

Андрей Наследов

IBM

SPSS 20

STATISTICS

и AMOS

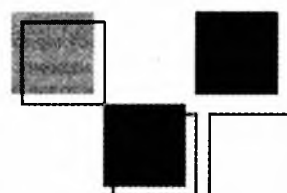


профессиональный
статистический
анализ данных



Андрей Наследов

IBM
SPSS 20
STATISTICS
и AMOS



профессиональный
статистический
анализ данных



Москва • Санкт-Петербург • Нижний Новгород • Воронеж
Ростов-на-Дону • Екатеринбург • Самара • Новосибирск
Киев • Харьков • Минск
2013

ББК 32.973.23-018.2
УДК 004.31
НЗ1

Наследов А.
НЗ1 IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. — СПб.: Питер, 2013. — 416 с.: ил.

ISBN 978-5-496-00107-6

Книга представляет собой практическое руководство по анализу данных с помощью самой мощной и популярной программы статистической обработки информации — IBM SPSS Statistics 20. В издании подробно описываются основы работы с пакетом SPSS, рассматривается большинство методов обработки и анализа данных, а также способов табличного и графического представления полученных результатов. Кроме того, книга содержит подробную информацию о программе IBM SPSS AMOS, позволяющей использовать в исследованиях популярную и эффективную методологию моделирования структурными уравнениями (SEM — structural equation modeling).

Материал книги организован таким образом, чтобы удовлетворить запросы как новичка, впервые приступающего к анализу данных на компьютере, так и опытного исследователя, желающего воспользоваться самыми современными методами. Основное содержание глав составляют пошаговые инструкции по реализации различных видов математико-статистического анализа в SPSS. Особое внимание уделяется полученным результатам и их интерпретации. В конце книги приведен глоссарий, содержащий определения большинства статистических терминов.

ББК 32.973.23-018.2
УДК 004.31

Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

Информация, содержащаяся в данной книге, получена из источников, рассматриваемых издательством как надежные. Тем не менее, имея в виду возможные человеческие или технические ошибки, издательство не может гарантировать абсолютную точность и полноту приводимых сведений и не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-5-496-00107-6

© ООО Издательство «Питер». 2013

Глава 5. Диаграммы.....	83
Графика в программе SPSS.....	83
Настройка диаграмм.....	84
Команды построения диаграмм.....	85
Редактирование диаграмм.....	94
Выход из программы.....	95
Глава 6. Частоты.....	96
Пошаговые алгоритмы вычислений.....	97
Представление результатов.....	102
Завершение анализа и выход из программы.....	105
Глава 7. Описательные статистики	106
Пошаговый алгоритм вычислений.....	108
Представление результатов.....	111
Завершение анализа и выход из программы.....	112
Глава 8. Таблицы сопряженности и критерий хи-квадрат	113
Таблицы сопряженности.....	113
Критерий независимости хи-квадрат.....	114
Пошаговый алгоритм вычислений.....	115
Представление результатов.....	121
Терминология, используемая при выводе.....	123
Завершение анализа и выход из программы.....	124
Глава 9. Корреляции	125
Понятие корреляции.....	125
Дополнительные сведения.....	126
Пошаговые алгоритмы вычислений.....	128
Представление результатов.....	132
Завершение анализа и выход из программы.....	133
Глава 10. Средние значения.....	134
Пошаговый алгоритм вычислений.....	134
Представление результатов.....	138
Глава 11. Сравнение двух средних и t-критерий	140
Уровень значимости.....	141
Пошаговые алгоритмы вычислений.....	141
Представление результатов.....	146
Завершение анализа и выход из программы.....	148
Глава 12. Непараметрические критерии.....	149
Параметрические и непараметрические критерии.....	149
Пошаговые алгоритмы и результаты вычислений.....	150
Завершение анализа и выход из программы.....	166

Глава 13. Однофакторный дисперсионный анализ.....	168
Пошаговые алгоритмы вычислений	169
Представление результатов	174
Терминология	177
Завершение анализа и выход из программы.....	177
Глава 14. Многофакторный дисперсионный анализ.....	179
Файлы данных для группы методов Общая линейная модель.....	180
Дисперсионный анализ с двумя факторами	181
Дисперсионный анализ с тремя и более факторами	182
Влияние ковариат	183
Пошаговые алгоритмы вычислений	183
Представление результатов	188
Терминология, используемая при выводе	192
Завершение анализа и выход из программы.....	192
Глава 15. Многомерный дисперсионный анализ.....	194
Пошаговые алгоритмы вычислений	195
Представление результатов	202
Завершение анализа и выход из программы.....	207
Глава 16. Дисперсионный анализ с повторными измерениями	208
Пошаговые алгоритмы вычислений	210
Представление результатов	216
Завершение анализа и выход из программы.....	220
Глава 17. Простая линейная регрессия	221
Простая регрессия	221
Оценка криволинейности.....	223
Пошаговые алгоритмы вычислений	225
Представление результатов	228
Терминология, используемая при выводе	230
Завершение анализа и выход из программы.....	231
Глава 18. Множественный регрессионный анализ.....	232
Уравнение множественной регрессии	232
Коэффициенты регрессии	233
Коэффициент детерминации и пошаговые методы.....	234
Условия получения приемлемых результатов анализа	235
Пошаговые алгоритмы вычислений	236
Представление результатов	242
Терминология, используемая при выводе	244
Завершение анализа и выход из программы.....	244
Глава 19. Анализ надежности	246
Коэффициент альфа.....	246
Надежность половинного расщепления.....	247

Пошаговые алгоритмы вычислений	247
Представление результатов	252
Завершение анализа и выход из программы	256
Глава 20. Факторный анализ	257
Вычисление корреляционной матрицы	257
Извлечение факторов	258
Выбор и вращение факторов	259
Интерпретация факторов	261
Пошаговые алгоритмы вычислений	261
Представление результатов	267
Терминология, используемая при выводе	271
Завершение анализа и выход из программы	272
Глава 21. Кластерный анализ	273
Сравнение кластерного и факторного анализа	273
Этапы кластерного анализа	275
Кластерный анализ матрицы различий (сходства)	276
Пошаговые алгоритмы вычислений	278
Представление результатов	285
Завершение анализа и выход из программы	288
Глава 22. Дискриминантный анализ	289
Этапы дискриминантного анализа	290
Пошаговые алгоритмы вычислений	292
Представление результатов	297
Терминология, используемая при выводе	301
Завершение анализа и выход из программы	303
Глава 23. Многомерное шкалирование	304
Квадратная асимметричная матрица различий	305
Квадратная симметричная матрица различий	306
Модель индивидуальных различий	306
Пошаговые алгоритмы вычислений	307
Представление результатов	313
Завершение анализа и выход из программы	317
Глава 24. Логистическая регрессия	318
Математическое описание логистической регрессии	319
Пошаговые алгоритмы вычислений	320
Представление результатов	323
Терминология, используемая при выводе	327
Завершение анализа и выход из программы	328
Глава 25. Логлинейный анализ таблиц сопряженности	329
Понятие логлинейной модели	329

Логлинейный метод подбора модели.....	330
Пошаговые алгоритмы вычислений	331
Представление результатов.....	335
Завершение анализа и выход из программы.....	339
Глава 26. AMOS: Моделирование структурными уравнениями	340
SEM и другие методы	341
Особенности и этапы SEM.....	343
Формирование модели и ее идентификация.....	345
Оценка, проверка согласованности и коррекция модели	348
Требования к данным	353
Пошаговые алгоритмы вычислений	356
Представление результатов.....	379
Терминология, используемая при выводе	384
Завершение анализа и выход из программы.....	386
Глоссарий	387
Англо-русский словарь терминов.....	400
Рекомендуемая литература	412

Предисловие

Эта книга продолжает серию учебных пособий и руководств по применению компьютерной программы SPSS. Первое издание этой серии, «SPSS: Компьютерный анализ данных в психологии и социальных науках», было опубликовано в 2005 году, предыдущее, «SPSS 19: Профессиональный статистический анализ данных», — в 2011 году. Изменения в содержании книги от издания к изданию на 90 % касались развития интерфейса программы SPSS. Принципиальным отличием этой книги от ее предшественниц является добавление объемной главы о программе SPSS AMOS, которая может применяться не только как один из модулей SPSS, но и автономно. AMOS — это весьма совершенный и удобный инструмент моделирования структурными уравнениями (SEM — structural equation modeling). SEM является результатом развития и объединения совокупности методов корреляционного, регрессионного, факторного и дисперсионного анализа. Но при этом SEM выходит за рамки конкретных методов анализа данных, представляя собой инструмент моделирования как дедуктивного подхода эмпирической верификации теоретических концепций. Широта применения SEM с легкостью варьируется от эмпирической проверки теоретических гипотез до data mining — поиска скрытых, заранее не известных закономерностей в имеющихся данных. Благодаря этим особенностям SEM получило широкое распространение за рубежом: англоязычные публикации по психологии, например, пестрят характерными для SEM блок-схемами, до 50 % публикаций ссылаются на применение SEM. В отечественных исследованиях эта методология пока редко применяется — во многом из-за отсутствия русскоязычных учебных пособий и руководств.

SPSS (Statistical Package for the Social Science — статистический пакет для социальных наук) — очень мощный и широко распространенный инструмент профессионального компьютерного анализа данных. Как следует из названия, SPSS представляет собой набор различных программ обработки данных. Эти программы облегчают процесс ввода информации, позволяют гибко менять структуру данных, использовать самые современные методы обработки (как по отдельности, так и совместно) и получать результаты в удобной и наглядной форме. Все это множество программ собрано в единую систему, обеспечивающую простую и дружелюбную диалоговую систему с исследователем и снабженную исчерпывающей справочной поддержкой. Благодаря такой дружелюбности систему SPSS легко освоить даже тому, кто имеет минимальные навыки владения компьютером.

Но если SPSS настолько дружелюбная, да еще и обладающая справочной поддержкой система, зачем нужна эта книга? Как показывает практика, оставшись один на один со справочной системой SPSS, даже опытный исследователь, не говоря о студентах, впадает в отчаяние, безуспешно пытаясь найти в мегабайтах документации нужную информацию. Усугубляет ситуацию не всегда достаточное знание пользователем специфической статистической и математической терминологии. Но даже те, кто применяет

SPSS для анализа данных, испытывают трудности при необходимости реорганизации исходных данных и редактирования результатов обработки (таблиц и графиков). Стремление помочь читателю избежать подобных трудностей и стало стимулом к написанию этой книги.

В книге содержится подробное описание большинства процедур SPSS, предназначенных для анализа данных, — от простейших до самых сложных, включая различные нюансы их применения. При описании любой процедуры основными подходами являлись максимально доступное изложение ее назначения, сведение действий по ее применению к четким пошаговым инструкциям и демонстрация ее применения на простейших примерах. Однако было бы неверно думать, что эта книга предназначена только для тех, кто хочет освоить азы компьютерного анализа данных. Она, без сомнения, будет полезна и как справочник для тех, кто уже имеет опыт работы с программой SPSS. Использовать книгу в качестве справочника помогает то, что в конце помещены глоссарий и англо-русский словарь терминов.

Содержание первых 12 глав адресовано широкой аудитории и не требует специальных знаний в области статистики, оставшаяся часть книги рассчитана на людей, имеющих такие знания. В начале каждой из глав обычно кратко излагается сущность статистической процедуры, которой она посвящена. При этом делается все возможное для того, чтобы исключить избыточные детали и оставить лишь достаточную для анализа данных информацию. Говорить простым языком о сложных вещах — весьма нелегкое занятие, и пришлось приложить немало усилий, чтобы написать эту книгу. Насколько успешными были эти усилия, судить вам, читатель.

Программа SPSS снабжена подробной справочной системой. На нескольких тысячах страниц можно найти практически любую информацию по работе с SPSS, которая обычно необходима даже опытному пользователю. Разумеется, такой объем информации нелегко было разместить в нескольких сотнях страниц данного издания. Однако подавляющее большинство методов нашло здесь отражение, хоть временами и поверхностное, но, хочется верить, достаточное для практического применения.

Цивилизованное использование SPSS, в том числе в учебных целях, невозможно без приобретения официальной лицензии. Собственником программного обеспечения SPSS с 2010 года является SPSS Inc., а IBM Company © со штаб-квартирой в США. В российском официальном представительстве этой компании можно купить лицензионные версии SPSS. Всю необходимую информацию вы найдете по адресу <http://www.spss.ru/price/index.htm>. Кроме того, можете воспользоваться бесплатной пробной версией SPSS, срок действия которой 14 дней. Скачать пробную 14-дневную версию SPSS последней модификации можно по адресу <http://www.spss.ru/products/spss/index.htm>.

От издательства

Ваши замечания, предложения, вопросы отправляйте по адресу электронной почты comp@piter.com (издательство «Питер», компьютерная редакция).

Мы будем рады узнать ваше мнение!

Все файлы данных, упомянутые в книге, вы сможете найти по адресу <http://www.piter.com/download>.

Подробную информацию о наших книгах вы найдете на веб-сайте издательства <http://www.piter.com>.

Введение

Анализ данных на компьютере

Необходимые знания

Версии и модули SPSS

Содержание книги

Файлы примеров

Структура глав и элементы описания

Основное содержание книги составляют пошаговые инструкции по применению программ SPSS Statistics и SPSS AMOS для самого разностороннего анализа данных. Она организована так, чтобы обеспечить как новичка, так и опытного пользователя исчерпывающей информацией для самостоятельной обработки данных.

В данной главе представлено краткое содержание книги, описана ее структура, приведены некоторые рекомендации по ее изучению. Если вы не знакомы с программами SPSS, рекомендуется прочитать эту главу полностью, поскольку это поможет лучше ориентироваться в материале книги.

Если вы уже имеете опыт обработки данных при помощи программ SPSS и вас интересует лишь описание конкретной статистической процедуры, достаточно найти это описание в одной из глав 6–26. Эти главы не зависят друг от друга: открыв любую из них и следуя инструкциям, вы получите результат и сможете в нем разобраться. Однако рекомендуем даже опытному пользователю все же обратить внимание на главы 2–5, в которых содержатся полезные рекомендации по таким важным темам, как управление исходными данными, использование командного языка Syntax, редактирование отчетов и таблиц результатов, создание и модификация графиков и диаграмм.

Анализ данных на компьютере

Анализ данных с применением компьютера включает в себя выполнение ряда шагов:

1. Определение структуры исходных данных.
2. Ввод данных в компьютер в соответствии с их структурой и требованиями программы. Редактирование и преобразование данных.
3. Задание метода обработки данных в соответствии с задачами исследования.
4. Получение результата обработки данных. Его редактирование и сохранение в нужном формате.
5. Интерпретация результата обработки.

