Significance), больше 0,05, то можно считать, что вид анализируемого выборочного распределения существенно не отличается от нормального (подробно смысл этого показателя рассматривается во второй части данного учебного пособия).

Идентичные результаты могут быть получены с помощью пакета STATISTICA, хотя и в менее удобной форме представления. Для этого нужно включить модуль *Basic Statistics* (выбирается при запуске пакета) и в меню *Analysis* выбрать опцию *Descriptive Statistics*. Для того чтобы указать переменную для обработки, нужно щелкнуть по кнопке *Variables* и в открывшемся окне выбрать соответствующую переменную (в данном случае — TIME).

После щелчка по кнопке *More Statistics* можно выбрать статистические показатели для расчета.

Valid N – число элементов выборки, пригодных для обработки;

Mean – среднее значение;

Sum – сумма значений переменной;

Median – медиана;

Standard Deviation – стандартное отклонение;

Variance – дисперсия;

Standard error of mean – стандартная ошибка среднего;

95% confidence limits of mean – 95%-ный доверительный интервал (уровень статистической значимости устанавливается в поле *Interval*);

Minimum & maximum – минимальное и максимальное значения выборки;

Lover & upper quartiles – верхний и нижний квартили;

Range – размах;

Quartiles range – разность между верхним и нижним квартилями;

Skewncss – коэффициент асимметрии;

Kurtosis – коэффициент эксцесса;

Standard error of skewness – стандартная ошибка коэффициента асимметрии;

Standard error of kurtosis – стандартная ошибка коэффициента эксцесса.

Результаты вычислений представляются в выходной форме (рис. 3.6).

Descriptive Statistics (lunch_time.sta)									
Continue...	Valid N	Mean	Confid. -95,000%	Confid. +95,000%	Minimum	Maximum	Variance	Std. Dev.	Standard Error
TIME	10	3,0000	2,1740	3,8260	1,0000	5,0000	1,3333	1,1547	,3651

Рис. 3.6. Результаты расчета описательных статистик в программе STATISTICA

Гистограмма частот может быть получена щелчком по кнопке *Histograms*. Кривая на гистограмме (*Expected Normal*) – график плотности нормального распределения со средним значением 3,0 и стандартным отклонением 1,155 (рис. 3.7). На гистограмме также представлены статистики критериев Колмогорова–Смирнова и Шапиро–Уилка.

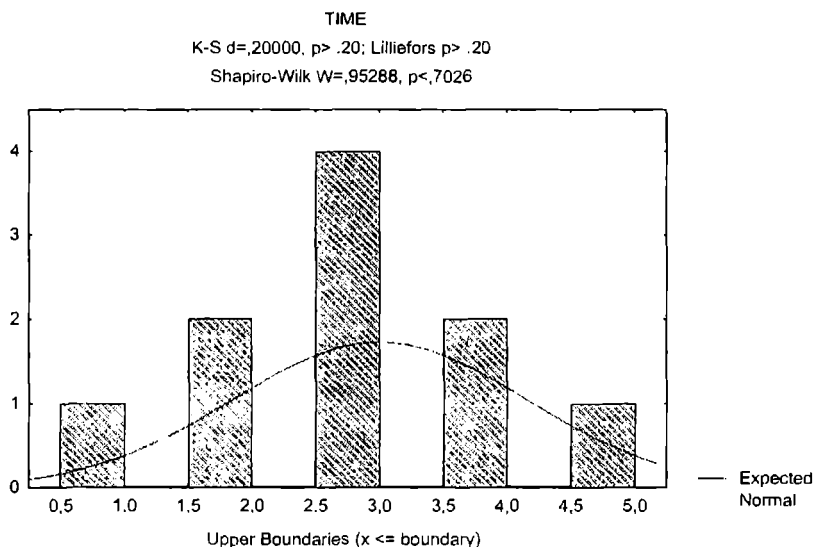


Рис. 3.7. Гистограмма и критерии для оценки нормальности распределения в программе STATISTICA

3.5. Определение необходимого объема выборки

Для определения необходимого количества элементов в выборке нужно задать желаемую точность, характеризующую предельной ошибкой среднего (σ_x). Поскольку этот параметр связан с объемом выборки, мы можем оценить необходимое количество элементов в выборке исходя из требуемой точности оценки исследуемого параметра генеральной совокупности.

Пример 3.4. Предположим, что необходимо обеспечить точность оценки исследуемого параметра (предпочтительное время обеда) ± 15 мин. с вероятностью 95% (см. подразд. 3.3, табл. 3.1). Тогда мы задаем предельную ошибку среднего $\delta = \pm 0,5$ (± 15 мин.) и соответственно $t = 2,26$ (так как $n = 10$).

Далее определяем минимально необходимый размер выборки:

$$\begin{aligned}\delta &= \sigma_x t; \\ \sigma_x &= \delta / t; \\ \delta / t &= \frac{s}{\sqrt{n}}; \\ n &= \left(\frac{s}{\delta / t} \right)^2\end{aligned}$$

По данным, полученным в результате выборочного исследования, находим

$$n = \left(\frac{1,155}{0,5 / 2,26} \right)^2 = 27,25.$$

Таким образом, объем выборки должен быть не менее 28 чел. для того, чтобы обеспечить точность оценки ± 15 мин. с вероятностью 95%. Если необходимо увеличить точность, например до ± 10 мин. с той же вероятностью, получим большее значение n (в качестве упражнения читателю предлагается самостоятельно решить эту задачу).

Если предварительное тестирование не проводится, но из предыдущих исследований известна величина стандартного отклонения или дисперсии значений изучаемого признака, однако

при этом не приводится объем выборки, по которому эта величина получена, значение минимального объема выборки для обеспечения необходимой точности можно получить следующим образом. Вначале выбирается значение z -критерия для заданной вероятности P (или, что то же самое, для уровня статистической значимости $p = 1 - P$), затем определяется первоначальное значение объема выборки (n_1) по формуле

$$n_1 = \left(\frac{s}{\delta / z} \right)^2$$

Если найденное значение меньше 100, то вместо z -критерия используется t -критерий (определяемый для найденного значения n_1 и заданного p), и минимальный объем выборки уточняется:

$$n_2 = \left(\frac{s}{\delta / t} \right)^2$$

Если при этом значение n_2 значительно больше n_1 , и при этом $n_2 < 100$, то можно провести еще одну итерацию уточнения минимального объема выборки с помощью t -критерия, который определяется для n_2 , p и т.д.

Данный подход к определению необходимой численности выборки предполагает, что выборка была сформирована по схеме повторного вероятностного отбора. При формировании выборки методом бесповторного отбора требуемый объем выборки определяется так (при использовании z -критерия):

$$n = \frac{s^2 N}{(\delta / z)^2 N + s^2}.$$

Читателю рекомендуется получить этот результат самостоятельно из формулы для определения предельной ошибки выборки при бесповторном отборе, приведенной в подразд. 3.3.

Пример 3.5. Рассмотрим еще один пример, связанный с гипотетическим студенческим проектом «Университетский ланч-клуб». В данном случае в качестве измеряемого параметра будет выступать «Частота посещений ланч-клуба»: «Как часто (сколько раз в месяц) Вы посещали бы ланч-клуб?»

Допустим, по результатам предварительного тестирования ($n = 64$) получено: $\bar{x} = 10,0$, $s = 4,1$. Отсюда находим

$$\sigma_x = \frac{s}{\sqrt{n}} = \frac{4,1}{\sqrt{64}} = 0,51.$$

Таким образом, с вероятностью 95% среднее по генеральной совокупности будет попадать в интервал 10 ± 1 визит в месяц, т.е. от 9 до 11 визитов (при расчете использовано значение $t(n = 64, p = 0,05) = 2,00$). С вероятностью 99% среднее по генеральной совокупности будет попадать в интервал $10 \pm 2,66$ 0,51 визит в месяц (т.е. на каждого посетителя будет приходиться от 8,64 до 11,36 визита в месяц).

Предположим, необходимо обеспечить точность оценки параметра $\pm 0,5$ с вероятностью 95%. Тогда $n = (4,1/0,25)^2 = 269$, т.е. необходимый объем выборки – 269 чел. Если исследователя устраивает меньшая точность, например ± 1 визит, тогда для выборочного исследования достаточно 65 чел., что для данного примера является более разумным решением с практической точки зрения.

Если необходимо определить требуемый объем выборки при одновременном измерении нескольких параметров (например, в опросный лист включены два пункта: «предпочтительное время обеда» и «предполагаемое количество посещений»), то определяется количество респондентов, требуемое для получения результатов заданной точности для каждого пункта. Максимальное из найденных значений и будет определять необходимую численность выборки.

Средняя ошибка выборки при формировании стратифицированной (типической) выборки и бесповторном случайном отборе внутри типических групп рассчитывается по формуле

$$\sigma_x = \frac{\bar{s}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}},$$

где \bar{s} – среднее стандартное отклонение по типическим группам.

Аналогично стандартная ошибка среднего при формировании стратифицированной (типической) выборки и повторном случайном отборе внутри типических групп рассчитывается по формуле

$$\sigma_x = \frac{\bar{s}}{\sqrt{n}}.$$

Тогда необходимый объем стратифицированной выборки при повторном случайном отборе определяем как

$$n = \left(\frac{\bar{s}}{\delta/z} \right)^2$$

Аналогично определяется необходимый объем стратифицированной выборки при бесповторном случайном отборе (читателю предлагается сделать это в качестве упражнения).

Определение численности каждой типической группы при условии пропорциональной численности единиц типических групп в стратифицированной выборке находим из условия

$$n_i = n \frac{N_i}{N},$$

где n_i – объем выборки из типической группы;

n – общий объем выборки;

N_i – объем типической группы;

N – объем генеральной совокупности.

Достаточно часто выборочное исследование используется для определения параметров, представляющих собой *пропорции* от генеральной совокупности. Для иллюстрации вновь вернемся к примеру со студенческим проектом «Университетский ланч-клуб». Предположим, инициаторы этого проекта задались вопросом: «Какой процент студентов университета будет посещать ланч-клуб?»

Стандартная ошибка вычисления пропорции при повторном отборе определяется как

$$\sigma_q = \sqrt{\frac{q_1 q_2}{n}},$$

где q_1 – доля элементов генеральной совокупности, обладающих измеряемым атрибутом;

q_2 – доля элементов генеральной совокупности, не обладающих измеряемым атрибутом; $q_1 = 1 - q_2$.

Сравнивая эту формулу с формулами для определения ошибки среднего, рассмотренными ранее, легко увидеть, что произведение $q_1 q_2$ имеет смысл дисперсии по выборке. Поэтому в литературе это произведение часто называется *групповой дисперсией*.

Отсюда получаем

$$n = \frac{q_1 q_2}{\sigma_q^2}.$$

Пример 3.6. Предположим, предварительное тестирование выявило, что 30% студентов выказали желание посещать клуб ($q_1 = 0,3$), но точное значение респондентов, принимавших участие в опросе, неизвестно. Установленная точность оценки: $\delta = \pm 0,1$ (10%), $z = 1,96$ (95%). Тогда $\sigma_q = 0,1/1,96 = 0,051$ и $n = 81$. Дальнейшее уточнение этого значения с использованием t -критерия дает $n = 83$.

В случае, если предварительное тестирование не проводится, можно руководствоваться следующей логикой: произведение $q_1 q_2$ никогда не превысит 0,25, поэтому размер выборки можно оценить как

$$n = \frac{0,25}{\sigma_q^2}.$$

При бесповторном отборе стандартная ошибка определения выборочной доли составит

$$\sigma_q = \sqrt{\frac{q_1 q_2}{n} \left(1 - \frac{n}{N}\right)}.$$

Отсюда легко может быть получено значение необходимого объема выборки для изучения выборочной доли (пропорции) в случае формирования выборки методом случайного бесповторного отбора:

$$n = \frac{q_1 q_2 N}{\left(\frac{\delta}{z}\right)^2 N + q_1 q_2}.$$

Аналогично при стратифицированном способе отбора с пропорциональным отбором внутри типических групп средняя ошибка выборки для пропорции определяется по формуле (отбор внутри групп – случайный повторный)

$$\sigma_q = \sqrt{\frac{q_1 q_2}{n}}.$$

Здесь $\overline{q_1 q_2}$ – среднее значение по группам (рассчитывается как среднее арифметическое взвешенное). Отсюда легко получаем необходимую численность выборки для определения пропорции

$$n = \frac{\overline{q_1 q_2}}{\sigma_q^2} = \frac{\overline{q_1 q_2}}{(\delta/z)^2}.$$

При случайном бесповторном отборе формула стандартной ошибки определения выборочной доли модифицируется:

$$\sigma_q = \sqrt{\frac{\overline{q_1 q_2}}{n} \left(1 - \frac{n}{N}\right)}.$$

Соответственно значение необходимого объема стратифицированной выборки при определении пропорции в случае бесповторного отбора определяется как

$$n = \frac{\overline{q_1 q_2} N}{(\delta/z)^2 N + \overline{q_1 q_2}}.$$

Таким образом, минимально необходимый объем выборки определяется не только требуемой точностью измерения, но и методом формирования выборочной совокупности.

Задания для самостоятельного выполнения

1. В табл. 3.4 приведен список элементов генеральной совокупности, представляющий собой возраст участников студенческих строительных отрядов, сформированных на базе вузов города в 2000 г. ($N = 100$).

Таблица 3.4

Данные для выполнения задания 1

№ п/п	Возраст	№ п/п	Возраст	№ п/п	Возраст	№ п/п	Возраст	№ п/п	Возраст
1	17	21	19	41	22	61	20	81	21
2	20	22	22	42	19	62	18	82	20
3	20	23	19	43	19	63	20	83	25
4	19	24	23	44	20	64	19	84	20
5	20	25	19	45	20	65	20	85	22
6	18	26	20	46	23	66	22	86	18

7	21	27	25	47	17	67	17	87	21
8	19	28	18	48	18	68	21	88	20
9	20	29	20	49	21	69	21	89	22
10	19	30	20	50	20	70	21	90	21
11	17	31	19	51	22	71	20	91	21
12	20	32	22	52	19	72	18	92	20
13	20	33	19	53	19	73	20	93	24
14	19	34	23	54	20	74	19	94	20
15	20	35	19	55	20	75	20	95	22
16	18	36	20	56	23	76	22	96	18
17	21	37	23	57	17	77	17	97	21
18	19	38	18	58	18	78	21	98	20
19	20	39	26	59	21	79	24	99	22
20	19	40	20	60	20	80	21	100	21

Методом, обеспечивающим случайный отбор (например, с помощью функции-генератора случайных чисел программы Excel), сформируйте по меньшей мере 10 выборок объемом 10 элементов, рассчитайте среднее для каждой выборки. С помощью стандартных графических инструментов Excel постройте по этим данным полигон частот (ломаную линию, получаемую при соединении точек, лежащих по оси абсцисс на верхних границах каждого класса, а по оси ординат – частот классов). Построенная кривая (ломаная) будет представлять собой распределение всех возможных выборочных средних. Соответствует ли форма полученной кривой центральной предельной теореме?

Рассчитайте среднее значение всех выборочных средних и среднее арифметическое генеральной совокупности. Сравните результаты. Соответствуют ли эти результаты центральной предельной теореме?

Рассчитайте стандартное отклонение для распределения выборочных средних и для генеральной совокупности. Сравните полученные значения. Соответствует ли этот результат центральной предельной теореме?

Цель задания – закрепление понимания центральной предельной теоремы. Согласно этой теореме распределение средних значений выборок, извлекаемых из одной и той же совокупности, при достаточно большом числе таких выборок соответствует нормальному распределению.

Пояснение. При выделении 10 или более выборок по 10 или более элементам может быть построен полигон частот, например следующего вида (рис. 3.8), на основе которого можно судить о приближительном соответствии нормальному распределению.

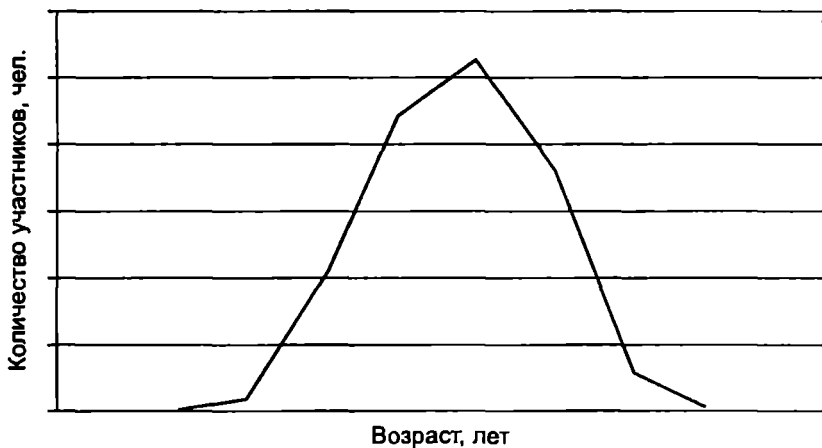


Рис. 3.8. Полигон частот

Данные из списка рекомендуется ввести в два столбца: идентификационный номер (номер по порядку) и возраст. Осуществлять случайный отбор из списка можно с помощью функции Excel СЛЧИС(). Необходимо учитывать диапазон генерируемых чисел (1 – 100). Для этого используем выражение $\text{СЛЧИС}()*99 + 1$. В результате может получиться дробное число. Для определения номера студента из списка это число нужно округлять, воспользовавшись функцией Excel ОКРУГЛ(). Для данной задачи округление проводится до целого, поэтому можно использовать выражение $= \text{ОКРУГЛ}(\text{СЛЧИС}()*99 + 1; 0)$, которое помещается, например, в ячейку С1.

Далее для автоматизации процесса выбора элементов из списка, чтобы не искать вручную, можно воспользоваться функцией определения элемента в таблице по значению (функция вертикального переноса – ВПР(значение; массив; номер столбца)). В качестве первого аргумента функции можно указать $\text{ОКРУГЛ}(\text{СЛЧИС}()*99 + 1; 0)$, массив – это набор идентификационных номеров студентов по списку и их возраст (например, $\$A\$1:\$B\100 ; номер столбца для данного случая – 2). В итоге получаем: $= \text{ВПР}(C1; \$A\$1:\$B\$100; 2)$.

Эта формула копируется в 10 ячеек (если формируется выборка из 10 элементов) и определяется среднее значение = СРЗНАЧ (массив из 10 ячеек). Щелкая по клавише F9, можно изменять выборку и получать каждый раз различное среднее значение по выборке из 10 элементов.

Для построения графика можно разбить диапазон возраста на равные промежутки, например 10 интервалов ((максимальный возраст – минимальный возраст) / 10), или выбрать фиксированную длину интервала, например 0,5. Затем подсчитывается частота попадания определяемого выборочного среднего в интервал. Чем больше случайных выборок вы сделаете и чем больше интервалов возьмете, тем ближе полигон частот будет напоминать нормальное распределение. Это упражнение будет более показательным при извлечении большего количества выборок (20, 50, 100) или при формировании выборки с большей численностью (например, 15 или 20). Однако в этом случае упражнение будет более трудоемким.

2. В одном из районов города проживает 2000 семей. С целью определения среднего размера семьи района методом случайного повторного отбора было проведено 3%-ное выборочное исследование, результаты которого представлены в табл. 3.5.

Таблица 3.5

Данные для выполнения задания 2

Размер семьи, чел.	1	2	3	4	5	6	7	8
Число семей	4	8	15	17	8	4	2	1

С вероятностью 95% определите пределы, в которых находится средний размер семьи в районе. (При расчете среднего значения и стандартного отклонения по выборке используйте метод расчета взвешенных показателей.)

Сделайте предположения о возможных сферах практического использования результатов, полученных в ходе решения задачи. С какой целью могли проводиться подобные исследования? Насколько полученные результаты могут быть полезны для достижения поставленных целей?

3. Определите необходимое количество респондентов для опроса по данным, приведенным в подразд. 3.5. Измеряемые параметры: «Предпочтительное время обеда» и «Количество посещений». Требуемая точность: ± 10 мин. и ± 1 визит соответственно с вероятностью 95%. Выполните это же задание при

предположении, что выборка формируется путем случайного бесповторного отбора (принять, что $N = 15000$). Сравните результаты.

4. Принимается решение о целесообразности строительства парка семейного отдыха в одном из районов города. Для определения среднего числа детей в семье проводится соответствующее исследование. Известно, что в районе проживает 2500 семей. Сформирована 2%-ная случайная бесповторная выборка семей. Данные, полученные в результате исследования, приведены в табл. 3.6.

Таблица 3.6

Данные для выполнения задания 4

Число детей в семье	Количество семей
0	10
1	20
2	12
3	4
4	2
5	2

С вероятностью 95% определите границы, в которых будет находиться среднее число детей в семье во всем районе. (При расчете среднего значения и стандартного отклонения по выборке используйте метод расчета взвешенных показателей.)

5. Для определения средней продолжительности поездки населения города на работу предполагается провести выборочное исследование методом случайного повторного отбора. Какова должна быть численность выборки, чтобы с вероятностью 95% предельная ошибка среднего не превышала 5 мин. при стандартном отклонении 20 мин. (данные предыдущего обследования). С какой целью могло проводиться подобное исследование?

6. В городе проживают около 10 тыс. семей (точные данные неизвестны). С помощью повторной случайной выборки предполагается определить долю семей в городе с числом детей три и более. Какова должна быть численность выборки, чтобы с вероятностью 95% предельная ошибка среднего не превышала 0,02 чел., если на основе предыдущих исследований известно, что дисперсия равна 0,2. Сделайте предположения о возможных сферах практического использования результатов, полученных

в ходе решения задачи. С какой целью могли проводиться подобные исследования? Насколько полученные результаты могут быть полезны для достижения поставленных целей?

7. На предприятии с числом рабочих 15 тыс. чел. предполагается определить долю рабочих со стажем работы 20 лет и более. Данные по рабочим еще не внесены в компьютер и в основном находятся в бумажных папках, поэтому определить за короткое время точное значение требуемого параметра невозможно. Предполагается сформировать выборку из имеющихся данных методом бесповторного случайного отбора. Какова должна быть численность выборки, чтобы с вероятностью 95% предельная ошибка среднего не превышала 0,03? Из предыдущих обследований известно, что дисперсия равна 0,2. Выполните это же задание при предположении, что выборка формируется методом повторного случайного отбора. Сравните результаты и сделайте выводы.

8. Для определения средней заработной платы работников трех различных категорий была проведена 20%-ная типическая выборка с отбором единиц пропорционально численности типических групп (внутри типов применялся метод случайного бесповторного отбора). Результаты выборки представлены в табл. 3.7.

Таблица 3.7

Данные для выполнения задания 8

Тип работы	Средняя заработная плата, руб.	Стандартное отклонение, руб.	Число работников, чел.
1	5100	650	150
2	6500	1900	500
3	7800	1150	350

С вероятностью 95% определите пределы, в которых находится средняя заработная плата всех работников каждой категории. Можно ли с вероятностью 95% считать среднюю заработную плату работников всех категорий одинаковой? Обоснуйте ответ, используя понятие «доверительный интервал».

9. Для выявления затрат времени на обработку деталей рабочими разных квалификаций на промышленном предприятии была проведена 10%-ная типическая выборка пропорционально численности выделенных групп (на основе случайного повторного отбора). Результаты исследования представлены в табл. 3.8.

Данные для выполнения задания 9

Квалификация рабочих	Численность выборки	Средние затраты времени на обработку одной детали, мин.	Стандартное отклонение, мин.
I	6	10	1
II	12	14	4
III	8	15	2
IV	4	22	5

В каких пределах с вероятностью 95% может варьироваться среднее время на обработку деталей рабочими разной квалификации? Можно ли считать, что средние затраты времени на обработку деталей значительно различаются у рабочих разной квалификации? Обоснуйте ответ, используя понятие «доверительный интервал».

10. В городе проживают 12 тыс. жителей, из них 7 тыс. женщин и 5 тыс. мужчин. С целью определения доли жителей в возрасте старше 60 лет предполагается провести типическую выборку жителей с пропорциональным отбором. Отбор внутри типов – случайный бесповторный. Какое количество жителей необходимо отобрать, чтобы с вероятностью 95% предельная ошибка не превышала 5%? На основе предыдущих исследований принимается, что среднегрупповая дисперсия составляет 1600 (указано с учетом того, что доли представляются в процентах).

11. В городе проживают 5 тыс. семей, из них 1 тыс. составляют семьи без детей и 4 тыс. – с детьми. С целью определения доли семей, имеющих отдельные квартиры, предполагается провести стратифицированную выборку семей с пропорциональным отбором внутри типических групп. Отбор внутри групп – случайный повторный. Какое количество семей необходимо отобрать, чтобы с вероятностью 95% предельная ошибка не превышала 5%? На основе предыдущих исследований принимается, что среднегрупповая дисперсия по типическим группам составляет 1000 (указано с учетом того, что доли представляются в процентах).

12. Проверьте нормальность распределения данных (см. табл. 3.4) графическим методом и с помощью статистических критериев. Задание выполните последовательно для первых 20, 50, 100 элементов таблицы. Постройте доверительный интервал для каж-

дого случая. Сравните результаты. Проследите, как количество элементов в выборке влияет на гистограмму частот, график SPSS «Normal Q-Q Plot» и уровень статистической значимости в отчетах SPSS о проверке выборочного распределения на нормальность.

13. Нормализуйте данные из табл. 3.4 (приведите к стандартной z -шкале) и повторите задание 12 для нормализованных данных. Сравните результаты и сделайте выводы.

14. По данным, представленным в табл. 3.9, проверьте нормальность распределения оценок. Нормализуйте данные (приведите к стандартной T -шкале) и повторите процедуру проверки. Сравните результаты и сделайте выводы.

15. Крупная компания, имеющая несколько тысяч клиентов, проводит ежегодный мониторинг, насколько клиенты удовлетворены предлагаемым сервисом. В 2005 г. опрос 280 случайно отобранных клиентов показал, что 54% клиентов высказали удовлетворенность, а в 2006 г. по результатам опроса 260 клиентов получена цифра 58%. Можно ли на основе данных результатов сделать вывод о том, что удовлетворенность клиентов качеством предоставляемого сервиса действительно выросла?

Таблица 3.9

Данные для выполнения задания 14

Студент	Предмет 1	Предмет 2	Студент	Предмет 1	Предмет 2
1	84	98	15	97	95
2	88	93	16	102	98
3	88	92	17	102	95
4	89	94	18	103	97
5	89	77	19	105	99
6	90	89	20	107	112
7	93	91	21	107	104
8	93	91	22	107	97
9	95	111	23	109	109
10	95	99	24	111	114
11	95	82	25	111	115
12	96	106	26	115	112
13	96	95	27	116	115
14	97	86	28	130	129

**Оценка качества измерений
и измерительных инструментов
в социально-экономических исследованиях****4.1. Показатели качества измерений**

Началом системных исследований в области качества измерений считаются работы Г. Гельмгольца (1821–1894), связанные с вычислением ошибок измерений в естественно-научных исследованиях. Большинство «классических» методов оценки качества измерений было разработано в первой половине XX в. Несмотря на то что история целенаправленных исследований проблем качества измерений насчитывает уже более ста лет, эти проблемы остаются актуальными и сегодня, особенно с методологической точки зрения. В теории измерений разработано довольно много методов и подходов к оценке качества измерений. Использование таких подходов для обоснования полученных результатов в зарубежной исследовательской практике считается обязательным. Проблема качества измерений приобретает актуальность не только после их осуществления, но и на этапе проектирования, в том числе на стадии разработки методов измерения. К сожалению, в нашей литературе этим методам и подходам уделяется недостаточное внимание, что является одной из основных причин и некачественных эмпирических экономических и социально-психологических исследований, описание которых иногда встречается в литературе, и создания сомнительных тестов и опросных листов, которые порой используются на практике. Именно поэтому данный вопрос стал предметом рассмотрения в рамках отдельной главы.

Как отмечалось в гл. 1, социально-экономические измерения часто являются косвенными. Непосредственно мы фиксируем некоторые признаки (время решения задачи, ответы на определенные вопросы, оценки респондентов и т.д.) и на этой основе делаем выводы о некотором скрытом (*латентном*) свойстве, недоступном прямому измерению (например, агрессивность, коммуникативность, креативность, благоприятные или неблагоприятные качества организационного климата для творчества и инноваций и т.п.). В частности, в социально-экономических

эмпирических исследованиях широко используется измерение отношений респондентов, на основе чего и принимается или отвергается та или иная исследовательская гипотеза. Наиболее распространенными инструментами для таких измерений выступают разнообразные анкеты, опросные листы и т.п. Косвенный характер измерений осложняет оценку их качества.

Качество измерений определяется их *обоснованностью* или *валидностью* (*validity*)¹ и *надежностью* (*reliability*). Качество полученных результатов социально-экономических исследований зависит также и от *точности* (*accuracy*) проводимых измерений. Стандартная ошибка среднего по истинным значениям (*стандартная ошибка измерений*), которая характеризует точность измерений, может быть определена исходя из тех же принципов, которые использовались для определения доверительного интервала и рассматривались в гл. 3. Для оценки точности измерений определяется среднее значение по оцениваемой группе показателей (пунктов анкеты) и стандартное отклонение средних по этой группе. Поскольку стандартное отклонение средних значительно меньше стандартного отклонения индивидуальных откликов, считается достаточным 68%-ный доверительный интервал.

В некоторых работах перечисленные термины (особенно «надежность») используются не только для характеристики качества *измерений*, но и для характеристики *моделей и методов измерений*. Подобное «расширенное толкование» понятий, характеризующих качество измерений, встречается преимущественно в «прикладных» статьях, где предлагается и описывается новый метод или инструмент для измерения тех или иных психологических или социальных феноменов. По качеству результатов измерений делают заключение о качестве примененного метода. Примерами такого подхода являются работы по измерению характеристик организационного климата и когнитивных стилей решения проблем [30; 32; 41; 46; 48].

¹ В русскоязычной литературе (преимущественно переводной), связанной с рассматриваемой темой, используется транслитерация термина «*валидность*». Поскольку этот термин представляется уже достаточно известным, закрепившимся и привычным, он употребляется и в данном учебном пособии.

В ряде зарубежных теоретико-методологических и учебных работ, посвященных вопросам социальных измерений, также можно встретить определения надежности и валидности как меры качества инструмента измерений или даже его составляющих, например вопросов анкеты или заданий теста [34; 39; 47; 51]. В этих работах термин «валидный», или «надежный», относят к инструменту в том смысле, что этот инструмент обеспечивает валидные (надежные) измерения.

Российский специалист в области теорий педагогических измерений профессор В.С. Аванесов полагает методологически неправомерным использование характеристик качества измерений для характеристики качества инструментов измерения. Он считает, что правильнее обсуждать вопрос не надежности или валидности педагогических тестов, а надежности или валидности тестовой информации (результатов), поскольку результаты зависят не только от качества теста, но и от условий, в которых он применяется, от выборки испытуемых и т.д. [2]. Действительно, критерии валидности, надежности и точности должны относиться к результатам, а не к методам и инструментам измерения. Характеристики методов оцениваются через анализ результатов измерений, но прямой перенос не всегда корректен. Хорошим методом могут быть получены некачественные измерения, и наоборот некачественный метод в ряде случаев может продуцировать результаты, которые с формально-статистической точки зрения можно считать хорошими. Поэтому все характеристики и методы оценки качества, рассматриваемые в данном учебном пособии, в первую очередь надо относить к результатам измерений.

4.2. Валидность как характеристика измерений

Обоснованность, или *валидность (validity)*, – это эквивалентность измерений (подходов к измерению, соответствующего инструментария и результатов измерения) характеристикам измеряемого объекта. Другими словами, это мера соответствия оценок, получаемых в процессе измерения, представлениям о сущности свойств исследуемых объектов и их роли в исследуемых процессах. Оценивая обоснованность измерений (а на этой основе – и качество измерительного инструмента), мы отвечаем на вопрос: «Действительно ли мы измеряем то, что предполагаем измерять?»

В общем случае валидность – это интерпретационная и неформализуемая характеристика [50]. Оценка и аргументация валидности измерений носят преимущественно описательный характер, хотя в некоторых случаях используются и математико-статистические методы. Тем не менее эта характеристика является чрезвычайно важной при оценке качества измерений и проектировании измерительного инструмента.

В литературе встречается свыше десяти терминов, обозначающих типы валидности. При этом их четкая классификация отсутствует. Иногда одним и тем же термином в разных источниках обозначается разное содержание, а иногда, наоборот, различные термины наполняются одинаковым содержанием, характеризующим валидность. В данном подразделе выделяются и рассматриваются следующие типы валидности: внешняя, содержательная, критериальная и концептуальная. Поскольку русскоязычная терминология в этой области окончательно не сложилась, в рассматриваемых ниже типах валидности указывается исходный англоязычный термин.

Внешняя валидность (*face validity*) характеризует восприятие вопросов анкеты непрофессионалами в той области, в которой планируется проводить исследования. В то же время это люди – объекты той генеральной совокупности, которую предполагается исследовать. Данное понятие характеризует, как пункты анкеты (вопросы, утверждения и т.д.) воспринимаются и понимаются респондентами.

Содержательная валидность (*content validity*) показывает, насколько утверждения или вопросы анкеты, проектируемой для проведения измерения, соответствуют сути (содержанию) измеряемых показателей. Это, как и внешняя валидность, неформализуемая характеристика, но в отличие от предыдущей она оценивается экспертами, т.е. специалистами в той области, в которой проводится исследование.

Содержательная валидность оценивается на основе квалификации, знания предмета, опыта (исследователя, экспертов или участников фокус-группы). И если исследователь, эксперт или участники фокус-группы делают обоснованное заключение о содержательной валидности, то это, как правило, считается достаточным.

Если, например, мы разрабатываем инструмент для оценки инновационности нового продукта, выводимого на рынок, то

для оценки содержательной валидности можно обратиться к специалистам в области творчества, инновационного менеджмента, маркетинга. Как правило, на первом этапе проектирования исследования и разработки соответствующего измерительного инструмента проводится интервьюирование (структурированное или полуструктурированное) нескольких специалистов в области предполагаемого исследования для того, чтобы определить основные факторы и характеристики, которые являются наиболее важными с точки зрения предмета предполагаемого исследования, т.е. для обеспечения содержательной валидности.

Проверка внешней и содержательной валидности — это первые и обязательные элементы при разработке любого измерительного инструмента. Возможно, что обоснование содержательной валидности будет предшествовать проверке внешней валидности.

Пример 4.1. Предположим, что отдел внеучебной работы намерен провести оценку того, насколько студенты довольны проживанием в общежитии университета. Для обеспечения содержательной валидности можно сформировать фокус-группу из нескольких (5–10) студентов, проживающих в общежитии, чтобы выяснить, какие факторы и аспекты наиболее важны для них. После этого может быть составлен черновой вариант анкеты и предложен этим студентам для ознакомления с тем, чтобы они высказали свое мнение.

Критериальная валидность (criterion-related validity) рассматривает качество измерений с позиций двух эмпирических критериев, а именно:

а) возможность предсказывать те или иные результаты на основе полученных измерений — прогностическая валидность (*predictive validity*);

б) соответствие полученных результатов неким «золотым стандартам», т.е. результатам измерений этого же свойства, полученным ранее (разного рода психометрические или социологические индексы, статистические распределения и т.п.); либо соответствие результатов измерений результатам, полученным уже испытанным и признанным инструментом, используемым параллельно проводимым измерениям, — *согласованная валидность (concurrent validity)*.

Эти два вида критериальной валидности могут оцениваться с помощью статистических показателей связи, например коэффициента корреляции. Считается, что значение коэффициента

корреляции, превышающее 0,7, свидетельствует в пользу валидности полученных измерений [47].

Пример 4.2. Последние 10–15 лет в практике управления персоналом достаточно широко используется такой инструмент оценки стилей решения проблем, как KAI (Kirton Adaption Innovation Inventory) [46]. Этот инструмент, разработанный английским исследователем М. Кертоном в 1970–1980-х гг., предварительно был протестирован на нескольких тысячах человек для оценки его надежности и обоснованности. В частности, данные, полученные с помощью этого инструмента, статистически значимо показывают, что люди разных профессий имеют разные стили решения проблем. Например, у бухгалтеров сравнительно более адаптивный стиль, а у маркетологов – более инновативный (в терминах концепции М. Кертона). Для разработки русскоязычной версии инструмента KAI автором данного учебного пособия проводилось тестирование адаптируемой версии в различных профессиональных группах. Соответствие полученных результатов тем данным, которые были получены ранее М. Кертоном и его коллегами, рассматривалось как одно из свидетельств валидности адаптируемого инструмента.

Концептуальная валидность (*construct validity*) обозначает соответствие результатов измерений тому концепту (свойству), для измерения или оценки которого проводилось исследование и разрабатывался соответствующий инструмент. В англоязычной специальной и учебной литературе подобный смысл передается термином *construct validity*. Транслитерация этого термина («конструктивная валидность»), которую употребляют некоторые авторы в русскоязычных изданиях, представляется не совсем подходящей. Поэтому автор этого учебного пособия, придерживаясь терминологии В.С. Аванесова [2], употребляет термин «концептуальная валидность».

Это наиболее важный и наиболее сложно оцениваемый аспект валидности. Другими словами, этот термин характеризует логическое соответствие измеряемых показателей изучаемому понятию (концепту): действительно ли с помощью выделяемых показателей мы можем характеризовать то свойство, которое изучаем. В литературе выделяются два вида концептуальной валидности:

а) **конвергентная валидность (*convergent validity*)** предполагает, что если с помощью различных методов измерений мы получаем близкие результаты, то эти измерения обоснованы (валидны);

б) *дивергентная* или *дифференцирующая валидность* (*discriminant validity*) связана с возможностью выделения и отделения различных показателей изучаемого свойства (концепта). Для оценки этого вида валидности может быть использован факторный анализ, но основное внимание здесь должно уделяться содержательному анализу изучаемого свойства или феномена.

Оценка и аргументация валидности носят преимущественно описательный характер, и в научных публикациях, в которых оценивается валидность измерений, соответствующий раздел может занимать несколько страниц.

4.3. Подходы к оценке надежности измерений

Надежность (*reliability*) – это характеристика, отражающая устойчивость и согласованность получаемых результатов измерения. В повседневном общении мы очень часто используем слово «надежность» (надежный человек, надежный компьютер, надежный автомобиль и т.д.). Основной смысл, который при этом вкладывается в эту характеристику, – стабильность, безотказность, повторяемость, предсказуемость, регулярность. Примерно такой же смысл вкладывается и в понятие «надежность» как характеристику измерения. Это, например, возможность получения для одной и той же выборки аналогичных результатов при условии, что измеряемая характеристика не изменилась.

Надежность является составляющей (необходимым, но недостаточным элементом) валидности измерений. Валидное измерение всегда должно быть надежным, но не всякое измерение, характеризующееся как надежное, является валидным (в смысле содержательной, критериальной и концептуальной валидности).

Связь этих двух характеристик может иллюстрировать простой пример с измерением веса. Если мы правильно измеряем наш вес (и проверяем это измерение разными весами), это измерение обоснованно и надежно. Если наши весы всегда показывают на 10 кг больше (что проверяется с помощью других весов), то эти измерения могут рассматриваться как надежные, но не обоснованные (не валидные). Если наши весы каждый день показывают заметно разные значения (что может зависеть, например, от температуры воздуха, атмосферного давления и

других факторов), то эти измерения не являются ни надежными, ни валидными.

Надежность характеризует, насколько измерения свободны от случайных ошибок. В отличие от оценки валидности оценка надежности измерений всегда осуществляется с помощью математико-статистических операций. Общий подход к оценке надежности заключается в оценке степени связанности результатов измерения с помощью либо *параллельных испытаний*, либо разнесения измерений во времени, либо соотнесения данных, полученных по разным фрагментам одного инструмента.

Для определения надежности используются три главных подхода, основанных на трех разных вариантах понимания надежности:

1) *надежность-устойчивость (stability)* характеризует стабильность результатов во времени;

2) *надежность-эквивалентность (equivalence)* характеризует идентичность результатов, полученных несколькими аналогичными инструментами;

3) *надежность-согласованность (internal consistency)* характеризует внутреннюю согласованность (гомогенность) результатов, полученных одним инструментом.

Рассмотрим эти подходы подробнее. Для оценки надежности в смысле устойчивости результатов во времени проводится повторное измерение тем же инструментом по той же выборке через определенный промежуток времени (*метод тест-ретест*). Результаты двух тестов, как правило, сравниваются путем определения коэффициента корреляции или другой меры связи (гл. 9), а также средних значений по двум испытаниям. В случае если получается высокий коэффициент корреляции (близкий к единице) и средние значения по первому и второму тестированиям близки, то это свидетельствует о надежности измерений в смысле их воспроизводимости и стабильности. Если по результатам первого и второго испытаний средние значения различаются достаточно сильно, но в целом те респонденты, которые имели высокие баллы при первом измерении, набрали также высокие баллы во втором, то в этом случае коэффициент корреляции принимает достаточно высокие значения, что указывает на определенную надежность инструмента. На практике статистически значимый коэффициент корреляции выше 0,8 считается свидетельством достаточной надежности измерений, хотя во многих случаях [47] в качестве приемлемого значения указывается 0,7.

При этом также следует указывать уровень статистической значимости полученного результата (методы и критерии проверки статистической значимости мер связи рассматриваются в гл. 9 учебного пособия). Плохая воспроизводимость результатов предыдущего тестирования приводит к низкой корреляции измерений, что свидетельствует об их низкой надежности.

Достоинство этого метода заключается в сравнительной простоте его использования, ясности основных посылок, лежащих в определении надежности, и простоте расчетов. Сложности возникают при определении временного интервала между двумя испытаниями. Если ретест проводится слишком рано, респонденты могут запомнить ответы, которые они давали при первом испытании. При слишком позднем проведении повторного испытания измеряемые характеристики могут измениться (например, знания, способности, опыт, отношения респондентов и т.д.). Приемлемым считается интервал между тестированиями от двух недель до двух месяцев. Кроме того, сама по себе высокая корреляция не может однозначно свидетельствовать о воспроизводимости результатов, поэтому результаты повторного тестирования рекомендуется контролировать другими методами. Например, можно сравнивать ранги респондентов, и если они в целом не изменились, то появляются дополнительные основания в пользу надежности измерений, но только в смысле их стабильности, так как возможен тренд, т.е. систематическое увеличение или уменьшение результатов от одного тестирования к другому. Возможно применение процедур проверки статистической гипотезы о равенстве средних значений и достоверности различий дисперсий по первому и повторному тестированиям (часть вторая учебного пособия).

Для оценки надежности-эквивалентности используется метод параллельного тестирования, или альтернативных тестов (*parallel forms*), проводимых либо одновременно, либо с небольшим интервалом. Данный метод оценки надежности применим только тогда, когда имеются параллельные (сходные, но не одинаковые) формы инструмента. Одной и той же группе респондентов предлагается вначале одна форма, затем сразу или после некоторого перерыва (до одной-двух недель) – другая. Коэффициент корреляции, полученный по результатам двух тестов, называется *коэффициентом эквивалентности измерения*. Если между

предъявлением обеих форм имеется значительный временной интервал (свыше двух недель), то полученный коэффициент корреляции называется *коэффициентом эквивалентности и стабильности измерений*.

Статистически значимый коэффициент корреляции $\geq 0,8$ считается свидетельством достаточной надежности измерений (минимально приемлемое значение — 0,7). Но вычисление коэффициента корреляции может оказаться недостаточным в случае больших различий в средних значениях и дисперсиях по параллельным тестам. Еще одна сложность, вызванная применением данного метода, заключается в невозможности обеспечить *полную эквивалентность* двух разных тестов. Рекомендуется в качестве альтернативного теста использовать тот же самый инструмент с переформулированными пунктами (вопросами или утверждениями). В педагогических измерениях в качестве альтернативных тестов используются сходные по сложности задания. Одним из вариантов метода параллельного тестирования является повторное измерение тем же самым инструментом (например, анкетой), но проводимое другим человеком.

Хотя и стабильность, и эквивалентность характеризуют надежность измерений как их воспроизводимость, стабильность измерений связана с их устойчивостью во времени и в разных ситуациях (проверяется методом повторного тестирования тем же инструментом в той же группе), а эквивалентность — с устойчивостью измерений к изменению формы представления инструмента и с сопоставимостью результатов измерения с «каноническими» результатами.

Надежность измерительных инструментов наиболее часто оценивается по *внутренней согласованности* полученных результатов. Такие измерительные инструменты, как анкеты и тесты, содержат большое число составляющих: вопросы, утверждения, задания и т.п. Каждый из пунктов направлен на косвенное выяснение какой-то стороны, фрагмента целого, вследствие чего он является частичным индикатором измеряемого фактора (свойства). Предполагается, что когда мы принимаем во внимание всю совокупность индикаторов и определенным образом интегрируем косвенную информацию, которую несет каждый из индикаторов, наши выводы становятся более надежными и обоснованными. Но при этом надежное измерение должно быть внутренне непротиворечиво.

Однако прежде чем интегрировать данные по индикаторам, необходимо соблюдение условия, что эти индикаторы отражают одно и то же, имеют нечто общее. Если это не так, то операция получения комплексной оценки просто не имеет смысла. Надежность-согласованность как раз и показывает, в какой степени результаты измерений внутренне согласованы (непротиворечивы).

Для оценки надежности-согласованности пунктов инструмента разработано несколько методов. Рассмотрим базовые предпосылки, лежащие в их основе. Убедиться в том, что пункты шкалы измеряют нечто общее, можно путем определения коэффициента корреляции между откликами на пункты инструмента. Достаточно высокое значение коэффициента корреляции ($r \geq 0,8$) между двумя переменными может свидетельствовать о том, что имеется какой-то скрытый (латентный) фактор, общая причина, которая стоит за каждой из них. Именно на этом соображении может строиться проверка такого качества шкалы, как согласованность: если измеряется один фактор, то результаты измерений должны коррелировать друг с другом. Но такой подход позволяет сравнивать измерения (отклики, ответы, баллы и т.д.) попарно. Как можно на этой основе получить универсальный показатель для результатов измерения в целом?

Одним из первых методов, разработанных для решения этой задачи, был *метод раздельного коррелирования (split-half)*. Он заключается в разбиении откликов по всем пунктам инструмента (например, ответов на все вопросы анкеты) пополам и расчете коэффициента корреляции по соответствующим двум наборам данных — суммарным результатам (например, баллам) по каждому пункту. Суммирование баллов в двух сформированных группах дает два набора данных, корреляция между которыми и характеризует надежность-согласованность инструмента. Если результаты измерений совершенно надежны, то следует ожидать, что обе части абсолютно коррелируют ($r = 1,0$). Впрочем, такая «абсолютная надежность» является гипотетической и на практике встречается исключительно редко. Если пункты инструмента не являются абсолютно согласованными, то коэффициент корреляции будет меньше единицы.

Преимущество метода раздельного коррелирования перед методом параллельного тестирования заключается в том, что он позволяет оценить надежность при однократном тестировании.

Однако использование этого метода предполагает допущение об эквивалентности не только отдельных форм, но и пунктов анкеты. Еще одна сложность использования метода раздельного коррелирования заключается в том, что два набора пунктов анкеты можно получить разными способами, причем количество возможных вариантов разделения «астрономически» возрастает с количеством заданий. Если у нас, например, 20 пунктов в анкете, можно первые 10 включить в одну группу, а вторые 10 — в другую; также можно в первую группу включать пункты с нечетными номерами, а во вторую — с четными (это наиболее распространенная на практике процедура) и т.д. Тест, состоящий, например, из 20 заданий, может быть поделен пополам для раздельного коррелирования $\frac{20!/2}{10!10!} = 92378$ разными способами.

Понятно, что коэффициент корреляции будет зависеть от способа разбиения.

Наиболее правильным считается разбиение, производимое случайным образом, что позволяет избежать искусственных эффектов. Тем не менее показатель надежности-согласованности, полученный таким методом, будет варьироваться всякий раз при формировании групп. Проблемами, связанными с этим подходом, являются также неэквивалентность пунктов (например, одни вопросы могут быть сложнее, чем другие, и наоборот), а также то обстоятельство, что респондент может вообще не ответить на какие-то вопросы или, дойдя до половины анкеты, ответить небрежно на оставшиеся вопросы.

Еще одной проблемой применения этого метода является то, что коэффициент корреляции рассчитывается не по всем пунктам анкеты, а по половине. Для корректировки полученного значения используется формула Спирмена—Брауна, предложенная независимо друг от друга К. Спирменом (С. Spearman) и У. Брауном (W. Brown) в 1910 г.:

$$r_{SB2} = \frac{2r}{1+r},$$

где r_{SB2} — скорректированный показатель надежности-согласованности по методу *split-half*;

r — коэффициент корреляции между двумя наборами пунктов инструмента.

Данная формула обобщается на случай инструмента, состоящего не из двух, а из k эквивалентных частей, по каждой из которых известен коэффициент корреляции r (обобщенная формула Спирмена–Брауна):

$$r_{SB} = \frac{kr}{1 + (k - 1)r}.$$

Эту формулу можно использовать для оценки согласованности измерений, предполагая, что k – количество всех пунктов инструмента (или пунктов в оцениваемом блоке анкеты), а r – усредненный коэффициент корреляции между этими пунктами [7]. В этом случае снимается проблема многообразия способов формирования групп.

Проиллюстрируем использование этой формулы на простом примере. Пусть мы имеем анкету из трех пунктов, на которые получены отклики четырех респондентов, кодируемые по шкале от 0 до 2 (табл. 4.1).

Таблица 4.1

Данные для оценки надежности-согласованности

Респондент	Пункт анкеты		
	1	2	3
1	0	1	1
2	1	2	2
3	2	1	2
4	0	1	1

Вычисляется корреляционная матрица (матрица коэффициентов корреляции). Это можно сделать, например, с помощью функции Excel для вычисления коэффициента парной корреляции КОРРЕЛ, а еще проще – с помощью специальной опции Excel «Анализ данных» («Сервис», «Анализ данных», «Корреляция»)¹. Входным интервалом при использовании этой опции будет выступать матрица откликов респондентов (способ группировки – по столбцам). В результате будет получена корреляционная матрица (табл. 4.2).

¹ Если в программе Excel такая опция отсутствует, необходимо зайти в меню Excel «Сервис», «Надстройки» и отметить опцию «Пакет анализа».

Корреляционная матрица для оценки надежности-согласованности

Корреляционная матрица	1	2	3
	1 0,25 0,918559	1 0,612372	1
Усредненный коэффициент корреляции	0,593644		

Для определения усредненного коэффициента корреляции необходимо посчитать среднее значение коэффициентов корреляции в одной половине матрицы (кроме диагональных, поскольку это корреляционные коэффициенты пунктов с собой).

Для нашего случая среднее арифметическое трех коэффициентов корреляции можно определить, например, с помощью функции СРЗНАЧ. (Для простоты использования этой функции можно предварительно удалить единицы, расположенные по диагонали корреляционной матрицы, и в качестве параметра функции указать массив коэффициентов 2×2 с одним пустым элементом.) Для рассматриваемого примера усредненный коэффициент корреляции равен 0,594. Далее по обобщенной формуле Спирмена–Брауна определяется индекс согласованности пунктов анкеты: 0,814 (читателям рекомендуется проверить это самостоятельно). Отметим, что в данном случае рассчитывается коэффициент корреляции Пирсона, предполагающий метрический характер шкалы. Если данные представлены в порядковой шкале, то более корректным будет использование коэффициента корреляции Спирмена (подробнее в гл. 9).

Такой подход предполагает равенство дисперсий в двух коррелируемых группах. Разработан аналог коэффициента раздельного коррелирования Спирмена–Брауна, который не предполагает такого равенства дисперсий. Он называется коэффициентом раздельного коррелирования Гутмана (*Guttman Split-half*) и также известен как коэффициент Рулона (*Rulon coefficient*) [53]:

$$r_R = 2 \frac{\sigma_I^2 - \sigma_1^2 - \sigma_2^2}{\sigma_I^2},$$

где σ_I^2 – общая дисперсия по всем данным (первой и второй группам);
 σ_1^2 – дисперсия по первой группе;
 σ_2^2 – дисперсия по второй группе.

Другие подходы к определению внутренней согласованности основаны на вычислении коэффициентов KR_{20} Кадера—Пичардсон, альфа Кронбаха и лямбда Гутмана. Рассмотрим базовые аксиомы, на основе которых разработаны формулы для вычисления этих коэффициентов.

Каждое измерение включает как истинное значение, так и частично неконтролируемую, случайную погрешность:

$$O = T + E,$$

где O — наблюдаемое (измеряемое) значение (*observed score*);
 T — истинное значение (*true score*);
 E — случайная ошибка (*random error*).

Более полно, с учетом систематической ошибки (B):

$$O = T + E + B.$$

Как будет связана изменчивость (вариация) слагаемых этого соотношения?

Изменчивость измеряемого признака может быть связана с «естественной» (истинной) изменчивостью самого признака (например, различие в подготовленности, квалификации и т.д.), но определенный вклад может внести то, как мы измеряем, т.е. изменчивость ошибки измерения. Запишем это как

$$\sigma_O^2 = \sigma_T^2 + \sigma_E^2$$

(систематическая ошибка не учитывается, так как считается, что ее изменчивость равна 0).

Тогда надежность измерения (ρ) может характеризоваться отношением изменчивости истинных значений к изменчивости наблюдаемых значений:

$$\rho = \frac{\sigma_T^2}{\sigma_O^2}.$$

Разумеется, мы не можем знать истинные значения и их изменчивость (иначе нам бы не пришлось проводить никакие измерения), но мы можем исключить их из рассмотрения, представив как $\sigma_T^2 = \sigma_O^2 - \sigma_E^2$. Тогда получаем

$$\rho = \frac{\sigma_O^2 - \sigma_E^2}{\sigma_O^2} = 1 - \frac{\sigma_E^2}{\sigma_O^2}.$$

В знаменателе мы имеем не что иное, как дисперсию измеряемых значений. Определить значение в числителе сложнее, но понятно, что оно должно иметь смысл дисперсии ошибок наших измерений. Однако заметим сразу, что использование подобного подхода не предполагает разбиения пунктов инструмента на группы, поэтому снимается проблема зависимости результата от способа разбиения. Это весьма абстрактная идея о надежности измерений воплотилась в нескольких конкретных вариантах расчетных моделей.

Впервые конкретная реализация подобных рассуждений была предложена Кадером и Ричардсон (1937 г.) и получила название формулы *Кадера—Ричардсон-20* (*Kuder—Richardson 20*), или KR_{20} . Необычное название формулы связано с тем, что авторы предложили несколько различных формул, обозначенных разными индексами; двадцатая оказалась наиболее удачной. Эта формула была предложена для вычисления индекса согласованности для дихотомической шкалы (т.е. для переменных, принимающих только два значения, например для ответов истинно/ложно):

$$KR_{20} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k p_i q_i}{\sigma_i^2} \right),$$

где k — количество вопросов;
 p_i — доля первого варианта ответа на i -й вопрос;
 $q_i = (1 - p_i)$ — доля второго варианта ответа на i -й вопрос;
 σ_i^2 — дисперсия сумм измеряемых значений (суммирование осуществляется по всем вопросам анкеты для каждого респондента).

Например, получены ответы пяти респондентов по трем пунктам анкеты с дихотомической шкалой. В табл. 4.3 приведены ответы, а также все промежуточные результаты для вычисления коэффициента KR_{20} . Полученное значение (0,86) свидетельствует о хорошей согласованности заданий данного теста.

Для порядковых шкал с большим количеством позиций (например, шкалы Лайкерта), а также для более мощных шкал (например, интервальных) Л. Кронбах предложил другую формулу для определения согласованности измерений (1951 г.). Показа-

Пример расчета коэффициента KR_{20}

Респондент	Пункт			Сумма	Показатель
	1	2	3		
1	0	1	1	2	—
2	1	1	1	3	—
3	0	0	0	0	—
4	1	1	1	3	—
5	1	1	0	2	—
σ_i (стандартное отклонение суммарных откликов)	—	—	—	—	1,22
σ_i^2 (дисперсия суммарных откликов)	—	—	—	—	1,5
p_i	0,6	0,8	0,6	—	—
q_i	0,4	0,2	0,4	—	—
$p_i q_i$	0,24	0,16	0,24	0,64	—
KR_{20}					0,86

тель согласованности, рассчитанный по этой формуле, получил название *коэффициент альфа Кронбаха (Cronbach's Coefficient Alpha)*. Большинство современных статистических пакетов (SPSS, SAS, STATISTICA и др.) включают процедуры вычисления коэффициента альфа Кронбаха. Несложно посчитать этот коэффициент и с помощью стандартных функций программы Excel. Формула выглядит следующим образом:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_f^2} \right),$$

где k — количество пунктов;

σ_i^2 — дисперсия откликов по каждому пункту анкеты;

σ_f^2 — дисперсия суммарной шкалы (дисперсия суммы откликов каждого респондента на пункты анкеты).

Формула Кронбаха является расширенной аналогией формулы Кадера—Ричардсон и отражает следующую идею. Если имеется несколько субъектов, отвечающих на вопросы анкеты, то можно вычислить дисперсию для каждого вопроса и суммарной шкалы. Дисперсия для суммарной шкалы будет меньше,

чем сумма дисперсий каждого отдельного вопроса в том случае, когда вопросы измеряют (оценивают) *одну и ту же* изменчивость между субъектами, т.е. если они измеряют некоторое истинное значение. Если не измеряется реальное значение, а только случайная погрешность в ответах на вопросы (следовательно, ответы полностью не коррелированы между субъектами), то дисперсия суммы будет такой же, как сумма дисперсий отдельных пунктов. Поэтому коэффициент альфа будет равен нулю. Если все вопросы измеряют один и тот же объект (истинную метку), то коэффициент альфа равен единице.

Пример 4.3. Возьмем те же данные, которые мы использовали для иллюстрации применения обобщенной формулы Спирмена–Брауна (см. табл. 4.1). Определим дисперсии по откликам на каждый вопрос и их сумму. Получим значение 1,3. Дисперсия суммарных оценок респондентов составит 2,7. Отсюда значение коэффициента альфа Кронбаха 0,778.

Если мы сравним полученное значение с индексом согласованности, определенным по обобщенной формуле Спирмена–Брауна, то увидим, что коэффициент альфа меньше: $0,778 < 0,814$. Это связано с тем, что обобщенная формула Спирмена–Брауна вычисляет коэффициент согласованности, как если бы измерения были стандартизованы, т.е. приведены к одной шкале с нулевым средним значением и единичной дисперсией (см. гл. 2). Часто (но не всегда) стандартизация исходных данных приводит к возрастанию надежности измерений.

Коэффициент альфа Кронбаха принимает значения в диапазоне от 0 до 1. Значения $\alpha \geq 0,8$ свидетельствуют о достаточной надежности-согласованности измерений. Однако следует принимать во внимание и объем выборки: чем меньше выборка, тем меньше может быть коэффициент альфа. Поэтому для небольших выборок (меньше 20 элементов) приемлемым может считаться значение $\alpha \geq 0,7$ [34]. Высокое значение коэффициента указывает на наличие общего основания у набора вопросов анкеты (заданий теста), но не говорит о том, что за ними стоит именно тот фактор, который предполагается измерять, поэтому предварительно необходимо обосновать валидность измерений.

Показатель надежности-согласованности, определяемый по формуле Кронбаха, будет зависеть также от количества и качества пунктов, входящих в анкету. При исключении любого пункта из

анкеты коэффициент альфа будет изменяться (уменьшаться или увеличиваться). При исключении пунктов, которые не противоречат другим пунктам анкеты (в том смысле, что все они направлены на измерение общего фактора), коэффициент альфа Кронбаха уменьшается. И напротив, при исключении пунктов, которые не согласуются с другими пунктами, значение коэффициента альфа будет увеличиваться. Теоретически при оценке надежности используемого (тестируемого) инструмента (и каждого пункта, входящего в этот инструмент) нужно определить коэффициент альфа Кронбаха при условии, что один из пунктов исключается (и так для всех пунктов). Это весьма трудоемкая задача, особенно если в измерительном инструменте много пунктов, поэтому для ее решения используют специальные статистические пакеты. Подробно технологию их применения мы рассмотрим в следующем подразделе, а пока лишь приведем пример, иллюстрирующий дальнейшее изложение последовательности улучшения надежности инструмента.

Пусть мы получили результаты анализа надежности-согласованности анкеты, состоящей из 10 пунктов (табл. 4.4)¹. Наибольший интерес представляют три крайних правых столбца таблицы. Они показывают корреляцию между откликами по соответствующему пункту и суммарным показателем по всем пунктам, квадрат множественного коэффициента корреляции между соответствующим пунктом и другими пунктами, внутреннюю согласованность пунктов (коэффициент альфа Кронбаха), если соответствующий пункт будет удален.

По этим показателям переменные ИТЕМ5 и ИТЕМ6 заметно выделяются: они не согласуются с остальными пунктами, поэтому эти два пункта можно исключить. Но исследователь может посчитать, что содержание этих пунктов важно для оценки изучаемого признака или что количество оставшихся пунктов может быть недостаточным для его полной оценки. Поэтому возможно, что пункты 5 и 6 потребуется переформулировать или заменить другими.

Пункты, при исключении которых коэффициент альфа увеличивается достаточно сильно, следует удалить из анкеты. К сожалению, однозначных критериев того, что значит «доста-

¹ Использован пример из мультимедийного учебника по пакету STATISTICA, доступного на сервере StatSoft Russia (www.statsoft.ru).

Результаты расчета коэффициентов альфа Кронбаха

Переменная	Итоги для шкалы: Среднее = 46.1100 Ст.откл. = 8.26444 N набл:100 Альфа Кронбаха: .794313 Стандартизованная альфа: .800491 Средняя межпозиционная корреляция: .297818					
	Среднее при удалении	Дисперсия при удалении	Ст. отклонение при удалении	Общ-поз. коррел.	Квадрат множественной регрессии	Альфа при удалении
ITEM1	41.61000	51.93790	7.206795	.656298	.507160	.752243
ITEM2	41.37000	53.79310	7.334378	.666111	.533015	.754692
ITEM3	41.41000	54.86190	7.406882	.549226	.363895	.766778
ITEM4	41.63000	56.57310	7.521509	.470852	.305573	.776015
ITEM5	41.52000	64.16961	8.010593	.054609	.057399	.824907
ITEM6	41.56000	62.68640	7.917474	.118561	.045653	.817907
ITEM7	41.46000	54.02840	7.350401	.587637	.443563	.762033
ITEM8	41.33000	53.32110	7.302130	.609204	.446298	.758992
ITEM9	41.44000	55.06640	7.420674	.502529	.328149	.772013
ITEM10	41.66000	53.78440	7.333785	.572875	.410561	.763314

точно сильное увеличение», не существует. Нельзя сказать, что если при удалении пункта коэффициент альфа увеличился на столько-то, то этот пункт должен быть исключен. Единственное общепринятое правило заключается в том, что если альфа (или другой показатель, характеризующий надежность-согласованность измерений) меньше 0,7, то измерение не может считаться надежным, т.е. если в анкете имеются пункты, при исключении которых коэффициент надежности увеличивается до 0,7 и выше, то такие пункты необходимо удалить и соответственно не учитывать результаты оценок или измерений по этим пунктам. Если, например, мы определили, что коэффициент альфа при исключении некоторого вопроса анкеты возрастает с 0,65 до 0,75, этот вопрос лучше исключить. Но если коэффициент альфа для исходного набора вопросов 0,8, а при исключении какого-то вопроса увеличивается до 0,95, нужно обратить особое внимание на этот вопрос (например, на то, как он сформулирован), но исключать его не обязательно, так как и с данным вопросом надежность-согласованность измерений приемлема.

На практике результаты, полученные по пункту анкеты, который плохо согласуется с другими, просто не учитывают при получении комплексных оценок и анализе результатов. Но если исследователь считает плохо согласованный пункт важным для проектируемого исследования или тестируемого измерительного инструмента, то этот пункт следует переформулировать. Однако после замены или переформулировки пункта необходимо вновь проводить тестирование инструмента и проверять надежность-согласованность измерений, поскольку после изменения любого из пунктов все значения коэффициентов надежности изменятся.

Коэффициент альфа Кронбаха можно рассматривать как оценку корреляции измерений данным инструментом с измерениями всеми другими инструментами, составленными из такого же числа пунктов, которые случайным образом извлекли из множества всех возможных индикаторов измеряемого свойства. Его можно также интерпретировать как корреляцию между измерениями данным инструментом и «истинными» измерениями, полученными, если бы респондент ответил на *все* возможные вопросы или пункты, направленные на измерение изучаемого свойства.

Коэффициент альфа может также применяться и для решения гораздо более широкого круга задач. Например, с его помощью можно измерять степень согласованности экспертов, оценивающих тот или иной объект, стабильность данных при многократных измерениях, качество различных шкал и т.д.

Еще один подход к оценке надежности-согласованности данных был предложен в 1945 г. Л. Гутманом, который предложил формулы для вычисления шести коэффициентов, наиболее важными из них являются первые три (L_1 , L_2 , L_3) [53]. Первый коэффициент определяет нижнюю границу надежности, второй коэффициент – «лучшую» из возможных оценок нижней границы надежности, а третий эквивалентен коэффициенту альфа Кронбаха. Доказано, что коэффициент L_2 всегда больше либо равен коэффициенту альфа Кронбаха. Мы не будем приводить формулы для расчета коэффициента Гутмана из-за их громоздкости, что делает весьма трудоемким расчет этих коэффициентов без специальных статистических пакетов. По-видимому, громоздкость формул и является основной причиной того, что этот подход получил значительно меньшее распространение на практике, чем формула Кронбаха, хотя подход Гутмана был описан в литературе на 6 лет раньше, чем подход Кронбаха. С помощью современных статистических программ коэффициенты Гутмана вычисляются так же просто, как и коэффициент альфа Кронбаха, но в силу «привычки» и того обстоятельства, что в литературе они описаны гораздо реже, коэффициенты Гутмана в исследовательской практике используются не так часто, как коэффициент альфа Кронбаха.

Отметим, что рассмотренные показатели (альфа Кронбаха, KR_{20} , L_1 , L_2 , L_3 и др.) не обязательно всегда неотрицательны. Возможны ситуации, когда один или каждый из этих коэффициентов будет иметь отрицательные значения, но в отличие, например, от коэффициента корреляции отрицательные значения коэффициентов надежности (внутренней согласованности) измерений не несут никакой дополнительной информации, кроме той, что из-за слабой согласованности измерения не могут считаться надежными.

При проверке валидности и надежности измерений может возникнуть сложность, связанная с тем, что инструменты, применяемые для измерения тех или иных интересующих исследователя признаков, часто включают несколько различных блоков

(например, пунктов анкеты), которые не только сформулированы по-разному, но и используют различные измерительные шкалы. Как в таком случае оценить валидность и надежность?

Рекомендуется измерительный инструмент делать гомогенным (однородным). Прежде всего необходимо, чтобы инструмент измерял некий единый концепт. Должны использоваться одинаковые шкалы для каждого пункта. Желательно, чтобы пункты анкеты были примерно одинаковы по форме (например, только вопросы или только утверждения), а также чтобы направление пунктов было одинаковым (например, только утвердительные или только отрицательно сформулированные утверждения), если только исследователь не намерен объединить в одном опросе два параллельных теста для оценки надежности-эквивалентности измерений. Если исследователь все же использует разнородный инструмент, то можно оценивать надежность измерений по блокам, определяемым смысловыми категориями (концептами), на изучение которых направлено исследование.

4.4. Использование статистических пакетов для оценки надежности измерений

Статистические пакеты SPSS и STATISTICA включают процедуры для оценки надежности несколькими разными методами. Данные должны вводиться таким образом, чтобы столбцы соответствовали пунктам анкеты, а строки — респондентам. В пакете SPSS для доступа к пакету оценки надежности следует в основном меню *Analyze* выбрать *Scale*, затем — опцию *Reliability Analysis*. В этом диалоговом окне выбираются необходимые пункты для анализа и переносятся в поле *Items*. В списке *Model* выбирается нужный метод оценки: *Alpha* (коэффициент альфа Кронбаха), *Split-Half* (метод раздельного коррелирования), *Parallel* (метод параллельного тестирования), *Strict Parallel* (метод параллельного тестирования при предположении равенства средних значений по каждому тесту), *Guttman* (оценка границ надежности методом Гутмана).

В окне *Statistics* можно установить режимы вывода базовых статистических показателей (среднее, стандартное отклонение и др.) для каждого пункта (*Items*) или для всего инструмента в целом (*Scale*). Полезно также выбрать режим, показывающий изменение базовых статистик и индикаторов, характеризующих

надежность в случае удаления одного из пунктов (*Scale if item deleted*).

Для иллюстрации работы и сравнения с результатами вычисления «вручную» используем данные из приведенного ранее примера (см. табл. 4.1). Из отчетов, сформированных пакетом SPSS после выполнения процедуры *Reliability Analysis* с использованием разных методов оценки, имеем следующие результаты (выходные формы по разным методам оценки объединены с сокращениями; пункты анкеты обозначены как ITEM1, ITEM2, ITEM3; знаком «*» отмечены комментарии автора данного учебного пособия, для сравнения результатов полужирным шрифтом отмечены значения, вычисленные ранее с помощью программы Excel).

* Базовая статистика для каждого пункта анкеты (среднее, стандартное отклонение, количество наблюдений):

		Mean	Std Dev	Cases
1.	ITEM1	,6000	,8944	5,0
2.	ITEM2	1,2000	,4472	5,0
3.	ITEM3	1,4000	,5477	5,0

* Корреляционная матрица пунктов анкеты:

Correlation Matrix			
	ITEM1	ITEM2	ITEM3
ITEM1	1,0000		
ITEM2	,2500	1,0000	
ITEM3	,9186	,6124	1,0000

* Базовая статистика по всему инструменту в целом (определяется на основе сумм откликов респондентов на пункты анкеты): среднее значение, дисперсия, стандартное отклонение, количество пунктов:

	Mean	Variance	Std Dev	N of Variables
Statistics for Scale	3,2000	2,7000	1,6432	3

* Статистика для средних значений пунктов анкеты:

Item Means					
Mean	Minimum	Maximum	Range	Max/Min	Variance
1,0667	,6000	1,4000	,8000	2,3333	,1733

Статистика для дисперсий по пунктам анкеты:

Item Variances

Mean	Minimum	Maximum	Range	Max/Min	Variance
,4333	,2000	,8000	,6000	4,0000	,1033

* Статистика для корреляций пунктов анкеты (усредненный коэффициент корреляции, минимальный и максимальный коэффициенты корреляции, диапазон изменения, отношение максимальный/минимальный коэффициенты, дисперсия):

Inter-item Correlations

Mean	Minimum	Maximum	Range	Max/Min	Variance
,5936	,2500	,9186	,6686	3,6742	,0896

* Статистика по инструменту при условии удаления одного из пунктов (среднее, дисперсия, коэффициент альфа):

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Alpha if Item Deleted
ITEM1	2,6000	,8000	,7500
ITEM2	2,0000	2,0000	,9000
ITEM3	1,8000	1,2000	,3333

* Коэффициент альфа Кронбаха:

Alpha = **,7778**

* Коэффициент альфа Кронбаха с учетом стандартизации измерений:

Standardized item alpha = **,8142**

* Коэффициенты Гутмана:

Lambda 1 = ,5185 Lambda 2 = ,8295 Lambda 3 = **,7778**

Пример 4.4. Предположим, что для оценки надежности-эквивалентности инструмента для оценки творческих стилей (стилей реше-

ния нестандартных проблем) [32; 46] использовались два инструмента, в которых вопросы имели приблизительно одинаковый смысл, но различались по формулировкам. Тестирование проводилось с интервалом в один день на одной и той же группе респондентов ($n = 20$). Результаты обрабатывались с помощью метода параллельного тестирования в программе SPSS.

Ниже представлен фрагмент отчета по выполнению процедуры оценки (некоторые показатели опущены).

		Mean	Std Dev	Cases
1.	TEST1	97,7000	14,4299	20,0
2.	TEST2	99,9500	14,9648	20,0

Correlation Matrix			
	TEST1	TEST2	
TEST1	1,0000		
TEST2	,9027	1,0000	

Item Variances					
Mean	Minimum	Maximum	Range	Max/Min	Variance
216,0829	208,2211	223,9447	15,7237	1,0755	123,6171

Parameter Estimates	
Estimated common variance =	216,0829
Error variance =	21,1513
True variance =	194,9316
Estimated common inter-item correlation =	,9021

Обратите внимание на то, что при выполнении этой процедуры помимо коэффициента парной корреляции вычисляются и приводятся в отчете дисперсия истинных значений σ_T^2 (*true variance*), дисперсия ошибок измерения σ_E^2 (*error variance*) и дисперсия наблюдаемых значений σ_O^2 или общая дисперсия (*common variance*). Эти величины, как было показано в подразд. 4.2, связаны друг с другом:

$$\sigma_O^2 = \sigma_T^2 + \sigma_E^2.$$

Как отмечалось ранее, надежность измерения может определяться как отношение

$$\rho = \frac{\sigma_T^2}{\sigma_O^2} = \frac{\sigma_O^2 - \sigma_E^2}{\sigma_O^2}$$

Как видно из отчета, это значение равно 0,9021, что очень близко к значению коэффициента корреляции (0,9027). Доказано, что эти два значения будут равны (или почти равны) при равенстве (или приблизительном равенстве) средних двух наборов данных. Соответственно чем больше различие средних значений, тем больше будет различаться коэффициент корреляции и оценка надежности. Это различие учитывается при использовании метода оценки надежности *Strict Parallel*. Для тех же самых данных мы получим следующие значения, выбрав метод *Strict Parallel* в окне *Statistics*.

		Mean	Std Dev	Cases
1.	TEST1	97,7000	14,4299	20,0
2.	TEST2	99,9500	14,9648	20,0

Correlation Matrix
TEST1 TEST2

TEST1	1,0000	
TEST2	,9027	1,0000

Item Variances

Mean	Minimum	Maximum	Range	Max/Min	Variance
216,0829	208,2211	223,9447	15,7237	1,0755	123,6171

Parameter Estimates

Estimated common mean	=	98,8250
Estimated common variance	=	217,3485
Error variance	=	22,6250
True variance	=	194,7235

Estimated common inter-item correlation = ,8910

Мы видим, что при использовании данной процедуры оценка надежности уменьшается (коэффициент корреляции, естественно, остается тем же). Если средние значения по результатам двух тестирований будут различаться в большей степени, то и различие между коэффициентом корреляции и оценкой надежности увеличится. Для того чтобы продемонстрировать это, увеличим среднее значение по первому тестированию на 20 (для этого достаточно к каждому элементу добавить 20). В этом случае получаем следующие результаты.

		Mean	Std Dev	Cases
1.	TEST1	117,7000	14,4299	20,0
2.	TEST2	99,9500	14,9648	20,0

Correlation Matrix
TEST1 TEST2

TEST1	1,0000			
TEST2	,9027	1,0000		
Statistics for Scale	Mean	Variance	Std Dev	Variables
	217,6500	822,0289	28,6710	2

Item Means

Mean	Minimum	Maximum	Range	Max/Min	Variance
108,8250	99,9500	117,7000	17,7500	1,1776	157,5313

Item Variances

Mean	Minimum	Maximum	Range	Max/Min	Variance
216,0829	208,2211	223,9447	15,7237	1,0755	123,6171

Parameter Estimates

Estimated common mean	=	108,8250
Estimated common variance	=	294,8485
Error variance	=	177,6250
True variance	=	117,2235
Estimated common inter-item correlation	=	,3940

В этом случае оценка надежности значительно снизилась (0,3940), тогда как коэффициент корреляции не изменился. Таким образом, вычисление лишь коэффициента корреляции может оказаться недостаточным при оценке надежности-эквивалентности методом параллельного тестирования. Методы *Parallel* и *Strict Parallel* также могут использоваться для проверки надежности-стабильности при повторном тестировании.

Пример 4.5. Для иллюстрации применения различных процедур оценки надежности-согласованности в пакете SPSS используем данные, полученные при оценке организационного климата в компании (30 респондентов, 8 пунктов анкеты). Данные представлены в 5-позиционной шкале Лайкерта (табл. 4.5) и соответствуют степени согласия респондентов с предложенными утверждениями относительно условий работы в компании.

Данные для оценки надежности измерений

№ п/п	Пункт анкеты								№ п/п	Пункт анкеты							
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
1	5	4	4	4	4	4	4	5	16	5	2	4	5	5	2	2	4
2	4	4	2	2	4	3	4	4	17	4	2	2	3	4	2	4	4
3	5	2	1	4	5	2	4	5	18	4	1	2	4	5	2	2	4
4	5	5	4	5	4	2	4	5	19	5	2	3	2	2	2	2	4
5	2	2	2	4	2	3	1	4	20	4	3	3	4	5	2	3	5
6	1	1	3	1	2	2	4	3	21	4	2	4	4	5	4	4	4
7	3	2	4	5	3	2	3	4	22	5	3	5	5	5	2	4	4
8	5	4	3	4	4	2	1	4	23	3	1	2	5	2	4	3	2
9	3	4	3	4	4	3	3	4	24	3	2	3	2	3	4	2	4
10	5	3	2	3	3	3	2	4	25	3	5	5	5	5	2	3	5
11	4	4	4	4	4	3	4	4	26	5	3	2	2	4	2	2	4
12	2	1	2	2	2	2	1	2	27	5	2	4	4	5	1	3	4
13	4	3	1	1	1	1	1	5	28	3	3	1	5	4	2	4	4
14	3	2	3	2	3	1	3	5	29	4	3	4	1	5	2	2	4
15	3	2	1	2	3	2	2	5	30	4	4	4	2	3	3	3	3

Будем рассматривать каждый пункт анкеты в качестве индикатора, характеризующего организационный климат компании, и оценим согласованность измерений по этим индикаторам. Ниже приведены результаты оценки согласованности методами *Reliability Analysis* пакета SPSS (пункты анкеты обозначены как ITEM1 – ITEM8; к отчету *Split-half* добавлены отчеты процедур *Alpha* и *Guttman*).