Шаги 1 (подготовительный) и 5 (заключительный) не способна выполнить ни одна компьютерная программа — исследователь делает их сам. Помощь компьютера (шаги 2–4) заключается в преобразовании длинной последовательности чисел в более компактную. На вход компьютера исследователь подает массив исходных данных, который недо-

ступен для осмысления, но пригоден для компьютерной обработки (шаг 2). Затем исследователь дает программе команду на обработку данных в соответствии с поставленной задачей и структурой данных (шаг 3). На выходе он получает результат обработки (шаг 4) — тоже массив данных, только меньший, доступный для осмысления и содержательной интерпретации. При этом исчерпывающий анализ данных обычно требует многократной их обработки с применением разных методов.

В этой книге очень подробно и последовательно рассматривается все, что связано с компьютерной обработкой данных (шаги 2–4), кратко излагаются математические основы каждого метода, а интерпретация результатов иллюстрируется конкретными примерами.

Необходимые знания

Эта книга будет для вас действительно эффективным руководством по компьютерной обработке данных, если вы имеете хотя бы самые общие знания о статистике и обладаете элементарными навыками работы с компьютером.

Статистика

В начале каждой из глав приводятся основные сведения, касающиеся той или иной статистической процедуры. Кроме того, в конце книги приведен глоссарий, содержащий определения большинства статистических терминов. И то и другое представляет собой информацию справочного характера и не рассчитано на новичков. Несмотря на то что практически любой человек с помощью этой книги может работать с пакетом SPSS и получать результаты, отсутствие базовых знаний по статистике не позволит самостоятельно решать даже простые задачи. Поэтому вы должны быть знакомы с основами статистики или, как минимум, изучать их.

От главы к главе требования к уровню знаний растут. А при описании сложных методов анализа данных (главы 13–26) даже подготовленному читателю, вероятно, потребуется обращение к специальной литературе¹.

Компьютер

Минимальные требования к навыкам работы на компьютере весьма скромные: вам необходимо лишь уметь включать компьютер, а также пользоваться клавиатурой и мышью. Существенно облегчат освоение SPSS навыки владения программами Microsoft Office и особенно Excel (в основе как Excel, так и SPSS лежат электронные таблицы).

На вашем компьютере должны быть установлены операционная система Windows и программа SPSS (желательно версии 12.0 и выше), а значок программы следует поместить на Рабочий стол.

Версии и модули SPSS

По всем параметрам SPSS является сложным и мощным статистическим пакетом. Но несмотря на сложность средства взаимодействия входящих в пакет программ с пользо-

¹ Математико-статистические основы большинства описываемых методов читатель может найти в книге: *Наследов А. Д. Математические методы психологического исследования. Анализ и интерпретация данных.* СПб.: Речь, 2008.

вателем весьма дружелюбны. С помощью пакета SPSS можно проводить практически любой анализ данных, а последние версии программы находят применение в самых разных научных областях и в мире бизнеса.

Основой для данной книги послужили русифицированная версия 20 IBM SPSS Statistics и англоязычная версия 20 IBM SPSS AMOS, распространяемые с 2011 года. Говоря точнее, снимки экрана, помещенные в книгу, соответствуют указанным версиям. Однако почти весь изложенный материал может быть с успехом применен и к более ранним версиям, начиная с SPSS 9.0 и AMOS 3. Основные различия будут наблюдаться в интерфейсе программ: названиях диалоговых окон, их виде и т. п. Кроме того, начиная с версии SPSS 13.0 более совершенными стали графические возможности программы. Наиболее значительное для отечественного пользователя программы новшество введено в SPSS 12.0 — стала доступной русифицированная версия интерфейса и окон вывода результатов.

Если вы пользуетесь нерусифицированными версиями программ, ориентироваться в англоязычной терминологии вам поможет англо-русский словарь терминов, приведенный в конце этой книги.

За последние два года семейство программ SPSS дважды меняло название. В 2009 году привычная аббревиатура *SPSS* сменилась на *PASW*, а в 2010 году традиционное название вернулось в виде *IBM SPSS*. Заметно расширилось и семейство программных модулей — до 22 программ, одна из которых (*IBM SPSS Statistics Base*) является основной, а остальные могут быть добавлены для расширения различных возможностей SPSS.

В книге речь пойдет преимущественно о статистических процедурах, реализованных в трех основных модулях SPSS Statistics и наиболее часто использующихся в социальных исследованиях. Мы опишем операции, относящиеся к основному системному модулю (*SPSS Base*), модулю дополнительных моделей (*SPSS Advanced Statistics*) и модулю регрессионных моделей (*SPSS Regression*). Отдельная, самая объемная глава посвящена программе *IBM SPSS AMOS*. Семейство программ SPSS сопровождается справочной системой, содержащей информацию обо всех процедурах, поддерживаемых этими программами. Однако для начинающего или недостаточно опытного пользователя справочная система слишком сложна, а ее объем, составляющий тысячи печатных страниц, затрудняет поиск нужной информации. В определенной степени это и послужило поводом к написанию книги, в которой в компактной форме представлено практически все, что может оказаться полезным на практике.

Содержание книги

Книга содержит описания большинства методов обработки данных. Решение задачи анализа данных, как отмечалось, происходит в три основных этапа:

1. Ввод данных в компьютер и их структурирование для дальнейшей обработки программой.
2. Определение типа анализа, необходимого для решения задачи.
3. Получение и расшифровка результатов обработки исходных данных.

В главе 2 речь пойдет об основах работы с пакетом SPSS: типах окон, использовании панелей инструментов и меню, просмотре, редактировании, сохранении и печати результатов обработки и т. д. Несмотря на то что глава в основном рассчитана на начинающих, в ней содержится немало информации, полезной и более опытным пользователям. Глава 3 посвящена первому из трех перечисленных этапов и включает в себя темы создания,

редактирования и форматирования файлов данных. Редактор данных SPSS является инструментом, позволяющим без труда создавать файлы данных и управлять их структурой. В главах 4 и 5 рассматриваются такие важные темы, как модификация и преобразование данных, а также создание диаграмм. В главе 4 описываются различные действия с данными: создание новых переменных, сортировка, изменение структуры, объединение файлов, выбор подмножества данных для анализа. В главе 5 рассмотрены основные процедуры работы с диаграммами. Заметим, однако, что большая часть материала, касающегося диаграмм, излагается в других главах книги по мере необходимости.

Все главы, начиная с 6-й, посвящены двум оставшимся этапам решения задачи анализа данных. В них речь идет о различных методах обработки данных и о том, каким образом интерпретировать их результаты. Каждая из глав 6–26 не зависит от других — это означает, что если вы захотите выполнить какой-либо вариант анализа данных, даже не обладая начальными знаниями, то сможете, открыв соответствующую главу и следуя представленным инструкциям, получить результат. При необходимости в инструкциях приводятся ссылки на другие главы.

Как было отмечено ранее, книга охватывает основное содержимое трех модулей SPSS Statistics (Base, Advanced Statistics и Regression Models) и модуля SPSS AMOS. Поскольку на практике иногда производится неполная установка пакета SPSS, некоторые из модулей могут оказаться недоступными (кроме Base, который устанавливается всегда). Для выполнения подавляющего большинства методов анализа достаточно модуля Base. Модули Advanced Models и Regression Models позволяют дополнительно выполнять следующие методы анализа:

- ▶ одномерный (многофакторный) и многомерный дисперсионный анализ (MANOVA) и многомерный ковариационный анализ (MANCOVA) — команды ОЛМ-одномерная и ОЛМ-многомерная группы Общая линейная модель;
- ▶ многомерный дисперсионный анализ с повторными измерениями (repeated-measures MANOVA) — команда ОЛМ-повторные измерения группы Общая линейная модель;
- ▶ логистическую регрессию;
- ▶ логлинейный анализ.

Программа SPSS AMOS может работать и без предварительной установки SPSS Statistics с файлами данных в других форматах, например Excel. Но при установке поверх SPSS Statistics AMOS встраивается в диалог SPSS наряду с другими модулями пакета. Будучи встроенным в состав IBM SPSS, модуль AMOS обеспечивает легкий доступ к широким возможностям современной методологии моделирования структурными уравнениями (SEM — structural equation modeling), выходящей далеко за привычные рамки статистического анализа данных.

Объем этой книги не позволяет поместить описание всех многочисленных возможностей SPSS Statistics и AMOS. Поэтому детально рассмотрены лишь наиболее востребованные методы и модели.

Файлы примеров

В большинстве глав книги при помощи разных методов анализируется один и тот же файл данных. Для последних глав были разработаны отдельные примеры, полнее демонстрирующие применение специфических возможностей пакета SPSS. Все файлы доступны на сайте издательства (<http://www.piter.com/download>). Наличие готовых ис-

ходных файлов (включая первый) избавит вас от необходимости набирать данные для обучения вручную. К тому же, выполняя различные процедуры, вы сможете сравнивать получаемые результаты с эталонными, приведенными в этой книге.

Первый файл (ex01.sav) создан таким образом, чтобы с его помощью можно было без труда продемонстрировать функционирование основных методов обработки. Он содержит информацию о 100 школьниках трех выпускных классов. О каждом из школьников имеются следующие данные:

- ▶ № (идентификационный номер);
- ▶ пол;
- ▶ класс;
- ▶ вуз (выбранный профиль вуза);
- ▶ хобби (систематическое внешкольное занятие);
- ▶ тест1;
- ▶ тест2;
- ▶ тест3;
- ▶ тест4;
- ▶ тест5;
- ▶ средняя годовая отметка за 10-й класс;
- ▶ средняя годовая отметка за 11-й класс.

Файл ex01.sav нам потребуется, начиная уже с вводных глав 2–5. Его содержимое частично приведено в конце главы 3. При желании вы можете вручную создать собственный файл и успешно использовать его в процессе изучения почти всей первой половины книги. Разумеется, все данные уже содержатся в готовом файле на сайте, однако целесообразно самостоятельно попрактиковаться во вводе, форматировании и других простых, но необходимых действиях с данными. Еще более полезным занятием была бы работа с собственными файлами данных.

Всего с упомянутого сайта издательства вы можете загрузить 21 файл примеров. Из них 18 файлов содержат данные, один файл (cross.spo) предназначен для демонстрации результатов обработки (окна вывода), один файл (Syn_clust.sps) содержит пример применения командного языка Syntax, и два файла содержат модели AMOS (КФА.amw, CM.amw). В табл. 1.1 приведен список файлов примеров с указанием глав книги, в которых они используются. В указанных главах вы можете найти либо подробные описания соответствующих примеров, либо ссылки на места в книге, содержащие эти описания.

Таблица 1.1. Файлы примеров и номера глав, в которых они используются

Файл примера	Глава книги
ex01.sav	2–13, 26
cross.spo	2
ex01a.sav, ex01b.sav	4
ex020.sav, ex021.sav, ex022.sav	14–16

продолжение ➤

Таблица 1.1 (продолжение)

Файл примера	Глава книги
exam.sav	17, 26
help.sav	18
TestA.sav	19
IQ.sav	20, 21, 26
mds1.sav, mds2.sav, mds3.sav	23
cars.sav	21
Syn_clust.sps	21
class.sav	22
helpLR.sav	24
helpLIM.sav	25
KFA.amw, CM.amw	26

В каждой из глав 5–26 имеются пошаговые процедуры, которые обязательно включают в себя открытие конкретного файла данных. При этом предполагается, что эти файлы данных сохранены в отдельной папке на жестком диске вашего компьютера.

Структура глав и элементы описания

Главы 2–5, являющиеся вводными, не имеют единой структуры, поскольку едва ли возможно систематизировать начальную информацию о работе с программой. Материал остальных глав излагается в соответствии с общей структурой книги, которая выглядит следующим образом:

- ▶ Вводный раздел содержит общее описание рассматриваемой процедуры. В зависимости от сложности анализа описание занимает от 1 до 20 страниц.
- ▶ Далее идет раздел пошаговых процедур, требующихся для выполнения одной или нескольких разновидностей статистического анализа.
- ▶ Раздел результатов анализа (как правило, в сокращенном виде). Здесь также приводится необходимая терминология, позволяющая понять смысл результатов.
- ▶ Раздел завершения анализа и выхода из программы посвящен, очевидно, завершающим операциям редактирования результатов, их сохранения и выхода из программы.

Снимки экрана

Каждая глава содержит снимки экрана компьютера с фрагментами выводимых данных, элементами окон или целыми окнами, появляющимися на экране компьютера в процессе работы. На рис. 1.1 в качестве примера показано диалоговое окно Частоты, полученное в ходе выполнения шага 4 пошаговой процедуры из главы 6.

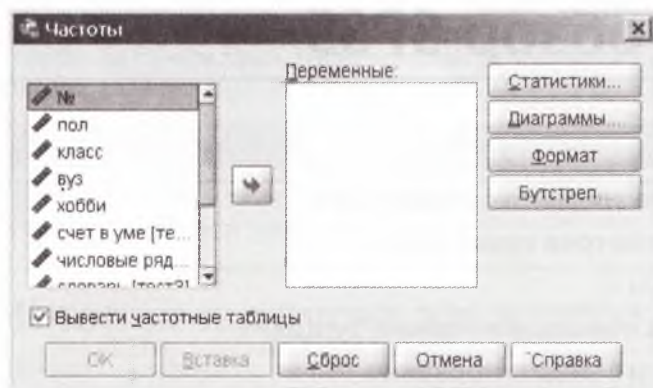


Рис. 1.1. Диалоговое окно Частоты

Иногда вы можете встретить рисунок, на котором будет представлено не целое окно программы, а его фрагмент, например такой, как показанная ниже строка меню. Подобный рисунок призван напомнить вам, как выглядит упомянутый в текущем абзаце элемент интерфейса или описываемый фрагмент окна.



Помимо окон и элементов программного интерфейса в разделах «Представление результатов» приведены большие фрагменты данных, генерируемые программой. Все они будут сопровождаться подробным описанием.

Пошаговые процедуры

Основные статистические операции оформлены в виде пошаговых процедур, чтобы вы точно знали, что нужно сделать для решения поставленной задачи. Например, далее приведен шаг 3 пошаговой процедуры из главы 4.

ШАГ 3

Откройте файл данных, с которым вы намерены работать (в нашем случае это файл ex01.sav). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню **Файл** команду **Открыть ▸ Данные** или щелкните на кнопке **Открыть данные** панели инструментов.
2. В открывшемся диалоговом окне дважды щелкните на имени ex01.sav или введите его с клавиатуры и щелкните на кнопке **ОК**.