	Mean	Std Dev	Cases
1. ITEM1	3,8333	1,0854	30,0
2. ITEM2	2,7000	1,1492	30,0
3. ITEM3	2,9000	1,1847	30,0
4. ITEM4	3,3333	1,3730	30,0
5. ITEM5	3,6667	1,1842	30,0
6. ITEM6	2,3667	,8503	30,0
7. ITEM7	2,8000	1,0635	30,0
8. ITEM8	4,0667	,7849	30,0

Statistics for	N of			
	Mean	Variance	Std Dev	Variables
Part 1	12,7667	10,5989	3,2556	4
Part 2	12,9000	5,6793	2,3831	4
Scale	25,6667	25,8851	5,0877	8

Item Means					
Mean	Minimum	Maximum	Range	Max/Min	Variance
Part 1					
3,1917	2,7000	3,8333	1,1333	1,4198	,2529
Part 2					
3,2250	2,3667	4,0667	1,7000	1,7183	,6069
Scale					
3,2083	2,3667	4,0667	1,7000	1,7183	,3688
Item Variances					
Mean	Minimum	Maximum	Range	Max/Min	Variance
Part 1					
1,4468	1,1782	1,8851	,7069	1,6000	,0940
Part 2					
,9681	,6161	1,4023	,7862	2,2761	,1330
Scale					
1,2075	,6161	1,8851	1,2690	3,0597	,1628
Inter-item Correlations					
Mean	Minimum	Maximum	Range	Max/Min	Variance
Part 1					
,2825	,1967	,3825	,1858	1,9445	,0067
Part 2					
,1277	-,2962	,4107 ^p	,7069	-1,3865	,0633
Scale					
,2258	-,2962	,4918	,7881	-1,6604	,0335
Item-total Statistics					
	Scale	Scale	Corrected		
	Mean	Variance	Item-	Squared	Alpha
	if Item	if Item	Total	Multiple	if Item
	Deleted	Deleted	Correlation	Correlation	Deleted
ITEM1	21,8333	20,6264	,4139	,3382	,6869
ITEM2	22,9667	19,6195	,4857	,3909	,6708
ITEM3	22,7667	19,7023	,4544	,3140	,6778
ITEM4	22,3333	18,5747	,4584	,2785	,6783
ITEM5	22,0000	18,0690	,6371	,5194	,6335
ITEM6	23,3000	24,7000	,0547	,2038	,7446
ITEM7	22,8667	20,6713	,4222	,2630	,6853
ITEM8	21,6000	23,0069	,3004	,4025	,7081
Correlation between forms = ,6191					
Equal-length Spearman-Brown = ,7648					
Guttman Split-half = ,7423					

Alpha for part 1 = ,6053	Alpha for part 2 = ,4242
4 items in part 1	4 items in part 2
Alpha = ,7164	Standardized item alpha = ,6999
Lambda 1 = ,6268	Lambda 2 = ,7405
	Lambda 3 = ,7164

Как видно из приведенных отчетов, при использовании метода раздельного коррелирования исходный набор измерений разделен на две группы по четыре пункта в каждой. Статистические показатели как по анкете в целом, так и по ее отдельным пунктам сгруппированы по двум сформированным группам. Коэффициент корреляции между двумя группами (*correlation between forms*): 0,6191. Коэффициент надежности-согласованности (коэффициент межгрупповой корреляции, скорректированный по формуле Спирмена—Брауна (*equal-length Spearman—Brown*)): 0,7648. Коэффициент раздельного коррелирования Гутмана (*Guttman Split-half*) несколько ниже: 0,7423 (различия в дисперсиях по группам довольно велики). Коэффициент альфа Кронбаха (*Alpha*): 0,7164; коэффициент альфа Кронбаха с учетом стандартизации измерений (*Standardized item alpha*): 0,8142. Коэффициент альфа отдельно рассчитывается также для каждой из двух групп (*Alpha for part 1* и *Alpha for part 2*): 0,6053 и 0,4242 соответственно. Коэффициенты лямбда Гутмана: 0,6268; 0,7405; 0,7164.

Полученные показатели надежности хотя и имеют значения на уровне приемлемых, но все же недостаточно высоки, чтобы можно было говорить о хорошей надежности-согласованности измерений. Очевидно, что оценки по пункту 6 (ИТЕМ6) в наименьшей степени согласованы с оценками по остальным пунктам. При удалении этого пункта из анкеты коэффициент альфа увеличивается с 0,7164 до 0,7446. Именно на этот пункт необходимо обратить внимание в первую очередь. Его следует удалить, заменить или переформулировать (как обсуждалось в подразд. 4.3).

Далее кратко опишем возможности пакета STATISTICA для оценки надежности измерений. После запуска пакета следует выбрать модуль *Reliability/Item Analysis* (из основного меню *Analysis, Other Statistics*), затем ввести (набрать, скопировать или загрузить из файла) исходные данные (рис. 4.1). При необходимости можно добавить или удалить число строк и (или) столбцов для ввода данных (*Edit, Cases* и (или) *Variables, Add* или *Delete* соответственно). В окне *Variables* указать переменные для обра-

ботки (пункты анкеты): ITEM1–ITEM8. В окне *Split-half reliability* выбираются переменные двух групп для использования метода раздельного коррелирования (если предполагается применение этого метода). Обычно в одну группу отбираются переменные с нечетными номерами, в другую – с четными.

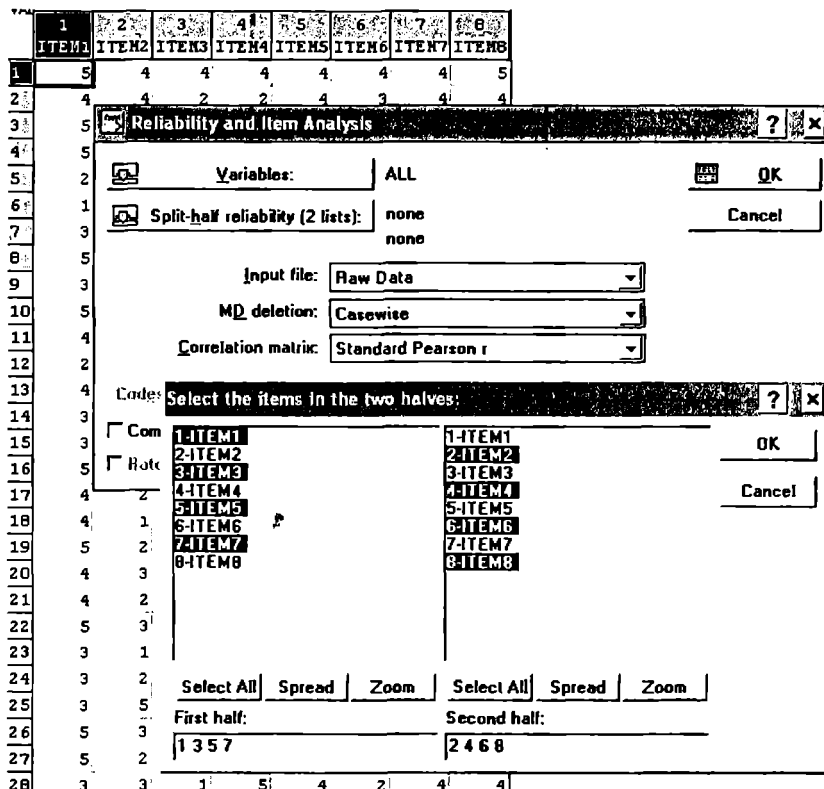


Рис. 4.1. Оценка надежности измерений в STATISTICA

Щелчком по кнопке *OK* открывается окно *Review Descriptive Statistics*, где, щелкнув по соответствующим кнопкам, можно вывести формы, содержащие базовые статистики для переменных (средние значения, стандартные отклонения, коэффициенты корреляции и др.) в табличной и графической формах. Если такие промежуточные результаты не требуются (они будут доступны и на следующем этапе), можно, щелкнув по кнопке *OK*, сразу перейти к итоговым результатам расчетов (рис. 4.2).

Number of items in scale: 8

Number of valid cases: 30

Number of cases with missing data: 0

Missing data were deleted: casewise

SUMMARY STATISTICS FOR SCALE

Mean: 25,666666667 Sum: 770,00000000
Standard Deviation: 5,087735987 Variance: 25,885057471
Skewness: -,298554656 Kurtosis: -,180071801
Minimum: 14,000000000 Maximum: 34,000000000
Cronbach's alpha: ,716366404 Standardized alpha: ,699941088
Average Inter-Item Correlation: ,232236047

Item-total statistics	?> Attenuation correction	
Split half reliability	?> What if more items	Cancel
Analysis of variance	?> How many more items	
Correlations	Graph	
Means & SD	Box & wh.	

Рис. 4.2. Показатели надежности измерений в отчете STATISTICA

Как видим, получены те же показатели оценки надежности, что и в программе SPSS. Так же, как в SPSS, можно проследить изменение надежности инструмента при удалении одного из пунктов (кнопка *Item-total statistics*) (рис. 4.3).

variable	Mean if deleted	Var. if deleted	Stdv. if deleted	Item-Total Correl.	Alpha if deleted
ITEM1	21,83333	19,93689	4,465298	,413871	,686932
ITEM2	22,96667	18,96556	4,354946	,485709	,670787
ITEM3	22,76667	19,04556	4,364121	,454442	,677771
ITEM4	22,33333	17,95555	4,237400	,458426	,678342
ITEM5	22,00000	17,46667	4,179314	,637086	,633503
ITEM6	23,30000	23,87667	4,886375	,054672	,744551
ITEM7	22,86667	19,98222	4,470148	,422184	,685313
ITEM8	21,60000	22,24000	4,715930	,300417	,708067

Рис. 4.3. Анализ надежности-согласованности в STATISTICA

Ценным дополнением к анализу надежности измерений в пакете STATISTICA является ряд интерактивных процедур «что-если», помогающих при разработке инструментов измерения. Например, при добавлении (удалении) некоторого количества вопросов пользователь может оценить ожидаемую надежность или определить количество вопросов, которые нужно внести в анкету или исключить из нее, чтобы добиться нужной надежности (опции *What if more items* и *How many more items* в окне *Reliability Results*). Если, например, планируется добавить в анкету два новых пункта (с согласованностью, соответствующей средней по анкете), то можно ожидать повышения коэффициента альфа до 0,76 (рис. 4.4).

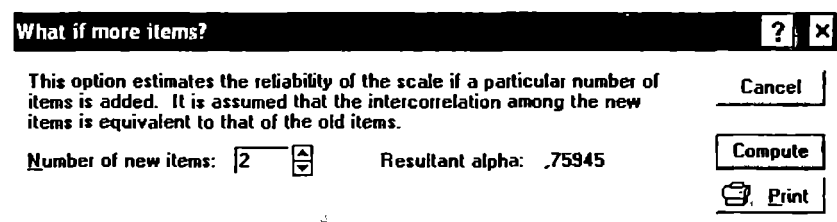


Рис. 4.4. Анализ изменения надежности измерений при изменении числа пунктов анкеты

Если, допустим, исследователь стремится обеспечить надежность измерений на уровне 0,85, легко определить, что необходимо добавить 10 дополнительных пунктов с согласованностью, соответствующей средней по анкете (рис. 4.5).

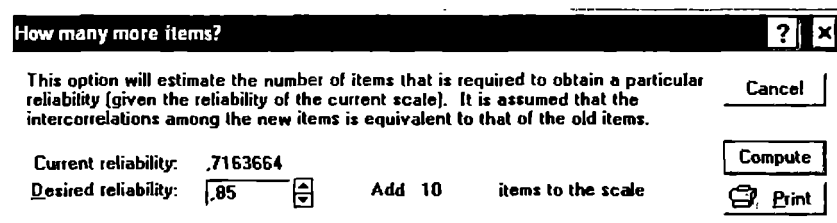


Рис. 4.5. Анализ «что-если» при оценке надежности-согласованности

Это предварительные оценки, поэтому после изменения инструмента требуются его тестирование и проверка согласованности измерений.

4.5. Оценка согласованности измерений на основе модели Раша

Рассмотренные выше процедуры оценки надежности-согласованности измерений были разработаны в рамках классической теории измерений. Серьезным ее недостатком является то, что во многих случаях при использовании процедур оценки не принимается во внимание вид измерительной шкалы. В частности, для данных в порядковых шкалах используются те же процедуры, что и для интервальных шкал.

Поэтому другой (не альтернативный, но дополняющий) подход к оценке качества измерений и измерительных инструментов может быть построен на основе модели Раша. С помощью этой модели можно ответить на вопросы: «Насколько пункты анкеты согласованы в плане измерения единого конструкта (концепта)?», «Измеряют ли они некий единый фактор или различные факторы?», «Насколько исходные данные подходят для измерения на основе используемой модели?»

Модель Раша показывает, насколько каждый пункт анкеты подходит (*fits*) для измерения той или иной характеристики предмета исследования [35]. Показатель надежности пункта (*item reliability index*) характеризует повторяемость результатов: если этот же пункт будет предложен другой группе респондентов, будут ли получены аналогичные результаты? Чем ближе к центральной оси диаграммы Раша расположены пункты анкеты, тем лучше они согласованы (рис. 4.6). Несогласованные пункты, расположенные за пунктирной линией, должны быть исключены или переформулированы.

Модель Раша также позволяет оценить надежность измерений для респондентов (*person reliability index*). Данный показатель характеризует повторяемость результатов: если этой же группе респондентов будут предложены вопросы, измеряющие тот же конструкт, будут ли получены аналогичные результаты?

На основе модели могут быть получены ошибки измерений по респондентам, а также степень соответствия откликов респондентов модели. Чем ближе к центральной оси диаграммы Раша расположены измерения, тем они точнее (см. рис. 4.6). Измерения, расположенные за пунктирной линией, не могут рассматриваться как надежные и должны быть исключены из анализа.

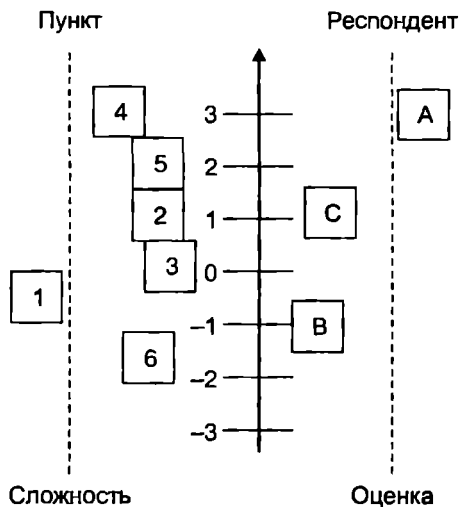


Рис. 4.6. Представление качества измерений на диаграмме Раша

Рассмотрим процедуру оценки результатов в WINSTEPS по данным, использованным в подразд. 4.4 (см. табл. 4.5). Исходный файл для обработки выглядит следующим образом.

ITEM1 = 1 ; Starting column of item responses

NI = 8 ; Number of items

SPSS Cases processed - 30

datum: 1 count: 20

datum: 2 count: 61

datum: 3 count: 46

datum: 4 count: 75

; datum: 5 count: 38

XWIDE = 1 this matches the biggest data value observed

CODES = "12345" ; matches the data

NAME1 = 10 ; Starting column for person label in data record

; Person Label variables: in label: in line

NAMLEN = 0 Length of person label

&END Item labels follow:

ITEM1 Item 1

ITEM2 Item 2

ITEM3 Item 3

ITEM4 Item 4

ITEM5 Item 5

ITEM6 Item 6
 ITEM7 Item 7
 ITEM8 Item 8

END NAMES

54444445 1

44224344 2

52145245 3

и т.д. (всего 30 респондентов).

Поскольку используется порядковая шкала с позициями 1 – 5, в поле CODES указаны соответствующие коды (12345). Результаты обработки данных (измерения по пунктам анкеты) представлены в табл. 4.6.

Таблица 4.6

Отчет WINSTEPS по пунктам анкеты

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEA CORR.	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD		
6	71	30	1.04	.21	1.13	.6	1.23	.9	.23	ITEM6
2	81	30	.62	.20	.88	-.5	.89	-.4	.65	ITEM2
7	84	30	.50	.20	.85	-.6	.96	-.1	.59	ITEM7
3	87	30	.39	.20	.93	-.2	.92	-.2	.63	ITEM3
4	100	30	-.12	.20	1.30	1.2	1.20	.9	.65	ITEM4
5	110	30	-.54	.21	.79	-.8	.78	-.8	.75	ITEM5
1	115	30	-.77	.22	1.03	.2	1.07	.4	.58	ITEM1
8	122	30	-1.13	.24	.86	-.5	.74	-.9	.46	ITEM8
MEAN	96.3	30.0	.00	.21	.97	-.1	.98	.0		
S.D.	17.0	.0	.71	.01	.16	.7	.17	.7		

В столбцах INFIT и OUTFIT указаны параметры, характеризующие соответствие данных модели Раша. В поле MNSQ (*mean-square statistics*) приводится статистика соответствия исходных данных измерительной модели, полученная на основе средних сумм квадратов отклонений теоретических значений от эмпирических. Значения MNSQ характеризуют степень «случайности» результатов или несоответствие данных используемой модели измерения. «Ожидаемые» значения MNSQ находятся вблизи 1 [55]. Высокие значения MNSQ OUTFIT могут быть связаны со «случайными» откликами респондентов. Высокие значения MNSQ INFIT обычно интерпретируются как свидетельство низкой валидности измерений, в том числе по причине низкой пригодности используемого инструмента для решения тех задач, для

решения которых он разрабатывался. Например, если в результате тестирования обнаруживаются высокие значения MNSQ INFIT, то это свидетельствует о том, что данный тест непригоден для группы испытуемых, в которой он предъявлялся.

Измерения с индексом MNSQ больше двух рассматриваются как несоответствующие измерительной модели, поэтому они не могут быть использованы при анализе результатов. Такие измерения рекомендуется исключать. Наиболее качественными и значимыми (*productive*) считаются измерения, для которых значения индекса MNSQ лежат в диапазоне от 0,5 до 1,5. Более высокие значения ($> 1,5$) указывают на неопределенность и «шум» в исходных данных. Слишком низкие значения ($< 0,5$) также не очень желательны, поскольку они свидетельствуют об информационной избыточности полученных измерений, в том числе по причине «информационной перегруженности» инструмента.

Более критичными для измерений считаются высокие значения MNSQ, поэтому анализ результатов следует начинать с пунктов, характеризуемых высоким значением индекса MNSQ. Более важными с точки зрения характеристики качества результатов считаются значения INFIT. Рассчитываемая статистика соответствия зависит от объема данных. Если количество наблюдений (респондентов) меньше 30, модель может оказаться не очень чувствительной к «неподходящим» пунктам («подходит все»). В случае, если количество наблюдений больше 300, модель, напротив, может оказаться слишком чувствительной («ничто не подходит»). В поле ZSTD приводятся стандартизованные значения MNSQ (со средним 0 и стандартным отклонением 1), по которым также можно судить о приемлемости полученных измерений.

Для данного примера статистика соответствия позволяет использовать измерения, полученные по всем пунктам анкеты, для дальнейшего анализа.

В поле PTMEA CORR. (или PTBIS CORR.) приведено значение *коэффициента бисериальной корреляции*, который показывает, насколько согласуются отклики всех респондентов по данному пункту с их откликами по всем пунктам. Данный коэффициент может принимать значения от -1 до 1 . Этот коэффициент рассматривается как показатель надежности-согласованности и может быть использован для определения и возможного исключения слабосогласованных измерений, а также (на этапе про-

ектирования измерительного инструмента) для выявления тех пунктов анкеты, которые следует переформулировать или исключить.

Аналогично интерпретируются результаты измерений по респондентам (табл. 4.7).

Таблица 4.7

Отчет WINSTEPS по респондентам (сокращен)

PERSON STATISTICS: MEASURE ORDER

ENTRY NUM- BER	RAW SCORE	COUNT	MEAS- URE	MODEL S.E.	INFIT		OUTFIT		PTMEA CORR.	PER- SON
					MNSQ	ZSTD	MNSQ	ZSTD		
1	34	8	1.70	.49	.20	-2.0	.27	-1.7	.77	1
4	34	8	1.70	.49	1.09	.4	.87	.0	.63	4
22	33	8	1.48	.46	1.07	.3	1.03	.2	.63	22
19	22	8	-.26	.39	.77	-.4	.67	-.6	.67	19
23	22	8	-.26	.39	2.42	2.4	2.61	2.5	-.12	23
5	20	8	-.57	.40	1.16	.5	1.22	.6	.34	5
15	20	8	-.57	.40	.60	-.8	.58	-.9	.80	15
6	17	8	-1.07	.43	2.10	1.9	2.43	2.2	-.09	6
13	17	8	-1.07	.43	2.26	2.0	2.08	1.8	.64	13
12	14	8	-1.71	.50	.34	-1.4	.50	-.9	.46	12
MEAN	25.7	8.0	.29	.41	.98	-.1	.98	-.1		
S.D.	5.0	.0	.80	.03	.56	1.2	.59	1.1		

Более высокие значения в столбце MEASURE соответствуют более «положительному» восприятию организационного климата в компании. По результатам статистики соответствия (INFIT и OUTFIT) отклики респондентов 23, 6, 13 должны быть исключены из анализа.

В заключение этого подраздела кратко остановимся на сопоставимости показателей качества измерений, получаемых на основе классической теории измерений и модели Раша. Характеристики качества измерений, определяемые «классическими» методами (в том числе альфа Кронбаха), и индикаторы качества измерений в модели Раша имеют разные смыслы, разную внутреннюю логику и разные вычислительные процедуры. Например, расчет коэффициент альфа Кронбаха основан на идее о согласованности результатов измерения, а модель Раша предлагает инструментарий для оценки соответствия данных модели (*fit statistics*). Поэтому прямо сопоставлять эти подходы нельзя, как нельзя напрямую сопоставить результаты их применения. В частности, нельзя сравнить согласованность измерений (например, вычислением коэффициента Кронбаха) по исходным данным и после их преобразования в шкалу Раша.

Дело в том, что применение модели Раша дает интегрированные результаты и по пунктам анкеты, и по респондентам;

в итоге осуществляется переход к вероятностным оценкам (например, может быть оценена вероятность определенного ответа на определенный вопрос определенным респондентом). Коэффициент альфа Кронбаха рассчитывается по фиксированным баллам каждого респондента по каждому заданию. Поэтому при оценке надежности измерений следует для полноты анализа не заменять, а дополнять одни подходы другими¹.

4.6. Оценка качества измерительных инструментов

Основные характеристики качества измерений (валидность, надежность и точность) не могут «напрямую» относиться к измерительным инструментам. Методологически некорректно говорить, например, о «валидном» или «надежном» тесте, поскольку эти характеристики определяются по результатам измерений, а их качество в свою очередь может зависеть от многих «неинструментальных» факторов.

Тем не менее качество измерительного инструмента всегда оценивается через анализ результатов, полученных с его помощью. Многократно проводя измерения, особенно при проектировании нового инструмента, мы оцениваем их качество. При этом, изменяя инструмент, например исключая или добавляя пункты анкеты, мы получаем измерения лучшего или худшего качества. Если мы стабильно получаем качественные измерения, то и используемый инструмент вызывает у нас больше доверия с точки зрения его качества. Например, если мы применим один и тот же инструмент для одной и той же выборки и получаем аналогичные результаты при условии, что измеряемая характеристика не изменилась, это дает нам возможность предполагать, что мы используем качественный инструмент. Иными словами, принимая во внимание условия проведения измерений и оценивая качество результатов измерений по показателям их валидности и надежности, мы можем судить о том, способен ли инструмент обеспечивать валидные и надежные измерения, но при этом не можем говорить с том, что этот инструмент «надежен и валиден».

¹ Такая несопоставимость в моделях в определенном смысле отражает нестыковку, несопоставимость научных парадигм, проанализированную Т. Куном [15].

Однако мы можем говорить об «*эффективности*» инструмента как комплексной характеристике его качества [2]. Эффективным мы можем называть измерительный инструмент, обеспечивающий качественные измерения с точки зрения их валидности, надежности и точности.

Кроме того, в понятие «*эффективность*» измерительного инструмента входит такая характеристика, как *практичность* (*practicality*), т.е. экономичность применения (низкая затратность), удобство и простота использования. Например, мы можем говорить о том, что один инструмент эффективнее другого, если он обеспечивает более качественные измерения при тех же затратах времени или, например, более удобен в использовании, обеспечивая измерения примерно того же качества, что и другой инструмент.

Понятие «*эффективность*» может относиться не только ко всему инструменту, но и к его составляющим, например пунктам анкеты. Выше мы уже обсуждали, как качество вопросов анкеты влияет на надежность-согласованность измерений. Исключение из инструмента «*плохих*» пунктов делает его более эффективным. Еще одним показателем качества пункта инструмента является коэффициент корреляции между откликами респондентов на этот пункт и общей суммарной шкалой (суммарным показателем по всем пунктам). Считается, что этот показатель не должен быть меньше 0,2–0,3. Пункты с меньшим коэффициентом «загромождают» анкету и делают ее менее эффективной. С другой стороны, если данный показатель очень высок (например, 0,95), то изменчивость откликов на один пункт на 90% (коэффициент детерминации $r^2 = 0,9$) повторяет или даже определяет изменчивость откликов по всей анкете. Это означает *избыточность* инструмента (либо данный пункт «*лишний*», либо все остальные пункты не несут дополнительной информации). Очевидно, что избыточность инструмента снижает его эффективность.

Эффективность инструмента зависит также от его *дифференцирующей способности*, связанной с отражением изменчивости измеряемых характеристик. Проверка дифференцирующей способности проводится для выделения и исключения пунктов, не обеспечивающих достаточную степень «уверенного» разделения ответов. Для иллюстрации приведем простой пример. Если на один вопрос теста *все* испытуемые отвечают правильно, а на

другой вопрос *все* отвечают неправильно, то такие вопросы не несут информации о фактической подготовленности испытуемых, поэтому такие задания не нужны в разрабатываемом тесте. Еще один пример. Предположим, проводится экзамен в студенческой группе и предлагается такой тест, по которому *все* студенты выполняют *все* задания и получают «отлично», затем предлагается другой тест, и в результате *никто* не выполняет *ни одного* задания, и все получают «неудовлетворительно». Способны ли такие «тесты» дать представление о знаниях студентов? Их нельзя считать эффективными. Аналогично, если на один вопрос, включенный в анкету, *все* респонденты отвечают «да», а на другой вопрос *все* отвечают «нет», то такие вопросы информации фактически не несут, поэтому они не вносят вклад в изучение того концепта, который интересует исследователя. Подобные вопросы должны исключаться из анкеты.

На практике дифференцирующая способность измерительного инструмента проверяется и оценивается по способности каждого пункта инструмента (вопроса или утверждения анкеты или задания теста) отражать изменчивость измеряемых характеристик. Для оценки дифференцирующей способности пунктов инструмента используются более или менее сложные математико-статистические процедуры, как правило, связанные с методами проверки статистических гипотез. Рассмотрим одну из таких процедур.

После анкетирования отклики всех респондентов по каждому из вопросов суммируются. Например, 1-й респондент по откликам на все пункты анкеты имеет 35 баллов, 2-й — 43, 3-й — 12 и т.д. Затем суммарные отклики ранжируются по величине. В итоге мы можем отобрать 20–25% респондентов с наименьшим суммарным откликом и столько же с наибольшим суммарным откликом. В контексте измерения отношений (восприятий, оценок и т.п.) первая группа соответствует респондентам с наименее благоприятным отношением к предмету исследования, вторая группа — с наиболее благоприятным отношением к предмету исследования. В контексте тестирования первая группа соответствует респондентам с наихудшими результатами, вторая группа соответствует респондентам с наилучшими результатами.

Таким образом, мы формируем две группы по n человек: группу с низким суммарным откликом (группа L) и группу с высоким суммарным откликом (группа H). Оставшиеся респон-

денты (50%) со «средним» откликом не рассматриваются. Далее для каждого пункта анкеты определяются следующие величины:

f – число респондентов, выбравших соответствующую позицию шкалы (например, в 5-позиционной шкале Лайкерта это 1, 2, 3, 4 или 5, что соответствует различной степени согласия с предлагаемым утверждением);

$\bar{X} = \frac{\sum fX}{n}$ – среднее взвешенное значение по откликам респондентов, где X – кодировка позиции шкалы (например, 1, 2, 3, 4 или 5), а $n = \sum f$ – число респондентов в группах L и H (в каждой группе это число должно быть одним и тем же).

Для каждого пункта анкеты определяется модифицированный t -критерий [39]¹:

Для каждого пункта анкеты определяется модифицированный t -критерий [39]¹:

$$t = \frac{\bar{X}_H - \bar{X}_L}{\sqrt{\frac{\left(\sum fX_L^2 - \frac{(\sum fX_L)^2}{n} \right) + \left(\sum fX_H^2 - \frac{(\sum fX_H)^2}{n} \right)}{n(n-1)}}}$$

В этой формуле индексы L (*low*) и H (*high*) относятся к первой и второй группам соответственно.

После определения t -критерия пункты ранжируются по величине t -критерия. Большее значение t -критерия соответствует лучшей дифференцирующей (разделяющей) способности пункта. В качестве критерия пригодности вопросов шкалы по степени различия принимается $t_{\text{крит}} = 1,73$ для $n \geq 10$. Пункты с $t < 1,73$ должны быть исключены из анкеты или переформулированы, в последнем случае инструмент вновь должен тестироваться на дифференцирующую способность. При $n < 10$ критическое значение t можно взять из таблицы t -распределения (табл. П.2) для числа степеней свободы $df = 2n - 2$ и выбранного уровня значимости p .

Если выборка состоит из нечетного числа респондентов (например, 71), то при формировании L - и H -групп не обязательно

¹ Этот подход к оценке дифференцирующей способности связан с методами проверки статистических гипотез на основе t -критерия, которые подробно разбираются во второй части учебного пособия.

добавлять или удалять респондентов, чтобы получить четное число, также не нужно включать в группы одних и тех же респондентов. Соотношение 25–25–50% – условное и может варьироваться. После ранжирования суммарных откликов всех респондентов отбирается равное количество респондентов «сверху» и «снизу» (приблизительно по 25% от численности всей выборки), какое именно количество остается в средней группе (четное или нечетное) – не принципиально. Отбираются *относительно* «высокие» и «низкие» суммарные отклики, т.е. определенное количество откликов из нижней и верхней частей списка, безотносительно к их абсолютным значениям.

Продемонстрируем использование этого подхода на примере оценки дифференцирующей способности одного из пунктов анкеты: «Я считаю свою работу интересной». В анкете используется 5-позиционная шкала Лайкерта, отражающая степень согласия с предлагаемыми утверждениями. Пусть анкета распространяется среди 292 человек, входящих в группу, сформированную для предварительного тестирования разрабатываемой анкеты (перед распространением анкеты в группе необходимо обеспечить по крайней мере ее внешнюю и содержательную валидность). После распространения анкеты и выделения в *L*- и *H*-группы по 25% респондентов (по 73 чел.) имеем следующие результаты (табл. 4.8).

Таблица 4.8

Оценка дифференцирующей способности пункта анкеты

Категория шкалы	Группа с наименьшим суммарным откликом (<i>L</i>)				Группа с наибольшим суммарным откликом (<i>H</i>)			
	<i>X</i>	f_L	fX_L	fX_L^2	<i>X</i>	f_H	fX_H	fX_H^2
Полностью согласен	5	3	15	75	5	22	110	550
Согласен	4	4	16	64	4	30	120	480
Нейтральное отношение	3	29	87	261	3	15	45	135
Не согласен	2	22	44	88	2	4	8	16
Полностью не согласен	1	15	15	15	1	2	2	2
–	–	73	177	503	–	73	285	1183
		<i>n</i>	ΣfX_L	ΣfX_L^2	–	<i>n</i>	ΣfX_H	ΣfX_H^2

По этим данным определяем t -критерий:

$$t = \frac{3,90 - 2,42}{\sqrt{\frac{\left(503 - \frac{177^2}{73}\right) + \left(1183 - \frac{285^2}{73}\right)}{73(73 - 1)}}} = \frac{3,90 - 2,42}{\sqrt{\frac{73,84 + 70,33}{73(73 - 1)}}} = 8,92.$$

На основе полученного результата можно сделать вывод о том, что анализируемый пункт анкеты обладает достаточно высокой дифференцирующей способностью. Аналогично можно провести анализ всех пунктов анкеты.

В заключение этого подраздела еще раз отметим, что если «плохой» (неэффективный) пункт измерительного инструмента представляется исследователю особо важным, то его необходимо переформулировать. Но после замены или переформулировки хотя бы одного из пунктов следует вновь проводить тестирование и оценивать качество измерений. На практике исследователь часто несколько раз проходит через этапы создания, удаления и переформулировки пунктов анкеты до тех пор, пока не придет к окончательному набору пунктов, обеспечивающих эффективный измерительный инструмент.

Задания для самостоятельного выполнения

1. Для проверки надежности-эквивалентности использовались два параллельных теста с интервалом в один день. Коэффициент корреляции по результатам тестирований составил 0,77. Поясните, как изменился бы результат, если бы временной интервал был увеличен до одной недели (коэффициент корреляции увеличился бы, уменьшился или остался прежним)?

2. Коэффициент альфа Кронбаха для некоторого набора измерений равен 0,6. Объясните, как изменится коэффициент при удвоении количества вопросов за счет добавления параллельного теста и почему?

3. Поясните, верны ли следующие утверждения.

- Валидность – более критичная характеристика для качества измерений, чем надежность.

- Оценить содержательную валидность сложнее, чем другие виды валидности.

• Стабильность и эквивалентность характеризуют одни и те же аспекты надежности измерений.

4. Представьте, что вам необходимо разработать анкету для опроса студентов, на основе которого можно было бы оценить качество преподаваемых курсов (см. вопрос 6 задания к гл. 1). Каким образом вы бы проверили (или обосновали) следующие характеристики измерений:

- внешняя валидность;
- содержательная валидность;
- критериальная валидность;
- концептуальная валидность;
- стабильность;
- эквивалентность;
- согласованность.

Дайте обоснованный и детальный ответ.

5. Представители одного из крупных автозаводов разработали инструмент для оценки уровня безопасности нового автомобиля. Анкета содержала 12 вопросов относительно нескольких функциональных систем автомобиля, которые обеспечивают безопасность пассажиров, например подушек безопасности, системы компьютерного тестирования и др. Проверка методами тест-ретест и параллельного тестирования показала высокую стабильность и эквивалентность измерений, однако коэффициент альфа Кронбаха оказался ниже 0,35. Какой вывод можно сделать?

6. Дисперсия «истинных» значений проведенных измерений равняется 190, а дисперсия наблюдаемых измерений равна 225. Можно ли только по этим данным судить о надежности измерений? Обоснуйте ответ.

7. Если стандартная ошибка измерений определена как 4,5, а стандартное отклонение истинных значений равно 10, можно ли, основываясь только на этих данных, судить о надежности измерений? Обоснуйте ответ.

8. По данным табл. П.9 оцените дифференцирующую способность каждого пункта анкеты для оценки организационного климата. Выполните задание: а) для данных по Компании 1; б) для данных по Компании 2; в) для объединенного массива данных. Сравните результаты и сделайте выводы.

9. По данным табл. П.9 оцените надежность-согласованность измерений с помощью обобщенной формулы Спирмена—Брау-

на. Выполните задание: а) для данных по Компании 1; б) для данных по Компании 2; в) для объединенного массива данных. Сравните результаты и сделайте выводы.

10. По данным табл. П.9 оцените надежность-согласованность измерений на основе коэффициента альфа Кронбаха. Выполните задание: а) для данных по Компании 1; б) для данных по Компании 2; в) для объединенного массива данных. Сравните результаты и сделайте выводы.

11. По данным табл. П.9 с помощью программы SPSS оцените надежность-согласованность измерений: а) методом раздельного коррелирования; б) на основе коэффициента альфа Кронбаха; в) с помощью коэффициентов лямбда Гутмана. Используйте данные для Компании 1. Сравните результаты и сделайте выводы.

12. Повторите условие задания 11 по Компании 2 и для объединенного массива данных. Сравните результаты и сделайте выводы.

13. По данным табл. П.9 оцените надежность-согласованность измерений на основе коэффициента альфа Кронбаха в программе SPSS. Используйте данные для Компании 1. По результатам, представленным в соответствующем отчете SPSS, проследите, как влияет исключение того или иного пункта анкеты на показатели согласованности измерений. Объясните причину этих изменений. Сколько пунктов нужно добавить в анкету, чтобы повысить коэффициент альфа Кронбаха до 0,8?

14. Повторите условие задания 13 по Компании 2 и для объединенного массива данных. Сравните результаты и сделайте выводы.

Часть вторая

Анализ результатов исследования и статистические выводы

Статистика – это прежде всего способ мышления, и для ее применения нужно лишь иметь немного здравого смысла и знать основы математики.

К.Р. МакКоннелл

Глава 5

Основные принципы проверки статистических гипотез

5.1. Статистическая достоверность связи и статистическая гипотеза

Тема проверки статистических гипотез является очень обширной, и вместе с тем она довольно хорошо разобрана в учебной и специальной литературе, доступной русскоязычному читателю. Существует свыше 100 различных методов и критериев для проверки статистических гипотез. Мы не ставим цель рассмотреть в этом издании все существующие методы и критерии. Следуя главному принципу данного учебного пособия, основное внимание уделяется логике и методологии проверки гипотез.

Любое исследование направлено на определение некоторой характеристики изучаемой генеральной совокупности или выявление связи между признаками (факторами). Такая связь часто исследуется в причинно-следственном аспекте, когда некоторые факторы рассматриваются как причины (*независимые переменные*), а другие – как следствия или результаты (*зависимые переменные*).

мые переменные). Связь может характеризоваться не только величиной (степенью связи) и направлением (что показывает, например, коэффициент корреляции), но и *статистической достоверностью* (*statistical confidence*). Последняя характеристика связи показывает, можно ли распространить результаты, полученные по данной выборке, на всю генеральную совокупность, из которой взята эта выборка. /

Любое заключение, полученное из статистического наблюдения, исследования и анализа, индуктивно и строится на конечном числе наблюдений, поэтому оно в принципе не полно и может быть недостоверно. Необходимо обоснование получаемого заключения, т.е. проверка результатов, на которых строится гипотеза, на статистическую достоверность. Достоверность непосредственно связана с репрезентативностью выборки, т.е. с тем, насколько уверенно данные, полученные по выборке, позволяют судить о соответствующих параметрах генеральной совокупности. Целью исследования почти никогда не является изучение данной конкретной выборки; выборка представляет интерес лишь постольку, поскольку она дает информацию обо всей генеральной совокупности. Статистическая достоверность связи определяется тем, насколько вероятно, что обнаруженная в выборке связь подтвердится (будет вновь обнаружена) на другой выборке той же генеральной совокупности. Оценивая статистическую достоверность связи, мы должны ответить на вопрос: какова вероятность случайного получения результата, подтверждающего наличие связи, которой нет в генеральной совокупности?

Фактически, оценивая статистическую достоверность связи, мы проверяем ту или иную статистическую гипотезу. **Статистическая гипотеза** — утверждение относительно неизвестного параметра (параметров) генеральной совокупности на основе выборочного исследования. Часто гипотеза формулируется как утверждение наличия или отсутствия связи между признаками (зависимыми и независимыми переменными).

Предположим, мы сформулировали некую гипотезу. Понятно, что измеряемые значения по выборке всегда отличаются от измеряемых значений по генеральной совокупности, поэтому нужно определить, насколько велико (значимо) это различие. Разница в измерениях может считаться значимой, если есть основания полагать, что она не представляет случайную флуктуа-

цию. Если мы имеем такие основания, то мы можем отклонить или признать невозможность отклонения сформулированной гипотезы.

Пример 5.1. Например, мы проверяем монетку по правилу «орел-решка»: фальшивая она или настоящая. Мы предполагаем, что нефальшивая монетка должна упасть «орлом» примерно столько же раз, сколько и «решкой». Предположим, монетка была подброшена 100 раз, из которых 57 раз выпал «орел», а 43 — «решка». Что можно сказать о ее подлинности? Достаточно ли такого количества подбрасываний, чтобы судить о подлинности монетки? С какой вероятностью мы можем судить о том, что монетка настоящая (или фальшивая)? Какова вероятность ошибки? Вот те вопросы, на которые мы должны уметь дать ответ при проверке гипотезы в данном случае о подлинности монетки и обосновании статистической достоверности ответа (исследовательского результата).

Пример 5.2. Предположим, директора завода интересуется, различается ли производительность труда у работников разного возраста на заводе при выполнении одних и тех же видов работ. Допустим, было осуществлено измерение производительности труда в двух разных возрастных группах (с объемом выборок 37 и 59 чел. соответственно). Получен следующий результат: средняя производительность труда в первой группе — 47 деталей/ч, во второй — 55 деталей/ч со стандартным отклонением 8 и 10 соответственно. Можно ли на основе полученных результатов сделать вывод о том, что производительность труда рабочих во второй группе выше, чем в первой? Какова вероятность того, что такой вывод может быть ошибочным?

Пример 5.3. Проверяется гипотеза о том, что женщины тратят больше времени на разговоры по телефону, чем мужчины. Предположим, что в исследовании принимали участие 52 мужчины и 43 женщины. Среднее время разговора составило 25 мин. в день у мужчин и 35 мин. в день у женщин (стандартное отклонение 7 и 15 мин. соответственно). На первый взгляд различия обнаружены, и эти результаты подтверждают гипотезу. Однако такой результат может быть получен случайно, даже если в генеральной совокупности различий нет, как и наоборот, когда различия на самом деле существуют. Поэтому закономерен вопрос: достаточно ли полученного различия в средних значениях для того, чтобы утверждать, что вообще *все* женщины в среднем говорят по телефону дольше, чем *все* мужчины? Какова вероятность, что это не так? Является ли это различие *статистически значимым*?

Как видно из приведенных примеров, *абсолютно точный* ответ о различиях или связях в отношении генеральной совокупности по результатам выборочного исследования получить невозможно. Поэтому необходимо определить, достаточно ли велика разность между средними значениями, чтобы эту разность можно было бы объяснить действием независимой переменной, а не случайностью, связанной с малым объемом выборки. Методы статистики позволяют оценить вероятность *случайного* получения такого различия при условии, что на самом деле различий в генеральной совокупности нет. Сферы применения этих методов (статистических тестов) очень широки и разнообразны. В качестве примеров можно назвать несколько приложений:

- сравнение числа отказов компьютеров разных производителей после 30-дневного тестирования;
- оценка уровня доходов разных групп населения;
- анализ предпочтения товаров в разных демографических группах;
- сравнение числа подписчиков на различные журналы;
- сравнение социально-экономических характеристик в разных регионах;
- сопоставление производительности труда в разных группах работников предприятия;
- анализ общественного мнения (выборы и др.).

5.2. Нулевая и альтернативная гипотезы

Первым шагом проверки статистической гипотезы является формулировка допущений. При решении любой задачи путем применения статистических методов требуется принять определенные предположения относительно данных, например о способе формирования выборки, об используемой измерительной шкале, виде распределения и т.д. Такие предположения называются *статистической моделью*. Очень важно обоснованно сформулировать статистическую модель, поскольку она обуславливает выбор конкретных статистических процедур, которые могут использоваться для проверки статистической гипотезы.

Нулевая гипотеза (null hypothesis) – это утверждение об *отсутствии* различий или связи. Согласно нулевой гипотезе (H_0) различие между значениями недостаточно велико, чтобы принять, что это различие вызвано неслучайными факторами. Точ-

ная формулировка нулевой гипотезы зависит от проводимой проверки. В случае с одной выборкой нулевая гипотеза может формулироваться следующим образом: изучаемая характеристика по выборке не отличается от определенной характеристики по генеральной совокупности.

Пример 5.4. Производительность труда в среднем по заводу составляет 50 деталей/ч. Допустим, выясняется, отличается ли производительность труда рабочих определенной возрастной группы от средних показателей в целом по заводу. Тогда нулевая гипотеза может быть сформулирована как $H_0: \mu = 50$. Такая формулировка означает, что проверяется равенство среднего значения по исследуемой группе рабочих (генеральной совокупности, из которой будет формироваться выборка) определенному значению (в данном случае – средней производительности труда).

Альтернативная гипотеза (alternative hypothesis) – гипотеза о значимости различий (утверждает наличие различий или существование связи). В соответствии с альтернативной гипотезой различия достаточно значимы и обусловлены влиянием неслучайных факторов. *Статистически значимое различие* – это различие, которое настолько велико, что вероятность его возникновения вследствие простой случайности крайне мала. Альтернативная гипотеза (H_A) является, как правило, «рабочей» гипотезой исследования, поскольку часто цель исследователя заключается в том, чтобы собрать доказательства, которые говорили бы в пользу альтернативной гипотезы и на основе которых можно было бы отклонить нулевую гипотезу.

Альтернативная гипотеза может быть *ненаправленной* и *направленной*. Если цель исследователя – показать, что значение изучаемого параметра по выборке «просто» отличается от значения параметра по генеральной совокупности, то альтернативная гипотеза символически формулируется с помощью знака « \neq ». Например, $H_A: \mu \neq 50$. В таком случае мы имеем дело с ненаправленной альтернативной гипотезой.

Если исследователя интересуют различия в определенном направлении, может использоваться *направленная* альтернативная гипотеза, учитывающая направление изменений (различий), т.е. если исследователь считает, что истинное значение генеральной совокупности выше или ниже, чем значение, указанное в нулевой гипотезе, то это мнение будет отражено в альтер-

нативной гипотезе. Например, мы можем предположить, что производительность труда в изучаемой возрастной группе больше, чем в среднем по заводу. Тогда альтернативная гипотеза будет иметь следующий вид: $H_A: \mu > 50$. Если мы предполагаем, что производительность труда в изучаемой возрастной группе меньше, чем в среднем по заводу, то альтернативная гипотеза формулируется как $H_A: \mu < 50$.

Если альтернативная гипотеза формулируется как направленная (например, $H_A: \mu > 0$), то нулевая гипотеза может формулироваться в следующем виде: $H_0: \mu \leq 50$. Однако символ строгого или нестрогого равенства (например, = или \leq , \geq) должен всегда присутствовать в формулировке нулевой гипотезы и никогда — в формулировке альтернативной гипотезы. В литературе существуют разные мнения по поводу возможности включения в формулировку нулевой гипотезы символов \leq и \geq , но с точки зрения процедуры проверки и содержательной интерпретации результатов такие, например, формулировки, как $H_0: \mu = 50$; $H_A: \mu > 0$ и $H_0: \mu \leq 50$; $H_A: \mu > 0$, полностью эквивалентны. Очень важно, чтобы выбор направления альтернативной гипотезы был основан на априорных теоретических предпосылках, а не только на эмпирических результатах. Поэтому если таких предпосылок нет, то следует формулировать альтернативную гипотезу как ненаправленную.

Нулевая и альтернативная гипотезы представляют полную группу несовместных событий: отклонение одной влечет принятие другой. Поэтому главный принцип проверки статистических гипотез состоит в том, что выдвигается нулевая гипотеза H_0 с тем, чтобы попытаться опровергнуть ее и тем самым подтвердить альтернативную гипотезу H_A . Если результаты статистического теста окажутся таковы, что позволят отклонить H_0 , это будет означать, что верна H_A , т.е. выдвинутая рабочая гипотеза подтверждается. Однако всегда следует помнить о том, что если мы не можем отклонить нулевую гипотезу, это не значит, что мы получили ее подтверждение, так как нулевая гипотеза никогда не может быть абсолютно подтверждена. В то же время из отклонения нулевой гипотезы всегда следует принятие альтернативной гипотезы.

При невозможности отклонения нулевой гипотезы нельзя декларировать отсутствие связи или различий, но выводом исследования будет являться именно то, что статистически значи-

мая связь или различия не обнаружены. С другой стороны, при отклонении нулевой гипотезы всегда остается вероятность того, что она верна. Поэтому нельзя *категорично* утверждать, что связь есть или что результаты *доказывают* наличие связи. Более корректным будет утверждение *статистически значимой* связи, что *свидетельствует* или *указывает* на наличие возможной связи.

Невозможность полного подтверждения нулевой гипотезы можно объяснить, например, следующими рассуждениями. Чем слабее зависимость между переменными, тем большего объема требуется выборка, чтобы ее обнаружить. Например, нужно сделать очень много бросков монеты, чтобы доказать, что отклонение от равной вероятности выпадения «орла» и «решки» составляет тысячные доли процента. Если различия или связь близки к нулю, то необходимый объем выборки для возможности их статистически значимого обнаружения приближается к объему всей генеральной совокупности.

Таким образом, при принятии нулевой гипотезы (точнее, невозможности ее отвергнуть) всегда остается вероятность того, что различия или связь все же существуют. Это означает лишь то, что не обнаружены статистически значимые различия. Некоторые авторы [18, 29] на основе этого утверждают, что невозможность отклонения нулевой гипотезы является отсутствием какого-либо результата. Но мы не согласны с тем, что невозможно сделать содержательный вывод на основе невозможности отклонения нулевой гипотезы. Действительно, «рабочей» гипотезой исследователя чаще всего становится альтернативная гипотеза. Однако бывают случаи, когда исследователь заинтересован в подтверждении (свидетельстве) отсутствия связи или различий. Например, проверка тест-ретест, оценка различий в заработной плате, эффективности работы, ожидаемых и фактических результатов, отклонений от нормального распределения и т.д. На основе необнаружения статистически значимых различий также можно принимать решения как в исследовательской работе, так и в бизнес-практике.

Еще раз обращаем внимание читателя на то, что нулевые гипотезы всегда формулируются в варианте «отсутствие различий» или «отсутствие связей», т.е. фактически в варианте нулевого различия (например, средних значений). При иной формулировке нулевые гипотезы было бы невозможно опровергнуть. Ведь если нулевая гипотеза формулировалась бы в терминах *на-*

личия связи или различий, то степень связи или различий была бы неопределенной, с бесконечным количеством вариантов. Предположим, что мы бы сформулировали $H_0: \mu \neq 0$. Что бы мы ожидали получить, если бы нулевая гипотеза была истинной? Все, что угодно! Нулевую гипотезу можно опровергнуть (с определенным уровнем значимости) только в том случае, если она сформулирована в виде строгого или нестрогого равенства.

5.3. Виды статистических ошибок и уровни статистической значимости

Вывод, полученный на основе выборочного исследования, с некоторой вероятностью может оказаться ошибочным, и обычно вероятность ошибки тем меньше, чем больше выборка. В общем случае чем больше получено результатов, тем в большей степени по различиям, например между двумя выборками, можно судить о том, что действительно имеет место в тех генеральных совокупностях, из которых взяты эти выборки. Однако обычно используемые выборки относительно невелики, и тогда вероятность ошибки может быть значительной.

Выделяют два главных типа статистических ошибок, возможных при проверке статистических гипотез. *Статистическая ошибка первого рода (Type I Error)*, или α -ошибка, — ошибка обнаружить различия или связи, которые на самом деле не существуют, т.е. это ошибка отклонения истинной нулевой гипотезы. *Статистическая ошибка второго рода (Type II Error)*, или β -ошибка, ошибка не обнаружить различия или связи, которые на самом деле существуют, т.е. это ошибка принятия решения, в результате которого ложная нулевая гипотеза не может быть отклонена¹.

Более «критичной» считается статистическая ошибка первого рода. В книге Д. Купера и П. Шиндлер [39] приводится «судебная» аналогия статистических ошибок двух видов. Предпо-

¹ М. Фокс, преподаватель Международного центра изучения творчества в г. Буффало (США), говорит еще об «ошибке третьего рода», которую он шутливо называет «Mike Fox Type Error» — «когда неправильная проблема решается правильными методами». Эта шутка прямого отношения к нашей теме не имеет, но она указывает на ситуацию, когда исследователи действительно получают ошибочный результат только потому, что уделили недостаточно внимания анализу разрабатываемой проблемы.

жим, присяжные заседатели выносят вердикт «не виновен» или «виновен». Согласно этой аналогии ошибкой первого рода будет ситуация, когда невинный обвинен, а ошибкой второго рода — когда виновный освобожден.

С понятиями ошибок первого и второго рода связано решение важного вопроса о том, что можно считать критерием отклонения или принятия статистической гипотезы, т.е. по каким критериям мы можем судить о том, что статистическая ошибка *маловероятна*, или о том, что связь *статистически значима*.

*Уровень значимости (level of significance)*¹ — это пороговая (критическая) вероятность ошибки, заключающейся в отклонении нулевой гипотезы, когда она верна. Другими словами, это допустимая (с точки зрения исследователя) вероятность совершения статистической ошибки первого рода — ошибки, связанной с тем, что различия сочтены существенными, а они на самом деле случайны. Обычно используют уровни значимости (обозначаемые α), равные 0,05; 0,01 и 0,001. Например, уровень значимости, равный 0,05, означает, что допускается не более чем 5%-ная вероятность ошибки.

Понятие уровня статистической значимости связано также со свойством нормального распределения выборочных средних, которое мы рассматривали в гл. 3. Из центральной предельной теоремы нам фактически заранее известно распределение средних для случая, когда верна нулевая гипотеза. Это распределение позволяет установить, насколько вероятно то или иное случайное отклонение среднего по выборке от среднего по генеральной совокупности. Например, мы знаем, что примерно 95% площади под кривой нормального распределения будет находиться в диапазоне $\pm 1,96 \sigma$ от среднего значения (рис. 5.1). Вероятность того, что выборочное значение попадет в этот диапазон, составляет 95%. Соответственно вероятность случайного отклонения выборочного среднего определяется площадью в «хвостах» под кривой нормального распределения — за пределами найденного диапазона. В контексте проверки статистических гипотез эта область в «хвостах» называется *критической*. Следовательно, вероятность того, что мы получим случайное отклонение выборочного среднего от истинного значения (при условии истин-

¹ Используются также термины *уровень достоверности*, *уровень надежности*, *доверительный уровень*, *вероятностный порог*.

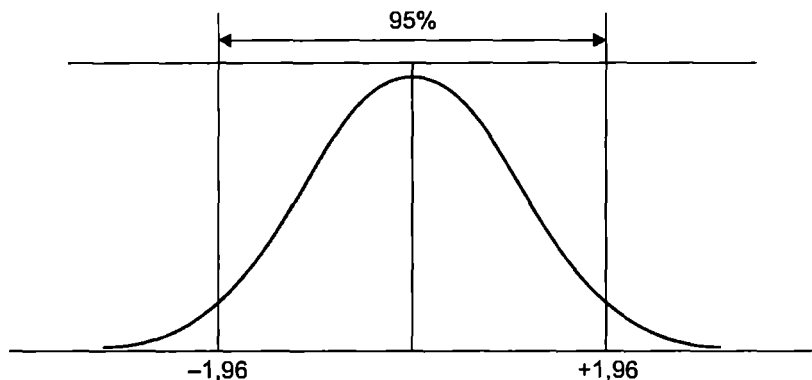


Рис. 5.1. Иллюстрация уровня значимости

ности нулевой гипотезы) составляет $1 - 0,95 = 0,05$. Это вероятность случайного получения результата, на основе которого может быть отклонена нулевая гипотеза, когда на самом деле для генеральной совокупности нулевая гипотеза является истинной. Для данного примера вероятность ошибки первого рода — $0,05$.

Таким образом, нулевую гипотезу можно отвергнуть в пользу альтернативной гипотезы, если по результатам статистического теста вероятность ошибки, т.е. вероятность случайного возникновения обнаруженного различия (*p-уровень*), не превышает 5 из 100 (имеется лишь 5 шансов из 100 ошибиться) для заданного уровня значимости в 5%. Если же этот уровень значимости не достигается (вероятность ошибки выше 5%), то считают, что разница вполне может быть случайной и поэтому нельзя отклонить нулевую гипотезу. Таким образом, *p-уровень значимости* (*p-value*) соответствует риску совершения ошибки первого рода (отклонения истинной нулевой гипотезы). Если $p < \alpha$, нулевая гипотеза отклоняется.

Чем меньше значение *p-уровня*, тем выше статистическая значимость и надежность результата исследования. Это показатель, представляющий собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю генеральную совокупность. Например, $p = 0,05$ показывает, что имеется 5%-ная вероятность ($1/20$), что найденная в выборке связь между переменными является лишь случайной особенностью данной выборки, т.е. если данная зависимость в генеральной совокупности отсутствует, то примерно в одном из

двадцати экспериментов на обнаружение связи можно было бы ожидать нахождения зависимости между переменными. Отметим, что это не означает, что исследуемая связь заведомо существует, но в среднем может быть воспроизведена в 5% случаев; речь идет именно о ее случайном или ошибочном обнаружении в 5% случаев.

Статистическая значимость возрастает (значение p -уровня уменьшается), если:

- величина связи или различия выборочных параметров выше;
- изменчивость признака меньше;
- объем выборки больше.

Соответствие некоторых значений α и z приведено в табл. 5.1, более подробно – в табл. П.1 приложения.

Таблица 5.1
Соотношение значений α и z

α	z
0,10	$\pm 1,65$
0,05	$\pm 1,96$
0,01	$\pm 2,58$
0,001	$\pm 3,29$

Критическая область включает все маловероятные и редкие значения выборочных результатов. Получение значения из этой области при обработке данных позволяет отклонить нулевую гипотезу. Остальная область состоит из всех значений выборочных результатов, являющихся «нередкими». Чем меньше значение альфа, тем меньше критическая область и тем больше расстояние между средним значением распределения выборочных статистик и граничными точками критической области.

Вопрос о приемлемом значении альфа, т.е. вопрос о том, при каком уровне можно отклонить H_0 , не имеет однозначного ответа. Для того чтобы свести к минимуму ошибку первого рода, следует использовать очень малые значения альфа. По мере уменьшения значения альфа критическая область становится меньшей и все больше удаляется от среднего значения распределения выборочных статистик, которому подчиняется выборка. Чем меньше значение альфа, тем меньше вероятность ошибки первого рода.

Однако по мере того, как уменьшается размер критической области (уменьшаются значения альфа), оставшаяся область (область между двумя критическими значениями) увеличивается. Следовательно, при увеличении значения альфа (например, с 0,01 до 0,05) вероятность ошибки второго рода уменьшается. Для установленного значения альфа вероятность ошибки второго рода уменьшается с ростом объема выборки. Если все остальное остается неизменным, тогда чем меньше значение альфа, тем меньше вероятность того, что значение выборочного результата попадет в критическую область. Это повышает вероятность принятия ошибки второго рода: неотклонение нулевой гипотезы, которая в действительности является ложной.

Итак, с уменьшением значения альфа снижается вероятность ошибки первого рода, но возрастает вероятность ошибки второго рода. Таким образом, обе ошибки обратно зависят друг от друга, и при одной проверке невозможно свести к минимуму обе ошибки. По мере уменьшения вероятности одной ошибки увеличивается вероятность другой, и наоборот. Поэтому проверка гипотез всегда содержит элемент риска, так как невозможно одновременно свести к минимуму вероятность ошибки и первого, и второго рода.

В конечном итоге это означает, что необходимо подбирать такое значение альфа, которое было бы балансом между двумя ошибками. Очень высокие значения альфа сведут к минимуму вероятность ошибки второго рода (признание ложного утверждения истинным), а очень низкие значения альфа сведут к минимуму вероятность ошибки первого рода (признание истинного утверждения ложным). Как мы отмечали, более критичной считается ошибка первого рода, поэтому обычно применяются низкие значения альфа (0,05; 0,01; 0,001 и ниже).

Не существует универсального способа определить, какой уровень значимости следует действительно считать «значимым» в каждом конкретном случае. Выбор определенного уровня значимости, выше которого результаты признаются незначимыми, является достаточно произвольным. На практике окончательное решение обычно зависит от того, был ли результат предсказан априори (т.е. до проведения опыта) или обнаружен апостериорно в результате многих анализов и сравнений, выполненных с множеством данных, а также от традиции, имеющейся в данной области исследований.

На практике значение α устанавливается исходя из «научных конвенций» – соглашений, принятых в научном сообществе на основе практического опыта в различных областях исследования. В частности, значение 0,05 считается общепринятым показателем значимого результата (т.е. $\alpha = 0,05$ рассматривается как «приемлемая граница» уровня ошибки). Результаты, значимые на уровне 0,01, обычно рассматриваются как высокозначимые. Традиционная интерпретация различных уровней статистической значимости, исходящая из $\alpha = 0,05$, приведена в табл. 5.2. Но не следует забывать, что данная классификация уровней значимости достаточно произвольна и является всего лишь неформальным соглашением, принятым на основе практического опыта.

Таблица 5.2

Интерпретация уровней статистической значимости

Уровень значимости	Решение	Возможный статистический вывод
$p > 0,1$	H_0 не может быть отклонена	Статистически достоверные различия не обнаружены
$p \leq 0,1$	Неопределенность в истинности H_0	Различия обнаружены на уровне статистической тенденции
$p \leq 0,05$	Отклонение H_0	Обнаружены статистически достоверные (значимые) различия
$p \leq 0,01$	Отклонение H_0	Различия обнаружены на высоком уровне статистической значимости

Следует также помнить, что чем больше число наблюдений, тем больше вероятность обнаружить некоторую связь или зависимость чисто случайно. Поэтому исследователь должен с осторожностью оценивать значимость «неожиданных» результатов. В связи с этим значение $\alpha = 0,05$ рекомендовано в качестве «границы значимости» для небольших выборок (когда высока вероятность ошибки второго рода). Если объемы выборок $n \geq 100$, то порог отклонения H_0 лучше снизить до $\alpha = 0,01$, т.е. принимать решение о наличии связи (различий) при $p \leq 0,01$.

Однако распространенное использование значения 0,05 является лишь общепринятой условностью, и нет причин, по которым нельзя было бы дать альфа любое другое разумное значение (например, 0,02, 0,03 или 0,07). Исследователь может выбирать и обосновывать такое значение альфа, которое выглядит прием-

лемым с точки зрения конкретного исследовательского проекта. С другой стороны, статистические пакеты (SPSS, STATISTICA и др.) вычисляют точное значение p -уровня (например, $p = 0,023$), и это значение в принципе можно указывать в исследовательских работах. Но с точки зрения принятых правил представления результатов статистических исследований корректнее указывать, что p -уровень меньше одного из принятых критических значений (0,01; 0,05 или 0,1), например $p < 0,05$.

Если при проверке гипотезы формулируется ненаправленная альтернативная гипотеза, то критические области, соответствующие отклонению нулевой гипотезы и принятию альтернативной, поровну распределяются по обоим «хвостам». Поэтому такие кривые называют *двусторонними (2-tailed)*. В этом случае если принят уровень статистической значимости α , то существуют два критических порога: $\alpha/2$ справа на кривой распределения и $-\alpha/2$ слева. Если проверяется направленная альтернатива, то говорят об *одностороннем (1-tailed)* критерии. В этом случае существует один критический порог α справа или $-\alpha$ слева в зависимости от направления альтернативы. При двусторонней альтернативе нулевая гипотеза отвергается при больших значениях силы связи (корреляции, различиях средних и т.д.), чем при односторонней, т.е. проверка ненаправленной гипотезы является более «строгой» при прочих равных условиях.

Существует простое соотношение между p -уровнями для направленных и ненаправленных альтернатив: для одного и того же эмпирического значения критерия p -уровень для направленной альтернативы в два раза меньше p -уровня для ненаправленной альтернативы. Критические значения z для одно- и двустороннего критериев для некоторых распространенных значений альфа приведены в табл. 5.3.

Таблица 5.3

Соотношение значений α и z для одно- и двустороннего критериев

α	Односторонний критерий	Двусторонний критерий
0,10	$\pm 1,29$	$\pm 1,65$
0,05	$\pm 1,65$	$\pm 1,96$
0,01	$\pm 2,33$	$\pm 2,58$
0,001	$\pm 3,10$	$\pm 3,29$

Значения z для других уровней значимости можно определить из табл. П.1 приложений или с помощью функции программы Excel НОРМСТОБР(P), возвращающей значение для *одностороннего* критерия z при заданной вероятности $P = 1 - \alpha$. Другая функция Excel – НОРМСТРАСП(z) возвращает значение P , из которого легко определяется α .

Например, в двусторонней проверке с альфа, равной 0,05, критическая область начинается со стандартизованного значения $z = \pm 1,96$. В случае одностороннего критерия при том же значении альфа критическое значение z равно +1,65, если задан правый «хвост» распределения, и –1,65, если задан левый «хвост» распределения. Критические значения для односторонних критериев расположены ближе к среднему значению распределения выборочных статистик. Это еще раз показывает, что при одностороннем критерии повышается вероятность отклонения нулевой гипотезы без изменения значения альфа. Как мы уже говорили, односторонние критерии следует использовать, когда направление различий можно с уверенностью предсказать, исходя из неких априорных теоретических предпосылок, позволяющих *заранее* (до эксперимента) предположить направленное различие. Если существуют сомнения при выборе между направленной и ненаправленной альтернативами, то всегда следует выбирать ненаправленную альтернативную гипотезу.

Из подтверждения альтернативной гипотезы на уровне $p < \alpha$ нельзя сделать вывод, что вероятность аналогичного результата при повторном исследовании будет больше $1 - \alpha$. О вероятности получения того или иного результата в повторных исследованиях сказать невозможно из-за соотношения ошибок первого и второго рода. На основе статистически значимой связи нельзя также делать вывод о наличии причинно-следственной связи между явлениями (признаками). Для утверждения причинно-следственной связи необходимо обоснование двух дополнительных признаков: а) одно явление предшествует другому во времени; б) отсутствует общая причина изменения признаков (явлений).

5.4. Общая процедура проверки статистических гипотез

Для принятия решений о том, можно ли отклонить нулевую гипотезу и принять альтернативную, используют *статистические критерии*, которые включают методы расчета определенного

показателя, на основе которого принимается решение об отклонении нулевой гипотезы, а также правила (условия) принятия решения. Этот рассчитываемый показатель называется *эмпирическим (или экспериментальным) значением критерия*. Найденное эмпирическое значение сравнивается с известным (например, заданным таблично или определенным с помощью той или иной статистической программы) эталонным числом, именуемым *критическим значением критерия*. В статистических таблицах критические значения приводятся, как правило, для нескольких уровней значимости: 5% (0,05), 1% (0,01) и др.

Статистический критерий зависит также от числа степеней свободы. **Число степеней свободы (*degrees of freedom*)** – это количество значений в распределении, которые свободны для изменения. Как правило, число степеней свободы линейно зависит от объема выборки (например, $df = n - 1$), а также от числа признаков или их градаций. Однако не существует единой формулы для определения числа степеней свободы для всех возможных случаев, поэтому статистический критерий также устанавливает формулу для расчета числа степеней свободы. Все показатели, определяемые для нахождения эмпирического и критического значений критерия, называют *статистиками критерия*.

Для большинства статистических критериев действует следующее правило: если эмпирическое значение критерия для данного числа степеней свободы оказывается ниже критического уровня, соответствующего выбранному значению α (порогу вероятности), то нулевая гипотеза не может считаться опровергнутой. Это означает, что выявленная разница (или связь) недостоверна. Другими словами, если эмпирическое значение критерия оказывается меньше или равно критическому, то можно сделать вывод, что характеристики распределений совпадают на уровне значимости α . Если эмпирическое значение критерия, напротив, оказывается строго больше критического, то нулевая гипотеза отклоняется и принимается альтернативная гипотеза – характеристики распределений считаются различными с достоверностью различий $1 - \alpha$. Например, если $\alpha = 0,05$ и принята альтернативная гипотеза, то достоверность различий равна 0,95, или 95%. Чем больше эмпирическое значение критерия критического значения, тем сильнее различаются характеристики сравниваемых объектов. Для некоторых статистических критериев

нулевая гипотеза отклоняется при ином соотношении эмпирического и критического значений (гл. 8).

Таким образом, общая процедура проверки статистической гипотезы включает следующие шаги.

1. Сформулировать статистическую модель (допущения относительно способа формирования выборки, используемой шкалы, вида распределения и т.д.).
2. Сформулировать нулевую и альтернативную гипотезы.
3. Выбрать соответствующий статистический критерий.
4. Выбрать требуемый уровень значимости ($\alpha = 0,1; 0,05; 0,01; 0,001$ и т.д.).
5. Вычислить эмпирическое значение критерия.
6. Сравнить эмпирическое значение критерия с критическим значением.
7. Принять решение об отклонении или невозможности отклонения нулевой гипотезы.

5.5. Статистические методы и критерии проверки гипотез

Для того чтобы судить о том, какова вероятность ошибки при отклонении или неотклонении нулевой гипотезы, применяют статистические критерии, соответствующие задачам исследования и особенностям выборки (выборок). Для данных, полученных в метрических шкалах (интервальных или абсолютных), при распределениях, близких к нормальным, используют *параметрические критерии*.

В частности, для определения достоверности разницы средних для двух выборок применяют *t*-критерий Стьюдента, а для того, чтобы судить о различиях между тремя или большим числом выборок, — дисперсионный анализ (ANOVA) или ANOVA с повторными измерениями. Корректное использование многих параметрических методов ограничивается определенными требованиями, накладываемыми на объем выборок, вид распределения, однородность дисперсий и др.

Если исследователь имеет дело с данными, полученными в неметрических шкалах, или если выборки слишком малы для уверенного заключения о том, что генеральные совокупности, из которых они взяты, подчиняются нормальному распределению, используют *непараметрические критерии* — χ^2 (хи-квадрат),

Манна—Уитни, Уилкоксона и др. Эти критерии довольно просты с точки зрения как расчетов, так и применения.

Непараметрические критерии менее чувствительны, чем параметрические, поэтому, если важно обнаружить даже слабые различия (например, оцениваются возможные негативные эффекты влияния тех или иных лекарственных средств на здоровье людей), следует особенно тщательно выбирать метод и уровень значимости. Преимуществом непараметрических критериев является то обстоятельство, что их можно использовать и при сравнительно небольших выборках, в отличие от многих параметрических критериев.

Выбор статистического критерия также зависит от того, являются ли выборки, средние которых сравниваются, независимыми (например, взятыми из двух разных групп испытуемых) или зависимыми (например, отражающими результаты одной и той же группы испытуемых до и после воздействия или после двух различных воздействий).

В зависимости от условий проверки можно использовать свыше 100 различных вариантов тестирования. Поэтому перед исследователем может возникнуть проблема выбора и корректного применения соответствующего статистического метода. Для демонстрации многообразия статистических критериев и различий их использования в зависимости от количества выборок и типов применяемых шкал в табл. 5.4 приводятся несколько основных критериев и методов.

Приведенная таблица может помочь в выборе необходимого метода проверки в зависимости от имеющихся у исследователя данных, поставленной цели и задач исследования, но она не определяет этот выбор полностью. Дело в том, что даже в одной и той же формальной ситуации (тип данных, количество выборок, характер связи, вид распределения и т.д.) может быть применено несколько методов и критериев, каждый из которых имеет как достоинства и недостатки, так и особенности применения. Например, критерий Уилкоксона предполагает, что можно ранжировать различия между сравниваемыми наблюдениями. Если это не так, то следует использовать критерий знаков. Такие нюансы будут оговариваться при рассмотрении параметрических и непараметрических методов в следующих главах, а пока можно предложить общее правило: если исследование является практически важным и экономически затратным, то всегда це-

Условия применимости статистических критериев

Тип шкалы	Одна выборка	Две выборки		Больше двух выборок	
		зависимые	независимые	зависимые	независимые
Номина- тивная	Одновы- борочный тест χ^2	Тест Мак—Не- мара	Двухвыбо- рочный тест χ^2	Q-тест Кокрена	Многовы- борочный χ^2
Порядко- вая	Одновы- борочный тест Кол- могоро- ва—Смир- нова	Тест Уил- коксона; критерий знаков	U-критерий Манна— Уитни; двухвыбо- рочный тест Колмогоро- ва—Смир- нова; критерий серий Уаль- да—Волфо- вица	Критерий χ^2 Фрид- мана	Тест Крас- кела—Уол- лиса
Метриче- ская	t-критерий Стьюдента; z-тест	t-тест для зависи- мых вы- борок	t-критерий для незави- симых вы- борок; z-тест	ANOVA с повтор- ными измере- ниями	ANOVA

лесообразно использовать различные методы, применение которых корректно в данной ситуации. Если результаты проверки различными методами будут существенно различаться, следует попытаться понять причины этих различий с точки зрения как содержания исследовательской ситуации, так и анализа примененных процедур (выполнены ли необходимые и достаточные предпосылки их использования, является ли данный критерий достаточно чувствительным, соответствует ли выбранный уровень значимости данной задаче и т.д.).

5.6. Статистическая значимость и практическая важность

Как уже было отмечено, многие исследователи иногда не уделяют должного внимания проверке и обоснованию статистической значимости исследовательских результатов. Но есть и дру-

гая крайность — исключительно формальная проверка полученных выводов. Корректно применив тот или иной статистический критерий проверки гипотез (с точки зрения выполнения всех предпосылок использования критерия) и подтвердив статистическую значимость искомой связи, различий и т.д., исследователь относится к такому формальному подтверждению как к некоему магическому знаку, который автоматически гарантирует качество результатов и выводов. Однако случается, что результаты, имеющие статистическую значимость, лишены содержательного смысла в контексте предмета исследования, т.е. они не имеют содержательной значимости. Это может быть вызвано многими причинами: плохо поставленными целями и задачами исследования; ошибками в сборе и обработке данных; неправильной интерпретацией результатов; революционностью получения результатов (выход за пределы той или иной научной парадигмы) и т.д. Но анализ таких вариантов не входит в задачи данного учебного пособия, поэтому рассмотрим лишь математические причины того, что результаты исследования могут быть статистически значимы, однако при этом не имеют содержательной значимости и практической важности.

Подтверждение статистически значимых различий осуществляется путем отклонения нулевой гипотезы. Мы знаем, что вероятность отклонения нулевой гипотезы связана с уровнем значимости, определяемым выбранным значением альфа. Чем выше значение альфа, тем больше критическая область и тем выше вероятность отклонения нулевой гипотезы. Например, при значении альфа 0,05 отклонить H_0 проще, чем при значении альфа 0,01, а при значении альфа 0,10 это сделать еще проще. С другой стороны, увеличение значения альфа ведет к повышению вероятности совершения ошибки первого рода, т.е. может оказаться, что мы сочтем несущественные различия статистически значимыми.

Нам также известно, что проверка ненаправленной гипотезы является более «строгой», чем направленной, поскольку в последнем случае уровень значимости фактически уменьшается ровно в два раза (значение альфа не распределяется пополам по «хвостам», а фиксирует критическую область с одной стороны). Поэтому неоправданное применение одностороннего критерия повышает вероятность отклонения нулевой гипотезы.

Еще одним фактором риска совершения ошибки первого рода является размер выборки: при увеличении объема выборки повышается вероятность отклонения нулевой гипотезы. Эмпирические значения критерия возрастают с увеличением количества элементов выборки, а чем сильнее отличается эмпирическое значение критерия от его критического значения, тем более значимыми полагаются различия, т.е. на больших выборках совершить ошибку первого рода более вероятно. Сказанное, конечно, не означает, что исследователь не должен работать с большими выборками. Напротив, большая выборка может быть лучшим приближением генеральной совокупности, которую она представляет, поэтому к решениям, основанным на выборках большего размера, можно относиться с большим доверием, чем к решениям, основанным на выборках малого размера. Но по большим выборкам даже сравнительно небольшие различия, не представляющие никакого интереса для исследователя, могут быть статистически значимыми. Именно поэтому при больших выборках ($n \geq 100$) значение альфа целесообразно уменьшить до 0,01.

Таким образом, и в случае больших, и в случае малых выборок обнаружение статистически значимых различий не гарантирует их теоретическую или практическую важность. С одной стороны, различие, которое не является статистически значимым, не может считаться важным, поскольку с точки зрения статистики этого различия вообще не существует, точнее, оно имеет случайный характер. Но, с другой стороны, статистически значимое различие может не иметь никакой важности для исследования. Поэтому помимо статистической значимости необходимо всегда обосновывать содержательное наполнение обнаруженных различий и связей.

Задания для самостоятельного выполнения

1. Ответьте на следующие вопросы:

- а) что общего и различного между параметрическими и непараметрическими тестами?
- б) какие общие предпосылки лежат в основе параметрических тестов?
- в) какие тесты можно назвать «независимыми от распределения» и почему?

г) почему невозможность отклонения нулевой гипотезы не означает ее принятие?

д) когда риск совершения ошибки первого рода является более критичным, чем риск совершения ошибки второго рода (приведите примеры)?

е) когда риск совершения ошибки второго рода является более критичным, чем риск совершения ошибки первого рода (приведите примеры)?

ж) каковы пути снижения вероятности ошибки первого рода?

з) каковы пути снижения вероятности ошибки второго рода?

и) в чем различие статистической и практической значимости? Приведите примеры;

к) каким образом могут быть проверены все предпосылки для применения дисперсионного анализа?

Пояснение. Для ответа на вопросы д) и е) рассмотрите, например, ситуации венчурного инвестирования (инвестирования в технологически новые и рискованные проекты), приобретения крупной партии оборудования и подобные; представьте, что исследуются такие вопросы, как: оказывает ли помощь людям дорогостоящая и болезненная терапия, является ли данная пищевая добавка опасной для людей и т.п. Сформулируйте для таких ситуаций соответствующие гипотезы и определите возможные ошибки первого и второго рода.

2. Для каждой из следующих ситуаций определите z критическое.

Альфа	Вид альтернативы	$z_{\text{крит}}$
0,05	Односторонняя	
0,10	Двусторонняя	
0,06	Двусторонняя	
0,01	Односторонняя	
0,02	Двусторонняя	

Пояснение. Критическое значение z -критерия можно посчитать в программе Excel с помощью функции НОРМСТОБР(P), возвращающей значение для одностороннего критерия z при заданном $P = 1 - \alpha$. Для двустороннего критерия следует в качестве аргумента функции подставлять $P = 1 - \alpha/2$ (объясните, почему). Например, для $\alpha = 0,06$ для одностороннего критерия имеем $P = 0,94$ и $z = 1,55$, для двустороннего – $P = 0,97$ и $z = 1,88$.

Параметрические методы проверки статистических гипотез

6.1. Применение z-критерия

Самым простым и известным параметрическим методом проверки статистических гипотез является *z-тест (z-критерий)*. Этот метод может использоваться при выполнении базовых предположений о нормальности распределения и представленности данных в метрической шкале.

Рассмотрим вначале *случай одной выборки*. Как мы уже знаем, эмпирическое значение z-критерия показывает, насколько выборочное среднее отличается от среднего по генеральной совокупности в единицах стандартного отклонения, и определяется по формуле

$$z = \frac{\bar{x} - \mu}{\sigma_x}$$

Для выборки, сформированной методом случайного повторного отбора, получаем (см. гл. 3):

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Вычислив эмпирическое значение z и сравнив его с критическим значением при заданном уровне значимости (табл. П.1), можно сделать вывод о том, является ли случайным различие средних значений по выборке и генеральной совокупности. Пусть, например, получено эмпирическое значение $z = 2$. Для этого значения определяется p – уровень значимости (из статистической таблицы или с помощью программы). Для данного случая $p = 0,046$, т.е. вероятность того, что данная выборка принадлежит генеральной совокупности со средним μ , составляет 4,6%. Другими словами, это значение соответствует вероятности того, что данный выборочный результат мог быть получен случайно, когда на самом деле в генеральной совокупности вер-

на H_0 . Следовательно, при $\alpha = 0,05$ нулевая гипотеза отклоняется, поскольку $p < \alpha$.

Во многих случаях измерение представляющей интерес характеристики элементов выборки проводится таким образом, что нельзя при этом подтвердить предположение о принадлежности полученных данных к метрической шкале измерений. Один из выходов в данной ситуации заключается в том, чтобы использовать в качестве критерия не среднее значение выборки, а выборочную пропорцию. Общая процедура проверки гипотезы в этом случае идентична процедуре, используемой при проверке средних значений. В этом случае основным для исследователя по-прежнему является вопрос: «Имеет ли генеральная совокупность, из которой была извлечена выборка, данную характеристику?» Следовательно, мы выполняем проверку, основываясь на предположении, что нулевая гипотеза истинна. Если полученное эмпирическое значение z попадает в критическую область (что маловероятно, если предположить, что H_0 истинна), тогда нулевая гипотеза отклоняется.

Формула для расчета экспериментального значения z при сравнении пропорций:

$$z = \frac{|P_s - P_p|}{\sqrt{P_p(1 - P_p)/n}},$$

где P_s — доля элементов выборочной совокупности, обладающих этим свойством;

P_p — доля элементов генеральной совокупности, обладающих определенным свойством.

Пример 6.1. В случайной повторной выборке из жителей самого неблагополучного района города 76% респондентов во время последних выборов голосовали за представителей партии «Партия». В целом по городу за эту партию проголосовало 66% электората. Можно ли сделать вывод, что среди жителей неблагополучного района значительно больше людей голосовало за «Партию»? Для проверки гипотезы имеем следующие данные: $P_s = 0,76$; $P_p = 0,66$; $n = 103$.

Нулевая гипотеза (H_0 : $P_p = 0,66$) утверждает, что выборка была получена из генеральной совокупности, 66% которой проголосовало за «Партию». Эмпирическое значение $z = 2,13$ (читателю предлагается проверить это самостоятельно). В данном

случае рассматривается гипотеза, действительно ли среди жителей неблагополучного района значительно *больше* людей голосовало за «Партию», поэтому должна использоваться односторонняя проверка. При значении альфа, равном 0,05, критическая область начнется с критического значения, равного 1,65. Критическая область располагается в правом «хвосте», так как мы предполагаем, что выборка извлечена из генеральной совокупности, в которой за «Партию» проголосовало *больше* людей, чем в целом по городу. Эмпирическое значение равно 2,13, поэтому нулевую гипотезу нужно отклонить. Следовательно, различие между жителями района и жителями города в целом является статистически значимым и имеет предсказанное направление.