Как можно видеть по приведенному фрагменту пошаговой процедуры, в книге используются некоторые элементы шрифтового выделения. Например, запись **Открыть ▸ Данные** означает, что в строке меню нужно щелкнуть мышью на пункте **Файл**, в открывшемся меню раскрыть (щелчком мыши) подменю **Открыть** и уже в этом подменю выбрать (снова щелчком мыши) команду **Данные**. Как вы, наверно, успели заметить, особым шрифтом набраны имена элементов интерфейса, файлов и переменных.

2 Общий обзор SPSS

20	Запуск программы
21	Кнопки и другие элементы управления
21	Настройка параметров программы
24	Окна программы
28	Окно редактора командного языка Syntax
30	Окно вывода и его редактирование
35	Сохранение, экспорт, перенос и печать результатов

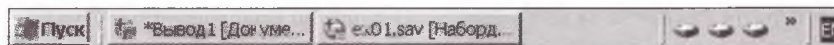
Как упоминалось в главе 1, необходимые начальные знания пользователя ограничиваются умением включать компьютер и представлением о Рабочем столе Windows. В этой главе описаны все действия, позволяющие работать с программой SPSS. Вы узнаете, как ее запускать и настраивать, как переключаться между программами с помощью Панели задач и т. д. В результате у вас появятся начальные навыки работы с программой SPSS. Если вы свободно владеете компьютером и уже знакомы с SPSS, часть материала этой главы можно просмотреть по диагонали. Тем не менее каждому читателю и даже искушенному пользователю SPSS следует обратить внимание на разделы «Настройка параметров программы», «Окно вывода и его редактирование» и «Сохранение, экспорт, перенос и печать результатов», поскольку в них содержится важная информация, касающаяся специфики SPSS 20. Особое значение при работе с программой имеют ее настройки, а также возможности редактирования и сохранения окна вывода, так как это относится к результатам анализа данных. Поэтому по мере освоения SPSS целесообразно возвращаться к соответствующим разделам этой главы.

Запуск программы

Пожалуй, основное, что вам необходимо (по крайней мере, в рамках этой книги), — знать, как запустить программу SPSS. Для этого на большинстве компьютеров вам следует, щелкнув на кнопке Пуск (Start), последовательно переместить указатель мыши сначала на пункт Программы (Programs), а затем на IBM SPSS Statistics. Когда последний окажется выделенным, раскроется программная группа пакета. Щелкните мышью на команде IBM SPSS Statistics 20. После этого программа SPSS запустится.

Помимо запуска SPSS вы должны уметь переключаться между окнами программ с помощью Панели задач. Для работы с пакетом SPSS это весьма актуально, поскольку он представляет собой набор из нескольких одновременно выполняющихся программ, имеющих собственные окна. При запуске SPSS вы видите на экране единственное *окно редактора данных*, но как только начнете работу с данными, на экране появится также *окно вывода*. В зависимости от выполняемых действий количество открытых окон может меняться, однако окна редактора данных и вывода, как правило, присутствуют на экране постоянно. Иногда пакет SPSS переключается между окнами автоматически, в других случаях вам потребуется переключать окна вручную.

Когда несколько программ выполняются одновременно, каждой из них соответствует кнопка на Панели задач. Ниже показана Панель задач с кнопками окна редактора данных и окна вывода.



Чтобы перейти в окно нужной программы, достаточно подвести указатель к соответствующей кнопке и щелкнуть на ней.

Кнопки и другие элементы управления

Кнопки используются почти в каждом окне. Они выглядят по-разному. Например, кнопка со стрелкой, направленной вправо или влево, между окнами (рис. 2.1) предназначена для перемещения элемента (зачастую имени переменной) из одного списка в другой. А кнопка ОК запускает процедуру обработки.

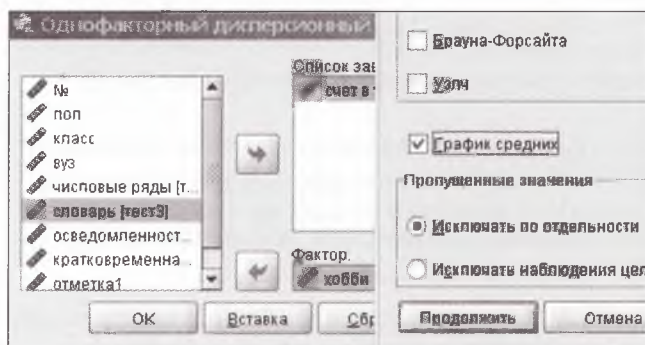


Рис. 2.1. Окно программы с кнопками, флажками и переключателями

Помимо кнопок в ходе работы с диалоговыми окнами вы можете столкнуться с еще двумя типами элементов управления. Первый элемент — *флажок*, который может быть установлен или сброшен. Если вы устанавливаете флажок, рядом с его названием появляется «галочка». Если «галочка» отсутствует, флажок считается сброшенным. Для изменения состояния флажка необходимо щелкнуть на месте расположения «галочки». Как правило (хотя и не всегда), флажки независимы, то есть установка или сброс одного из них никак не влияет на состояние остальных.

Область окна Пропущенные значения (см. рис. 2.1) представляет собой *переключатель*. Программный переключатель по своему смыслу аналогичен физическим переключателям, например переключателю скоростей в автомобиле или бытовом приборе: его можно установить только в одно положение (отмечено точкой).

Настройка параметров программы

После установки программы целесообразно задать минимальные параметры, облегчающие работу.

Прежде всего создайте на Рабочем столе ярлык программы SPSS. Предполагается, что программа SPSS еще не запущена.

ШАГ 1

1. Выделите в Главном меню Windows команду Пуск ► Программы ► IBM SPSS Statistics ► IBM SPSS Statistics 19 и щелкните на ней правой кнопкой мыши. В открывшемся меню выберите команду Копировать.
2. Переместите указатель на свободное место Рабочего стола, щелкните правой кнопкой мыши и выберите команду Вставить ярлык.

Теперь вы можете запускать программу SPSS при помощи значка на Рабочем столе.

ШАГ 2

Запустите программу SPSS при помощи значка на Рабочем столе, дважды щелкнув на нем левой кнопкой мыши, и в открывшемся диалоговом окне выбора режима работы SPSS («Что требуется выполнить?») щелкните на кнопке Отмена (Cancel).

После выполнения этого шага на экране появится окно редактора данных SPSS. В зависимости от версии, интерфейс SPSS может быть русскоязычным или англоязычным. В последних версиях программы (начиная с SPSS 17) возможно задание русскоязычного интерфейса и вывода в рамках диалогового окна Параметры (Options).

ШАГ 3

В командной строке (вверху) выберите команду Правка ► Параметры (Edit ► Options).

Откроется окно Параметры (Options) (рис. 2.2), имеющее множество вкладок.

На вкладке Общие свойства (General) задайте русский язык интерфейса в области Интерфейс пользователя (User interface) и русский язык вывода в области Вывод (Output). Переключатель Роли рекомендуется установить в положение Задавать самостоятельно.

На вкладке Местоположение файлов следует указать папку, к которой программа SPSS будет обращаться по умолчанию при открытии и сохранении файлов. Для файлов примеров создайте папку SPSSWORK на диске D:\ и укажите пути для нее, а также файлов данных и других файлов, в области Папки по умолчанию для диалоговых окон открытия и сохранения файла.

Особое внимание следует уделить вкладке Мобильные таблицы (рис. 2.3), так как на ней задаются параметры для таблиц результатов в окне Вывод. В разделе Визуализация таблиц обязательно установите флажок Вывести, используя прежний стиль таблиц, иначе некоторые таблицы могут отображаться некорректно. В раскрывающемся списке Режим редактирования по умолчанию выберите Открывать все таблицы в отдельных окнах. Эта установка позволяет использовать широкие возможности редактора SPSS для всех таблиц, независимо от их размера. В раскрывающемся списке Копирование широких таблиц в буфер обмена в формате RTF можно выбрать Уменьшить до ширины окна. Тогда при переносе широкой таблицы в документ Word она будет уменьшена до ширины окна, а не разбита на несколько таблиц. В области Показать блоки строк снимите флажок Показать таблицу в виде блоков строк, чтобы большие таблицы в окне вывода отображались целиком, а не частично. По желанию можно выбрать предпочтительный вид таблиц в области Шаблон таблиц..

На вкладке Окно результатов снимите флажок Выводить команды в журнал, тем самым избавляясь от излишне частого появления окна вывода. По мере освоения программы вы можете вновь установить этот флажок — это обеспечит постоянную регистрацию и сохранение всех манипуляций с данными

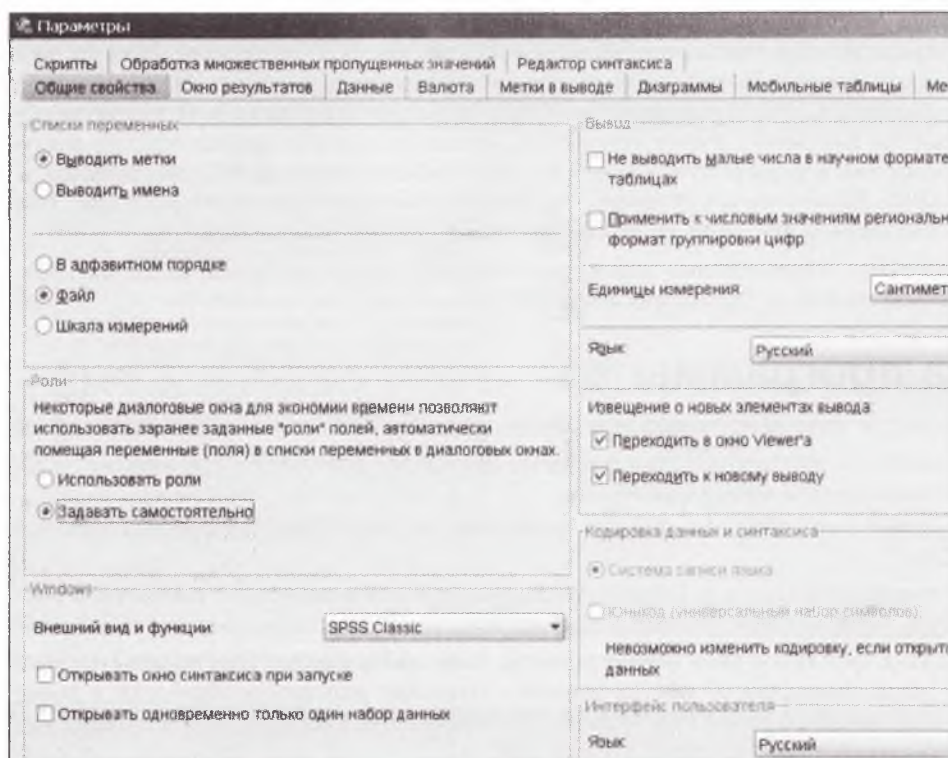


Рис. 2.2. Вкладка Общие свойства окна Параметры

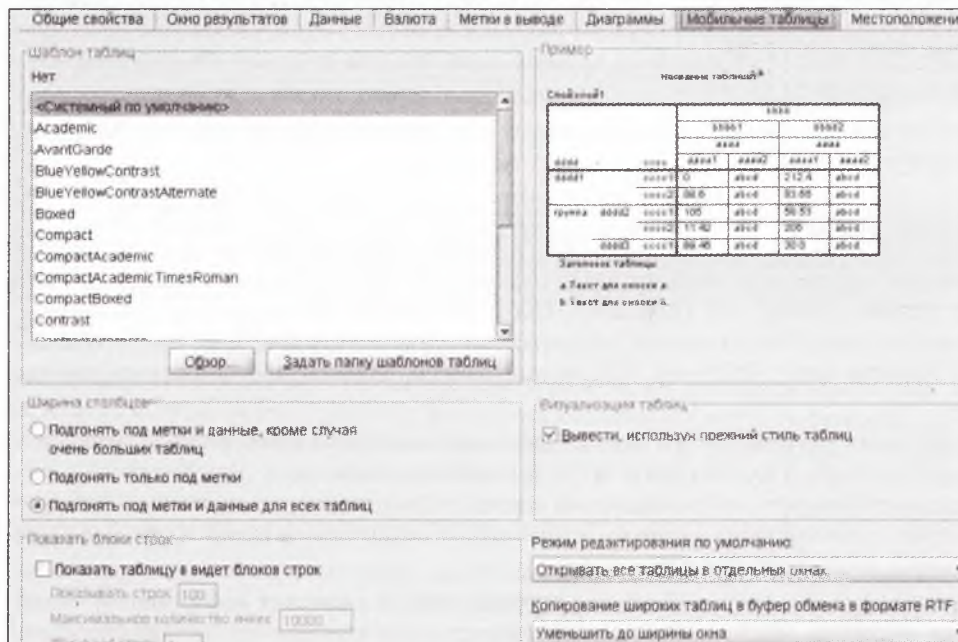


Рис. 2.3. Вкладка Мобильные таблицы окна Параметры

На вкладке Диаграммы в области Текущие установки для Параметры цикла стиля желательно установить Цикл только по узорам — тогда все графики будут черно-белыми, различаясь лишь узором. Можно также задать стиль графиков, например, в соответствии с рекомендациями APA (Американской психологической ассоциации). Для этого в области Шаблон диаграмм следует установить переключатель в положение Использовать шаблон диаграмм и указать путь к месту, где этот шаблон расположен. Если установлена версия SPSS 20, это путь C:\Program Files\IBM\SPSS\Statistics\19\Looks\APA_Styles.sgt. Этот шаблон под именем APA_Styles.sgt находится также в папке файлов примеров.

Для того чтобы установленные параметры вступили в силу, щелкните на кнопке Применить и закройте окно Параметры, нажав ОК.

Окна программы

Существует три типа окон, которые при работе с пакетом SPSS используются чаще других:

- ▶ главное окно программы (окно редактора данных) — появляется при запуске SPSS, с чего начинается работа с пакетом;
- ▶ диалоговое окно Открытие файла — позволяет получить доступ к ранее созданным файлам;
- ▶ базовые диалоговые окна (обработки данных) — хотя и зависят от конкретной процедуры, имеют схожие элементы интерфейса.

Помимо этих трех типов окон особое значение имеет окно вывода. Оно появляется каждый раз после окончания обработки данных и содержит результаты обработки, а также краткие пояснения по их интерпретации. Мы подробно рассмотрим работу с окном вывода и печатью результатов далее в этой главе.

Дополнительные возможности управления данными и их обработки предоставляет окно редактора командного языка Syntax. Работа с ним — удел весьма продвинутых пользователей SPSS. Однако мы дадим лаконичное описание этого окна, так как оно позволяет выполнять действия, выходящие за пределы того, что обеспечивает стандартный оконный интерфейс SPSS.

А сейчас рассмотрим окно редактора данных, или главное окно, которое появляется на экране при запуске SPSS.

Главное окно

При запуске программы SPSS на фоне неактивного главного окна появляется окно выбора режима работы SPSS (рис. 2.4). Чтобы это окно не появлялось в дальнейшем, установите флажок Не показывать это диалоговое окно в будущем. Для того чтобы перейти в главное окно SPSS (рис. 2.5), в окне выбора режима работы нажмите кнопку Отмена.

В верхней части окна редактора данных расположены строка меню с командами и панель инструментов. В момент запуска SPSS редактор данных пуст. Для того чтобы ввести данные, необходимо либо набрать их вручную, заполнив нужные ячейки, либо открыть существующий файл данных. Первый способ подробно описан в главе 3, а здесь мы займемся открытием файлов.

Панель инструментов расположена под строкой меню и содержит набор кнопок с различными значками. Щелчок на кнопке приводит к выполнению некоторой операции.

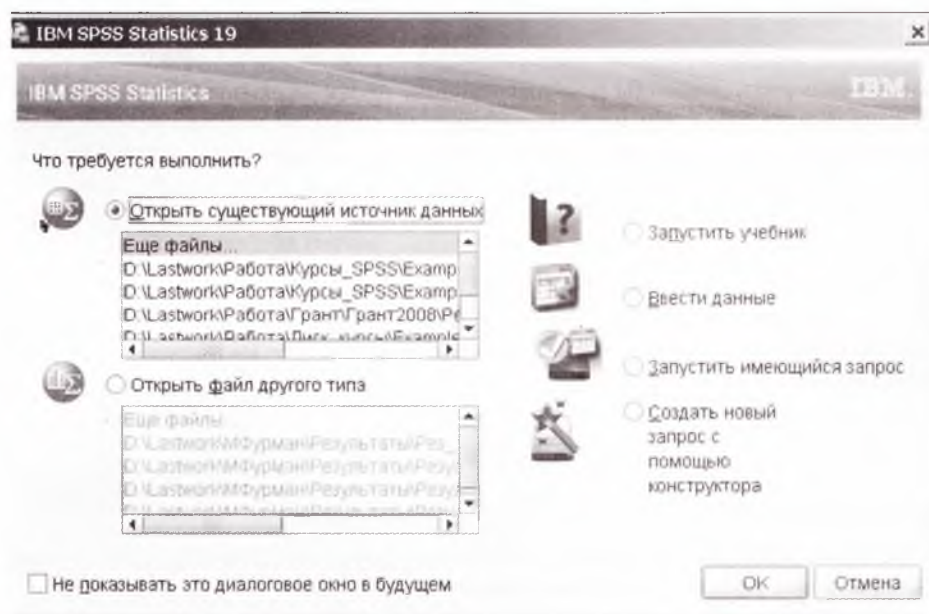


рис. 2.4. Окно выбора режима работы

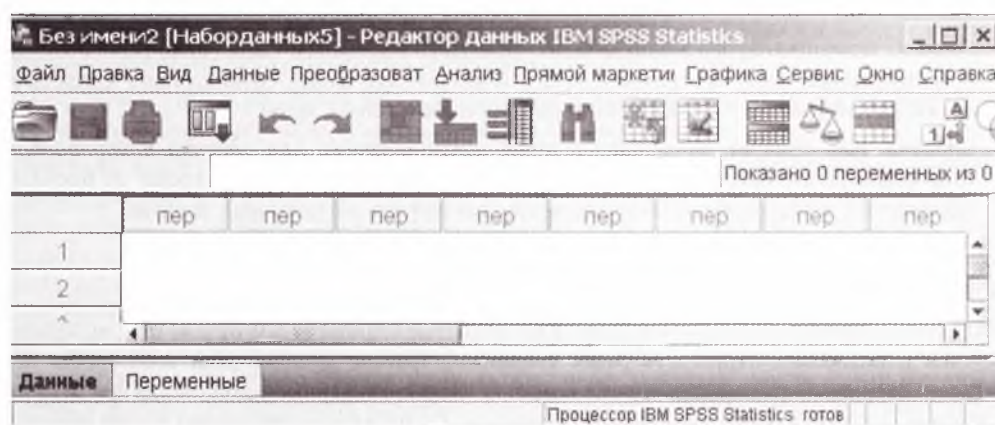


рис. 2.5. Окно редактора данных SPSS

Назначение кнопки высвечивается при наведении на нее указателя мыши. Как правило, панель инструментов позволяет выполнять без помощи меню те операции, которые требуются наиболее часто. Обратите внимание на то, что в различных приложениях SPSS вид панели инструментов может быть разным. Чтобы убедиться в этом, достаточно сравнить панели инструментов окна данных и окна вывода (см. раздел «Окно вывода и его редактирование»).

Как вы, вероятно, успели заметить, некоторые кнопки выглядят блекло. Это означает, что соответствующая команда в данный момент недоступна пользователю. Так, кнопка Печать недоступна потому, что в программе нет данных, которые можно было бы напечатать. Как только вы введете данные, кнопка Печать станет доступной. Мы рекомендуем

вам, для того чтобы приобрести навыки работы с панелями инструментов, самостоятельно пощелкать на различных кнопках и понаблюдать за реакцией программы.

Строка меню содержит команды для выполнения почти всех операций, предусмотренных в программе SPSS. По мере изучения этой книги вы познакомитесь с большим количеством команд. Как правило, выполнение команды начинается с появления диалогового окна, в котором пользователю предлагается установить значения параметров. Далее приводится краткое описание основных меню, которые будут рассмотрены:

- ▶ **Файл.** Содержит команды, предназначенные для открытия, чтения и сохранения файлов, а также команду выхода из программы SPSS.
- ▶ **Правка.** В этом меню находятся команды редактирования — копирования, вставки, замены, поиска и т. п.
- ▶ **Вид.** Содержит набор команд, влияющих на представление информации на экране. Наиболее часто используются команды **Метки значений** и **Шрифты**.
- ▶ **Данные.** Здесь находятся команды, предназначенные для управления вводом и представлением данных.
- ▶ **Преобразовать.** Содержит команды, модифицирующие введенные данные, а также создающие новые данные на основе существующих.
- ▶ **Анализ.** С этого меню начинаются все процедуры анализа данных.
- ▶ **Графика.** Здесь находятся команды, позволяющие создавать различные диаграммы. Иногда с помощью команд этого меню задают параметры статистических процедур.
- ▶ **Окно.** С помощью этого меню можно управлять взаимным расположением и статусом открытых окон программы SPSS. Фактически, меню **Окно** служит средством переключения между окнами, альтернативными **Панели задач**.
- ▶ **Справка.** Как следует из названия, меню предназначено для доступа к справочной информации. Весьма полезно при изучении этой книги было бы время от времени обращаться к справочной системе, чтобы расширять получаемые знания.

Диалоговое окно открытия файла

Диалоговое окно **Открыть данные** (рис. 2.6) является стандартным окном операционной системы и позволяет открывать ранее созданные файлы данных. Для того чтобы вызвать это окно, выберите в меню **Файл** команду **Открыть** ▶ **Данные** либо щелкните мышью на кнопке **Открыть данные** панели инструментов. Содержимое открытого таким образом окна будет соответствовать папке, с которой вы работали ранее, либо папке, заданной в качестве рабочего каталога (открываемой программой по умолчанию).

Чтобы сделать описание окна более наглядным, воспользуемся файлом данных первого из примеров с именем `ex01.sav`. Обратите внимание на то, что все файлы, созданные с помощью редактора данных SPSS, имеют расширение `.sav`, которое в зависимости от настройки вашего компьютера может отображаться или не отображаться в диалоговом окне.

Существует несколько способов открытия файлов. Если нужный вам файл находится в списке диалогового окна **Открыть данные**, для его открытия достаточно дважды щелкнуть на нем мышью. Вы также можете выделить файл в списке и щелкнуть на кнопке **Открыть**.

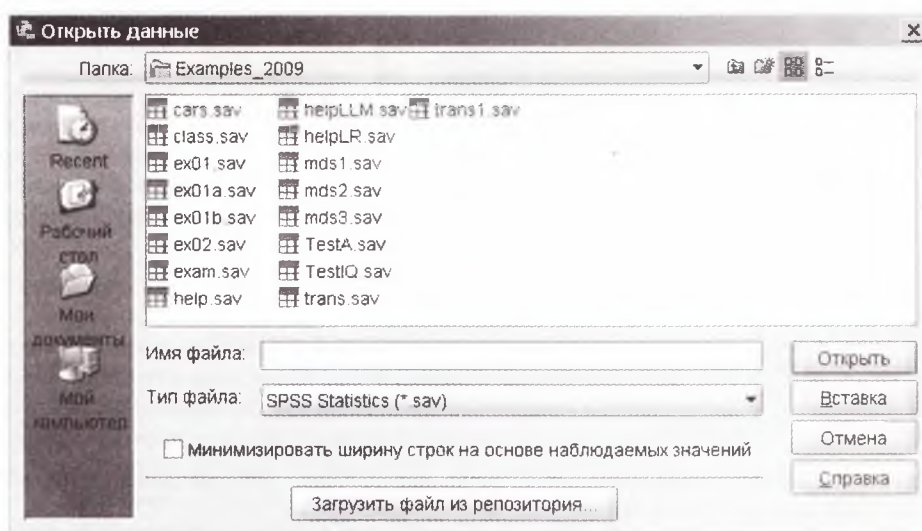


Рис. 2.6. Диалоговое окно Открыть данные

Если файл создан не в редакторе данных SPSS, перед открытием необходимо указать его тип. Для этого щелкните на стрелке раскрывающегося списка Тип файла и выберите тип, соответствующий вашему файлу. К примеру, если файл создавался программой Excel, нужно выбрать пункт Excel (*.xls). После задания типа файла вы можете воспользоваться одним из двух описанных ранее способов его открытия.

В результате открытия файла часть ячеек окна редактора данных будет заполнена данными, которые можно обрабатывать с помощью команд меню.

Начиная с SPSS версии 15.0 появилась возможность открывать одновременно несколько файлов данных (до этого для открытия нового или другого файла данных необходимо было завершить работу с предыдущим файлом). При этом каждому открытому файлу данных соответствует свое окно редактора данных. Например, можно сначала открыть файл ex01.sav, а затем ex01b.sav. Тогда окажутся открытыми два окна редактора данных: ex01.sav [Наборданных1] — Редактор данных IBM SPSS Statistics (неактивное окно) и ex01b.sav [Наборданных2] — Редактор данных IBM SPSS Statistics (активное окно). Нажимая соответствующие кнопки на Панели задач, можно переключаться между этими окнами редактора данных. Возможность работать одновременно с несколькими файлами данных существенно повышает эффективность применения SPSS.

Диалоговое окно процедуры обработки

Каждая процедура обработки имеет собственное диалоговое окно. Почти все диалоговые окна построены по одному и тому же принципу. Мы продемонстрируем это на примере диалогового окна процедуры Частоты (рис. 2.7) (см. также главу 6).

В левой части окна расположен список, содержащий все доступные переменные файла, справа могут находиться от одного до трех изначально пустых списков. Несмотря на то что заголовки этих списков различны, все они предназначены для указания переменных, участвующих в процедуре обработки. Между списками находится кнопка со стрелкой, позволяющая перемещать переменные из одного списка в другой. Стрелка на кнопке

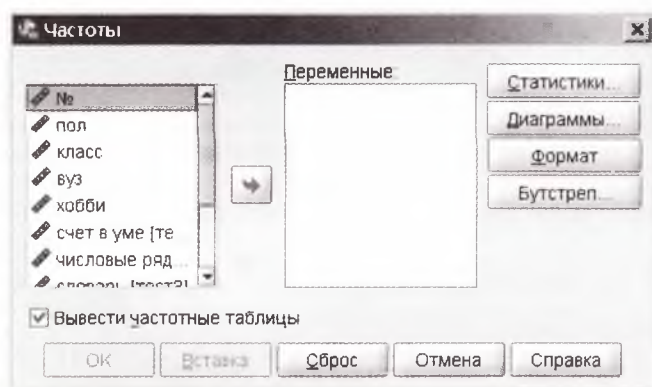


Рис. 2.7. Диалоговое окно Частоты

соответствует направлению перемещения. Если выделить пункт в левом списке, стрелка будет направлена вправо, если в правом списке — влево.

В диалоговых окнах всех процедур обработки имеются 5 кнопок:

- ▶ ОК — инициирует процедуру обработки выбранных переменных;
- ▶ Вставка — открывает окно командного языка Syntax, рассмотренное далее в этой главе;
- ▶ Сброс — возвращает параметры к значениям по умолчанию. Каждый раз при выполнении обработки набор исходных данных сохраняется, что позволяет использовать его многократно с разными значениями параметров. Это удобно, если вы проводите серию статистических опытов с одними и теми же переменными. Если же необходимо изменить начальные условия, вы сможете очистить список переменных и установить значения параметров, принятые по умолчанию, щелкнув на кнопке Сброс;
- ▶ Отмена — позволяет немедленно закрыть окно без выполнения процедуры обработки;
- ▶ Справка — открывает доступ к контекстной помощи. Это означает, что, щелкнув на этой кнопке в окне Частоты, вы получите описание именно процедуры Частоты, а не общую справку по всем процедурам обработки.

Наконец, кнопки Статистики, Диаграммы, Формат и Бутстреп позволяют задавать различные параметры, управляющие представлением результатов процедуры и ее выполнением. Почти все диалоговые окна имеют подобные кнопки. Количество кнопок и параметры, которые они позволяют задавать, меняются в зависимости от конкретной процедуры.

Окно редактора командного языка Syntax

Работа с окном редактора командного языка Syntax предоставляет более широкие возможности, чем настройка параметров команды с помощью кнопок диалогового окна. Как правило, подобные возможности доступны только опытным пользователям в особых случаях. А для решения подавляющего большинства задач, особенно в начале освоения SPSS, достаточно стандартных средств интерфейса программы. Тем не менее по мере освоения SPSS мы будем обращаться к возможностям синтаксиса (в главах 9 и 21).

В дальнейшем, освоив обработку данных при помощи стандартных средств интерфейса SPSS, вы сможете открыть для себя поистине безграничные возможности управления данными, все чаще обращаясь к командному языку Syntax. В этом вам могут

помочь многочисленные советы и рекомендации, содержащиеся на сайте А. Балабанова «Raynald's SPSS Tools по-русски» (<http://www.spsstools.ru/>).

Окно редактора командного языка Syntax предназначено для ввода, редактирования и выполнения команд Syntax. Оно имеет текстовый формат записи команд и позволяет открывать, создавать и сохранять командные файлы, которые можно использовать в дальнейшем. Такие файлы имеют расширение *.sps, и один из них содержится среди файлов примеров к этой книге (файл Syn_clust.sps, пример к главе 21). Командный файл может быть создан или открыт тремя способами.