Теперь рассмотрим возможности применения z-теста для *случая двух выборок*. В случае с одной выборкой предполагается, что она была получена на основе вероятностного (случайного) отбора. Это подразумевает, что у всех наблюдений в генеральной совокупности должны быть равные шансы на то, чтобы попасть в выборку. В ситуации с двумя выборками они должны быть сформированы не только по схеме случайного отбора, но и еще независимо друг от друга. Это требование удовлетворяется в случаях, когда выбор конкретного объекта для одной выборки не оказывает влияния на вероятность того, что некоторый другой объект будет включен в другую выборку. Для того чтобы удовлетворить этому условию, можно составить одну случайную выборку из общего списка генеральной совокупности, а затем разделить ее элементы на группы. Если исходная выборка была получена методом случайного отбора, то все подвыборки будут удовлетворять предположению о независимых случайных выборках.

Нулевая гипотеза в случае с двумя выборками по-прежнему представляет собой утверждение об отсутствии различий. Однако теперь вместо того, чтобы проверять значимость различия между средними по выборке и генеральной совокупности, из которой эта выборка была извлечена, проверяется значимость различия между средними по двум выборкам. Нулевая гипотеза утверждает, что между двумя генеральными совокупностями, из которых извлечены эти две выборки, нет различий. Если статистика критерия попадает в критическую область, то нулевую гипотезу, утверждающую отсутствие различий между генеральными совокупностями, можно отклонить.

Если нулевая гипотеза истинна, то разность между средними значениями генеральных совокупностей должна быть близкой к нулю и различия между средними значениями выборок будут небольшими. Чем больше различия между средними значениями выборок, тем больше вероятность того, что действительно существует различие между генеральными совокупностями, представленными этими выборками.

Если выборки имеют сравнительно большой размер (суммарное количество наблюдений в двух выборках превышает 100), то распределение разностей средних значений выборок будет очень близко к нормальному и для установления критических областей можно использовать кривую нормального распределения.

В случае двух выборок эмпирическое значение z определяется по формуле

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_{x_1 - x_2}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s_1^2 / (n_1 - 1) + s_2^2 / (n_2 - 1)}},$$

где $\bar{x}_1 - \bar{x}_2$ — разность средних значений выборок;

$\sigma_{x_1 - x_2}$ — стандартное отклонение распределения значений разностей средних значений выборок;

s_1, s_2 — стандартные отклонения по первой и второй выборкам;

n_1, n_2 — объемы первой и второй выборок.

В остальном процедура тестирования не отличается от случая одной выборки.

Проверка значимости различий между двумя пропорциями выборок аналогична проверке значимости различий между выборочными средними. Нулевая гипотеза утверждает, что между генеральными совокупностями, из которых были извлечены выборки, нет различий по проверяемой переменной. При больших размерах выборок (сумма двух выборок превышает 100) для нахождения границы критической области можно по-прежнему использовать z -распределение. В данном случае эмпирическое значение z определяется по формуле

$$z = \frac{|P_{s_1} - P_{s_2}|}{\sigma_{s_1 - s_2}} = \frac{|P_{s_1} - P_{s_2}|}{\sqrt{P_p(1 - P_p)} \sqrt{(n_1 + n_2) / (n_1 n_2)}},$$

$$\text{где } P_p = \frac{n_1 P_{s1} + n_2 P_{s2}}{n_1 + n_2},$$

$P_{s1} - P_{s2}$ – разность между пропорциями выборок;

σ_{s1-s2} – стандартное отклонение распределения разностей выборочных пропорций;

n_1, n_2 – объемы выборок.

Дальнейшая процедура тестирования аналогична случаю с одной выборкой.

6.2. Применение *t*-критерия Стьюдента для независимых выборок

В больших выборках (содержащих 100 и более наблюдений) выборочное стандартное отклонение дает адекватную оценку σ_x . Для выборок небольшого объема (сумма элементов выборок меньше 100) и при неизвестном σ_x *z*-критерий применять нельзя, так как распределение средних соответствует другому теоретическому распределению – *t*-распределению Стьюдента. Форма *t*-распределения изменяется в зависимости от размера выборки. При небольших размерах выборки кривая *t*-распределения намного более пологая, чем кривая *z*-распределения, но при увеличении размера выборки *t*-распределение все больше напоминает *z*-распределение. При объеме выборки больше 120 эти распределения практически идентичны, поэтому для больших выборок *z*- и *t*-тесты обеспечивают почти одинаковые результаты. По этой причине в большинство статистических пакетов не включается процедура проверки гипотез на основе *z*-критерия, но там обязательно присутствует процедура *t*-тестирования.

Наиболее часто *t*-критерий Стьюдента применяется для обнаружения различий между средними значениями двух выборок. Важное преимущество *t*-критерия – возможность его использования на малых выборках (с 10 и даже меньшим количеством элементов). Для применения *t*-критерия необходимо рассчитать число степеней свободы (*df*) с целью правильного определения критических значений *t*-критерия. В случае одной выборки $df = n - 1$, для двух выборок $df = n_1 + n_2 - 2$, где n_1, n_2 – объемы выборок. Критические значения для *t*-распределения при за-

данных уровнях значимости и числе степеней свободы можно взять из соответствующих статистических таблиц (табл. П.2) или рассчитать, например, с помощью функции программы Excel **СТЮДРАСПОБР**(α , df).

Общая логика проверки и даже формула расчета эмпирического значения t -критерия остаются аналогичными тем, что и при использовании z -теста. Для небольших выборок критическое значение t будет больше, чем критическое значение z при том же уровне значимости. Это связано с тем, что t -распределение более пологое, чем z -распределение. Поэтому при использовании t -распределения начальные точки критических областей будут находиться дальше от среднего значения выборочного распределения, следовательно, нулевую гипотезу будет сложнее отклонить. Более того, чем меньше размер выборки (чем меньше степеней свободы), тем более высоким будет эмпирическое значение t , необходимое для отклонения нулевой гипотезы.

Рассмотрим использование t -теста в *случае одной выборки*. Эмпирическое значение t определяется по формуле (при повторном случайном отборе)

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}},$$

где \bar{x} – среднее значение изучаемой характеристики, определенное по выборочной совокупности;

μ – среднее значение по генеральной совокупности (предполагаемое и проверяемое в ходе тестирования);

s – стандартное отклонение выборки;

n – объем выборки.

Пример 6.2. Известно, что производительность труда в среднем по заводу составляет 50 деталей/ч. Необходимо установить, отличается ли производительность труда рабочих определенной возрастной группы от этого значения. Формулируем нулевую и альтернативную гипотезы:

H_0 : $\mu = 50$ – нулевая гипотеза;

H_A : $\mu \neq 50$ – альтернативная гипотеза.

Данные, полученные в ходе исследования, представлены в табл. 6.1.

Данные выборочного исследования

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	61	80	40	71	39	20	40	80	80	60	80	29	51	43	47
№ п/п	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
x_i	40	41	37	41	60	40	60	31	40	65	61	40	59	60	80

Находим $\bar{x} = 52,5$ – среднее по выборке, $s = 17,1$ – стандартное отклонение по выборке, $n = 30$ – объем выборки. Подставляя эти значения в формулу t -критерия, получаем $t = 0,81$ – эмпирическое значение критерия. Для уровня значимости $\alpha = 0,05$ и числа степеней свободы $df = n - 1 = 29$ критическое значение t -критерия составляет 2,05. Поскольку эмпирическое значение меньше критического, то H_0 не может быть отклонена (проверить эти расчеты читателю предлагается самостоятельно).

Аналогичные результаты можно получить, воспользовавшись соответствующей процедурой пакета SPSS. Для этого необходимо в меню *Analyze* выбрать *Compare Means*, затем – *One-Sample T Test*. В поле *Test Value* устанавливается значение, с которым сравнивается выборочное среднее (в данном случае значение 50), затем нужно щелкнуть по кнопке *OK*. В результате получаем базовые статистики по выборке (*One-Sample Statistics*) (среднее, стандартное отклонение и стандартную ошибку среднего) и статистику одновыборочного t -критерия (*One-Sample Test*). Мы видим, что эмпирическое значение критерия составляет 0,81, а p -уровень значимости при двусторонней проверке – 0,425 ($df = 29$), что значительно выше $\alpha = 0,05$. Следовательно, H_0 не может быть отклонена (табл. 6.2).

При использовании t -критерия для случая с двумя независимыми выборками помимо предположения о нормальности распределения изучаемого признака в выборках вводится дополнительное предположение о равенстве (не слишком большом различии) дисперсий генеральных совокупностей, из которых формируются выборки. Для проверки нормальности распределения можно применить подходы, описанные в гл. 3. Предположение о равенстве дисперсий генеральных совокупностей можно проверить с помощью критериев однородности дисперсий, которые доступны в SPSS и STATISTICA, например крите-

**Отчет SPSS с результатами проверки статистической гипотезы
на основе *t*-критерия Стьюдента
One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
SAMPLE1	30	52,5333	17,12980	3,12746

One-Sample Test

Test Value = 50				
	t	df	Sig. (2-tailed)	Mean Difference
SAMPLE1	,810	29	,425	2,5333

рий Левена (*Levene test*) или критерий Брауна–Форсайта (*Brown–Forsythe test*). Однако если проверка этого предположения специально не проводится, можно считать, что имеет место равенство дисперсий генеральных совокупностей, если размеры выборок приблизительно одинаковы. В случае, если распределение признака хотя бы в одной из выборок заметно отличается от нормального и/или дисперсии статистически значимо различаются, альтернативой данному методу может быть непараметрический критерий Манна–Уитни (описан в гл. 8).

Формула для расчета экспериментального значения *t*-критерия:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{x_1 - x_2}},$$

$$\text{где } \sigma_{x_1 - x_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}};$$

s_1, s_2 – стандартные отклонения по первой и второй выборкам;

n_1, n_2 – объемы выборок.

Эта формула применяется для точных расчетов, когда выборки заметно различаются по численности. Для приближенных расчетов в случае, если численность выборок примерно одинакова, можно воспользоваться более простой формулой, в которой

$$\sigma_{x_1-x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Предположим, что менеджер крупной торговой компании намерен сравнить эффективность обучения персонала двумя разными методами (*A* и *B*), причем у него есть основания полагать, что первый метод должен обеспечивать лучший результат для данной сферы деятельности. Были отобраны 22 вновь поступивших сотрудника, которые случайным образом поделены на две группы. Одна группа прошла обучение по методу *A*, другая — по методу *B*. В конце года менеджер компании получил данные об объемах продаж в каждой группе на одного сотрудника (табл. 6.3).

Таблица 6.3

Средние объемы продаж, приходящиеся на одного сотрудника, тыс. руб.

Показатель	Группа <i>A</i>	Группа <i>B</i>
Среднемесячные продажи	1500	1300
Стандартное отклонение	225	251

Следуя стандартной процедуре проверки статистических гипотез, проверяем, действительно ли метод *A* дает лучший результат по сравнению с методом *B*.

H_0 : объемы продаж за рассматриваемый период не различаются в двух группах (эффективность методов одинакова);

H_A : объемы продаж в первой группе больше, чем во второй (метод *A* обеспечивает лучшие продажи по сравнению с методом *B*).

Используем *t*-критерий Стьюдента для проверки гипотезы, поскольку данные представлены в метрической шкале и выборки рассматриваются как независимые. Выбранный уровень значимости — 0,05, проверяется направленная альтернатива. Подстановка имеющихся данных в формулу для расчета экспериментального значения *t*-критерия дает значение $t = 200/101,63 = 1,97$ (читателю рекомендуется получить это значение самостоятельно). Критическое значение *t*-критерия — 1,725, при $df = (11 - 1) + (11 - 1) = 20$. Это значение можно определить из табл. П.1 приложения или с помощью функции программы Excel СТЬЮДРАСПОБР(0,1; 20). Обратите внимание на то, что выб-

ран уровень значимости $\alpha = 0,05$ для односторонней альтернативы, поэтому в функцию Excel в качестве первого параметра подставляем значение 0,1. Таким образом, мы можем отклонить нулевую гипотезу и принять, что метод *A* обеспечивает лучшую эффективность тренингов по сравнению с методом *B*. Если бы проверялась ненаправленная альтернативная гипотеза, то критическое значение *t*-критерия для $\alpha = 0,05$ составило бы 2,086, и нулевая гипотеза не могла быть отклонена на этом уровне значимости, хотя могла быть отклонена на уровне $\alpha = 0,1$.

Рассмотрим, как может быть осуществлена процедура проверки статистической гипотезы на основе *t*-критерия (для двух независимых выборок) с помощью статистических пакетов. Допустим, известны результаты измерения производительности труда рабочих двух разных возрастных групп (табл. 6.4).

Таблица 6.4

Производительность труда в двух возрастных группах

№ п/п	Группа 1	Группа 2	№ п/п	Группа 1	Группа 2	№ п/п	Группа 1	Группа 2
1	59	61	11	56	80	21	58	40
2	56	80	12	37	29	22	71	60
3	85	40	13	56	51	23	31	31
4	92	71	14	42	43	24	42	40
5	48	39	15	49	47	25	34	65
6	59	20	16	70	40	26	51	61
7	42	40	17	30	41	27	70	40
8	87	80	18	58	37	28	42	59
9	42	80	19	79	41	29	67	60
10	73	60	20	84	60	30	56	80

Для обработки этих данных в пакете SPSS необходимо сгруппировать их, а именно представить в виде двух переменных. Первая (зависимая) переменная соответствует изучаемому признаку (в данном случае – производительность труда), а вторая (группирующая) – принадлежности к той или иной группе. Объединим результаты измерения производительности труда в одну переменную, содержащую 60 значений (назовем эту переменную, например, *perfom*), в качестве второй переменной будем использовать столбец, первые 30 элементов которого будут иметь значение (код) 1, а оставшиеся 30 – значение 2 (дадим этой переменной имя *group*).

После ввода данных в меню *Analyze* нужно выбрать опцию *Compare Means*, затем – *Independent Samples T Test*. В поле *Test Variable(s)* переносится переменная *perform*, а в поле *Grouping Variable* – переменная *group* (рис. 6.1). В окне *Define Groups* указывается кодировка группирующей переменной (в данном примере одна группа обозначена кодом 1, другая – 2).

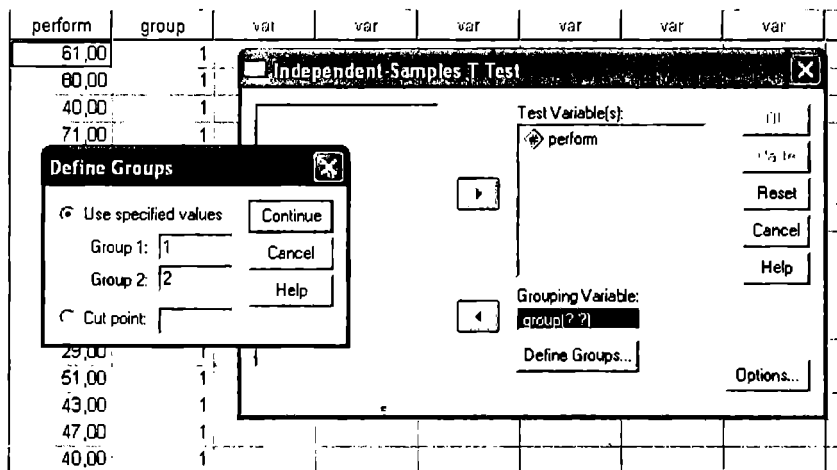


Рис. 6.1. Использование *t*-критерия для независимых выборок в пакете SPSS

Отчет SPSS показывает (табл. 6.5), что, несмотря на, казалось бы, довольно большое различие в средних (52,5 и 57,5), это различие не является статистически значимым: $p = 0,256 > \alpha = 0,05$. Результатом теста Левена является то, что гипотеза однородности дисперсий двух выборок не может быть отклонена (Sig. $> 0,05$), поэтому применение *t*-критерия вполне корректно. Незначительные различия выборочных стандартных отклонений приводят к практически неразличимым результатам при предположении равных дисперсий (*Equal variances assumed*) и отсутствии такого предположения (*Equal variances not assumed*).

Такие же результаты, но представленные в другой форме, могут быть получены в пакете STATISTICA. Для этого необходимо запустить модуль *Basic Statistics and Tables*, где в меню *Analysis* выбрать *t-test for independent samples* (рис. 6.2). В этом пакете данные могут быть представлены как в сгруппированном виде (*One record per case (use a grouping variable)*), подобно тому, как это

Статистика использования *t*-критерия для независимых выборок
Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
PERFORM	1	30	52,5333	17,12980	3,12746
	2	30	57,5333	17,27812	3,15454

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
Equal variances assumed	,176	,676	-1,126	58	,265
Equal variances not assumed			-1,126	57,996	,265

T-Test for Independent Samples (Groups)

Input file:

Variables (groups):

First list: SAMPLE1
Second list: SAMPLE2

I-tests

Box & whisker plot

Options

Display long variable names

t-test with separate variance estimates

Levene's test (homogeneity of variances)

Brown & Forsythe (homog. of variances)

It is assumed that each variable contains the data for one group. Each variable (group) in the first list will be compared to each variable (group) in the second list.

	1 SAMPLE1	2 SAMPLE2
1	61,000	59,000
2	80,000	56,000
3	40,000	85,000
4	71,000	92,000
5	39,000	48,000
6	20,000	59,000
7	40,000	42,000
8	80,000	87,000
9	80,000	42,000
10	60,000	73,000
11	80,000	56,000
12	29,000	37,000
13	51,000	56,000

Рис. 6.2. Использование *t*-критерия для независимых выборок в пакете STATISTICA

было сделано для работы в пакете SPSS, так и в исходном, когда каждая переменная содержит измерения по соответствующей выборке (*Each variable contains the data for one group*). В режиме *Variables (groups)* в левом и правом окнах выбирается по одной переменной для анализа. Дополнительно можно включить в выходную форму результаты тестов на однородность дисперсий выборок: тест Левена (*Levene's test*) и тест Брауна–Форсайта (*Brown & Forsythe*). Интерпретация результатов теста Брауна–Форсайта полностью аналогична интерпретации результатов теста Левена. Выходная форма с результатами этих тестов

и статистикой t -критерия представляется после щелчка по кнопке T -test. Мы не приводим эту форму, чтобы не загромождать текст.

Полезной опцией для анализа результатов являются диаграммы размаха (*Box & Whisker plot*). С их помощью можно визуально оценить степень различия между двумя переменными путем сравнения средних и различных мер отклонения от среднего (стандартная ошибка среднего, стандартное отклонение, размах и др.) в выборках. Можно, например, сравнить 95%-ные доверительные интервалы по двум выборкам (рис. 6.3). Из графика также следует, что нулевая гипотеза не может быть отклонена на данном уровне значимости (область пересечения доверительных интервалов велика).

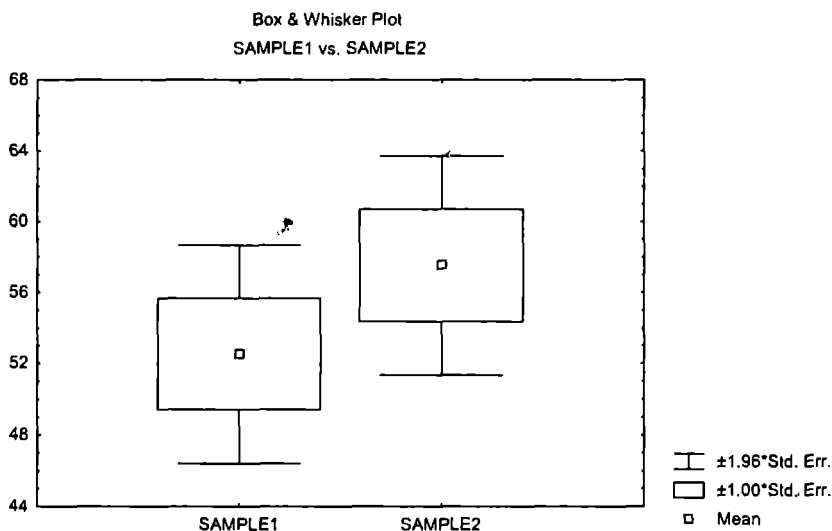


Рис. 6.3. Доверительные интервалы для двух выборок

На этом графике указаны и 68%-ные доверительные интервалы (определяемые одной стандартной ошибкой среднего), пересечение которых убедительно свидетельствует об отсутствии статистически значимых различий.

6.3. Применение t -критерия Стьюдента для связанных выборок

Для связанных (зависимых) выборок подход к определению эмпирического значения t -критерия незначительно отличается от случая независимых выборок. Допущение зависимости выборок чаще всего означает, что признак измерен на одной и той же выборке дважды, например до воздействия и после него (состояние пациентов до и после лечения; производительность труда рабочих до и после отпуска и т.п.). Применение t -критерия для зависимых выборок может также оказаться полезным при сравнении результатов предварительного и повторного тестирования для оценки надежности измерений (метод тест-ретест).

Исходные данные должны быть таковы, чтобы имелось по два значения изучаемого признака для каждого объекта (при повторных измерениях) или для каждой пары связанных (зависимых) объектов. Очевидно, что измерения по обеим выборкам должны быть содержательно сопоставимы, т.е. должны использоваться аналогичные методы получения и обработки данных и одинаковые единицы измерения.

Для расчета t -критерия для зависимых выборок рассматриваются разности между двумя измерениями (до и после) для каждого элемента выборки (респондента). Таким образом, в формуле t -критерия для зависимых выборок основным элементом является разность значений для каждой из n пар объектов: $d_i = x_{1i} - x_{2i}$. Соответственно определяются среднее значение (M_d) и стандартное отклонение (s_d) этих разностей, после чего используется формула

$$t = \frac{|M_d|}{s_d / \sqrt{n}}, \quad df = n - 1,$$

где $M_d = \frac{\sum d_i}{n}$;

$$s_d = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n - 1}}.$$

Для иллюстрации этого подхода используем данные по объемам продаж 10 американских компаний, случайно выбранных из списка журнала «Forbes Super 500» в 1988 и 1989 гг. (табл. 6.6) [39].

Таблица 6.6

Объемы продаж, млн долл.

Компания	Продажи в 1989 г.	Продажи в 1988 г.
General Motors	126932	123505
General Electric	54574	49662
Exxon	86656	78944
IBM	62710	59512
Ford	96146	92300
AT&T	36112	35173
Mobil	50220	48111
DuPont	35099	32427
Sears	53794	49975
Amoco	23966	20779

Есть ли статистически значимые различия в объемах продаж ведущих компаний за год? Для ответа на этот вопрос можно применить t -тест для двух зависимых выборок, поскольку имеются результаты повторных измерений для каждой компании, а данные представлены в метрической шкале.

Для анализа используем соответствующую процедуру пакета SPSS (*Analyze, Compare Means, Paired Samples T Test*). В данном случае группировать переменные не нужно, достаточно указать две переменные, по которым проводится сравнение, в поле *Paired Variables*.

В итоге мы получаем (табл. 6.7), что средние объемы продаж (62620,9 и 59038,8 млн долл.) статистически значимо различаются ($p < 0,001$). Поэтому нулевая гипотеза отклоняется, что позволяет сделать вывод о росте торгового оборота ведущих американских компаний за анализируемый период.

Процедура вычисления статистики t -критерия для зависимых выборок в пакете STATISTICA (*t-test for dependent samples*) аналогична процедуре вычисления статистик этого критерия для независимых выборок. В результате выполнения данной процедуры мы получаем такие же эмпирические значения t -критерия (рис. 6.4).

Статистика использования t -критерия для связанных выборок
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	S1989	62620,9000	10	31777,64941	10048,97508
	S1988	59038,8000	10	31072,87052	9826,10443

Paired Samples Test

		Paired Differences			t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Er- ror Mean			
S1989	S1988	3582,1000	1803,1592	570,2090	6,282	9	,000

T-test for Dependent Samples (new-2.sta)								
Marked differences are significant at p < ,05000								
Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
SAMPLE1	62620,90	31777,65						
SAMPLE2	59038,80	31072,87		1803,159	6,282062			,000144

Рис. 6.4. Статистика t -критерия для связанных выборок в программе STATISTICA

Предположения и ограничения на использование t -критерия для зависимых выборок аналогичны статистической модели t -критерия для независимых выборок. Поскольку предполагается, что распределение исследуемого признака существенно не отличается от нормального, то в данном случае это означает, что попарные разности тоже должны быть нормально распределены. Если эти условия не выполняются, то можно воспользоваться одним из альтернативных непараметрических критериев, например T -критерием Уилкоксона (гл. 8).

Применение t -критерия для зависимых выборок также предполагает, что данные двух измерений положительно коррелируют (это свидетельствует о связанности выборок). Если данные для двух выборок не коррелируют положительно, можно использовать t -критерий Стьюдента для независимых выборок.

Закономерен вопрос о том, как будут различаться результаты проверки при замене t -критерия для зависимых выборок соответствующим критерием для независимых выборок. В случае

зависимых выборочных совокупностей допустимо применение критерия Стьюдента для независимых выборок, но не наоборот. Причем это рекомендуется осуществлять, если корреляция между двумя измерениями отрицательна. Если корреляция положительна, то такая замена приведет к существенной недооценке достоверности различий. Это связано с тем, что в критерии для зависимых выборок парное вычитание значений по двум выборкам и последующий анализ этих разностей позволяют исключить ту часть вариации, которая является результатом различия в исходных уровнях респондентов. Поэтому по сравнению с t -критерием для независимых выборок критерий Стьюдента для зависимых выборок является гораздо более чувствительным. Действительно, если в рассмотренном примере (см. табл. 6.6) к двум выборкам, характеризующим объемы продаж, применить t -критерий для независимых выборок, нулевая гипотеза не может быть отклонена ($p = 0,802$), т.е. статистически значимые различия не обнаруживаются. Читателю рекомендуется самостоятельно проверить этот результат. Данный пример еще раз иллюстрирует необходимость корректного использования статистических методов.

Задания для самостоятельного выполнения

1. Для каждой ситуации определите t критическое.

Альфа	Вид альтернативы	df	$t_{\text{крит}}$
0,10	Односторонняя	20	
0,02	Двусторонняя	30	
0,01	Двусторонняя	30	
0,01	Односторонняя	25	
0,05	Двусторонняя	15	

2. Студенты университета посещают клубы в среднем 3,3 раза в месяц. Результаты случайной выборки, состоящей из 97 студентов экономического факультета, показывают, что студенты посещают клубы в среднем 3,8 раза в месяц со стандартным отклонением 0,53. Можно ли сделать вывод, что студенты-экономисты отличаются от студентов университета в целом в пристрастиях к посещению клубов?

Если сформулировать вопрос исследования по-другому: «Посещают ли студенты-экономисты большее количество вечеринок, чем студенты университета в целом?», что изменится в процедуре проверки и в результатах?

3. Средняя заработная плата 45 работников филиала компании 24375 руб./месяц со стандартным отклонением 523 руб. Средняя заработная плата всех работников этой компании составляет 22230 руб./мес. Является ли это различие статистически значимым? Можно ли сказать, что работники филиала получают больше?

4. Ниже приведены данные двух выборок о среднем количестве часов в неделю, которое студенты посвящают внеурочной деятельности.

Иногородные студенты	Городские студенты
$\bar{x} = 12,4$	$\bar{x} = 10,2$
$s = 2,0$	$s = 1,9$
$n = 48$	$n = 33$

Можно ли сказать, что иногородные студенты, живущие в общежитиях, принимают большее участие в жизни университета, чем городские студенты? Является ли это различие значимым?

5. Для выявления затрат времени на обработку деталей рабочими двух разных квалификационных разрядов на предприятии было проведено выборочное исследование, показавшее, что среднее время на выполнение одной стандартной производственной операции специалистами 4-го разряда составляет 11 мин. со стандартным отклонением 1 мин. (численность выборки – 7 чел.); среднее время на выполнение этой же операции специалистами 3-го разряда – 14 мин. со стандартным отклонением 4 мин. (численность выборки – 12 чел.). Можно ли считать, что средняя производительность труда значимо различается у рабочих, имеющих разные квалификационные разряды?

6. По данным управления внутренних дел области, раскрывается 55% случаев грабежа и 62% случаев разбоя. В областном центре зафиксировано 207 случаев грабежа и 178 случаев разбоя в течение года. Было раскрыто 103 случая грабежа и 90 случаев разбоя. Можно ли сказать, что между раскрываемостью преступлений по области и областному центру имеют место значимые различия?

7. Специалист в области финансового консалтинга анализирует изменение ситуации на национальном рынке услуг связи за определенный период времени. Ему известны данные о цене акций 11 компаний, работающих на этом рынке, на соответ-

ствующие даты (табл. 6.8). Выборка компаний осуществлена случайным образом.

Таблица 6.8

Цена за одну акцию

Номер компании	Дата1	Дата2	Номер компании	Дата1	Дата2
1	218,9	361,0	7	105,0	124,9
2	79,4	82,3	8	446,8	465,6
3	99,1	125,2	9	120,7	122,6
4	49,0	84,7	10	128,2	117,7
5	171,5	157,8	11	124,9	148,8
6	224,8	203,2	—	—	—

Проверьте, увеличилась ли средняя цена акций компаний за анализируемый период. Примите $\alpha = 0,05$.

Пояснение. Для выполнения задания может быть применен однонаправленный t -критерий для зависимых выборок, поскольку данные представлены в шкале отношений, а выборки на две даты являются связанными. Критическое значение t -критерия: 1,81 ($df = 10$; $\alpha = 0,05$); эмпирическое значение t -критерия: 1,53. Нулевая гипотеза не может быть отклонена. Следовательно, статистически значимые различия в ценах акций компаний, работающих на рынке услуг связи, за анализируемый период не обнаруживаются.

Глава 7

Дисперсионный анализ

7.1. Особенности применения дисперсионного анализа

Главные идеи дисперсионного анализа (*Analysis of variance* – ANOVA) были разработаны и предложены в 1920 г. Р. Фишером (R. Fisher), английским ученым, основоположником математической генетики, внесшим значительный вклад в разработку теории математической статистики и ее приложений для анализа экспериментальных исследований. В определенном смысле ANOVA может рассматриваться как развитие t -критерия Стью-

дента. Но в отличие от критерия Стьюдента, который применим, когда независимая переменная имеет только две категории (вспомните процедуру группирования данных для обработки в SPSS), дисперсионный анализ используется для независимых переменных, имеющих более чем две категории. Фактически в случае сравнения двух групп дисперсионный анализ дает результаты, идентичные t -критерию. Но при сложных исследовательских планах следует использовать дисперсионный анализ, а не последовательность t -тестов.

Дисперсионный анализ применяется, когда зависимая переменная измеряется в метрической шкале. Независимые переменные (факторы) могут измеряться в неметрических шкалах и представляться в нескольких градациях (уровнях).

В зависимости от количества факторов, влияние которых на зависимую переменную анализируется, выделяют одно- (*one-way ANOVA*), двух- (*two-way ANOVA*) и трехфакторный дисперсионный анализ (*three-way ANOVA*) и т.д. Многофакторный дисперсионный анализ позволяет проверить гипотезы не только о влиянии каждого фактора по отдельности, но и об их взаимодействии, когда влияние одного фактора на зависимую переменную проявляется по-разному при разных уровнях других факторов. Иными словами, многофакторный дисперсионный анализ позволяет проследить, насколько влияние одного фактора зависит от градаций других факторов. Это свойство делает многофакторный метод уникальным и часто незаменимым.

Процедура использования дисперсионного анализа значительно сложнее, чем применение рассмотренных выше методов, в силу того, что ANOVA предполагает проведение большого количества довольно громоздких вычислений. Большинство существующих статистических пакетов (SPSS, SAS, STATISTICA, MICROCASE, NCSS, STATA и др.) включают процедуры ANOVA. Но работа с этими процедурами и, главное, правильная интерпретация результатов требуют ясного понимания того, как этот метод работает и что означают основные показатели, получаемые в результате его использования. Поэтому ниже приводятся много простых примеров и решений для иллюстрации «внутренних» механизмов работы метода, а затем эти результаты сравниваются с теми, которые получены на основе применения статистических пакетов.

7.2. Однофакторный дисперсионный анализ

Метод ANOVA основан на соотношении межгрупповой и внутригрупповой изменчивости изучаемой зависимой переменной, т.е. на сравнении величины изменчивости между категориями с величиной изменчивости внутри категорий. Чем больше различия между категориями по отношению к различиям внутри категорий, тем больше вероятность того, что нулевая гипотеза об *отсутствии различий* является ложной и может быть отклонена.

Пример 7.1. Предположим, необходимо выяснить, зависит ли время выполнения определенного задания (операции 1) работниками предприятия от возможных методов выполнения этой операции. Случайным образом были сформированы четыре группы работников, в каждой из которых демонстрировался один из четырех методов. Результаты представлены в табл. 7.1.

Таблица 7.1

Время выполнения технической операции 1, мин.

	Метод 1	Метод 2	Метод 3	Метод 4
Среднее значение	11,2	10,1	10,5	9,5
Стандартное отклонение	2,7	1,5	2,3	1,7

Средние значения и стандартные отклонения по разным методам отличаются незначительно. Вероятность нулевой гипотезы об отсутствии различий между методами достаточно велика.

Пример 7.2. При исследовании затрат времени на выполнение другой операции (2) получены следующие результаты (табл. 7.2).

Таблица 7.2

Время выполнения технической операции 2, мин.

	Метод 1	Метод 2	Метод 3	Метод 4
Среднее значение	16,2	12,1	9,5	6,5
Стандартное отклонение	1,1	0,9	0,8	0,6

В этом случае средние значения заметно различаются, а стандартные отклонения невелики и сравнительно одинаковы, что указывает на небольшую изменчивость внутри каждой категории. Такие результаты с большой вероятностью позволяют от-

клонить нулевую гипотезу и указывают на действительные различия эффективности методов выполнения задания.

Нулевая гипотеза, проверяемая методом ANOVA, утверждает равенство *всех* средних значений (для *всех* изучаемых генеральных совокупностей). Для приведенного выше примера нулевая гипотеза формулируется в виде $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$. При отклонении нулевой гипотезы принимается альтернативная гипотеза о том, что по крайней мере два средних значения различаются (среднее по крайней мере для одной генеральной совокупности отличается от остальных). Часто альтернативная гипотеза формулируется как $H_A: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$. Такая формулировка не означает, что *все* средние значимо отличаются друг от друга. Дело в том, что если нулевая гипотеза отклоняется, то без дополнительных методов все равно невозможно определить, какое именно среднее (или средние) значимо отличается от остальных. Для этого используются критерии множественных сравнений, рассматриваемые в следующем подразделе.

Изучаемый признак (зависимая переменная) измеряется для объектов, каждый из которых принадлежит к *одной* из нескольких сравниваемых выборок. Для принятия решения на основе дисперсионного анализа необходимо вычислить следующие величины.

Общая сумма квадратов, или SS_T (*total sum of squares*):

$$SS_T = \sum_{i=1}^n (x_i - \bar{x})^2$$

Эта величина складывается из двух компонентов:

$$SS_T = SS_B + SS_W.$$

Первый компонент характеризует случайную изменчивость внутри категорий и называется внутригрупповой суммой квадратов SS_W (*sum of squares within groups*). Второй компонент характеризует изменчивость между категориями и называется межгрупповой суммой квадратов SS_B (*sum of squares between groups*).

Внутригрупповая сумма квадратов определяется как

$$SS_W = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2,$$

где \bar{x}_j – среднее значение признака по j -й группе (категории);
 n_j – количество наблюдений в j -й категории;
 x_{ij} – i -й элемент j -й группы;
 k – количество категорий (групп).

Межгрупповая сумма квадратов определяется как

$$SS_B = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$$

Отношение межгрупповой и общей сумм квадратов определяет долю общей дисперсии зависимой переменной, обусловленную влиянием исследуемого фактора. Этот показатель имеет тот же смысл, что и коэффициент детерминации, т.е. квадрат коэффициента корреляции

$$r^2 = \frac{SS_B}{SS_T}.$$

Если нулевая гипотеза верна (различий между категориями нет), то изменчивость от категории к категории не должна быть заметно больше изменчивости внутри категорий, т.е. SS_W и SS_B должны быть примерно равны. Чем больше различие между этими двумя оценками, тем с большей вероятностью нулевая гипотеза должна быть отклонена. Если средние значения в группах приблизительно одинаковы, то тогда различия не будут значимыми. Чем больше различие между средними значениями категорий и чем более однородны категории, тем больше вероятность, что различия статистически значимы.

Критерием принятия решения в методе дисперсионного анализа является *F-отношение (критерий Фишера)*:

$$F = \frac{MS_B}{MS_W}, \quad df_B = k - 1, \quad df_W = n - k,$$

где MS_B – межгрупповой средний квадрат (*mean square between*):

$$MS_B = \frac{SS_B}{df_B};$$

MS_W – внутригрупповой средний квадрат (*mean square within*):

$$MS_W = \frac{SS_W}{df_W}.$$

Величины df_B и df_W являются числом степеней свободы для соответствующих сумм квадратов (SS_B и SS_W) и рассчитываются следующим образом: число степеней свободы для межгрупповой суммы квадратов $df_B = k - 1$; число степеней свободы для внутригрупповой суммы квадратов $df_W = n - k$.

Таким образом, F -отношение показывает, во сколько раз величина изменчивости между категориями превышает величину изменчивости внутри категорий. Чем больше изменчивость между категориями по сравнению с изменчивостью внутри категорий, тем больше будет значение критерия Фишера и тем больше вероятность того, что нулевая гипотеза будет отклонена.

Для определения критических значений F -распределения (при соответствующих числах степеней свободы df_B и df_W и уровне значимости) можно воспользоваться специальными статистическими таблицами (табл. П.3) или рассчитать их с помощью функции ФРАСПОБР(α , df_B , df_W). Если проверка гипотезы на основе ANOVA осуществляется без программных продуктов, реализующих соответствующую процедуру, то для упрощения расчетов можно использовать другую, более удобную для вычислений формулу для общей суммы квадратов:

$$SS_T = \sum_{i=1}^n x_i^2 - n(\bar{x})^2$$

После определения SS_B значение SS_W находится по формуле

$$SS_W = SS_T - SS_B.$$

Пример 7.3. Сформированы четыре выборки по 4 чел. в каждой и получены следующие результаты по выполнению задания (табл. 7.3). Использование метода дисперсионного анализа для такой группы, как будет показано ниже, не совсем корректно из-за малой численности

выборок, но в данном случае это сделано исключительно для иллюстрации работы метода, чтобы не утруждать читателя повторением излишне громоздких расчетов.

Таблица 7.3

Среднее время выполнения задания, мин.

№ п/п	Метод 1	Метод 2	Метод 3	Метод 4
1	10	25	14	16
2	13	19	16	17
3	15	20	16	23
4	16	17	24	15

Проверим гипотезу о том, что время на выполнение задания не зависит от метода (примем уровень значимости $\alpha = 0,05$).

H_0 : $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (среднее время выполнения задания разными методами статистически значимо не различается).

H_A : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ (среднее время выполнения задания статистически значимо отличается по крайней мере для одной выборки).

Таблица 7.4

Расчет F -отношения

№ п/п Метод	x_i				x_i^2			
	1	2	3	4	1	2	3	4
1	10	25	14	16	100	625	196	256
2	13	19	16	17	169	361	256	289
3	15	20	16	23	225	400	256	529
4	16	17	24	15	256	289	576	225
Сумма	54	81	70	71	750	1675	1284	1299
\bar{x}	17,25							
\bar{x}_k	13,5	20,25	17,5	17,75				

Сделаем необходимые вычисления для определения F -отношения (табл. 7.4):

$$SS_T = (750 + 1675 + 1284 + 1299) - 16(17,25)^2 = 247;$$

$$SS_B = 4(13,5 - 17,25)^2 + 4(20,25 - 17,25)^2 + 4(17,5 - 17,25)^2 + 4(17,75 - 17,25)^2 = 93,5;$$

$$SS_W = 247 - 93,5 = 153,5;$$

$$df_W = 16 - 4 = 12;$$

$$df_B = 4 - 1 = 3;$$

$$MS_B = 93,5/3 = 31,17;$$

$$MS_W = 53,5/12 = 12,79;$$

$$F_{\text{эксп}} = 31,17/12,79 = 2,44;$$

$$F_{\text{крит}} = F_{\text{РАСПОБР}}(0,05;3;12) = 3,49.$$

Сравнение $F_{\text{эксп}}$ и $F_{\text{крит}}$ ($F_{\text{эксп}} < F_{\text{крит}}$) не позволяет отклонить нулевую гипотезу, поэтому необходимо сделать вывод о том, что наблюдаемые различия выборочных средних не являются статистически значимыми и, следовательно, в данном случае производительность труда при применении разных методов статистически значимо не различается.

Статистической моделью при проверке гипотез методом ANOVA являются предположения о том, что выборки случайные и независимые, зависимая переменная измеряется в интервальной шкале или шкале отношений, распределение зависимой переменной в сравниваемых генеральных совокупностях характеризуется нормальным законом и равными дисперсиями. (Последнее предположение называется предположением об *однородности дисперсий*.) Формально численность выборок не должна быть меньше двух объектов, но рекомендуется иметь не менее пяти объектов в каждой выборке.

В теории статистики показано, что F -критерий довольно устойчив к небольшим отклонениям от статистической модели, в частности к отклонению от нормальности. Если число наблюдений достаточно большое, то отклонение от нормальности не имеет особого значения в силу *центральной предельной теоремы*, в соответствии с которой распределение среднего значения при большом объеме выборки близко к нормальному независимо от начального распределения.

Предположение об однородности дисперсий связано с тем, что вычисление суммы квадратов ошибок производится внутри каждой группы. Если дисперсии в двух группах значительно отличаются друг от друга, то сложение их некорректно и не дает верной оценки общей внутригрупповой дисперсии. Доказано, что F -критерий также достаточно устойчив относительно нарушения предположений однородности дисперсий. Поэтому считается, что пока объемы выборок равны или приблизительно равны между собой, метод дисперсионного анализа можно использовать без дополнительной проверки однородности дисперсий выборок, иначе необходимо проверять однородность дис-

персий в сравниваемых выборках. Для этого применяется тест Левена (*Levene test of homogeneity of variances*), который реализован в SPSS, STATISTICA и других статистических пакетах.

Несмотря на то что все названные ограничения на применение метода дисперсионного анализа не являются критическими и на практике этот метод используется при небольших отклонениях от указанной статистической модели, следует помнить, что такие ограничения существуют. И это обстоятельство нужно всегда принимать во внимание как при планировании и обработке результатов собственных исследований, так и при вынесении суждений о корректности результатов, полученных другими исследователями. Если статистическая модель нарушается в значительной степени или исследователь не уверен в том, что отклонения от модели незначительны, то рекомендуется применить непараметрический аналог этого метода (критерий χ^2 или *H*-критерий Краскела–Уоллиса).

7.3. Критерии множественных сравнений

Альтернативная гипотеза, формулируемая при применении метода ANOVA и принимаемая в случае отклонения нулевой гипотезы, неконкретна. Она лишь утверждает, что среднее по крайней мере для одной генеральной совокупности отличается от остальных. Очевидно, что исследователю важно знать, по каким именно категориям различия значимы, какие из групп значимо отличаются от других. Принятие решения только по величине средних значений может привести к неверным выводам. Поэтому для корректного заключения о том, какие средние значимо отличаются от остальных, необходимо использовать дополнительные статистические критерии.

Казалось бы, что для решения этой задачи можно провести серию *t*-тестов Стьюдента и путем попарного сравнения выборок определить значимые различия для всех возможных пар средних. Однако использование *t*-критерия Стьюдента в данном случае не совсем корректно. Дело в том, что *t*-критерий проверяет, являются ли два средних значения значимо отличающимися друг от друга, в предположении, что рассматриваются всего две выборки. Многократное применение критерия Стьюдента к одной выборке в парных сравнениях увеличивает шансы случайного получения значимого результата.

Разработаны процедуры для более корректного сравнения пар средних с целью выявления статистически значимых различий. Эти процедуры называются *критериями множественных сравнений*. Некоторые из этих критериев могут применяться *после* использования ANOVA и отклонения нулевой гипотезы, поэтому они называются *апостериорными критериями (post hoc tests)*, от лат. *post hoc* – после того). Другие критерии могут применяться и *до* использования ANOVA, поэтому они называются *априорными критериями (a priori tests)*.

Пример апостериорных процедур – *тест Шеффе (Scheffe's test)*. Этот метод должен использоваться в случае, если нулевая гипотеза уже отклонена на основе применения ANOVA. Оценка значимости различий основана на вычислении *t-критерия Шеффе*, являющегося модификацией *t-критерия* Стьюдента:

$$t_s = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{MS_W}{n_1} + \frac{MS_W}{n_2}}},$$

где \bar{x}_1, \bar{x}_2 – сравниваемые средние значения для двух анализируемых выборок;

MS_W – внутригрупповой средний квадрат;

n_1, n_2 – объемы соответствующих выборок.

Критическое значение *t-критерия Шеффе* определяется как $t'_s = \sqrt{F'(k-1)}$, где F' – критическое значение *F-критерия*, определяемое для заданного уровня значимости α и соответствующих чисел степеней свободы df_B и df_W . Тест Шеффе имеет те же допущения по использованию, что и ANOVA.

Апостериорные критерии особенно полезны в случаях, когда дисперсионный анализ обнаружил различия, которые являются неожиданными для исследователя, поэтому требуется определить, по каким именно факторам эти различия наблюдаются. Апостериорные критерии позволяют надежно идентифицировать значимые различия между парами выборочных средних.

Если исследователь заранее (до осуществления процедуры проверки методом ANOVA) решает, что необходимо проводить множественную проверку выборок, то можно воспользоваться методом *априорных контрастов (a priori contrasts)*.

Контрастом называется сумма

$$C = \sum_{j=1}^k c_j \bar{x}_j,$$

где k – количество групп (категорий изменения фактора);

c_j – коэффициенты контраста, такие, что $\sum_{j=1}^k c_j = 0$;

\bar{x}_j – средние значения исследуемой зависимой переменной по группам (категориям изменения фактора).

Если, например, фактор имеет три градации, то коэффициентами контраста могут быть: $c_1 = 2, c_2 = -1, c_3 = -1$ или $c_1 = 1, c_2 = -0,5, c_3 = -0,5$. В случае игнорирования какой-либо категории соответствующему коэффициенту присваивается нулевое значение. Несмотря на возможность выбора различных сочетаний значений коэффициентов контраста, эмпирическое значение критерия, на основе которого проверяется гипотеза, не изменяется.

Существуют две модификации этого метода. Одна из них (реализованная в пакете SPSS) основана на использовании статистики t -критерия Стьюдента. Эмпирическое значение t -критерия определяется по формуле

$$t_c = \frac{C}{\sqrt{MS_W \sum_{j=1}^k \frac{c_j^2}{n_j}}}.$$

При изменении (ином выборе) коэффициентов контраста величина t_c изменяться не будет. Найденное значение сравнивается с критическим значением двустороннего t -распределения Стьюдента для $df = df_W = n - k$.

Другая модификация метода контрастов (реализованная в STATISTICA) основана на использовании статистики F -критерия Фишера. Эмпирическое значение F -критерия вычисляется по формуле

$$F_c = \frac{C^2}{MS_W \sum_{j=1}^k \frac{c_j^2}{n_j}}$$

и сравнивается с критическим значением Фишера для $df_B = k - 1$ и $df_W = n - k$.

Проиллюстрируем использование метода контраста на рассмотренном выше примере проверки различий времени выполнения задания четырьмя методами (см. табл. 7.3). Проверим статистическую значимость различия времени выполнения задания методом 2 от результатов, полученных другими методами (это имеет смысл сделать, поскольку среднее значение по второй выборке в наибольшей степени отличается от средних значений по другим выборкам).

Выберем, например, следующие коэффициенты контраста: 0, 1, 0, -1 (сравниваются второй и четвертый методы). Значение контраста в этом случае составит 2,5. Величина MS_W была определена в предыдущем подразделе. Подстановка всех величин в формулу дает нам эмпирическое значение t -критерия: 0,99. Критическое значение $t = 2,18$ ($\alpha = 0,05$ и $df = 16 - 4 = 12$). Сравнение эмпирического и критического значений ($0,99 < 2,18$) не позволяет отклонить нулевую гипотезу, поэтому приходим к выводу о том, что среднее значение по второй выборке не отличается статистически значимо от среднего по четвертой выборке (различие в эффективности методов не обнаружено). Расчет в SPSS дает значение p -уровня 0,342.

Вычисление эмпирического значения F -критерия Фишера дает величину 0,98. Критическое значение F -критерия: 3,49 ($\alpha = 0,05$, $df_B = 3$, $df_W = 12$). Таким образом, нулевая гипотеза не может быть отклонена. Расчет в пакете STATISTICA также дает значение p -уровня 0,342.

А теперь сравним средние значения первой и второй выборок (между средними этих выборок максимальное различие). Определим коэффициенты контраста, например, как -1, 1, 0, 0. Значение контраста в этом случае составит 6,75. Эмпирическое значение t -критерия: 2,67; критическое значение $t = 2,18$; эмпирическое значение F -критерия: 7,12; критическое значение F -критерия: 3,49. Нулевая гипотеза отклоняется на уровне $\alpha = 0,05$ ($p = 0,02$).

Этот пример демонстрирует важную особенность применения априорных методов множественного сравнения. Мы видим, что в данном случае эмпирические значения критериев по двум вариантам метода контрастов превышают критические значения. Казалось бы, можно отклонить нулевую гипотезу и сделать

вывод о том, что среднее по второй выборке статистически значимо отличается от среднего по первой выборке (эффективность соответствующих методов различается). Но такой вывод будет не совсем корректным. Очень важно помнить, что мы используем априорный метод, который дает нам информацию о возможных различиях еще до применения метода дисперсионного анализа. И если в результате проведенного дисперсионного анализа нулевая гипотеза о равенстве всех средних отклоняется, т.е. подтверждается статистическая значимость различий хотя бы по одной паре, то тогда мы можем рассматривать информацию, полученную на основе метода контрастов, как достоверную. Но если в результате дисперсионного анализа нулевая гипотеза не может быть отклонена, различия, обнаруженные методом контрастов, не могут рассматриваться как статистически значимые. Поэтому целесообразнее использовать апостериорные критерии множественных сравнений, которые являются более консервативными и надежными. Несмотря на такой нюанс, метод априорных контрастов полезен тем, что его применение увеличивает вероятность обнаружения различий, если они действительно существуют.

7.4. Реализация метода однофакторного дисперсионного анализа в статистических пакетах

Рассмотрим процедуру использования методов однофакторного дисперсионного анализа и множественных сравнений в пакетах SPSS и STATISTICA. Для сравнения результатов проведем обработку данных уже рассмотренного примера. Предварительно необходимо сгруппировать данные подобно тому, как мы это делали при использовании *t*-критерия для независимых выборок. Формируется столбец из 16 элементов, соответствующий зависимой переменной (время выполнения операции), и столбец группирующей переменной (фактор), принимающей четыре значения (например, 1, 2, 3 и 4), которые соответствуют применяемому методу. Обозначим зависимую переменную как *perform* (англ. *performance* – выполнение), а группирующую (факторную) переменную – как *method* (метод). После подготовки массива данных в меню *Analyze* выбираются опция *Compare Means*, затем *One-Way Anova*. В окне *Dependent Lists* переносим зависимую переменную (*perform*), в поле *Factor* – группирующую переменную (*method*).

В окне *Contrasts* указывается, что для вычисления контраста будет использоваться линейный полином (выбрать опции *Polynomial Coefficients* и *Linear*). Затем, щелкнув по кнопке *Add*, нужно ввести коэффициенты контраста (рис. 7.1). Если предполагается вычисление нескольких контрастов, то, завершив ввод всех коэффициентов одного контраста, нужно щелкнуть по кнопке *Next* и ввести коэффициенты следующего контраста и т.д.

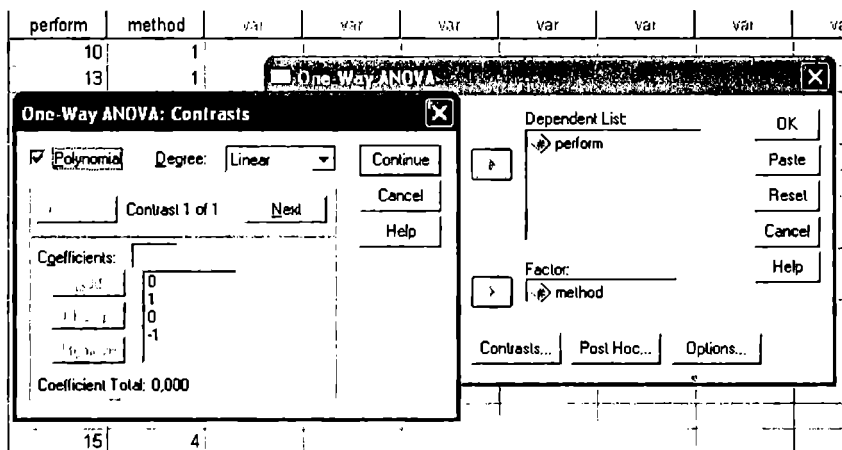


Рис. 7.1. Метод контрастов в пакете SPSS

Для выбора апостериорных критериев множественного сравнения нужно щелкнуть по кнопке *Post Hoc* в окне *One-Way Anova*. И пакет SPSS, и STATISTICA предлагают широкий выбор апостериорных критериев, позволяющих определить различия между изучаемыми группами. Наиболее часто используемыми являются тест Шеффе (*Scheffe*) при предположении равенства дисперсий и T2-тест Тэмхейна (*Tamhane's T2*) при предположении о неравенстве дисперсий выборок. В поле *Significance level* указывается уровень значимости. Переход в окно *One-Way Anova* осуществляется щелчком по кнопке *Continue* (рис. 7.2).

Для выбора параметров, которые будут выведены в отчет, необходимо щелкнуть по кнопке *Options* в окне *One-Way Anova*. В появившемся окне выбрать *Descriptive Statistics* для вывода базовых статистических показателей и *Homogeneity of variance tests* для вывода результатов теста на однородность дисперсий. При необходимости можно выбрать и другие тесты (Брауна—Форсайта и Уэлша).

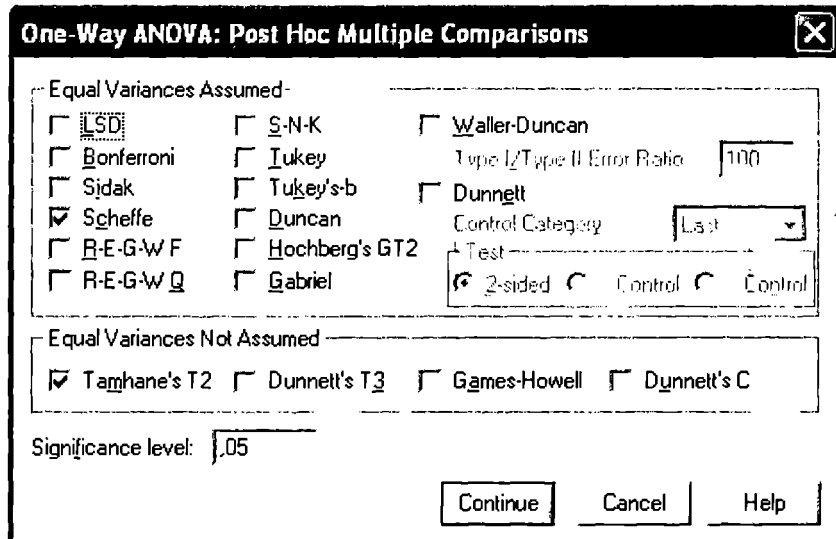


Рис. 7.2. Апостериорные критерии в SPSS

Отчеты о результатах генерируются после щелчка по кнопке *OK*. Результаты обработки данных приведены ниже. В отчете *Descriptives* (табл. 7.5) для каждой из четырех выборок указаны их объемы (*N*), средние значения (*Mean*), стандартные отклонения (*Std. Deviation*), стандартная ошибка среднего (*Std. Error*), нижняя (*Lower Bound*) и верхняя (*Upper Bound*) границы 95%-ного доверительного интервала, минимальное (*Min*) и максимальное (*Max*) значения элементов выборки.

Таблица 7.5

Отчет *Descriptives*

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
1	4	13,50	2,646	1,323	9,29	17,71	10	16
2	4	20,25	3,403	1,702	14,83	25,67	17	25
3	4	17,50	4,435	2,217	10,44	24,56	14	24
4	4	17,75	3,594	1,797	12,03	23,47	15	23
Total	16	17,25	4,058	1,014	15,09	19,41	10	25

В отчете *Test of Homogeneity of Variances* (табл. 7.6) представлены результаты проверки равенства дисперсий критерием Левена. На основе статистики критерия получен p -уровень (*Sig.*). Это значение намного выше уровня значимости $\alpha = 0,05$, поэтому нулевая гипотеза не может быть отклонена. Следовательно, предположение о равенстве дисперсий выполняется и использование метода дисперсионного анализа корректно (при условии, что все другие предпосылки для его применения также выполнены).

Таблица 7.6

Отчет *Test of Homogeneity of Variances*

Levene Statistic	df1	df2	Sig.
,294	3	12	,829

В отчете *ANOVA* (табл. 7.7) приведены результаты вычисления F -критерия (сравните значения, представленные в таблице, с результатами, полученными ранее при расчете вручную). Значение p -уровня (0,115) превышает $\alpha = 0,05$, поэтому нулевая гипотеза не может быть отклонена. Следовательно, наблюдаемые различия выборочных средних не являются статистически значимыми.

Таблица 7.7

Отчет *ANOVA*

		Sum of Squares	df	Mean Square	F	Sig.	
Between Groups	(Combined)	93,500	3	31,167	2,436	,115	
	Linear Term	Contrast	20,000	1	20,000	1,564	,235
		Deviation	73,500	2	36,750	2,873	,096
Within Groups		153,500	12	12,792			
Total		247,000	15				

Таблица 7.8

Результаты применения метода контрастов

Contrast Coefficients

Contrast	1	2	3	4
1	0	1	0	-1
2	-1	1	0	0

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Assume equal variances	1	2,50	2,529	,989	12	,342
	2	6,75	2,529	2,669	12	,020
Does not assume equal variances	1	2,50	2,475	1,010	5,982	,352
	2	6,75	2,155	3,132	5,656	0,22

В табл. 7.8 приведены результаты применения метода контрастов. В разделе *Contrast Coefficients* указаны введенные коэффициенты двух контрастов. В разделе *Contrast Tests* приведены уже знакомые нам значения статистик *t*-критерия: 0,342 для второй и четвертой выборок и 0,020 для первой и второй (мы проводили расчеты при предположении о равенстве дисперсий, а в отчете даны также значения статистик при предположении неравенства дисперсий выборок).

Таблица 7.9

Отчет *Multiple Comparisons*

	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	1	2	-6,75	2,529	,121	-14,93	1,43
		3	-4,00	2,529	,501	-12,18	4,18
		4	-4,25	2,529	,451	-12,43	3,93
	2	1	6,75	2,529	,121	-1,43	14,93
		3	2,75	2,529	,760	-5,43	10,93
		4	2,50	2,529	,807	-5,68	10,68
	3	1	4,00	2,529	,501	-4,18	12,18
		2	-2,75	2,529	,760	-10,93	5,43
		4	-,25	2,529	1,000	-8,43	7,93
	4	1	4,25	2,529	,451	-3,93	12,43
		2	-2,50	2,529	,807	-10,68	5,68
		3	,25	2,529	1,000	-7,93	8,43
Tamhane	1	2	-6,75	2,155	,125	-15,26	1,76
		3	-4,00	2,582	,703	-14,96	6,96
		4	-4,25	2,231	,502	-13,16	4,66
	2	1	6,75	2,155	,125	-1,76	15,26
		3	2,75	2,795	,935	-8,31	13,81
		4	2,50	2,475	,926	-7,03	12,03

	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
	3	1	4,00	2,582	,703	-6,96	14,96
		2	-2,75	2,795	,935	-13,81	8,31
		4	-,25	2,854	1,000	-11,43	10,93
	4	1	4,25	2,231	,502	-4,66	13,16
		2	-2,50	2,475	,926	-12,03	7,03
		3	,25	2,854	1,000	-10,93	11,43

В отчете *Multiple Comparisons* (табл. 7.9) приведены результаты парных сравнений средних значений всех выборок на основе двух апостериорных критериев множественных сравнений — Шеффе (*Scheffe*) и Тэмхейна (*Tamhane*). Здесь представлены статистики этих тестов при сравнении *i*-й выборки с *j*-й выборкой (различие средних, стандартная ошибка различия, *p*-уровень значимости, нижняя и верхняя границы доверительного интервала). Наиболее информативен столбец, представляющий *p*-уровень значимости (*Sig.*). Как следует из отчета, по всем парным сравнениям это значение превышает $\alpha = 0,05$, поэтому нулевая гипотеза о различии средних в парах выборок отклоняется, что согласуется с полученным ранее выводом о том, что наблюдаемые различия выборочных средних не являются статистически значимыми.

В пакете STATISTICA задача решается аналогично. В модуле *Basic Statistics* выбирается режим *Breakdown & One-Way ANOVA*, затем *Variables* (рис. 7.3). Здесь выбираются группирующая и зависимая переменные.

Процедура определения переменных завершается щелчком по кнопке *OK* в окнах *Select grouping variables and the dependent variables* и *Descriptive Statistics and Correlation by Groups (Breakdown)*.

В диалоговом окне *Descriptive Statistics and Correlation by Groups — Results* (рис. 7.4) выбираются результаты, которые будут представлены в соответствующих выходных формах. Могут быть рассчитаны все базовые статистики (*Summary table of means and Statistics*), статистики для расчета *F*-критерия (*Analysis of Variance*), апостериорные критерии множественных сравнений (*Post-hoc comparisons of means*), критерии Левена и Брауна—Форсайта (*Levene and Brown—Forsythe*). В правой нижней части окна можно выбрать

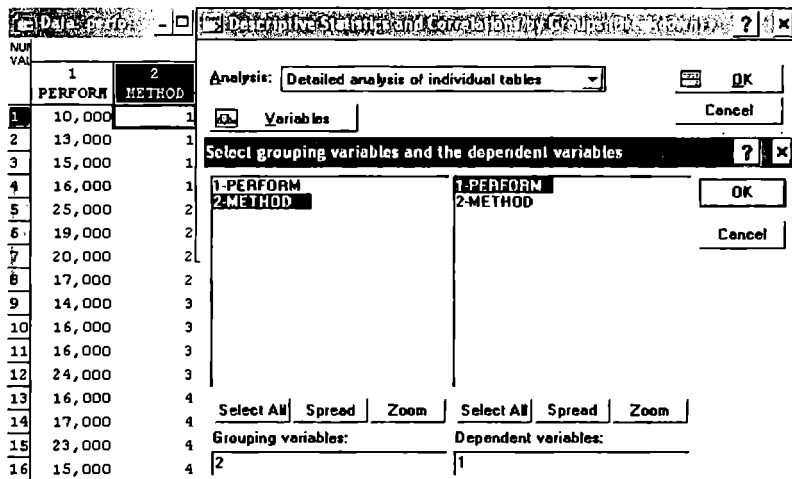


Рис. 7.3. Однофакторный дисперсионный анализ в пакете STATISTICA

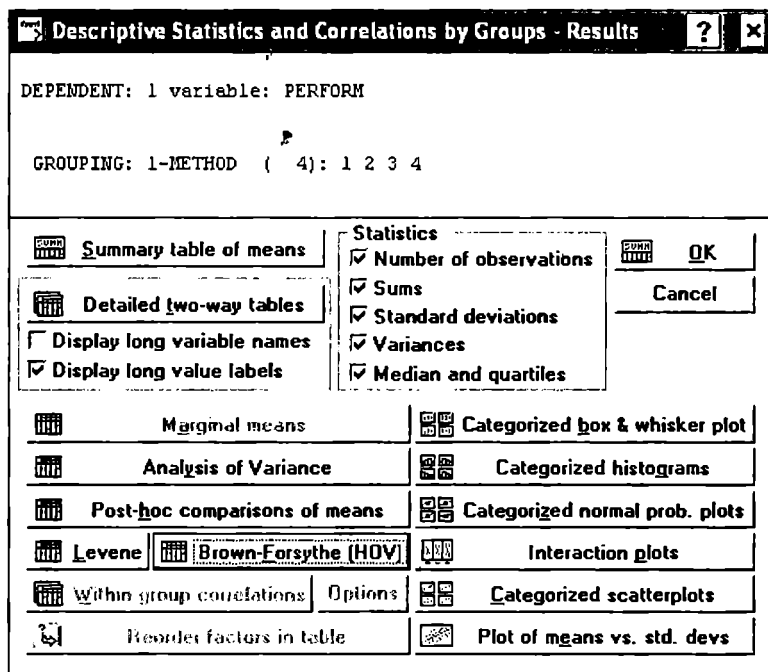


Рис. 7.4 Опции анализа ANOVA в пакете STATISTICA

различные инструменты для визуального анализа результатов (например, диаграмму размаха и др.). Мы не приводим результаты расчетов, поскольку они полностью эквивалентны результатам, полученным ранее в среде SPSS.

Более широкий набор аналитических и графических инструментов для проведения дисперсионного анализа, в том числе метод контрастов (*Planned Comparisons*), доступен в модуле ANOVA/MANOVA пакета STATISTICA. Последовательность этапов осуществления анализа в этом модуле в целом аналогична процедуре *Breakdown & One-Way ANOVA*.

Пример 7.4. Крупное маркетинговое агентство получило заказ, связанный с оценкой качества сервиса трех ведущих авиакомпаний страны, совершающих международные перелеты. Случайным образом было отобрано по 20 пассажиров каждой авиакомпании (10 пассажиров экономического класса и 10 пассажиров бизнес-класса), которые по определенной методике оценивали качество сервиса соответствующей компании (Рейтинг 1). Через месяц после исследования всем опрошенным ранее пассажирам был доставлен аналогичный опросный лист для повторной оценки, чтобы проследить изменение результатов во времени (Рейтинг 2). Результаты опросов приведены в табл. 7.10. (Категория «класс» означает класс полета: 1 – экономический, 2 – бизнес.)

Нужно определить, различаются ли статистически значимо рейтинги авиакомпаний (по результатам первой оценки). Принимается, что оценки представлены в интервальной шкале; уровень значимости – 0,05. Оценка различия сервисов авиакомпаний осуществляется без учета класса полета.

Проведем решение этой задачи в пакете SPSS (зависимая переменная – «рейтинг 1», факторная переменная – «авиакомпания»). Результаты анализа (табл. 7.11) показывают, что средние значения оценок для трех авиакомпаний (38,95; 58,90; 72,90) различаются статистически значимо. Полученное значение F -критерия (28,3) значительно превышает критическое ($p < 0,001$). Тест Левена не позволяет отклонить гипотезу о равенстве дисперсий ($p = 0,747 > \alpha = 0,05$), поэтому применение метода ANOVA корректно. Рекомендуем читателю обратить внимание на значения межгрупповой, внутригрупповой и общей суммы квадратов и соотнести их с полученным значением критерия Фишера.

Таблица 7.10

Оценки качества сервиса авиакомпаний

Класс	Рейтинг 1	Рейтинг 2	Респондент	Авиакомпания	Класс	Рейтинг 1	Рейтинг 2
1	40	36	31	2	2	52	65
1	28	28	32	2	2	70	80
1	36	30	33	2	2	73	79
1	32	28	34	2	2	72	88
1	60	40	35	2	2	73	89
1	12	14	36	2	2	71	72
1	32	26	37	2	2	55	58
1	36	30	38	2	2	68	67
1	44	38	39	2	2	81	85
1	36	35	40	2	2	78	80
2	40	42	41	3	1	92	95
2	68	49	42	3	1	56	60
2	20	24	43	3	1	64	70
2	33	35	44	3	1	72	78
2	65	40	45	3	1	48	65
2	40	36	46	3	1	52	70
2	51	29	47	3	1	64	79
2	25	24	48	3	1	68	81
2	37	23	49	3	1	76	69
2	44	41	50	3	1	56	78
1	56	67	51	3	2	88	92
1	48	58	52	3	2	79	85
1	64	78	53	3	2	92	94
1	56	68	54	3	2	88	93
1	28	69	55	3	2	73	90
1	32	74	56	3	2	68	67
1	42	55	57	3	2	81	85
1	40	55	58	3	2	95	95
1	61	80	59	3	2	68	67
1	58	78	60	3	2	78	83

Результаты анализа качества сервиса авиакомпаний

Descriptives
RATING1

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
1	20	38,95	14,006	3,132	32,40	45,50	12	68
2	20	58,90	15,089	3,374	51,84	65,96	28	81
3	20	72,90	13,902	3,108	66,39	79,41	48	95
Total	60	56,92	19,902	2,569	51,78	62,06	12	95

Test of Homogeneity of Variances
RATING1

Levene Statistic	df1	df2	Sig.
,293	2	57	,747

ANOVA
RATING1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11644,033	2	5822,017	28,304	,000
Within Groups	11724,550	57	205,694		
Total	23368,583	59			

Таким образом, нулевая гипотеза об отсутствии различий в рейтингах может быть отклонена. Поэтому можно сделать вывод: гипотеза о том, что различия между тремя агентствами вызваны только случайными отклонениями, очень маловероятна. Рейтинг авиакомпаний статистически значимо различается.

В данном случае у нас не было *теоретических* причин *a priori* (заранее) предполагать различия в уровне сервиса, предоставляемого авиакомпаниями. Поэтому используем апостериорный критерий множественных сравнений Шеффе для определения, по каким именно авиакомпаниям существуют различия. По результатам теста Шеффе (табл. 7.12) мы видим, что все попарные комбинации (по всем компаниям) значимо различаются между собой.

Результаты теста Шеффе

(I) AIRLINE	(J) AIRLINE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-19,95	4,535	,000	-31,35	-8,55
	3	-33,95	4,535	,000	-45,35	-22,55
2	1	19,95	4,535	,000	8,55	31,35
	3	-14,00	4,535	,012	-25,40	-2,60
3	1	33,95	4,535	,000	22,55	45,35
	2	14,00	4,535	,012	2,60	25,40

7.5. Многофакторный дисперсионный анализ

Многофакторный дисперсионный анализ, как уже отмечалось, позволяет оценить не только влияние на исследуемую (зависимую) переменную каждого фактора по отдельности (так называемые *главные эффекты*), но и зависимость влияния одних факторов от уровня других (*взаимодействие факторов*). Это преимущество метода ANOVA иногда дает возможность получать интересные исследовательские результаты, которых нельзя достичь другими методами.

Рассмотрим случай двух факторов, для чего вернемся к задаче об авиакомпаниях. Если в модель задачи добавить еще один фактор, например класс полета, мы получим двухфакторную модель для дисперсионного анализа. Данная модель позволяет ответить на три вопроса.

1. Есть ли статистически значимое различие в уровнях сервиса, предоставляемого тремя авиакомпаниями?

2. Есть ли статистически значимое различие в уровнях сервиса, предоставляемого при полете экономическим и бизнес-классом?

3. Влияют ли *совместно* вид авиакомпании и класс полета на оценку сервиса?

Общая логика проведения двухфакторного дисперсионного анализа принципиально не отличается от логики однофакторного анализа, за исключением того, что межгрупповая изменчивость включает также изменчивость взаимодействия факторов:

$$SS_B = SS_a + SS_b + SS_{ab},$$

где SS_a, SS_b – суммы квадратов для факторов a и b ;
 SS_{ab} – сумма квадратов для взаимодействия факторов.

Для каждого источника изменчивости вычисляются соответствующие суммы квадратов и числа степеней свободы, на основе чего определяется эмпирическое значение F -отношения, которое сравнивается с критическим значением.

На аналогичном принципе построен дисперсионный анализ, в котором количество факторов превышает два. Отличие заключается лишь в том, что помимо двухфакторных взаимодействий также анализируются взаимодействия больших порядков. Например, в случае трех факторов (a, b, c) вычисляются $SS_a, SS_b, SS_c, SS_{ab}, SS_{bc}, SS_{ac}$, а также сумма квадратов для трехфакторного взаимодействия SS_{abc} . Мы не будем приводить формулы для расчета этих сумм из-за их громоздкости и нецелесообразности использования без применения статистических пакетов.

Программные продукты, реализующие многофакторный метод ANOVA, существенно упростили проблему громоздкости вычислений, но не проблему содержательной интерпретации результатов в случае многофакторного анализа. Дело в том, что количество формулируемых и проверяемых гипотез (h) растет с увеличением числа факторов (k) в геометрической прогрессии: $h = 2^k - 1$, т.е. в двухфакторной модели проверяются три гипотезы, в трехфакторной – семь и т.д. Именно по причине сложности формирования исследовательской модели и содержательной интерпретации результатов дисперсионный анализ с количеством факторов больше двух используется редко.

Для иллюстрации процедуры применения двухфакторного дисперсионного анализа и интерпретации результатов рассмотрим двухфакторную модель задачи об авиакомпаниях (см. табл. 7.10). Для этого воспользуемся возможностями, предоставляемыми SPSS. В главном меню *Analyze* выбирается опция *General Linear Model*, затем *Univariate*. Как и в случае однофакторного анализа, определяется зависимая переменная (*Dependent Variable*). В данном случае это переменная «рейтинг 1» (*rating1*). В качестве анализируемых факторов (*Fixed Factor(s)*) выбираются переменные «авиакомпания» (*airline*) и «класс полета» (*class*). В окне *Options* рекомендуется выбрать опцию *Descriptive Statistics* для вывода

базовых статистик и *Homogeneity tests* для вывода результатов проверки однородности выборочных дисперсий, а также установить требуемый уровень значимости (по умолчанию $\alpha = 0,05$).

Для визуализации результатов в окне *Plots* в качестве горизонтальной оси графика (*Horizontal Axis*) рекомендуется указать фактор *class*, а для отображения рейтингов авиакомпаний отдельными линиями (*Separate Lines*) – фактор *airline*. После выбора факторов нужно шелкнуть по кнопкам *Add*, а затем – *Continue*.

Для использования апостериорных критериев множественных сравнений в окне *Post Hoc* из поля *Factor(s)* выбирается фактор *Airline* и переносится в поле *Post Hoc Tests for*. В качестве критерия проверки выбирается тест Шеффе. (Фактор *class* не выбирается, так как он содержит лишь две категории, а для факторов с количеством категорий меньше трех критерий множественных сравнений применять нет смысла.)

Базовая статистика по выборкам приведена в отчете *Descriptive Statistics* (табл. 7.13).

Таблица 7.13

Отчет *Descriptive Statistics* для переменной **RATING1**

AIRLINE	CLASS ^Р	Mean	Std. Deviation	N
1	1	35,60	12,140	10
	2	42,30	15,550	10
	Total	38,95	14,006	20
2	1	48,50	12,501	10
	2	69,30	9,166	10
	Total	58,90	15,089	20
3	1	64,80	13,037	10
	2	81,00	9,603	10
	Total	72,90	13,902	20
Total	1	49,63	17,165	30
	2	64,20	20,035	30
	Total	56,92	19,902	60

В отчете по тесту Левена (табл. 7.14) значение *Sig.* больше 0,05, поэтому нулевая гипотеза о равенстве дисперсий не может быть отклонена и ANOVA может применяться.

Проверка однородности дисперсий

F	df1	df2	Sig.
,712	5	54	,617

Из отчета *Tests of Between-Subjects Effects* (табл. 7.15) по значениям F -критерия и p -уровня значимости мы можем получить ответы на все три поставленных вопроса о раздельном и совместном влиянии на оценку сервиса компаний двух анализируемых факторов.

Таблица 7.15

Отчет *Tests of Between-Subjects Effects*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	15343,883	5	3068,777	20,650	,000
Intercept	194370,417	1	194370,417	1307,962	,000
AIRLINE	11644,033	2	5822,017	39,178	,000
CLASS	3182,817	1	3182,817	21,418	,000
AIRLINE * CLASS	517,033	2	258,517	1,740	,185
Error	8024,700	54	148,606		
Total	217739,000	60			
Corrected Total	23368,583	59			

R Squared = ,657 (Adjusted R Squared = ,625)

Оценим вначале эффект взаимодействия факторов, поскольку анализ воздействия главных эффектов по отдельности будет неполным, если обнаружится эффект взаимного влияния. Как следует из отчета, эффект взаимодействия не существен на уровне значимости 0,05 ($F = 1,74$; $p = 0,185$). Нулевая гипотеза относительно взаимодействия факторов (точнее, их отсутствия) не может быть отклонена. Поэтому теперь можно оценить влияние главных эффектов. Как и в однофакторной модели, нулевая гипотеза о неразличимости по компаниям отклоняется на уровне 0,001, различия по классам также значимы на уровне 0,001. В отчете приведено значение $R^2 = 0,657$ (*R Squared*), которое показывает, что изменчивость зависимой переменной на 66% определяется изменчивостью анализируемых факторов (подробнее этот показатель рассматривается в третьей части учебного пособия).

В пакете STATISTICA получаются полностью аналогичные результаты (*ANOVA/MANOVA – Variables – OK – All effects*) (рис. 7.5).

Summary of all Effects: design: (2-way anova.sta)						
Continue... 1-AIRLINE, 2-CLASS						
Effect	df Effect	SS Effect	df Error	MS Error	F	p-level
1	2	5822,017	54	148,6056	39,17765	,000000
2	1	3182,817	54	148,6056	21,41788	,000024
32	2	258,517	54	148,6056	1,73962	,185283

Рис. 7.5. Результаты двухфакторного дисперсионного анализа в STATISTICA

График (рис. 7.6) подтверждает, что между категориями «класс полета» и «авиакомпания» нет взаимодействия, иначе линии были бы не параллельными, а пересекались. Это могло бы означать, например, что по экономическому классу рейтинг одной из компаний выше рейтинга другой (других), а по бизнес-классу – ниже.

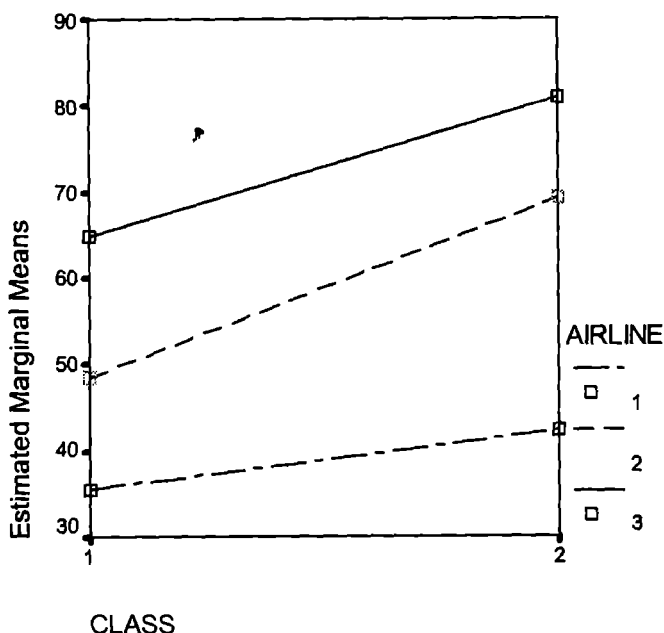


Рис. 7.6. Средние значения зависимой переменной (rating1) при разных уровнях факторных признаков (class и airline)

Отчет *Multiple Comparisons* (табл. 7.16) представляет результаты применения теста Шеффе. Эти результаты показывают статистически значимое различие среднего рейтинга всех компаний на уровне $p < 0,05$.