Самый простой способ создания командного файла — кнопка Вставка любого диалогового окна процедуры обработки (см. рис. 2.7). Эта возможность рассмотрена в главе 9 при создании нестандартной таблицы корреляций. После задания в диалоговом окне процедуры обработки всех необходимых параметров (переменных, процедур, статистик и пр.) достаточно нажать кнопку Вставка, чтобы открылось окно редактора командного файла. Например, если в диалоговом окне команды Частоты задать переменные пол, класс, вуз (глава 6, шаг 5), но вместо кнопки ОК, запускающей эту команду, нажать кнопку Вставка, откроется окно редактора командного файла (рис. 2.8). Выбор через командную строку этого окна команды Запуск ► Всё приведет к выполнению действия, аналогичного нажатию кнопки ОК соответствующего диалогового окна процедуры обработки.

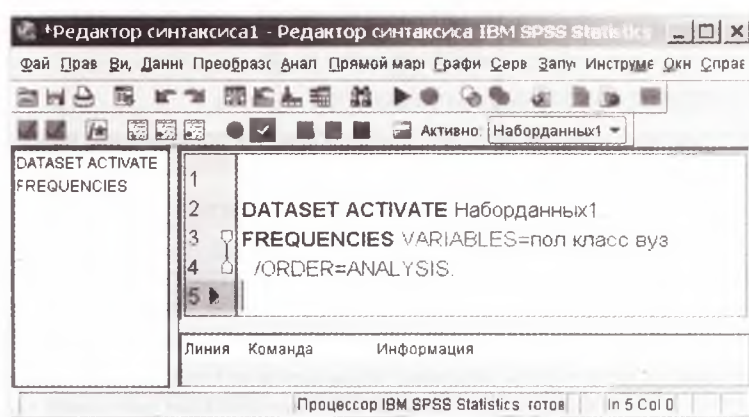


Рис. 2.8. Диалоговое окно редактора синтаксиса

Можно воспользоваться уже готовым командным файлом, выбрав в меню Файл главного окна (редактора данных) команду Открыть ► Синтаксис. Таким образом вы сможете открыть файл примера для кластерного анализа Syn_clust.sps (глава 21).

Наконец, можно самому создать новый командный файл, выбрав в меню Файл главного окна (редактора данных) команду Новый ► Синтаксис. В открывшемся (пустом) окне редактора вы можете вручную ввести необходимую синтаксическую конструкцию. А можно вставить ее, предварительно скопировав из любого текстового редактора или с сайта «Raynald's SPSS Tools по-русски» (<http://www.spsstools.ru/>).

Для выполнения открытого командного файла достаточно выбрать через командную строку редактора командного языка Syntax команду Запуск ► Все. Перед выполнением командный файл можно отредактировать по своему усмотрению, добавляя или убирая переменные, команды, процедуры и т. д. Созданный командный файл Syntax при желании можно сохранить для дальнейшего использования.

Следует отметить, что необходимость обращения к командному языку SPSS обычно возникает лишь при уверенном освоении стандартного интерфейса программы — для облегчения управления данными, выполнения длинных последовательностей команд или реализации редких методов анализа данных. Тем не менее с самого начала освоения SPSS следует делать первые шаги по изучению командного языка. Для начала достаточно при выполнении статистических процедур чаще пользоваться кнопкой Вставка с последующим выбором команды Запуск ► Все — это позволит поближе познакомиться с языком Syntax.

Окно вывода и его редактирование

Термином «вывод» в этой книге, как правило, обозначается совокупность результатов любой процедуры обработки. Каждая глава, в которой упоминается вывод данных, обязательно снабжена указаниями о содержании окна вывода. Поскольку это окно отражает главный результат обработки данных, уделим ему особое внимание.

Пример окна вывода представлен на рис. 2.9. Подобные результаты обработки можно использовать и без редактирования, сразу выводя их на печать, однако нужно иметь в виду следующие моменты:

- При выводе больших объемов данных требуется большое количество бумаги и пространство печати зачастую используется неэффективно.
- Как правило, не вся выводимая информация оказывается нужной.
- Большие таблицы, создаваемые программой SPSS, неудобны для восприятия, поэтому желательна их реорганизация.
- В отчет, который, как правило, готовится в формате Word, включается лишь часть информации, представленной в выводе.

Рассмотрим подробно возможности работы с окном вывода SPSS и его панелью инструментов. Это позволит вам в дальнейшем удалять ненужные фрагменты полученных данных, добавлять комментарии и реорганизовывать порядок вывода.

Для знакомства с файлом вывода используем результат обработки из главы 8 — таблицу сопряженности пол × хобби × класс. Желательно, чтобы все описываемые операции вы проделали сами, открыв файл примера cross.sps из меню Файл ► Открыть ► Вывод.

Некоторые кнопки на панели инструментов окна вывода совпадают с кнопками окна редактора данных. Эти кнопки выполняют аналогичные действия, разница состоит лишь в том, что в редакторе данных обрабатываются исходные данные, а в окне вывода — результаты.

Строго говоря, окно вывода на самом деле не является окном. Как видно на рис. 2.9, результаты анализа отображены в окне Вывод — Viewer IBM SPSS Statistics специальной программы просмотра результатов анализа. Окно Вывод разделено на две области. Правая область представляет собой то, что мы называли окном вывода, то есть окно результатов анализа данных. В левой области отображена иерархическая структура объектов окна вывода. В сущности, весь вывод программы — заголовки, таблицы, диаграммы и т. п. — на самом деле являются не чем иным, как разного рода объектами, имеющими иерархическую структуру. Таким образом, вместе с результатами анализа мы видим их внутреннее программное представление.

Вывод [Документ9] - Viewer IBM SPSS Statistics

Файл Правка Вид Данные Преобразовать Вставка Формат Анализ Прямой маркетинг Графика Сервис Окно Справка

Вывод

- Таблицы сопряженности
 - Заголовок
 - Примечания
 - Активный набор данных
 - Сводка обработки
 - Таблица сопряженности

→ Таблицы сопряженности

[Набор данных:1] D:\Lastwork\Работа\SPSS_2010\New\Примеры\ex01.sav

Сводка обработки наблюдений

	Наблюдения					
	Валидные		Пропущенные		Итого	
	N	Процент	N	Процент	N	Процент
пол * хобби * класс	100	100,0%	0	,0%	100	100,0%

Таблица сопряженности пол * хобби * класс

Частота

класс	пол	жен	хобби			Итого
			спорт	компьютер	искусство	
А	пол	жен	6	5	3	14
		муж	11	7	1	19
	Итого		17	12	4	33
Б	пол	жен	6	6	9	21
		муж	5	7	2	14
	Итого		11	13	11	35
В	пол	жен	3	8	15	26
		муж	2	4	0	6
	Итого		5	12	15	32
Итого	пол	жен	15	19	27	61
		муж	18	18	3	39
	Итого		33	37	30	100

Процессор IBM SPSS Statistics 19.0.0

Рис. 2.9. Окно вывода SPSS

Если вы внимательно посмотрите на содержимое окна Вывод, то заметите, что объект Примечания, отображенный в иерархической структуре, отсутствует в окне вывода. Это объясняется тем, что по умолчанию он является *скрытым*. Слева от объектов расположены значки, принимающие вид открытой или закрытой книги в зависимости от того, скрыты или отображены эти объекты в данный момент. Если дважды щелкнуть на таком значке, состояние объекта изменится на противоположное.

Иерархическая структура делает удобной навигацию внутри окна вывода. Так, если вам нужно просмотреть содержимое какой-либо таблицы, достаточно щелкнуть на объекте, чтобы соответствующее изображение появилось на экране. Как нетрудно догадаться, операция удаления фрагментов вывода сводится к удалению соответствующих объектов. Для удаления объекта щелкните на его имени в левой части окна, а затем в меню Правка выберите команду Удалить. Если вы хотите изменить порядок следования объектов, выделите один или несколько объектов, которые намереваетесь переместить, а затем в меню Правка выберите команду Вырезать. После этого выделите объект, за которым хотите их вставить, и выберите команду Правка ► Вставить после.

Уже после нескольких операций обработки одних и тех же данных навигация внутри окна вывода становится затруднительной даже при иерархической структуре объектов.

Для решения проблемы используется сворачивание объектов. Как можно видеть на рис. 2.9, слева, около названий объектов имеется значок «-» (минус). Если вы не хотите, чтобы объекты, относящиеся к какому-либо из заголовков, присутствовали на экране, щелкните на этом значке. После этого названия соответствующих объектов исчезнут, на схеме останется лишь заголовок, а вместо значка «-» (минус) появится значок «+» (плюс). Для того чтобы снова развернуть список объектов, достаточно щелкнуть на этом значке, и содержимое схемы будет восстановлено.

Теперь, когда вы получили элементарные знания о том, что представляет собой окно вывода программы SPSS и как с ним работать на объектном уровне, пришло время заглянуть внутрь самих объектов и научиться изменять их. Нас будет интересовать структура таблиц и диаграмм, поскольку от того, насколько удобно представлены эти элементы вывода, зависит удобочитаемость выводимых данных в целом. В этой книге мы не будем рассматривать все процедуры, изменяющие представление таблиц и диаграмм, однако знаний, которые вы получите, вполне достаточно для комфортной работы с результатами исследований. Редактирование диаграмм описывается в главе 5.

В первую очередь необходимо уяснить принцип построения таблиц. Обычно SPSS создает таблицы автоматически, самостоятельно решая, какие показатели разместить в строках, а какие — в столбцах. Конечно, не всегда удобно полагаться на выбор программы, и для этих случаев SPSS предлагает средства редактирования таблиц. Следует отметить, что средства редактирования таблиц SPSS значительно совершеннее возможностей Word.

Чтобы попрактиковаться в редактировании таблиц, снова обратимся к таблице сопряженности пол × хобби × класс (из файла примера cross.spv). Если вы успели поработать с этим файлом, закройте его без сохранения и вновь откройте, чтобы вернуть его прежний вид. В таблице сопряженности (из файла cross.spv), приведенной на рис. 2.10, SPSS по умолчанию размещает данные о хобби учащихся в столбцах, а о поле и классе — в строках. Несмотря на то что в подобном представлении нет ничего неправильного, могут существовать ситуации, когда для наглядности потребовалось бы расположить данные в таблице по-другому. При этом сами данные необходимо оставить неизменными, поскольку они являются результатом анализа.

Таблица сопряженности пол 'хобби' класс						
Частота			хобби			Итого
класс			спорт	компьютер	искусство	
1	пол	ЖЕН	6	5	3	14
		МУЖ	11	7	1	19
	Итого		17	12	4	33
2	пол	ЖЕН	8	6	9	21
		МУЖ	5	7	2	14
	Итого		11	13	11	35
3	пол	ЖЕН	3	8	15	26
		МУЖ	2	4	0	6
	Итого		5	12	15	32

Рис. 2.10. Исходная таблица сопряженности

В столбцах рассматриваемой таблицы перечислены хобби учащихся (спорт, компьютер, искусство), а каждая строка содержит два уровня заголовков: номер класса (1, 2, 3) и пол ученика (ЖЕН, МУЖ). Для каждого заголовка первого уровня перебраны все возможные

варианты заголовков второго уровня. Данная таблица является однослойной (слой — класс), поэтому мы сразу перейдем к организации столбцов и строк.

Для того чтобы выполнять какие-либо действия с таблицей, необходимо открыть ее для редактирования. Следует навести на нее указатель мыши и щелкнуть правой кнопкой. В открывшемся меню выбрать Изменить содержимое. Доступны два режима редактирования: В средстве просмотра — непосредственно в окне вывода, В отдельном окне — в открывшемся окне Мобильная таблица. В окне вывода удобно редактировать небольшие таблицы, а для больших таблиц лучше выбирать пункт В отдельном окне. При выборе любого из этих пунктов на экране исчезнут панели инструментов, а в строке меню появится новое меню Мобильная таблица. Кроме того, изменится набор доступных команд меню. Далее будут рассмотрены команды Повернуть внутренние метки столбцов, Повернуть наружные метки строк, Транспонировать строки и столбцы и Поля вращения меню Мобильная таблица и Формат.

Вероятно, вы уже обратили внимание на то, что ширина столбцов рассматриваемой таблицы гораздо больше, чем необходимо для отображения содержащихся в ней чисел. Это объясняется тем, что дополнительное пространство занято именами (метками) столбцов. Если изменить направление текста меток столбцов, то есть применить к таблице команду Формат ► Повернуть внутренние метки столбцов, формат таблицы станет более компактным (рис. 2.11).

Для того чтобы восстановить предыдущий формат, достаточно выбрать указанную команду еще раз. Аналогичным образом можно изменять формат строк, только в этом случае нужно в меню Формат выбирать команду Повернуть наружные метки строк.

Как уже говорилось, при создании таблицы содержимое ее строк и столбцов назначается программой автоматически. Если вам необходимо изменить расположение категорий в таблице, следует воспользоваться командой Транспонировать строки и столбцы или Поля вращения. Первая меняет местами строки и столбцы, результат ее применения к рассматриваемой таблице показан на рис. 2.12.

			хобби			Итого
			спорт	компьютер	искусство	
класс	пол	ЖЕН	6	5	3	14
		МУЖ	11	7	1	19
	Итого		17	12	4	33
2	пол	ЖЕН	6	6	9	21
		МУЖ	5	7	2	14
	Итого		11	13	11	35
3	пол	ЖЕН	3	8	15	26
		МУЖ	2	4	0	6
	Итого		5	12	15	32

Рис. 2.11. Таблица сопряженности после поворота меток столбцов

Таблица сопряженности пол 'хобби' класс

Частота

		класс								
		1			2			3		
		пол		Итого	пол		Итого	пол		Итого
		ЖЕН	МУЖ		ЖЕН	МУЖ		ЖЕН	МУЖ	
хобби	спорт	6	11	17	6	5	11	3	2	5
	компьютер	5	7	12	6	7	13	8	4	12
	искусство	3	1	4	9	2	11	15	0	15
Итого		14	19	33	21	14	35	26	6	32

Рис. 2.12. Таблица сопряженности после перестановки строк и столбцов

Вторая команда позволяет осуществить более гибкую реорганизацию таблицы. При ее выборе на экране появляется новое окно Поля вращения (рис. 2.13), в котором вы можете задать содержимое будущих строк и столбцов. Команда Поля вращения также предназначена для реорганизации слоев.