Таблица 7.16

Отчет *Multiple Comparisons*

(I) AIRLINE	(J) AIRLINE	Mean Dif- ference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-19,95	3,855	,000	-29,65	-10,25
	3	-33,95	3,855	,000	-43,65	-24,25
2	1	19,95	3,855	,000	10,25	29,65
	3	-14,00	3,855	,003	-23,70	-4,30
3	1	33,95	3,855	,000	24,25	43,65
	2	14,00	3,855	,003	4,30	23,70

7.6. Дисперсионный анализ с повторными измерениями

В случае, когда осуществляются повторные измерения (хотя бы по одному фактору) по одной и той же группе объектов (т.е. выборки связанные), целесообразно применять специальную модификацию ANOVA – *дисперсионный анализ с повторными измерениями (Repeated Measures ANOVA)*. В принципе в данной ситуации можно использовать и рассмотренный выше вариант многофакторного ANOVA для независимых выборок, но метод с повторными измерениями более чувствителен к влиянию изучаемых факторов на зависимую переменную, поэтому с большей надежностью обнаруживает эффекты такого воздействия. В остальном процедура проверки и анализа результатов аналогична рассмотренным выше.

Дисперсионный анализ с повторными измерениями можно рассматривать как обобщение *t*-критерия Стьюдента для зависимых выборок, позволяющее сделать анализ не только корректным, но и более чувствительным и гибким. Например, можно одновременно контролировать влияние нескольких факторов.

Для примера воспользуемся данными повторной оценки сервиса трех авиакомпаний (в данном случае фактор «класс полета»

не будем принимать во внимание). В SPSS в меню *Analyze* выбирается *General Linear Model*, затем *Repeated measures* (рис. 7.7). В окне *Repeated measures Define Factor(s)* определяем переменные с повторными измерениями. Для этого задается имя фактора, включающего переменные с повторными измерениями. Имя фактора выбирается произвольно. Для нашего примера возьмем имя «*ratings*». В поле *Number of levels* указывается количество уровней (повторных измерений). Для данного случая – два. После щелчка по кнопкам *Add* и *Define* осуществляется переход в окно *Repeated measures*. В левом поле из списка выбираются переменные с повторными измерениями (в данном случае – «*rating1*» и «*rating2*»), затем они переносятся в поле *Within-Subject Variables*.

В поле *Between Subject Factor* переносится анализируемый фактор (в данном случае – переменная «*airline*»). В окне *Options* отмечаются опции *Descriptive Statistics* и *Homogeneity tests*, а также указывается требуемый уровень значимости. В окне *Plots* для визуализации результатов в качестве оси абсцисс (*Horizontal Axis*) в данном примере указан фактор «*airline*», а в поле *Separate Lines* – новый фактор («*ratings*»), объединяющий рейтинги авиакомпаний. Далее щелкаем по кнопкам *Add*, *Continue*, *OK* и получаем выходную форму с несколькими отчетами (табл. 7.17).

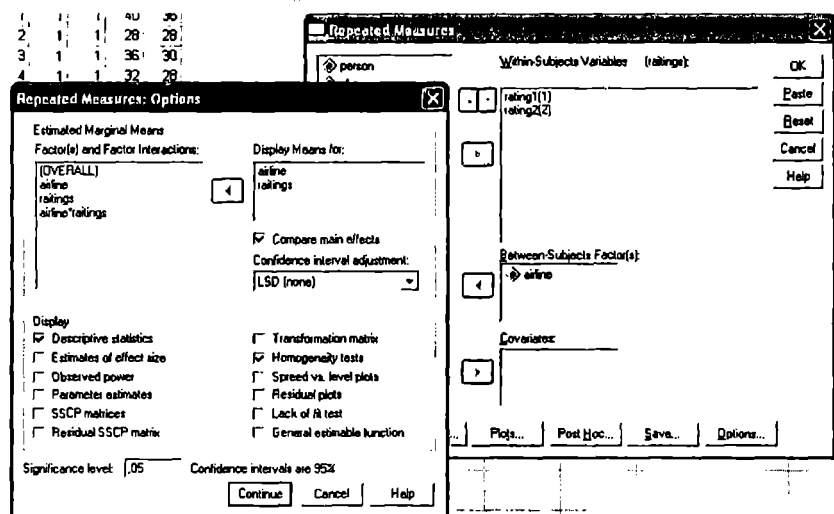


Рис. 7.7. ANOVA с повторными измерениями

В отчете *Descriptive Statistics* приведены средние значения и стандартные отклонения для трех компаний для первого и повторного измерений.

Тест Левена демонстрирует отсутствие статистически достоверных различий дисперсий зависимой переменной по факторным группам для двух измерений (первый и повторный опросы). Тест Бокса эквивалентности ковариационных матриц (*Test of Equality of Covariance Matrices*) также важен с точки зрения корректности применения используемого метода. Как и в тесте Левена, если значение *Sig.* превышает установленный уровень статистической значимости (например, 0,05), то применение данного подхода является корректным.

Таблица 7.17

Результаты ANOVA с повторными измерениями

Descriptive Statistics

	AIRLINE	Mean	Std. Deviation	N
RATING1	1	38,95	14,006	20
	2	58,90	15,089	20
	3	72,90	13,902	20
	Total	56,92	19,902	60
RATING2	1	32,40	8,268	20
	2	72,25	10,572	20
	3	79,80	11,265	20
	Total	61,48	23,208	60

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
RATING1	,293	2	57	,747
RATING2	1,472	2	57	,238

Box's Test of Equality of Covariance Matrices

Box's M	7,500
F	1,187
df1	6
df2	80975,077
Sig.	,310

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
RATINGS	Pillai's Trace	,201	14,318	1,000	57,000	,000
	Wilks' Lambda	,799	14,318	1,000	57,000	,000
	Hotelling's Trace	,251	14,318	1,000	57,000	,000
	Roy's Largest Root	,251	14,318	1,000	57,000	,000
RATINGS * AIRLINE	Pillai's Trace	,453	23,592	2,000	57,000	,000
	Wilks' Lambda	,547	23,592	2,000	57,000	,000
	Hotelling's Trace	,828	23,592	2,000	57,000	,000
	Roy's Largest Root	,828	23,592	2,000	57,000	,000

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	420556,800	1	420556,800	1590,941	,000
AIRLINE	35527,550	2	17763,775	67,199	,000
Error	15067,650	57	264,345		

Для содержательной интерпретации результатов наиболее важным является отчет *Multivariate Tests*. В этом отчете приводятся результаты многомерного тестирования данных по нескольким критериям. Из отчета следует, что статистически значимо ($Sig. < 0,001$) различаются оценки, сдвинутые во времени (раздел табл. 7.17 RATINGS). Тем самым доказывається, что фактор времени в значительной степени обуславливает различия в оценках сервиса авиакомпаний. Вторая половина этого отчета (раздел табл. 7.17 RATINGS * AIRLINE) демонстрирует, что взаимодействие факторов «время измерения» и «авиакомпания» также достоверно на высоком уровне статистической значимости. Другими словами, сочетание эффекта времени и авиакомпании, сервис которой оценивается, значительно определяет различия в рейтингах.

Главный эффект влияния фактора «airline» на оценки сервиса компаний можно оценить из отчета *Tests of Between-Subjects Effects*, показывающего, что и этот фактор является значимым источником различий в оценках ($F = 67,199; p < 0,001$).

Таким образом, получено, что оценки услуг значимо зависят от компаний; временной интервал оказывает значимое влияние на оценки; сочетание временного интервала и авиакомпании также оказывает значимое влияние на оценки.

Полученные результаты иллюстрирует график средних значений (рис. 7.8). Прямая 1 соответствует первоначальным измерениям, прямая 2 – повторным. Видно, что оценки услуг компаний отличаются во времени. Оценки второй и третьей компаний по прошествии времени улучшились, а оценки первой – ухудшились, что и отражается пересечением прямых на графике.

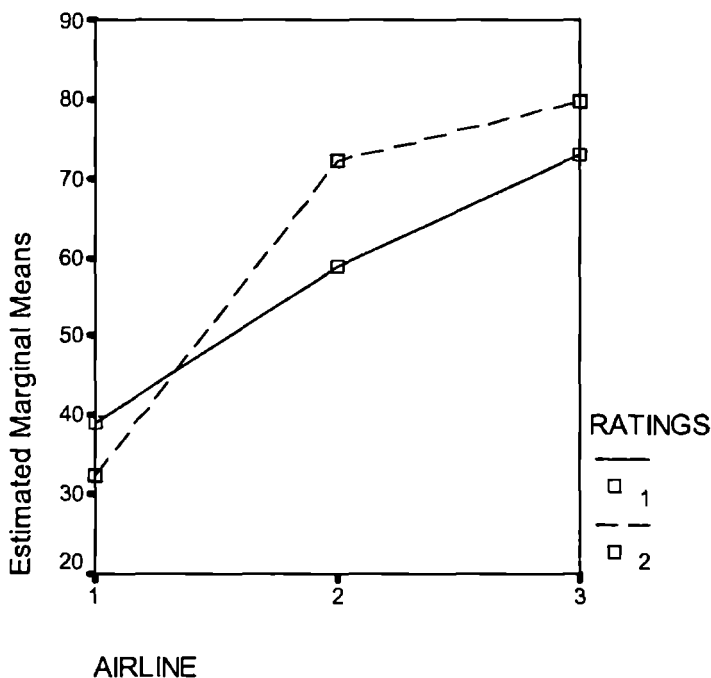


Рис. 7.8. График средних значений при повторных измерениях

В пакете STATISTICA аналогичные результаты могут быть получены с помощью модуля *ANOVA/MANOVA*. В окне *Variables* в качестве независимого фактора (*Independent factor*) указывается переменная «airline», в качестве зависимых (*Dependent*) переменных – переменные «rating1» и «rating2». (Выделить сразу две

и более переменные можно, удерживая клавишу *Ctrl* на клавиатуре компьютера.) Щелчком по кнопке *Repeated measures (within SS) design* можно определить факторы, объединяющие зависимые переменные, подобно тому, как это было сделано при решении задачи в SPSS. С помощью кнопки *All effects* можно получить выходную форму, содержащую результаты, аналогичные полученным в SPSS (рис. 7.9).

Summary of all Effects design (2-Way Anova (a))						
GENERAL MANOVA 1-AIRLINE, 2-FACTOR1						
Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	17763,78	57	264,3447	67,19928	,000000
2	1	625,63	57	43,6956	14,31799	,000373
12	2	1030,86	57	43,6956	23,59180	,000000

Рис. 7.9. Результаты ANOVA с повторными измерениями в пакете STATISTICA

Из данной формы следует, что все эффекты (влияние фактора авиакомпаний ($F = 67,199$; $p < 0,000001$), влияние фактора времени ($F = 14,318$; $p < 0,0004$) и совместное влияние этих факторов ($F = 23,592$; $p < 0,000001$)) достоверны на высоком уровне значимости. Графики средних значений для всех эффектов могут быть построены после щелчка по кнопке *Mean/Graphs* и выбора анализируемого эффекта в окне *Table of All Effects*.

Задания для самостоятельного выполнения

1. Компания по производству лакокрасочных изделий разработала новый вид краски для внешних (уличных) покрытий. Прежде чем выпускать новую продукцию на рынок, компания намерена проверить устойчивость краски к внешним воздействиям. Компания обладает соответствующим тестирующим оборудованием, которое воздействует на окрашенную поверхность специальными щетками и фиксирует количество воздействий, необходимых для полного стирания краски. С помощью этого оборудования было проведено пять испытаний, в ходе которых сравнивалось качество краски, разработанной компанией, с аналогичными продуктами, предлагаемыми двумя конкурентами. Результаты представлены в табл. 7.18.

Результаты испытаний для проверки качества краски, тыс. воздействий

Номер испытания	Продукт компании	Продукт конкурента А	Продукт конкурента В
1	37	34	24
2	30	19	25
3	34	22	23
4	28	31	20
5	29	27	20

Различается ли оцениваемое качество продуктов? Проверьте гипотезу о различии средних на уровне $\alpha = 0,05$.

Пояснение. Для выполнения задания можно использовать однофакторный дисперсионный анализ для k независимых выборок (объясните, почему). Предварительно необходимо проверить выполнение предпосылок для применения этого метода (однородность дисперсий в выборках и нормальность распределения). Первая предпосылка подтверждается тестом Левена, вторая – Q–Q-графиками для каждой выборки и тестом Шапиро–Уилка. Критическое значение F -критерия: 3,89 при $df = (2;12)$ и $\alpha = 0,05$; эмпирическое значение F -критерия: 5,5. Нулевая гипотеза отклоняется, $p = 0,02$. Далее требуется использовать апостериорные сравнения (метод Шеффе). Статистически значимое различие обнаруживается между компанией и конкурентом В.

2. Производитель компьютерной и оргтехники разработал новый продукт, позиционируемый преимущественно на использование в «домашних» условиях, а не в офисе. Организованы тестовые продажи продукта в магазинах, торгующих бытовой электроникой, универмагах и специализированных компьютерных магазинах (случайно отобраны по пять торговых точек каждой категории). Количество покупок продукта в магазинах каждого типа приведено в табл. 7.19.

Таблица 7.19

Количество покупок продукта

Тип магазина	Точка продаж				
	1	2	3	4	5
Магазины электроники	5	4	3	3	3
Универмаги	9	7	8	6	5
Компьютерные магазины	7	4	8	4	3

Определите, в магазинах какого типа следует организовать основные продажи. Проверьте гипотезу о различиях средних на уровне $\alpha = 0,05$.

Пояснение Поскольку данные представлены в шкале отношений и имеется k независимых выборок, может быть использован однофакторный дисперсионный анализ при условии выполнения остальных необходимых предпосылок (требуется проверить). $F_{\text{крит}}: 3,89$ при $df = (2; 12)$ и $\alpha = 0,05$; $F_{\text{эмп}}: 5,43$. Нулевая гипотеза отклоняется, $p = 0,021$. Дальнейшее применение метода Шеффе обнаруживает статистически значимое различие между магазинами электроники и универмагами.

3. Необходимо оценить эффективность, с которой каждая из трех социальных служб (агентств) реализует определенную социально значимую городскую программу. В частности, заказчица этого исследования интересуется скоростью, с которой данные агентства работают с документами, определяя, подходит ли потенциальный клиент под условия этой программы. Для выполнения оценки работы агентств была собрана информация о количестве дней, требуемых для обработки документов. Информация была получена на основе формирования случайной выборки из 10 наблюдений для каждого агентства. Имеющиеся данные приведены в табл. 7.20 (в днях). Существует ли значимое различие между тремя агентствами по скорости обработки документов? Принять уровень значимости $\alpha = 0,05$.

Пояснение. Для выполнения задания целесообразно применить процедуру ANOVA и критерии множественных сравнений (объясните, почему). Средние значения по выборкам: 7, 13, 15 дней соответственно. Эмпирическое значение F -критерия: 12,66; критическое значение F -критерия: 3,35 ($\alpha = 0,05$; $df_B = 2$, $df_W = 27$). Для $\alpha = 0,01$ критическое значение F -критерия: 5,49. Нулевая гипотеза об отсутствии различий может быть отклонена при $p < 0,01$. Используя функцию FРАСПОБР (α ; 2; 57), показать, что нулевая гипотеза может быть отклонена и при $p < 0,001$. Поэтому можно сделать вывод: гипотеза о том, что различия между тремя агентствами вызваны только случайными отклонениями, маловероятна; скорость обработки документов статистически значимо различается по агентствам. Используя критерии множественных сравнений, выясните, по каким именно агентствам наблюдаются различия.

Время обработки документов в агентствах

№ п/п	Агентство 1	Агентство 2	Агентство 3
1	5	12	9
2	7	10	8
3	8	19	12
4	10	20	15
5	4	12	20
6	9	11	21
7	6	13	20
8	9	14	19
9	6	10	15
10	6	9	11

4. Центр социально-политических исследований проводит оценку различий активности избирателей при проведении выборов разного уровня (выборы депутатов в органы законодательной власти региона, выборы депутатов Государственной Думы и выборы Президента страны). Случайная выборка из избирательных участков показывает процент явки избирателей в зависимости от типа выборов (табл. 7.21). Изменяется ли значительно процент явки избирателей в зависимости от типа выборов?

Таблица 7.21

Процент явки избирателей на выборы

Региональные выборы	Выборы в Думу	Выборы Президента
33	35	42
78	56	40
32	35	52
28	40	66
10	45	78
12	42	62
61	65	57
28	62	75
29	25	72
45	47	51
44	52	69
41	55	59

5. Какие методы проверки статистических гипотез и почему могут быть корректно применены в следующих ситуациях?

а) проверка зависимости карьерного роста сотрудников государственных организаций от пола ($n = 100$);

б) партия уплотнителей проверяется на соответствие техническим требованиям по толщине, которая должна составлять 2 мм ($n = 25$);

в) проверяется различие общего расходования горючего в двух службах такси, принадлежащих одному владельцу (отобрано по 10 автомобилей от каждой компании);

г) оценивается результативность тренинговой программы «Эффективные продажи». В тренинге принимали участие 15 торговых представителей компании. Требуется выяснить, изменились ли статистически значимо объемы продаж через месяц после проведения тренинга;

д) компания предполагает определить, связан ли объем продаж с повышением качества предлагаемой продукции, затратами на рекламу или совместным воздействием качества продукции и затрат на рекламу. Зависимая переменная – объем продаж (в тыс. руб.), независимыми переменными (влияющими факторами) являются качество продукции и объем рекламы. Качество продукции ранжируется как «высокое», «среднее» и «низкое», а рекламный бюджет представлен в диапазонах «0–10000 руб.», «10000–20000 руб.», «свыше 20000 руб.»);

е) для оценки успешности карьерного старта выпускников двух ведущих университетов города проводится выборочный опрос о заработной плате (по уровню оплаты) и престижности работы (в баллах) выпускников этих университетов, получивших диплом год назад. Количество опрошенных выпускников одного университета – 20 чел., другого – 30 чел.

б. По данным табл. П.9 проверьте, различается ли организационный климат в двух компаниях по показателям. Например, различается ли значение показателя 1 в двух компаниях, и т.д. (проверьте различие каждого из семи показателей). В каком направлении имеются различия?

ба. Решите задачу б по данным, представленным в шкале Лайкерта.

бб. Решите задачу б по данным, представленным в интервальной шкале.

бв. Решите задачу б после преобразования данных в шкалу Раша. Сравните результаты.

7. С целью оценки надежности-стабильности нового инструмента для оценки когнитивных стилей (И1) использовался метод тест-ретест (см. подразд. 4.2) с интервалом в 1 мес. (объем выборки – 28 респондентов). Данные тестирования представлены в табл. 7.22.

Таблица 7.22

Данные тестирования (И1)

Респондент	Тест	Ретест	Респондент	Тест	Ретест
1	84	98	15	97	95
2	88	93	16	102	98
3	88	92	17	102	95
4	89	94	18	103	97
5	89	77	19	105	99
6	90	89	20	107	112
7	93	91	21	107	104
8	93	91	22	107	97
9	95	111	23	109	109
10	95	99	24	111	114
11	95	82	25	111	115
12	96	106	26	115	112
13	96	95	27	116	115
14	97	86	28	130	129

С помощью методов проверки статистических гипотез определите, существует ли статистически значимое различие в результатах двух тестирований. Предположите, что: а) шкала порядковая; б) шкала интервальная. Сравните результаты. Оцените на этой основе надежность измерений. Объясните преимущества и недостатки такого подхода к оценке надежности по сравнению с подходом, основанном на определении коэффициента корреляции между измерениями. Обоснуйте, какие из известных вам методов проверки статистических гипотез могут быть применимы для решения данной задачи, а какие нет. Объясните преимущества и недостатки использования каждого метода для решения поставленной задачи.

Непараметрические методы проверки статистических гипотез

8.1. Критерий χ^2 Пирсона

Критерий χ^2 (хи-квадрат) – один из самых широко используемых критериев для проверки статистических гипотез в социально-экономических и гуманитарных исследованиях. Это наиболее простой критерий проверки значимости связи между категоризованными переменными. Популярность этого критерия связана как с его простотой, так и с возможностями очень гибкого применения. В отличие от параметрических методов для применения критерия χ^2 не требуется выполнение многих допущений и ограничений, с которыми связано использование, например, z - и t -тестов. Данные могут представляться в любых шкалах, в том числе номинативной шкале; форма распределения также несущественна для этого теста (поэтому такие критерии еще называются *критериями, свободными от распределения (distribution free tests)*).

В качестве примеров использования этого критерия можно назвать следующие исследовательские задачи.

- Кто чаще обращается в службу знакомств: мужчины или женщины?
- Влияет ли на частоту посещений магазина уровень доходов и статус покупателей?
- Кто чаще совершает аварии: мужчины или женщины?
- Зависит ли количество аварий от дня недели?

Различаются два варианта применения этого критерия. Первый вариант, который называется *критерием согласия*, предназначен для проверки различий эмпирического и теоретического распределений исследуемого свойства по одной выборке. Сравнивается численность элементов выборки, обладающих определенным свойством, с предполагаемой (теоретической) численностью согласно определенному виду распределения случайной величины. На практике чаще всего сопоставляются численности долей откликов по выборке с равномерным распределением, хотя вид теоретического распределения, с которым сравнивается эмпирическое распределение, может быть и любым другим.

Эмпирические частоты откликов по определенной градации изменения изучаемого свойства называются *наблюдаемыми*, а теоретические частоты — *ожидаемыми*.

Второй вариант, который называется *критерием независимости*, применяется для сравнения распределений *нескольких независимых выборок*. В этом случае нулевая гипотеза формулируется как отсутствие зависимости (связи) между изучаемыми переменными. Применительно к критерию хи-квадрат понятие независимости имеет следующее значение: две переменные являются независимыми, если отнесение наблюдения к определенной категории одной переменной не оказывает влияния на вероятность того, что это наблюдение попадет в определенную категорию другой переменной. Иными словами, при формулировке нулевой гипотезы в терминах независимости истинная нулевая гипотеза означает, что влияние одной выборки на другую (или другие) случайно.

В случае использования критерия согласия эмпирическое значение хи-квадрат вычисляется путем сравнения эмпирического, или наблюдаемого (*observed*), распределения частот (*O*) и теоретического, или ожидаемого (*expected*), распределения (*E*) по следующей формуле:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i},$$

где *k* — количество градаций изменения изучаемого признака, т.е. количество сравниваемых категорий (или количество сравнений по определенным категориям).

Следуя общей логике и процедуре проверки статистических гипотез, эмпирическое значение хи-квадрат сравнивается с критическим при выбранном уровне значимости α и числе степеней свободы для исследуемой ситуации. Число степеней свободы для критерия согласия хи-квадрат определяется как $df = k - 1$. Критическое значение хи-квадрат можно получить из специальных статистических таблиц (П. 1.4), а также с помощью функции Excel ХИ2ОБР(α ; df).

Проиллюстрируем процедуру вычисления статистики критерия согласия хи-квадрат на примере, связанном с мониторингом проекта создания «Университетского ланч-клуба» (см. гл. 3). Предположим, что поставлена задача оценить возможное посе-

шение клуба студентами в зависимости от условий их проживания. Допустим, что было опрошено 200 студентов. Нулевая гипотеза H_0 формулируется следующим образом: желание посещать ланч-клуб не зависит от условий проживания, т.е. наблюдаемое распределение частот соответствует ожидаемому (в данном случае – равномерному) распределению ($O = E$).

В результате опроса получены данные, приведенные в табл. 8.1, где указаны частоты откликов респондентов по определенным группам (категориям). В данном случае имеются четыре категории изменения признака, т.е. $k = 4$.

Таблица 8.1

Результаты опроса студентов

	Намерены посещать (O)	Опрошено человек	Ожидаемое значение (E)
Общежитие	16	90	27
Квартира в центре	13	40	12
Квартира на окраине	16	40	12
За городом	15	30	9
Всего	60	200	60

В данном случае принцип вычисления ожидаемых частот (E) очевиден: количество опрошенных по каждой из четырех категорий умножается на отношение суммы наблюдаемых частот O (сколько всего человек намерены посещать «Университетский ланч-клуб») к общему количеству опрошенных (предполагается равномерное распределение). Рассчитывая статистику χ^2 , имеем эмпирическое значение критерия $\chi^2 = 9,898$ (рекомендуется проверить самостоятельно). Количество степеней свободы $df = 4 - 1 = 3$. При $\alpha = 0,05$ критическое значение $\chi^2 = 7,815$. Сравнивая критическое и эмпирическое значения критерия, принимаем решение об отклонении нулевой гипотезы и делаем вывод о статистически значимом влиянии условий проживания студентов на их вероятное посещение «Университетского ланч-клуба».

Задачи такого рода можно легко решить в среде STATISTICA. В модуле *Nonparametrics/Distrib.* выбирается процедура *Observed versus expected XI*, затем в окне *Variables* – переменные, содержащие наблюдаемые и ожидаемые частоты, после чего нужно щелк-

нуть по кнопке *OK*. Исходные данные из рассмотренного выше примера и результаты вычисления критерия согласия χ^2 представлены на рис. 8.1.

Observed vs. Expected Frequencies (xi2.sta)				
NONPAR		Chi-Square = 9,898149 df = 3 p < ,019464		
STATS		observed	Expected	
Case		O	E	
C: 1	16,00000	27,00000	-11,00000	4,481482
2	13,00000	12,00000	1,00000	,0833333
3	16,00000	12,00000	4,00000	1,3333333
4	15,00000	9,00000	6,00000	4,0000000
Sum	60,00000	60,00000	0,00000	9,898149

Рис. 8.1. Расчет критерия χ^2 в пакете STATISTICA

При вычислении критерия согласия χ^2 в SPSS исходные данные должны быть представлены в виде столбца, содержащего значения номинативной переменной по нескольким категориям (кодировка значений каждой категории может быть произвольной). Процедура расчета критерия согласия запускается из меню *Analyze, Nonparametric tests, Chi-square*. В открывшемся окне выбирается необходимая переменная и переносится в поле *Test Variable List*. После щелчка по кнопке *OK* вычисляются наблюдаемые и ожидаемые частоты, статистика хи-квадрат, уровень значимости, которые представляются в соответствующих отчетах формируемой выходной формы.

Применение критерия независимости χ^2 основано на использовании *таблицы сопряженности*, которая отражает перекрестное распределение частот представленности изучаемого признака по градациям его изменения в одной или нескольких сопоставляемых выборках.

Пример 8.1. Для иллюстрации понятия таблицы сопряженности предположим, что поставлена задача оценить различие в предполагаемом посещении «Университетского ланч-клуба» студентами и студентками. Всего было опрошено 50 студентов и 50 студенток. Предпочтения мужчин и женщин по посещению «Университетского ланч-клуба» отражены в табл. 8.2 (положительный ответ о предполагаемом посещении дали 26 студентов и 35 студенток, а отрицательный ответ – 24 студента и 15 студенток).

Таблица сопряженности «пол – посещение»

	Вероятно, буду посещать	Вероятно, не буду посещать	Всего
Мужчины	26	24	50
Женщины	35	15	50
Всего	61	39	100

Это таблица сопряженности 2×2 , в которой количество сопоставляемых выборок равно двум (мужчины и женщины) и количество градаций изучаемого признака тоже равно двум («буду посещать» и «не буду посещать»). Если бы все студентки выбрали, предположим, вариант «буду посещать», а все студенты – вариант «не буду посещать», то было бы очевидно, что между предпочтениями студентов и студенток есть явно выраженные различия. В данном случае различия в предпочтениях выражены слабее, поэтому наблюдаемое различие необходимо проверять на статистическую значимость.

Для проверки статистической значимости различий такого рода может быть использован критерий независимости χ^2 , который, как и критерий согласия, тестирует значимость различия между *наблюдаемыми* и *ожидаемыми* частотами. В нашем примере если между предпочтениями студентов разного пола нет различий (или, что то же самое, между предпочтением и полом нет связи), то естественно предполагать равное количество частот выбора вариантов «посещать» и «не посещать» для каждой выборки. Относительно малые отклонения наблюдаемых частот от ожидаемых будут доказывать значимость, если число наблюдений велико.

Формула для определения экспериментального значения критерия независимости χ^2 незначительно отличается от формулы для критерия согласия. Отличие заключается в том, что сравнение наблюдаемых (O) и ожидаемых частот (E) осуществляется по всем ячейкам таблицы сопряженности

$$\chi^2 = \sum_i^k \sum_j^l \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

где k – число градаций изменения изучаемого признака;

l – количество сопоставляемых распределений (выборок); индексы i, j соответствуют ячейкам таблицы сопряженности.

Ожидаемые частоты в таблице сопряженности при гипотезе «нет различий/нет связи» можно вычислить непосредственно. С этой целью для каждой ячейки таблицы предварительно определяется так называемая краевая, или маргинальная (*marginal*), частота по каждой строке и каждому столбцу таблицы (в таблице рассматриваемого примера она обозначена как «Всего»). Ожидаемая частота для каждой ячейки таблицы равна произведению краевой частоты по столбцу на краевую частоту по строке, деленному на общее количество наблюдений в таблице (n). В нашем случае $n = 100$. Тогда для левой верхней ячейки имеем $61 \cdot 50/100 = 30,5$. В общем случае ожидаемая частота в ячейке ij рассчитывается как

$$E_{ij} = \frac{f_i \cdot f_j}{n},$$

где f_i – сумма частот i -й строки;
 f_j – сумма частот j -го столбца таблицы сопряженности.

Таким образом, значение статистики χ^2 зависит от общего числа наблюдений и количества ячеек в таблице сопряженности. Число степеней свободы для критерия независимости χ^2 определяется как $df = (k - 1)(l - 1)$, где k – число градаций изменения изучаемого признака, l – количество сопоставляемых распределений.

Пример 8.2. Директор фабрики, который пытается ввести на фабрике запрет на курение, намерен установить, связано ли курение с браком на производстве. По результатам учета бракованных изделий случайным образом сформированы списки рабочих, соответственно допускаявших и не допускаявших брак в работе. После интервьюирования рабочих об их пристрастии к курению (случайная выборка) и сопоставления со сведениями о браке построена таблица сопряженности, в которой указаны также определенные ожидаемые частоты (табл. 8.3).

Таблица 8.3

Таблица сопряженности «брак – курение»

	Брак	Нет брака	Всего
Заядлый курильщик	12 8,24	4 7,75	16 –
Умеренный курильщик	9 7,73	6 7,27	15 –
Не курит	13 18,03	22 16,97	35 –
Всего	34	32	66

Читателю рекомендуется проверить, что эмпирическое значение χ^2 составляет 6,86, а критическое значение – 5,59 ($\alpha = 0,05$, $df = 2$). Нулевая гипотеза отклоняется, и принимается, что случаи брака связаны с курением.

В случае, когда количество сравниваемых выборок превышает две, использование критерия независимости χ^2 не отличается от случая двух независимых выборок. Все расчеты проводятся аналогично.

Альтернативная гипотеза, принимаемая при отклонении нулевой гипотезы об отсутствии связи, неконкретна, в том смысле, что можно утверждать, что связь проявляется по крайней мере для одной ячейки таблицы сопряженности, но неизвестно, для какой именно. Поэтому, как и в случае применения метода ANOVA, возникает необходимость конкретизации результатов. Для этого исходную таблицу сопряженности обычно разбивают на подтаблицы 2×2 , после чего проводят множественные сравнения с помощью критерия χ^2 .

Имеется существенное ограничение на использование критерия χ^2 (кроме предположения о случайном выборе наблюдений и независимости выборок). Это ограничение состоит в том, что ожидаемые частоты не должны быть очень малы. Существует эмпирическое правило, помогающее определить минимальное значение ожидаемых частот, при котором применение критерия хи-квадрат можно считать корректным. Если число степеней свободы $df = 1$, то ожидаемые частоты в каждой ячейке не должны быть меньше пяти. Если число степеней свободы $df > 1$, то в таблице сопряженности не должно быть свыше 20% ячеек с ожидаемыми частотами ниже пяти и ни одной ячейки с ожидаемой частотой ниже единицы.

Для таблиц сопряженности 2×2 с малыми частотами (меньше 10) рекомендуется рассчитывать эмпирическое значение критерия χ^2 , используя поправку на непрерывность Йетса (*Yates' correction for continuity*):

$$\chi^2 = \sum_i \sum_j \frac{(|O_{ij} - E_{ij}| - 0,5)^2}{E_{ij}}.$$

Поскольку поправка Йетса применяется лишь для таблиц 2×2 , то $df = 1$. Поправка Йетса делает оценку связей более умеренной, она может использоваться, когда объем выборки n

находится между 20 и 40, а если хотя бы одно из значений ожидаемых частот меньше 5, то для $n > 40$. Если эти требования не выполняются, можно применить так называемый *точный критерий Фишера (Fisher's exact test)*, расчет которого целесообразно выполнять с помощью статистических пакетов.

Точный критерий Фишера, как и поправка Йетса, применим только для таблиц сопряженности 2×2 . Этот критерий основан на вычислении вероятности получения наблюдаемых частот исходя из краевых частот, заданных в таблице сопряженности. Такая вероятность *точно* вычисляется подсчетом всех таблиц, которые можно построить исходя из известных краевых частот. Таким образом, критерий Фишера вычисляет вероятность появления наблюдаемых частот при нулевой гипотезе (отсутствие связи между табулированными переменными).

Для таблиц размерности большей, чем 2×2 , не существует формулы для корректировки эмпирического значения критерия χ^2 в случае малых выборок. Выход может заключаться в том, чтобы объединить некоторые категории переменной, тем самым увеличив ожидаемые частоты. Но такой подход можно использовать только тогда, когда исследователь считает возможным (с точки зрения здравого смысла или теоретического обоснования) игнорировать различия в тех или иных категориях. Если категории не удастся объединить, а количество ячеек с частотами пять и менее небольшое, то критерий можно применять, но результаты следует интерпретировать с осторожностью.

Как и все статистические тесты, χ^2 чувствителен к объему выборки даже в большей степени, чем другие тесты (так как эмпирическое значение критерия изменяется пропорционально объему выборки). Поэтому использование больших выборок может привести к решению об отклонении нулевой гипотезы, хотя связи между переменными могут быть несущественны. Это обстоятельство еще раз подчеркивает различие между статистической значимостью и практической важностью. Критерий χ^2 , как и все другие статистические критерии, может служить индикатором того, являются ли результаты статистически значимыми. Но он не может свидетельствовать о том, *важны* ли результаты в любом другом смысле. Отчасти это может разрешить оценка мер связи, рассматриваемая в третьей части учебного пособия.

В SPSS для вычисления критерия независимости χ^2 , а также других критериев, рассчитываемых по таблицам сопряженности, используется процедура *Crosstabs*, запускаемая из меню *Analyze, Descriptive Statistics*. Исходные данные для работы должны быть представлены в виде столбцов, соответствующих изменяемым номинативным переменным. Например, для данных табл. 8.3 первый столбец (DEFECT) может включать выборки работников, допускавших и не допускавших брак. В этот столбец могут входить значения 1 и 2, которые соответствуют работе с браком и без брака. В столбце будут 34 элемента со значением 1 и 32 элемента со значением 2. Второй столбец (SMOKE) – степень пристрастия работника к курению (например, 1 – «заядлый курильщик», 2 – «умеренный курильщик», 3 – «не курит»). Всего в этом столбце будет 16 элементов со значением 1, 15 элементов со значением 2 и 35 элементов со значением 3 (всего 66 элементов). В поля *Row(s)* и *Column(s)* необходимо перенести переменные, соответствующие строкам и столбцам таблицы сопряженности. В данном случае переменная DEFECT переносится в поле *Column(s)*, а переменная SMOKE – в поле *Row(s)*. Щелчком по кнопке *Statistics* выбирается режим *Chi-square*. После щелчка по кнопкам *Continue* и *OK* осуществляется подсчет наблюдаемых и ожидаемых частот и статистик критерия хи-квадрат. Результаты расчетов представляются в отчетах выходной формы, они полностью совпадают с результатами, полученными выше (табл. 8.4).

Таблица 8.4

Результаты расчета критерия χ^2 в SPSS

SMOKE*DEFECT Crosstabulation

Count

		DEFECT		Total
		1	2	
SMOKE	1	12	4	16
	2	9	6	15
	3	13	22	35
Total		34	32	66

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6,860 ^a	2	.032
Likelihood Ratio	7,070	2	.029
Linear-by-Linear Association	6,688	1	.010
N of Valid Cases	66		

a. 0 cells (0%) have expected count less than 5. The minimum expected count is 7.27.

В пакете STATISTICA процедура работы с таблицами сопряженности осуществляется в модуле *Basic statistics* (меню *Tables and banners*). Исходные данные также задаются в виде столбцов, содержащих значения номинативной переменной по категориям изменения. Для таблиц сопряженности 2×2 статистики всех критериев выводятся в одной выходной форме после ввода наблюдаемых частот (модуль *Nonparametrics/Distrib.*, процедура 2×2 *Tables...*).

8.2. Тест Мак-Немара

Если гипотеза проверяется для зависимых (связанных) выборок с данными, представленными в неметрической шкале, то используется *тест Мак-Немара* (*McNemar test*). Особенно полезен этот тест при повторных измерениях. Он позволяет сравнить долю респондентов, которые не имели некоторой характеристики (0) во время первого обследования (например, до воздействия), но стали обладать ею во время повторного обследования (1), с долей тех, кто имел это свойство (1), но перестал иметь его после (0) (табл. 8.5).

Таблица 8.5

Иллюстрация теста Мак-Немара

До	После	
	0	1
1	<i>A</i>	<i>B</i>
0	<i>C</i>	<i>D</i>

Например, метод может применяться для оценки предпочтений некоторого товара одними и теми же респондентами *до* и *после* рекламы. В таком случае 0 может означать, что товар не нравится, а 1 – что товар нравится.

Вычисляются два значения χ^2 : для ячеек *A* и *D* (1,0 и 0,1) и для ячеек *B* и *C* (1,1 и 0,0). *A/D* хи-квадрат проверяет гипотезу о том, что частоты в ячейках *A* и *D* (верхняя левая, нижняя правая) одинаковы. *B/C* χ^2 проверяет гипотезу о равенстве частот в ячейках *B* и *C* (верхняя правая, нижняя левая). Метод не может использоваться для выборок, содержащих меньше 10 элементов.

Для ячеек *A* и *D* эмпирическое значение χ^2 для теста Мак-Немара вычисляется по формуле

$$\chi^2 = \frac{(|A - D|)^2}{A + D}, \quad df = 1.$$

Нулевая гипотеза формулируется как $H_0: P(A) = P(D)$.

Соответственно для ячеек B и C эмпирическое значение хи-квадрат определяется как

$$\chi^2 = \frac{(|B - C|)^2}{B + C},$$

а нулевая гипотеза формулируется как $H_0: P(B) = P(C)$.

Подобно тому, как для независимых выборок вводится поправка Йетса, при применении теста Мак-Немара также рекомендуется использовать модифицированную формулу для эмпирического значения хи-квадрат, дающую более умеренную оценку. Например, для ячеек A и D эмпирическое значение χ^2 должно определяться следующим образом:

$$\chi^2 = \frac{(|A - D| - 1)^2}{A + D}, \quad df = 1.$$

В пакетах SPSS и STATISTICA метод реализован с использованием критерия χ^2 , но реализация метода Мак-Немара может быть основана и на применении z -критерия. В этом случае проверка гипотез (а) (1,0 и 0,1) и б) (1,1 и 0,0) осуществляется на основе одной из формул для соответствующих гипотез:

$$z = \frac{|A - D|}{\sqrt{A + D}};$$

$$z = \frac{|B - C|}{\sqrt{B + C}}.$$

Найденные по этим формулам эмпирические значения z -критерия сравниваются с критическими, на основе чего делается вывод об отклонении или невозможности отклонения нулевой гипотезы.

Из данных формул следует, что значение z , возведенное в квадрат, будет равняться значению χ^2 , определенному по фор-

муле без корректировки. Для иллюстрации взаимосвязи результатов, полученных при использовании χ^2 и z -критерия в методе Мак-Немара, приведем два примера.

Пример 8.3. В компании проводился опрос, насколько сотрудники ценят коллективную работу. Была случайным образом сформирована выборка из 200 сотрудников. По результатам опроса сотрудники были разделены на две группы с условным названием «Нравится коллективная работа» и «Не нравится коллективная работа». В компанию был приглашен специалист, который провел тренинг для сотрудников, включающий элементы совместного принятия решений. После тренингов по той же выборке был осуществлен повторный опрос. Результаты представлены в табл. 8.6. Нулевая и альтернативная гипотезы соответственно формулируются как $H_0: P(A) = P(D)$ и $H_A: P(A) \neq P(D)$.

Таблица 8.6

Результаты опроса сотрудников

До	После	
	не нравится	нравится
Нравится	10	90
Не нравится	60	40

Используется тест Мак-Немара, поскольку данные представлены в номинативной шкале, две выборки являются зависимыми (проводятся повторные измерения). При подсчете вручную получаем эмпирическое значение χ^2 , равное $29^2/50 = 16,82$. Критическое значение (при $\alpha = 0,05$ и $df = 1$) равно 3,84. Нулевая гипотеза отклоняется. Фактически эмпирическое значение таково, что $p < 0,001$. Отметим, что при применении формулы без корректировки получим $30^2/50 = 18$.

Определим теперь эмпирическое значение z -критерия:

$$z = \frac{|10 - 40|}{\sqrt{10 + 40}} = 4,243.$$

При $\alpha = 0,05$ критическое значение z составляет 1,96. Следовательно, и на основе этого критерия мы отклоняем нулевую гипотезу. Заметим, что если возвести полученное эмпирическое значение z в квадрат, то мы получим эмпирическое значение хи-квадрат, определенное по формуле без корректировки: $4,243^2 = 18$.

Пример 8.4. С помощью этого примера продемонстрируем еще один случай, когда выборки должны рассматриваться как связанные. Компьютерная компания тестирует две различные антивирусные программы (*A* и *B*) на определенном количестве зараженных файлов (156). Результаты тестирования (количество файлов, «излеченных» от вирусов) представлены в табл. 8.7.

Таблица 8.7

Результаты тестирования антивирусных программ

«Излечено» программой <i>A</i>	«Излечено» программой <i>B</i>	
	нет	да
Да	33	45
Нет	20	58

Можно ли, основываясь на этих испытаниях, сказать, что программы работают приблизительно с одинаковой эффективностью?

Данные представлены в номинативной шкале («излечен» — «не излечен»), но применять в этом случае критерий χ^2 для независимых выборок нельзя, поскольку программы тестируются на одной и той же выборке зараженных файлов. Фактически мы работаем с одной выборкой, подвергая ее разным воздействиям, поэтому мы должны принять предположение связанности выборки и, следовательно, использовать тест Мак-Немара.

В качестве индикатора эффективности работы антивирусной программы выберем показатель «количество файлов, излеченных одной программой, но не излеченных другой». В таком случае мы должны сравнивать ситуации (1,0) и (0,1), т.е. нас интересуют ячейки *A* и *D*. Нулевая гипотеза формулируется как $H_0: P(A) = P(D)$ — показатели эффективности программ равны. Соответственно альтернативная гипотеза формулируется как $H_A: P(A) \neq P(D)$ — показатели эффективности программ не равны.

Определяем эмпирическое значение z :

$$z = \frac{|33 - 58|}{\sqrt{33 + 58}} = 2,62.$$

При $\alpha = 0,05$ критическое значение z составляет 1,96. Следовательно, мы можем отклонить нулевую гипотезу и признать, что эффективность программ статистически значимо различается.

Определим теперь эмпирическое значение χ^2 по формуле без корректировки:

$$\chi^2 = \frac{(|33 - 58|)^2}{33 + 58} = 6,87.$$

Отметим, что квадратный корень из этого значения равен эмпирическому значению z : $\sqrt{6,87} = 2,62$.

Критическое значение χ^2 при $df = 1$ и $\alpha = 0,05$ равно 3,84. Таким образом, нулевая гипотеза отклоняется и на основе этого критерия. Применяв модифицированную формулу, получим

$$\chi^2 = \frac{(|33 - 58| - 1)^2}{33 + 58} = 6,33.$$

Нулевая гипотеза также отвергается, но уже при несколько меньшем уровне значимости.

В случае, если результаты измерений сразу представлены в виде таблицы сопряженности, то удобнее и быстрее определить эмпирические значения χ^2 или z по формулам метода Мак-Немара вручную или с помощью программы STATISTICA, чем с помощью SPSS. Дело в том, что, как отмечалось в подразд. 8.1, исходные данные для работы с таблицами сопряженности в SPSS должны представляться в виде столбцов, соответствующих изменяемым номинативным переменным (например, состояниям «до» и «после»). Эти переменные содержат значения, например 0 или 1, в зависимости от наличия или отсутствия интересующей исследователя характеристики. На основе данных, введенных таким образом, программа SPSS формирует таблицы сопряженности и вычисляет необходимые статистические показатели, которые можно использовать для принятия решения об отклонении или невозможности отклонения нулевой гипотезы.

Программа STATISTICA позволяет работать с таблицами сопряженности 2×2 непосредственно по заданию наблюдаемых частот в окне 2×2 Tables, выводимом выбором процедуры 2×2 Tables... модуля Nonparametrics/Distrib. (рис. 8.2).

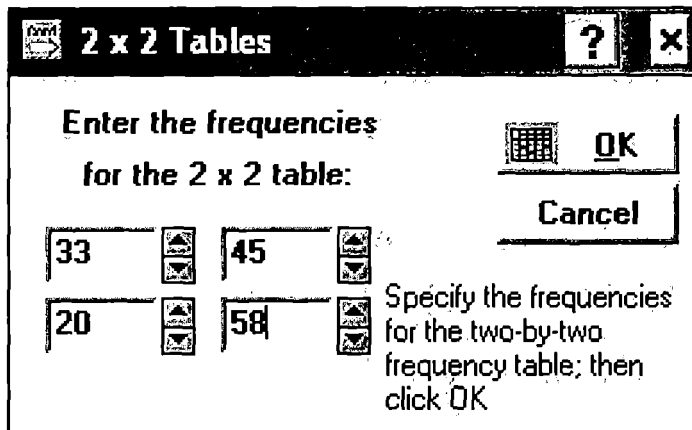


Рис. 8.2. Использование таблиц сопряженности в пакете STATISTICA

В отчете выводятся все статистические критерии, рассчитываемые для таблиц сопряженности 2×2 , в том числе статистика теста Мак-Немара (рис. 8.3).

2 x 2 Table (xi2.sta)			
Continue...	Column 1	Column 2	Row Totals
Frequencies, row 1	33	45	78
Percent of total	21,154%	28,846%	50,000%
Frequencies, row 2	20	58	78
Percent of total	12,821%	37,179%	50,000%
Column totals	53	103	156
Percent of total	33,974%	66,026%	
Chi-square (df=1)	4,83	p= ,0280	
V-square (df=1)	4,80	p= ,0285	
Yates corrected Chi-square	4,12	p= ,0425	
Phi-square	,03096		
Fisher exact p, one-tailed		p= ,0210	
two-tailed		p= ,0420	
McNemar Chi-square (A/D)	6,33	p= ,0119	
Chi-square (B/C)	8,86	p= ,0029	

Рис. 8.3. Результаты обработки данных из таблицы сопряженности

Если исходные данные представлены в виде двух наборов данных в номинативной шкале, то целесообразно использовать программу SPSS. Для этого в меню *Analyze* необходимо выбрать *Descriptive Statistics*, затем *Crosstabs*. Одна из переменных переносится в поле *Row(s)*, другая – в поле *Column(s)*. В *Statistics* выбирается нужный тест (*McNemar*). Тест Мак-Немара можно запустить также через *Analyze, Nonparametric Tests, 2 related samples, McNemar*.

8.3. Критерий Манна–Уитни

Для сравнения двух независимых выборок может быть применен *U-критерий Манна–Уитни (Mann–Whitney U-test)*. Этот статистический критерий является самым мощным (чувствительным) среди непараметрических аналогов *t*-критерия Стьюдента для независимых выборок. Для использования этого критерия данные должны быть представлены, по меньшей мере, в порядковой шкале (но тест может быть применен и для шкал большей мощности).

Эмпирическое значение критерия показывает, насколько совпадают два ряда значений измеряемого признака. Для вычисления статистики критерия значения двух выборок ранжируются: представляются в виде одной упорядоченной последовательности значений – от наименьших к наибольшим. Наибольшие отрицательные значения имеют наименьший ранг. Если несколько элементов имеют одинаковое значение, им приписываются средние арифметические значения рангов. Например, два элемента имеют значение 1300 и находятся на 8-м и 9-м местах. Каждому из этих элементов приписывается ранг 8,5.

Эмпирическое значение *U*-критерия вычисляется по формуле

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

или

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2,$$

где n_1, n_2 – соответственно объем первой и второй выборок;
 R_1, R_2 – соответственно сумма рангов по первой и второй выборкам.

Правильность вычислений проверяется по формуле

$$U_1 + U_2 = n_1 n_2.$$

В качестве эмпирического значения принимается наименьшее значение из U_1, U_2 . Критические значения этого критерия можно взять из соответствующих статистических таблиц (табл. П.5). Обратите внимание, что для этого теста нулевая гипотеза отклоняется, если эмпирическое значение *меньше* критического.

Пример 8.5. Предположим, оценивается эффективность работы персонала филиалов компании в двух регионах. Для такой оценки среди клиентов распространяется специально разработанная анкета. Результаты анкетирования представляются в порядковой шкале по нескольким показателям. (Это может быть и метрическая шкала, например объемы продаж, но будем считать, что по каким-то причинам t -критерий Стьюдента неприменим.) Результаты оценки даны в табл. 8.8, а также приведены ранги оцененных показателей.

Таблица 8.8

Эффективность работы персонала в филиалах

Филиал 1	Ранг	Филиал 2	Ранг
1500	15	1340	10
1540	16	1300	8,5
1860	22	1620	18
1230	6	1070	3
1370	12	1210	5
1550	17	1170	4
1840	21	1770	20
1250	7	950	1
1300	8,5	1380	13
1350	11	1460	14
1710	19	1030	2

Предположим, что в филиале 1 эффективность работы выше. В этом случае нулевая гипотеза: рейтинги двух филиалов не различаются; альтернативная гипотеза: рейтинг филиала 1 выше рейтинга филиала 2. Определим суммы рангов ($R_1 = 154,5$; $R_2 = 98,5$), эмпирические значения критерия:

$$U_1 = 11 \cdot 11 + (11(11 + 1))/2 - 154,5 = 32,5;$$

$$U_2 = 11 \cdot 11 + (11(11 + 1))/2 - 98,5 = 88,5.$$

Критическое значение критерия составляет 34 ($\alpha = 0,05$, $n_1 = n_2 = 11$, однонаправленный тест). Наименьшее значение эмпирического значения критерия меньше критического значения, следовательно, нулевая гипотеза отклоняется.

Для использования U -критерия Манна–Уитни в статистических пакетах данные из двух выборок объединяются в одну переменную, в качестве другой (группирующей) переменной берутся идентифицирующие значения (например, 1 и 2). В пакете SPSS из меню *Analyze* выбирается опция *Nonparametric Tests*, затем *2-Independent Samples*.

Независимая переменная (в данном примере – rating) переносится в поле *Test Variable(s)*, группирующая переменная (в данном примере – group) – в поле *Grouping Variable(s)* (рис. 8.4). Для того чтобы задать коды группирующей переменной, соответствующие двум выборкам, необходимо щелкнуть по кнопке *Define Groups* и установить идентифицирующие значения (в данном случае – 1 и 2).

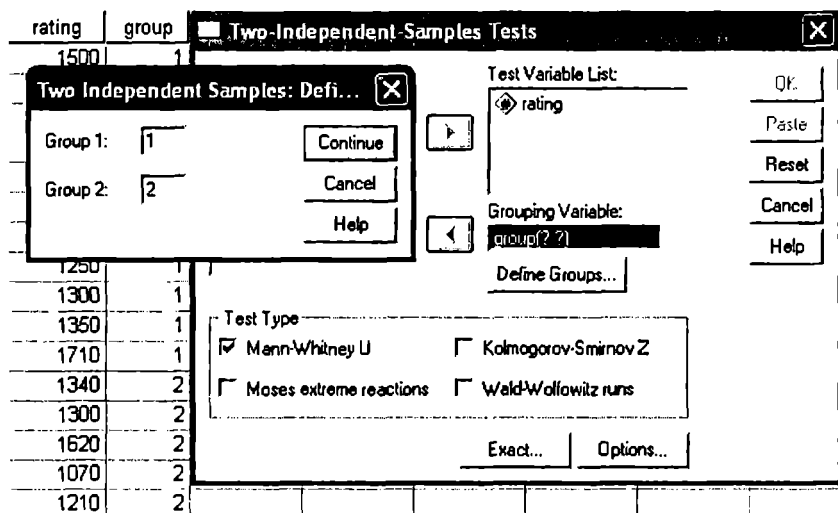


Рис. 8.4. Расчет U -критерия Манна–Уитни пакете SPSS

После щелчка по кнопкам *Continue* и *OK* генерируются отчеты (табл. 8.9). В отчете *Ranks* приводятся объемы выборок (*N*), средние ранги (*Mean Rank*) и суммы рангов (*Sum of Ranks*) по выборкам (сравните эти значения с определенными нами выше).

Таблица 8.9

Результаты применения *U*-критерия Манна–Уитни

Ranks

	GROUP	N	Mean Rank	Sum of Ranks
RATING	1	11	14,05	154,50
	2	11	8,95	98,50
	Total	22		

Test Statistics

	RATING
Mann-Whitney U	32,500
Wilcoxon W	98,500
Z	-1,839
Asymp. Sig. (2-tailed)	,066
Exact Sig. [2*(1-tailed Sig.)]	,065

В отчете *Test Statistics* приводятся минимальное значение из эмпирических значений *U*-критерия Манна–Уитни (*Mann–Whitney U*) и *p*-уровень для двусторонней проверки (*Asymp. Sig. (2-tailed)*). В нашем примере альтернативная гипотеза формулировалась как односторонняя, поэтому полученный *p*-уровень должен быть поделен пополам. Если бы осуществлялась двусторонняя проверка (не было бы оснований предполагать более высокую эффективность работы персонала первого филиала), то нулевая гипотеза не могла бы быть отклонена на уровне значимости 0,05.

Вычисления в пакете STATISTICA осуществляются аналогично. После запуска модуля *Nonparametrics/Distrib.* в меню *Analysis* запускается *Startup Panel* и выбирается процедура *Mann–Whitney U test*. В окне *Variables* в качестве группирующей переменной (*Indep. (grouping) variable*) указывается переменная *group*, а в качестве зависимой (*Dependent variable list*) – переменная *rating* (рис. 8.5).

Статистические показатели, которые приводятся в выходной форме (рис. 8.6), идентичны полученным в пакете SPSS.

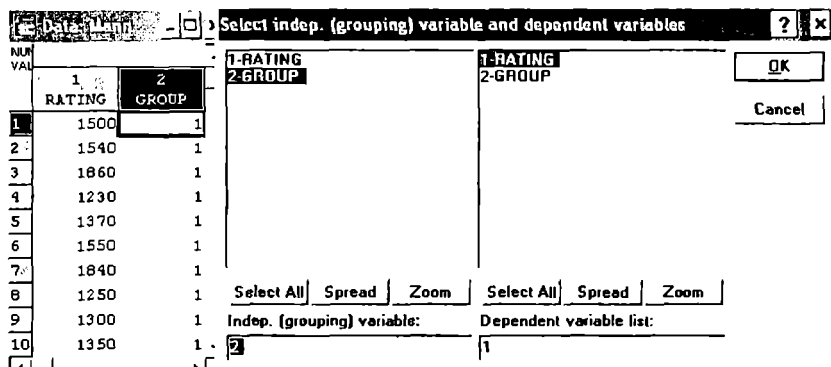


Рис. 8.5. Расчет U -критерия Манна–Уитни в STATISTICA

Mann-Whitney U Test (mann-whitney.sta)					
Continue...					
By variable GROUP					
Group 1: 1 Group 2: 2					
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-level
RATING	154,5000	98,50000	32,50000	1,838620	,065980

Рис. 8.6. Статистика U -критерия Манна–Уитни

8.4. Тест Колмогорова–Смирнова

Тест Колмогорова–Смирнова (*Kolmogorov-Smirnov test*) используется для данных, представленных по крайней мере в порядковой шкале, для сравнения наблюдаемого и теоретического распределений, например нормального, равномерного, пуассоновского или биномиального. Одновыборочный тест Колмогорова–Смирнова (*Kolmogorov-Smirnov one-sample test*) более мощный, чем хи-квадрат, и, более того, он может быть применен для выборок, объем которых даже меньше, чем минимально допустимый объем выборки для корректного использования хи-квадрат. Тест Колмогорова–Смирнова основан на принципе подсчета *накопленных* теоретических (F_T) и наблюдаемых (F_0) частот и определения их максимального расхождения (D):

$$D = \max | F_0(x) - F_T(x) |$$

Найденное значение D сравнивается с критическим значением (табл. П.6).

Проиллюстрируем использование этой процедуры на примере знакомой ситуации с проектом «Университетский ланч-клуб». Предположим, необходимо оценить, влияет ли курс обучения на желание студентов посещать ланч-клуб.

Нулевая гипотеза формулируется как H_0 : нет связи между курсом обучения студентов и их желанием посещать ланч-клуб. Результаты опроса (60 чел.) и обработки данных по методу Колмогорова–Смирнова представлены в таблице. Принцип определения накопленных наблюдаемых и теоретических частот (предполагается равномерное распределение) наглядно показан в табл. 8.10.

Таблица 8.10

Исходные данные и промежуточные расчеты для использования метода Колмогорова–Смирнова

Курс обучения	1	2	3	4	5
Количество студентов, выразивших желание посещать «Университетский ланч-клуб»	5	9	11	16	19
Накопленные наблюдаемые F_n^* частоты	5/60	14/60	25/60	41/60	60/60
Накопленные теоретические частоты	12/60	24/60	36/60	48/60	60/60
$ F_n(x) - F_T(x) $	7/60	10/60	11/60	7/60	0

По данным табл. 8.10 определяем максимальное расхождение накопленных наблюдаемых и теоретических частот $D = 11/60 = 0,183$. Из таблицы критических значений (табл. П.6) находим критическое значение критерия, которое для $\alpha = 0,05$ и $n = 60$ составляет $1,36/\sqrt{60} = 0,175$. Эмпирическое значение критерия превышает его критическое значение, поэтому нулевая гипотеза должна быть отклонена.

Работа с одновыборочным тестом Колмогорова–Смирнова в среде SPSS осуществляется с помощью процедуры *1-Sample K-S* (меню *Analyze, Nonparametric Tests*).

Двухвыборочный тест Колмогорова–Смирнова (Kolmogorov–Smirnov two-sample test) применяется для проверки гипотезы о том, что две независимые выборки $x = (x_1, x_2, \dots, x_{n1})$ и $y = (y_1,$

y_2, \dots, y_{n2}) получены из одной генеральной совокупности, т.е. функции распределения $F_1(x)$ и $F_2(y)$ двух генеральных совокупностей одинаковы (генеральные совокупности *однородны*). Для использования теста должны выполняться предположения о независимости выборок и представленности данных в порядковой шкале.

Эмпирическим значением критерия является максимальное расхождение частот

$$D = \max | F_1(x) - F_2(x) |.$$

Значение D вычисляется так, как и в случае одновыборочного теста, но критические значения критерия будут другими (табл. П.7).

Процедура работы с двухвыборочным тестом Колмогорова–Смирнова в пакетах SPSS и STATISTICA не отличается от процедуры использования U -критерия Манна–Уитни, рассмотренной в подразд. 8.3, за исключением того, что вместо *Mann–Whitney U test* выбирается *Kolmogorov–Smirnov test*.

Обратите внимание на то, что если применить тест Колмогорова–Смирнова к данным примера, рассмотренного в подразд. 8.3 (см. табл. 8.8), то статистически значимое различие между рейтингами двух филиалов компании не обнаруживается (табл. 8.11) ни при направленной, ни при ненаправленной альтернативной гипотезе ($p = 0,206$), в отличие от использования U -критерия Манна–Уитни ($p = 0,66$).

Таблица 8.11

Результаты применения теста Колмогорова–Смирнова
к данным табл. 8.8

		RATING
Most Extreme Differences	Absolute	,455
	Positive	,000
	Negative	-,455
Kolmogorov-Smirnov Z		1,066
Asymp. Sig. (2-tailed)		,206

Мы видим, что U -критерий более мощный, т.е. он лучше «чувствует» различия, чем критерий Колмогорова–Смирнова. Но для проверки предположений о характере распределения выборочных данных с целью вывода о возможности использования

того или иного статистического метода обычно применяется критерий Колмогорова–Смирнова именно из-за его меньшей чувствительности, так как в большинстве случаев различающая способность этого критерия оказывается вполне достаточной для решения об обоснованности применения подходящего метода.

8.5. Критерий знаков и T -критерий Уилкоксона

Для сравнения *двух зависимых* выборок, элементы которых представлены в неметрических шкалах, или вследствие других причин, определяющих невозможность использования соответствующих параметрических критериев, применяются *критерий знаков* и *T -критерий Уилкоксона*.

Критерий знаков (*sign test*) используется для проверки гипотезы об однородности генеральных совокупностей по *парно связанным выборкам*. Такая задача возникает, когда по одной и той же выборке сравниваются эффективность двух методов, точность измерения двумя приборами и т.д. Критерий знаков основан на подсчете количества раз, когда элементы одной выборки превышают парные элементы другой выборки ($x_a > x_b$), и наоборот ($x_b > x_a$). Нулевая гипотеза предполагает равенство этих случаев.

Более чувствительным, чем критерий знаков, при сравнении двух связанных выборок является T -критерий Уилкоксона (*Wilcoxon matched pairs test*¹). Его целесообразно использовать в случаях, когда неприменим t -критерий Стьюдента.

Критерий Уилкоксона основан на вычислении разности (d_i) измерений в каждой i -й паре элементов связанных выборок. Затем эти разности ранжируются от меньших к большим по модулю безотносительно знака. Если несколько разностей имеют одинаковое значение, им приписываются средние арифметические значения рангов. Нулевые значения не учитываются. Затем подсчитываются суммы рангов положительных (T^+) и отрицательных разностей (T^-). Меньшая из сумм принимается в качестве эмпирического значения и сравнивается с критическим значением. Нулевая гипотеза отклоняется, если эмпирическое значение *меньше* критического.

¹ Другое название, встречающееся в литературе, — *Wilcoxon signed-rank test*.

Пример 8.6. Предположим, исследуется эффект влияния известных торговых марок (brand name) на восприятие качества продукта. Десяти респондентам предлагается попробовать, сравнить и оценить в порядковой шкале (баллы) два вида алкогольных напитков, один из которых представляется как продукт известного производителя, другой – как продукт компании, недавно вышедшей на рынок. В действительности оба образца идентичны. Полученные данные и результаты их обработки представлены в табл. 8.12.

Таблица 8.12

Данные для оценки эффекта влияния бренда

№ п/п	Балл		d_i	Ранг d_i
	бренд	новый продукт		
1	52	48	-4	-4
2	37	32	-5	-5,5
3	50	52	+ 2	+ 2
4	45	32	-13	-9
5	56	59	+ 3	+ 3
6	51	50	-1	-1
7	40	29	-11	-8
8	59	54	-5	-5,5
9	38	38	0	-
10	40	32	-8	-7

Сумма отрицательных разностей $T^- = 40$, сумма положительных разностей $T^+ = 5$. Следовательно, за эмпирическое значение критерия Уилкоксона принимается $T = 5$. Критическое значение находится из табл. П.8 при $\alpha = 0,05$ и n , равном количеству ненулевых разностей (для данного примера $n = 10 - 1 = 9$). Критическое значение критерия равно шести. В этом тесте эмпирическое значение должно быть *меньше* критического для отклонения нулевой гипотезы. Следовательно, нулевая гипотеза отклоняется, и делается содержательный вывод о статистически значимом влиянии бренда на восприятие качества продукта.

В пакете SPSS расчет критерия знаков и критерия Уилкоксона осуществляется с помощью процедуры *2 Related Samples* (из меню *Analyze, Nonparametric Tests*). В окне *Two-Related-Samples Tests* необходимо в поле *Test Pair(s) List* перенести переменные, соответствующие двум связанным выборкам, из списка пере-

менных справа (для переноса следует выделить *две* переменные). В разделе *Test Type* нужно выбрать требуемый тест: критерий знаков (*Sign*) или критерий Уилкоксона (*Wilcoxon*). Щелчок по кнопке *OK* приводит к генерации выходной формы с отчетами (табл. 8.13).

Таблица 8.13

Статистика критерия Уилкоксона

Ranks

		N	Mean Rank	Sum of Ranks
NEW - BRAND	Negative Ranks	7	5,71	40,00
	Positive Ranks	2	2,50	5,00
	Ties	1		
	Total	10		

Test Statistics

		NEW - BRAND
	Z	-2,075
	Asymp. Sig. (2-tailed)	,038

В отчете Ranks указаны количество отрицательных (7), положительных (2) и нулевых (1) разностей, средние ранги (Mean Rank) и суммы отрицательных и положительных разностей (Sum of Ranks) (40 и 5 соответственно).

В отчете Test Statistics указан *p*-уровень для двусторонней проверки критерием Уилкоксона (*Asymp. Sig. (2-tailed)*). Он меньше выбранного уровня значимости ($\alpha = 0,05$), что позволяет отклонить нулевую гипотезу.

В отчетах Frequencies и Test Statistics (табл. 8.14) приведены результаты использования критерия знаков для рассматриваемого примера. Как видим, этот критерий не обнаружил статистически значимых различий: *p*-уровень (*Exact Sig. (2-tailed)*) превышает $\alpha = 0,05$, что не позволяет отклонить нулевую гипотезу, что служит примером более низкой чувствительности этого критерия по сравнению с критерием Уилкоксона. Поэтому при решении подобных задач следует ориентироваться на результаты, полученные с помощью критерия Уилкоксона, а не критерия знаков.

Статистика критерия знаков

Frequencies

		N
NEW - BRAND	Negative Differences	7
	Positive Differences	2
	Ties	1
	Total	10

Test Statistics

	NEW - BRAND
Exact Sig. (2-tailed)	,180

8.6. Критерии Краскела–Уоллиса и χ^2 Фридмана

Критерий Краскела–Уоллиса (*Kruskal–Wallis test*) может использоваться как альтернатива однофакторному дисперсионному анализу, поэтому еще одним названием этого критерия является следующее: однофакторный дисперсионный анализ Краскела–Уоллиса (*Kruskal–Wallis one-way analysis of variance*). Он может применяться для трех и более независимых выборок с данными в порядковой или интервальной шкале, если не удовлетворяются предположения об использовании дисперсионного анализа.

Критерий Краскела–Уоллиса является обобщением критерия Манна–Уитни для случая, когда исследуются более двух выборок по уровню выраженности признака. Как и критерий Манна–Уитни, критерий Краскела–Уоллиса основан на положении о том, что чем меньше совпадений, тем больше различаются сравниваемые выборки. Как и в случае критерия Манна–Уитни, все значения сравниваемых выборок представляются в виде одной последовательности упорядоченных значений с последующим вычислением среднего ранга для каждой из выборок.

Эмпирическое значение критерия определяется как

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1),$$

где k – количество анализируемых выборок;

R_j – сумма рангов для j -й выборки;

N – суммарная численность всех анализируемых выборок;

n_j – объем j -й выборки.

Если объем каждой изучаемой выборки больше пяти или количество выборок больше трех, то эмпирическое значение критерия сравнивается с критическим (табличным) значением χ^2 для $df = k - 1$, иначе эмпирическое значение сравнивается с критическим значением критерия Краскела–Уоллиса, приведенным в специальных таблицах. Такие таблицы представлены во многих доступных учебных изданиях по статистике.

Пример 8.7. Определяется, есть ли различие в количестве покупок, совершаемых ежедневно в трех магазинах, принадлежащих одной компании (табл. 8.15).

Таблица 8.15

Количество покупок в магазинах

Магазин 1	Ранг	Магазин 2	Ранг	Магазин 3	Ранг
60	1	80	5	90	8,5
70	2,5	90	8,5	90	8,5
80	5	80	5	110	14
70	2,5	100	11,5	100	11,5
90	8,5	110	14	140	18
110	14	130	16,5	130	16,5
R_j	33,5	R_j	60,5		77

Эмпирическое значение критерия $H = 5,66$. Критическое значение хи-квадрат для $df = 2$ и $\alpha = 0,05$ составляет 5,99. Следовательно, нулевая гипотеза об отсутствии различий не может быть отклонена при заданном уровне значимости.

Для расчета критерия Краскела–Уоллиса в пакетах SPSS и STATISTICA данные должны быть сгруппированы в одну переменную, содержащую значения признака (для рассмотренного примера – переменная *sale*), и группирующую переменную, содержащую коды выборок (для данного примера – переменная *group*, включающая значения 1–3). В SPSS из меню *Analyze, Nonparametric Tests* выбирается процедура *K Independent Samples*. В открывшемся диалоговом окне *Tests for Several Independent Samples* необходимо в поле *Test Variable List* перенести зависимую переменную (*sale*), а в поле *Grouping Variable* – группирующую переменную (*group*). В разделе *Test Type* нужно отметить *Kruskal-Wallis H*, а в окне *Define Range* – указать минимальное и максимальное значения группирующей переменной (в данном случае 1 и 3).

Результаты расчетов содержатся в отчетах *Ranks* и *Test Statistics* (табл. 8.16). Показатели, приведенные в отчетах, аналогичны тем, которые даны в отчетах других непараметрических тестов, которые были рассмотрены в предыдущих подразделах, поэтому на них мы останавливаться не будем. Отметим лишь, что эмпирическое значение критерия, рассчитанное в SPSS (5,765), несколько выше, чем рассчитанное вручную (5,66). Это связано с тем, что в пакете используется дополнительная корректировка критерия в случае, если несколько элементов выборки имеют один ранг. Тем не менее полученное значение p -уровня (0,056) также не позволяет отклонить нулевую гипотезу.

Таблица 8.16

Статистика критерия Краскела–Уоллиса в SPSS

Ranks

	GROUP	N	Mean Rank
SALE	1	6	5,58
	2	6	10,08
	3	6	12,83
	Total	18	

Test Statistics

	SALE
Chi-Square	5,765
df	2
Asymp. Sig.	,056

Аналогичные вычисления в пакете STATISTICA осуществляются в модуле *Nonparametrics/Distrib.* с помощью процедуры *Kruskal–Wallis ANOVA* (выбирается из меню *Analysis, Startup Panel*). Работа с этой процедурой аналогична работе с процедурой *Mann–Whitney U test*. Результаты (рис. 8.7) идентичны результатам, полученным в SPSS.

Как и в случае использования ANOVA, при отклонении нулевой гипотезы по критерию Краскела–Уоллиса нельзя сделать конкретный вывод о направлении различий. В такой ситуации можно применять критерий Манна–Уитни для парного сравнения выборок.

Kruskal-Wallis ANOVA by Ranks (kruskal.sta)			
NONPAR	Independent (grouping) variable: GROUP		
STATS	Kruskal-Wallis test: $H(2, N=18) = 5,765296$ $p = ,0560$		
Depend. :		Valid:	Sum of
SALE	Code	N	Ranks
Group 1	1	6	33,50000
Group 2	2	6	60,50000
Group 3	3	6	77,00000

Рис. 8.7. Статистика критерия Краскела–Уоллиса в пакете STATISTICA

Для сравнения *более двух зависимых* выборок используется критерий *хи-квадрат Фридмана (Friedman test)*, который является непараметрическим аналогом однофакторного дисперсионного анализа с повторными измерениями. Этот критерий может быть даже более эффективен, чем ANOVA, при небольших выборках. Работа с критерием *хи-квадрат Фридмана* в статистических пакетах аналогична работе с рассмотренными выше методами. В SPSS эта процедура носит название *K Related Samples*, в STATISTICA – *Friedman ANOVA*. В случае отклонения нулевой гипотезы по критерию *хи-квадрат Фридмана* для парного сравнения выборок используется *T-критерий Уилкоксона*.

Задания для самостоятельного выполнения

1. Автодилер исследует предпочтения покупателей, в частности связана ли повторная покупка автомобиля со страной производителя (страна 1 и страна 2). Результаты исследований (количество покупок) приведены в табл. 8.17. При проверке гипотезы примите $\alpha = 0,05$.

Таблица 8.17

Данные для решения задачи 1

	Производитель	
	страна 1	страна 2
Повторная покупка	40	20
Первая покупка	8	32

Пояснение. Нулевая гипотеза в этом задании может быть сформулирована следующим образом: предпочтения покупателем страны-изготовителя не связаны с тем, имел ли он машину ранее, т.е. наблюдаемые частоты равны ожидаемым: $H_0: O_i = E_i$. Для проверки этой гипотезы может быть использован критерий χ^2 (объясните, почему). Критическое значение критерия: 3,84 ($df = 1, \alpha = 0,05$); эмпирическое значение критерия: 20,94. В случае использования поправки на непрерывность получим 19,11. В любом случае нулевая гипотеза отклоняется.

2. Определите, существует ли статистически значимое различие между студентами и студентками в их предпочтениях по посещению «Университетского ланч-клуба» (по данным, приведенным в табл. 8.18).

Таблица 8.18

Данные для решения задачи 2

	Посещают	Не посещают	Всего
Мужчины	26	24	50
Женщины	35	15	50
Всего	61	39	100

3. Предложите варианты заданий 1 и 2, при которых должны использоваться другие непараметрические критерии проверки статистических гипотез. Поясните ответ.

Часть третья

Корреляционно-регрессионный анализ и статистическое прогнозирование

Цель расчетов – не числа, а понимание.

Р. Хэмминг

Глава 9

Меры связи и их проверка на значимость

9.1. Виды связей исследуемых признаков

Статистические критерии проверки гипотез (см. вторую часть) позволяют выявить наличие статистически значимых связей, но не оценить их степень (величину). Поэтому в исследовательской практике очень часто определяется мера связи между двумя или несколькими признаками, для чего, например, вычисляются коэффициент корреляции, коэффициент детерминации, строятся регрессионные уравнения разного вида и т.д.

Основой экономико-статистического прогнозирования является корреляционно-регрессионный анализ. Первоначально методы корреляционного анализа разрабатывались в биологии (вторая половина XIX – начало XX в.), а позднее распространились и на другие области, в том числе на социально-экономическую сферу исследований. Почти одновременно с методами корреляционного анализа начали разрабатываться и использоваться методы регрессионного анализа. В настоящее время для анализа бизнес-ситуаций и принятия управленческих решений на основе эмпирических исследований наиболее часто используются именно методы корреляционно-регрессионного анализа.

Анализ и прогнозирование состояния изучаемого явления предполагают его изучение во взаимосвязи с другими явлениями или величинами. «Основной» изучаемый признак называют *зависимой переменной* или *результативным признаком* (y), другие, связанные с y , признаки называют *независимыми переменными* или *факторными признаками*, или *факторами* (x_1, \dots, x_n). Результативный и факторный признаки могут быть представлены в различных шкалах, и от вида шкалы будет зависеть, какие методы можно использовать для установления наличия связи, оценки степени связи, определения ее вида и т.д.

Выделяют *функциональные (жестко детерминированные)* и *вероятностные (стохастически детерминированные)* связи. Функциональные связи называют также *явными*, поскольку они связывают показатели, которые получены вычислением по заранее известным формулам и законам. Поэтому значение зависимой переменной становится известным, как только известны значения независимых переменных. Примерами являются связь между стоимостью единицы товара и ценой, уплаченной за партию; зависимость производительности труда и затрат рабочего времени, индексы темпов роста и т. д.

Жесткий детерминизм связей проявляется в том, что любое действие вызывает строго определенный результат; случайными (непредвиденными заранее) воздействиями при этом пренебрегают. Поэтому при заданных начальных условиях состояние такой системы может быть определено с вероятностью, равной 1.

Связь признаков y и x называется *функциональной*, если каждому возможному значению независимого признака x_i соответствует одно или несколько строго определенных значений зависимого признака y_i :

$$y_i = f(x_i),$$

где y_i – результативный признак;
 x_i – факторный признак, $i = 1, \dots, n$.

Характерной особенностью функциональных связей является то, что в каждом случае известны полный перечень факторов, определяющих значение результативного признака, а также точный механизм их влияния, выраженный определенным уравнением.

В реальной действительности ввиду неполноты информации часто возникает неопределенность в характере воздействия фак-

торов на результативную переменную и даже в наличии самих влияющих факторов. В этом случае связь между признаками называется *вероятностной* или *стохастической*. Это означает, что результативный признак кроме рассматриваемых факторов подвержен влиянию ряда неучтенных или неконтролируемых (случайных) факторов, а также некоторых ошибок измерения. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть определены абсолютно точно, но могут быть предсказаны с определенной вероятностью.

При стохастических связях не известен ни полный перечень факторов, определяющих значение результативного признака, ни точный механизм их взаимодействия с результативным признаком, т.е. всегда имеет место влияние случайных факторов. Модель стохастической связи может быть представлена в общем виде уравнением

$$\hat{y} = f(x_i) + \varepsilon,$$

где \hat{y} — прогнозируемое значение результативного признака;
 $f(x_i)$ — функция, отражающая влияние известных и учтенных факторных признаков ($i = 1, \dots, n$);
 ε — влияние неконтролируемых или неучтенных факторов, а также случайных ошибок измерения.

При стохастических связях по значениям известных факторов можно установить лишь некоторую «среднюю» тенденцию в значениях зависимой переменной. Так, между ростом человека и его весом существует некоторая зависимость, но эта зависимость не жестко детерминирована, она действует для «среднего» человека. При анализе стохастических связей зависимая переменная рассматривается как случайная величина, а независимые переменные можно прямо или косвенно контролировать.

В зависимости от направления функциональные и стохастические связи могут быть *прямыми* и *обратными*. При прямой связи направление изменения результативного признака совпадает с направлением изменения признака-фактора, т.е. с увеличением факторного признака увеличивается и результативный, и, наоборот, с уменьшением факторного признака уменьшается и результативный признак. В противном случае между рассматриваемыми величинами существуют обратные связи. Например, чем выше квалификация рабочего (разряд), тем выше уровень

производительности труда — прямая связь. А чем выше производительность труда, тем ниже себестоимость единицы продукции — обратная связь.

По аналитическому выражению (форме) связи могут быть *прямолинейными* и *криволинейными*. При прямолинейной связи с возрастанием значения факторного признака происходит непрерывное возрастание (или убывание) значений результативного признака. Математически такая связь представляется уравнением прямой, а графически — прямой линией. Отсюда ее другое название — *линейная связь*. При криволинейных связях с возрастанием значения факторного признака возрастание (или убывание) результативного признака происходит неравномерно или же направление его изменения меняется на обратное. Геометрически такие связи представляются кривыми линиями (гиперболой, параболой и т.д.).

По количеству факторов, действующих на результативный признак, связи классифицируются как *одно-* и *многофакторные* (два и более факторов). Однофакторные, или *простые*, связи обычно называются *парными* (так как рассматривается пара признаков), например корреляционная связь между прибылью и производительностью труда. В случае моделирования многофакторной связи предполагается, что изучаемая переменная зависит от нескольких факторов. Например, спрос зависит от цены и уровня доходов населения; объем выпускаемой продукции зависит от объема инвестиций и числа рабочих, занятых в процессе производства; производительность труда зависит от уровня организации труда, уровня автоматизации производства, квалификации рабочих, их производственного стажа и др.

Частным случаем стохастической связи является *корреляционная связь* (лат. *correlatio* — соотношение). При такой связи среднее значение (математическое ожидание) случайной величины результативного признака у изменяется в зависимости от изменения другой случайной величины x_i или набора случайных величин x_1, x_2, \dots, x_n . Корреляционная связь отражает изменение средней величины результативного признака при изменении средних величин факторных признаков, тогда как в общем случае стохастическая связь может проявляться и в изменениях других характеристик изучаемых признаков.