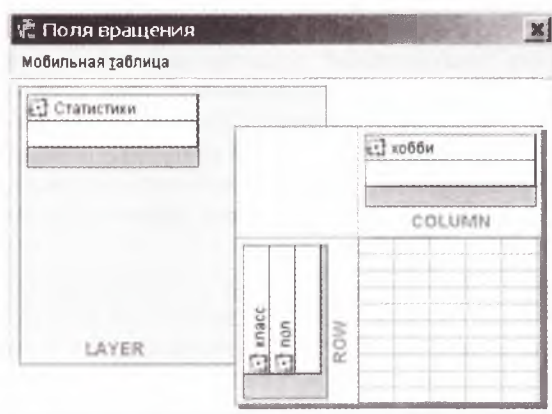


Рис. 2.13. Окно Поля вращения

Реорганизация таблицы производится с помощью *осей*. Под *осью строк* будем понимать столбец, содержащий метки строк (ROW), а под *осью столбцов* — строку, содержащую метки столбцов (COLUMN). В нашем примере на оси столбцов находится информация о номере класса и поле учащихся, а на оси строк — об их хобби.

В окне Поля вращения имеются панели ROW (Строка) и COLUMN (Колонка), соответствующие оси строк и оси столбцов. Вы можете менять положение меток, перетаскивая мышью (удерживая нажатой ее левую кнопку) цветной квадратик с названием переменной (пол, класс, хобби) в пределах одной оси или с одной оси на другую. Если вы хотите, к примеру, поменять местами класс и пол учащихся, достаточно перетащить в окне Поля вращения значок переменной пол в пустую ячейку правее переменной класс. В этом случае пол окажется заголовком первого уровня и каждая метка переменной пол будет содержать 3 метки переменной класс.

Можно также реорганизовывать переменные, перемещая соответствующие квадратные значки с именами переменных между осями строк и столбцов. Так, если вам необходимо сделать пол переменной второго уровня для каждого значения переменной хобби, следует перетащить переменную пол под значок переменной хобби.

Часто возникает необходимость удаления лишних столбца или строки таблицы. Для этого достаточно выделить заголовок щелчком левой кнопки мыши и затем открыть меню щелчком правой кнопки. В раскрывшемся меню нужно выбрать команду Выбрать ► Ячейки данных и метки. Будет выделена соответствующая область таблицы. Наведя указатель на выделенную область, правой кнопкой мыши снова откройте меню и выберите Вырезать или Очистить для удаления.

Как вы уже успели заметить, даже для такой простой таблицы, как эта, существует большое количество вариантов реорганизации. Самый простой способ научиться реорганизовывать таблицы — поэкспериментировать с ними, воспользовавшись примером cross.srv или своими результатами.

Сохранение, экспорт, перенос и печать результатов

После редактирования содержимого окна вывода в соответствии с вашими предпочтениями у вас, вероятно, возникнет желание сохранить результат для дальнейшего использования.

Сохранение результатов

В отличие от файла данных, файл вывода имеет расширение .spv. Для сохранения файла вывода в меню File (Файл) выберите команду Save (Сохранение), в появившемся диалоговом окне задайте имя файла и щелкните на кнопке Save (Сохранить)¹.

Помимо сохранения программа SPSS позволяет *экспортировать* вывод в другие форматы (например, Word), *перенести* его в другой документ или *напечатать* весь файл вывода или его выделенные фрагменты. Перед этим настоятельно рекомендуется обдумать, стоит ли переносить или печатать те или иные фрагменты вывода, поскольку искать полезную информацию среди большого объема ненужных таблиц, как правило, довольно сложно.

Чтобы *выделить* для экспорта, переноса или печати фрагмент выводимых данных, выполните в окне вывода любое из трех действий:

- ▶ Щелкните на заголовке или объекте иерархической структуры, расположенной в левой части окна вывода. При щелчке на заголовке выбранными оказываются все объекты, расположенные под этим заголовком.
- ▶ Способом, описанным в предыдущем пункте, выделите один объект, а затем, удерживая клавишу Shift, выделите другой объект. В результате выбранными окажутся все объекты, расположенные между ними.
- ▶ Выделите один объект, а затем, удерживая клавишу Ctrl, поочередно выделите все объекты, предназначенные для печати. Этот способ отличается от предыдущего тем, что позволяет выбирать несмежные объекты.

Экспортирование результатов

Начиная с версии 11.5, программа SPSS предоставляет возможность сохранения файлов вывода и в более привычном для многих формате текстового редактора Word. Для этого в меню Файл выберите команду Экспортировать. На экране появится диалоговое окно Экспортировать вывод (рис. 2.14). В этом окне необходимо в области Документ в раскрывающемся списке Тип выбрать пункт Word/RTF file (*.doc). В поле Имя файла ввести имя нового файла и задать место его расположения, щелкнув на кнопке Обзор. Дополнительно с помощью переключателя Экспортируемые объекты можно выбрать объекты для экспорта (обычно Все видимые объекты или Выделенные объекты). А при помощи кнопки Изменить параметры (в области Документ) установить более тонкие настройки параметров экспортирования документа. После окончания настройки в этом диалого-

¹ В более ранних версиях файлы вывода имели расширение .spo. Для того чтобы иметь возможность открывать такие файлы, необходимо установить модуль Legacy Viewer. Его можно скачать с сайта www.spss.com по ссылке http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/LegacyViewer/LegacyViewer_English.zip.

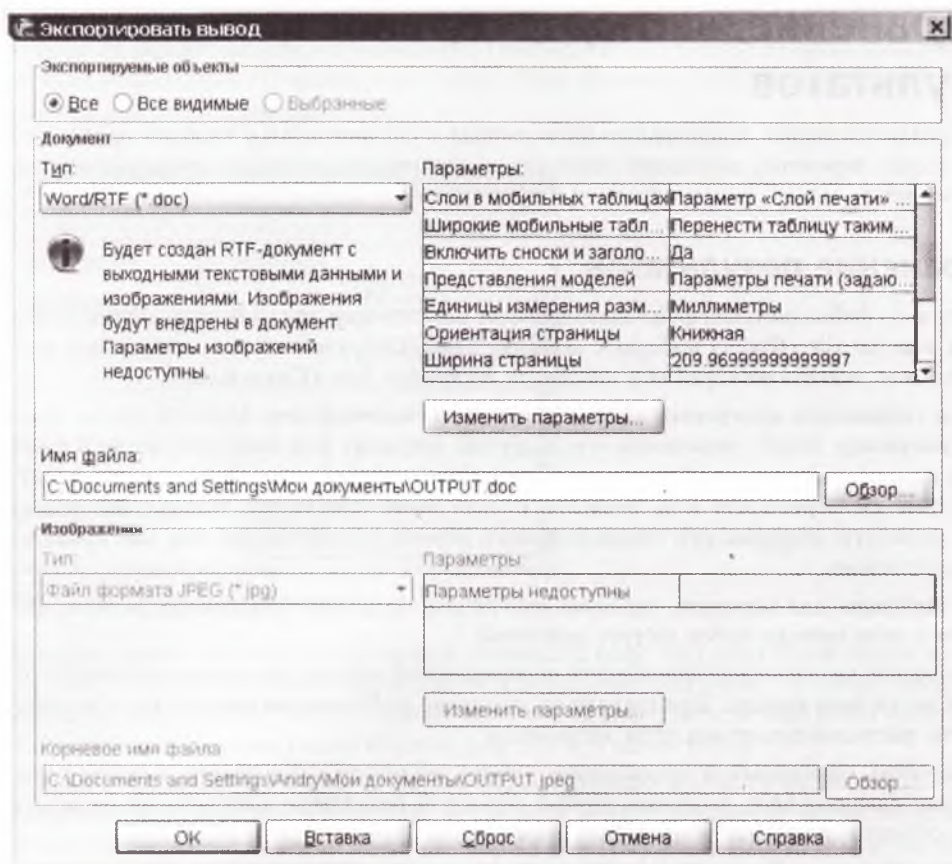


Рис. 2.14. Диалоговое окно Экспортировать вывод

вом окне щелкните на кнопке ОК. Будет создан файл с расширением .doc, содержащий выбранные объекты файла вывода. Этот файл можно будет открывать и редактировать в программе Word.

Перенос результатов в документ Word

Наиболее популярный способ сохранения содержимого окна вывода — перенос его выделенной части в документ Word. Для этого сначала следует скопировать выделенную часть окна вывода в буфер обмена, например, при помощи щелчка правой кнопкой мыши и выбора в контекстном меню команды Копировать.

Доступны два варианта вставки скопированного фрагмента в документ Word. При выборе команды Вставить в документе Word таблица будет вставлена как обычная таблица Word, доступная для редактирования. Недостаток этого способа заключается в том, что средства редактирования таблиц SPSS заметно совершеннее возможностей Word, поэтому перенос может нарушить структуру таблицы. Вторым вариантом — перенести таблицу как рисунок. Для этого после копирования таблицы вывода следует в документе Word выбрать в меню Правка (в командной строке) команду Специальная вставка и указать

в перечне объектов Точечный рисунок. Перенос таблицы из SPSS в Word как рисунка гарантирует сохранность ее формата, но лишает возможности редактировать таблицу.

Печать результатов

После выделения предназначенных для печати объектов в меню Файл выберите команду Предварительный просмотр для проверки расположения результатов на бумаге. При необходимости поменяйте ориентацию листов с книжной на альбомную. Щелкните на кнопке Печатать. На экране появится системное диалоговое окно, с помощью которого можно выбрать принтер (в случае, если на компьютере установлено более одного принтера), а также диапазон печати, определяющий, следует печатать все данные или только выделенные фрагменты. Когда все параметры установлены нужным образом, щелкните на кнопке ОК для завершения процесса.

Создание и редактирование файлов данных

38	Структура файла данных
48	Ввод данных
49	Редактирование данных
51	Пример файла данных

Эта глава поможет вам сделать первый шаг к статистическому анализу — ввести данные в программу. Как вы уже знаете из предыдущей главы, в SPSS для управления данными используется специальная программа — редактор данных. Окно редактора данных представляет собой электронную таблицу, в которой столбцы отражают переменные, а строки — объекты (случаи). На пересечении строки и столбца находится ячейка, в которой сохраняется значение переменной (столбца) для данного объекта (строки). Для работы с таблицами, не уместяющимися в пределах окна, предусмотрены вертикальные и горизонтальные полосы прокрутки, а также автоматические функции перехода к заданным переменной или объекту.

Данная глава в книге первая из содержащих пошаговые процедуры. В них приведена последовательность действий, необходимых для ввода данных в компьютер. В качестве примера мы будем использовать файл `ex01.sav`, упоминавшийся в главе 2 и содержащий разнообразную информацию о 100 школьниках, обучающихся в трех классах. Как уже упоминалось, готовые файлы примеров можно загрузить с веб-сайта <http://www.piter.com/download>.

Структура файла данных

Перед тем как начать ввод данных, необходимо определить структуру будущего файла. Для этого вы должны ответить себе на вопрос: как будут использоваться в анализах те или иные переменные? К сожалению, многие пользователи начинают задумываться об этом гораздо позже, чем следует, и в этом кроется причина многих ошибок при проведении анализа данных: чем сложнее план исследования, тем больше вероятность того, что из-за необдуманных действий исследователя оно завершится неудачей.

Итак, первое, что следует сделать, — определить последовательность действий при обработке данных. Кроме того, необходимо четко представлять себе структуру и взаимосвязи переменных в вашем плане исследования. Далее перечислены наиболее характерные для файлов данных ошибки и недостатки:

- ▶ Отсутствуют ключевые переменные (пол, возраст и т. п.), являющиеся основой для анализа.
- ▶ Переменная плохо отражает содержание соответствующей реальной величины (например, на сложный вопрос предусмотрено только два варианта ответа: «да» и «нет»).

- При большом количестве независимых переменных отсутствуют зависимые переменные, отражающие цель исследования (или наоборот).
- Недостаточно независимых переменных, влияющих на заданную зависимую переменную.

Эти примеры демонстрируют, что залогом успеха как исследования в целом, так и создания файла данных в частности является тщательно продуманный выбор структуры данных.

Следует отметить, что структура файла данных должна соответствовать плану исследования. В то же время план исследования должен быть составлен так, чтобы его исходные данные можно было обработать в соответствии с задачами и гипотезами исследования. Самый оптимальный и простой путь обеспечения этих соответствий — определение структуры данных на этапе планирования исследования, еще до их сбора. Это позволит избежать большинства типичных ошибок, относящихся как к планированию исследования, так и к организации данных.

На этапе планирования исследования структура данных может быть задана в виде предварительного списка переменных с указанием их типов и диапазонов возможных значений, например так, как в табл. 3.1.

Таблица 3.1. Предварительный список переменных

№	Название	Тип	Диапазон возможных значений
1	Идентификационный номер	Номинальная	1–100
2	Пол	Номинальная	1 — жен; 2 — муж
3	Класс	Номинальная	1 — А; 2 — Б; 3 — В
4	Предполагаемый для поступления вуз	Номинальная	1 — гуманитарный; 2 — экономический; 3 — технический; 4 — естественнонаучный
5	Внешкольные увлечения	Номинальная	1 — спорт; 2 — компьютер; 3 — искусство
6–10	Показатели тестов 1–5	Количественная	1–20
11	Средний балл за 10-й класс	Количественная	3–5
12	Средний балл за 11-й класс	Количественная	3–5

Каждая переменная — это значимое для исследователя основание, позволяющее отличать объекты друг от друга. На предварительном этапе следует выделять два типа переменных: количественные и категориальные (номинальные). Количественная переменная позволяет различать объекты по уровню выраженности некоторого свойства, например: средний балл отметки, тестовый показатель и пр. Идентификация количественных переменных на предварительном этапе не составляет труда: обычно они соответствуют тому, что исследователь намеревается измерить. Второй тип — категориальные (номинальные) переменные. Обычно они используются как основания для деления объектов (испытуемых) на группы или категории: пол, класс, возрастная категория, уровень дохода и пр. Типичная ошибка начинающего исследователя — игнорирование возможных

оснований для деления объектов на группы как самостоятельных номинальных переменных в структуре данных.

Важным свойством номинальных переменных является возможность их представления в виде набора целых чисел. Например, трем видам внешкольных увлечений (хобби) учащихся (спорт, компьютер, искусство) можно сопоставить числа 1, 2 и 3 соответственно. Числовое представление данных в компьютерных программах всегда предпочтительнее символьного, поскольку обработка чисел происходит быстрее, проще и с меньшей вероятностью ошибок. Кроме того, числовое представление легко модифицировать: вы можете переназначить числа, соответствующие созданным элементам, а также (что часто требуется на практике) без проблем включить в анализ новые элементы. Например, если в группе окажется учащийся, увлечение которого не соответствует перечисленным, будет полезно включить в переменную хобби категорию с названием другие и присвоить ей число 4. Эта операция рассмотрена подробнее в главе 4.

Порядок создания переменных также важен при вводе данных. Здесь следует придерживаться простого правила: наиболее важные и часто используемые переменные должны находиться в начале файла, для остальных данных порядок следования не столь важен, однако рекомендуется объединять их в группы по их «физическому смыслу». Чаще всего в начало файла следует помещать категориальные переменные, которые далее предполагается использовать для деления объектов (испытуемых) на группы, например: пол, семейное положение и пр. Далее можно перечислять остальные сведения, а логическое объединение переменных производить в зависимости от того, какие аспекты они отражают.

Далее приведем инструкции, с помощью которых вы сможете приступить к освоению программы SPSS.

ШАГ 1

Первое, что необходимо сделать, — запустить программу SPSS. В открывшемся диалоговом окне выбора режима работы щелкните на кнопке Отмена. Вы получите доступ к окну редактора данных (рис. 3.1).

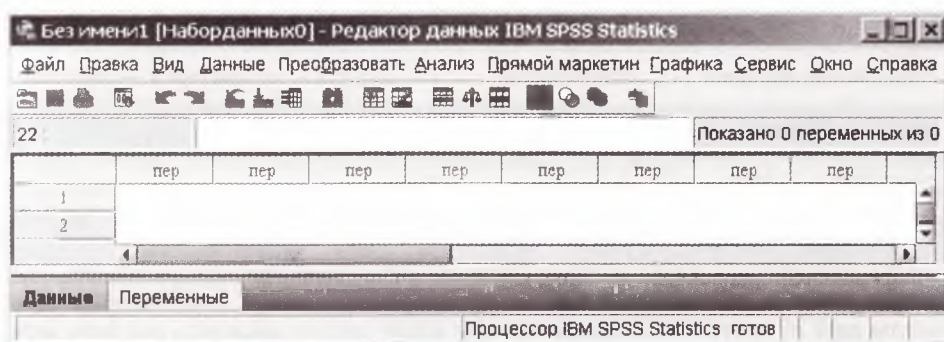


Рис. 3.1. Исходный вид окна редактора данных SPSS

ШАГ 2

Перейдите на вкладку Переменные, щелкнув на ее ярлычке мышью (рис. 3.2).

Вкладка Данные, которая отображается сразу после запуска редактора, предназначена для ввода значений в создаваемый файл данных. Вкладка Переменные позволяет задать структуру файла данных, то есть определить имена, метки и структуры переменных. Заголовки столбцов представляют собой параметры каждой из переменных.

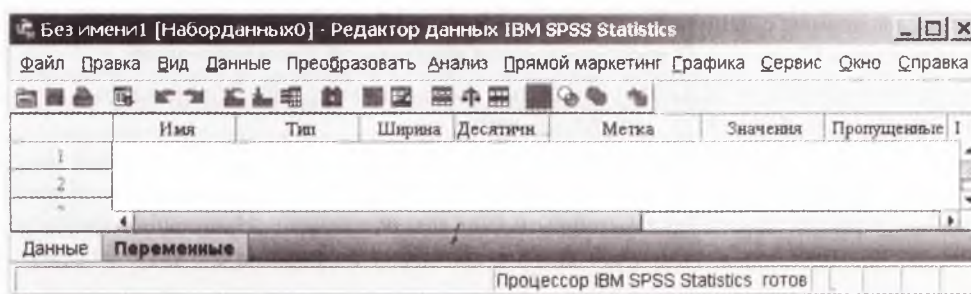


Рис. 3.2. Вкладка просмотра переменных окна редактора данных

В следующих пошаговых инструкциях возьмем за основу данные из файла ex01.sav. Часть его содержимого приведена в конце данной главы. В процессе создания файла можно выделить три основных действия: задание имени переменной, определение ее параметров и ввод данных. Необязательно последовательно выполнять каждый из этих этапов для каждой переменной; такой порядок выбран лишь в качестве примера. На практике вам может оказаться удобнее создавать файл данных порциями: сначала полностью задать параметры части переменных, ввести их значения, затем задать новую группу переменных и т. д. Мы же сейчас обратимся к рассмотрению параметров переменных.

Имя переменной

Параметр Имя определяет имя переменной. Чтобы задать имя первой переменной, просто введите его с клавиатуры в текущую ячейку. Имя второй переменной вводится в том же столбце под именем первой, то есть во второй строке, имя третьей переменной — в третьей строке и т. д. Для перемещения между строками пользуйтесь клавишами ↓ и ↑. Если вы пользуетесь одной из ранних версий SPSS (до 12-й), возможно, вам не удастся задавать имена переменных кириллицей. Ничего страшного, набирайте имена и параметры переменных латинскими буквами.

ШАГ 3

Для задания имен всем переменным файла ex01.sav выполните следующие действия (должна быть открыта вкладка Переменные редактора данных):

1. Введите символ № и нажмите клавишу ↓.
2. Введите слово пол и нажмите клавишу ↓.
3. Введите слово класс и нажмите клавишу ↓.

Аналогичным образом вводятся оставшиеся имена переменных.

Как вы, наверное, уже поняли, при переходе на следующую строку (то есть при окончании ввода имени переменной) оставшиеся параметры переменной автоматически заполняются значениями по умолчанию.

Имя переменной не является произвольным. Существует ряд соглашений, которым оно должно удовлетворять:

- ▶ длина имени — не более 64 символов (в ранних версиях — до 8 символов);
- ▶ в имени могут использоваться любые буквы, цифры, символы @, #, ., -, \$, однако оно всегда должно начинаться с буквы, а символ «.» (точка) не может стоять в его конце;
- ▶ имена переменных должны быть разными;

- ▶ буквы, введенные с использованием верхнего и нижнего регистров, не различаются, то есть имена ID, id, Id и iD воспринимаются программой как идентичные;
- ▶ имена переменных не должны совпадать с каким-либо из зарезервированных слов (all, ne, eq, to, le, lt, by, or, gt, and, not, ge, with).

Чтобы не запоминать все ограничения, проще пользоваться следующим правилом именования переменных. Имя должно быть коротким (не более 8 символов), начинаться с буквы и содержать только буквы, цифры и знак подчеркивания. И, конечно, имена не должны повторяться.

Тип переменной

Параметр Тип определяет тип переменной. Если фокус ввода находится в столбце Имя какой-либо из созданных переменных, переместите его в столбец Тип, например, нажав клавишу Tab, тогда в правой части ячейки появится кнопка с многоточием. Щелчок на ней приводит к появлению на экране диалогового окна Тип переменной (рис. 3.3).

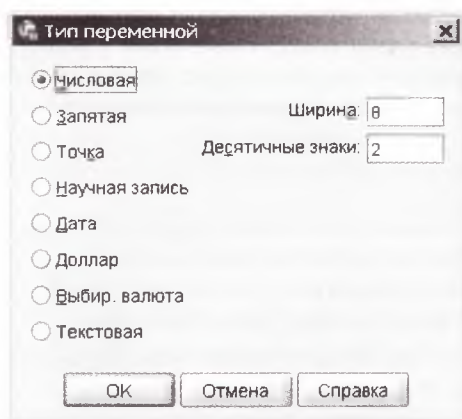


Рис. 3.3. Диалоговое окно Тип переменной

Как видите, текущим типом переменной является Числовая. В подавляющем большинстве случаев вам придется иметь дело именно с числовыми данными. В тех редких случаях, когда значения переменных представляют собой буквы или буквосочетания (слова), необходимо установить переключатель в положение Текстовая. Текстовые данные, в отличие от числовых, могут включать в себя буквы и другие символы, то есть нести текстовую информацию. Это могут быть ФИО респондентов, название города и т. п. В частном случае текстовая переменная может хранить число, однако обработка такого «числа» будет производиться так, как будто оно является текстом. *Как правило, текстовые переменные не подлежат обработке. Поэтому их следует избегать, за исключением редких случаев, например, когда данная переменная содержит имена людей или названия городов.*

Остальные шесть типов переменных, которые можно задать с помощью данного переключателя, практически не используются при обработке данных, поэтому мы не будем на них останавливаться.

Поскольку все переменные имеют установленный по умолчанию тип Числовая, никаких дополнительных действий с ними производить не нужно.

Ширина переменной

Параметр Ширина позволяет задать максимальное количество знаков, которое может иметь значение переменной, включая дробную часть. Например, ширина переменной № (идентификатор) составляет 3 знака, поскольку числа, которые мы намерены вводить, содержат от 1 до 3 знаков. На практике заранее определить ширину переменной гораздо труднее, поскольку не всегда известно, какие данные нам понадобятся вводить в будущем. Поэтому следует задавать ширину переменной с гарантированным запасом: ее можно ограничить потом, после ввода данных.

В большинстве случаев нет необходимости менять заданную по умолчанию ширину переменной.

Зададим для всех переменных ширину 3 знака.

ШАГ 4

Для задания ширины переменных выполните следующие действия:

1. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Ширина, соответствующую переменной №, и задайте значение 3. В ячейке должно остаться значение ширины, равное 3.
2. Наведите указатель мыши на ячейку столбца Ширина, соответствующую переменной №, нажмите правую кнопку мыши и выберите Копировать.
3. При помощи мыши вставьте скопированное значение для всех остальных переменных. Для этого выделите ячейки столбца Ширина соответствующих переменных, нажмите правую кнопку мыши и выберите Вставить.

Дробная часть числа

Параметр Десятичные предназначен для задания числа десятичных знаков после запятой в случае, если тип переменной допускает использование дробных чисел. Для строковых переменных значение в ячейке Десятичные автоматически устанавливается равным 0, а для числовых переменных — 2. У строковых переменных значение параметра Десятичные недоступно для изменения.

Когда фокус ввода оказывается в ячейке столбца Десятичные, справа появляются две кнопки счетчика. Щелкая на этих кнопках, вы можете изменять текущее количество знаков после запятой. Изначально у всех числовых переменных оно равно 2.

В файле ex01.sav знаки после запятой необходимы для переменных отметка1 и отметка2, причем количество этих знаков равно 2 и задано программой по умолчанию. Таким образом, задача заключается в том, чтобы установить для всех остальных числовых переменных значение параметра Десятичные равным 0. Отметим, что это не является необходимым, а служит для удобства отображения данных на экране или в отчете.

ШАГ 5

Для задания числа десятичных знаков в переменных выполните следующие действия:

1. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Десятичные, соответствующую переменной №, и дважды щелкните на нижней кнопке счетчика. В ячейке должно остаться значение, равное 0.
2. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Десятичные, соответствующую переменной пол, и дважды щелкните на нижней кнопке счетчика. В ячейке должно остаться значение, равное 0.
3. Аналогичные действия повторите для остальных переменных, кроме переменных отметка1 и отметка2.

Метки переменных

С помощью параметра Метка можно создать метку переменной. Как правило, она используется в тех случаях, когда смысл переменной недостаточно точно отражен в ее имени. По сути, метка — это комментарий к имени переменной. Заданные метки переменных отображаются в окне вывода вместо их имен.

Ячейки столбца Метка представляют собой обычные текстовые поля, в которые вы можете вводить текст меток. Длина метки не должна превышать 256 символов. Ограничений на используемые символы нет. Но помните о том, что слишком длинные метки ухудшают удобочитаемость фрагментов, в которых они присутствуют. Как показывает практика, 20–30 символов вполне достаточно, чтобы описать назначение переменной. И чем короче метка, тем лучше.

Если какая-либо из надписей окажется длиннее ширины столбца, последняя автоматически увеличится до необходимого размера. Если вам не обязательно видеть все надписи целиком, можете задать ширину столбца вручную. Для этого подведите указатель мыши к правой границе ячейки с меткой столбца — указатель примет вид двух стрелок, направленных в разные стороны. Нажмите левую кнопку мыши и, удерживая ее нажатой, перетащите правую границу столбца на новое место. Аналогичным способом можно перетаскивать границы любых столбцов и тем самым оформлять вкладку Переменные по своему вкусу.

ШАГ 6

Для ввода меток выполните следующие действия:

1. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Метка, соответствующую переменной тест1, введите словосочетание *счет в уме* и нажмите клавишу Enter.
2. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Метка, соответствующую переменной тест2, введите словосочетание *числовые ряды* и нажмите клавишу Enter.
3. Для переменной тест3 введите слово *словарь*, для переменной тест4 — слово *осведомленность*, для переменной тест5 — словосочетание *кратковременная память*.

Метки значений переменных

Параметр Значения позволяет управлять наименованиями уровней (категорий) переменной. Под уровнем, или категорией, понимается целочисленное значение переменной, имеющее определенный смысл. Например, переменная пол имеет два уровня: 1 — жен (женский) и 2 — муж (мужской). Буквосочетания или слова, поставленные в соответствие уровням переменной, например жен и муж, называются *метками значений* и отражают смысл разных значений переменной для исследователя. Наличие меток значений исключительно важно для удобочитаемости результатов анализа, поскольку очень трудно запомнить смысл значений всех переменных, даже если их сравнительно немного. В крупных демографических исследованиях количество переменных может превышать 100, поэтому метки значений являются практически незаменимым средством расшифровки результатов статистического анализа. SPSS также позволяет включать метки значений в файл данных, что очень удобно при его создании. Длина метки значения ограничивается 60 символами, однако, как правило, пользователям достаточно 1–10 символов.

Как и в случае со столбцом Тип, перемещение в ячейку столбца Значения приводит к появлению кнопки с многоточием. Если щелкнуть на ней, на экране появится диалоговое окно Метки значений, в котором можно задать метки для различных значений переменной (рис. 3.4).

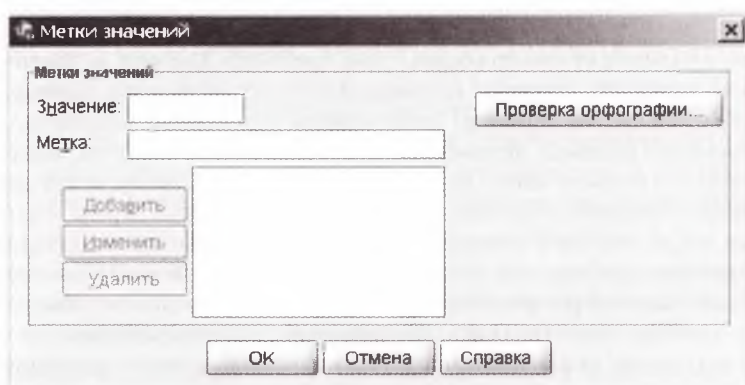


Рис. 3.4. Диалоговое окно Метки значений

ШАГ 7

Далее в качестве примера приведены действия, позволяющие задать метки значений для переменной пол:

1. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Значения, соответствующую переменной пол, и щелкните мышью на кнопке с многоточием.
2. В поле Значение введите число 1, нажмите клавишу Tab для перехода в поле Метка, введите туда слово жен и щелкните на кнопке Добавить.
3. В поле Значение введите число 2, нажмите клавишу Tab для перехода в поле Метка, введите туда слово муж и щелкните на кнопке Добавить.
4. Для закрытия диалогового окна Метки значений щелкните на кнопке ОК.