Корреляционные связи — это связи соотносительные, а не причинные, т.е. они не обязательно отражают влияние (зависи-

мость) одного фактора на другой. Под *причинной зависимостью* понимается такая связь между явлениями и процессами, когда изменение одного из них является следствием изменения другого. Корреляционные методы представляют лишь инструмент для изучения *статистической зависимости*. Представления и гипотезы о причинной связи могут быть получены с помощью теоретического анализа, *содержательно* объясняющего изучаемое явление.

Особое значение при исследовании причинно-следственных связей имеет выявление временной последовательности: причина всегда должна предшествовать следствию, однако не каждое предшествующее событие следует считать причиной, а последующее следствием.

9.2. Сущность и задачи корреляционного анализа

Корреляция – это статистическая зависимость между случайными величинами, не имеющая строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Парная корреляция характеризует связь между двумя признаками (результативным и факторным или двумя факторными). *Частная корреляция* характеризует связь между признаками при исключении влияния (фиксированном значении) других признаков. *Множественная корреляция* характеризует связь результативного признака и двух или более факторных признаков.

Корреляционный анализ является одним из методов статистического анализа взаимосвязи нескольких признаков и используется для количественной оценки взаимосвязи двух наборов данных. Корреляционный анализ дает возможность установить, ассоциированы ли наборы данных по величине, т.е. большие значения из одного набора данных связаны с большими значениями другого набора (*положительная корреляция*), или, наоборот, малые значения одного набора связаны с большими значениями другого (*отрицательная корреляция*), или данные двух диапазонов никак не связаны (*нулевая корреляция*).

Корреляционное поле, или *диаграмма распределения*, – эмпирическое распределение значений исследуемых признаков, представленное на графике. Корреляционное поле является вспомогательным средством при анализе выборочных данных. По характеру расположения точек поля можно составить предвари-

тельное мнение о форме зависимости случайных величин (например, о том, что одна величина в среднем возрастает или убывает при возрастании другой). На рис. 9.1 представлены примеры корреляционных полей, отражающих различные виды связей.

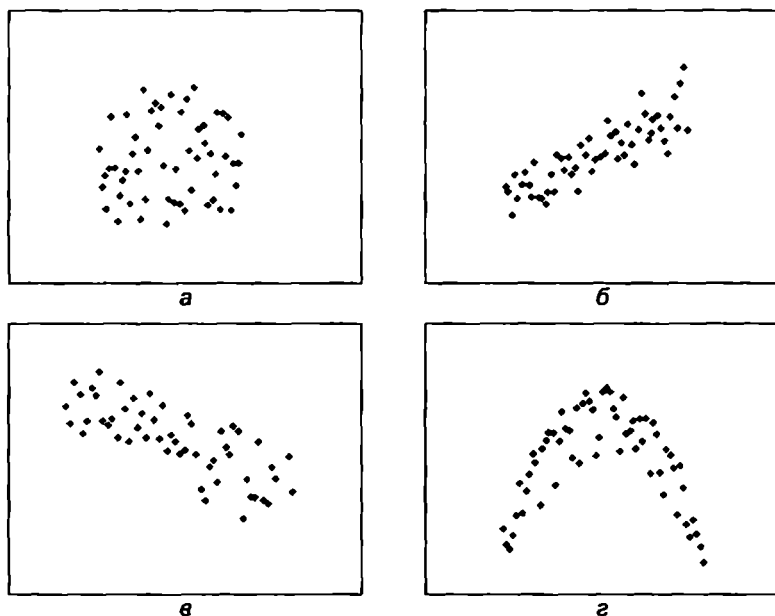


Рис. 9.1. Примеры корреляционных полей:
а – связь отсутствует; б – связь сильная, линейная, прямая;
в – связь сильная, линейная, обратная; г – связь сильная,
криволинейная

В практике социально-экономических исследований довольно часто определяется *частная корреляция*¹. Это необходимо тогда, когда изучается связь, например, между тремя переменными x , y , z (рис. 9.2) и нужно оценить связь между двумя признаками без учета влияния третьего признака на другие (например, изучается связь x и y без учета влияния z).

Выявление частной корреляции оказывается полезным во многих случаях. Часто обнаруживаются так называемые ложные корреляции, т.е. логически необъяснимые и даже абсурдные.

¹ Термин был впервые введен в работе Юла (Yule) в 1907 г.

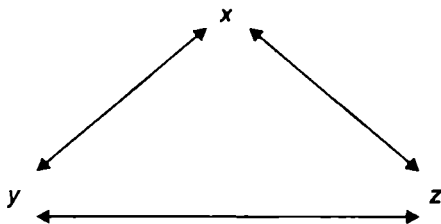


Рис. 9.2. Возможные взаимосвязи между признаками x , y , z

Конкретным примером может служить результат исследования, обнаружившего связь размера обуви с уровнем образованности респондентов. Была обнаружена также корреляция между ростом и уровнем образованности. Отсюда можно сделать нелепый вывод о том, что человек тем более образован, чем большего размера обувь он носит или чем он выше ростом. С другой стороны, существует вполне объяснимая и логичная связь между ростом и размером обуви. Эти показатели связаны с еще одной переменной – возрастом. Вместе эти корреляции приводят к ложным корреляциям «образованность – размер обуви», «образованность – рост». В результате статистически значимый коэффициент корреляции оказывается индикатором не некоторой причинной связи между переменными «образованность» и «размер обуви», а обусловлен влиянием на эти переменные «третьей» переменной (возрастом респондентов). Для исключения «искажающих» переменных и необходим расчет коэффициентов частной корреляции. В данном случае коэффициенты частной корреляции «образованность – рост» и «образованность – размер обуви» при исключенной переменной «возраст» оказываются близкими к нулю.

Достаточно высокое значение коэффициента меры связи рассматривается многими исследователями как индикатор сильной связи, что как будто предполагает ее безусловное присутствие. Однако, с другой стороны, для корректного анализа необходимо не только определить вид, степень и направление связи, но и проверить статистическую значимость полученного результата: существует ли связь, обнаруженная в выборке и характеризуемая коэффициентом корреляции, во всей генеральной совокупности, из которой эта выборка была извлечена. Поэтому корреляционный анализ направлен на решение двух задач:

1) количественная оценка силы (тесноты) связи между двумя признаками (при парной связи) и между результативным признаком и множеством факторных признаков (при многофакторной связи);

2) проверка статистических гипотез о наличии и силе корреляционной связи.

Поэтому корреляционный анализ должен основываться как на показателях тесноты связи, так и на показателях ее статистической значимости.

Проверяя любую статистическую гипотезу, исследователь ищет ответ на вопрос, существуют ли в генеральной совокупности те характерные особенности, которые наблюдаются в выборках. Проверке на значимость могут (и должны) быть подвергнуты и меры связи. Это абсолютно необходимая процедура с точки зрения научной строгости анализа и корректности представления исследовательских результатов.

Величина (сила) связи и ее значимость (достоверность) представляют две различные характеристики связи. Тем не менее нельзя сказать, что они совершенно независимы. В общем случае чем сильнее связь, тем более значимой она является. Это связано с тем, что если предполагать отсутствие связи между соответствующими переменными в генеральной совокупности, то наиболее вероятно ожидать, что в исследуемой выборке связь между этими переменными также будет отсутствовать. Таким образом, чем более сильная связь обнаружена в выборке, тем менее вероятно, что этой связи нет в генеральной совокупности, из которой она извлечена.

Однако указанная зависимость между силой связи и ее значимостью имеет место только при фиксированном объеме выборки, поскольку при различных объемах выборки связь одинаковой силы может оказаться как высоко значимой, так и незначимой вовсе. Поэтому, определив меру связи, необходимо всегда указывать и уровень значимости, например $p < 0,01$.

Согласно общей логике проверки статистических гипотез, в данном случае нулевая гипотеза утверждает, что для генеральной совокупности, из которой была извлечена выборка, связь между переменными полностью отсутствует, т.е. значение проверяемого показателя меры связи для генеральной совокупности равно нулю. Дальнейшая логика проверки на статистическую значимость аналогична общей последовательности этапов проверки любой статистической гипотезы.

Базовым условием применения корреляционного анализа является предположение о случайном характере выборки из генеральной совокупности. Однако на практике исследователи иногда отклоняются от этого условия, используя систематические выборки (см. подразд. 3.2) или другие схемы отбора исходных данных, нарушающие принципы равновероятностного отбора. Тем не менее результаты корреляционного анализа и в этом случае могут быть очень полезны для выявления связей между признаками. Поэтому на практике такое отклонение считается не критическим для проведения корреляционного анализа.

В зависимости от вида шкал, в которых измерены эмпирические данные, используются разные методы определения степени и направления связи. Например, для номинативных шкал в качестве меры связи вычисляют коэффициент ϕ (фи) и коэффициент V Крамера, для порядковых – коэффициент γ (гамма), ρ -коэффициент Спирмена, τ -коэффициент Кендалла. Для метрических шкал вычисляется коэффициент корреляции (r) Пирсона. В качестве критерия для проверки значимости используется t -критерий (для ρ и r) или z -критерий (для γ и τ). Особенности использования этих критериев рассматриваются ниже.

• 9.3. Меры связи для переменных в метрических шкалах

Для оценки тесноты корреляционной связи между двумя признаками x и y по данным, представленным в метрической шкале, используется *коэффициент корреляции Пирсона*. Это один из наиболее широко используемых статистических показателей. Коэффициент корреляции Пирсона вычисляется по формуле

$$r_{xy} = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{(n - 1)s_x s_y},$$

где \bar{x}, \bar{y} – средние значения признаков;
 s_x, s_y – их стандартные отклонения по выборке.

По знаку коэффициента корреляции можно судить о направлении связи (положительное значение соответствует прямой связи, отрицательное – обратной). По абсолютной величине

коэффициента корреляции можно судить о степени связи между признаками (табл. 9.1).

Таблица 9.1

Интерпретация модуля коэффициента корреляции

Связь	Модуль коэффициента корреляции
Отсутствует	0 – 0,1
Слабая	0,1 – 0,4
Заметная (умеренная)	0,4 – 0,7
Сильная (высокая)	0,7 – 0,9
Значительная (очень высокая), близкая к функциональной	0,9 – 1

Коэффициент корреляции, возведенный в квадрат, называется *коэффициентом детерминации* и показывает, в какой степени (на сколько процентов) изменчивость одного признака определяется (детерминируется) изменчивостью другого, т.е. на сколько процентов резульативный признак зависит от анализируемого признака. Коэффициент детерминации в отличие от коэффициента корреляции отражает связь между признаками линейно, он является аддитивным, т.е. допускается его усреднение для нескольких выборок.

Значение коэффициента корреляции не зависит от единиц измерения и выбора начала отсчета. Это означает, что если переменные x и y уменьшить или увеличить в k раз либо на одно и то же число c , то коэффициент корреляции не изменится. Коэффициент корреляции определяет степень *линейной* зависимости. Если связь между признаками существует, но она имеет нелинейный характер (см., например, рис. 9.1г), коэффициент корреляции может иметь значение, близкое к 0.

Для определения коэффициента корреляции и выполнения проверки на значимость требуется принять ряд допущений. Предполагается следующее: во-первых, метрический характер измеряемых данных (данные представлены в интервальной шкале или шкале отношений); во-вторых, обе переменные подчиняются нормальному закону распределения; в-третьих, зависимость между переменными приблизительно линейна; в-четвертых, *гомоскедастичность*, т.е. однородность дисперсий. Это понятие означает, что дисперсия значений у равномерна для всех значе-

ний x . На графике в осях $x - y$ это должно проявляться следующим образом: значения y должны быть приблизительно равномерно распределены выше и ниже линии регрессии $y(x)$ по всей ее длине, т.е. точки на графике должны образовывать достаточно симметричную овалообразную форму без значительных выбросов. Существенные признаки нелинейности или отклонения от гомоскедастичности свидетельствуют о необходимости использовать другую меру связи и другой критерий значимости.

Предположим, что все допущения выполняются и коэффициент корреляции определен. Тогда нулевая гипотеза $H_0: r_{\text{попул}} = 0$ утверждает отсутствие линейной связи для генеральной совокупности; альтернативная гипотеза утверждает наличие такой связи ($H_A: r_{\text{попул}} \neq 0$ для ненаправленной альтернативной гипотезы; $H_A: r_{\text{попул}} > 0$ или $H_A: r_{\text{попул}} < 0$ для направленной альтернативной гипотезы).

Для оценки статистической значимости коэффициента корреляции (r) по выборке из n элементов используется следующая формула для определения эмпирического значения t -критерия:

$$t = r \sqrt{\frac{n-2}{1-r^2}}.$$

Критическое значение t -критерия вычисляется с помощью функции Excel СТЬЮДРАСПОБР(α , df) или определяется из таблицы значений t -распределения для выбранного уровня значимости α и числа степеней свободы $df = n - 2$ (табл. П.2). Затем на основе сравнения критического и экспериментального значений принимается решение об отклонении или невозможности отклонения нулевой гипотезы по тем же правилам, что и для других критериев (см. вторую часть этого пособия).

Аналогично можно оценить значимость коэффициента частной корреляции. Если изучается связь между тремя переменными x , y , z и нужно оценить связь между, например, x и y без учета влияния z , то коэффициент частной корреляции вычисляется по формуле

$$r_{xy-z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}.$$

Нулевая гипотеза $H_0: r_{xy-z} = 0$. Эмпирическое значение t -критерия для проверки этой гипотезы определяется как

$$t = \frac{r_{xy-z} \sqrt{n-3}}{\sqrt{1-r_{xy-z}^2}}.$$

Критическое значение t -критерия определяется для выбранного уровня значимости α и числа степеней свободы $df = n - 3$.

Квадрат коэффициента частной корреляции r_{xy-z} называется *коэффициентом частной детерминации*. Он определяет долю дисперсии, которая объясняется влиянием y на x (или x на y) при фиксированном значении z .

Коэффициент *парной корреляции* показывает тесноту связи между двумя исследуемыми признаками, а коэффициент *множественной корреляции* характеризует степень влияния нескольких признаков на результирующий признак. Расчет коэффициента множественной корреляции осуществляется после построения регрессионного уравнения (гл. 10).

В состав Microsoft Excel входит набор средств анализа данных («Сервис», «Анализ данных»), предназначенный для решения статистических, экономических, инженерных и других задач. С целью расчета коэффициента корреляции для нескольких признаков с помощью этого пакета (модуль «Корреляция») следует указать входные данные и выбрать параметры представления результатов. Для определения коэффициента парной корреляции в Excel используется функция КОРРЕЛ(X, Y), аргументами которой являются массивы данных по паре исследуемых признаков X, Y .

Для расчета коэффициента парной корреляции в SPSS необходимо выбрать пункт меню *Analyze* (анализ), затем *Correlate* (корреляция) и *Bivariate* (парная). В открывшемся диалоговом окне из списка переменных выбираются две переменные (или несколько переменных) для анализа. По умолчанию установлены параметры *Pearson* (расчет коэффициента корреляции Пирсона) и *Flag significant correlation* (выделение статистически значимых коэффициентов). Результаты расчетов представляются в отчете *Correlations* в виде таблицы, в которой указываются коэффициенты корреляции для каждой пары выбранных переменных (строка *Pearson Correlation*), их статистическая значимость (*Sig*) и число наблюдений, по которому вычисляется коэффициент корреляции (N).

Параметры расчета коэффициента частной корреляции в пакете SPSS устанавливаются в диалоговом окне *Partial* (меню *Analyze, Correlate*). Из списка доступных переменных выбираются две переменные, между которыми обнаружена «странная» корреляция. Они переносятся в поле *Variables* (переменные). Переменную или переменные, с которыми коррелируют оба исследуемых признака и влияние которой или которых нужно исключить, переносят в поле *Controlling for*. Результаты представляются в отчете *Partial Corr*, где указываются коэффициент частной корреляции для пары выбранных переменных без учета влияния одной или нескольких исключенных переменных, а также его статистическая значимость.

Построить корреляционное поле и проверить предположение о линейности и гомоскедастичности можно с помощью графических инструментов в SPSS, например *Graphs – Scatter – Simple* (в поля *Y Axis* и *X Axis* переносятся анализируемые переменные).

Поскольку корреляционный анализ на практике используется для выявления сильных и статистически значимых связей, то такие предпосылки и допущения по его применению, как линейность связи, нормальность распределения и гомоскедастичность, проверяются лишь для тех признаков, для которых такие связи обнаружены. Если эти допущения не выполняются, то полученные результаты нельзя считать корректными.

9.4. Меры связи для переменных в неметрических шкалах

В случае, если хотя бы одна из переменных представлена в порядковой шкале, могут использоваться коэффициенты ρ Спирмена, γ (гамма) и τ Кендалла. Это непараметрические показатели связи, использование которых не требует обязательного выполнения предпосылки о нормальности распределения данных.

Коэффициент ρ Спирмена интерпретируется аналогично коэффициенту корреляции Пирсона и может принимать значения в таком же диапазоне (от -1 до $+1$). Фактически это и есть коэффициент корреляции Пирсона, вычисленный по рангам переменных (после предварительного ранжирования). При отсутствии повторяющихся рангов коэффициент ρ Спирмена определяется по формуле

$$\rho = \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

где d_i – разность рангов двух признаков i -го элемента выборки.

Проверка статистической значимости коэффициента ρ Спирмена ничем не отличается от аналогичной процедуры для коэффициента Пирсона (при числе наблюдений в выборке 10 и больше распределение выборочных значений ρ Спирмена приближается к t -распределению, поэтому для проверки используется t -критерий).

Коэффициент γ основан на проверке различия между вероятностью того, что наблюдаемые данные расположены в том же порядке для двух величин (*совпадение*), и вероятностью того, что они расположены в другом порядке (*инверсия*). Для вычисления этого коэффициента между элементами двух выборок попарно определяется число совпадений (P) и инверсий (Q).

Коэффициент γ вычисляется по формуле

$$\gamma = \frac{P - Q}{P + Q}.$$

Коэффициент τ Кендалла является обобщением коэффициента γ на случай возникновения одинаковых рангов (*tied ranks*) у элементов. Формула для вычисления этого коэффициента включает поправки к коэффициенту γ , но принцип расчета остается тем же. Обычно значения коэффициентов Спирмена и Кендалла различаются, так как у них разные принцип и процедура вычисления. Соотношение между этими коэффициентами выражается неравенством

$$-1 \leq 3\tau - 2\rho \leq 1.$$

Значения коэффициентов Пирсона, Спирмена и Кендалла могут быть определены с помощью статистических пакетов. Для этого, например, в пакете SPSS из меню *Analyze* выбирается процедура *Correlate*. Затем для расчета парной корреляции выбирается опция *Bivariate*, где указывается пара переменных для расчета и выбираются нужный коэффициент, а также вид проверки на значимость (одно- или двунаправленная). Для расчета частной корреляции выбирается опция *Partial*.

Пример 9.1. Рассмотрим следующую задачу. По выборке из восьми семей (табл. 9.2) необходимо определить, существует ли связь между престижностью работы отца и престижностью работы детей (старших сыновей и дочерей). Оценка престижности представлена в баллах (порядковая шкала), поэтому нужно использовать непараметрический метод определения меры связи. Воспользуемся для решения задачи коэффициентом ρ Спирмена, для расчета которого будем применять соответствующие процедуры пакета SPSS.

Таблица 9.2

Оценки престижности работы членов семьи, баллы

Семья	Престижность работы отца	Престижность работы сына	Престижность работы дочери
1	80	85	82
2	78	80	77
3	75	70	68
4	70	75	77
5	69	72	60
6	66	60	52
7	64	48	48
8	52	55	57

Результаты расчетов представлены в табл. 9.3. Здесь переменные father, son, daughter соответствуют исследуемым признакам (престижность работы отца, сына и дочери).

Таблица 9.3

Коэффициенты ρ Спирмена

		FATHER	SON	DAUGHTER
FATHER	Correlation Coefficient	1,000	,905	,886
	Sig. (2-tailed)	,	,002	,003
	N	8	8	8
SON	Correlation Coefficient	,905	1,000	,946
	Sig. (2-tailed)	,002	,	,000
	N	8	8	8
DAUGHTER	Correlation Coefficient	,886	,946	1,000
	Sig. (2-tailed)	,003	,000	,
	N	8	8	8

Как видим, обнаруживается сильная ($p > 0,88$) и значимая ($p < 0,01$) связь между престижностью работы всех членов семьи. Для сравнения определим коэффициенты корреляции Пирсона (табл. 9.4). Они также демонстрируют наличие сильной ($r > 0,73$) и значимой ($p < 0,05$) связи между всеми исследуемыми признаками. Это может рассматриваться в качестве косвенного свидетельства о том, что используемая шкала, в которой измеряется престижность работы, близка к интервальной.

Таблица 9.4

Коэффициенты корреляции Пирсона

		FATHER	SON	DAUGHTER
FATHER	Pearson Correlation	1	,825	,733
	Sig. (2-tailed)	,	,012	,038
	N	8	8	8
SON	Pearson Correlation	,825	1	,927
	Sig. (2-tailed)	,012	,	,001
	N	8	8	8
DAUGHTER	Pearson Correlation	,733	,927	1
	Sig. (2-tailed)	,038	,001	,
	N	8	8	8

Но если мы обратимся к результатам расчетов коэффициентов частной корреляции (табл. 9.5), то увидим, что статистически значимая связь между престижностью работы отца и престижностью работы детей отсутствует даже на уровне значимости 0,1. Сохраняется лишь сильная частная корреляция между престижностью работы сына и дочери ($r_{xy-z} = 0,84$; $p < 0,05$).

Для оценки силы связи между двумя переменными в номинальных шкалах наиболее часто используется коэффициент ϕ (ϕ_i). Его вычисление основано на статистике χ^2 для таблиц сопряженности 2×2 (см. гл. 8):

$$\phi = \sqrt{\frac{\chi^2}{n}},$$

где n — общее количество наблюдений.

Значения коэффициента ϕ изменяются от 0 (нет зависимости между переменными, $\chi^2 = 0$) до 1 (абсолютная зависимость между двумя факторами в таблице). В случае таблиц сопряжен-

Кoeffициенты частной корреляции

Controlling for..	DAUGHTER	
	FATHER	SON
FATHER	1,0000 (0) P = ,	,5706 (5) P = ,181
SON	,5706 (5) P = ,181	1,0000 (0) P =
Controlling for..	SON	
	FATHER	DAUGHTER
FATHER	1,0000 (0) P = ,	-,1492 (5) P = ,750
DAUGHTER	-,1492 (5) P = ,750	1,0000 (0) P = ,
Controlling for..	FATHER	
	DAUGHTER	SON
DAUGHTER	1,0000 (0) P = ,	,8377 (5) P = ,019
SON	,8377 (5) P = ,019	1,0000 (0) P = ,

ности большего размера верхний предел коэффициента ϕ может превысить 1, тогда интерпретация результатов затрудняется. Поэтому используется более общий показатель, который получил название *коэффициент V Крамера (Cramer's V)*:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

В этой формуле k означает наименьшее значение из числа строк и столбцов в таблице сопряженности.

Кoeffициенты ϕ и V Крамера вычисляются в пакете SPSS с помощью процедуры *Crosstabs* (таблицы сопряженности) из меню *Analyze* (анализ) и *Descriptive Statistics* (описательная статистика). После выбора необходимых переменных для формирования строк и столбцов таблицы сопряженности (как это осуществляется при расчете критерия χ^2) следует в окне *Statistics* выбрать необходи-

мые коэффициенты для расчета. В выходной форме представляются значения коэффициентов и p -уровень значимости.

В частности, для примера о связи курения и брака на производстве, рассмотренного в гл. 8 (см. табл. 8.3), имеем следующие результаты (табл. 9.6).

Таблица 9.6

Результаты оценки связи курения и производственного брака

Symmetric Measures

		Value	Approx. Sig.
Nominal by	Phi	,322	,032
Nominal	Cramer's V	,322	,032
N of Valid Cases		66	

- Not assuming the null hypothesis.
- Using the asymptotic standard error assuming the null hypothesis.

Эти результаты демонстрируют слабую значимую связь между двумя изучаемыми признаками: степень связи характеризуется коэффициентами ϕ и V Крамера, которые равны 0,322 ($p < 0,05$).

Задания для самостоятельного выполнения¹

1. По данным табл. 9.7 установите, есть ли статистически значимая связь между объемом инвестиций в производство и ценой за акцию предприятия. Объясните смысл p -уровня значимости для данного случая.

Таблица 9.7

Данные для решения задачи 1

Объем инвестиций, млн руб.	108	4,4	3,5	3,6	39	68,4	7,5	5,5	375	12	51
Цена за акцию, руб.	12	4	5	6	13	19	8,5	5	15	6	12

2. По данным табл. 9.8 определите наличие статистически значимой связи.

¹ Для работы над заданиями 1–5 проверьте выполнение предпосылок с целью применения корреляционного анализа, обоснуйте выбор используемого показателя силы связи.

Данные для решения задачи 2

Доход, тыс. руб.	400	620	980	580	540	850	350	620
Издержки производства, тыс. руб.	65	77	42	49	68	80	50	40

3. По данным табл. 9.9 определите, существует ли статистически значимая связь между размером основных фондов и выпуском продукции.

Таблица 9.9

Данные для решения задачи 3

Размер основных фондов, млн руб.	4,3	5,4	3,6	6,9	3,9	4,7	4,0	6,4	5,5	6,8
Выпуск продукции, тыс. руб.	22,4	18,6	13,1	25,1	10,2	19,2	15,7	23,4	16,0	21,5

4. По данным табл. 9.10 охарактеризуйте связь между явкой избирателей на выборы, уровнем безработицы и зафиксированными случаями нарушения избирательного законодательства в ходе предвыборной кампании. Является ли эта связь статистически значимой? Определите коэффициенты частных корреляций. Сделайте выводы. Какой фактор в большей степени влияет на решение избирателей участвовать в выборах?

Таблица 9.10

Данные для решения задачи 4

Избирательный участок	Явка избирателей, %	Уровень безработицы, %	Количество нарушений
1	45	5	59
2	50	8	63
3	55	9	69
4	58	9	41
5	60	10	58

5. В табл. 9.11 представлены показатели эффективности производства по данным 15 однотипных машиностроительных предприятий [8]. Охарактеризуйте связь этих показателей. Определите коэффициенты частных корреляций. Сделайте выводы.

Данные для решения задачи 5

Выработка валовой продукции на одного работника, млн руб.	Выпуск валовой продукции на 1 руб. среднегодовой стоимости производственных фондов (фондоотдача), руб.	Стоимость материалов в валовой продукции (материалоемкость), %
6,0	2,0	25
4,9	0,8	30
7,0	2,7	20
6,7	3,0	21
5,8	1,0	28
6,1	2,1	26
5,0	0,9	30
6,9	2,6	22
6,8	3,0	20
5,9	1,1	29
5,0	0,8	27
5,6	2,2	25
6,0	2,4	24
5,7	2,2	25
5,1	1,3	30

6. В табл. 9.12 приведена информация о предложении, спросе и цене товара А за 12 мес. прошлого года. Выясните, существуют ли связи между этими показателями, определите их силу. Связи спроса и предложения с ценой оцените с учетом и без учета индекса потребительских цен. Определите коэффициенты частных корреляций. Предложите экономическую интерпретацию полученных результатов. Выполняются ли в данном случае предпосылки для расчета коэффициента корреляции? Объясните смысл p -уровня значимости для данного случая. Обоснуйте возможность использования корреляционного анализа для решения данной задачи.

7. По данным табл. 9.13 определите наличие, силу и статистическую значимость связи между ценой на различные продукты и коэффициенты частных корреляций. Предложите экономическую интерпретацию полученных результатов. Связано ли изменение цен с изменением цены на какой-то один продукт или существует некая «общая» причина изменения цен? Объясните смысл p -уровня значимости для данного случая. Обоснуйте возможность использования корреляционного анализа для решения данной задачи.

Таблица 9.12

Данные для решения задачи 6

Период	Предложение, тыс. ед.	Спрос, тыс. ед.	Цена продажи, руб./ед.	Среднемесяч- ный индекс потребитель- ских цен, %
Январь	1000	800	50	—
Февраль	1900	3000	90	105
Март	2000	2500	110	101
Апрель	2100	2000	110	102,5
Май	2000	1500	110	104
Июнь	2000	1300	120	105
Июль	1500	1700	135	106
Август	1600	1800	150	103,5
Сентябрь	1900	1850	148	103,5
Октябрь	2500	2000	148	105
Ноябрь	2300	2000	150	108,5
Декабрь	2500	1200	150	111

Таблица 9.13

Данные для решения задачи 7

Дата	Мясо (говядина)	Масло животное	Масло растительное	Сахар- песок
03.07	114,08	190,58	78,53	40,00
10.07	114,46	191,12	78,90	40,16
17.07	115,61	190,52	80,37	41,67
24.07	116,91	189,86	81,75	43,27
31.07	117,47	189,38	82,99	44,17
07.08	118,05	189,00	85,05	45,11
14.08	118,86	188,64	87,45	45,94
21.08	119,63	188,61	89,64	46,47
28.08	120,16	188,62	92,05	46,74
04.09	121,31	187,67	94,41	46,79
11.09	121,66	187,50	97,06	46,86
18.09	122,00	187,96	98,07	46,85
25.09	122,94	188,10	100,00	46,84
02.10	123,05	188,10	102,98	45,75
09.10	123,06	188,54	105,04	46,75
16.10	123,42	189,30	106,09	46,65
23.10	123,66	190,31	106,92	46,47
30.10	123,73	191,35	107,06	46,46

8. По данным табл. 3.9 определите, существует ли статистически значимая связь между успешностью освоения студентами двух учебных предметов, выраженная в балльных оценках. О чем может свидетельствовать наличие такой связи?

9. После завершения учебного курса по обработке данных студентам предлагалась тестовая задача. Время решения задачи и показатель успешности освоения курса представлены в табл. 9.14. Определите наличие связи. Сделайте выводы.

Таблица 9.14

Данные для решения задачи 9

Студент	Время решения задачи, мин	Успешность прохождения курса, балл
1	12	47
2	10	45
3	10	44
4	15	37
5	13	36
6	9	46
7	16	41
8	17	44
9	12	41
10	15	36
11	17	31
12	11	48

10. Результаты социологического опроса (выборка из 15 человек) представлены в табл. 9.15. Существуют ли статистически значимые связи между признаками? Сделайте выводы.

Таблица 9.15

Данные для решения задачи 10

Престижность работы, балл	Количество детей	Возраст, лет	Количество посещений ресторана (кафе) в мес.	Длительность телепросмотров в день, ч
32	3	34	3	1
50	0	41	4	3
17	0	52	4	2
69	3	67	0	5
17	0	40	1	5
52	0	22	8	3

Престижность работы, балл	Количество детей	Возраст, лет	Количество посещений ресторана (кафе) в мес.	Длительность телепросмотров в день, ч
32	3	31	1	4
50	0	23	4	4
19	9	64	0	6
37	4	55	1	2
14	3	66	0	5
51	0	22	7	1
45	0	19	2	7
44	0	21	2	1
46	4	58	1	2

11. Крупная компания по продаже бытовой электроники получила следующие данные об объемах продаж (y) и затратах на рекламу (x) в 50 регионах (табл. 9.16). Оцените силу и значимость связи между этими признаками.

Таблица 9.16

Данные для решения задачи 11

№	1	2	3	4	5	6	7	8	9	10	11	12	13
x	77	110	110	93	90	95	100	85	96	83	88	94	104
y	85	103	102	109	85	103	110	86	92	87	99	101	109
№	14	15	16	17	18	19	20	21	22	23	24	25	26
x	89	95	94	96	92	93	98	103	95	103	89	97	97
y	81	107	95	109	113	84	98	112	96	93	97	92	105
№	27	28	29	30	31	32	33	34	35	36	37	38	39
x	99	101	113	86	100	96	105	101	89	90	102	106	116
y	102	105	118	82	101	97	101	104	91	99	112	104	111
№	40	41	42	43	44	45	46	47	48	49	50		
x	117	100	96	88	109	109	116	98	100	95	96		
y	119	94	81	92	108	103	113	104	98	108	87		

12. По данным табл. П.11 проведите корреляционный анализ данных. Дайте интерпретацию парных и частных коэффициентов корреляции. Обнаруживаются ли «ложные» корреляции? Объясните полученные результаты.

13. По данным табл. П.13 проведите корреляционный анализ данных. Дайте интерпретацию парных и частных коэффициентов корреляции. Обнаруживаются ли «ложные» корреляции? Объясните полученные результаты.

14. По данным табл. П.12 с помощью корреляционного анализа этих данных выявите факторы, в наибольшей степени влияющие на размер заработной платы.

Глава 10

Регрессионный анализ

10.1. Модели парной и множественной регрессии

Корреляционный анализ может дать представление о степени связи, но не о ее виде. Для анализа воздействия на результативный признак одного или нескольких факторных признаков используется регрессионный анализ – основной метод математической статистики для выявления неявных и завуалированных связей между результатами измерений.

Если выявленные на основе корреляционного анализа связи между изучаемыми признаками окажутся *существенными* (т.е. достаточно сильными и статистически значимыми), то целесообразно найти их математическое выражение в виде регрессионной модели и оценить ее адекватность. Адекватная регрессионная модель может использоваться для прогнозирования изучаемого явления или показателя.

Таким образом, задачами *регрессионного анализа* являются определение формы связи (выбор и построение модели), установление степени влияния факторов на результативный признак и прогнозирование значений зависимой переменной.

Регрессионный анализ заключается в подборе соответствующей аппроксимирующей функции для имеющегося набора наблюдений. *Аппроксимация* (от лат. *approximo* – приближаюсь) – это приближенное выражение эмпирических данных в виде функции. Полученная функциональная зависимость называется *уравнением регрессии*, или просто *регрессией*. Более строго, регрессия – это зависимость *среднего* значения какого-либо признака от *среднего* значения других (одного или нескольких) признаков.

Регрессия называется *парной*, если она описывает зависимость между функцией и одной переменной, т.е. имеет вид $y = f(x)$. Регрессия называется *множественной*, если она включает несколько переменных и имеет вид $y = f(x_1, x_2, \dots, x_n)$.

Если уравнение регрессии линейное, т.е. имеет вид $y = a_1x + a_0$ (для одной переменной) или $y = a_1x_1 + a_2x_2 + \dots + a_nx_n + a_0$ (для n переменных), регрессия называется *линейной*, иначе – *нелинейной*.

Важнейшим этапом построения *регрессионной модели* (уравнения регрессии) является установление в исходных эмпирических данных некой закономерности и ее выражение в виде математической функции. Сложность заключается в том, что из множества функций необходимо найти такую, которая лучше других выражает реально существующие связи между анализируемыми признаками, т.е. нужно найти такую функцию, которая наилучшим образом аппроксимирует имеющиеся данные. Выбор типа функции может опираться на теоретические знания об изучаемом явлении, опыт предыдущих аналогичных исследований или осуществляться эмпирически – перебором и оценкой функций разных типов.

Для подбора аппроксимирующей функции решается следующая задача (для простоты рассмотрим случай парной регрессии, когда имеется лишь один факторный признак). Известен набор из n измерений результативного признака y при разных значениях факторного признака x : (x_i, y_i) , $i = 1, \dots, n$. Пусть имеется ряд функций (линейные, квадратичные, экспоненциальные и т.д.). Требуется выбрать функцию $\hat{y} = f(x)$ из имеющихся, чтобы ее значения \hat{y}_i для набора x_i наилучшим образом приближали значения y_i . Значения y_i называются *теоретическими (ожидаемыми) значениями*, а значения y_i – *эмпирическими (наблюдаемыми)*. Что значит «наилучшим образом»? В качестве критерия, определяющего, насколько одна функция лучше другой, используется набор разностей $\hat{y}_i - y_i$. Очевидно, что выбором функции $\hat{y} = f(x)$ их нужно сделать как можно меньшими. Но для сравнения качества аппроксимации просто складывать разности нельзя, ведь они могут иметь разные знаки, и тогда ошибки могут взаимно компенсироваться. Поэтому выбирают сумму квадратов разностей, которая должна быть минимизирована:

$$\sum (y_i - \hat{y}_i)^2 \rightarrow \min.$$

По этой причине такой метод называется *методом наименьших квадратов*¹ (МНК). Фактически этот метод реализован как метод решения систем уравнений.

Проиллюстрируем использование МНК на простейшем примере парной линейной регрессии. Уравнение однофакторного линейного регрессионного уравнения имеет вид

$$\hat{y} = a_0 + a_1x,$$

где \hat{y} — теоретические значения результативного признака, полученные по уравнению регрессии;

a_0, a_1 — коэффициенты (параметры) уравнения регрессии.

Значение a_0 соответствует теоретическому значению y при $x = 0$. Коэффициент a_1 показывает среднюю величину изменения результативного признака y при изменении факторного признака x на одну единицу, т.е. вариацию y , приходящуюся на единицу вариации x . Знак a_1 указывает направление этого изменения. Параметры регрессионного уравнения a_0, a_1 находят из условия минимальности сумм квадратов отклонений эмпирических данных y_i от теоретических \hat{y} :

$$\sum (y_i - \hat{y})^2 = \sum (y_i - a_0 - a_1x_i)^2 \rightarrow \min.$$

Графическая иллюстрация этого подхода представлена на рис. 10.1.

Таким образом, должна минимизироваться функция двух

переменных $f(a_0, a_1) = \sum_{i=1}^n (y_i - a_1x_i - a_0)^2$. Для этого первые част-

ные производные функции $f(a_0, a_1)$ по аргументам a_0 и a_1 приравниваются нулю, и в результате получается система двух линейных уравнений:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases}$$

¹ Метод был разработан немецким математиком К. Гауссом в 1794 г.

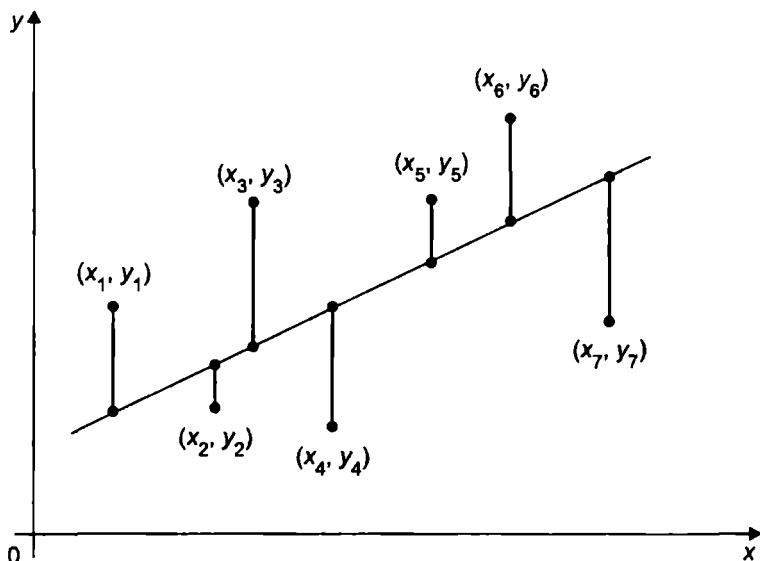


Рис. 10.1. Иллюстрация метода МНК

Решение этой системы дает значение коэффициентов регрессии в общем виде:

$$a_0 = \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - \sum x \sum y};$$

$$a_1 = \frac{n \sum xy - \sum y \sum x}{n \sum x^2 - \sum x \sum y}.$$

Технически более удобно вычислять параметры уравнения парной линейной регрессии по следующим формулам, дающим тот же результат:

$$a_1 = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sum (x - \bar{x})^2};$$

$$a_0 = \bar{y} - a_1 \bar{x}.$$

Метод легко обобщается на случай множественной линейной и нелинейной регрессии. Если, например, в качестве апп-

роксимирующей функции выбирается квадратическая зависимость вида

$$y = a_0 + a_1x + a_2x^2,$$

то минимизироваться должна функция трех переменных

$$f(a_0, a_1, a_2) = \sum_{i=1}^n (y_i - a_2x_i^2 - a_1x_i - a_0)^2$$

Для того чтобы найти значения коэффициентов регрессии, нужно вычислить первые частные производные функции $f(a_0, a_1, a_2)$ по аргументам a_0 , a_1 , и a_2 , приравнять их нулю и решить систему трех уравнений с тремя неизвестными.

Но при использовании МНК для определения параметров регрессии многих переменных возрастают техническая сложность и объем расчетов из-за увеличения числа переменных и уравнений, поэтому регрессионный анализ целесообразно осуществлять с помощью общепользовательских или специальных программных средств. Например, можно использовать соответствующие стандартные функции Excel или процедуры пакета SPSS, более удобные и содержательные с точки зрения отчетов о результатах (см. следующие подразделы этой главы).

Если зависимая и независимые переменные представлены в стандартной z -шкале (см. подразд. 2.2), то коэффициенты уравнения регрессии $y = \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + a_0$ называются *стандартными коэффициентами регрессии*, или *β -коэффициентами*. Знак β -коэффициента соответствует знаку корреляции между зависимой переменной и данной независимой переменной. Если β -коэффициент положителен, то связь этой переменной с зависимой переменной прямая, если β -коэффициент отрицателен, то связь обратная. По абсолютной величине β -коэффициента можно судить о влиянии данной независимой переменной на зависимую: чем больше коэффициент по абсолютному значению, тем сильнее это влияние, т.е. тем выше информативность этой переменной для прогнозирования значений результативного признака. Если β -коэффициент близок к нулю, связь между переменными отсутствует. В случае парной регрессии β -коэффициент в точности равен коэффициенту корреляции между результативным и факторным признаками.

При изучении связи социальных и экономических показателей помимо линейных моделей используют различного вида нелинейные уравнения. Во многих случаях нелинейные связи *линеаризуют*, т.е. приводят к линейному виду (путем логарифмирования, замены переменных и т.д.). Например, при полиномиальной связи можно добавить дополнительную переменную.

С помощью соответствующих трансформаций в линейную модель можно перевести и другие исходно нелинейные связи. К примеру, очень часто встречающуюся экспоненциальную связь $y = ae^{bx}$ можно преобразовать в линейную путем логарифмирования: $\ln(y) = \ln(a) + bx$.

Если в качестве аппроксимирующей функции выбирается зависимость

$$y = \frac{1}{a_0 + a_1x},$$

то замена переменных $z = 1/y$ дает линейную зависимость $z = a_0 + a_1x$. Если $y = (a_0 + a_1x)^2$, то замена $z = \sqrt{y}$ приводит к линейной зависимости $z = a_0 + a_1x$.

В микроэкономике используется производственная функция Кобба–Дугласа, связывающая объем выпуска продукции f с затратами капитала K и труда L :

$$f(K, L) = K^\alpha L^\beta$$

При этом предполагается, что параметры α и β одни и те же для предприятий одной отрасли. Поэтому знание этих параметров позволит прогнозировать объем производства для разных значений K и L . Для оценки этих параметров можно собрать информацию по n предприятиям отрасли об объемах производства, затратах капитала и труда на этих предприятиях и по этим данным построить регрессионную модель. При этом функцию Кобба–Дугласа можно линеаризовать путем ее логарифмирования:

$$\ln f(K, L) = \alpha \ln K + \beta \ln L.$$

После замены переменных и соответствующего преобразования исходных данных $x_i = \ln K_i$, $y_i = \ln L_i$, $z_i = \ln f_i$ для всех $i = 1, \dots, n$, получаем линейное регрессионное уравнение $z = \alpha x + \beta y$, из которого получают оценки α и β .

Связи, которые с помощью соответствующих трансформаций могут быть переведены в линейную модель, называются *линейными по существу (intrinsically linear model)*. Параметры регрессионных моделей, линейных по существу, вычисляются методом МНК, т.е. непосредственно решением соответствующей системы уравнений. Для нелинейных моделей, которые нельзя линеаризовать, может использоваться итерационная процедура оценки коэффициентов нелинейной регрессионной модели (подразд. 10.4).

Применение регрессионного анализа предполагает выполнение ряда предпосылок. Хотя многие из этих предпосылок нельзя проверить *точно*, исследователь должен выполнить хотя бы приблизительную их оценку и сделать обоснованный вывод о величине отклонений от этих предпосылок и возможности проведения регрессионного анализа. Чем сильнее отклонение от предпосылок-допущений, тем менее надежными могут оказаться результаты анализа и тем с большей осторожностью к ним следует относиться.

Первая предпосылка – отсутствие сильных линейных взаимосвязей между независимыми переменными регрессионной модели (не случайно они называются *независимыми*). Наличие таких связей называется *мультиколлинеарностью*. Эффект мультиколлинеарности означает, что по крайней мере одна из независимых переменных, включенных в модель, является лишней (модель *избыточна*). Кроме того, мультиколлинеарность может привести к ложным корреляциям, которые рассматривались в подразд. 9.2. При мультиколлинеарности снижается точность оценки параметров регрессионной модели и прогноза, сделанного на ее основе, возрастают ошибки оценок некоторых параметров модели, эти оценки становятся неустойчивыми: добавление или исключение нескольких наблюдений приводит к значительным изменениям в оценках. Интерпретация коэффициентов регрессионного уравнения при эффекте мультиколлинеарности также может быть затруднена.

Проиллюстрируем эффект мультиколлинеарности на очень простом, но весьма характерном примере. Предположим, определяется наличие связи между ростом человека и длиной ступни. В качестве результативного признака выберем рост, а в качестве факторного признака – длину ступни. Очевидно, что такая связь должна наблюдаться, но что получится, если в качестве

переменных регрессионной модели использовать длину левой ступни и длину правой ступни? Если строится линейная регрессионная модель и коэффициенты перед переменными положительны, то их следует интерпретировать так: при увеличении длины левой ступни на 1 см рост увеличивается в среднем на a_1 см, а при увеличении длины правой ступни на 1 см рост увеличивается в среднем на a_2 см. Если a_1 и a_2 отличаются друг от друга (что весьма вероятно), то подобная интерпретация выглядит абсурдной. Еще более абсурдна интерпретация, если коэффициенты будут иметь разные знаки (что также не исключено). В таком случае из модели будет следовать, что с увеличением длины одной ступни рост увеличивается, а с увеличением длины другой — уменьшается! Такие абсурдные выводы получаются из-за тесной связанности переменных, включенных в модель, т.е. две или более переменных содержат одинаковую (как в данном примере) или приблизительно одинаковую информацию для объяснения резульативного признака.

Самый простой способ выявить эффект мультиколлинеарности — определить коэффициенты корреляции между переменными, которые предполагается включить в регрессионную модель в качестве факторных признаков. Следует избегать включения в модель переменных, между которыми обнаруживается сильная корреляционная связь ($r > 0,7$). Существует довольно много других статистических индикаторов мультиколлинеарности (избыточности); некоторые из них рассматриваются в подразд. 10.4.

Вторая предпосылка — переменные модели должны иметь распределение, близкое к нормальному. Хотя отклонение от этого предположения не критично для проведения регрессионного анализа, тем не менее, прежде чем сделать окончательные выводы, следует рассмотреть распределения представляющих интерес переменных с помощью соответствующих статистических критериев и графиков распределения (см. подразд. 3.4).

Третья предпосылка — резульативный и факторные признаки должны быть измерены в метрической шкале. Однако на практике это требование не всегда можно выполнить. Поэтому допустимым считается, чтобы резульативный признак измерялся в метрической шкале, а факторные — по крайней мере в порядковой шкале.

Довольно часто в регрессионную модель включаются переменные и в номинативной шкале (например, переменная «пол»,

принимая значения 0 и 1). Такие переменные называются *фиктивными переменными* (*dummy variables*) или *индикаторами*. Коэффициенты перед соответствующей фиктивной переменной показывают, насколько в среднем изменяется значение резуль- тативного признака при включении того или иного факторного признака, измеренного в номинативной шкале, по сравнению с отсутствием этого признака (присутствием другого признака).

На практике также распространены случаи использования номинативной или порядковой переменной, изменяющейся по более чем двум категориям. Например, необходимо построить модель продаж с учетом изменения по кварталам (четыре кате- гории изменения переменной), модель изменения цены на не- движимость в городе с учетом городского района, модель изме- нения заработной платы с учетом уровня образования и т.п. В этом случае создают столько фиктивных переменных, сколько имеется категорий изменения признака, но в регрессионную модель включаются не все фиктивные переменные, а на одну меньше (не включается любая из фиктивных переменных). В этом случае коэффициент перед каждой фиктивной переменной показывает, насколько изменяется среднее значение резуль- тативного признака для соответствующей переменной по срав- нению с переменной, не включенной в модель. Например, при моделировании продаж по кварталам создаются четыре фиктив- ные переменные (q_1, q_2, q_3, q_4), а в модель включаются любые три переменные, например q_2, q_3, q_4 . Коэффициенты перед эти- ми переменными будут показывать, насколько в среднем изме- няется объем продаж по сравнению с первым кварталом.

Осуществляемое на основе регрессионного анализа прогно- зирование значений резуль- тативного признака у обеспечивает, как правило, лучшие результаты при *интерполяции* (определе- ние у внутри интервала изменения эмпирических значений фак- торных признаков), чем при *экстраполяции* (за пределами этого интервала). Ошибка прогнозирования возрастает тем сильнее, чем дальше значение любого из факторных признаков x_i , по ко- торым строится прогноз, выходит за пределы минимального и максимального значений эмпирических данных [$x_{i\min}; x_{i\max}$] (под- разд. 10.4).

Как и в случае корреляционного анализа, любая регрессион- ная модель позволяет обнаружить только количественные зави- симости, которые не обязательно отражают причинные зависи-

мости, т.е. влияние одного фактора на другой. Гипотезы о причинной связи признаков должны дополнительно обосновываться с помощью теоретического анализа, содержательно объясняющего изучаемое явление или процесс. Интерпретация полученных результатов, т.е. перевод их с языка статистики на язык экономики или другой дисциплины, является даже более сложным этапом, чем применение математико-статистических процедур регрессионного анализа.

10.2. Проверка обоснованности регрессионной модели

Регрессионный анализ – это не только выбор аппроксимирующей функции и расчет ее коэффициентов на основе метода наименьших квадратов. Построенная модель должна быть *обоснованной* с точки зрения ее соответствия исследуемому объекту (процессу, явлению и т.д.) и эмпирическим данным. Проверка обоснованности модели включает три момента:

1) качественный анализ возможного вида функции с точки зрения содержания и законов развития объекта исследования (если, например, динамика роста цен на определенный товар не предполагает резкого изменения, то для прогнозирования этого процесса не следует использовать квадратическую или кубическую регрессионную модель);

2) проверка *адекватности* модели, т.е. ее соответствия имеющимся эмпирическим данным;

3) оценка *статистической значимости* модели в целом и ее параметров, т.е. оценка их «неслучайности»: насколько параметры модели характерны для всей генеральной совокупности, не являются ли они результатом стечения случайных обстоятельств.

При этом, разумеется, предполагается, что все предпосылки применения регрессионного анализа, рассмотренные в подразд. 10.1, выполняются, т.е. проведена проверка их выполнения.

Первый этап проверки обоснованности модели предполагает привлечение знаний из конкретных экономических, социологических и других дисциплин (например, микро- или макроэкономики). На этом этапе делается заключение о том, насколько вид используемой функциональной зависимости и характер влияния факторных признаков на результативный могут соответ-

ствовать исследуемому процессу. О характере влияния факторных признаков на резульативный признак в первую очередь свидетельствуют знак и величина соответствующего коэффициента регрессии, а также степень соответствующей независимой переменной. Например, если коэффициент перед независимой переменной имеет положительное значение, то с увеличением данного фактора резульативный признак возрастает. Но интерпретация этих величин должна соответствовать социально-экономическому содержанию моделируемого признака. Если теория указывает, что факторный признак должен иметь положительное значение, а в уравнение регрессии он входит с отрицательным знаком, то необходимо проверить расчеты параметров уравнения регрессии и (или) выбрать другую регрессионную модель.

Адекватность и значимость модели проверяются с помощью математико-статистических подходов, методов и показателей, которые и являются главным предметом рассмотрения данного и следующих подразделов этой главы.

Адекватность регрессионных моделей – это их соответствие фактическим статистическим данным. Регрессионная модель считается адекватной, если теоретические значения зависимой переменной (т.е. предсказанные на основе модели) согласуются с результатами наблюдений.

Исходное предположение для проверки адекватности регрессионной модели заключается в том, что зависимость между прогнозируемым (теоретическим) значением резульативного признака (\hat{y}) и факторами (x_i) имеет вид

$$\hat{y} = f(x_i) + \varepsilon,$$

где ε – некоторая случайная величина, связанная с влиянием неконтролируемых или неучтенных факторов, случайных ошибок измерения или с тем, что самой «природе» объекта исследования свойственны случайные процессы.

Именно из-за величины ε возникают ненулевые остатки, т.е. разности между теоретическими и эмпирическими значениями ($y_i - \hat{y}_i$). Предполагается также, что эти остатки независимы (некоррелированы) и распределены по нормальному закону с нулевым средним и одинаковой дисперсией. Это предположение легко проверить путем построения диаграммы остатков.

Поэтому общий подход к проверке адекватности полученной модели заключается в нахождении остатков, а точнее – значения суммы квадратов разностей между наблюдаемыми и предсказанными моделью значениями переменной y . Этот показатель обычно обозначается как SSe (от *Sum of Squares*). Далее вычисляются статистические показатели, на основе которых можно судить об адекватности полученной модели (основным здесь является F -критерий (критерий Фишера)).

Для адекватной модели, кроме некоррелированности остатков и их нормального распределения, должно выполняться условие *гомоскедастичности*, т. е. постоянства дисперсии ошибок для всех наблюдений. Оценка выполнимости этого условия проводится по графику остатков: если все остатки укладываются в симметричную относительно нулевой линии полосу, то можно считать, что дисперсия ошибок наблюдений постоянна. На графике распределения значений зависимой переменной от одной из независимых переменных не должно быть сильных «раздуваний». Значительное отклонение от этого условия называется *гетероскедастичностью*. Для оценки гетероскедастичности разработаны и специальные статистические тесты.

Таким образом, основным показателем отклонений данных от модели может служить сумма квадратов разностей теоретических и эмпирических значений зависимой переменной y . Чем этот показатель меньше, тем приближение лучше, следовательно, построенная модель лучше описывает реальные данные. На этой основе вычисляется такой показатель адекватности регрессионной модели, как *остаточная дисперсия*:

$$\hat{\sigma}_y^2 = \frac{SSe}{n-1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}.$$

Но при этом следует учитывать, что если в модель добавляется новый параметр, то сумма квадратов, как правило, оказывается меньше. Иными словами, многочлен второй степени лучше соответствует данным, чем линейная функция, многочлен третьей степени лучше приближает исходные данные, чем многочлен второй степени, и т.д. В конце концов можно дойти до многочлена степени $(n - 1)$ с n коэффициентами, который про-

ходит через *все* заданные точки, т.е. разница между эмпирическими и теоретическими значениями y будет равна нулю [21]. Но прогностические возможности такой модели существенно меньше, чем у линейной функции. Излишнее усложнение статистических моделей вредно. Даже если адекватность простых функций незначительно меньше, чем более сложных, рекомендуется выбирать более простую функцию.

Поэтому показатель адекватности регрессионной модели, использующий оценку *остаточной дисперсии*, корректируется на число параметров модели (m) [21]:

$$\hat{\sigma}_y^2 = \frac{SSe}{n - m}.$$

В случае парной регрессии скорректированная оценка остаточной дисперсии имеет вид

$$\hat{\sigma}^2 = \frac{SSe}{n - 2},$$

поскольку число оцениваемых параметров $m = 2$.

Корень квадратный из этого показателя называется *стандартной ошибкой оценки* (*standard error of estimate*) и на практике используется для характеристики точности, с которой можно оценивать и прогнозировать значение результативного признака. При этом данный показатель интерпретируется аналогично стандартной ошибке среднего при построении доверительных интервалов (см. подразд. 3.3). Например, среднее предсказанное значение зависимой переменной при определенном значении независимой переменной (или переменных в случае множественной регрессии) с вероятностью 95% будет лежать в пределах приблизительно двух величин стандартной ошибки оценки по обе стороны от значения y , предсказанного на основе регрессионной модели $\hat{y} \pm 1,96\hat{\sigma}$

Другими важными показателями адекватности регрессионной модели являются коэффициент множественной корреляции (R) и коэффициент множественной детерминации, равный квадрату коэффициента множественной корреляции (R^2).

Расчет этих коэффициентов базируется на следующих рассуждениях. Чем меньше остаточная дисперсия $\hat{\sigma}_y^2$ по отно-

шению к общей дисперсии σ_y^2 результативного признака

$$\left(\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \right), \text{ тем, очевидно, лучше прогноз.}$$

Например, если связь между переменными x и y полностью отсутствует, то отношение остаточной дисперсии переменной y к исходной дисперсии результативного признака равно 1. Если модель полностью адекватна, то остаточная дисперсия отсутствует и отношение дисперсий будет равно 0. В большинстве случаев это отношение будет лежать в диапазоне от 0 до 1. Коэффициент множественной детерминации определяется как

$$R^2 = 1 - \frac{SSe}{SSy},$$

где $SSy = \sum_{i=1}^n (y_i - \bar{y})^2$ – сумма квадратов отклонений зависимой переменной от среднего.

Иначе R^2 можно определить как

$$R^2 = \frac{SS_R}{SSy},$$

где $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ – сумма квадратов отклонений прогнозных (теоретических) значений зависимой переменной от среднего (*сумма квадратов, обусловленная регрессией*);

$$SSy = SS_R + SSe.$$

Если, например, $R^2 = 0,6$, то 60% изменчивости результативного признака может быть объяснено (предсказано) на основе регрессионной модели, а 40% остаточной дисперсии остаются необъясненными (неучтенными). Таким образом, значение R^2 является индикатором степени адекватности регрессионной модели (значение R^2 , близкое к 1, показывает, что модель объяс-

няет почти всю изменчивость результативного признака). Иногда R^2 называют показателем прогностической силы регрессионной модели.

В случае множественной регрессионной модели коэффициент множественной детерминации R^2 всегда больше коэффициентов парной детерминации r^2 , полученных для результативного признака и каждой из независимых переменных на основе связи такого же вида, что и построенная модель. Но обычно коэффициент множественной детерминации меньше, чем сумма соответствующих парных коэффициентов. Однако иногда возможны ситуации «синергетического эффекта», когда комбинация нескольких переменных в регрессионной модели объясняет изменчивость результативной переменной лучше, чем простое суммирование эффектов влияния на результативный признак каждой из независимых переменных по отдельности.

При включении дополнительных независимых переменных в регрессионную модель R^2 всегда возрастает, даже если включаемая переменная не имеет «смысловой» связи или значимой корреляционной связи. Подобное «техническое» увеличение показателя R^2 может исказить представление об адекватности конструируемой регрессионной модели, поэтому на практике используют также «скорректированный R^2 » (*adjusted R^2*), учитывающий возможное «техническое» влияние дополнительных переменных на изменчивость результативного признака:

$$R_{adj}^2 = 1 - \frac{SSe / (n - m)}{SSy / (n - 1)} = 1 - \frac{\hat{\sigma}_y^2}{\sigma_y^2},$$

где n — число наблюдений;

m — число параметров модели.

Если при включении дополнительной переменной в регрессионную модель значение этого показателя уменьшается или остается примерно тем же, это в большинстве случаев означает, что данная переменная реально не объясняет изменчивость результативного признака, поэтому ее не следует включать в регрессионное уравнение.

При конструировании регрессионной модели следует исходить из принципа *парсимонии* (*parsimony*): нужно объяснять изменчивость результативного признака с помощью минимального

числа независимых переменных, т.е. чем проще модель, обеспечивающая приемлемую адекватность, тем она лучше. Более простой модели легче придать «разумную» интерпретацию, т.е. объяснить изменение результативного признака с точки зрения содержательно-смысловых, а не абстрактно-математических связей.

Коэффициент множественной корреляции по определению равен корню квадратному из коэффициента детерминации. Это неотрицательная величина, принимающая значения между 0 и 1. Коэффициент множественной корреляции — это мера линейной связи зависимой переменной с множеством независимых переменных, включенных в модель. На практике коэффициент множественной корреляции можно определить как коэффициент парной корреляции между теоретическими и эмпирическими значениями результативного признака. Для парной регрессии R равен коэффициенту корреляции между y и x .

Более тщательная проверка адекватности регрессионной модели может быть проведена, если для зависимой переменной выполнены повторные наблюдения. В этом случае используется следующая процедура. Если модель адекватна, то эмпирические значения y должны быть близки значениям, предсказанным регрессионной моделью. Мерой неадекватности модели будет сумма квадратов отклонений этих значений. Чем меньше она будет, тем лучше результаты наблюдений согласуются с моделью.

Иногда для оценки адекватности регрессионную модель строят не по всему имеющемуся набору данных, а по меньшему числу наблюдений. Оставшиеся наблюдения используются для проверки адекватности, как и в случае повторных измерений. Если при этом модель признается достаточно адекватной, то для уточнения параметров модели вновь строится уравнение регрессии на основе выбранного вида аппроксимирующей функции.

Регрессионный анализ, как и корреляционный, обычно проводится по выборке из генеральной совокупности. В связи с этим показатели регрессии и корреляции — параметры уравнения регрессии, коэффициенты множественной корреляции и детерминации могут быть искажены действием случайных факторов. Для того чтобы проверить, насколько эти показатели характерны для всей генеральной совокупности, необходимо проверить их статистическую значимость (иногда говорят «*существование*» или «*достоверность*»). При этом выясняют, насколько

каждый из вычисленных параметров регрессии (включая коэффициент a_0 – постоянный член регрессии) характерен для генеральной совокупности, т.е. не является ли полученное значение результатом действия случайных причин или стечения случайных обстоятельств.

Логика проверки коэффициентов множественной корреляции и детерминации, а также параметров регрессии в целом аналогична логике проверки статистических гипотез (вторая часть учебного пособия). Например, значимость коэффициентов парной линейной регрессии проверяют с помощью t -критерия Стьюдента. Эмпирические значения t -критерия для параметра a_0 вычисляются по формуле

$$t_{a_0} = |a_0| \frac{\sqrt{n-2}}{\hat{\sigma}_y},$$

а для параметра a_1 – по формуле

$$t_{a_1} = |a_1| \frac{\sqrt{n-2}}{\hat{\sigma}_y} \sigma_x,$$

где $\hat{\sigma}_y$ – среднее квадратическое отклонение эмпирических значений результативного признака y от теоретических значений \hat{y} (корень квадратный из остаточной дисперсии):

$$\hat{\sigma}_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}};$$

σ_x – среднее квадратическое отклонение факторного признака x от его среднего значения \bar{x} (корень квадратный из дисперсии x):

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}},$$

где n – объем выборки.

Эмпирические значения t , вычисленные по вышеприведенным формулам, сравнивают с критическими t , которые определяют из соответствующей статистической таблицы t -распределения Стьюдента (табл. П.2) с учетом принятого уровня значимости α и числом степеней свободы $df = n - 2$. Параметр регрессионной модели признается значимым при условии, если $t_{\text{эмп}} > t_{\text{крит}}$. В этом случае маловероятно, что найденные значения параметров обусловлены только случайными совпадениями.

В общем случае при проверке статистической значимости коэффициентов регрессии β_i ($i = 0, \dots, n$) нулевая гипотеза формулируется как $H_0: \beta_i = 0$, а альтернативная — $H_A: \beta_i \neq 0$ (т.е. проверяется ненаправленная альтернативная гипотеза). Если нулевая гипотеза отклоняется, соответствующая переменная может быть включена в регрессионное уравнение, иначе — переменную не следует включать. На практике статистическая значимость параметров модели оценивается с помощью статистических пакетов, определяющих p -уровень значимости, на основе которого принимают решение об отклонении или невозможности отклонения нулевой гипотезы.

Еще раз отметим, что статистическая значимость параметров — не единственный критерий принятия окончательного решения о включении переменной в модель. Другими критериями являются экономический смысл переменных и их возможных связей, изменение R^2 и скорректированного R^2 , наличие и степень мультиколлинеарности и др.

В заключение этого подраздела обратим внимание читателя на соотношение числа наблюдений и количества параметров регрессионной модели. Этот вопрос особенно важен, когда число наблюдений относительно мало по сравнению с числом независимых переменных. Например, понятно, что «неразумно» делать выводы из анализа модели с 50 переменными на основе 10 наблюдений. Минимально необходимое число наблюдений для построения регрессионной модели (n) должно быть не меньше числа *всех* признаков, включенных в модель (m), в том числе результативного. Но рекомендуется использовать не менее 10 наблюдений на одну независимую переменную. Если, например, предполагается оценить влияние на результативный признак двух факторов, то желательно сформировать выборку, включающую не менее 20 элементов.

10.3. Построение регрессионных моделей в программе Excel

Наиболее доступным и распространенным программным продуктом, позволяющим проводить базовый статистический анализ эмпирических данных, является программа Excel. В этом подразделе мы рассмотрим применение стандартных статистических функций и инструментов Excel, которые можно использовать для регрессионного анализа, и проиллюстрируем смысл параметров регрессионной модели, а также основные индикаторы ее адекватности и статистической значимости на «хрестоматийном» примере.

Пример 10.1. Известны данные о рыночном спросе на товар (Q , тыс. шт.), рыночной цене товара (P , ден. ед.) и объеме средств, направленных на рекламу данного товара (A , тыс. ден. ед.). Необходимо построить модель зависимости спроса от цены товара и затрат на рекламу, оценить ее обоснованность и определить прогнозируемое значение спроса при следующих значениях факторных признаков:

P	A
13	15
9	30
12	20

Для наглядности исходные данные и результаты расчетов приведены в одной таблице, с адресацией ячеек, принятой в Excel (табл. 10.1).

В ячейках A19:C21 табл. 10.1 приведены предельные и средние значения для исходных данных, полученные с использованием функций МИН, МАКС, СРЗНАЧ.

В ячейках F5:H5 приведены коэффициенты парной корреляции между всеми изучаемыми показателями. Для их расчета использовалась функция КОРРЕЛ. Например, коэффициент корреляции для пары Q и P рассчитывается как КОРРЕЛ (C3:C17; A3:A17). В ячейках F7:H7 приведены коэффициенты детерминации. Полученные коэффициенты указывают на то, что в рассматриваемой ситуации цена является основным фактором, определяющим спрос. Связь между переменными P и A отсутствует (эффект мультиколлинеарности не наблюдается).

Для дальнейшего анализа построим регрессионную зависимость $Q = f(P, A)$. Будем исходить из предположения о линей-

Исходные данные и результаты регрессионного анализа

	A	B	C	D	E	F	G	H
1	Исходные данные							
2	P	A	Q	Тенденция				
3	11,5	21	1170	1185,66	Коэффициенты корреляции			
4	9,3	18,5	2010	1652,75	$r(OP)$	$r(OA)$	$r(PA)$	
5	12,76	20	720	730,13	-0,775	0,4043	-0,048	
6	9,87	18,2	1170	1458,45	Коэффициенты детерминации			
7	10,1	18	1000	1373,95	0,6006	0,1635	0,0023	
8	10,5	17,5	1250	1215,22				
9	10,61	17	770	1143,55	ЛИНЕЙН			
10	11	16	1200	949,16	77,3171	-300,17	3013,97	
11	11,2	15,5	980	850,47	31,1842	58,85685	888,061	
12	9,54	20	1520	1696,68	0,735909	252,9565	#Н/Д	
13	9,1	22	2310	1983,39	16,71941	12	#Н/Д	
14	11,96	15,8	850	645,54	2139649	767844	#Н/Д	
15	12,1	22,5	1140	1121,54				
16	12,24	18	770	731,59	$p(F)$	0,00034		
17	11,7	19,4	880	1001,92	$p(t)$	0,02899	0,00026	0,00533
18	Предельные и сред. значения				P	A	Q	
19	9,1	15,5	720	Мин	13	15	271,50	
20	12,76	22,5	2310	Макс	9	30	2631,94	
21	10,8987	18,6267	1182,667	Средн.	12	20	958,26	
22	ЛГРФРИБЛ				Прогноз			
23	1,056	0,795	4959,805		P	A	Q	
24	0,023	0,043	0,650		13	15	565,99	
25	0,746	0,185	#Н/Д		9	30	3194,46	
26	17,661	12,000	#Н/Д		12	20	933,51	
27	1,210	0,411	#Н/Д					

ном характере связи и используем функцию ЛИНЕЙН. Эта функция рассчитывает коэффициенты множественной линейной регрессии и соответствующую статистику с применением метода наименьших квадратов. Функция возвращает массив значений следующего вида (табл. 10.2).

Таблица 10.2

Возвращаемые значения функции ЛИНЕЙН

m_n	m_{n-1}		m_1	b
$\sigma(m_n)$	$\sigma(m_{n-1})$		$\sigma(m_1)$	$\sigma(b)$
R^2	$\sigma(y)$	—	—	—
F	df	—	—	—
SS_R	SSe	—	—	—

В данной таблице используются следующие обозначения: m_1, m_2, m_n, b – коэффициенты регрессионного уравнения

$$y = m_1x_1 + m_2x_2 + \dots + m_nx_n + b;$$

$\sigma(m_1), \dots, \sigma(m_n), \sigma(b)$	– стандартные значения ошибок оценки величин коэффициентов m_1, \dots, m_n и b ;
R^2	– коэффициент множественной детерминации;
$\sigma(y)$	– стандартная ошибка оценки y ;
F	– F -критерий (используется для определения того, является ли зависимость между результативным и факторными признаками статистически значимой);
df	– число степеней свободы, определяемое по формуле $df = n - (k + 1)$, где n – объем выборки (для данного примера $n = 15$), k – число независимых переменных (в данном случае $k = 2$);
SS_R, SS_e	– сумма квадратов, обусловленная регрессией, и остаточная сумма квадратов.

Синтаксис функции:

ЛИНЕЙН(известные_y; известные_x; конст; статистика),

- где «известные_y» – массив эмпирических значений результативного признака;
- «известные_x» – массив независимых переменных (если этот массив имеет несколько столбцов, то каждый столбец интерпретируется как отдельная переменная);
- «конст» – параметр, принимающий значения 0 или 1 и указывающий, требуется ли, чтобы постоянный член регрессионного уравнения b был равен 0 («конст» = 0), по умолчанию параметр «конст» принимается равным 1;
- «статистика» – параметр, принимающий значения 0 или 1 и указывающий, требуется ли возвращать статистику регрессии ($\sigma(m_1), \sigma(m_n), \sigma(b), R^2, \sigma(y)$ и др.). Если «статистика» имеет значение 1, то функция **ЛИНЕЙН** возвращает эти показатели, в противном случае функция **ЛИНЕЙН** возвращает только коэффициенты m_i и b .

Поскольку возвращается массив значений, функция должна задаваться в виде формулы массива, т.е. предварительно должен

быть выделен выходной диапазон ячеек, а после указания всех аргументов функции в соответствующем диалоговом окне «Мастер функций» нужно щелкнуть не по кнопке *OK*, а нажимать комбинацию клавиш *Ctrl – Shift – Enter*.

В данном примере для размещения результатов указан диапазон F10:H14 и функция имеет следующие аргументы: ЛИНЕЙН(C3:C17; A3:B17; 1;1). В итоге получено уравнение регрессии:

$$Q = 3013,97 - 300,17P + 77,32A.$$

Смысл коэффициентов данной модели следующий: константа (3014) определяет значение спроса при нулевых значениях цены и затрат на рекламу (предположение ограниченности спроса); коэффициент перед переменной *P* показывает, что с увеличением цены товара на 1 ден. ед. спрос в *среднем* уменьшается на 300 тыс. шт.; коэффициент перед переменной *A* показывает, что с увеличением затрат на рекламу товара на 1 тыс. ден. ед. спрос в *среднем* увеличивается на 77 тыс. шт.

Коэффициент множественной детерминации (0,74) можно интерпретировать следующим образом: на основе данной модели изменение спроса на 74% определяется изменением цены на товар и затрат на рекламу. Этот коэффициент превышает коэффициенты парной детерминации (0,60 и 0,16), но меньше значения их суммы.

С помощью статистики, возвращаемой функцией ЛИНЕЙН, можно оценить статистическую значимость полученных результатов. Нулевая гипотеза формулируется так: зависимость между результативным признаком и факторными признаками отсутствует. Альтернативная гипотеза, утверждающая, что такая зависимость существует, проверяется на основе *F*-критерия и в данном случае с помощью функции ФРАСП, возвращающей *p*-уровень значимости.

В итоге в ячейку F16 (см. табл. 10.1) заносится значение функции ФРАСП(F13;2;G13): 0,00034. Это значит, что нулевая гипотеза может быть отклонена на уровне $\alpha = 0,001$. Данный результат свидетельствует о высокой статистической значимости (достоверности) связи $Q = f(P, A)$.

Значимость коэффициентов регрессионного уравнения (m_1, m_2, \dots, m_n, b) можно оценить, используя функцию распределения Стьюдента. Для этого предварительно нужно вычис-

лить величины $t_i = |m_i|/\sigma(m_i)$, затем использовать функцию СТЬЮДРАСП(t_i ; df ; 2). Последний аргумент означает, что применяется двустороннее распределение Стьюдента (проверка ненаправленной альтернативной гипотезы). Функция возвращает p -уровень значимости (ячейки F17:H17).

Нулевая гипотеза утверждает, что соответствующий коэффициент регрессионного уравнения равен 0. Полученные значения p -уровня позволяют отклонить нулевую гипотезу на уровне $\alpha = 0,05$ для коэффициента m_2 и $\alpha = 0,01$ для двух других коэффициентов.

На основе полученного уравнения можно предсказать спрос (Q) при различных значениях P и A , т.е. определить точки, лежащие на регрессионной прямой $Q = 3013,97 - 300,17P + 77,32A$. Для этого в ячейку таблицы Excel можно записать данную формулу либо использовать специальную функцию ТЕНДЕНЦИЯ, которая возвращает значения в соответствии с линейным трендом. Как и функция ЛИНЕЙН, она аппроксимирует прямой линией (по методу наименьших квадратов) массивы «известные_у» и «известные_х», но возвращает прогнозные значения y для заданного массива «новые_х». Синтаксис функции:

ТЕНДЕНЦИЯ(известные_у; известные_х; новые_х; конст).

Если массив «новые_х» опущен, то предполагается, что он совпадает с массивом «известные_х». Смысл параметра «конст» такой же, как в функции ЛИНЕЙН. В табл. 10.1 в столбце «Тенденция» приведены результаты использования функции ТЕНДЕНЦИЯ(C3:C17; A3:B17) для тех же данных P и A , по которым строилась регрессионная модель. В ячейках H19:H21 приведены результаты прогнозирования спроса при значениях цены и расходов на рекламу, указанных в ячейках F19:G21. Для того чтобы получить эти значения, использовалась функция ТЕНДЕНЦИЯ(C3:C17; A3:B17; F19:G21).

Для анализа нелинейных зависимостей можно применять функцию ЛГРФПРИБЛ, которая аппроксимирует исходные данные на основе показательной зависимости:

$$y = b \cdot m_1^{x_1} \cdot m_2^{x_2} \dots \cdot m_n^{x_n}$$

Значения m_i являются основанием, возводимым в степень x_i , а b – константа. Данная зависимость может быть линеаризована путем логарифмирования (этим и объясняется название функции):

$$\ln y = x_1 \ln m_1 + \dots + x_n \ln m_n + \ln b.$$

Функция ЛГРФПРИБЛ работает аналогично функции ЛИНЕЙН, имеет такой же синтаксис и возвращает аналогичный массив результатов. Единственным исключением является смысл аргумента «конст». Если «конст» = 0, b принимается равным 1, иначе подбирается соответствующее значение b из условия минимизации остатков (по умолчанию «конст» = 1).

Для рассматриваемого примера возвращаемые значения функции ЛГРФПРИБЛ(С3:С17; А3:В17; 1;1) приведены в ячейках А23:С27. При использовании нелинейной зависимости коэффициент множественной детерминации незначительно увеличился ($R^2 = 0,746$), что свидетельствует скорее о нелинейном характере связи $Q = f(P, A)$. В каждом конкретном случае на основе этого и других показателей необходимо решать, какой из видов связи в наибольшей степени подходит к исходным данным (подробнее в подразд. 10.4).

В случае нелинейных зависимостей для прогнозирования значений результативного признака используется функция РОСТ, по синтаксису полностью совпадающая с функцией ТЕНДЕНЦИЯ. На основе применения этих функций можно сравнить вычисленные значения с фактическими значениями и построить графики для визуального сравнения. В ячейках Н24:Н26 приведены результаты использования функции РОСТ(С3:С17; А3:В17; F24:G26).

Функции ЛИНЕЙН и ТЕНДЕНЦИЯ можно применять для аппроксимации и нелинейных зависимостей после линеаризации исходных данных путем соответствующих преобразований исходных переменных и введения новых. Например, при аппроксимации исходных данных полиномом третьей степени проводят регрессионный анализ для той же переменной, возведенной в различные степени. Допустим, столбец A содержит значения y , а столбец B – значения x . Тогда можно ввести x^2 в столбец C , x^3 в столбец D , а затем использовать функции линейной регрессии, указав в качестве независимой переменной массив столбцов от B до D , а в качестве зависимой переменной – столбец A .

Для выбора требуемого вида регрессионного уравнения можно воспользоваться и графическими инструментами Excel. Для рассматриваемого примера остановимся на анализе связи спроса с ценой – наиболее важным и определяющим фактором.

С помощью пункта меню *Мастер диаграмм* по исходным данным построим график, показывающий зависимость спроса на товар от его цены. (Предварительно данные по цене необходимо отсортировать с помощью инструмента «Сортировка» в меню «Данные». Тип диаграммы: «Точечная».)

Установив курсор мыши на область графика, выберите команду **Добавить линию тренда** в меню *Диаграмма*. В появившемся окне *Линия тренда* выберите нужный тип линии тренда. Затем перейдите на вкладку *Параметры* и отметьте поля *Показывать уравнение на диаграмме* и *Поместить на диаграмму величину коэффициента аппроксимации*, а также при необходимости установите пределы экстраполяции данных.

Результат линейной и полиномиальной (полиномом второй степени) аппроксимации представлен на рис. 10.2. Полиномиальная аппроксимация несколько улучшает коэффициент множественной детерминации, что связано с нелинейным характером зависимости «спрос-цена», но очевидно, что такая функция не может обеспечить надежную экстраполяцию (на ее основе прогнозируется резкий рост спроса, тогда как эмпирические данные свидетельствуют скорее о его плавном снижении или приближении к некоторому постоянному значению).

Выбирая другие типы парной регрессии, с помощью данного инструмента можно сразу получать регрессионные уравнения и значение коэффициента множественной детерминации как индикатора адекватности выбранной модели.

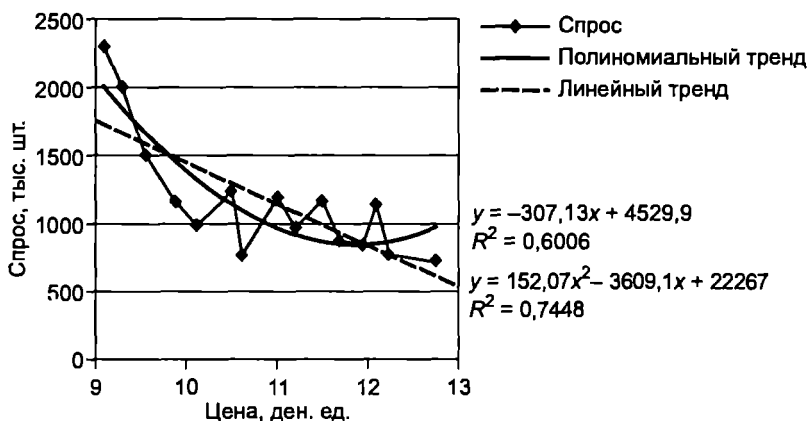


Рис. 10.2. Результаты линейной и полиномиальной аппроксимации

10.4. Построение регрессионных моделей в пакете SPSS

По сравнению с Excel, пакеты SPSS и STATISTICA обеспечивают гораздо более удобный и гибкий инструментарий для проведения регрессионного анализа, а также более содержательные отчеты о его результатах. Рассмотрим решение задачи построения адекватной модели «спрос – цена – реклама» в SPSS. (Описание процедур регрессионного анализа в пакете STATISTICA читатель может найти, например, в [8; 68]. При этом общая последовательность работы и интерпретации результатов аналогична работе в пакете SPSS.)

Для построения линейной множественной регрессии в пакете SPSS нужно выбрать из меню *Analyze* опцию *Regression*, затем *Linear* (рис. 10.3). Зависимая переменная переносится в поле *Dependent*. Независимые переменные (факторные признаки) – в поле *Independent(s)*.

цена	реклама	спрос
11,50	21,00	1170,00
9,30	18,50	2010,00
12,76	20,00	720,00
9,87	18,20	1170,00
10,10	18,00	1000,00
10,50	17,50	1250,00
10,61	17,00	770,00
11,00	16,00	1200,00
11,20	15,50	980,00
9,54	20,00	1520,00
9,10	22,00	2310,00
11,96	15,80	650,00
12,10	22,50	1140,00
12,24	18,00	770,00
11,70	19,40	680,00
13,00	15,00	
9,00	30,00	
12,00	20,00	

The image shows a screenshot of the SPSS 'Linear Regression' dialog box. On the left, a data table is visible with columns for 'цена', 'реклама', and 'спрос'. The dialog box has the following settings: 'Dependent' is 'спрос', 'Independent(s)' are 'цена' and 'реклама', and 'Method' is 'Enter'. The 'Statistics...' button is highlighted.

Рис. 10.3. Диалоговое окно Linear Regression

После щелчка по кнопке *Statistics* появляется диалоговое окно *Linear Regression: Statistics*. Здесь выбираются параметры:

Regression Coefficient Estimates – оценки значимости коэффициентов регрессионной модели;

Regression Coefficient Confidence Intervals – доверительные интервалы для коэффициентов регрессионной модели;

Model fit – параметры соответствия модели эмпирическим данным (коэффициенты множественной корреляции, множественной детерминации и др.);

Descriptives – описательная статистика по эмпирическим данным (среднее арифметическое, стандартное отклонение и объем выборки);

Collinearity diagnostics – параметры для оценки мультиколлинеарности (связанность независимых переменных).

Параметры регрессионной модели и показатели ее адекватности приводятся в отчете *Regression* (сравните нижеприведенные результаты с результатами, полученными в программе Excel и рассмотренными в подразд. 10.3).

В разделе *Descriptive Statistics* (табл. 10.3) указываются среднее арифметическое (mean), стандартное отклонение (Std. Deviation) и объем выборки (N).

Таблица 10.3

Отчет *Descriptive Statistics*

	Mean	Std. Deviation	N
СПРОС	1182,6667	455,71712	15
ЦЕНА	10,8987	1,14995	15
РЕКЛАМА	18,6267	2,17041	15

В разделе *Correlation* (табл. 10.4) представлены корреляционная матрица результативного и факторных признаков (Pearson Correlation), уровень статистической значимости коэффициентов корреляции (Sig), количество элементов, используемых при расчете коэффициентов корреляции (N).

Таблица 10.4

Отчет *Correlation*

		СПРОС	ЦЕНА	РЕКЛАМА
Pearson Correlation	СПРОС	1,000	-,775	,404
	ЦЕНА	-,775	1,000	-,048
	РЕКЛАМА	,404	-,048	1,000
Sig. (1-tailed)	СПРОС	,	,000	,067
	ЦЕНА	,000	,	,433
	РЕКЛАМА	,067	,433	,
N	СПРОС	15	15	15
	ЦЕНА	15	15	15
	РЕКЛАМА	15	15	15

Раздел *Model Summary* (табл. 10.5) содержит статистику соответствия модели эмпирическим данным: коэффициент множественной корреляции R , коэффициент множественной детерминации (R Square), скорректированный R -квадрат ($Adjusted R$ Square), стандартная ошибка оценки зависимой переменной ($Std.$ Error of the Estimate).

Таблица 10.5

Отчет *Model Summary*

R	R Square	Adjusted R Square	Std. Error of the Estimate
,858	,736	,692	252,95652

Раздел *ANOVA* (табл. 10.6) показывает суммы квадратов отклонений (сумму, обусловленную регрессией, – $Regression$, остаточную – $Residual$, общую – $Total$), F -критерий Фишера, уровень значимости модели (Sig), по которому можно судить о достоверности построенной связи переменных.

Таблица 10.6

Отчет *ANOVA*

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2139649,328	2	1069824,664	16,719	,000
Residual	767844,006	12	63987,000		
Total	2907493,333	14			

В разделе *Coefficients* (табл. 10.7) приводятся значения параметров регрессионной модели и показатели их статистической значимости (постоянный член регрессионного уравнения обозначен $Constant$):

B – значения коэффициентов регрессионного уравнения ($Unstandardized$ Coefficients);

$Std.$ Error – стандартная ошибка коэффициентов;

$Standardized$ Coefficients Beta – стандартные β -коэффициенты регрессионной модели (см. подразд. 10.1);

t – эмпирическое значение t -критерия для проверки статистической значимости соответствующего коэффициента;

Sig – p -уровень значимости коэффициентов (вероятность ошибочного принятия гипотезы о существовании ненулевых коэффициентов регрессии);

Confidence Interval – доверительный интервал изменения соответствующего коэффициента регрессии (по умолчанию строится 95%-ный интервал). По величине стандартной ошибки и доверительному интервалу можно судить о том, в каких пределах могут изменяться значения коэффициентов и соответствующие им реальные экономические показатели (например, эластичность спроса) в генеральной совокупности.

В последнем столбце табл. 10.7 (Collinearity Statistics) приводятся показатели мультиколлинеарности. Значение показателя VIF должно быть меньше 10, чтобы считать, что эффект мультиколлинеарности не наблюдается, и модель с этой точки зрения можно считать приемлемой. Чем выше этот показатель, тем в большей степени связаны переменные. Если для какой-либо переменной значение показателя VIF больше 10, то ее следует исключить из регрессионной модели. В этом случае значение коэффициента множественной детерминации R^2 может уменьшиться, но интерпретация модели будет более осмысленной.

Таблица 10.7

Отчет *Coefficients*

	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B		Collinearity Statistics VIF
	B	Std. Error				Lower Bound	Upper Bound	
Constant	3013,971	888,061		3,394	,005	1079,05	4948,89	
ЦЕНА	-300,171	58,857	-,757	-5,100	,000	-428,41	-171,93	1,002
РЕКЛАМА	77,317	31,184	,368	2,479	,029	9,373	145,262	1,002

Программа SPSS также позволяет построить линию тренда (для парной регрессии). В меню *Graphs* нужно выбрать *Scatter – Simple*, в поле *Y Axis* перенести зависимую переменную (в данном случае – «спрос»), а в поле *X Axis* – независимую переменную (например, «цена»). Для построения линии тренда нужно дважды щелкнуть мышью по полученной диаграмме. В появившемся окне *SPSS Chart Editor* в меню *Chart* выбрать *Options*, затем в разделе *Fit Line* отметить *Total* и щелкнуть по кнопке *Fit Options*. В диалоговом окне *Fit Line* (рис. 10.4) выбрать тип аппроксимирующей функции (например, линейную функцию – *Linear Regression*), при необходимости отметить параметры *Include*

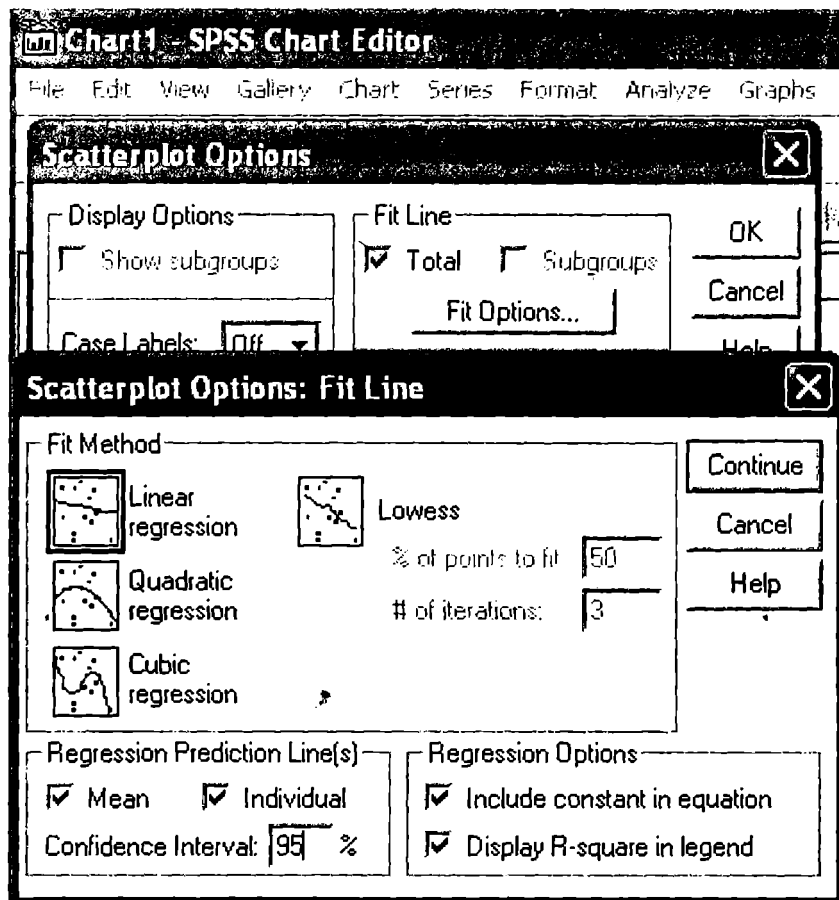


Рис. 10.4. Построение линии тренда в пакете SPSS

constant in equation (включение константы в регрессионное уравнение) и *Display R-square in legend* (вывод на график значения R^2). Центральная линия на рис. 10.5 представляет собой линейный тренд, построенный по методу наименьших квадратов.

Для построения доверительных интервалов прогнозных значений зависимой переменной выбираются параметры раздела *Regression Prediction Line(s)*. По умолчанию строятся 95%-ные доверительные интервалы. При необходимости этот параметр можно изменить, задав нужное значение в окне *Confidence Interval*. Выбор параметра *Mean* приведет к построению интервала *сред-*

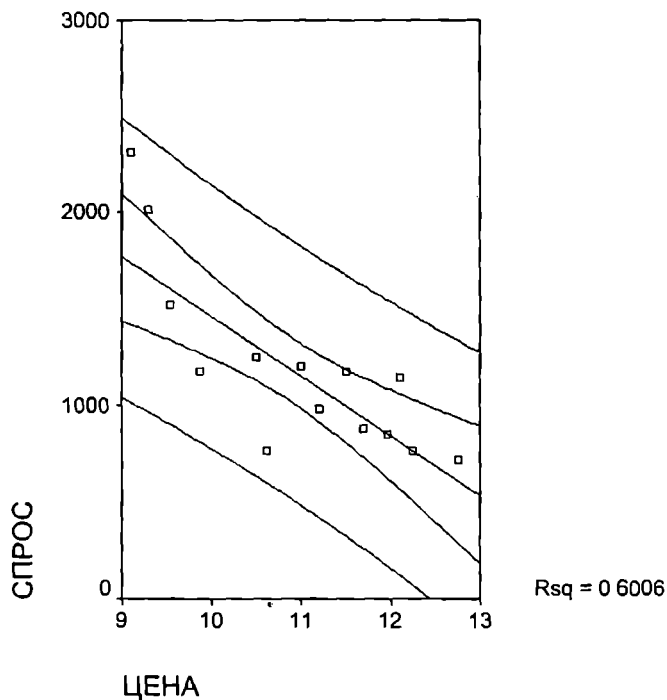


Рис. 10.5. Линейный тренд и доверительные интервалы

них значений результативного признака (см. рис. 10.5). Это означает, что если бы для всех элементов генеральной совокупности, из которой извлечена данная выборка, имеющих некоторое определенное значение x , измерить значение признака y , то *среднее* значение прогнозируемого признака y для данного x будет находиться в данном интервале с вероятностью 95%. Обратите внимание, что ширина доверительного интервала заметно увеличивается у границ изменения факторного признака по выборке. Соответственно в случае экстраполяции (т.е. при выходе за эти границы) точность прогнозирования значительно уменьшается.

Выбор параметра *Individual* в разделе *Regression Prediction Line(s)* (см. рис. 10.4) приведет к построению более широкого доверительного интервала. В этом интервале с вероятностью 95% будет находиться значение прогнозируемого признака y для определенного (заданного) x при выборе из генеральной совокуп-

ности любого элемента с данным значением x . Исследователя, как правило, в большей степени интересует среднее значение прогнозируемого признака (в целом по генеральной совокупности) для некоторого заданного x , поэтому чаще рассматривается доверительный интервал средних в качестве одной из характеристик адекватности модели и точности прогноза на ее основе.

Для сравнения на рис. 10.6 показан тренд, построенный на основе квадратической функции, и соответствующие доверительные интервалы. Сравните результаты с трендом, построенным в Excel (см. рис. 10.2). Данный рисунок еще более наглядно демонстрирует невозможность применения полиномиального тренда для экстраполяции спроса, так как доверительный интервал при $x > 12$ расширяется очень сильно, т.е. точность прогноза значительно снижается.

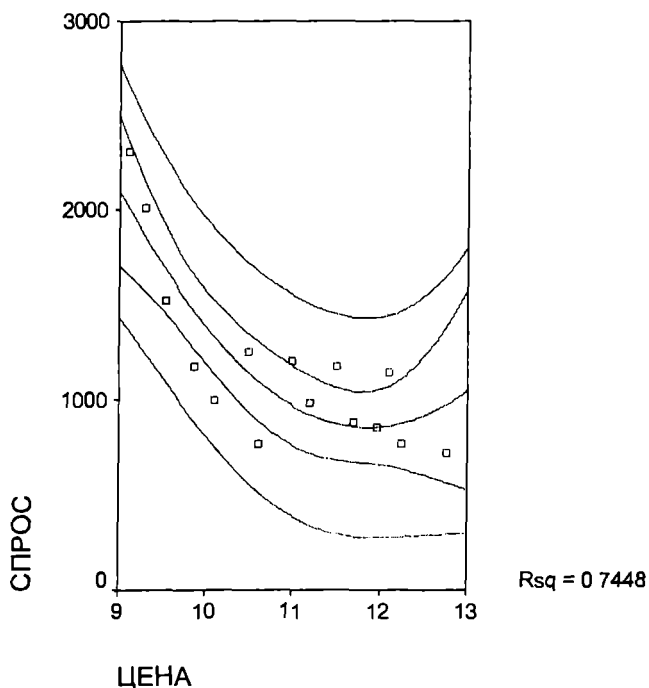


Рис. 10.6. Полиномиальный тренд и доверительные интервалы

На рис. 10.7 представлен линейный тренд для связи «спрос – затраты на рекламу». Эта связь может характеризоваться как прямая линейная умеренная по силе ($r > 0,4$), значимая ($p < 0,1$).

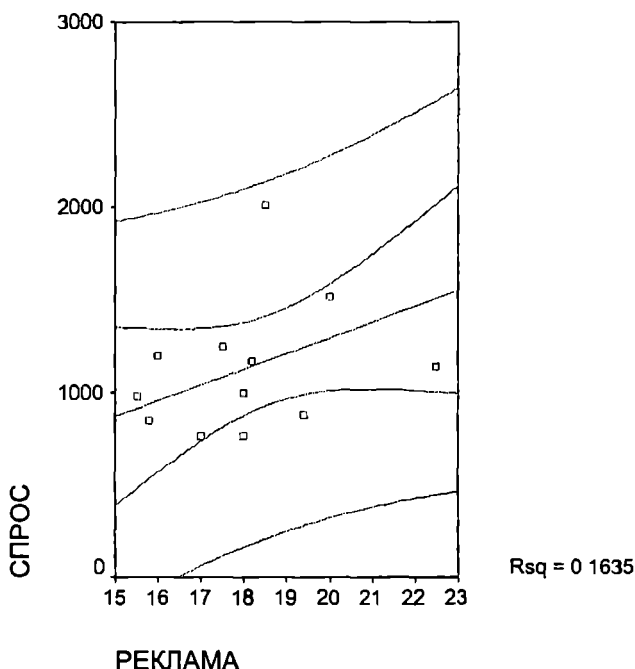


Рис. 10.7. Линейный тренд «спрос – затраты на рекламу»

Более широкие возможности для оценки вида парной регрессионной зависимости между результативным признаком и каждым из факторных признаков обеспечивает инструмент *Curve Estimation* (меню *Analyze – Regression*). В соответствующих полях диалогового окна (рис. 10.8) вводятся зависимая переменная (*Dependent(s)*) и одна из независимых переменных (*Variable*). Отмечаются виды регрессионных моделей, которые должны оцениваться на соответствие эмпирическим данным и статистическую значимость: линейная (*Linear*), квадратичная (*Quadratic*), кубическая (*Cubic*), логарифмическая (*Logarithmic*), экспоненциальная (*Exponential*), S-кривая (*S*), обратная (*Inverse*), логистическая (*Logistic*), показательная – аналог ЛГРФПРИБЛ в Excel

(*Compound*) и др. По умолчанию установлены опции *Include constant in equation* (включать константу в уравнение) и *Plot model* (строить график). Для проверки статистической значимости связи выбирается также *Display ANOVA table* (выводить результаты анализа ANOVA).

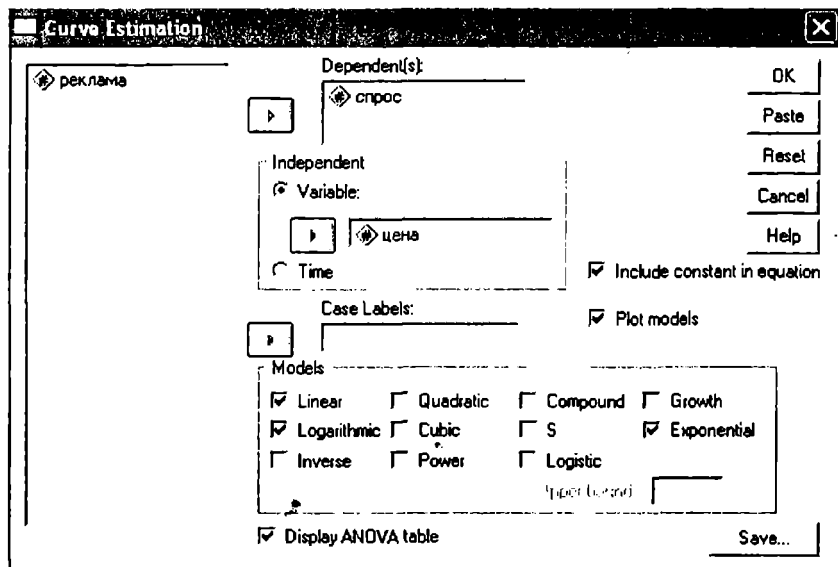


Рис. 10.8. Инструмент Curve Estimation

В отчете *Curve Fit* для каждой независимой переменной приводятся параметры модели и основные статистические показатели, характеризующие ее адекватность и значимость (R , R^2 , Standard Error, Sig. и др.). Ниже приведен фрагмент отчета для трех регрессионных моделей (линейной, логарифмической и экспоненциальной). Содержание всех показателей в отчете рассматривалось ранее. Читателю рекомендуется самостоятельно сравнить все показатели адекватности и значимости трех регрессионных моделей.

Dependent variable..	СПРОС	Method..	LINEAR
Multiple R	,77500		
R Square	,60062		
Adjusted R Square	,56990		
Standard Error	298,86821		

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	1746304,6	1746304,6
Residuals	13	1161188,7	89322,2
F =	19,55062	Signif F = ,0007	

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig
ЦЕНА	-307,126452	69,460347	-,774998	-4,422	,0007
(Constant)	4529,935492	760,948036		5,953	,0000

Dependent variable..	СПРОС	Method..	LOGARITH
Multiple R	,79252		
R Square	,62809		
Adjusted R Square	,59948		
Standard Error	288,40865		

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	1826159,2	1826159,2
Residuals	13	1081334,1	83179,5
F =	21,95443	Signif F = ,0004	

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig
ЦЕНА	-3384,890636	722,409750	-,792519	-4,686	,0004
(Constant)	9250,109506	1723,378481		5,367	,0001

Dependent variable..	СПРОС	Method..	EXPONENT
Multiple R	,79200		
R Square	,62726		
Adjusted R Square	,59859		
Standard Error	,21558		

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	1,0167806	1,0167806
Residuals	13	,6041956	,0464766
F =	21,87727	Signif F = ,0004	

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig
ЦЕНА	-,234353	,050104	-,792000	-4,677	,0004
(Constant)	14348,67320	7875,971378		1,822	,0916

Примечание. Разделы отчета, соответствующие применению разных моделей, обозначаются LINEAR, LOGARITH и EXPONENT.

Из отчета следует, что все рассматриваемые модели имеют приблизительно одинаковую адекватность и значимость¹. Но при выборе вида связи нужно ориентироваться не только на статистические показатели, такие, как R^2 , но и учитывать возможности содержательной интерпретации выбираемой связи. В частности, для данного примера динамика изменения спроса (рис. 10.9) позволяет предположить выход на некоторое плато (постоянный уровень), что лучше отражает модель на основе экспоненциальной зависимости. Использование для данного случая степенной модели вида $Q = a_0 P^{a_1}$ также может быть оправдано, поскольку коэффициент a_1 легко интерпретируется как эластичность спроса: на сколько процентов изменится спрос при изменении цены на 1%. Читателю рекомендуется самостоятельно определить параметры модели и оценить ее адекватность.

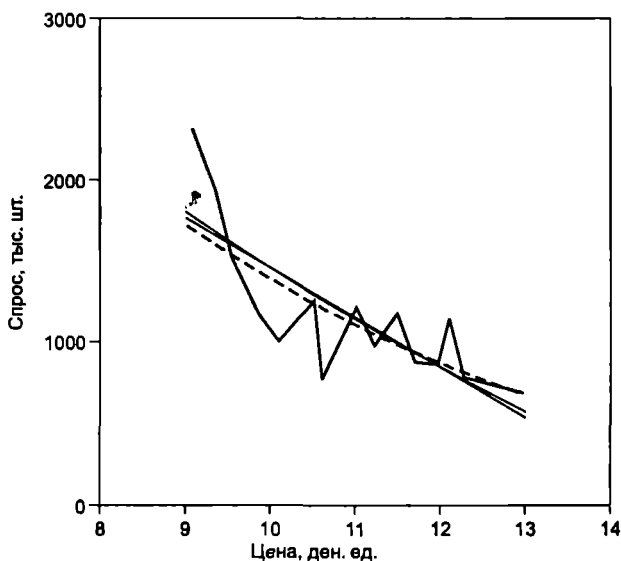


Рис. 10.9. Оценка регрессионной кривой с помощью инструмента Curve Estimation

¹ Обращаем внимание читателя на то, что адекватность моделей разных видов некорректно сравнивать по величине стандартной ошибки оценки (Standard Error), поскольку результативный признак в разных моделях может иметь разный «масштаб» и разные единицы измерения (например, руб. и ln (руб.)).

Прогнозные значения результативного признака, полученные на основе каждой модели, вычисляются с помощью инструмента *Save*. Для этого в таблицу исходных данных должны быть предварительно включены значения факторных признаков, для которых определяются прогнозные значения y . Так, для этого примера к основным данным для построения модели добавлены три пары значений для переменных «цена» и «реклама» (см. рис. 10.3). Значения спроса в этих наблюдениях неизвестны (должны прогнозироваться), поэтому в SPSS они представляются как пропущенные с помощью символа «,».

Инструмент *Save* активизируется после щелчка по кнопке *Save* в диалоговых окнах, где устанавливаются параметры построения регрессионной модели, например в окне *Linear Regression* (см. рис. 10.3) или окне *Curve Estimation* (см. рис. 10.8). Предположим, необходимо построить прогноз спроса для трех пар значений «цена» и «реклама» на основе множественной линейной регрессии (см. рис. 10.3). Для этого в меню *Analyze* нужно выбрать опцию *Regression – Linear*, указать зависимую и независимые переменные и щелкнуть по кнопке *Save*.

В появившемся диалоговом окне *Linear Regression Save* (рис. 10.10) в разделе *Predicted values* (прогнозируемые значения) выбирается опция *Unstandardized* (нестандартизованные значения). Для оценки границ доверительного интервала, в котором будет находиться прогнозируемое значение, можно выбрать параметр *Mean* и (или) *Individual* из раздела *Prediction Intervals* (смысл этих параметров обсуждался выше), а также установить уровень доверительного интервала (в данном случае строится 90%-ный доверительный интервал). После щелчка по кнопкам *Continue* и *OK* в рабочей области данных появляется дополнительная переменная *pre_1*, содержащая теоретические значения зависимой переменной, вычисленные по построенной регрессионной модели (в данном случае линейной), в том числе в точках, в которых требуется построить прогноз. Переменные *lmsci_1* и *umsci_1* показывают нижнюю и верхнюю границы доверительного интервала (рис. 10.11).

Если в диалоговом окне *Linear Regression Save* (см. рис. 10.10) отметить *Residuals Unstandardized*, то в рабочей области данных появляется дополнительная переменная *res_1*, содержащая так называемые остатки (отклонения теоретических значений результативного признака от эмпирических). Это может быть по-

<input type="checkbox"/> Predicted Values <input checked="" type="checkbox"/> Unstandardized <input type="checkbox"/> Standardized <input type="checkbox"/> Adjusted <input type="checkbox"/> S.E. of mean predictions	<input type="checkbox"/> Residuals <input type="checkbox"/> Unstandardized <input type="checkbox"/> Standardized <input type="checkbox"/> Studentized <input type="checkbox"/> Deleted <input type="checkbox"/> Studentized deleted	<input type="button" value="Continue"/> <input type="button" value="Cancel"/> <input type="button" value="Help"/>
<input type="checkbox"/> Distances <input type="checkbox"/> Mahalanobis <input checked="" type="checkbox"/> Cook's <input checked="" type="checkbox"/> Leverage values	<input type="checkbox"/> Influence Statistics <input type="checkbox"/> DfBeta(s) <input type="checkbox"/> Standardized DfBeta(s) <input type="checkbox"/> DfFit <input type="checkbox"/> Standardized DfFit <input type="checkbox"/> Covariance ratio	
<input type="checkbox"/> Prediction Intervals <input checked="" type="checkbox"/> Mean <input type="checkbox"/> Individual Confidence Interval: <input type="text" value="90 %"/>		
<input type="checkbox"/> Save to New File <input type="checkbox"/> Coefficient statistics <input type="text" value="1%"/>		
<input type="checkbox"/> Export model information to XML file <input type="text" value=""/> <input type="button" value="Browse"/>		

Рис. 10.10. Диалоговое окно Linear Regression Save

лезным при проверке качества построенной регрессионной модели. Остатки должны «появляться» случайно (не систематически) и подчиняться нормальному распределению. Это можно проверить по гистограмме остатков, которую можно построить с помощью инструмента *Histogram* из меню *Graph*.

Такую же гистограмму (но для стандартизованных значений со средним 0 и стандартным отклонением 1), а также график для проверки нормальности остатков можно получить, щелкнув по кнопке *Plots* в диалоговом окне *Linear Regression* (см. рис. 10.3) и выбрав в появившемся окне *Linear Regression Plots* параметры *Histogram* и *Normal Probability* раздела *Standardized Residual Plots*. В случае проверки качества парной регрессии можно также построить график, по оси ординат которого откладываются

	цена	реклама	спрос	pre_1	lmci_1	umci_1
1	11,50	21,00	1170,00	1185,663	996,6591	1374,667
2	9,30	18,50	2010,00	1652,747	1448,208	1857,285
3	12,76	20,00	720,00	730,1305	487,3935	972,8675
4	9,87	18,20	1170,00	1458,454	1297,207	1619,701
5	10,10	18,00	1000,00	1373,951	1225,422	1522,480
6	10,50	17,50	1250,00	1215,224	1075,688	1354,760
7	10,61	17,00	770,00	1143,547	992,2122	1294,882
8	11,00	16,00	1200,00	949,1632	762,5403	1135,786
9	11,20	15,50	980,00	850,4705	640,1737	1060,767
10	9,54	20,00	1520,00	1696,681	1500,078	1893,284
11	9,10	22,00	2310,00	1983,390	1698,909	2267,872
12	11,96	15,80	850,00	645,5356	424,2666	866,8046
13	12,10	22,50	1140,00	1121,536	841,6044	1401,468
14	12,24	18,00	770,00	731,5853	546,9383	916,2323
15	11,70	19,40	880,00	1001,922	850,8971	1152,946
16	13,00	15,00	.	271,5041	-42,3963	585,4046
17	9,00	30,00	.	2631,944	1968,021	3295,867
18	12,00	20,00	.	958,2604	775,0547	1141,466

Рис. 10.11. Эмпирические и теоретические значения спроса (линейная модель)

остатки, а по оси абсцисс – значения независимой переменной (с помощью инструмента *Graphs – Scatter – Simple*). Полученная диаграмма распределения должна свидетельствовать об отсутствии корреляции, а точки должны выстраиваться в симметричную относительно нулевой линии полосу.

Для моделирования связей, которые нельзя линеаризовать, в пакете SPSS используется итерационная процедура оценки коэффициентов нелинейной регрессионной модели *Nonlinear* (меню *Analyze – Regression*).

В диалоговом окне *Nonlinear Regression* зависимую переменную нужно перенести в соответствующее поле (*Dependent*). В поле *Model Expression* вводится формула предполагаемой связи зависимой переменной и одной или нескольких независимых переменных, с использованием соответствующих символов и функций. В формулу связи кроме имен независимых переменных должны быть включены коэффициенты – параметры регрессионной модели, которые будут оцениваться с помощью итерационной процедуры.

Продemonстрируем использование этой процедуры на следующем примере. Предположим, зависимость спроса (Q) от цены (P) и затрат на рекламу (A) моделируется следующей функцией, включающей как линейную, так и нелинейную составляющие:

$$Q = a_0 e^{-a_1 P} + a_2 A.$$

Данный вид связи выбран в предположении, что цена является основным фактором, определяющим спрос и оказывающим на него обратное нелинейное влияние (см. рис. 10.9 и табл. 10.7), а затраты на рекламу оказывают прямое умеренное линейное влияние на спрос (см. рис. 10.7 и табл. 10.7).

Хотя связь такого вида является «линейной по существу», т.е. может быть линеаризована, с практической точки зрения удобнее воспользоваться процедурой нелинейной регрессии. В диалоговом окне *Nonlinear Regression* определяем зависимую переменную и вводим соответствующую формулу (рис. 10.12).

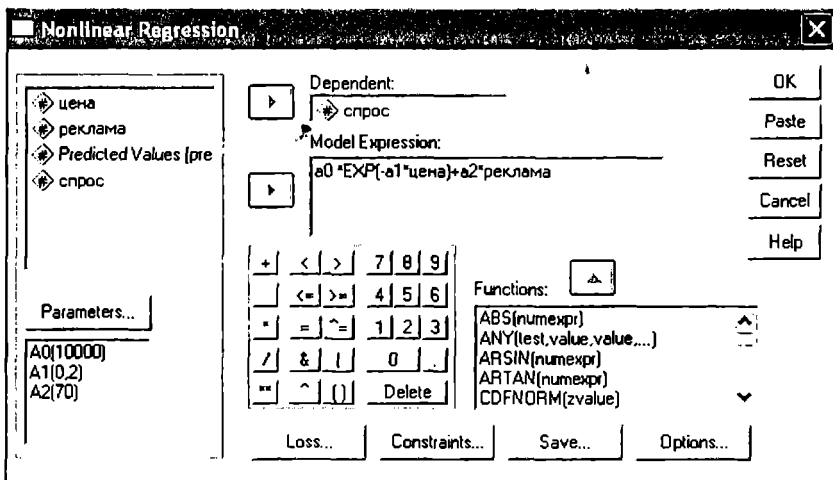


Рис. 10.12. Диалоговое окно *Nonlinear Regression*

Необходимо также задать начальные значения параметров регрессии (коэффициенты a_0 , a_1 , a_2), щелкнув по кнопке *Parameter* (параметр). В появившемся диалоговом окне укажите в поле имен (*Name*) имя первого параметра, а в поле *Starting value* — его начальное значение, затем щелкните по кнопке *Add* (Добавить), и так для каждого параметра регрессионной модели (рис. 10.13).

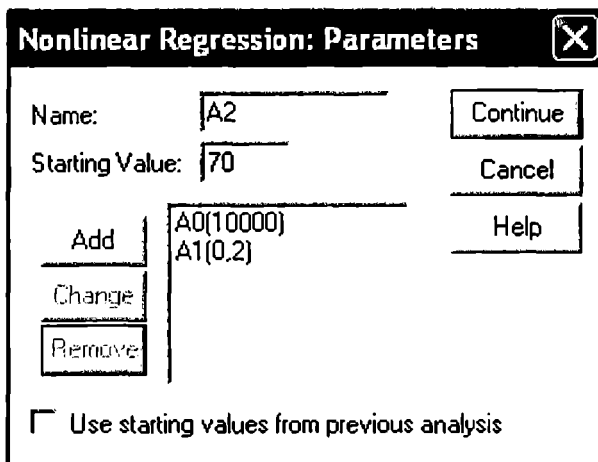


Рис. 10.13. Установка начальных значений параметров регрессии

Если необходимо вывести прогнозные значения результативного признака на основе данной модели, щелкните по кнопке *Save* и в открывшемся окне отметьте опцию *Predicted Values* (прогнозируемые значения).

В отчете *Nonlinear Regression* отражаются результаты итерационного процесса подбора значений коэффициентов регрессионного уравнения:

- остаточная сумма квадратов и значения коэффициентов на каждой итерации;
- стандартная ошибка и доверительный интервал для каждого коэффициента;
- значение R^2 и корреляционная матрица оценок параметров. Очень высокие абсолютные значения корреляций ($> 0,8$) между коэффициентами перед *независимыми переменными* в модели указывают на то, что модель содержит неоправданно большое количество параметров, т.е. модель и с меньшим количеством независимых переменных может обеспечить столь же хорошее приближение.

Ниже приведен фрагмент отчета *Nonlinear Regression* для рассматриваемого примера конструирования и тестирования нелинейной модели влияния цены и рекламных затрат на спрос.

Iteration	Residual SS	A0	A1	A2
1	25381002,13	10000,0000	,200000000	70,0000000
1.1	2107040,121	28737,1753	,469608583	63,3285167
2	2107040,121	28737,1753	,469608583	63,3285167

68	431746,9159	787680905	1,46527148	48,2784085
68.1	431746,9150	788124305	1,46533292	48,2788194

Run stopped after 155 model evaluations and 68 derivative evaluations. Iterations have been stopped because the relative reduction between successive residual sums of squares is at most $SSCON = 1,000E-08$

Nonlinear Regression Summary Statistics

Dependent Variable СПРОС

Source	DF	Sum of Squares	Mean Square
Regression	3	23456253,0850	7818751,02835
Residual	12	431746,91496	35978,90958
Uncorrected Total	15	23888000,0000	
(Corrected Total)	14	2907493,33333	

R squared = $1 - \text{Residual SS} / \text{Corrected SS} = ,85151$

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval	
			Lower	Upper
A0	788124305,02*	3254856317,7	-6303598399	7879847008
A1	1,465332919	,451685509	,481194737	2,449471101
A2	48,278819419	4,495160997	38,484704966	58,07293387

Из отчета следует, что на первой итерации приняты начальные значения параметров модели, для которой оценивается остаточная сумма квадратов. Далее параметры модели изменяются, и остаточная сумма квадратов уменьшается. После этого осуществляется переход к следующей итерации (на каждой итерации может быть сделано несколько оценок для улучшения значений параметров). В данном примере для достижения заданной точности потребовалось 68 итераций и 155 оценок параметров модели. R^2 для данной модели составляет 0,85, что превышает соответствующий показатель для линейной регрессионной модели с данными факторными признаками (0,74) (см. табл. 10.5). Это свидетельствует о лучшей адекватности модели и, с учетом сделанных предположений о характере связи спроса с ценой и затратами на рекламу, о ее большей пригодности для прогноза. Значение R^2 для данной модели также превышает сумму коэффициентов

парной детерминации для независимых переменных на основе соответствующих моделей (0,63 для парной экспоненциальной модели «спрос – цена» и 0,16 для парной линейной регрессии «спрос – затраты на рекламу»). Но в данном случае это может быть связано не с «синергетическим эффектом», а с «техническим» увеличением значения показателя R^2 (см. подразд. 10.2).

Недостатком рассмотренной процедуры нелинейной регрессии является то, что не для всех оцениваемых моделей в отчетах приводятся показатели статистической значимости модели и ее параметров, а также показатели, позволяющие сделать более полное заключение о прогнозной силе конструируемых регрессионных моделей (хотя в принципе об этом можно судить и на основе приводимых значений сумм квадратов и ошибок коэффициентов регрессии). Поэтому в случае необходимости проведения более полного анализа рекомендуется использовать линеаризованные регрессионные модели.

Аналогичная процедура построения нелинейной регрессии (даже с несколько большими возможностями для анализа) реализована в пакете STATISTICA. Для ее запуска нужно переключиться в модуль *Nonlinear Estimation* (нелинейная оценка) и в открывшемся меню выбрать *User-specified regression* (определяемая пользователем регрессия). В окне *Function to be estimated...* (оцениваемая функция) вводится формула требуемого регрессионного уравнения. В окне *Loss function* (функция потерь) по умолчанию введена функция (OBS-PRED)**2, что соответствует квадрату разностей наблюдаемых и предсказанных значений, т.е. по умолчанию используется метод наименьших квадратов для оценки параметров. В окне *Model Estimation* (модель оценки) задаются начальные значения параметров (*Start value*), начальные величины их изменения (*Initial step size*), количество итераций (*Maximum number of iterations*). Для всех этих параметров задаются некоторые значения по умолчанию, но с заданными по умолчанию начальными значениями процедура определения параметров не всегда может быть успешно завершена (в этом случае появится сообщение *Error in function: Change start values/precision/step size*, предупреждающее о том, что нужно изменить начальные значения параметров). Для получения ошибок оценок определяемых параметров нужно выбрать опцию *Asymptotic standard errors*.

В случае множественной регрессии с большим числом факторов необходимо классифицировать эти переменные по степени их важности для предсказания зависимой переменной (y). Затем следует исключить из анализа факторы, наименее существенные для предсказания y , а также переменные, сильно коррелированные с другими, уже включенными в модель факторами. Эту задачу можно определить как выбор «наилучшей» регрессии, т.е. определение минимального набора факторов, достаточно точно предсказывающих значение резульативного признака.

Одним из методов отбора наиболее существенных факторов является пошаговая регрессия (*Stepwise Regression*). Рассмотрим две процедуры пошаговой регрессии, реализованные в пакете SPSS: обратная (*Backward*) и прямая пошаговая регрессия (*Forward*).

Обратная пошаговая регрессия заключается в том, что последовательно исключаются наименее значимые факторы. На нулевом шаге проводится регрессионный анализ для всех факторов. Каждый фактор проверяется на значимость. Если статистический показатель значимости больше критического значения, определяющего *величину F-удаления (F-to remove)*, то фактор исключается из анализа и строится новое уравнение регрессии по оставшимся факторам (по умолчанию критическое значение для удаления задается на уровне 0,1).

Прямая пошаговая регрессия организована в противоположном направлении: на первом шаге в уравнение регрессии включается фактор, имеющий наибольший коэффициент корреляции с y , и проверяются адекватность и значимость модели. Если модель значима, включается следующий фактор и вычисляется *F-статистика* для каждой переменной модели. Если статистический показатель значимости какой-либо переменной больше критической величины, то фактор исключается, если меньше — сохраняется, и в уравнение включается следующая переменная. Поскольку проверка всех выбранных переменных осуществляется на каждом шаге, может оказаться, что переменная, включенная в уравнение на предыдущем шаге, может быть исключена на следующих шагах.

Процедура пошаговой регрессии позволяет значительно сократить объем работы при конструировании адекватной и значимой регрессионной модели. Проиллюстрируем работу этой процедуры на примере.

Пример 10.2. Специалист агентства по недвижимости запросил данные о 12 недавних продажах коттеджей (цена, площадь дома, количество этажей, количество ванн и возраст дома). Эти данные приведены в табл. 10.8 [8] (значение 1,5 в столбце «Количество ванн комнат» означает, что в доме имеется одна «полная» ванная комната и один дополнительный санузел). Необходимо построить на основе этих данных модель множественной регрессии для оценки влияния основных факторов и прогнозирования цены на коттеджи.

Таблица 10.8

Данные для построения множественной регрессии

Цена, млн руб.	Площадь, м ²	Этажность	Количество ванн комнат	Возраст дома, лет
4,965	89	1	1,0	2
6,795	95	1	1,0	6
8,115	126	2	1,5	11
8,160	129	2	1,5	8
9,150	190	2	1,0	22
9,525	176	1	1,0	17
10,035	200	2	1,5	12
10,425	206	2	1,5	11
11,265	205	1	2,0	9
14,970	251	2	2,0	8
16,065	227	2	2,0	18
23,250	408	3	4,0	12

Построение регрессионных моделей на основе пошаговой регрессии в пакете SPSS практически не отличается от рассмотренной ранее процедуры построения множественной линейной регрессии. Выбор режима пошаговой регрессии осуществляется в окне *Linear Regression* (см. рис. 10.3) в поле *Method*, все остальные действия аналогичны рассмотренным выше.

Корреляционная матрица (табл. 10.9) показывает, что переменные «площадь», «этаж», «ван_комн» имеют сильную значимую связь с зависимой переменной «цена», но в то же время все эти переменные обнаруживают сильные корреляции и друг с другом, поэтому они не могут считаться независимыми. В регрессионное уравнение должны включаться лишь какие-то из них для исключения эффекта мультиколлинеарности. Для переменной «возраст» не обнаруживается даже умеренная и значимая

корреляция с ценой, поэтому данная переменная при включении ее в линейную регрессионную вряд ли будет заметно улучшать адекватность модели.

Таблица 10.9

Коэффициенты корреляции и их значимость

		ЦЕНА	ПЛОЩАДЬ	ЭТАЖ	ВАН_КОМН	ВОЗРАСТ
Pearson Correlation	ЦЕНА	1,000	,962	,706	,924	,296
	ПЛОЩАДЬ	,962	1,000	,716	,895	,337
	ЭТАЖ	,706	,716	1,000	,701	,318
	ВАН_КОМН	,924	,895	,701	1,000	,026
	ВОЗРАСТ	,296	,337	,318	,026	1,000
Sig. (1- tailed)	ЦЕНА		,000	,005	,000	,175
	ПЛОЩАДЬ	,000		,004	,000	,142
	ЭТАЖ	,005	,004		,006	,157
	ВАН_КОМН	,000	,000	,006		,468
	ВОЗРАСТ	,175	,142	,157	,468	

Проверим эти предположения и построим регрессионную модель методом пошаговой регрессии (режим Backward). Отчеты пакета SPSS с результатами регрессионного анализа полностью аналогичны рассмотренным ранее за исключением того, что все статистические показатели приводятся для каждого шага построения регрессионной модели. Из раздела *Variables Entered/Removed* (табл. 10.10) следует, что вся процедура состояла из четырех шагов, т.е. фактически оценивались четыре регрессионные модели. В первую модель (первый шаг) входили все четыре переменные, рассматриваемые как независимые. Это соответствует методу *Enter* (включение в регрессию всех переменных). На каждом следующем шаге исключается по одной переменной, для которой уровень значимости, определяемый на основе *F*-статистики, равен или превышает 0,1.

В разделе *Model Summary* (табл. 10.11) показано, как изменяются показатели адекватности модели на каждом шаге. Видно, что при исключении каждой следующей переменной значение R^2 уменьшается незначительно, а скорректированный R^2 может даже увеличиться.

Отчет *Variables Entered/Removed*

Model	Variables Entered	Variables Removed	Method
1	ВОЗРАСТ, ВАН_КОМН, ЭТАЖ, ПЛО- ЩАДЬ		Enter
2		ЭТАЖ	Backward (criterion: Probability of F-to-remove > = ,100).
3		ВОЗРАСТ	Backward (criterion: Probability of F-to-remove > = ,100).
4		ВАН_КОМН	Backward (criterion: Probability of F-to-remove > = ,100).

Таблица 10.11

Отчет *Model Summary*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,976	,952	,925	1,35073
2	,975	,952	,933	1,27557
3	,972	,945	,933	1,28398
4	,962	,925	,918	1,41988

Статистическая значимость модели (табл. 10.12) на каждом шаге остается очень высокой ($p < 0,001$). При этом, судя по возрастающим значениям F -критерия, исключение одной из переменных увеличивает значимость модели.

Таблица 10.12

Отчет *ANOVA*

Model		Sum of Squares	df	Mean Square	F	Sig
1	Regression	256,085	4	64,021	35,091	,000
	Residual	12,771	7	1,824		
	Total	268,857	11			
2	Regression	255,840	3	85,280	52,413	,000
	Residual	13,017	8	1,627		
	Total	268,857	11			
3	Regression	254,019	2	127,010	77,041	,000
	Residual	14,837	9	1,649		
	Total	268,857	11			
4	Regression	248,696	1	248,696	123,35	,000
	Residual	20,161	10	2,016		
	Total	268,857	11			

Причины исключения переменных из регрессионной модели можно понять из отчета *Coefficients* (табл. 10.13). На первом шаге обнаруживается, что коэффициент перед переменной «этаж» незначим ($p = 0,725$), при этом стандартный коэффициент β для этой переменной очень мал по абсолютному значению (0,047), поэтому данная переменная исключена (на втором шаге она отсутствует в модели). По этой же причине на следующих шагах исключаются переменные «возраст» и «ван_комн». На четвертом шаге остается одна переменная «площадь». Но в данной модели константа не является статистически значимым параметром ($p = 0,716$), поэтому ее следует исключить из модели (т.е. пересчитать параметры модели с включенной переменной «площадь», но без включения константы в регрессионное уравнение).

Таблица 10.13

Отчет *Coefficients*

Model	Unstandardized Coefficients		Standardized Coefficients Beta	Sig.	Collinearity Statistics VIF
	B	Std. Error			
1 (Constant)	-,138	1,336		,921	
ПЛОЩАДЬ	2,852E-02	,015	,493	,100	10,003
ЭТАЖ	-,371	1,013	-,047	,725	2,388
ВАН_КОМН	3,029	1,556	,511	,093	10,178
ВОЗРАСТ	,117	,110	,131	,323	2,226
2 (Constant)	-,324	1,167		,788	
ПЛОЩАДЬ	2,867E-02	,014	,496	,079	9,996
ВАН_КОМН	2,824	1,372	,477	,074	8,865
ВОЗРАСТ	,104	,098	,116	,321	1,986
3 (Constant)	,410	,945		,675	
ПЛОЩАДЬ	3,927E-02	,010	,679	,004	5,037
ВАН_КОМН	1,870	1,041	,316	,106	5,037
4 (Constant)	,390	1,044		,716	
ПЛОЩАДЬ	5,562E-02	,005	,962	,000	1,000

Полезным для анализа также является раздел *Excluded Variables* (табл. 10.14), где приводятся значения стандартных коэффициентов β (Beta In), p -уровня значимости (Sig), коэффициента частной корреляции соответствующей переменной с результативным признаком (Partial Correlation) и других показателей для исключенных переменных, если бы они все же были включены в модель.

Отчет *Excluded Variables*

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics VIF
2	ЭТАЖ	-,047	-,367	,725	-,137	2,388
3	ЭТАЖ	-,002	-,018	,986	-,006	2,130
	ВОЗРАСТ	,116	1,058	,321	,350	1,986
4	ЭТАЖ	,037	,281	,785	,093	2,052
	ВОЗРАСТ	-,032	-,336	,745	-,111	1,128
	ВАН_КОМН	,316	1,797	,106	,514	5,037

В данном примере, рассматривая причины исключения переменных, мы обращали внимание лишь на статистические показатели. В реальном исследовании необходимо обращаться и к смысловому содержанию переменных. Исследователь может принять решение остановиться не на последнем шаге конструирования модели, а на одном из предыдущих шагов, если, по его мнению, некоторые переменные очень важны для объяснения и прогнозирования результативного признака. Но в этом случае, возможно, понадобятся дополнительные данные (наблюдения, измерения и т.п.) для достижения приемлемой статистической значимости модели и ее параметров. Необходимо помнить также, что для принятия окончательных выводов по результатам регрессионного анализа должны дополнительно проверяться предпосылки его применения (см. подразд. 10.1).

Подобную процедуру последовательного исключения переменных можно осуществить, последовательно выполняя процедуры линейной или нелинейной (после линеаризации уравнений) регрессии по методу *Enter*, как было рассмотрено ранее в данном подразделе. Но метод пошаговой регрессии сокращает трудоемкость и время выполнения данных процедур, представляя в одном отчете результаты сразу нескольких этапов.

Теоретически процедуры прямой и обратной пошаговой регрессии должны приводить к одной и той же регрессионной модели при условии независимости всех рассматриваемых переменных. В случае, если какие-либо переменные (например, x_i и x_j) сильно коррелируют, возможна ситуация, когда процедура прямой пошаговой регрессии приведет к включению в модель переменной x_i , а обратной — x_j . Более часто на практике используют процедуру обратной пошаговой регрессии.

В заключение еще раз обратим внимание читателя на то, что методы корреляционного и регрессионного анализа тесно связаны между собой: с помощью первых оценивается сила (теснота) связи, с помощью вторых исследуется ее форма. И те, и другие методы служат для установления соотношений между признаками и для определения вида и силы связи между ними. Поэтому эти методы используются комплексно.

Корреляционно-регрессионному анализу связей исследуемых признаков должен предшествовать качественный анализ изучаемого социального или экономического явления с позиций, например, экономической теории, социологии или социальной психологии: связи какого вида и какой силы возможны в исследуемой ситуации. На следующем этапе данные должны быть проверены на предмет аномальности или наличия «выбросов», т.е. наблюдений, резко выделяющихся из массива данных. После этого строится модель связи с помощью математико-статистических методов, выбор которых зависит от целей исследования и характера поставленных задач, а также от того, насколько имеющиеся данные удовлетворяют условиям для использования тех или иных статистических инструментов. Заключительный этап — интерпретация результатов — вновь связан с качественным анализом особенностей изучаемого явления. Таким образом, статистическое решение любой задачи должно основываться на корректности и объективности сбора исходных данных и неразрывном сочетании математико-статистического анализа с содержательным анализом.

Задания для самостоятельного выполнения¹

1. Автосервисное предприятие имеет следующие данные по стоимости ежегодного технического обслуживания автомобилей определенной категории в зависимости от времени эксплуатации (табл. 10.15). Постройте по этим данным наиболее адекватную регрессионную модель для последующей предварительной оценки стоимости технического обслуживания.

¹ Для работы над заданиями 1–8 проверьте выполнение предпосылок с целью применения регрессионного анализа, обоснуйте выбор модели, оцените ее адекватность и статистическую значимость, объясните смысл всех параметров модели.

Данные для решения задачи 1

Стоимость технического обслуживания, тыс. руб.	5,3	5,2	6,0	5,7	6,6	6,8	8,1	6,9	10,3	4,0	2,5
Время эксплуатации, лет	5	4	5	6	7	8	10	8	11	3	2

2. В табл. 10.16 приведены данные [8], характеризующие стоимость продукции в зависимости от количества исходного сырья. Постройте по этим данным наиболее подходящую регрессионную модель для последующей предварительной оценки стоимости продукции.

Таблица 10.16

Данные для решения задачи 2

Стоимость продукции, тыс. руб.	10	7	5	6	7	6	4	3	8	9
Количество исходного сырья, т	0,25	0,20	0,16	0,17	0,19	0,18	0,15	0,14	0,21	0,21

3. В табл. 10.17 представлен индекс цен на группу товаров в стране 1 по сравнению с ценами на аналогичную группу товаров в стране 2, а также отношение национальных валют этих стран. Можно ли с помощью регрессионного анализа выявить и объяснить закономерность влияния роста цен на валютный курс?

Таблица 10.17

Отношение цен и курсов национальных валют

Отношение цен	Отношение курсов валют	Отношение цен	Отношение курсов валют
0,6	0,5	1,5	1,3
0,65	0,4	1,8	2
0,8	0,45	1,9	1,3
0,78	0,52	2,9	2,6
1,1	1,3	2,95	2,05
1,4	1,05		

4. По данным табл. 10.18 с помощью регрессионного анализа определите зависимость между выпуском продукции и размером основных фондов.

Таблица 10.18

Данные для решения задачи 4

Размер основных фондов, млн руб.	4,3	5,4	3,6	6,9	3,9	4,7	4,0	6,4	5,5	6,8
Выпуск продукции, тыс. руб.	22,4	18,6	13,1	25,1	10,2	19,2	15,7	23,4	16,0	21,5

5. Для 13 торговых предприятий получены данные о розничном товарообороте в млн руб. (y) и издержках обращения (x):

x	640	405	200	425	570	472	250	380	550	750	660	450	563
y	39	26	15	34	37	28	19	24	38	44	36	27	34

Подберите адекватную регрессионную модель для этих данных.

6. По результатам отчетов о работе 20 предприятий одного профиля (табл. 10.19) постройте адекватную регрессионную модель для определения зависимости накладных расходов от объема производства.

Таблица 10.19

Данные для решения задачи 6, млн руб.

Объем работ	Накладные расходы	Объем работ	Накладные расходы
12,5	4,0	9,3	2,6
8,0	2,0	8,0	3,2
7,0	2,5	10,2	2,3
5,2	2,2	10,0	3,0
6,4	2,5	12,0	2,6
9,5	2,7	15,0	3,0
14,0	4,0	16,0	5,0
13,0	4,0	17,0	4,3
5,0	2,0	21	5,0
7,4	2,6	19,0	4,8

7. По данным табл. 10.20 постройте обоснованную модель связи цены на ЖК-телевизоры и комплексного показателя качества (качество изображения, процент брака, дополнительные функции и т.д.).

Таблица 10.20

Цена телевизора и показатель его качества

Цена, у.е.	Качество, балл	Цена, у.е.	Качество, балл	Цена, у.е.	Качество, балл
429	4	486	8	465	6
449	4	455	5	433	3
473	8	473	7	425	4
485	8	479	9	459	5
446	4	467	8	417	2
423	2	442	5	418	5
423	3	441	5	487	8
463	7	499	9	491	9
428	3	478	9	508	9
479	7	466	7	463	6

8. По данным табл. 10.21 постройте прогноз (90%) численности населения региона на период 2000–2010 гг.

Таблица 10.21

Численность населения региона

Год	Население, тыс. чел.	Год	Население, тыс. чел.
1940	2264,9	1991	2654,5
1950	2261,5	1995	2697,2
1960	2594,1	1997	2678,4
1970	2502,0	1998	2679,6
1980	2506,1	1999	2682,3
1990	2640,5		

9. Дирекция электростанции, обслуживающей южный курортный город, с целью прогнозирования потребности электроэнергии определяет, как зависит количество электроэнергии, потребляемой организациями и жителями города, от температуры воздуха (предполагается, что в жаркие дни возрастают расходы электроэнергии на кондиционирование помещений). Данные по 25 случайно отобраным летним дням приведены в табл. 10.22. Постройте прогноз потребления электроэнергии при температуре 20, 25, 30, 35 и 40°C с 90 и 95%-ными доверительными интервалами.

Потребление электроэнергии и температура воздуха

Потребление, МВт	Температура, °С	Потребление, МВт	Температура, °С
89	29,5	84	27
94	32	79	24,5
100	35	85	27,5
97	33,5	89	29,5
95	32,5	68	19
100	35	98	34
92	31	86	28
106	38	87	28,5
67	19,5	108	39
68	19,5	96	33
74	22	76	23
88	29	71	20,5
90	30		

10. По данным табл. 9.12 определите возможные значения всех показателей (с 90%-ным доверительным интервалом) в первые три месяца следующего года.

11. По данным табл. 9.13 проанализируйте ретроспективную информацию по изменению цен на продукты и постройте прогноз (90%-ный доверительный интервал) изменения цены каждого продукта в последующие периоды времени. Прогнозный период примите равным $1/3$ ретроспективного периода.

12. По данным табл. 10.23 определите, как влияет цена порции мороженого в кафе на количество заказов. В каком интервале будет лежать количество заказов при цене мороженого 55 руб. с вероятностью 95%?

Таблица 10.23

Цена порции мороженого и количество заказов

Цена, руб.	45	48	50	40	47	53	56	60	62
Заказ, шт.	6040	6544	8575	9120	5876	6975	7021	5287	5962

13. По данным табл. 9.16 постройте адекватную регрессионную модель для прогнозирования спроса. Дайте интерпретацию параметров модели с учетом доверительных интервалов. Постройте 95%-ный интервальный прогноз продаж при следующих затратах на рекламу: 112, 99, 130, 75 ден. ед.

14. В табл. 10.24 представлены ежемесячные объемы продаж новой модели DVD-проигрывателя (тыс. руб.) за два года. Можно ли считать рост продаж линейным? Постройте адекватную регрессионную модель, на основе которой спрогнозируйте объем продаж на три последующих месяца (с 95%-ным доверительным интервалом).

Таблица 10.24

Данные для решения задачи 14

x	1	2	3	4	5	6	7	8	9
y	23	156	330	482	1209	1756	2000	2512	2366
x	10	11	12	13	14	15	16	17	18
y	2942	2872	2937	3136	3241	3149	3524	3542	3312
x	19	20	21	22	23	24			
y	3547	3376	3375	3403	3697	3495			

15. В табл. 10.25 представлена выборка значений цены акций некоторой компании за 1 ч работы биржи. Спрогнозируйте изменение цены на период 12:10–12:30 с интервалом 5 мин.

Таблица 10.25

Изменение цены акций

Время	Цена сделки, руб./акция	Время	Цена сделки, руб./акция
11:16:45	44,5	11:53:51	36,6
11:21:53	44	11:55:05	36,5
11:23:09	43,1	11:55:24	37
11:23:37	43,1	11:58:18	38
11:24:49	40	11:58:20	37
11:24:57	40	11:58:24	36,5
11:48:40	36,1	11:58:35	38
11:49:45	39,9		

Пояснение. При выполнении задания в статистических пакетах установите соответствующий формат переменной «Время». При выполнении задания в Excel следует перевести значения в столбце Время в интервальную шкалу с условным началом отсчета 0 и с единицей измерения 1 с, используя для этого функции МИНУТЫ() и СЕКУНДЫ().

16. Проведите регрессионный анализ данных, представленных в табл. 10.1, и сравните с результатами, полученными в подразд. 10.3 и 10.4.

17. По результатам отчетов (млн руб.) о работе 30 малых предприятий (табл. 10.26) [8] постройте адекватную регрессионную модель для последующей предварительной оценки производства продукции.

Таблица 10.26

Данные для решения задачи 17

Среднегодовая стоимость основных производственных фондов	Производство продукции	Среднегодовая стоимость основных производственных фондов	Производство продукции
47	40	27	21
36	35	28	35
56	60	41	38
57	48	44	46
45	43	55	51
39	45	33	30
46	48	37	38
60	46	49	50
55	57	56	61
53	34	37	34
42	42	38	35
41	47	49	39
35	30	26	19
33	41	29	36
46	27	20	24

18. Торговый дом, специализирующийся на аукционной продаже антиквариата, предоставил данные (табл. 10.27) о продажах на последнем аукционе: окончательная цена (y), возраст объекта (x_1), количество заявок на покупку (x_2). На основе соответствующей регрессионной модели проследите, как цена объекта изменяется с его возрастом и количеством заявок на покупку. Связано ли количество заявок с возрастом объекта?

Данные для решения задачи 18

x_1	113	126	115	182	150	127	159	117	175	168
x_2	9	10	7	11	9	13	9	13	8	7
y	2365	3340	1860	4947	3805	3087	3707	2880	3862	3155
x_1	182	156	179	108	143	187	111	137	194	156
x_2	8	6	9	6	6	8	15	15	5	12
y	3875	2617	4480	1822	2135	3982	2937	4282	3390	4555
x_1	127	108	132	137	137	162	117	170	111	153
x_2	7	14	10	9	8	11	11	14	7	6
y	2112	2637	3132	3242	2867	4710	2560	5327	1962	2730

19. Компании грузоперевозок необходимо построить модель для оценки стоимости технического обслуживания автомобилей. Предложите факторы, которые могут влиять на стоимость технического обслуживания (5–10 факторов). Какие из этих факторов могут рассматриваться в качестве независимых переменных регрессионной модели затрат? Предположите характер и степень влияния этих факторов на результативный признак и предложите возможные варианты регрессионной модели.

20. По данным табл. 10.1 постройте мультипликативную модель вида $Q = a_0 P^{a_1} A^{a_2}$. Дайте экономическую интерпретацию коэффициентов уравнения и их доверительных интервалов.

21. На основе соответствующей регрессионной модели определите, какое влияние оказывают вес бандероли и расстояние до пункта назначения на цену за доставку бандероли, устанавливаемую почтовой службой (табл. 10.28). Насколько хорошо эти признаки объясняют изменчивость результативного признака? Сравните значения R^2 и скорректированного R^2 . Во сколько обойдется отправка бандероли весом 1,5 кг на расстояние 300 км?

Таблица 10.28

Данные для решения задачи 21

Цена доставки, руб.	Вес, кг	Расстояние до пункта назначения, км	Цена доставки, руб.	Вес, кг	Расстояние до пункта назначения, км
66	1,64	190	120	2,48	230
40	0,12	320	34	0,44	180
220	2,04	480	290	2,6	480

Цена доставки, руб.	Вес, кг	Расстояние до пункта назначения, км	Цена доставки, руб.	Вес, кг	Расстояние до пункта назначения, км
52	2,36	94	280	3	380
38	1,8	106	184	2,64	320
160	1,4	500	22	1,08	90
310	2,8	520	242	3,24	320
100	0,96	418	30	0,28	160
20	0,24	200	160	1,76	404
88	0,3	560	78	1,28	290

22. Используя данные табл. 10.29 [54], транспортное управление города выявляет наиболее значимые факторы, определяющие пассажиропоток в метро. До проведения регрессионного анализа сделайте предположение о возможном характере связи признаков и их влиянии на среднее количество пассажиров метро. Постройте подходящую регрессионную модель. Насколько хорошо модель объясняет изменение пассажиропотока?

Используя в качестве независимой переменной время (год), постройте соответствующие однофакторные регрессионные модели, на основе которых сделайте прогноз изменения всех переменных в 2005–2008 гг. Оцените точность этих прогнозов.

Таблица 10.29

Данные для решения задачи 22

Год	Еженедельный пассажиропоток, тыс. чел.	Стоимость проезда, долл.	Население, тыс. чел.	Среднегодовой доход, долл.	Стоимость парковки, долл.
1978	1200	0,35	1800	11600	0,50
1979	1190	0,35	1790	12400	0,50
1980	1195	0,35	1780	12800	0,60
1981	1100	0,45	1778	13000	0,60
1982	1105	0,45	1750	13100	0,60
1983	1115	0,45	1740	13160	0,70
1984	1130	0,45	1725	16400	0,75
1985	1095	0,50	1725	17200	0,75
1986	1090	0,50	1720	17600	0,75
1987	1087	0,50	1705	18400	0,80
1988	1080	0,50	1710	19260	0,80

Год	Еженедельный пассажиропоток, тыс. чел.	Стоимость проезда, долл.	Население, тыс. чел.	Среднегодовой доход, долл.	Стоимость парковки, долл.
1989	1020	0,60	1700	21140	0,80
1990	1010	0,60	1695	22660	0,85
1991	1010	0,60	1695	23200	1,00
1992	1005	0,60	1690	23600	1,05
1993	995	0,60	1630	23660	1,05
1994	930	0,95	1640	25300	1,05
1995	915	0,95	1635	26000	1,10
1996	920	0,95	1630	26448	1,25
1997	940	0,95	1620	27532	1,30
1998	950	0,95	1615	28020	1,50
1999	910	1,20	1605	28936	1,55
2000	930	1,20	1590	30000	1,65
2001	933	1,20	1595	30400	1,75
2002	940	1,20	1590	31200	1,75
2003	948	1,20	1600	32000	1,90
2004	955	1,20	1610	32400	2,00

23. На предприятии работают 16 цехов, выпускающих различные виды продукции. В табл. 10.30 приведены показатели работы каждого цеха. Можно ли построить адекватную регрессионную модель для прогнозирования прибыли? Какие факторы оказывают наибольшее влияние на прибыль?

Таблица 10.30

Данные для решения задачи 23, д.е.

№ п/п	Производственные фонды	Средне-месячный объем работ	Средне-месячная прибыль	№ п/п	Производственные фонды	Средне-месячный объем работ	Средне-месячная прибыль
1	699	190	16	9	293	391	66
2	532	211	42	10	300	396	87
3	650	152	14	11	73	160	22
4	768	216	17	12	862	199	22
5	67	106	32	13	112	136	29
6	322	397	52	14	289	388	74
7	736	180	18	15	512	195	58
8	501	239	60	16	490	201	65

24. Найдите оценки параметров a и b функции Кобба–Дугласа:

$$Q = Q_0 \left(\frac{L}{L_0} \right)^a \left(\frac{K}{K_0} \right)^b$$

где Q – объем производства;
 L – трудовые ресурсы;
 K – капитал;
 Q_0 , L_0 и K_0 – фиксированные значения этих переменных: $Q_0 = 51$,
 $L_0 = 19$, $K_0 = 16$.

Значения Q при определенных трудовых ресурсах L и стоимости капитала K заданы в табл. 10.31 [8].

Таблица 10.31

Данные для решения задачи 24

$L \backslash K$	6	11	16	21	26
9	6	12	19	27	36
14	11	12	35	50	66
19	17	32	51	72	96
24	22	42	67	95	126
29	27	52	83	118	156

25. Проведите регрессионный анализ данных, представленных в табл. 10.32 [8], используя методы Backward и Forward. Сравните и проанализируйте результаты. Рассчитайте минимальное число факторов x_i , достаточно точно определяющих значение зависимой переменной y .

Таблица 10.32

Данные для решения задачи 25

x_1	x_2	x_3	x_4	y
7	26	6	60	78,5
1	29	15	52	74,3
11	56	8	20	104,3
11	31	8	47	87,6
7	52	6	33	95,9
11	55	9	22	109,2
3	71	17	6	102,7

x_1	x_2	x_3	x_4	y
1	31	22	44	72,5
2	54	18	22	93,1
21	47	4	26	115,9
1	40	23	34	83,8
11	66	9	12	113,3
10	68	8	12	109,4

26. Проведите регрессионный анализ по данным табл. 10.8 и сравните результаты с полученными в подразд. 10.4.

27. По данным табл. П.11 определите факторы, в наибольшей степени влияющие на цену дома. Наблюдается ли эффект мультиколлинеарности? Постройте регрессионную модель влияния этих факторов на цену. Оцените ее обоснованность. Объясните смысл показателей адекватности и статистической значимости модели, а также параметров модели и их доверительных интервалов.

28. Почтовая компания использует классификацию груза по трем категориям: 1 – «прочный», 2 – «полупрочный», 3 – «хрупкий». Стоимость доставки груза зависит от его категории. Сформирована выборка (табл. 10.33) грузов разных категорий, но приблизительно одного веса, предназначенных для перевозки в пределах одной зоны (оплата за расстояние одинакова). Постройте регрессионную модель с использованием фиктивных переменных и оцените ее адекватность. Дайте интерпретацию параметров модели с учетом доверительных интервалов. Как категория товара влияет на оплату его пересылки?

Таблица 10.33

Данные для решения задачи 28

Тип груза	Цена доставки, руб.	Тип груза	Цена доставки, руб.	Тип груза	Цена доставки, руб.
2	140	3	344	3	218
1	40	2	230	3	240
3	276	2	130	1	150
3	222	2	170	1	68
2	200	1	42	1	26

29. По данным табл. П.12 с помощью регрессионного анализа определите, от каких факторов в наибольшей степени зависит размер заработной платы. Можно ли сделать вывод о различии заработной платы сотрудников разного пола? Если да, то какими причинами вызвано это различие?

Пояснение. При выполнении задания переменная «пол» должна интерпретироваться как фиктивная переменная.

30. В табл. П.13 приведены данные о цене однокомнатных квартир в разных районах города.

30а. Проведите множественный регрессионный анализ этих данных, взяв в качестве зависимой переменной цену квартиры. При построении регрессионных моделей вначале используйте метод Enter. Определите влияние различных факторов. Постройте модели с учетом и без учета «фиктивных переменных» («дом», «этаж»). Проверьте значимость и адекватность регрессионной модели. Обозначая район города фиктивной переменной, определите, как средняя цена квартиры зависит от района города.

30б. Повторите регрессионный анализ, используя методы Backward и Forward. Сравните и проанализируйте результаты. Определите минимальное число факторов, достаточно точно определяющих цену квартиры. Выясните, изменяются ли параметры регрессионной модели, если использовать не все данные, а следующие подмножества данных:

- 1) данные о квартирах, расположенных на первом или последнем этаже;
- 2) данные о квартирах, не расположенных на первом и последнем этажах;
- 3) данные о квартирах, расположенных в районе 3.

30в. Определите предполагаемую цену однокомнатной квартиры на первом этаже кирпичного дома в районе 1, если ее общая площадь равна 40 м^2 ; площадь кухни равна $9,5 \text{ м}^2$, а площадь комнаты составляет 21 м^2 , при условии, что до ближайшей остановки 5 мин ходьбы. Найдите 95%-ные доверительные интервалы предсказанного значения.

Пояснение. При выполнении задания переменная «район» должна интерпретироваться как представленная в номинативной, а не в порядковой шкале.

Моделирование и прогнозирование на основе анализа временных рядов

11.1. Основные характеристики и общие модели временных рядов

Регрессионный анализ позволяет выявить закономерности в динамике тех или иных процессов на уровне общей (средней) тенденции. Если в исследуемом процессе имеется дополнительная изменчивость, связанная с регулярными изменениями тех или иных признаков во времени (например, сезонное изменение спроса на прохладительные или слабоалкогольные напитки, мороженое и др.), то регрессионный анализ может эту изменчивость не отследить. Для более тщательного анализа и точного прогнозирования динамики процессов, характеристики которых определенным образом зависят от времени, применяют методы анализа временных рядов.

Временным рядом называется последовательность наблюдений, упорядоченная по времени: y_1, y_2, \dots, y_n , где y_t — измерения некоторой переменной в n равностоящих моментах времени $t = 1, 2, \dots, n$. Примерами временных рядов являются регулярно фиксируемые (каждый час, ежедневно, еженедельно и т.п.) данные об объеме продаж и цене на товар, количестве заказов в кафе и ресторанах, количестве звонков в сервисный центр, потреблении электроэнергии, биржевой активности, котировках валют, транспортных перевозках. Все эти процессы *меняются во времени и подвержены случайным колебаниям*.

График типичного процесса, заданного временным рядом, приведен на рис. 11.1 [68]. На графике виден возрастающий тренд, в то же время характер процесса повторяется с определенной периодичностью.

Основные задачи анализа временных рядов:

- 1) анализ характерных особенностей ряда для понимания закономерностей процесса, отраженного рядом;
- 2) прогнозирование характеристик ряда в будущие периоды времени.

Прогнозирование на основе анализа временных рядов связано даже с большими сложностями, чем прогнозирование на ос-

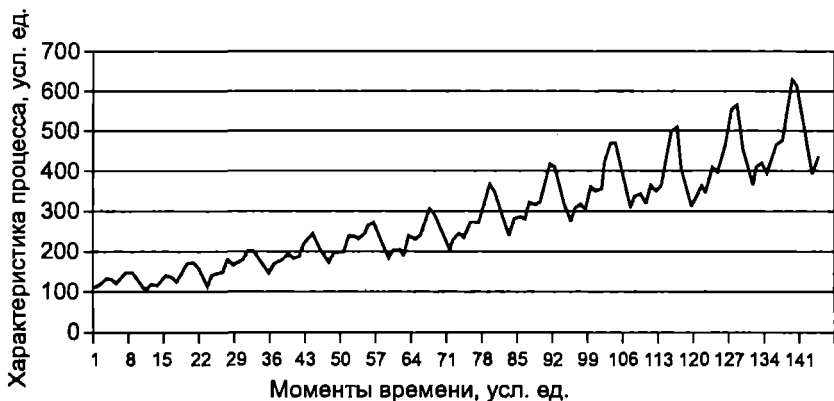


Рис. 11.1. Пример процесса, заданного временным рядом

нове регрессионного анализа. Во-первых, прогнозирование характеристик ряда — это всегда экстраполирование, поскольку исследователя интересует динамика процесса в будущем. Во-вторых, не всегда просто выявить «скрытые» закономерности и регулярности в истории развития процесса, на основе которых можно было бы строить прогноз. В частности, эти закономерности (если они действительно есть) необходимо отделять от «шума» — случайных изменений изучаемого признака. В-третьих, то или иное развитие процесса в прошлом отнюдь не гарантирует аналогичное поведение в будущем, которое может значительно измениться из-за влияния некоторых факторов, не отраженных в истории процесса. Например, динамика роста цен на нефть на мировом рынке может резко измениться из-за решения стран ОПЕК изменить квоты на добычу нефти или появления признаков смены политического курса одной из ведущих «нефтяных» стран. Но если у исследователя нет дополнительной информации о возможном поведении изучаемого процесса в будущем, принимается предположение о продолжении действия закономерностей, выявленных в прошлые периоды (*принцип «самоповторения истории»*).

Помимо методов регрессионного анализа в теории временных рядов разработаны и используются специальные подходы и методы анализа и экстраполирования: спектральный анализ, методы сглаживания и фильтрации, модели авторегрессии и др. [4; 5; 8; 12]. Отметим сразу, что как и в случае регрессионного

анализа, применение более сложной модели или метода прогнозирования не всегда обеспечивает лучшую точность прогноза.

В отличие от регрессионных моделей, в которых изменение исследуемого признака анализируется и прогнозируется через изменение других (факторных) признаков, анализ временных рядов предполагает изучение и прогнозирование признака на основе изменения *этого же* признака в предыдущие периоды.

Прогнозирование на основе анализа временных рядов предполагает построение модели, адекватной данным прошлых периодов. Что для этого случая может означать понятие адекватности модели?

Пусть имеется временной ряд y_1, y_2, \dots, y_n . Предположим, мы построили некоторую модель, позволяющую предсказывать значение y_t по предыдущим значениям ряда. Как и в случае регрессионного анализа, здесь можно использовать понятие *ошибки прогноза* e_t , определяемой как разность предсказанного и действительного значений признака в момент времени t . Далее можно предположить, что если модель обеспечивает небольшие суммарные ошибки в прошлом (для моментов времени $t = 1, 2, \dots, n$), то ошибки останутся небольшими и для будущих моментов времени ($t > n$).

В качестве реальных показателей адекватности, разработанных на основе сделанных предположений, используются следующие величины: *MAE* (*mean absolute error* – средняя абсолютная ошибка), *MSE* (*mean square error* – среднеквадратическая ошибка), *MAPE* (*mean absolute percentage error* – средняя абсолютная процентная ошибка). Эти показатели определяются по формулам

$$MAE = \frac{\sum_{t=1}^n |e_t|}{n};$$

$$MSE = \frac{\sum_{t=1}^n e_t^2}{n};$$

$$MAPE = 100\% \cdot \frac{\sum_{t=1}^n |e_t / y_t|}{n}$$

Показатель MSE эквивалентен *остаточной дисперсии*, используемой для оценки адекватности регрессионных моделей (см. подразд. 10.2). Соответственно, корень квадратный из MSE эквивалентен *стандартной ошибке оценки*. Показатель MAPE, характеризующий суммарные *относительные* ошибки, позволяет учесть тот факт, что при прогнозировании периодически изменяющихся показателей их абсолютные величины могут меняться достаточно сильно и приводить к значительным изменениям ошибок прогноза, поэтому абсолютные значения в этом показателе заменяются относительными.

Очевидно, что при подборе соответствующей модели временного ряда нужно ориентироваться на минимальные значения данных показателей. Следует помнить, что минимальные значения этих показателей свидетельствуют об адекватности модели для известных моментов времени, но *не гарантируют* точность прогноза для будущих периодов.

В анализе временных рядов, как и в большинстве статистических методов, предполагается, что исходные данные содержат детерминированную и случайную составляющие. В общем случае детерминированная составляющая может быть представлена в виде комбинации следующих компонентов:

1) *тренд* – общая систематическая (линейная или нелинейная) тенденция временного ряда, которая может изменяться во времени;

2) *сезонная составляющая* – периодически повторяющиеся колебания значений временного ряда¹.

Эти компоненты могут присутствовать одновременно. Например, продажи компании могут возрастать из года в год, но они также содержат сезонную составляющую (продажи в определенные месяцы года заметно превышают среднегодовые).

Таким образом, временной ряд в общем случае представляется в виде

¹ В более сложных моделях временных рядов может присутствовать и так называемая *циклическая составляющая* (см. [4; 5; 8; 12]). Циклическая компонента отличается от сезонной тем, что она обычно имеет большую временную протяженность и проявляется через неравные промежутки времени. Так, некоторый товар может быть особенно популярным в течение определенного времени после проведения интенсивной рекламы или показа телефильма, где герои пользуются данным товаром, и т.д.

$$y_t = f(u_t, S_t, \varepsilon_t),$$

где y_t – значение изучаемого признака в момент t ;

u_t – тренд;

S_t, ε_t – соответственно сезонная и случайная составляющие ряда.

Временной ряд может быть представлен различными математическими моделями, наиболее распространенными из которых являются *аддитивная модель*

$$y_t = u_t + S_t + \varepsilon_t$$

и мультипликативная модель

$$y_t = u_t S_t \varepsilon_t,$$

которая может быть сведена к аддитивной модели путем логарифмирования.

Представление временного ряда в виде той или иной композиции компонентов естественно приводит к идее их последовательного выделения и прогнозирования будущих значений ряда на основе полученной модели.

Фактически аддитивные модели предполагают прогнозирование данных путем рекурсивного прибавления (или вычитания) определенных значений к известным значениям, а мультипликативные модели – путем умножения известных значений на определенные коэффициенты.

Пример 11.1. Допустим, по имеющимся данным определено, что среднемесячное увеличение спроса составляет 20 единиц. Тогда в следующем месяце прогнозное значение спроса должно составить предыдущее значение плюс 20 единиц. Если мы определили, что, предположим, среднемесячное увеличение спроса составляет 5%, то в следующем месяце прогнозное значение составит предыдущее значение, умноженное на 1,05.

Сезонная закономерность, повторяющаяся в ряду с определенной периодичностью, также может быть аддитивной или мультипликативной. Если, например, каждый год объем продаж некоторого товара может увеличиваться в декабре на 1 млн руб., можно учесть эти сезонные изменения, *прибавляя* к прогнозу на декабрь этот 1 млн (*аддитивная сезонность*). Если каждый год в декабре объем продаж данного товара увеличивается примерно

на 40%, то прогноз нужно *умножить* на 1,4, т.е. если средний объем продаж этого товара невелик, то абсолютное (в денежном выражении) увеличение этого объема в декабре также будет относительно небольшим (но в процентном исчислении оно будет постоянным); если товар продается хорошо, то и абсолютный рост объема продаж будет значительным (*мультипликативная сезонность*). На графиках временных рядов различие между этими двумя видами сезонности будет проявляться следующим образом: при аддитивной сезонности ряд будет иметь постоянные сезонные колебания, величина которых не зависит от общего уровня значений ряда; при мультипликативной сезонности величина сезонных колебаний будет меняться в зависимости от общего уровня значений ряда (см. рис. 11.1).

Приведенные примеры – это элементарная иллюстрация общего различия аддитивных и мультипликативных моделей. Модели, применяемые в реальной практике анализа и прогнозирования временных рядов, значительно сложнее и могут включать как аддитивные, так и мультипликативные элементы.

11.2. Простейшие модели временных рядов и автокорреляционный анализ

Простейшая модель предполагает, что значения изучаемого показателя изменяются случайно относительно постоянного среднего значения (нет восходящего или нисходящего тренда) с постоянной дисперсией и не зависят друг от друга. Примером данных, адекватно описываемых такой моделью, будут результаты работы генератора случайных чисел.

Если при этом значения временного ряда можно считать нормально распределенными (само по себе это предположение не входит в модель), то для прогнозирования данных с помощью этой модели достаточно определить на основе имеющихся наблюдений среднее значение \bar{y} и дисперсию s^2 , после чего построить доверительный интервал с заданной вероятностью. Например, с вероятностью 95% любое последующее наблюдение будет лежать в интервале приблизительно $\bar{y} \pm 2s$

Проверить, соответствует ли модель данным предположениям, можно, построив соответствующий график. Например, на рис. 11.2 представлены ежемесячные объемы продаж за период

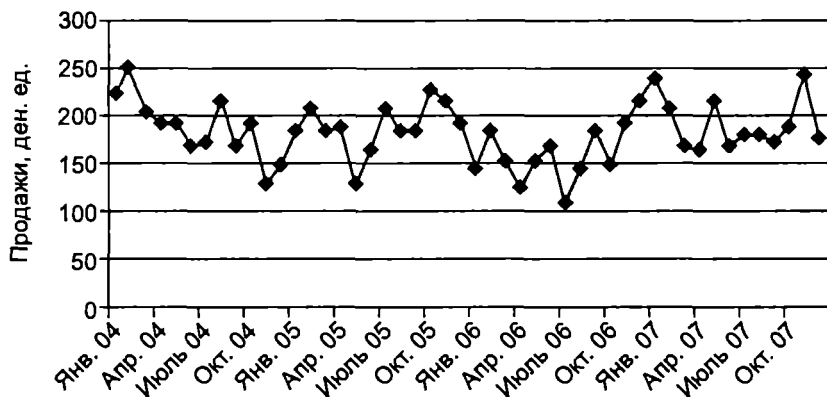


Рис. 11.2. Ежемесячные объемы продаж

с января 2004 г. по декабрь 2007 г. ($n = 48$). В этих данных не наблюдается тренд, какие-либо регулярности или закономерности отсутствуют (если бы какие-либо закономерности существовали, то они определялись бы визуально, через каждые k временных интервалов).

Более точно проверить предположение о связанности данных и наличии периодических составляющих можно с помощью *коррелограммы* – набора графической и численной информации, отражающей результат вычисления *автокорреляции*. Понятие автокорреляции поясним с помощью следующего рассуждения. Если во временном ряду имеется некоторая регулярность, т.е. повторяющееся с определенным периодом изменение, то между наблюдениями через этот период повторения должна наблюдаться корреляция, т.е. такие наблюдения должны коррелировать. Если, например, объем продаж подарков для мужчин возрастает в феврале, то должна обнаруживаться корреляция между февральскими объемами продаж каждый год.

Следовательно, если во временном ряду есть периодическая зависимость, повторяющаяся через k временных интервалов, то такая зависимость может быть формально определена как корреляционная зависимость порядка k между каждым i -м элементом ряда и $(i-k)$ -м элементом. Эту зависимость можно обнаружить и определить ее силу, вычисляя корреляцию между соответствующими элементами ряда, отстоящими друг от друга на k интервалов.

Корреляция между членами ряда называется *автокорреляцией*, а период k между коррелируемыми элементами ряда называется *лагом* (иногда используют термин «сдвиг»).

Коррелограмма показывает численно и графически автокорреляционную функцию (АКФ), т.е. коэффициенты автокорреляции и их стандартные ошибки для последовательности лагов из определенного диапазона (например, от 1 до 30). На коррелограмме обычно отмечается диапазон в размере двух стандартных ошибок на каждом лаге, что позволяет оценить и статистическую значимость коэффициентов автокорреляции. Интерес представляют сильные и значимые автокорреляции, поскольку они указывают на наличие периодических связей.

Построим коррелограмму для рассматриваемого примера. В пакете SPSS коррелограмма может быть построена выбором из меню *Graphs* опций *Time Series* и *Autocorrelations*. В диалоговом окне *Autocorrelations* выбирается переменная, соответствующая значениям временного ряда (в данном случае – переменная «продажи»), указывается параметр *Display Autocorrelations* (вывести значения коэффициентов автокорреляции), затем – *OK*. Результат представляется в отчете SPSS в двух вариантах: графическом и текстовом. На рис. 11.3 представлена коррелограмма с указанием 95%-ного доверительного интервала.

Текстовый вариант (рис. 11.4) также представляет значения коэффициентов автокорреляции для каждого лага, их ошибки и p -уровни значимости. Коэффициент автокорреляции для лага k показывает связь между значениями элементов ряда через k временных интервалов. Например, для лага 3 значение коэффициента составляет всего 0,081 при $p > 0,05$, что в данном случае свидетельствует об отсутствии какой-либо связи между продажами через три месяца.

Интерес представляют те лаги, для которых значение коэффициента автокорреляции выходит за пределы 95%-ного доверительного интервала, т.е. превышает величину, приблизительно равную двум значениям стандартной ошибки. Значение стандартной ошибки приблизительно определяется как $1/\sqrt{n}$, т.е. для данного случая 0,14. Более точно определенные значения стандартных ошибок и, соответственно, доверительные интервалы приводятся на коррелограмме. В данном случае единственный лаг, для которого значение коэффициента автокорреляции

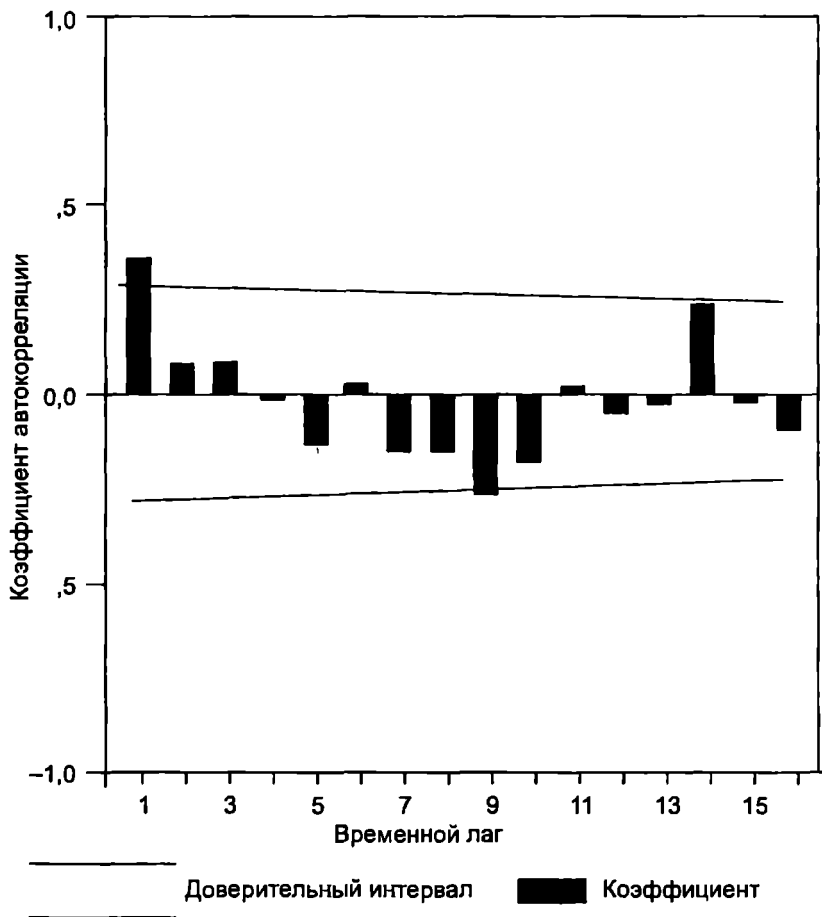
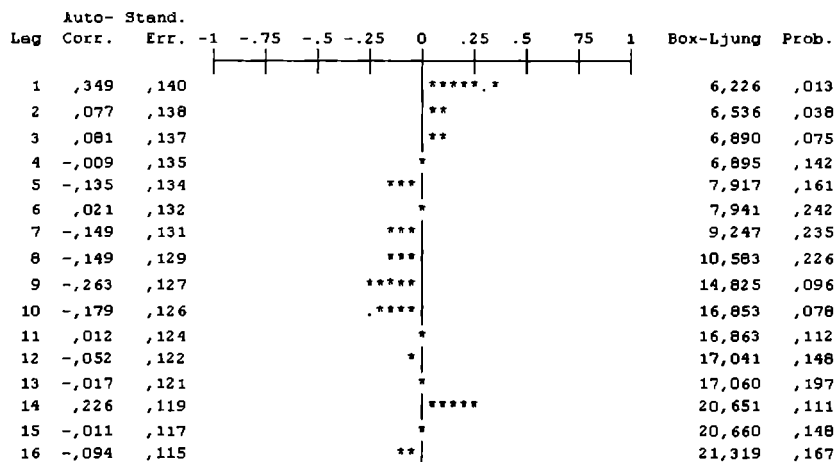


Рис. 11.3. Коррелограмма, построенная в пакете SPSS

выходит за пределы доверительного интервала, это лаг 1 (0,349). Это означает, что обнаруживается слабая тенденция увеличения объемов продаж на следующий месяц, если в предыдущем месяце продажи увеличивались, и уменьшения — если в предыдущем месяце продажи уменьшались. Для остальных лагов значения коэффициентов не превышают двух стандартных ошибок, поэтому любая кажущаяся тенденция по другим временным периодам — не более чем случайный «шум».



Plot Symbols: Autocorrelations * Two Standard Error Limits

Total cases: 48 Computable first lags:

Рис. 11.4. Текстовый вариант коррелограммы, подготовленной в пакете SPSS

Аналогичную коррелограмму можно построить в пакете STATISTICA. Для этого при запуске пакета нужно выбрать модуль *Time Series/Forecasting* (временные ряды/прогнозирование). В окне *Time Series Analysis* (анализ временных рядов) выбирается переменная, содержащая значения элементов ряда (данные должны быть заранее введены или скопированы), затем нужно щелкнуть по кнопке *OK* и в окне *Transformations of Variables* – по кнопке *Autocorrelations* (рядом с этой кнопкой в поле *Alpha* указывается уровень значимости для построения доверительного интервала (по умолчанию $\alpha = 0,05$). Построенная коррелограмма приведена на рис. 11.5.

Обычно коррелограмма строится для такого числа лагов, которое составляет не более 1/3 от общего числа наблюдений (рекомендуется даже 1/4). Иными словами, если в рассматриваемом примере имеется 48 наблюдений, то коррелограмму достаточно строить для 12 лагов и не более чем 16. Для указания максимального количества лагов в пакете SPSS выбирается кнопка *Options* в диалоговом окне *Autocorrelations* и затем в поле *Maximum Number of Lags* указывается нужное число. В пакете STATISTICA это число указывается в поле *Number of Lags*.

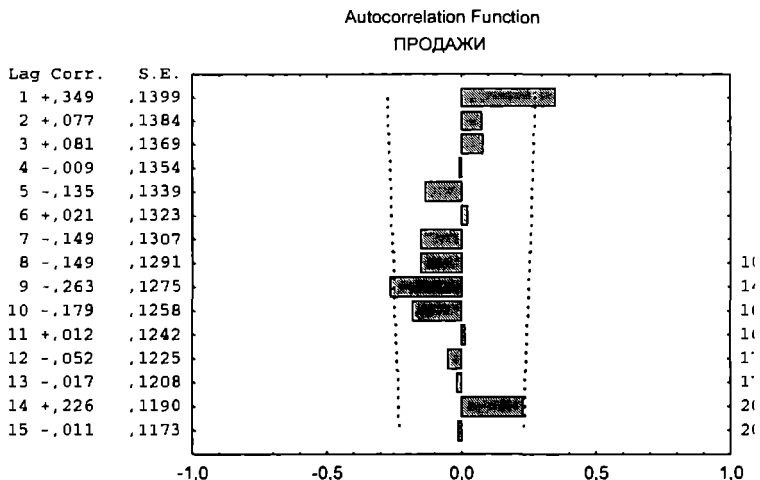


Рис. 11.5. Коррелограмма, подготовленная в программе STATISTICA

Несколько более сложный случай аддитивной модели временного ряда — модель случайных блужданий (*random walk model*):

$$y_t = y_{t-1} + \mu + \varepsilon_t.$$

Модель предполагает присутствие тренда, но отсутствие сезонной составляющей. В этой модели изменение значений элементов ряда не случайно, но *разности* значений описываются моделью случайных изменений:

$$dy_t = y_t - y_{t-1} = \mu + \varepsilon_t.$$

Прогнозирование на основе данной модели осуществляется посредством нахождения среднего значения ($\overline{dy_t}$) и стандартного отклонения для разностей dy_t по всем элементам ряда. Тогда прогнозное значение изучаемого признака для следующих периодов определяется как $y_{t+1} = y_t + \overline{dy_t}$.

Рассмотрим применение данной модели на примере динамики индекса Доу-Джонса (среднее ежемесячное значение) за период с января 1988 г. по март 1992 г. (табл. П.14). График изменения индекса (рис. 11.6) позволяет предположить наличие

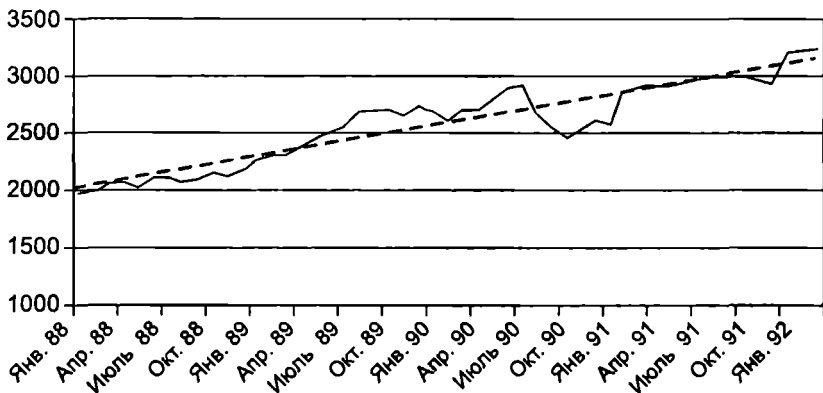


Рис. 11.6. Динамика индекса Доу-Джонса

тренда, близкого к линейному. Каких-либо заметных периодических колебаний не наблюдается.

Коррелограмма для этого ряда представлена на рис. 11.7. Она показывает, что все коэффициенты до лага 11 включительно положительны и их значения выходят за пределы доверительного интервала. Однако из этого еще не следует наличие периодических компонент в ряду. Дело в том, что автокорреляции последовательных лагов могут быть связаны между собой. Например, если первый элемент ряда тесно связан со вторым, а второй с третьим, то первый элемент должен также каким-то образом зависеть от третьего, и т.д. Именно это и получается в рассматриваемом примере.

Исключить «ложные» периодические составляющие ряда можно с помощью вычисления *частных автокорреляций* (сравните с показателем частной корреляции, рассмотренным в гл. 9). Частная автокорреляция на данном лаге аналогична обычной автокорреляции, за исключением того, что при вычислении из нее удаляется влияние автокорреляций с меньшими лагами. На лаге 1 частная автокорреляция равна «обычной» автокорреляции. Частная автокорреляция дает более отчетливое представление о периодических зависимостях временного ряда.

Построение коррелограммы для частных автокорреляций не отличается от рассмотренной выше процедуры вычисления автокорреляции за исключением того, что в диалоговом окне *Autocorrelations* программы SPSS указывается параметр *Display*

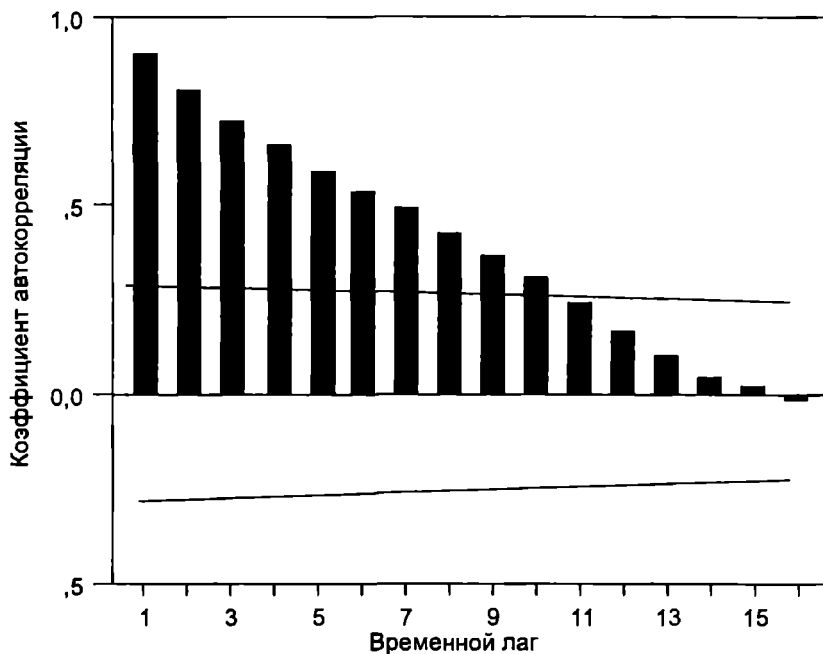


Рис. 11.7. Коррелограмма для ряда, описываемого моделью случайных блужданий

Partial autocorrelations (вывести значения коэффициентов частной автокорреляции), а в соответствующем окне программы STATISTICA – кнопка *Partial autocorr.*

Частные автокорреляции демонстрируют отсутствие каких-либо периодических составляющих временного ряда (рис. 11.8). Интерпретация коэффициента на лаге 1 аналогична ситуации, рассмотренной ранее (см. рис. 11.3).

Другой способ исключить формальную зависимость последовательных лагов между собой при автокорреляционном анализе временных рядов заключается в удалении автокорреляций первого порядка, т.е. взятии разностей с лагом 1 (фактически это означает, что строится новый ряд с числом элементов $n-1$ путем вычитания из каждого элемента ряда значения предыдущего элемента). Построить автокорреляционную функцию по разностям первого и других порядков можно, используя ранее рассмотренные инструменты SPSS, выбирая при этом в диало-

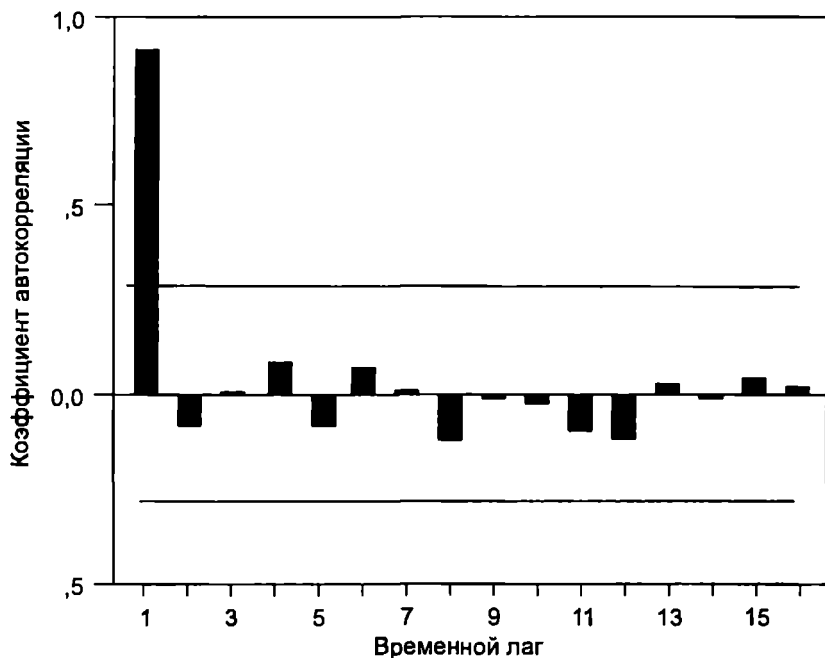


Рис. 11.8. Частные автокорреляции

говом окне *Autocorrelations* параметр *Difference* (разность) и указывая лаг (по умолчанию устанавливается 1). Коррелограмма, построенная по разности первого порядка для рассматриваемого примера (динамика индекса Доу-Джонса), представлена на рис. 11.9.

Взятие разности соответствующего порядка используется также для удаления периодических составляющих (например, при взятии разности k -го порядка из каждого i -го элемента ряда вычитается $(i-k)$ -й элемент). Таким образом можно определить скрытые периодические составляющие ряда. Поскольку автокорреляции на последовательных лагах зависимы, удаление некоторых автокорреляций может изменить другие автокорреляции, которые, возможно, подавляли их, и сделает некоторые сезонные составляющие более заметными. В частности, на рис. 11.9 видно, что автокорреляция на лаге 11 незначительно выходит за пределы доверительного интервала. Это могло бы означать, что в изменении индекса есть некоторая регулярность, проявляющаяся

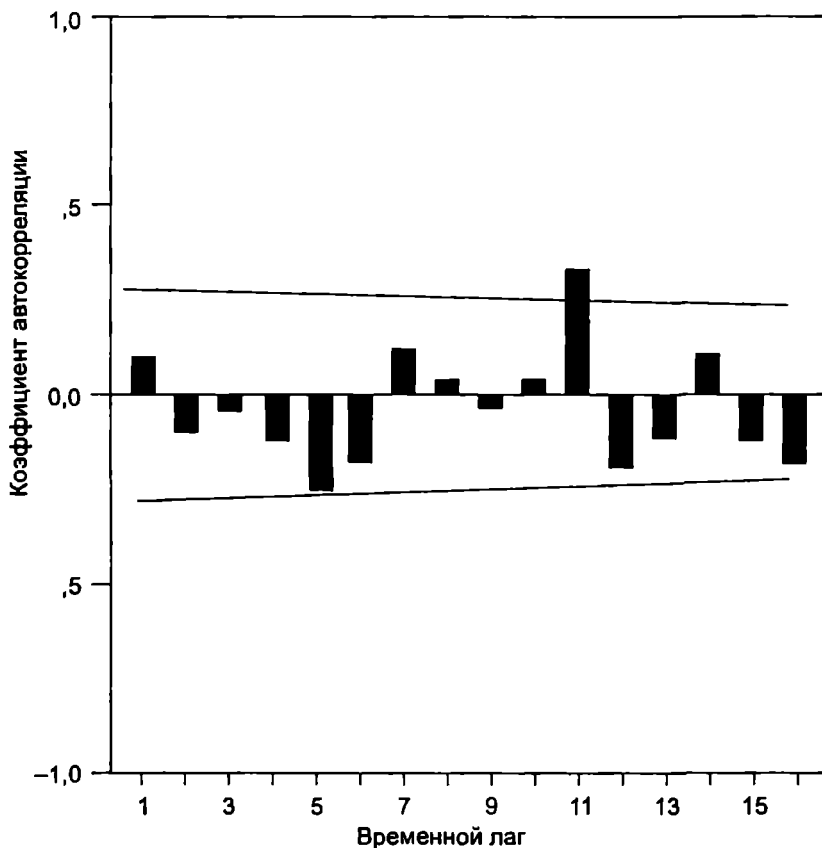


Рис. 11.9. Коррелограмма разностей первого порядка

каждые 11 месяцев. Но для подтверждения или опровержения данной гипотезы требуется более глубокий анализ, в том числе содержательный.

Удаление сезонных составляющих делает ряд *стационарным*, т.е. имеющим постоянные по времени среднее значение, дисперсию и автокорреляции. Выполнение условия стационарности требуется для применения некоторых методов анализа временных рядов.

Разности dy_t по элементам ряда ($t = 2, \dots, n$) могут быть вычислены в Excel, SPSS или другой программе. В пакете SPSS это может быть сделано выбором из меню *Transform* (преобразовать) – *Create Time Series* (создать временной ряд) – *Difference* (разность).

Для рассматриваемого примера среднее значение по остаточным разностям первого порядка равняется 26, а стандартное отклонение – около 85. Следовательно, прогнозируемое на основе данной модели значение индекса Доу-Джонса, например, на апрель 1992 г. должно составить 3247 (значение на март 1992 г.) плюс 26, т.е. 3273; а с учетом возможной ошибки, определяемой по величине стандартного отклонения: 3273 ± 170 с вероятностью 95%.

11.3. Методы сглаживания временных рядов

При анализе временных рядов довольно часто применяется *сглаживание на основе скользящего среднего (moving averages)*. Сглаживание временного ряда означает представление тренда в данной точке посредством среднего значения ряда, вычисленного в окрестности данной точки. Используемое для сглаживания количество точек в окрестности называют *базой (span)*. Как правило, в качестве окрестности точки принимаются значения ряда, *предшествующие* данной точке, т.е. сглаженный ряд меньше исходного ряда на величину базы. Если значения ряда из окрестности данной точки, входят с одним и тем же весом, то такая операция сглаживания ряда называется *методом простого скользящего среднего*.

При использовании этого метода большое значение имеет выбор базы. Если, например, имеются данные о ежемесячных объемах продаж и в качестве базы выбрано 12 точек (12 месяцев), то тренд может оказаться излишне сглаженным (слишком усредненным). Если в качестве базы выбраны три точки, то излишнее влияние могут оказывать случайные «выбросы». При выборе в качестве базы одной точки эффекта сглаживания нет совсем. В этом случае предполагается, что значение характеристики ряда в последующий период будет таким же, что и в текущем периоде, поэтому такой подход к прогнозированию, когда база равна 1, называется *наивным*. В качестве индикаторов выбора оптимальной базы следует использовать графический вид ряда, по которому можно первоначально судить о его динамике и «зашумленности», а также показатели MAE, MSE и MAPE.

Метод простого скользящего среднего легко реализуется в Excel. Для этого достаточно воспользоваться функцией СРЗНАЧ, указав в качестве аргументов массив предыдущих значений ряда

(например, если база равна трем точкам, то в качестве аргумента функции будет выступать массив из трех ячеек), а затем скопировать формулу в нижеследующие ячейки. После этого нужно определить разности значений исходного и сглаженного рядов в каждой точке и найти показатели MAE, MSE и MAPE. Это простейший метод сглаживания временного ряда, используемый на практике. Как правило, данный метод применяется однократно, но иногда (при очень «зашумленных» данных) процедура сглаживания используется несколько раз: сглаживается исходный ряд, затем — сглаженный ряд и т.д.

Недостатки этого метода очевидны. Во-первых, значения ряда за предыдущие периоды, используемые для прогнозирования, имеют одинаковый вес, но на последующее развитие процесса текущие его состояния могут оказывать более сильное влияние, чем относительно более ранние. Во-вторых, для применения этого метода, как правило, требуется довольно большой массив данных (сотни наблюдений).

Другим, более качественным методом сглаживания и прогнозирования, избавленным от указанных недостатков, является *метод экспоненциального сглаживания временного ряда*. Эта процедура использует веса, убывающие по геометрическому или экспоненциальному закону. Экспоненциальное сглаживание — это на сегодняшний день наиболее популярная серия методов прогнозирования временных рядов. Впервые серия этих методов была разработана и использована для решения военных, а затем и экономических задач в 1940–1950-х гг.

Существует несколько видов экспоненциального сглаживания, реализованных в статистических пакетах. Самая простая процедура называется *простым (simple) экспоненциальным сглаживанием*. Она предполагает, что ряд не содержит ни тренда, ни сезонной составляющей. Если наблюдается тренд, но нет сезонной составляющей, используется *метод Холта (Holt's method)*. Если ряд содержит и сезонную компоненту, применяется *метод Винтера (Winter's method)*. Разработаны и другие процедуры экспоненциального сглаживания, но названные три метода применяются наиболее часто, и для большинства практических случаев их вполне достаточно для построения качественного прогноза.

Модель простого экспоненциального сглаживания представляется в следующем виде:

$$l_t = \alpha y_t + (1 - \alpha)l_{t-1};$$

$$f_{t+k} = l_t,$$

где l_t — переменная, называемая *уровень (level) ряда*;

α — *постоянная сглаживания (smoothing constant)*, принимающая значения от 0 до 1;

y_t — значение ряда в момент времени t ;

f_{t+k} — прогнозное значение y_{t+k} .

В данной модели текущий уровень ряда представляется как взвешенное среднее текущего значения ряда y_t (взятого с весом α) и предыдущего уровня ряда l_{t-1} (взятого с весом $1 - \alpha$), а прогнозное значение на k периодов вперед определяется по последней оценке уровня ряда (прогнозное значение, следовательно, будет одним и тем же для любого $k > 1$).

Учитывая, что для момента времени t ошибка прогноза ϵ_t определяется как $\epsilon_t = y_t - f_t = y_t - l_{t-1}$, модель простого экспоненциального сглаживания может быть переписана как $l_t = l_{t-1} + \alpha \epsilon_t$. Из этого следует, что текущее значение уровня ряда определяется на основе предыдущего значения уровня ряда с учетом текущего значения ошибки прогноза. Иными словами, если предыдущий прогноз был завышенным ($\epsilon_t < 0$), текущее значение уровня ряда уменьшается, и наоборот, при заниженном прогнозе ($\epsilon_t > 0$) текущее значение уровня ряда увеличивается. Величина корректировки зависит от значения α : чем больше α , тем значительнее корректировка, т.е. тем чувствительнее прогноз к изменениям ряда.

Эффект влияния величины постоянной сглаживания проявляется более наглядно, если переписать исходную модель, рекурсивно подставляя в нее значения l_{t-1} , l_{t-2} и т.д. В результате получается следующее выражение (читателю рекомендуется проверить это самостоятельно):

$$l_t = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \alpha(1 - \alpha)^3 y_{t-3} +$$

Такой вид модели отчетливо демонстрирует, что текущий прогноз значения ряда зависит от *всех* предыдущих значений ряда, но берущихся с уменьшающимися весами. Если значение α близко к 0, а значение $1 - \alpha$, соответственно, близко к 1, то весовые коэффициенты перед предыдущими значениями ряда уменьшаются незначительно, т.е. предыдущие значения ряда

продолжают оказывать довольно сильное влияние на прогнозируемое значение; в результате сглаживание (усреднение) ряда получается довольно сильным.

Если значение α близко к 1, то весовые коэффициенты перед предыдущими значениями ряда уменьшаются очень быстро, т.е. лишь ближайшие значения ряда оказывают наиболее заметное влияние на прогнозируемое значение; в результате прогнозные значения весьма заметно реагируют на изменения значений ряда. В пределе, если α точно равно 1, то предыдущие наблюдения полностью игнорируются. Если α точно равно 0, то модель полностью нечувствительна к изменениям уровня ряда.

Какое значение следует выбрать при использовании этой модели? Некоторые специалисты рекомендуют значения $\alpha = 0,1$ или $\alpha = 0,2$ [54], или во всяком случае не больше 0,3 [68]. Другие исследования показывают, что значения $\alpha > 0,3$ для некоторых рядов обеспечивают лучший прогноз, чем меньшие значения α . Поэтому лучше оценить качество прогноза для нескольких различных значений α . Многие статистические пакеты включают в себя оптимизационные процедуры для подбора такого значения α , которое минимизирует значения MSE или MAPE. Простейшая из таких процедур (*поиск по сетке*) работает следующим образом: возможные значения параметра α разбиваются сеткой с определенным шагом. Например, рассматривается сетка значений от $\alpha = 0,1$ до $\alpha = 0,9$, с шагом 0,1. Затем выбирается α , для которого значение MSE или MAPE является минимальным. Существуют и более сложные процедуры подбора оптимального значения α . Однако, как уже неоднократно подчеркивалось, минимизация данных показателей может свидетельствовать о хорошем приближении к исходным данным, но не обязательно гарантирует наилучший прогноз будущих значений ряда.

Формула простого экспоненциального сглаживания показывает, что для вычисления первого уровня ряда (l_1) нужно знать значение l_0 , но оно не известно. Выбор начального значения (*initial value*) уровня ряда (l_0) называется *инициализацией*. В зависимости от выбора параметра α (особенно если α близко к 0), начальное значение сглаженного процесса может оказать существенное воздействие на прогноз для многих последующих наблюдений. Обычно начальное значение l_0 принимается равным значению y_1 . В статистических пакетах можно задавать и любые другие начальные значения уровня. Как и в других рекоменда-

циях по применению экспоненциального сглаживания, рекомендуется брать начальное значение, дающее наилучший прогноз. Статистические пакеты включают и специальные процедуры автоматического поиска «оптимального» начального значения. Влияние эффекта инициализации значительно уменьшается с длиной ряда и становится некритичным при большом числе наблюдений.

Покажем использование процедуры простого экспоненциального сглаживания в среде SPSS. Воспользуемся данными, характеризующими ежеквартальные объемы продаж некоторой гипотетической компании. В качестве первой переменной («кварталы») в SPSS вводятся номера кварталов (1–42), в качестве второй переменной («продажи») — объемы продаж за соответствующий квартал (тыс. руб.). В меню *Analyze* (анализ) программы SPSS выбирается опция *Time Series* (временные ряды), затем — *Exponential Smoothing* (экспоненциальное сглаживание). В появившемся диалоговом окне (рис. 11.10) переменная «продажи» переносится в поле *Variables* (переменные). В разделе *Model* (Модель) отмечается опция *Simple* (простое сглаживание).

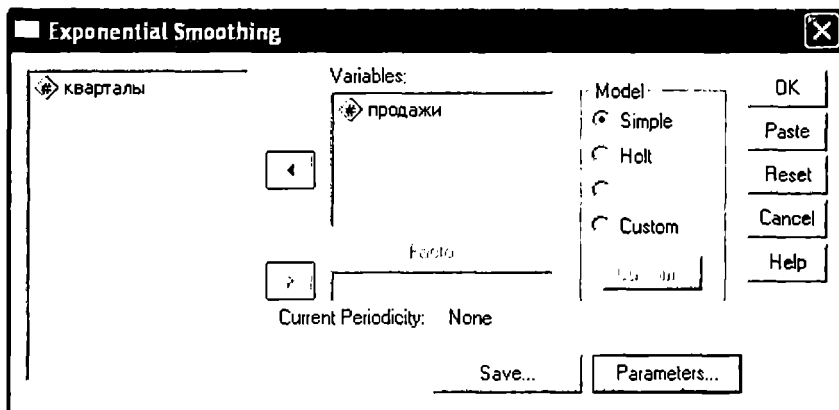


Рис. 11.10. Диалоговое окно *Exponential Smoothing*

Параметры сглаживания устанавливаются в диалоговом окне *Exponential Smoothing: Parameters* (вызывается щелчком по кнопке *Parameters*). В разделе *General (Alpha)* в поле *Value* по умолчанию указывается значение $\alpha = 0,1$. Если требуется подбор оптимального значения α , в этом разделе нужно выбрать параметр

Grid Search (поиск по сетке) и указать параметры метода поиска по сетке: *Start* (начальное значение α), *Stop* (конечное значение α) и *Step* (шаг). Для данного примера в поле *Value* было указано значение $\alpha = 0,2$. Установив параметры сглаживания, нужно щелкнуть по кнопке *Continue* (продолжить), чтобы вернуться в окно *Exponential Smoothing*. Если требуется построить прогноз на несколько периодов вперед, необходимо щелкнуть по кнопке *Save*, после чего в разделе *Predict Cases* выбрать параметр *Predict through* (прогнозировать до) и в поле *Observation* (наблюдения) указать последний период, для которого строится прогноз. Для данного примера (42 наблюдения) указано значение 46 (т.е. прогноз должен строиться на четыре квартала вперед). Если параметр *Predict through* не выбирается, то прогнозные значения ряда строятся для имеющегося диапазона наблюдений ($t = 1, \dots, n$).

После щелчка по кнопке *OK* в окне *Exponential Smoothing* в области данных SPSS появляются две новые переменные, значения которых соответствуют прогнозным значениям ряда и ошибкам прогноза. Результаты сглаживания ряда (исходные и прогнозные значения) представлены на рис. 11.11.

Визуальное сравнение показывает, что значения сглаженного ряда (т.е. прогнозные значения) не очень хорошо соответствуют исходному распределению данных, но если рассматривать «зигзаги» в исходном ряду лишь как «случайный шум», то, очевидно, и не следует стремиться к близкому повторению этого шума (т.е. стремиться к модели, слишком чувствительной к случайным колебаниям).

Аналогичные результаты можно получить в программе STATISTICA. Отметим, что данная программа предоставляет более широкие возможности для анализа временных рядов, чем SPSS. В частности, для подбора параметра α может быть использован не только метод поиска по сетке, но и другие процедуры, график исходного и сглаженного рядов строится автоматически, в выходной форме выводятся все параметры, характеризующие адекватность прогноза, и др.

На стартовой панели модуля *Time Series Analysis* (анализ временных рядов) программы STATISTICA нужно выбрать опцию *Exponential Smoothing & Forecasting* (экспоненциальное сглаживание и прогнозирование). В появившемся окне *Select the variables for time series analysis* (выберите переменные для анализа временных рядов) указывается переменная (переменные), соответствующую

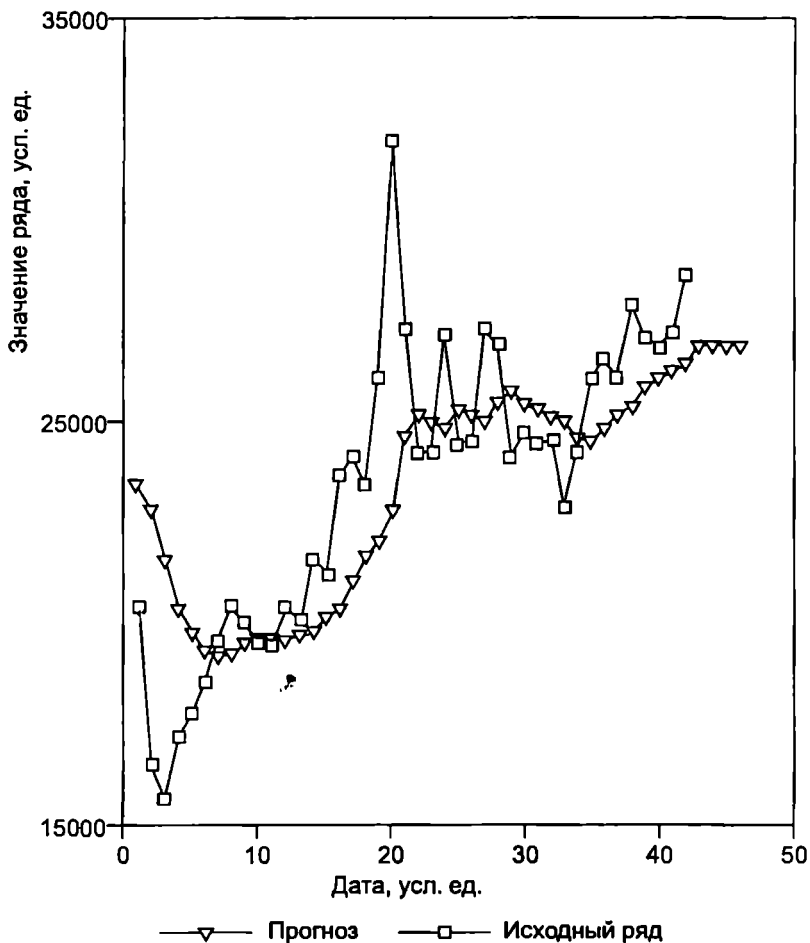


Рис. 11.11. Исходный и сглаженный ряды ($\alpha = 0,2$)

шая анализируемому временному ряду (для данного примера это переменная sales). После этого на экран выводится диалоговое окно *Seasonal and Non-Seasonal Exponential Smoothing* (сезонное и несезонное экспоненциальное сглаживание), в котором устанавливаются параметры экспоненциального сглаживания (рис. 11.12).

Простое экспоненциальное сглаживание без учета тренда и сезонной составляющей выполняется при выборе параметра *single*. Требуемое значение α устанавливается в поле *Alpha*. При необ-

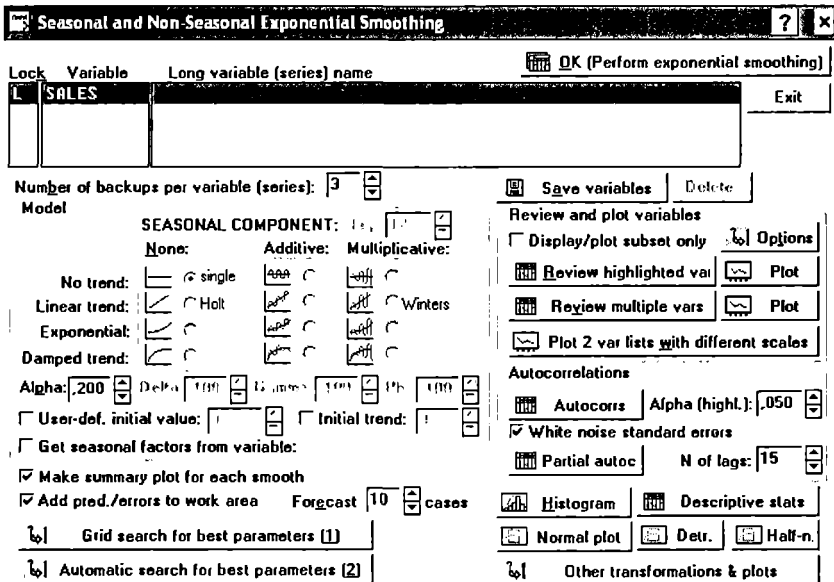


Рис. 11.12. Диалоговое окно *Seasonal and Non-Seasonal Exponential Smoothing*

ходимости определения оптимального значения α выбирается опция *Grid search for best parameters* (поиск по сетке) или *Automatic search for best parameters* (автоматический поиск наилучших параметров). В первом случае устанавливаются параметры сетки, как и в рассмотренном выше примере для SPSS. Во втором случае можно в качестве критерия оптимизации установить один из параметров адекватности модели: *Mean squared error* (средне-квадратическая ошибка – MSE), *Mean absolute error* (средняя абсолютная ошибка – MAE) или *Mean abs. pers. error* (средняя абсолютная процентная ошибка – MAPE). Запуск процедуры сглаживания осуществляется щелчком по кнопке *OK (Perform exponential smoothing)*.

Помимо прогнозных значений ряда и ошибок прогноза, а также графика сглаженного ряда, аналогичного представленному на рис. 11.11, выводятся показатели адекватности модели (рис. 11.13).

Для данной модели средняя абсолютная процентная ошибка (MAPE) составляет 8,2%. Попытаться улучшить этот результат можно, «запустив» процедуры поиска оптимального значения

TIME SERIES	No trend, no season	Alpha=,200
Summary of error		Error
Mean error	404,698035	
Mean absolute error	1861,918610	
Sums of squares	276953559,198027	
Mean square	6594132,361858	
Mean percentage error	,431473	
Mean abs. perc. error	8,237383	

Рис. 11.13. Показатели адекватности модели экспоненциального сглаживания

параметра α . Соответствующие процедуры программ SPSS и STATISTICA дают для α результат, близкий к 0,9. Однако в этом случае сглаженный ряд хотя и обеспечивает меньшее значение ошибки (MAPE = 5,9), вряд ли может считаться лучшей прогностической моделью, так как прогнозные значения получаются очень сильно зависящими от ближайшего предыдущего периода, т.е. модель слишком чувствительна к шуму (рис. 11.14). Этот пример показывает, что технически «оптимальное» значение α не всегда обеспечивает наилучшую прогностическую модель.

Если временной ряд содержит тренд, для сглаживания ряда должна использоваться модель Холта:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + m_{t-1});$$

$$m_t = \gamma(l_t - l_{t-1}) + (1 - \gamma)m_{t-1};$$

$$f_{t+k} = l_t + km_t.$$

По сравнению с моделью простого экспоненциального сглаживания здесь используются два новых параметра: m_t — значение тренда в момент времени t ; γ — постоянная сглаживания для тренда.

Все рассуждения, приведенные ранее для константы α , верны и для константы γ . Эта константа определяет, насколько модель чувствительна к изменениям тренда. При небольших значениях γ модель менее чувствительна (более медленно реагирует на изменения тренда), и наоборот. Очевидно, что при использовании данной модели вместо выбора (или подбора) значений одной константы в модели сглаживания требуется подбирать

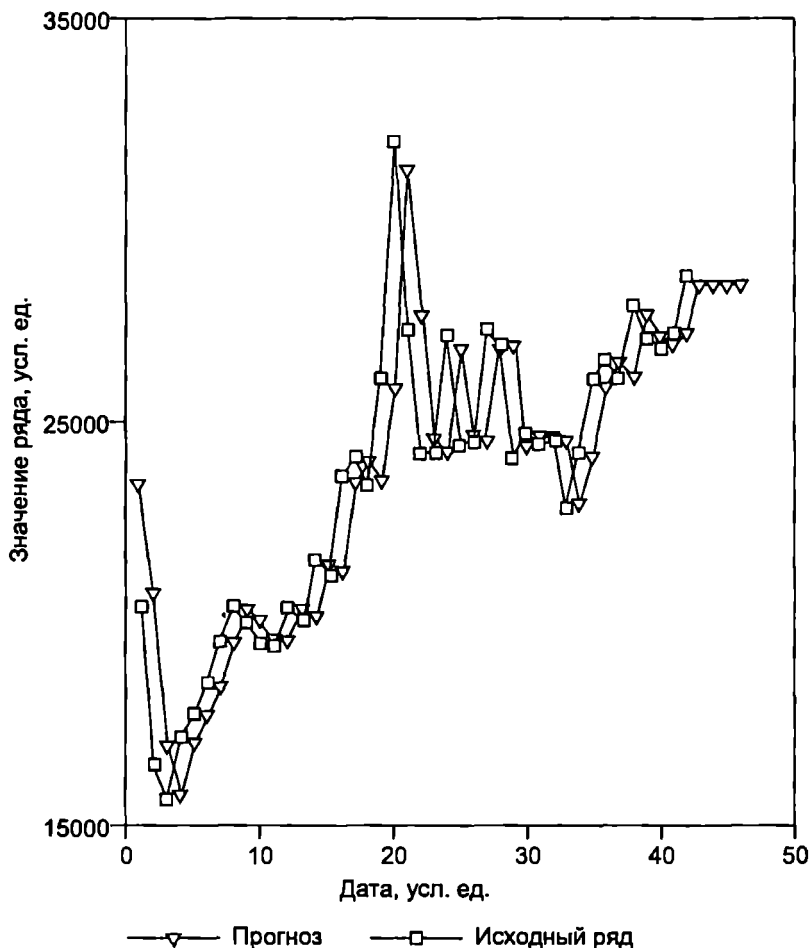


Рис. 11.14. Исходный и сглаженный ряды ($\alpha = 0,9$)

значения двух констант. Некоторые специалисты советуют устанавливать значения α и γ равными друг другу и небольшими (0,1 или 0,2) [54], а другие — использовать оптимизационные процедуры для подбора этих параметров.

Работа с моделью Холта в статистических пакетах аналогична работе с моделью простого экспоненциального сглаживания за исключением того, что в соответствующих разделах диалоговых окон выбирается режим *Holt*. Постоянная сглаживания тренда обозначается в этих пакетах как *Gamma*.

Для примера вновь обратимся к временному ряду, отражающему динамику индекса Доу-Джонса (см. рис. 11.6). Если сглаживать этот ряд на основе модели простого экспоненциального сглаживания (с $\alpha = 0,2$), получим результат, представленный на рис. 11.15 (МАРЕ = 5,3). На этом же рисунке для сравнения приведен результат применения модели Холта ($\alpha = \gamma = 0,2$) (МАРЕ = 3,9). Для данного примера модель Холта обеспечивает несколько лучшую точность прогноза и «улавливает» линейный тренд.

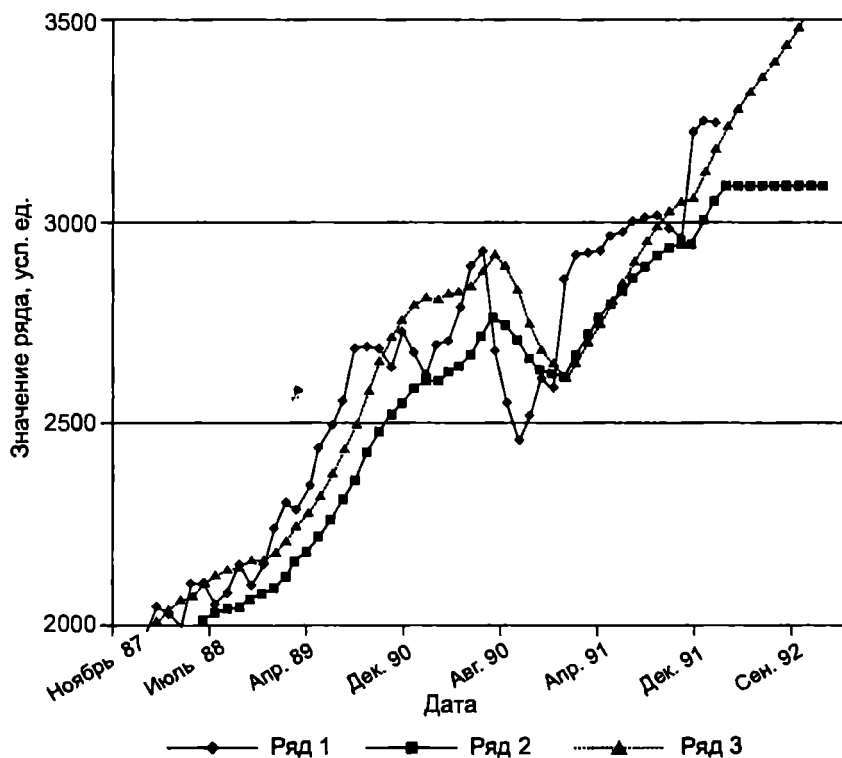


Рис. 11.15. Исходные значения ряда (ряд 1), модель простого экспоненциального сглаживания (ряд 2), модель Холта (ряд 3)

Процедура поиска оптимальных значений для параметров модели Холта предлагает $\alpha = 1, \gamma = 0$ (МАРЕ = 2,3). Исходный и сглаженный ряды приведены на рис. 11.16. Модель Холта с

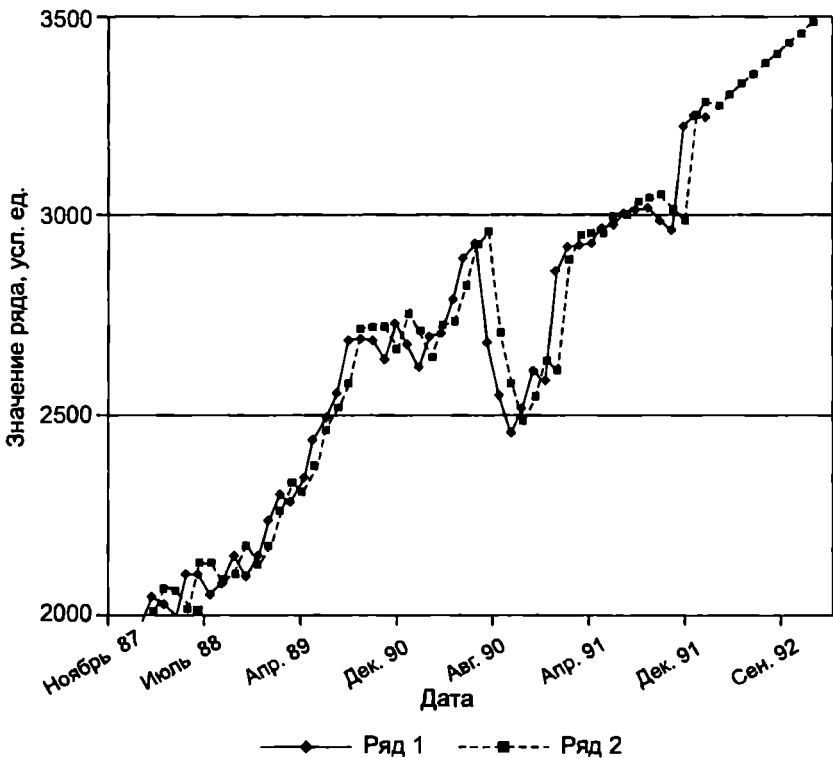


Рис. 11.16. Исходные (ряд 1) и прогнозные (ряд 2) значения ряда при оптимальных параметрах модели Холта

такими параметрами эквивалентна модели случайных блужданий. Другими словами, наилучшее предсказание для данного ряда обеспечивается по предыдущему значению ряда с добавлением постоянного тренда. Обращаем внимание читателя на то, что такой результат получен именно для данного ряда. Для другого набора данных оптимальные значения α и γ могут быть другими.

Теперь рассмотрим модель Винтера — модель экспоненциального сглаживания, учитывающую сезонность. Как обсуждалось в подразд. 11.1, сезонная составляющая временного ряда представляет собой некоторые регулярные паттерны, повторяющиеся с определенной периодичностью. К примеру, объемы продаж елочных игрушек могут резко возрастать в канун новогодних праздников с ежегодной повторяемостью.

Существуют два основных подхода к прогнозированию с учетом сезонной составляющей. Первый подход учитывает сезонность непосредственно (модель Винтера представляет данный подход). Второй подход предполагает первоначальное исключение сезонной составляющей из ряда, затем построение прогноза и последующую корректировку прогноза с учетом сезонной составляющей (этот подход рассматривается ниже).

Модель Винтера является усложнением модели Холта и представляется в следующем виде:

$$l_t = \alpha y_t + s_{t-g} + (1 - \alpha)(l_{t-1} + m_{t-1});$$

$$m_t = \gamma(l_t - l_{t-1}) + (1 - \gamma)m_{t-1};$$

$$s_t = \delta y_t + (1 - \delta)s_{t-g};$$

$$f_{t+k} = (l_t + km_t)s_{t+k-g}.$$

По сравнению с моделью Холта в модели Винтера используются несколько новых параметров: s_t — значение сезонной компоненты в момент времени t ; δ — постоянная сглаживания для сезонной компоненты; g — сезонный лаг, который соответствует временному периоду, через который наблюдается сезонное повторение значения ряда (например, $g = 4$ для данных, фиксируемых каждый месяц, с периодом повторения через четыре месяца).

Константа δ определяет чувствительность модели к изменениям сезонной составляющей. Чем меньше эта величина, тем менее чувствительной будет модель к сезонным паттернам. Рекомендуемые специалистами « типовые » значения параметров модели: $\alpha = \gamma = 0,2$; $\delta = 0,5$ [54]. Но, как и в случае ранее рассмотренных моделей сглаживания, можно использовать специальные статистические процедуры определения « оптимальных » параметров модели.

Работа с моделью Винтера в пакете STATISTICA осуществляется аналогично работе с моделью простого экспоненциального сглаживания и моделью Холта. В диалоговом окне *Seasonal and Non-Seasonal Exponential Smoothing* (см. рис. 11.12) выбирается режим *Winter*. Константа δ в пакете STATISTICA обозначена как (*Delta*). В поле *SEASONAL COMPONENT: lag* необходимо указать сезонный лаг.

Продемонстрируем работу этой процедуры на примере. В табл. 11.1 приведены значения временного ряда [8]. Если представить эти данные на графике (рис. 11.17), наблюдается очевидная сезонность с лагом 4.

Исходные данные для сглаживания ряда

Период	Значение	Период	Значение	Период	Значение
1	8,8	9	10,9	17	13,5
2	13,5	10	17,8	18	23,1
3	18,9	11	24,4	19	33,1
4	15	12	18,5	20	21,9
5	9,8	13	12,3	21	13,7
6	16	14	20,2	22	24
7	22,1	15	27,8	23	33,5
8	16,9	16	20,2	24	22,1

Exp smoothing Multipl season (4) S0=13,12 T0=4638
 Lin trend, mult season, Alpha=0,662 Delta=1,00 Gamma=0,00

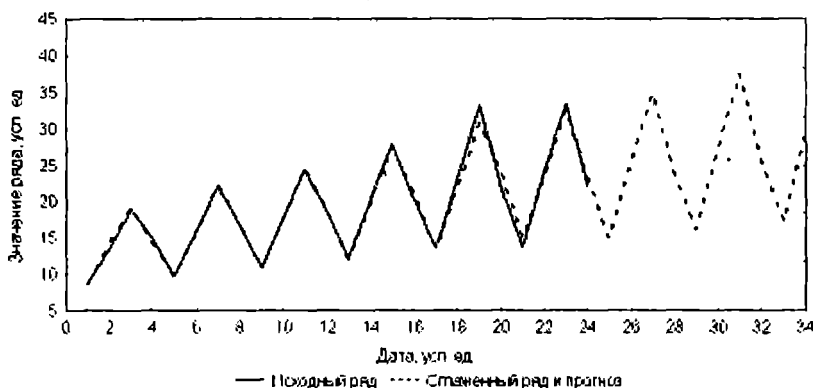


Рис. 11.17. Исходный и сглаженный ряды

Если временной ряд зашумлен, то сезонный лаг может быть не столь очевиден. В этом случае для его обнаружения может быть полезным построение коррелограммы (см. подразд. 11.2). Коррелограмма анализируемого ряда приведена на рис. 11.18. Она показывает, что ряд имеет сезонную составляющую с периодом, равным 4.

Ряд, сглаженный по методу Винтера, с прогнозными значениями на 10 следующих периодов также приведен на рис. 11.17. Значения оптимальных параметров модели, полученные в результате автоматического определения (критерий оптимальности — минимум MSE), следующие: $\alpha = 0,66$; $\delta = 1,0$; $\gamma = 0$. При

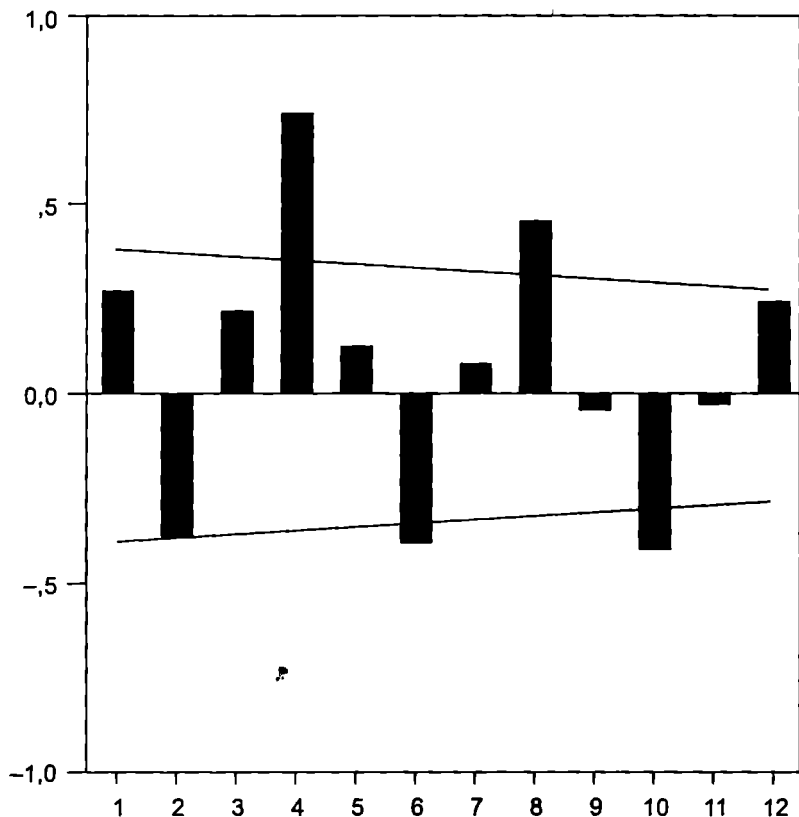


Рис. 11.18. Коррелограмма временного ряда с сезонной составляющей

таких параметрах $MSE = 0,84$, $MAE = 0,67$, $MAPE = 3,3$, что является индикатором хорошего приближения модели к эмпирическим данным. При таких параметрах модель достаточно чувствительна к изменениям уровня ряда и быстро реагирует на них ($\alpha = 0,66$), нечувствительна к изменениям тренда¹ ($\gamma = 0$) и максимально чувствительна к сезонным паттернам ($\delta = 1,0$).

Сезонные факторы (*seasonal factors*), также определяемые программой STATISTICA, имеют следующие значения: 63,48; 100,74; 136,43; 99,35. Это означает, что значение ряда в первом периоде составляет приблизительно 63,5% от среднего значения

¹ Это не означает, что тренд отсутствует.

за четыре периода (т.е. приблизительно на 36,5% меньше среднего за четыре периода, через которые наблюдается сезонность), значение ряда во втором периоде на 0,74% превышает среднее значение за четыре периода, и т.д.

Если в качестве критерия оптимальности выбрать минимум значения параметра MAPE, то можно получить даже несколько лучшие показатели адекватности модели: $MSE = 0,82$, $MAE = 0,54$, $MAPE = 2,6$. При этом параметры модели принимают следующие значения: $\alpha = 0$; $\delta = 0,38$; $\gamma = 0,09$, т.е. данная модель нечувствительна к изменению уровня ряда, но реагирует на изменение сезонной компоненты и отчасти на изменения тренда. В то же время полученный сглаженный ряд визуально практически не отличается от показанного на рис. 11.17, поэтому график ряда, сглаженного по методу Винтера с этими параметрами, мы не приводим.

Аналогичные результаты можно получить и в SPSS. Но для применения модели Винтера данные предварительно нужно определить как значения временного ряда. Для этого из меню *Data* (данные) нужно выбрать опцию *Define Dates* (определить временные периоды). В разделе *Cases Are* можно выбрать возможные временные составляющие ряда (например, изменение по годам (*Years*), изменение по годам и кварталам (*Years, quarters*), изменение по годам и месяцам (*Years, months*) и т.д.). Для рассматриваемого гипотетического ряда реальные временные составляющие неизвестны, но в ряду очевидно наблюдается периодичность с периодом 4. Поэтому необходимо определить этот период. Для данного случая можно выбрать, например, изменение по годам и кварталам (*Years, quarters*). Это будет означать, что повторяющийся паттерн наблюдается ежегодно с минимальным значением в первом квартале и максимальным значением в третьем квартале. После определения временной составляющей ряда в меню *Analyze* (анализ) выбирается опция *Time Series* (временные ряды), затем – *Exponential Smoothing* (экспоненциальное сглаживание), и в соответствующем окне – параметр *Winter*. Дальнейшая последовательность действий полностью аналогична работе с моделью Холта.

Как показывают результаты, модель Винтера очень хорошо отслеживает сезонные составляющие и демонстрирует возможности прогнозировать сезонные изменения в будущем. Для сравнения на рис. 11.19 приведены результаты применения модели

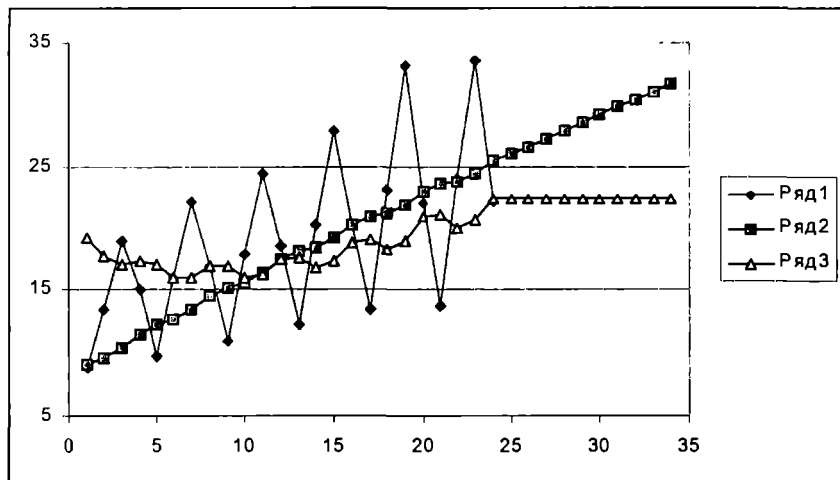


Рис. 11.19. Исходные значения ряда (ряд 1), модель Холта (ряд 2), модель простого экспоненциального сглаживания (ряд 3)

простого экспоненциального сглаживания и сглаживания по Холту. Очевидно, что эти две модели не могут считаться адекватными для анализа рассматриваемого временного ряда.

Помимо трех моделей экспоненциального сглаживания, которые рассмотрены в данном подразделе, в пакетах SPSS и STATISTICA реализованы и другие (более сложные) модели сглаживания временных рядов: модель экспоненциального тренда (*exponential trend*) и модель демпфированного (затухающего) тренда (*damped trend*), которые работают как без учета, так и с учетом сезонной составляющей (см. рис. 11.12). Эти модели хорошо описаны в литературе [5; 12], но обычно точность прогноза, полученного с их помощью, лишь незначительно улучшается по сравнению с применением рассмотренных в данном подразделе моделей [54]. Порядок работы с моделями экспоненциального и затухающего трендов принципиально не отличается от работы с рассмотренными моделями сглаживания.

В заключение этого подраздела рассмотрим еще один подход к анализу ряда с сезонной составляющей, называемый *сезонная декомпозиция* (*seasonal decomposition* или *deseasonalizing*). Данный метод используется как альтернативный методу сглаживания по Винтеру, и обычно (хотя и не всегда) эти два метода дают примерно одинаковые результаты.

Сезонная декомпозиция заключается в том, чтобы исключить влияние сезонности из каждого значения временного ряда. Затем ряд можно сгладить и построить прогноз без учета сезонной компоненты, после чего скорректировать прогноз, сделав поправку на сезонную составляющую.

Подход к исключению сезонной компоненты весьма прост, он основан на определении *индексов (факторов) сезонности*, которые показывают, во сколько раз значение ряда больше (или меньше) в тот или иной период по сравнению со средним за весь период сезонности. Например, если сезонность проявляется с периодом 12 месяцев (ежегодная сезонность) и мы обнаружили, что, например, в июне значение ряда на 30% превышает среднегодовое значение (среднее за 12 месяцев), т.е. индекс сезонности составляет 1,3, то для исключения сезонной компоненты (получения *десезонализованного* значения) июньское значение ряда нужно *разделить* на 1,3. И так для всех периодов для всех месяцев в данном примере): определяется индекс сезонности, с учетом которого корректируется соответствующее значение ряда. После сглаживания ряда, например по методу Колта, десезонализованное значение каждого периода нужно *умножить* на соответствующий индекс сезонности, чтобы вновь учесть сезонную компоненту.

Таким образом, процедура сезонной декомпозиции начинается с вычисления скользящего среднего (*ratio-to-moving-averages*) для временного ряда, при этом ширина окна берется равной периоду сезонности. Если период сезонности — четное число (12 мес., 4 квартала и т.п.), можно выбрать один из двух вариантов: брать скользящее среднее с одинаковыми весами или же с неравными весами так, что первое и последнее наблюдения в окне имеют усредненные веса (эта процедура называется *центрированием*).

Далее определяется отношение исходных данных к скользящим средним (в %), на основе чего вычисляется индекс сезонности как усреднение отношений, приходящихся на каждый сезон:

$$I_l = \left(\sum_{j=1}^k \frac{y_j}{\bar{y}_j} \right) / k \cdot 100,$$

где k — количество l -х сезонов в ряду;

y_j — значение ряда в момент времени j ;

\bar{y}_j — номер сезона;

\bar{y}_j — скользящее среднее на момент времени j .

Покажем, как работает процедура сезонной декомпозиции в пакете STATISTICA на рассмотренном ранее модельном примере временного ряда с сезонной составляющей. После запуска программы и ввода значений ряда нужно выбрать модуль *Time series/Forecasting*, ввести соответствующую переменную (кнопка *Variables*) и выбрать *Seasonal decomposition (Census I)* – сезонная декомпозиция. В появившемся диалоговом окне (рис. 11.20) задаются параметры модели сезонной декомпозиции.

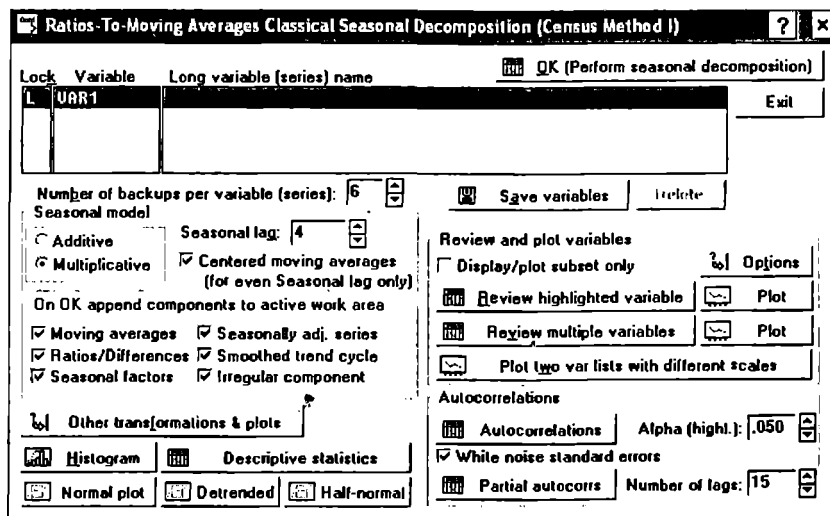


Рис. 11.20. Параметры модели сезонной декомпозиции (программа STATISTICA):

Additive – аддитивная модель;

Multiplicative – мультипликативная модель;

Seasonal lag – период сезонности (сезонный лаг);

Centered moving averages – использование процедуры центрирования (для четного сезонного лага);

Moving averages – вывод скользящих средних;

Ratios / Differences – вывод отношений (для мультипликативных моделей) или разностей (для аддитивных моделей);

Seasonal factors – вывод индексов сезонности;

Seasonal adj. series – вывод значений ряда без сезонной составляющей;

Smoothed trend cycle – вывод значений сглаженного тренда;

Irregular components – вывод значений остаточной (неучтенной / нерегулярной / случайной) компоненты.

Запуск процедуры сезонной декомпозиции (*Perform seasonal decomposition*) осуществляется щелчком по кнопке *OK*

В результате выполнения этой процедуры в рабочей области данных появляются новые переменные (рис. 11.21).

Case	VAR1	Moving Average	Ratios	Seasonal Factors	Adjusted Series	Smoothed Trend-c	Irreg. Compon.
1	8,80			63,1	13,95	13,47	1,035
2	13,50			101,4	13,31	13,72	,970
3	18,90	14,18	133,3	136,0	13,90	14,22	,978
4	15,00	14,61	102,7	99,5	15,07	14,80	1,019
5	9,80	15,33	63,9	63,1	15,53	15,38	1,010
6	16,00	15,96	100,2	101,4	15,78	15,88	,993
7	22,10	16,34	135,3	136,0	16,25	16,34	,994
8	16,90	16,70	101,2	99,5	16,98	16,81	1,010
9	10,90	17,21	63,3	63,1	17,27	17,23	1,002
10	17,80	17,70	100,6	101,4	17,55	17,63	,996
11	24,40	18,08	135,0	136,0	17,94	18,10	,991
12	18,50	18,55	99,7	99,5	18,59	18,68	,995
13	12,30	19,28	63,8	63,1	19,49	19,32	1,009
14	20,20	19,91	101,4	101,4	19,92	19,84	1,004
15	27,80	20,28	137,1	136,0	20,44	20,30	1,007
16	20,20	20,79	97,2	99,5	20,30	20,81	,976
17	13,50	21,81	61,9	63,1	21,40	21,68	,987
18	23,10	22,69	101,8	101,4	22,78	22,46	1,014
19	33,10	22,93	144,4	136,0	24,34	22,86	1,065
20	21,90	23,06	95,0	99,5	22,01	22,73	,968
21	13,70	23,23	59,0	63,1	21,71	22,83	,951
22	24,00	23,30	103,0	101,4	23,66	23,10	1,024
23	33,50			136,0	24,64	23,50	1,048
24	22,10			99,5	22,21	23,70	,937

Рис. 11.21. Результаты применения модели сезонной декомпозиции:

Moving Average – центрированные скользящие средние по периоду сезонности (окрестность из четырех ближайших точек ряда);

Ratios – отношения значений ряда к скользящему среднему (в %);

Seasonal Factors – индексы сезонности (в %);

Adjusted Series — значения ряда с исключенной сезонной компонентой (вычисляется делением значений исходного ряда на индексы сезонности и умножением на 100);

Smoothed Trend-c. — значения ряда с исключенной сезонной компонентой после сглаживания (после применения модели экспоненциального сглаживания);

Irreg. Compon. — остаточная компонента (вычисляется делением значений переменной *Adjusted Series* на значения переменной *Smoothed Trend-c.*).

На рис. 11.22 приведены графики исходного и сглаженного рядов с исключенной сезонной компонентой.

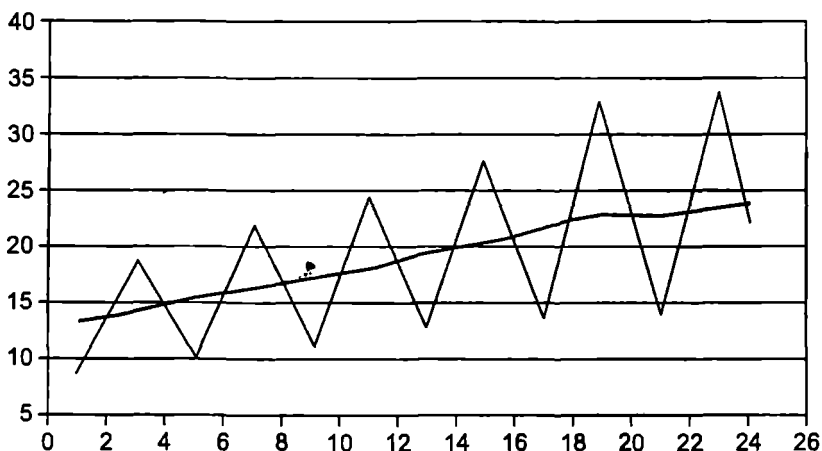


Рис. 11.22. Исходный и десезонализованный ряды

Точно такие же результаты могут быть получены в пакете SPSS, но исходные данные должны быть предварительно определены как значения временного ряда, как было показано при рассмотрении модели Винтера (меню *Data*, опция *Define Dates*). Для выбора параметров модели сезонной декомпозиции в SPSS из меню *Analyze* (анализ) нужно выбрать опцию *Time Series* (временные ряды), затем — опцию *Seasonal Decomposition*. В появившемся диалоговом окне (рис. 11.23) задаются параметры модели сезонной декомпозиции.

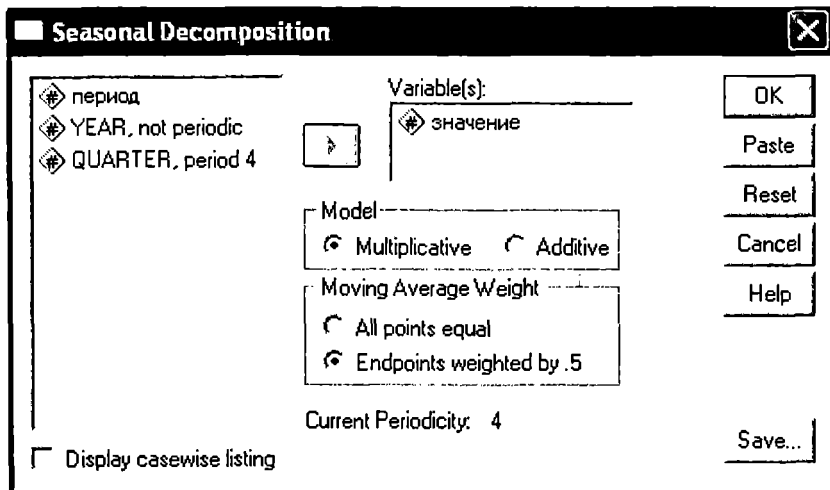


Рис. 11.23. Параметры модели сезонной декомпозиции (среда SPSS):

Additive – аддитивная модель;

Multiplicative – мультипликативная модель;

All points equal – расчет скользящего среднего с одинаковыми весами;

Endpoints weighted by .5 – использование процедуры центрирования при расчете скользящего среднего.

После учета сезонной компоненты (умножения значений сглаженного ряда на соответствующий индекс сезонности) и прогнозирования значений в последующих 10 временных периодах получаем результат, представленный на рис. 11.24. Как видим, данный результат визуально очень похож на тот, что был получен с помощью модели Винтера.

Процедура сезонной декомпозиции технически осуществляется намного проще, чем сглаживание ряда и прогнозирование на основе модели Винтера. Однако это преимущество не является существенным при использовании статистических пакетов.

Модель Винтера обеспечивает возможность более гибкого моделирования ряда с учетом показателей адекватности. Обращение к сезонной декомпозиции может быть полезно при необходимости детального учета изменений индексов сезонности.

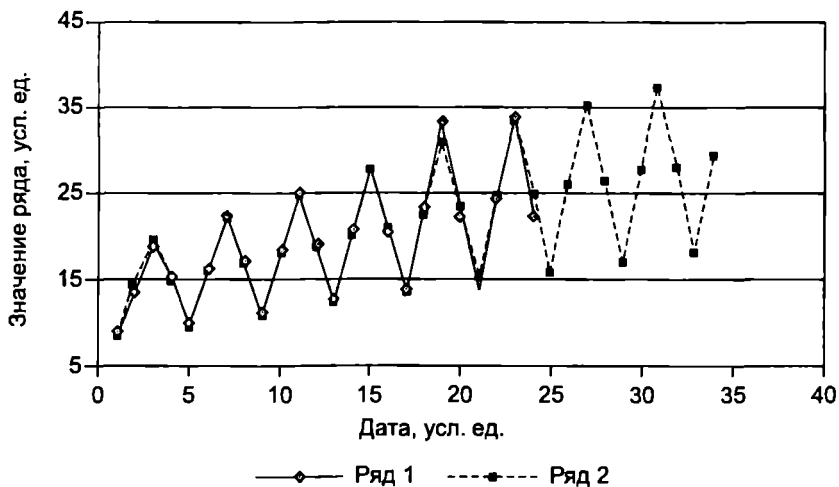


Рис. 11.24. Значения исходного (ряд 1) и восстановленного (ряд 2) рядов

11.4. Регрессионные и авторегрессионные модели временных рядов

Для выявления и оценки тренда при анализе временного ряда могут использоваться регрессионные модели. В таких моделях независимым фактором выступает время. Например, линейная регрессионная модель для анализа временного ряда имеет вид $y_t = \alpha + \beta t + \epsilon_t$. В данной модели параметр β показывает предполагаемые изменения характеристики ряда от одного периода времени к другому. Если этот параметр положительный, то тренд возрастающий, если отрицательный – тренд убывающий. Интерпретация параметра α в таких моделях менее важна или не важна совсем. Этот параметр соответствует величине характеристики ряда в момент времени $t = 0$. Если кодирование периода времени таково, что первое наблюдение соответствует $t = 1$, то параметр α показывает величину изучаемой характеристики ряда в момент, предшествующий первому наблюдению. Если анализируется ряд с другой кодировкой времени, то α может не иметь никакого реального практического смысла. Например, известны данные о ежегодных объемах продаж товара за период с 1997 по 2007 г. Если для анализа этого ряда использовать линейную

регрессионную модель, то параметр α показывает объем продаж на начало нашей эры, что, разумеется, представляется нелепым. Тем не менее в ряде случаев простые регрессионные модели могут быть полезны при анализе временных рядов с достаточно очевидным трендом и без сезонной компоненты.

Соответственно все подходы и статистические инструменты регрессионного анализа могут быть использованы и для анализа временного ряда. Однако далеко не всегда регрессионные модели обеспечивают лучший прогноз, чем, например, методы сглаживания или другие специальные методы анализа временных рядов, особенно при наличии сезонной компоненты.

Пример 11.2. На рис. 11.25 приведен график ежеквартальных доходов компании (с I квартала 1996 г. по II квартал 2006 г., всего 42 наблюдения).

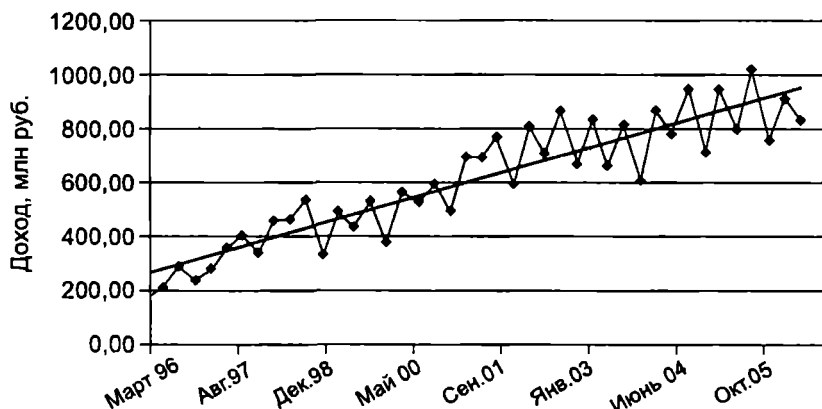


Рис. 11.25. Временной ряд с линейным трендом

Линейная регрессионная модель для данного ряда имеет вид

$$\text{Доход} = 244,82 + 16,53t \quad (\text{период времени кодируется как } t = 1, \dots, 42).$$

Оба коэффициента значимы на уровне $p < 0,001$; $R^2 = 0,83$; стандартная ошибка оценки — 90,38. Согласно этой модели ежеквартальный рост дохода составляет в среднем 16,53 млн руб. Но, несмотря на достаточно высокое значение R^2 , эта модель далека от совершенства. Во-первых, стандартная ошибка оценки (т.е.

фактически ошибка прогноза) весьма велика – 90,38 млн руб., т.е. свыше 11% от дохода за последний квартал. Во-вторых, «зигзаги», наблюдаемые в динамике ряда, вряд ли можно считать случайным «шумом». Скорее всего, здесь наблюдается некая сезонная составляющая, неотслеживаемая регрессионной моделью.

Помимо линейных моделей для анализа временных рядов могут применяться и нелинейные регрессионные модели. Пусть, например, ежеквартальные доходы компании за тот же период (I квартал 1996 г. – II квартал 2006 г.) изменяются так, как показано на рис. 11.26.

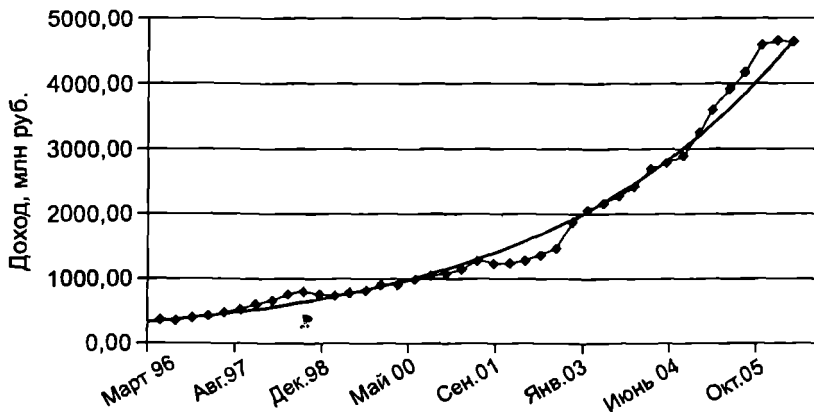


Рис. 11.26. Временной ряд с экспоненциальным трендом

Очевидно, что в данном случае наблюдаемый тренд явно нелинейный. Для его аппроксимации воспользуемся экспоненциальной регрессионной моделью

$$y_t = \alpha t^{\beta t} + \varepsilon_t.$$

В этой модели коэффициент β характеризует не абсолютное изменение дохода за один период времени (в данном случае – квартал), а относительное изменение. Если, например, $\beta = 0,05$, то это будет соответствовать среднему росту дохода на 5% каждый квартал, $\beta = -0,05$ будет означать ежеквартальное уменьшение дохода в среднем на 5%.

Для данного ряда экспоненциальная регрессионная модель имеет вид $y_t = 295,38e^{0,066t}$ (получено с помощью инструмента

Curve Estimation программы SPSS). Оба коэффициента значимы на уровне $p < 0,001$; $R^2 = 0,983$. Согласно этой модели ежеквартальный рост дохода составляет в среднем 6,6%, или около 29% в год. Эта модель хорошо описывает динамику доходов за прошлые периоды, но не отслеживает ситуацию «замораживания» доходов за последние три квартала, что может привести к ошибочному прогнозу доходов на ее основе.

Теперь приведем более сложный пример временного ряда. На рис. 11.27 представлен график еженедельных продаж некоторого товара. Рассмотрим на этом примере возможности метода авторегрессии для прогнозирования.

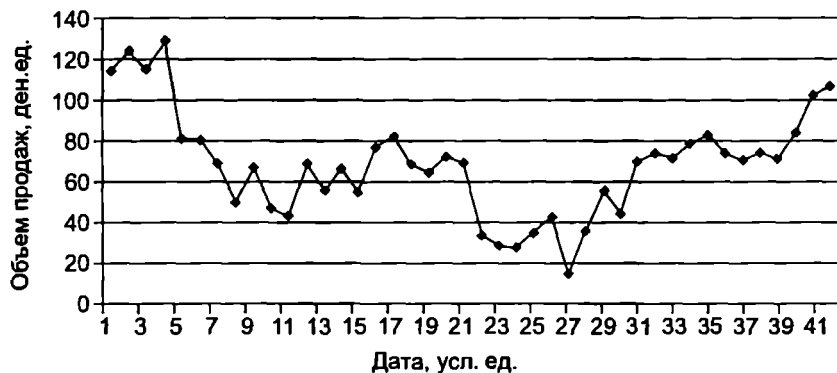


Рис. 11.27. Еженедельные продажи

Авторегрессионная модель – это линейное регрессионное уравнение, в котором в качестве результативного признака выступает значение временного ряда в момент времени t , а в качестве факторных признаков – значение ряда в моменты времени $t-1$ (авторегрессия первого порядка), $t-2$ (авторегрессия второго порядка), $t-3$ (авторегрессия третьего порядка) и т.д. Иными словами, *авторегрессией первого порядка* называется выражение

$$y_t = \xi + \alpha_1 y_{t-1} + \varepsilon_t,$$

авторегрессией второго порядка – выражение вида

$$y_t = \xi + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t,$$

и т.д. для авторегрессий последующих порядков.

Здесь α_1, α_2 – коэффициенты авторегрессии; ξ – константа (постоянный член авторегрессии); ε_t – случайная компонента.

Коррелограмма для рассматриваемого примера приведена на рис. 11.28. Из нее следует, что первые три лага заметно выходят за пределы доверительного интервала. На основе этого, пользуясь рассмотренным в гл. 10 инструментарием построения множественной линейной регрессионной модели, построим авторегрессию третьего порядка:

$$y_t = \xi + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \varepsilon_t.$$

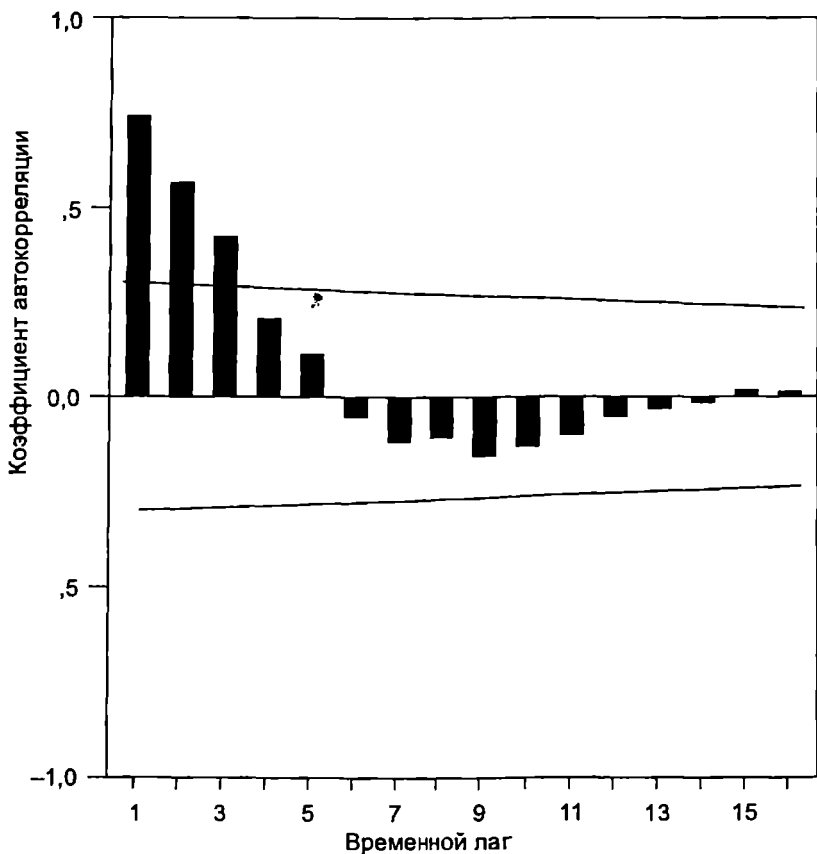


Рис. 11.28. Коррелограмма для временного ряда еженедельных продаж

Исходные значения временного ряда выступают в качестве зависимой переменной. Для создания независимых переменных воспользуемся опцией *Create Time Series* (создать временной ряд) из меню *Transform* (преобразовать) программы SPSS (хотя это можно сделать и в Excel). В диалоговом окне *Create Time Series* исходная переменная (sales) переносится в поле *New Variables* (рис. 11.29).

	sales	sales_1	sales_2	sales_3	var	var	var
1	115,00						
2	125,00	115,00					
3	116,00	125,00	115,00				
4	130,00	116,00	125,00	115,00			
5	82,00	130,00	116,00	125,00			

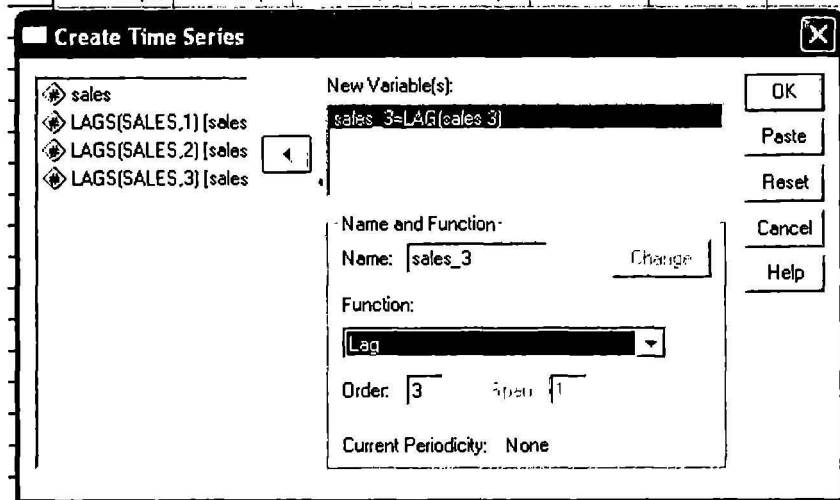


Рис. 11.29. Создание новых переменных для авторегрессии

В поле *Name* (раздел *Name and Function*) задается новое имя переменной, например sales_1. В поле *Function* этого же раздела выбирается функция *Lag* (лаг), в поле *Order* (порядок) задается значение лага (1). После этого нужно щелкнуть по кнопке *Change*, а затем – по *OK*. В рабочей области данных появляются значения новой переменной со значениями исходного ряда, но «сдвинутыми» вниз на один период. Аналогично создаются еще две переменные (sales_2 и sales_3) со значениями, смещенными соответственно на два и три периода (см. рис. 11.29). Эти пере-

менные будут выступать в качестве независимых переменных регрессионной модели.

Построенная регрессионная модель (меню *Analyze – Regression – Linear*) обеспечивает значение $R^2 = 0,574$ и стандартную ошибку оценки 15,72, но при этом статистически значимым (на уровне $\alpha = 0,05$) является лишь первый коэффициент регрессионной модели α_1 ($p = 0,01$), а также (на уровне $\alpha = 0,1$) константа ξ ($p = 0,057$). Коэффициенты α_2 и α_3 не являются статистически значимыми ($p = 0,448$ и $0,830$ соответственно), поэтому переменные y_{t-2} и y_{t-3} не могут быть включены в регрессионное уравнение.

После пересчета параметров регрессионного уравнения $y_t = \xi + \alpha_1 y_{t-1}$ получаем $R^2 = 0,646$ и стандартную ошибку оценки 15,44. Значение $\alpha_1 = 0,793$ ($p < 0,001$), $\xi = 13,76$ ($p = 0,05$). Следовательно, в итоге получаем авторегрессионное уравнение вида $y_t = 13,76 + 0,793y_{t-1}$ для прогнозирования последующих значений ряда.

Значения исходного ряда и прогнозные значения, полученные на основе построенного авторегрессионного уравнения, приведены на рис. 11.30. Как видим, данная авторегрессионная модель достаточно хорошо аппроксимирует исходный ряд, но не гарантирует уверенный прогноз на последующие периоды. Так, в периоды, следующие за 42-м, данная модель будет давать только нисходящий тренд.

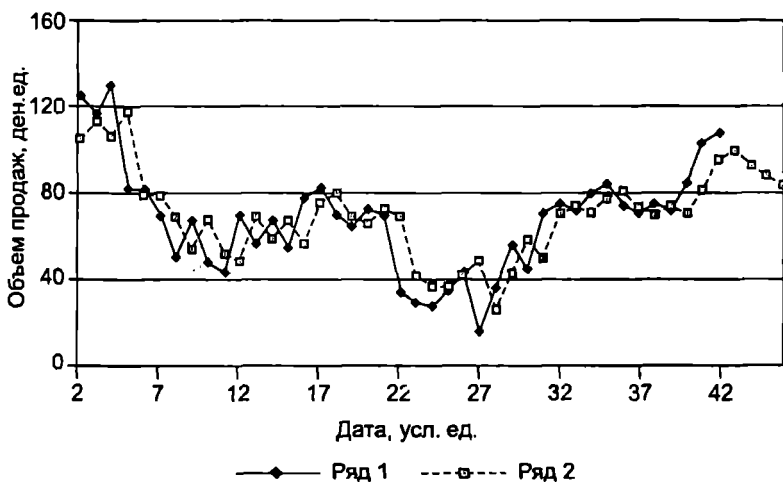


Рис. 11.30. Исходный ряд (ряд 1) и прогноз на основе авторегрессионного уравнения (ряд 2)

Для реализации процедуры построения авторегрессионной модели в SPSS существует специальная опция *Autoregression* (меню *Analyze – Time series*). В соответствующем диалоговом окне в качестве зависимой переменной (*Dependent*) задается переменная, соответствующая исходному ряду (в данном случае это переменная *sales*), в качестве независимой (*Independent*) переменной (для авторегрессии первого порядка) или независимых переменных (для авторегрессии более высокого порядка) указываются переменные, полученные из исходного ряда со значениями, смещенными на соответствующее число периодов. По умолчанию в рабочую область данных в качестве новых переменных выводятся прогнозные значения, абсолютная ошибка прогноза для каждого периода, значения стандартной ошибки оценки, а также нижние и верхние границы 95%-ного доверительного интервала. Полученные результаты будут полностью аналогичны тем, что получаются при соответствующем использовании процедуры множественной линейной регрессии, рассмотренной выше.

Пакеты SPSS и STATISTICA включают и более комплексную процедуру для анализа временных рядов, которая фактически объединяет все рассмотренные ранее относительно простые подходы. Эта модель получила название *ARIMA model* (*AutoRegressive Integrated Moving Average model*), а в русскоязычной литературе – модель АРПСС (модель авторегрессии и интегрированного скользящего среднего). Как следует из названия данной модели, она позволяет объединить процессы авторегрессии и скользящего среднего, а также выделить сезонную компоненту ряда. Соответственно модель содержит три типа параметров: параметры авторегрессии (p), порядок разности (d) и параметры скользящего среднего (q). Например, модель $(0, 1, 2)$ содержит нуль параметров авторегрессии (p) и два параметра скользящего среднего (q), которые вычисляются для ряда после взятия разности с лагом 1. Взятие разности необходимо, чтобы обеспечить стационарность ряда и исключить сезонные составляющие.

Для задания параметров данной модели в программе SPSS нужно выбрать опцию *ARIMA* (меню *Analyze – Time series*); в программе STATISTICA – опцию *ARIMA & Autocorrelation functions* (модуль *Time series/Forecasting*).

Данный метод весьма гибкий и мощный, но в то же время довольно сложный, для овладения им требуется серьезная практика¹.

¹ Детальному описанию данного метода посвящены специальные учебные и научные издания, например [4, 5].

Задания для самостоятельного выполнения

1. Табл. 11.2 содержит данные о еженедельных продажах книжного магазина (тыс. руб.). Можно ли считать изменения продаж случайными? Подходит ли для описания этих данных модель случайных изменений? Подтвердите ответ результатами автокорреляционного анализа. Постройте прогноз продаж на следующие четыре недели.

Таблица 11.2

Данные для решения задачи 1

Неделя	Продажи	Неделя	Продажи	Неделя	Продажи
1	151,07	9	158,32	17	143,59
2	147,31	10	143,36	18	160,60
3	154,64	11	137,37	19	160,96
4	155,60	12	154,64	20	136,81
5	161,35	13	149,84	21	151,90
6	149,01	14	156,83	22	155,29
7	147,10	15	141,69	23	140,05
8	161,73	16	143,31	24	146,57

2. Табл. 11.3 содержит данные о ежемесячных обращениях в банк для получения ипотечного кредита за последние два года. Можно ли считать изменения продаж случайными? Подходит ли для описания этих данных модель случайных изменений? Подтвердите ответ результатами автокорреляционного анализа. С помощью соответствующей модели постройте прогноз количества обращений на следующие три месяца.

Таблица 11.3

Данные для решения задачи 2

Месяц	Количество обращений	Месяц	Количество обращений	Месяц	Количество обращений
1	96	9	105	17	95
2	114	10	110	18	122
3	112	11	117	19	113
4	81	12	96	20	100
5	97	13	74	21	102
6	103	14	94	22	96
7	120	15	100	23	116
8	99	16	96	24	98

3. С помощью автокорреляционного анализа определите, действительно ли функция Excel СЛЧИС генерирует случайный ряд значений (используйте не менее 100 значений).

4. Используйте модель случайных блужданий для анализа и прогнозирования (на последующие четыре месяца) динамики индекса Доу-Джонса (табл. П. 14). Сравните с результатами, приведенными в подразд. 11.2. Задание выполните в Excel и SPSS (STATISTICA).

5. Используйте Excel для генерации 100 элементов временного ряда, описываемого моделью $y_t = y_{t-1} + 500 + \epsilon_t$, где ϵ_t – случайная величина со средним 0 и стандартным отклонением 10. Используйте модель случайных блужданий для анализа ряда. Сравните коррелограммы автокорреляций, частных автокорреляций и коррелограмму разностей первого порядка и сделайте выводы.

6. Используйте Excel для генерации 100 элементов временного ряда, описываемого моделью $y_t = y_{t-1} + \epsilon_t$, где ϵ_t – случайная величина со средним 0 и стандартным отклонением 1. Примените метод простого скользящего среднего для анализа полученного ряда. Покажите эффект влияния базы на результат сглаживания ряда. Какое значение этого параметра можно считать наилучшим для данного ряда? Задание выполните в Excel и SPSS (STATISTICA).

7. Используйте Excel для генерации 100 элементов временного ряда, описываемого моделью $y_t = 100 + \epsilon_t$, где ϵ_t – случайная величина со средним 0 и стандартным отклонением 15. Примените метод простого экспоненциального сглаживания для анализа полученного ряда с параметром сглаживания $\alpha = 0,1$. Какое значение этого параметра можно считать наилучшим для данного ряда? Является ли, по вашему мнению, данный метод сглаживания подходящим для данного ряда и почему?

8. Используйте Excel для генерации 100 элементов временного ряда, описываемого моделью $y_t = y_{t-1} + \mu + \epsilon_t$, где ϵ_t – случайная величина со средним 0 и стандартным отклонением 4. Примите значения $y_0 = 100$, $\mu = 1$. Используйте метод Холта для сглаживания полученного ряда с параметрами сглаживания $\alpha = \gamma = 0,1$. Какие значения этих параметров можно считать наилучшим для данного ряда? Является ли, по вашему мнению, данный метод сглаживания подходящим для данного ряда и почему?

Примените для данного ряда метод простого экспоненциального сглаживания с параметром $\alpha = 0,1$. Какое значение этого параметра можно считать наилучшим для данного ряда? Какой из двух используемых методов может обеспечить лучшую точность прогноза для данного ряда и почему?

9. В табл. 11.4 содержатся данные о сезонном посещении небольшого курортного комплекса за последние 5 лет [8]. Вычислите сезонный индекс для каждого квартала и исключите из данных сезонную составляющую. Если весной на курорте работало 50 человек, сколько человек нужно нанять в другие сезоны?

Таблица 11.4

Данные для решения задачи 9

Год	Весна	Лето	Осень	Зима
1	200	300	125	325
2	175	250	150	375
3	225	300	200	450
4	200	350	225	375
5	175	300	200	350

10. Для данных, представленных в табл. 11.5, определите оптимальные значения параметров α , γ , δ при различных критериях оптимизации (минимум MSE, MAE и MAPE). Сравните результаты и предложите их интерпретацию.

Таблица 11.5

Данные для решения задачи 10

t	y_t	t	y_t	t	y_t	t	y_t	t	y_t
Янв.03	29	Янв.04	35	Янв.05	34	Янв.06	34	Янв.07	45
Фев.03	32	Фев.04	43	Фев.05	40	Фев.06	29	Фев.07	49
Мар.03	41	Мар.04	46	Мар.05	43	Мар.06	36	Мар.07	62
Апр.03	44	Апр.04	46	Апр.05	42	Апр.06	42	Апр.07	62
Май 03	49	Май 04	43	Май 05	43	Май 06	43	Май 07	58
Июн.03	47	Июн.04	41	Июн.05	44	Июн.06	44	Июн.07	59
Июл.03	46	Июл.04	44	Июл.05	39	Июл.06	44	Июл.07	64
Авг.03	47	Авг.04	47	Авг.05	40	Авг.06	48	Авг.07	62
Сен.03	43	Сен.04	41	Сен.05	33	Сен.06	45	Сен.07	50
Окт.03	45	Окт.04	40	Окт.05	32	Окт.06	44	Окт.07	52
Ноя.03	34	Ноя.04	32	Ноя.05	31	Ноя.06	40	Ноя.07	50
Дек.03	31	Дек.04	32	Дек.05	28	Дек.06	37	Дек.07	44

11. В табл. 11.6 содержатся данные о ежеквартальных (t) продажах (S) продукции компании за период с 1998 по 2007 г. Используйте модель Винтера для анализа ряда. На трех-четырех примерах покажите, как влияют параметры модели на точность прогноза. Определите оптимальные значения параметров модели. Постройте прогноз объемов продаж на четыре квартала 2008 г.

Таблица 11.6

Ежеквартальные объемы продаж

t	S	t	S	t	S	t	S
Мар.98	1735	Сен.00	2128	Мар.03	2481	Сен.05	3629
Июн.98	2245	Дек.00	2027	Июн.03	3040	Дек.05	3373
Сен.98	2534	Мар.01	1911	Сен.03	3172	Мар.06	3352
Дек.98	2155	Июн.01	2331	Дек.03	2879	Июн.06	4342
Мар.99	1548	Сен.01	2207	Мар.04	2772	Сен.06	4461
Июн.99	2104	Дек.01	2174	Июн.04	3550	Дек.06	4017
Сен.99	2014	Мар.02	2148	Сен.04	3508	Мар.07	3854
Дек.99	1992	Июн.02	2739	Дек.04	3244	Июн.07	4936
Мар.00	1869	Сен.02	2793	Мар.05	3056	Сен.07	4895
Июн.00	2314	Дек.02	2556	Июн.05	3899	Дек.07	4333

12. Используя модель сезонной декомпозиции, проведите анализ временного ряда, представленного в табл. 11.6. Постройте прогноз объемов продаж на четыре квартала 2008 г. Сравните результаты с результатами, полученными в задаче 11.

13. В табл. 11.7 содержатся данные о среднегодовой процентной ставке по ипотечному кредитованию в банках США за период с 1973 по 1996 г. [54]. Определите наилучшую авторегрессионную модель для анализа данных и построения прогноза на 1997, 1998 и 1999 гг.

Таблица 11.7

Процентная ставка по ипотечному кредитованию

Год	Процент	Год	Процент	Год	Процент	Год	Процент
1973	8,04	1979	11,20	1985	12,43	1991	9,25
1974	9,19	1980	13,74	1986	10,19	1992	8,39
1975	9,05	1981	16,63	1987	10,21	1993	7,31
1976	8,87	1982	16,04	1988	10,34	1994	8,38
1977	8,85	1983	13,24	1989	10,32	1995	7,93
1978	9,64	1984	13,88	1990	10,13	1996	7,81

14. В табл. 11.8 содержатся данные о ежемесячных объемах продаж (количество упаковок) нового лекарства в течение двух лет после его вывода на рынок. Определите наилучшую авторегрессионную модель для анализа данных на основе автокорреляционного и авторегрессионного анализа и постройте прогноз на последующие два месяца.

Таблица 11.8

Продажи нового лекарства (шт.)

Месяц	Продажи	Месяц	Продажи	Месяц	Продажи
1	75	9	19188	17	4920002
2	150	10	38410	18	9832214
3	300	11	76779	19	19654552
4	600	12	153675	20	39358998
5	1200	13	307445	21	78620833
6	2402	14	614617	22	157442052
7	4799	15	1229335	23	314614938
8	9603	16	2457354	24	629107171

15. В табл. 11.9 содержатся данные о ежеквартальных продажах продукции компании за шестилетний период.

Таблица 11.9

Ежеквартальные объемы продаж (тыс. руб.)

Квартал	Продажи	Квартал	Продажи	Квартал	Продажи
1	50147	9	50862	17	56405
2	49325	10	53028	18	60031
3	57048	11	58849	19	71486
4	76781	12	79660	20	92183
5	48617	13	51640	21	60800
6	50898	14	54119	22	64900
7	58517	15	65681	23	76997
8	77691	16	85175	24	103337

15а. Постройте авторегрессионную модель, наилучшим образом описывающую данные. Постройте прогноз на два последующих квартала. Оцените точность прогноза.

15б. Постройте множественную линейную регрессию с использованием фиктивных переменных, соответствующих квар-

талам. Постройте прогноз на два последующих квартала. Оцените точность прогноза.

15в. Постройте нелинейную регрессионную модель вида

$y_t = \alpha\beta_1^t\beta_2^{X_1}\beta_3^{X_2}\beta_4^{X_3}$, где X_1, X_2, X_3 – фиктивные переменные: $X_1 = 1$, если t относится к I кварталу; $X_2 = 1$, если t относится ко II кварталу; $X_3 = 1$, если t относится к III кварталу. Постройте прогноз на два последующих квартала. Оцените точность прогноза. Сравните результаты 15а – 15в.

16. По данным табл. 9.12 с помощью известных вам методов анализа временных рядов проведите анализ всех показателей и постройте прогноз на первый месяц следующего года. Сравните результаты и сделайте выводы.

17. По данным табл. 9.13 с помощью известных вам методов анализа временных рядов проанализируйте динамику цен на продукты и постройте прогноз изменения цены каждого продукта на последующие три недели. Сравните результаты и сделайте выводы.

18. Примените различные модели сглаживания временного ряда для анализа и прогнозирования (на последующие четыре месяца) динамики индекса Доу-Джонса (табл. П.14). Сравните с результатами, приведенными в подразд. 11.3. Используйте авторегрессионную модель для анализа и прогнозирования (на последующие четыре месяца) значений ряда. Сравните результаты применения методов сглаживания и авторегрессионной модели и сделайте выводы.

19. Выполните задание 18, используя данные табл. 11.1.

20. В табл. 11.10 содержатся данные о динамике курса доллара США за период с 23 января по 19 февраля 2008 г. (данные приведены в обратной последовательности, курс на пропущенные дни принимается равным ближайшему предшествующему значению). Используя различные модели для анализа временных рядов, спрогнозируйте курс доллара на период с 20 по 23 февраля 2008 г. Сравните результаты применения разных методов анализа.

21. По данным табл. 10.29 с помощью известных вам методов анализа временных рядов сделайте прогноз пассажиропотока, стоимости проезда, численности населения, среднегодового дохода и стоимости парковки в 2005–2008 гг. Оцените точность этих прогнозов. Сравните результаты использования разных методов и сделайте выводы.

Динамика курса доллара США (2008 г.)

Дата	Курс	Дата	Курс	Дата	Курс
19.02	24,5767	08.02	24,6466	30.01	24,475
16.02	24,5861	07.02	24,6706	29.01	24,5952
15.02	24,6392	06.02	24,5211	26.01	24,4386
14.02	24,6655	05.02	24,4543	25.01	24,6349
13.02	24,6537	02.02	24,4201	24.01	24,6325
12.02	24,6715	01.02	24,4262	23.01	24,8917
09.02	24,7813	31.01	24,4764		

22. В табл. 11.11 содержатся данные о ежеквартальных доходах компании за период с I квартала 2003 г. по IV квартал 2007 г. Определите, содержит ли этот ряд тренд и сезонную компоненту? Какие из известных вам методов анализа временных рядов наиболее подходят для анализа и прогнозирования в данном случае? Поясните ответ. С помощью выбранных методов постройте прогноз на четыре квартала 2008 г. Сравните результаты и сделайте выводы.

Таблица 11.11

Ежеквартальные доходы компании (тыс. руб.)

Период	Доход	Период	Доход	Период	Доход	Период	Доход
1	1026	6	1249	11	1449	16	4200
2	1056	7	1346	12	3893	17	1776
3	1182	8	3402	13	1462	18	1808
4	2861	9	1286	14	1452	19	1942
5	1172	10	1317	15	1631	20	4129

23. В табл. П.15 представлены данные о ежеквартальных доходах ведущих американских компаний за период 1991–1995 гг. [54]. Определите, какие из методов анализа временных рядов наиболее подходят для анализа и прогнозирования доходов каждой компании. С помощью выбранных методов постройте прогноз на четыре квартала 1996 г. Оцените точность прогноза.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Аванесов В.С.* Применение тестовых форм в Rasch Measurement / В.С. Аванесов // Педагогические измерения. – 2004. – № 4. <http://testolog.narod.ru>
2. *Аванесов В.С.* Проблема качества педагогических измерений / В.С. Аванесов // Педагогические измерения. – 2004. – № 2. – С. 3–27.
3. *Айвазян С.А.* Прикладная статистика. Основы моделирования и первичная обработка данных: Справочное издание / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1983.
4. *Бокс Дж.* Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – Вып. 1, 2. – М.: Мир, 1974.
5. *Боровиков В.П.* Прогнозирование в системе STATISTICA в среде Windows: Основы теории и интенсивная практика на компьютере: учеб. пособие / В.П. Боровиков, Г.И. Ивченко. – М.: Финансы и статистика, 2006.
6. *Бююль А.* SPSS: искусство обработки информации / А. Бююль, П. Цефель. – М.: Diasoft, 2005.
7. *Виноградов А.Г.* Надежность / HR-Лаборатория Human Technologies. www.ht.ru/on-line/forum/?forum=7&topic=190
8. *Вуколов Э.А.* Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL / Э.А. Вуколов. – М.: ФОРУМ: Инфра-М, 2004.
9. *Дубнов П.Ю.* Обработка статистической информации с помощью SPSS / П.Ю. Дубнов. – М.: АСТ, 2005.
10. *Дубина И.Н.* Использование матрицы приоритетов улучшения качества в маркетинговой деятельности предприятия / И.Н. Дубина, А.Е. Ошкало // Социально-экономическая политика государства и возможности ее реализации на региональном уровне. – Барнаул: Изд-во Алт. ун-та, 2006.
11. *Дюк В.А.* Конструирование психодиагностических тестов: традиционные математические модели и алгоритмы. <http://psyfactor.org/lib/>
12. *Кендалл М.* Временные ряды / М. Кендалл. – М.: Финансы и статистика, 1981.

13. *Ковалева Г.Д.* Анализ социологических данных с применением статистического пакета SPSS: учебно-метод. пособие / Г.Д. Ковалева, П.С. Ростовцев. – Новосибирск: НГУ, 2002.
14. *Крыштановский А.О.* Анализ социологических данных с помощью пакета SPSS / А.О. Крыштановский. – М.: ГУ ВШЭ, 2006.
15. *Кун Т.* Структура научных революций / Т. Кун. – М.: Прогресс, 1975.
16. *Лавренов С.М.* Excel: Сборник примеров и задач / С.М. Лавренов. – М.: Финансы и статистика, 2003.
17. *Наследов А.Д.* SPSS. Компьютерный анализ данных в психологии и социальных науках / А.Д. Наследов. – СПб.: Питер, 2005.
18. *Наследов А.Д.* Математические методы психологического исследования. Анализ и интерпретация данных / А.Д. Наследов. – СПб.: Речь, 2006.
19. *Нейман Ю.М.* Введение в теорию моделирования и параметризации педагогических тестов / Ю.М. Нейман, В.А. Хлебников. – М.: Прометей, 2000.
20. *Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: учебник / под ред. О.Э. Башиной, А.А. Спирина.* – М.: Финансы и статистика, 2007.
21. *Орлов А.И.* Прикладная статистика / А.И. Орлов. – М.: Экзамен, 2004.
22. *Орлов А.И.* Репрезентативная теория измерений и ее применения / А.И. Орлов // Заводская лаборатория. – 1999. – Т. 65. – № 3.
23. *Орлов А.И.* Часто ли распределение результатов наблюдений является нормальным? / А.И. Орлов // Заводская лаборатория. – 1991. – Т. 57. – № 7.
24. *Пациорковский В.В.* SPSS для социологов / В.В. Пациорковский, В.В. Пациорковская. – М.: ИСЭПН РАН, 2005.
25. *Плис А.И.* Практикум по прикладной статистике в среде SPSS: В 2-х ч. / А.И. Плис, Н.А. Сливина. – М.: Финансы и статистика, 2004.
26. *Практикум по общей теории статистики / под ред. Н.Н. Ряузова.* – М.: Финансы и статистика, 1981.
27. *Таганов Д.Н.* SPSS: Статистический анализ в маркетинговых исследованиях / Д.Н. Таганов. – СПб.: Питер, 2005.
28. *Тюрин Ю.Н.* Статистический анализ данных на компьютере / Ю.Н. Тюрин, А.А. Макаров. – М.: Инфра-М, 1998.
29. *Хили Дж.* Статистика: Социологические и маркетинговые исследования / Дж. Хили. – СПб.: Питер, 2005.

30. *Amabile T.M., Burnside R.M. and Gryskiewicz S.S.* User's manual for assessing the climate for creativity: A survey from the Center for Creative Leadership. Greensboro, NC, 1999.
31. *Anastasi A.* Psychological testing (6th ed.). New York: MacMillan Publishing Company, 1988.
32. *Basadur M., Graen G. and Wakabayashi M.* Identifying individual differences in creative problem solving style, *Journal of Creative Behavior*. – 1990. – №. 24. – Pp. 111–131.
33. *Baker F.* The Basics of Item Response Theory. University of Maryland, 2001.
34. *Black T.R.* Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics. SAGE Publications, 1999.
35. *Bond T.G. and Fox C.M.* Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Lawrence Erlbaum, 2001.
36. *Campbell D.T.* Social measurement. Thousand Oaks: SAGE Publications, 2001.
37. *Cannel C., Fowler F.J., Kalton G., Oksenberg L. and Bishoping K.* New quantitative techniques for pretesting survey questions, in *Bulmer M.* (Ed) Questionnaires. Thousand Oaks: SAGE Publications. – 2004. – Vol. 4. – Pp. 187–201.
38. *Carlson M.* GTE Directories: Customer Focus and Satisfaction. The Quest for Excellence VII, The Official Conference of the Malcolm Baldrige National Quality Award, February 6–8, 1995, Washington, DC, 1995.
39. *Cooper D.R. and Shindler P.S.* Business Research Methods. Irwin/McGraw-Hill, 1995.
40. *Dubina I.N. and Umpleby S.A.* «Agenda Setting and Improvement Monitoring in a University Department». The Proceedings of the 12th International Deming Research Seminar. New York: The W. Edwards Deming Institute, 2006.
41. *Ekvall G.* «Organizational climate for creativity and innovation». *European Journal of Work and Organizational Psychology*. – V. 5. – 1996.
42. *Feick L.F.* «Latent class analysis of survey questions that include don't know responses», in *Bulmer, M.* Questionnaires. – V. 4. – Thousand Oaks: SAGE Publications, 2004.
43. *Forman E.H. and Selly M.A.* Decision By Objectives. World Scientific, 2001. <http://www.expertchoice.com/dbo/>

44. *Hess J., Singer E. and Bushrey J.* «Predicting Test-Retest reliability from behavior coding», in Bulmer, M. Questionnaires. – V. 4. – Thousand Oaks: SAGE Publications, 2004.

45. *Karabatsos G.* «Axiomatic measurement theory as a basis for model selection in item response theory». Paper presented at 32nd annual conference of the Society for Mathematical Psychology, Santa Cruz, CA, 1999.

46. *Kirton M.J.* Adaptors and innovators: a description and measure, in Puccio, G.J. and Murdock, M.C. (Eds.) Creativity Assessment: Readings and Resources, Buffalo: Creative Education Foundation Press, 1999.

47. *Litwin M.S.* How to measure survey reliability and validity. Thousand Oaks: SAGE Publications, 1995.

48. *Mathisen G.R. and Einarsen S.* «A review of instruments assessing creative and innovative environments within organizations», Creativity Research Journal. – V 16. – № 1. – 2004.

49. *Maxwell J.A.* Qualitative research design: An interactive approach. Thousand Oaks: SAGE Publications, 1996.

50. *Parry H.J. and Crossley H.M.* «Validity of responses to survey questions», in Bulmer, M. Questionnaires. – V. 3. – Thousand Oaks: SAGE Publications, 2004.

51. *Shuman H.* «The random probe: A technique for evaluating the validity of closed questions», in Bulmer, M. Questionnaires. – V. 3. – Thousand Oaks: SAGE Publications, 2004.

52. *Tewari J., Cooper D.R. and Shindler P.S.* Instructor's Resource Guide to accompany Business Research Methods. Irwin/McGraw-Hill, 1995.

53. *Traub R.E.* Reliability for the social sciences: theory and applications. Thousand Oaks: SAGE Publications, 1994.

54. *Winston W.L., Albright S.C.* Practical Management Science. Duxbury, 2001.

55. *Wright B.D. and Masters G.N.* Rating Scale Analysis, MESA Press, 1982.

Интернет-ресурсы

56. Образовательный портал «Экономика. Социология. Менеджмент: методы сбора и анализа социологических данных». <http://ecsocman.edu.ru/db/sect/124/36.html>

57. Сайт В.С. Аванесова. <http://testolog.narod.ru>

58. Социологические исследования. <http://ecsocman.edu.ru/socis/>

59. Социология и маркетинг в сети. <http://socio.net.narod.ru/stat.html>

60. Социология: методология, методы, математические модели. <http://www.nir.ru/socio/scipubl/4M.htm>
61. Статистика и обработка данных в психологии. <http://psyfactor.org/lib/>
62. Статистические методы. Сайт А.И. Орлова. <http://orlovs.pp.ru/stat.php>
63. Центр маркетинга и информационного менеджмента РГГУ. Словарь терминов. <http://marketing.rsuh.ru>
64. Экономика и математические методы. <http://www.cemi.rssi.ru/emm/>
65. Cumulative Item Response Theory Models. <http://www.education.umd.edu/Depts/EDMS/tutorials/CIRT.html>
66. Exponenta.ru. Образовательный математический сайт. <http://www.exponenta.ru>
67. Institute for Objective Measurement. <http://www.rasch.org>
68. StatSoft Russia. <http://www.statsoft.ru>
69. Wikipedia, the free encyclopedia. <http://en.wikipedia.org>

ПРИЛОЖЕНИЯ

1. СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Таблица П.1

Стандартное нормальное распределение

Для значений z указана площадь (S_R) под кривой стандартного нормального распределения между средним значением (0) и значением z (правая половина). В силу симметричности кривой нормального распределения площади слева (0 и z) будут иметь такое же значение. Общая площадь под кривой определяется как $2S_R$. В таблице также приведены значения α (для ненаправленной проверки альтернативной гипотезы). При односторонней (направленной) проверке значения α следует делить пополам.

z	S_R	α	z	S_R	α	z	S_R	α
0,60	0,226	0,549	0,86	0,305	0,390	1,12	0,369	0,263
0,61	0,229	0,542	0,87	0,308	0,384	1,13	0,371	0,258
0,62	0,232	0,535	0,88	0,311	0,379	1,14	0,373	0,254
0,63	0,236	0,529	0,89	0,313	0,373	1,15	0,375	0,250
0,64	0,239	0,522	0,90	0,316	0,368	1,16	0,377	0,246
0,65	0,242	0,516	0,91	0,319	0,363	1,17	0,379	0,242
0,66	0,245	0,509	0,92	0,321	0,358	1,18	0,381	0,238
0,67	0,249	0,503	0,93	0,324	0,352	1,19	0,383	0,234
0,68	0,252	0,497	0,94	0,326	0,347	1,20	0,385	0,230
0,69	0,255	0,490	0,95	0,329	0,342	1,21	0,387	0,226
0,70	0,258	0,484	0,96	0,331	0,337	1,22	0,389	0,222
0,71	0,261	0,478	0,97	0,334	0,332	1,23	0,391	0,219
0,72	0,264	0,472	0,98	0,336	0,327	1,24	0,393	0,215
0,73	0,267	0,465	0,99	0,339	0,322	1,25	0,394	0,211
0,74	0,270	0,459	1,00	0,341	0,317	1,26	0,396	0,208
0,75	0,273	0,453	1,01	0,344	0,312	1,27	0,398	0,204
0,76	0,276	0,447	1,02	0,346	0,308	1,28	0,400	0,201
0,77	0,279	0,441	1,03	0,348	0,303	1,29	0,401	0,197
0,78	0,282	0,435	1,04	0,351	0,298	1,30	0,403	0,194
0,79	0,285	0,430	1,05	0,353	0,294	1,31	0,405	0,190
0,80	0,288	0,424	1,06	0,355	0,289	1,32	0,407	0,187
0,81	0,291	0,418	1,07	0,358	0,285	1,33	0,408	0,184
0,82	0,294	0,412	1,08	0,360	0,280	1,34	0,410	0,180
0,83	0,297	0,407	1,09	0,362	0,276	1,35	0,411	0,177
0,84	0,300	0,401	1,10	0,364	0,271	1,36	0,413	0,174
0,85	0,302	0,395	1,11	0,367	0,267	1,37	0,415	0,171

z	S_R	α	z	S_R	α	z	S_R	α
1,38	0,416	0,168	1,78	0,462	0,075	2,18	0,485	0,029
1,39	0,418	0,165	1,79	0,463	0,073	2,19	0,486	0,029
1,40	0,419	0,162	1,80	0,464	0,072	2,20	0,486	0,028
1,41	0,421	0,159	1,81	0,465	0,070	2,21	0,486	0,027
1,42	0,422	0,156	1,82	0,466	0,069	2,22	0,487	0,026
1,43	0,424	0,153	1,83	0,466	0,067	2,23	0,487	0,026
1,44	0,425	0,150	1,84	0,467	0,066	2,24	0,487	0,025
1,45	0,426	0,147	1,85	0,468	0,064	2,25	0,488	0,024
1,46	0,428	0,144	1,86	0,469	0,063	2,26	0,488	0,024
1,47	0,429	0,142	1,87	0,469	0,061	2,27	0,488	0,023
1,48	0,431	0,139	1,88	0,470	0,060	2,28	0,489	0,023
1,49	0,432	0,136	1,89	0,471	0,059	2,29	0,489	0,022
1,50	0,433	0,134	1,90	0,471	0,057	2,30	0,489	0,021
1,51	0,434	0,131	1,91	0,472	0,056	2,31	0,490	0,021
1,52	0,436	0,129	1,92	0,473	0,055	2,32	0,490	0,020
1,53	0,437	0,126	1,93	0,473	0,054	2,33	0,490	0,020
1,54	0,438	0,124	1,94	0,474	0,052	2,34	0,490	0,019
1,55	0,439	0,121	1,95	0,474	0,051	2,35	0,491	0,019
1,56	0,441	0,119	1,96	0,475	0,050	2,36	0,491	0,018
1,57	0,442	0,116	1,97	0,476	0,049	2,37	0,491	0,018
1,58	0,443	0,114	1,98	0,476	0,048	2,38	0,491	0,017
1,59	0,444	0,112	1,99	0,477	0,047	2,39	0,492	0,017
1,60	0,445	0,110	2,00	0,477	0,046	2,40	0,492	0,016
1,61	0,446	0,107	2,01	0,478	0,044	2,41	0,492	0,016
1,62	0,447	0,105	2,02	0,478	0,043	2,42	0,492	0,016
1,63	0,448	0,103	2,03	0,479	0,042	2,43	0,492	0,015
1,64	0,449	0,101	2,04	0,479	0,041	2,44	0,493	0,015
1,65	0,451	0,099	2,05	0,480	0,040	2,45	0,493	0,014
1,66	0,452	0,097	2,06	0,480	0,039	2,46	0,493	0,014
1,67	0,453	0,095	2,07	0,481	0,038	2,47	0,493	0,014
1,68	0,454	0,093	2,08	0,481	0,038	2,48	0,493	0,013
1,69	0,454	0,091	2,09	0,482	0,037	2,49	0,494	0,013
1,70	0,455	0,089	2,10	0,482	0,036	2,50	0,494	0,012
1,71	0,456	0,087	2,11	0,483	0,035	2,51	0,494	0,012
1,72	0,457	0,085	2,12	0,483	0,034	2,52	0,494	0,012
1,73	0,458	0,084	2,13	0,483	0,033	2,53	0,494	0,011
1,74	0,459	0,082	2,14	0,484	0,032	2,54	0,494	0,011
1,75	0,460	0,080	2,15	0,484	0,032	2,55	0,495	0,011
1,76	0,461	0,078	2,16	0,485	0,031	2,56	0,495	0,010
1,77	0,462	0,077	2,17	0,485	0,030	2,57	0,495	0,010

z	S_R	α	z	S_R	α	z	S_R	α
2,58	0,495	0,010	2,79	0,497	0,005	3,00	0,499	0,003
2,59	0,495	0,010	2,80	0,497	0,005	3,01	0,499	0,003
2,60	0,495	0,009	2,81	0,498	0,005	3,02	0,499	0,003
2,61	0,495	0,009	2,82	0,498	0,005	3,03	0,499	0,002
2,62	0,496	0,009	2,83	0,498	0,005	3,04	0,499	0,002
2,63	0,496	0,009	2,84	0,498	0,005	3,05	0,499	0,002
2,64	0,496	0,008	2,85	0,498	0,004	3,06	0,499	0,002
2,65	0,496	0,008	2,86	0,498	0,004	3,07	0,499	0,002
2,66	0,496	0,008	2,87	0,498	0,004	3,08	0,499	0,002
2,67	0,496	0,008	2,88	0,498	0,004	3,09	0,499	0,002
2,68	0,496	0,007	2,89	0,498	0,004	3,10	0,499	0,002
2,69	0,496	0,007	2,90	0,498	0,004	3,11	0,499	0,002
2,70	0,497	0,007	2,91	0,498	0,004	3,12	0,499	0,002
2,71	0,497	0,007	2,92	0,498	0,004	3,13	0,499	0,002
2,72	0,497	0,007	2,93	0,498	0,003	3,14	0,499	0,002
2,73	0,497	0,006	2,94	0,498	0,003	3,15	0,499	0,002
2,74	0,497	0,006	2,95	0,498	0,003	3,16	0,499	0,002
2,75	0,497	0,006	2,96	0,498	0,003	3,17	0,499	0,002
2,76	0,497	0,006	2,97	0,499	0,003	3,18	0,499	0,001
2,77	0,497	0,006	2,98	0,499	0,003	3,19	0,499	0,001
2,78	0,497	0,005	2,99	0,499	0,003	3,20	0,499	0,001

t-распределение Стьюдента

df	Уровни значимости для одностороннего критерия					
	0,1	0,05	0,025	0,01	0,005	0,0005
	Уровни значимости для двустороннего критерия					
	0,2	0,1	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,656	636,578
2	1,886	2,920	4,303	6,965	9,925	31,600
3	1,638	2,353	3,182	4,541	5,841	12,924
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,408
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,768
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,689
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,660
30	1,310	1,697	2,042	2,457	2,750	3,646
35	1,306	1,690	2,030	2,438	2,724	3,591
40	1,303	1,684	2,021	2,423	2,704	3,551
45	1,301	1,679	2,014	2,412	2,690	3,520

df	Уровни значимости для одностороннего критерия					
	0,1	0,05	0,025	0,01	0,005	0,0005
	Уровни значимости для двустороннего критерия					
	0,2	0,1	0,05	0,02	0,01	0,001
50	1,299	1,676	2,009	2,403	2,678	3,496
55	1,297	1,673	2,004	2,396	2,668	3,476
60	1,296	1,671	2,000	2,390	2,660	3,460
65	1,295	1,669	1,997	2,385	2,654	3,447
70	1,294	1,667	1,994	2,381	2,648	3,435
75	1,293	1,665	1,992	2,377	2,643	3,425
80	1,292	1,664	1,990	2,374	2,639	3,416
85	1,292	1,663	1,988	2,371	2,635	3,409
90	1,291	1,662	1,987	2,368	2,632	3,402
95	1,291	1,661	1,985	2,366	2,629	3,396
100	1,290	1,660	1,984	2,364	2,626	3,390
150	1,287	1,655	1,976	2,351	2,609	3,357
200	1,286	1,653	1,972	2,345	2,601	3,340
250	1,285	1,651	1,969	2,341	2,596	3,330
300	1,284	1,650	1,968	2,339	2,592	3,323

F-распределение

 $\alpha = 0,05$

dfw	dfn										
	2	3	4	5	6	8	10	15	20	∞	
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,40	19,43	19,45	19,50
3	10,128	9,552	9,277	9,117	9,013	8,941	8,845	8,785	8,703	8,660	8,527
4	7,709	6,944	6,591	6,388	6,256	6,163	6,041	5,964	5,858	5,803	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,818	4,735	4,619	4,558	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,147	4,060	3,938	3,874	3,669
7	5,591	4,737	4,347	4,120	3,972	3,866	3,726	3,637	3,511	3,445	3,230
8	5,318	4,459	4,066	3,838	3,688	3,581	3,438	3,347	3,218	3,150	2,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,230	3,137	3,006	2,936	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,072	2,978	2,845	2,774	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	2,948	2,854	2,719	2,646	2,405
12	4,747	3,885	3,490	3,259	3,106	2,996	2,849	2,753	2,617	2,544	2,297
13	4,667	3,806	3,411	3,179	3,025	2,915	2,767	2,671	2,533	2,459	2,207
14	4,600	3,739	3,344	3,112	2,958	2,848	2,699	2,602	2,463	2,388	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,641	2,544	2,403	2,328	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,591	2,494	2,352	2,276	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,548	2,450	2,308	2,230	1,961
18	4,414	3,555	3,160	2,928	2,773	2,661	2,510	2,412	2,269	2,191	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,477	2,378	2,234	2,155	1,879
20	4,351	3,493	3,098	2,866	2,711	2,599	2,447	2,348	2,203	2,124	1,844
21	4,325	3,467	3,072	2,840	2,685	2,573	2,420	2,321	2,176	2,096	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,397	2,297	2,151	2,071	1,784
23	4,279	3,422	3,028	2,796	2,640	2,528	2,375	2,275	2,128	2,048	1,758
24	4,260	3,403	3,009	2,776	2,621	2,508	2,355	2,255	2,108	2,027	1,734
25	4,242	3,385	2,991	2,759	2,603	2,490	2,337	2,236	2,089	2,007	1,712
30	4,171	3,316	2,922	2,690	2,534	2,421	2,266	2,165	2,015	1,932	1,623
35	4,121	3,267	2,874	2,641	2,485	2,372	2,217	2,114	1,963	1,878	1,559
40	4,085	3,232	2,839	2,606	2,449	2,336	2,180	2,077	1,924	1,839	1,510
45	4,057	3,204	2,812	2,579	2,422	2,308	2,152	2,049	1,895	1,808	1,471
50	4,034	3,183	2,790	2,557	2,400	2,286	2,130	2,026	1,871	1,784	1,439
60	4,001	3,150	2,758	2,525	2,368	2,254	2,097	1,993	1,836	1,748	1,390
70	3,978	3,128	2,736	2,503	2,346	2,231	2,074	1,969	1,812	1,722	1,354
80	3,960	3,111	2,719	2,486	2,329	2,214	2,056	1,951	1,793	1,703	1,326
90	3,947	3,098	2,706	2,473	2,316	2,201	2,043	1,938	1,779	1,688	1,303
100	3,936	3,087	2,696	2,463	2,305	2,191	2,032	1,927	1,768	1,676	1,284
∞	3,842	2,996	2,605	2,372	2,214	2,099	1,939	1,831	1,666	1,571	1,000

dfiv	uH										
	1	2	3	4	5	6	8	10	15	20	∞
2	98,50	99,00	99,16	99,25	99,30	99,33	99,38	99,40	99,43	99,45	99,46
3	34,12	30,82	29,46	28,71	28,24	27,91	27,49	27,23	26,87	26,69	26,61
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,55	14,20	14,02	13,97
5	16,26	13,27	12,06	11,39	10,97	10,67	10,29	10,05	9,722	9,553	9,521
6	13,75	10,93	9,780	9,148	8,746	8,466	8,102	7,874	7,559	7,396	7,366
7	12,25	9,547	8,451	7,847	7,460	7,191	6,840	6,620	6,314	6,155	6,126
8	11,26	8,649	7,591	7,006	6,632	6,371	6,029	5,814	5,515	5,359	5,330
9	10,56	8,022	6,992	6,422	6,057	5,802	5,467	5,257	4,962	4,808	4,779
10	10,04	7,559	6,552	5,994	5,636	5,386	5,057	4,849	4,558	4,405	4,376
11	9,646	7,206	6,217	5,668	5,316	5,069	4,744	4,539	4,251	4,099	4,070
12	9,330	6,927	5,953	5,412	5,064	4,821	4,499	4,296	4,010	3,858	3,829
13	9,074	6,701	5,739	5,205	4,862	4,620	4,302	4,100	3,815	3,663	3,634
14	8,862	6,515	5,564	5,035	4,695	4,456	4,140	3,939	3,656	3,505	3,476
15	8,683	6,359	5,417	4,893	4,556	4,318	4,004	3,805	3,522	3,372	3,343
16	8,531	6,226	5,292	4,773	4,437	4,202	3,890	3,691	3,409	3,259	3,230
17	8,400	6,112	5,185	4,669	4,336	4,101	3,791	3,593	3,312	3,162	3,133
18	8,285	6,013	5,092	4,579	4,248	4,015	3,705	3,508	3,227	3,077	3,048
19	8,185	5,926	5,010	4,500	4,171	3,939	3,631	3,434	3,153	3,003	2,974
20	8,096	5,849	4,938	4,431	4,103	3,871	3,564	3,368	3,088	2,938	2,909
21	8,017	5,780	4,874	4,369	4,042	3,811	3,506	3,310	3,030	2,880	2,851
22	7,945	5,719	4,817	4,313	3,988	3,758	3,453	3,258	2,978	2,827	2,798
23	7,881	5,664	4,765	4,264	3,939	3,710	3,406	3,211	2,931	2,780	2,751
24	7,823	5,614	4,718	4,218	3,895	3,667	3,363	3,168	2,889	2,738	2,709
25	7,770	5,568	4,675	4,177	3,855	3,627	3,324	3,129	2,850	2,699	2,670
30	7,562	5,390	4,510	4,018	3,699	3,473	3,173	2,979	2,700	2,549	2,520
35	7,419	5,268	4,396	3,908	3,592	3,368	3,069	2,876	2,597	2,445	2,416
40	7,314	5,178	4,313	3,828	3,514	3,291	2,993	2,801	2,522	2,369	2,340
45	7,234	5,110	4,249	3,767	3,454	3,232	2,935	2,743	2,464	2,311	2,282
50	7,171	5,057	4,199	3,720	3,408	3,186	2,890	2,698	2,419	2,265	2,236
60	7,077	4,977	4,126	3,649	3,339	3,119	2,823	2,632	2,352	2,198	2,169
70	7,011	4,922	4,074	3,600	3,291	3,071	2,777	2,585	2,306	2,150	2,121
80	6,963	4,881	4,036	3,563	3,255	3,036	2,742	2,551	2,271	2,115	2,086
90	6,925	4,849	4,007	3,535	3,228	3,009	2,715	2,524	2,244	2,088	2,059
100	6,895	4,824	3,984	3,513	3,206	2,988	2,694	2,503	2,223	2,067	2,038
∞	6,635	4,605	3,787	3,310	3,017	2,807	2,511	2,321	2,040	1,878	1,849

χ^2 -распределение

df	Уровень значимости			
	0,1	0,05	0,01	0,001
1	2,706	3,841	6,635	10,827
2	4,605	5,991	9,210	13,815
3	6,251	7,815	11,345	16,266
4	7,779	9,488	13,277	18,466
5	9,236	11,070	15,086	20,515
6	10,645	12,592	16,812	22,457
7	12,017	14,067	18,475	24,321
8	13,362	15,507	20,090	26,124
9	14,684	16,919	21,666	27,877
10	15,987	18,307	23,209	29,588
11	17,275	19,675	24,725	31,264
12	18,549	21,026	26,217	32,909
13	19,812	22,362	27,688	34,527
14	21,064	23,685	29,141	36,124
15	22,307	24,996	30,578	37,698
16	23,542	26,296	32,000	39,252
17	24,769	27,587	33,409	40,791
18	25,989	28,869	34,805	42,312
19	27,204	30,144	36,191	43,819
20	28,412	31,410	37,566	45,314
21	29,615	32,671	38,932	46,796
22	30,813	33,924	40,289	48,268
23	32,007	35,172	41,638	49,728
24	33,196	36,415	42,980	51,179
25	34,382	37,652	44,314	52,619
26	35,563	38,885	45,642	54,051
27	36,741	40,113	46,963	55,475
28	37,916	41,337	48,278	56,892
29	39,087	42,557	49,588	58,301
30	40,256	43,773	50,892	59,702
35	46,059	49,802	57,342	66,619
40	51,805	55,758	63,691	73,403
45	57,505	61,656	69,957	80,078
50	63,167	67,505	76,154	86,660
60	74,397	79,082	88,379	99,608
70	85,527	90,531	100,425	112,317
80	96,578	101,879	112,329	124,839
90	107,565	113,145	124,116	137,208
100	118,498	124,342	135,807	149,449
200	226,021	233,994	249,445	267,539
300	331,788	341,395	359,906	381,424
400	436,649	447,632	468,724	493,131
500	540,930	553,127	576,493	603,446

**Критические значения критерия Манна–Уитни
(для ненаправленных проверок) [18, 368]**

n_2	n_1													
	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	34	41	48	55	62	69	76	83	90	98	105	112	119	127

**Критические значения D одновыборочного теста
Колмогорова–Смирнова [39, 687]**

	Уровни значимости	
	0,05	0,01
1	0,98	0,99
2	0,78	0,93
3	0,64	0,83
4	0,56	0,73
5	0,51	0,67
6	0,47	0,62
7	0,44	0,58
8	0,41	0,54
9	0,39	0,51
10	0,37	0,49
11	0,35	0,47
12	0,34	0,45
13	0,33	0,43
14	0,31	0,42
15	0,30	0,40
17	0,29	0,38
19	0,27	0,36
25	0,24	0,32
30	0,22	0,29
35	0,21	0,27
Свыше 35	$1,36/\sqrt{n}$	$1,63/\sqrt{n}$

**Критические значения D двухвыборочного теста
Колмогорова–Смирнова [39; 688, 689]**

N	Уровень значимости			
	направленная проверка		ненаправленная проверка	
	0,05	0,01	0,05	0,01
3	3	–	–	–
4	4	–	–	–
5	4	5	5	5
6	5	6	5	6
7	5	6	6	6
8	5	6	6	7
9	6	7	6	7
10	6	7	7	8
11	6	8	7	8
12	6	8	7	8
13	7	8	7	9
14	7	8	8	9
15	7	9	8	9
20	8	10	9	11
25	9	11	10	12
30	10	12	11	13
35	11	13	12	–
40	11	14	13	–
Свыше 40	–	–	$1,36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	$1,63 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$

Таблица П.8

Критические значения T -критерия Уилкоксона [39, 686]

n	Направленная проверка			n	Направленная проверка		
	0,025	0,01	0,005		0,025	0,01	0,005
	Ненаправленная проверка				Ненаправленная проверка		
	0,05	0,02	0,01		0,05	0,02	0,01
7	2	0	–	17	35	28	23
8	4	2	0	18	40	33	28
9	6	3	2	19	46	38	32
10	8	5	3	20	52	43	38
11	11	7	5	21	59	49	43
12	14	10	7	22	66	56	49
13	17	13	10	23	73	62	55
14	21	16	13	24	81	69	61
15	25	20	16	25	89	77	68
16	30	24	20				

2. ДАННЫЕ ДЛЯ ВЫПОЛНЕНИЯ ПРАКТИЧЕСКИХ ЗАДАЧ

Таблица П.9

Фрагмент массива данных по оценке организационного климата в компаниях

№ п/п	Компания 1							Компания 2						
	Пункт анкеты							Пункт анкеты						
	п1	п2	п3	п4	п5	п6	п7	п1	п2	п3	п4	п5	п6	п7
1	4	3	5	4	3	3	4	4	4	5	2	2	2	2
2	2	3	4	3	4	1	5	4	3	4	4	4	4	4
3	4	2	5	1	1	1	4	3	4	5	4	3	2	3
4	4	4	4	4	4	4	4	3	5	5	5	5	5	5
5	3	5	3	1	3	1	2	3	4	4	4	3	2	2
6	1	3	1	4	2	2	2	4	5	4	5	3	2	3
7	3	4	4	5	3	2	3	4	3	3	3	3	3	4
8	2	5	3	2	2	3	2	2	1	3	5	1	5	2
9	3	5	4	1	4	4	4	3	5	4	4	4	4	3
10	4	4	4	2	3	3	3	3	5	4	1	2	2	3
11	4	5	4	3	3	4	4	3	4	2	3	2	1	2
12	1	2	1	3	2	3	1	3	4	4	5	3	5	4
13	5	3	1	1	1	1	1	2	5	5	5	1	2	4
14	3	4	1	3	3	2	3	4	5	4	4	3	3	3
15	3	3	3	1	2	1	1	3	5	4	5	5	4	4
16	4	4	4	4	4	2	4	3	2	2	1	1	2	1
17	3	3	4	4	3	3	2	4	4	4	4	4	4	4
18	4	4	2	1	2	2	2	1	2	1	1	1	1	2
19	3	5	3	4	3	3	3	3	4	3	4	3	2	3
20	4	4	4	4	4	4	4	4	4	4	4	3	4	3
21	4	4	4	4	4	4	4	3	4	3	3	2	3	3
22	5	4	5	3	3	4	4	3	4	5	5	3	5	2
23	2	4	2	4	2	2	1	5	5	5	5	3	5	4
24	2	2	2	3	3	3	2	3	3	5	5	3	5	4
25	3	5	3	4	4	1	1	3	3	4	2	3	1	3
26	3	3	2	3	2	1	1	4	1	5	3	2	4	4
27	5	4	5	4	4	4	4	3	3	3	4	3	3	3
28	3	4	2	5	3	4	4	1	4	4	4	1	4	1
29	1	5	4	4	3	3	5	4	4	5	4	3	5	4
30	3	4	3	2	2	3	3	3	4	5	1	4	5	3

Данные для построения матрицы QIPM

Фактор	Респондент																					
	1		2		3		4		5		6		7		8		9		10		11	
	I	P	I	P	I	P	I	P	I	P	I	P	I	P	I	P	I	P	I	P	I	P
1	9	7	-	-	9	6	9	7	5	5	6	7	7	6	6	7	6	6	8	8	7	7
2	8	6	-	-	9	6	9	7	8	5	7	7	7	5	8	8	6	7	8	8	7	7
3	8	8	5	2	9	1	9	5	9	1	9	9	4	5	7	4	6	6	7	8	7	5
4	8	5	-	-	8	2	9	2	9	1	6	5	5	5	5	5	6	7	7	2	7	4
5	7	7	7	3	9	3	9	5	7	1	8	5	7	3	9	3	7	7	8	5	6	6
6	8	7	9	6	8	5	9	9	9	4	8	7	5	5	7	7	7	8	8	7	6	8
7	8	7	2	9	1	1	7	7	3	5	8	4	6	6	5	5	6	6	2	2	6	6
8	3	6	1	5	5	1	5	5	6	3	5	5	5	6	6	4	5	2	2	4	5	5
9	1	7	1	5	2	1	5	5	5	3	5	5	5	4	4	5	5	2	2	3	3	
10	7	5	5	1	7	6	9	4	9	1	8	7	6	4	7	2	6	7	8	6	7	3
11	8	7	9	8	7	7	9	5	5	5	5	5	-	-	-	-	6	7	8	8	6	6
12	9	8	7	9	8	6	9	2	9	1	7	4	6	4	9	5	6	8	8	1	7	7
13	8	7	9	1	9	1	9	2	9	1	7	8	6	4	9	4	7	8	9	6	7	5
14	8	8	9	1	9	1	9	8	9	1	8	8	6	6	9	7	6	7	9	9	7	5
15	8	8	1	5	7	4	8	8	8	3	7	7	5	9	9	8	6	8	8	8	8	7
16	5	9	1	8	6	4	8	5	7	5	9	3	7	7	6	6	5	5	6	9	5	5
17	8	7	3	4	8	5	6	6	9	5	9	7	7	5	9	7	8	9	8	5	7	7
18	7	7	-	-	6	3	8	2	9	3	6	4	7	3	6	5	7	7	7	2	9	5
19	8	8	6	1	9	7	9	8	8	1	9	8	5	5	7	7	7	7	9	8	9	5
20	8	8	1	7	2	3	9	9	7	3	9	4	5	7	7	7	7	8	8	8	8	5
21	8	6	9	1	9	1	7	7	7	5	9	3	5	5	6	4	8	8	7	7	9	3
22	7	7	7	2	9	3	9	5	8	2	9	3	5	5	4	1	7	6	8	2	7	7
23	8	7	5	2	7	3	9	2	6	3	9	8	3	3	-	-	8	6	5	1	8	6
24	8	2	4	6	8	8	9	5	8	6	7	7	5	5	9	6	6	7	8	6	8	4
25	6	6	2	2	6	5	9	8	8	5	9	5	7	5	9	9	5	6	2	2	1	1
26	3	6	5	6	8	5	7	5	6	1	9	4	6	6	2	2	8	8	1	1	5	5
27	9	7	7	2	9	1	9	5	7	3	9	1	6	4	7	5	8	8	7	3	7	7
28	9	7	6	7	9	5	9	8	8	3	9	3	6	4	5	5	7	6	8	7	6	5
29	8	2	9	1	6	1	7	5	9	1	9	1	7	2	5	2	8	8	1	6	6	6
30	8	3	9	1	6	1	5	5	9	1	9	1	7	2	-	-	7	7	8	1	5	4
31	9	1	6	7	5	1	5	5	6	5	9	1	6	4	-	-	7	5	9	1	7	8
32	8	6	1	-	6	5	8	7	6	5	8	3	3	3	-	-	8	6	4	7	5	7
33	7	7	1	-	5	5	8	6	7	5	8	3	3	3	-	-	7	5	4	4	6	6

Данные о продажах домов в пригородной зоне

Цена, тыс. руб.	Площадь дома, м ²	Площадь участка, га	Количество комнат	Количество важных комнат
2160	54	0,10	3	1,0
3489	95	0,09	5	1,5
4560	162	0,14	7	1,5
2415	83	0,06	5	1,0
4257	176	0,15	8	2,5
3720	160	0,04	8	1,5
3510	91	0,10	6	1,0
4977	166	0,23	8	2,0
4605	333	0,22	10	3,0
3795	98	0,05	6	1,0
3660	176	0,10	7	1,5
4200	126	0,10	6	2,0
6690	151	2,91	8	2,0
2985	90	0,10	4	1,0
6357	208	0,09	8	2,5
3657	117	0,16	6	1,0
5070	174	0,61	9	3,0
4680	270	0,10	11	2,5
3705	123	0,08	7	2,0
4080	158	0,10	7	2,0
5847	187	0,20	8	2,5
3855	121	0,19	6	2,0
9060	192	0,41	8	1,5
4260	135	0,08	7	1,0
4380	216	0,08	7	2,5
5400	204	0,08	9	2,5
3795	95	0,20	5	1,0
4185	145	0,09	8	1,5
3747	112	0,04	7	1,0
3990	180	0,10	8	2,0
3300	131	0,06	6	2,0
3555	112	0,05	7	1,0
5835	229	0,06	9	2,5
8097	225	0,19	8	3,0
5097	257	0,16	9	3,0
5700	157	0,12	7	2,5
6090	239	0,36	9	3,0

Данные о средней заработной плате сотрудников

Пол (1 – жен, 0 – муж)	Воз- раст, лет	Предыдущий стаж работы по специаль- ности, лет	Стаж работы в компании, лет	Образова- ние (лет обучения в вузе)	Заработ- ная пла- та, руб.
1	39	5	12	4	12817
0	44	12	8	6	16971
0	24	0	2	4	9785
1	25	2	1	4	9250
0	56	5	25	8	36428
1	41	9	10	4	16147
1	33	6	2	6	13402
0	37	11	6	4	14110
1	51	12	16	6	29163
0	23	0	1	4	8706
0	31	5	4	6	13985
1	27	0	8	0	5813
0	47	11	9	4	16546
1	35	5	5	6	14395
1	29	5	4	0	4790
0	46	4	15	6	18204
1	50	10	17	4	25642
0	30	3	6	4	12852
1	34	10	1	4	12394
1	42	11	8	4	16904
1	51	10	15	8	30158
0	63	16	20	4	31196
0	28	0	5	4	12300
1	32	4	1	4	11033
0	55	11	16	6	29956
1	45	20	2	4	17086
0	34	2	12	2	6369
0	33	2	7	4	13075
1	23	0	1	4	9581
0	40	4	13	6	18322
1	48	6	15	4	17796
1	27	0	6	0	6005
1	36	5	5	6	13068
0	58	9	22	4	29588
0	31	1	1	6	11943

Цены на однокомнатные квартиры в районах города

№ п/п	Район	Удаленность от останов-ки, мин	Общая площадь, м ²	Кухня, м ²	Жилая площадь, м ²	Этаж (0 – первый или последний, 1 – нет)	Дом (1 – кирпичный, 0 – панельный)	Цена, млн руб.
1	1	5	34,00	7,50	19,00	1	1	2,24
2	2	4	36,00	10,00	20,00	0	0	1,44
3	2	6	45,00	13,00	20,00	1	1	2,45
4	3	5	35,30	10,00	20,00	1	0	1,43
5	2	3	33,00	5,50	22,00	1	0	1,32
6	1	3	33,00	8,50	18,00	1	1	2,37
7	1	5	37,00	10,00	19,00	1	0	1,72
8	3	2	38,00	8,50	19,10	1	0	1,66
9	1	8	54,00	9,20	27,20	1	1	2,81
10	2	8	35,00	6,00	20,00	0	1	1,77
11	1	4	38,00	8,00	19,00	0	0	1,38
12	2	5	31,60	8,80	14,00	0	0	1,29
13	1	3	32,00	8,00	17,00	1	0	1,58
14	1	5	37,00	10,00	19,00	1	0	1,74
15	2	3	32,00	8,00	17,00	1	1	1,56
16	3	5	37,00	8,00	19,00	1	1	2,02
17	3	10	32,20	6,50	17,00	0	1	1,29
18	3	3	33,00	8,00	19,00	1	0	1,28
19	3	3	37,50	9,60	19,80	1	0	1,46
20	3	5	33,00	7,00	19,00	1	0	1,39
21	2	8	32,00	6,00	21,50	1	0	1,49
22	1	8	36,00	10,00	19,00	0	0	1,39
23	2	2	31,60	6,00	21,60	1	1	1,48
24	2	3	52,00	12,00	34,00	1	1	3,02

Динамика индекса Доу-Джонса [54]

Дата	Значение	Дата	Значение	Дата	Значение
Янв. 1988	1947,35	Июнь 1989	2494,9	Нояб. 1990	2518,56
Фев. 1988	1980,65	Июль 1989	2554,03	Дек. 1990	2610,92
Март 1988	2044,31	Авг. 1989	2691,11	Янв. 1991	2587,6
Апр. 1988	2036,13	Сен. 1989	2693,41	Фев. 1991	2863,04
Май 1988	1988,91	Окт. 1989	2692,01	Март 1991	2920,11
Июнь 1988	2104,94	Нояб. 1989	2642,49	Апр. 1991	2925,54
Июль 1988	2104,22	Дек. 1989	2728,47	Май 1991	2928,42
Авг. 1988	2051,29	Янв. 1990	2679,24	Июнь 1991	2968,14
Сен. 1988	2080,06	Фев. 1990	2614,18	Июль 1991	2978,19
Окт. 1988	2144,31	Март 1990	2700,13	Авг. 1991	3006,09
Нояб. 1988	2099,04	Апр. 1990	2708,26	Сен. 1991	3010,35
Дек. 1988	2148,58	Май 1990	2793,81	Окт. 1991	3019,74
Янв. 1989	2234,68	Июнь 1990	2894,82	Нояб. 1991	2986,12
Фев. 1989	2304,3	Июль 1990	2934,23	Дек. 1991	2958,64
Март 1989	2283,11	Авг. 1990	2681,89	Янв. 1992	3227,06
Апр. 1989	2348,91	Сен. 1990	2550,69	Фев. 1992	3257,27
Май 1989	2439,55	Окт. 1990	2460,54	Март 1992	3247,42

Таблица П.15

Данные о ежеквартальных доходах ведущих американских компаний за период 1991–1995 гг. (млн долл. США) [54]

McDonalds	GM	Microsoft	AT&T	Nike	GE	CocaCola	Ford
1558	29200	369	15270	947	13330	2480	21340
1671	31300	461	15720	734	14770	3040	23850
1737	28900	487	15630	867	14580	3170	21120
1730	33600	527	16460	848	17550	2880	21980
1618	32000	581	15380	1098	12430	2770	24560
1774	35200	682	15850	876	14200	3550	26840
1913	29400	681	16180	972	14210	3510	23330
1828	35800	815	17500	983,3	16230	3240	25410
1654	35000	818	15719	1108	12900	3060	26760
1878	36658	938	16316	805,8	14761	3899	29420
1944	30138	958	18459	871,8	14858	3629	24500
1932	37268	1039	17097	1004	18087	3373	27840
1796	37495	983	18238	1170	12657	3352	30400
2029	40392	1129	18649	1054	14768	4342	33770
2225	34510	1244	21110	1125	14481	4461	30660
2270	42553	1293	18262	1412	17791	4017	33643
2161	43285	1247	19512	1615	15126	3854	34783
2468	44146	1482	19704	1443	17809	4936	36389
2580	37463	1587	22131	1492	17341	4895	31418
2586	43935	1621	79609	1921	19752	4333	34547

Учебное издание

Дубина Игорь Николаевич

**МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ
В ЭМПИРИЧЕСКИХ СОЦИАЛЬНО-
ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ**

Заведующая редакцией *Л.А. Табакова*
Редактор *Н.А. Кузнецова*
Художественный редактор *Ю.И. Артюхов*
Технический редактор *В.Ю. Фотиева*
Корректоры *Т.М. Васильева, Н.Б. Вторушина*
Компьютерная верстка *И.В. Витте*
Оформление художника *А.Л. Чирикова*

ИБ № 5004

Подписано в печать 28.05.2009. Формат 60 × 90/16.
Гарнитура «Таймс». Печать офсетная.
Усл. п. л. 26,0. Уч. изд. л. 24,5.
Тираж 1500 экз. Заказ № 2458 «С» 039.

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

Издательский Дом «ИНФРА-М»
127282, Москва, ул. Полярная, д. 31в
Тел.: (495) 380-05-40, 380-05-43. Факс (495) 363-92-12
E-mail: books@infra-m.ru <http://www.infra-m.ru>

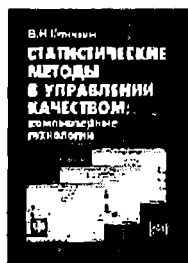
Отпечатано с готовых диапозитивов в ОАО ордена «Знак Почета»
«Смоленская областная типография им. В. И. Смирнова».
214000, г. Смоленск, проспект им. Ю. Гагарина, 2.

Издательство
“ФИНАНСЫ И СТАТИСТИКА”
предлагает учебное пособие

В.Н.Клячкин

**СТАТИСТИЧЕСКИЕ МЕТОДЫ
В УПРАВЛЕНИИ КАЧЕСТВОМ:
КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ**

304 с.



Рассматриваются статистические методы, используемые в задачах управления качеством продукции, компьютерные технологии и их реализация в среде электронных таблиц Excel и системе STATISTICA: излагаются семь простых инструментов текущего контроля качества, статистическое управление процессами, приемочный контроль продукции, статистические методы анализа качества. Включены лабораторные работы.

Для студентов вузов, обучающихся по специальностям "Управление качеством" и "Прикладная математика". Для повышения квалификации работников инженерного и экономического профиля, занимающихся вопросами обеспечения качества. Представляет интерес для специалистов, разрабатывающих и использующих статистические методы контроля и анализа качества продукции.

По вопросам приобретения литературы
обращайтесь в Издательство по адресу:

101000, Москва, ул.Покровка, 7

(метро "Китай-город", выход на ул.Маросейка)

Тел.: (495) 625-35-02, 623-80-42. Факс (495) 625-09-57

E-mail: mail@finstat.ru <http://www.finstat.ru>

При Издательстве работает киоск:

понедельник – четверг с 9.00 до 19.30,

пятница с 9.00 до 18.30

Тел.(495) 621-86-57

Система "Книга-почтой"

Стоимость пересылки почтовыми бандеролями –
30% от стоимости заказа

МАТЕМАТИКО- СТАТИСТИЧЕСКИЕ МЕТОДЫ

В ЭМПИРИЧЕСКИХ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ
ИССЛЕДОВАНИЯХ

В учебном пособии:

- даны основы теории социально-экономических измерений, подходы к проектированию измерительных шкал и оценке качества измерений;
- описаны методы проверки исследовательских гипотез и построения статистических прогнозов;
- рассмотрены возможности математико-статистических пакетов SPSS, STATISTICA и WINSTEPS для обработки и анализа результатов количественных исследований в экономических и социальных областях.

Дубина И. И. Математико-статистические
методы. СПб.



9 785279 031078

Цена: 509 руб.

КТК: 0910