После задания меток значений переменных появляется еще одна удобная возможность изменения вида окна вкладки Данные. Если в меню Вид установить флажок рядом с командой Метки значений, численные значения переменных будут заменены заданными метками. Для того чтобы вернуться к численным значениям переменных, достаточно снять флажок рядом с командой Метки значений.

Пропуски

Параметр Пропуски используется очень редко, поскольку программа и так позволяет учитывать пропуски в данных. Необходимость в этом параметре возникает, когда требуется различать причины пропусков значений. Например, пропуск в данных может быть обусловлен тем, что респондент еще не опрошен, а может быть, он отказался отвечать на данный вопрос. Так, в отношении переменной вуз (предполагаемый для поступления вуз) необходимо различать тех учащихся, которых не успели опросить, и тех, которые еще не определились. Для еще не опрошенных учащихся мы будем оставлять пустую ячейку, а не определившихся обозначим цифрой 9. После щелчка на кнопке с многоточием, появляющейся при перемещении фокуса ввода в ячейку столбца Пропуски, открывается диалоговое окно Пропущенные значения, в котором можно задать обозначения

пропусков. Если поставить переключатель Отдельные пропущенные значения и ввести число, например 9, это значение наряду с пустыми ячейками не будет использоваться в дальнейшем при обработке переменной.

Столбцы

Параметр Ширина столбца, в отличие от своего соседа слева, требуется для всех переменных. С его помощью можно управлять шириной (в символах) столбцов вкладки Данные. Если вы зададите широкие столбцы, это позволит видеть имена переменных полностью, однако количество одновременно видимых переменных будет невелико. Напротив, узкие столбцы позволяют наблюдать на экране много переменных сразу, но их имена могут не уместиться в столбце целиком. Например, при ширине столбцов 3 символа на экране можно увидеть 33 переменные, а при ширине 8 символов — всего 12. Опытные пользователи знают, что работа с данными тем удобнее, чем больше их видно одновременно. Поэтому вы часто будете сталкиваться с задачей размещения на экране как можно большего объема данных с сохранением их удобочитаемости. Пожалуй, основной принцип, которым следует руководствоваться, следующий: чем меньше знаков занимает большинство значений некой переменной, тем меньшей следует задавать ширину соответствующего столбца. Для изменения параметра Ширина столбца используется счетчик. По умолчанию для всех переменных значение равно 8.

ШАГ 8

Для примера далее приведены действия по изменению ширины столбцов для переменных №, пол, класс:

1. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Ширина столбца, соответствующую переменной №, и несколько раз щелкните на нижней кнопке счетчика, пока не появится значение 4.
2. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Ширина столбца, соответствующую переменной пол, и несколько раз щелкните на нижней кнопке счетчика, пока не появится значение 6.
3. С помощью клавиш со стрелками переместите фокус ввода в ячейку столбца Ширина столбца, соответствующую переменной класс, и три раза щелкните на нижней кнопке счетчика. В ячейке должно появиться значение 5. Таким же образом подберите ширину оставшихся столбцов, проверяя результат на вкладке Данные.

Выравнивание

Параметр Выравнивание позволяет управлять расположением данных внутри ячейки. Доступ к раскрывающемуся списку возможен, только если фокус ввода оказался в соответствующей ячейке, — теперь можно выбрать один из трех типов выравнивания: По правому краю, По левому краю и По центру. По умолчанию значения числовых переменных выравниваются по правому краю, строковых — по левому. Для каждой переменной вы можете самостоятельно задать тип выравнивания.

Шкала измерения

Как и значение предыдущего параметра, значение параметра Шкала выбирается в раскрывающемся списке из трех доступных: Количественная, Порядковая и Номинальная. Каждое из этих значений несет дополнительную информацию о типе данных переменной:

- ▶ Значение Количественная указывается для количественных типов данных, допускающих арифметические операции. Примером может служить возраст человека: для него имеют смысл операции вычитания, сложения и деления. В файле ex01.sav это значение параметра Шкала можно указать для переменных тест1, ..., тест5, отметка1, отметка2.
- ▶ Значение Порядковая указывается для количественных данных, отражающих измеренное качество на уровне порядка и не допускающих применения к ним основных арифметических операций. Например, упорядочивание учащихся по сообразительности или кодирование ответов по степени выраженности (1 — «не нравится», 2 — «нейтрально», 3 — «нравится») — это измерения в порядковой шкале. К подобным типам можно отнести и многие дихотомические переменные (переменные, имеющие лишь одно из двух значений: 0 или 1, да или нет и т. п.).
- ▶ Значение Номинальная указывается для категориальных (неколичественных) типов данных, не отражающих количество измеряемого свойства. В файле ex01.sav к таким данным относятся переменные №, пол, класс, вуз, хобби.

Иногда может оказаться затруднительно сделать выбор между значениями Количественная и Порядковая. Не отчаивайтесь, если попадете в такую ситуацию: во всех статистических процедурах оба этих значения параметра Шкала сказываются на результатах анализа одинаково.

Чаще всего нет необходимости менять значение параметра Количественная, заданное по умолчанию. Тем более что выбор значения Номинальная ограничивает набор возможных методов обработки данных.

Роль

Параметр Роль позволяет задавать роли, которые могут быть использованы для предварительного выбора переменных для анализа. При открытии некоторых из диалоговых окон анализа переменные, которые удовлетворяют требованиям роли, будут автоматически отображаться в списках назначения. Доступные роли:

- ▶ Входная. Переменная будет использоваться в качестве входной (например, предиктор, независимая переменная);
- ▶ Целевая. Переменная будет использоваться в качестве выходной или целевой (например, независимая переменная);
- ▶ Двойного назначения. Переменная будет использоваться в качестве входной и выходной;
- ▶ Не создавать. Переменная не имеет назначения роли;
- ▶ Разделения. Переменная будет использоваться, чтобы разделить данные на отдельные выборки для обучения, испытания и проверки;
- ▶ Расщепления. Включены для двустороннего совмещения с SPSS Modeler.

По умолчанию всем переменным, включая данные из форматов внешних файлов и файлов данных из более ранних версий SPSS, назначается роль Входная. Назначение роли влияет только на диалоговые окна, которые поддерживают назначение роли, и не оказывает влияния на командный синтаксис.

Задание некоторых ролей ограничивает использование переменной для анализа. Поэтому оставим для всех переменных роль, заданную по умолчанию: Входная.

Ввод данных

Ввод самих данных гораздо проще, чем задание параметров переменных, поэтому самое трудное в этой главе уже позади. В данном разделе мы опишем два способа непосредственного ввода данных в программу, а также процедуру переноса данных из программы Excel. Но сначала рассмотрим действия, необходимые для сохранения файла.

В процессе ввода рекомендуется время от времени выполнять сохранение файла во избежание случайной порчи или утери введенных данных. Обратите внимание на то, что перед вводом данных следует перейти на вкладку Данные (см. рис. 3.1).

ШАГ 9

В случае если вы сохраняете файл в первый раз, потребуется задать его имя:

1. В меню Файл выберите команду Сохранить как. На экране появится диалоговое окно Сохранить данные как.

Если в этом диалоговом окне нажать кнопку Сохранить, будет сохранено все содержимое файла данных, а если нажать кнопку Переменные, появится возможность выбрать переменные для сохранения.

2. Введите имя файла и щелкните на кнопке Сохранить.

Перечисленные действия позволяют сохранять файлы лишь в текущей папке. Если вы создали рабочий каталог, текущей является папка D:\SPSSWORK, если нет — папка Мои документы. При необходимости вы можете сохранить файл в любой другой папке обычным способом.

ШАГ 10

Если ваш файл уже был сохранен хотя бы один раз, достаточно выбрать команду Файл ► Сохранить или щелкнуть на кнопке Сохранить панели инструментов.

Существует два метода непосредственного ввода данных в ячейки электронной таблицы: *по строкам (переменным)* и *по столбцам (объектам)*. Предпочтительным один из них оказывается в зависимости от конкретных данных.

Ввод данных по переменным предполагает последовательное заполнение всех строк одного столбца значениями одной переменной, затем другой переменной и т. д. Вы вводите значение в левую верхнюю ячейку окна, затем нажимаете клавишу ↓ или Enter, переходя на следующую строку, заполняете еще одну ячейку и продолжаете до тех пор, пока не заполните весь первый столбец. Затем заполняете столбцы, соответствующие второй, третьей и всем остальным переменным. Таким образом, визуальное введение данных по переменным осуществляется по вертикали.

Ввод данных по объектам заключается в последовательном заполнении каждой строки значениями всех переменных. Сначала значение вводится в левую верхнюю ячейку окна, затем с помощью клавиши → или Tab выполняется переход в соседний столбец (переменную), вводится значение переменной, и так продолжается до тех пор, пока вся строка не будет заполнена. После этого тем же способом заполняются вторая, третья и все остальные строки. В отличие от предыдущего способа ввод данных по объектам осуществляется по горизонтали.

Некоторым пользователям привычнее вводить, хранить и редактировать данные в электронной таблице программы Excel. Следует отметить, что программа SPSS лучше приспособлена для этого. Тем не менее перенос данных из Excel в SPSS не составляет труда,

если в таблице данных программы Excel строки соответствуют объектам, а столбцы — переменным. Перенос можно осуществить двумя способами.

Первый способ самый простой. Он предполагает предварительную подготовку специального файла Excel. Столбцы этого файла должны соответствовать переменным, строки — объектам. В первой строке должны содержаться имена переменных, удовлетворяющие требованиям SPSS (см. ранее). Достаточно открыть такой файл Excel из редактора данных SPSS, чтобы переместить содержащиеся в нем данные в SPSS. Для этого при помощи последовательности команд Файл ► Открыть ► Данные необходимо в диалоговом окне Открыть данные в списке Тип файла выбрать пункт Excel (*.xls).

Второй способ переноса данных из Excel в SPSS несколько сложнее и сводится к копированию значений переменных в таблице Excel и вставке их в таблицу редактора данных SPSS. Сначала необходимо подготовить структуру файла исходных данных в SPSS, как было описано ранее. Прежде всего обращайте внимание на те переменные, которые имеют буквенные значения. Для них вместо принятого по умолчанию числового типа переменной необходимо задать строковый тип, установив переключатель в окне Тип переменной (см. рис. 3.3) в положение Текстовая. После задания структуры данных, не закрывая подготовленный файл SPSS, откройте файл Excel, содержащий данные. Выделите и скопируйте при помощи мыши необходимый блок данных (только значения переменных) в таблице Excel (если эта процедура вызывает у вас затруднения, изучите сначала раздел «Редактирование данных»). Перейдите к файлу SPSS и выделите верхнюю левую ячейку пока еще пустой таблицы данных и в меню Правка выберите команду Вставить. Таблица SPSS будет заполнена теми данными, которые вы ранее скопировали из таблицы Excel. Сходным образом можно также пополнять данные, перенося их из Excel в SPSS и наоборот.

Редактирование данных

Иногда у вас может возникнуть необходимость изменить данные, уже введенные в файл. Мы рассмотрим, каким образом скорректировать содержимое ячейки, создать новый объект (строку) или переменную (столбец), а также скопировать, удалить и вставить данные.

Изменение содержимого ячейки

Щелкните на ячейке, содержимое которой вы намерены отредактировать, введите новое значение, а затем перейдите в любую соседнюю ячейку с помощью клавиш Enter, Tab или клавиш со стрелками.

Вставка нового объекта

Если вам необходимо вставить новый объект (строку) между двумя соседними строками, щелкните сначала на нижней из них, а затем на кнопке Вставить наблюдения панели инструментов. В результате будет создана пустая строка, а номера строк, находящихся ниже, увеличатся на единицу.

Вставка новой переменной

Чтобы вставить новую переменную между двумя соседними, щелкните сначала на правой из них, а затем на кнопке Вставить переменную. Будет создан пустой столбец, а все переменные, находящиеся справа, окажутся сдвинутыми на один столбец.

Копирование и вырезание содержимого ячеек

Для того чтобы скопировать или вырезать содержимое одной или нескольких ячеек, сначала необходимо их выделить. Активная ячейка (то есть ячейка, в которой находится фокус ввода) всегда является выделенной. Чтобы выделить группу ячеек, нажмите левую кнопку мыши на угловой ячейке будущей группы и, удерживая ее, перетащите указатель к противоположной угловой ячейке группы, после чего кнопку отпустите. Если вам необходимо выделить целиком строку или столбец, щелкните на нужном объекте или имени переменной соответственно.

Когда ячейки выделены, в меню Правка выберите команду Копировать или Вырезать. В обоих случаях содержимое ячеек будет скопировано в буфер обмена, однако команда Копировать оставит исходные ячейки нетронутыми, а команда Вырезать очистит их.

Вставка ячеек

Для того чтобы вставить предварительно скопированные в буфер обмена данные, нужно переместить фокус ввода в левую верхнюю ячейку группы, в которую будет осуществляться вставка, и в меню Правка выбрать команду Вставить. Следует помнить о двух неприятностях, связанных с операцией вставки:

- ▶ Если в области, в которую планируется вставить данные, уже содержатся какие-либо другие данные, после выполнения вставки они будут утеряны.
- ▶ При копировании и вставке данных их смысл не анализируется программой, и вы можете случайно заменить значения одной переменной значениями любой другой переменной. Последствия такой вставки могут быть весьма плачевными, поскольку разобраться в возникшей путанице, как правило, весьма непросто.

Приведем несколько рекомендаций, которые могут быть полезными при выполнении вставки:

- ▶ Перед удалением и вставкой важных фрагментов данных сохраняйте текущий файл. В случае ошибки вы сможете открыть его снова.
- ▶ По возможности создавайте новые переменные и строки для вставки данных.
- ▶ Как правило, вам придется иметь дело с копированием, вырезанием и вставкой столбцов (переменных) или строк (объектов) целиком. Однако особую аккуратность следует проявлять при вставке *блоков* столбцов или строк.
- ▶ Будьте особенно внимательны при вставке значений переменных разных типов, например, не допускайте вставки значений строковых переменных в столбцы, соответствующие числовым переменным. Это может привести к трудно идентифицируемым ошибкам в процессе статистического анализа.

Поиск данных

Очень удобным вспомогательным средством работы с данными является функция поиска. Выделив заголовок (имя) переменной, в меню Правка выберите команду Найти или щелкните на кнопке Поиск панели инструментов. На экране появится диалоговое окно Найти и заменить — Данные, с помощью которого можно найти заданное значение и (или) заменить его другим значением (рис. 3.5).

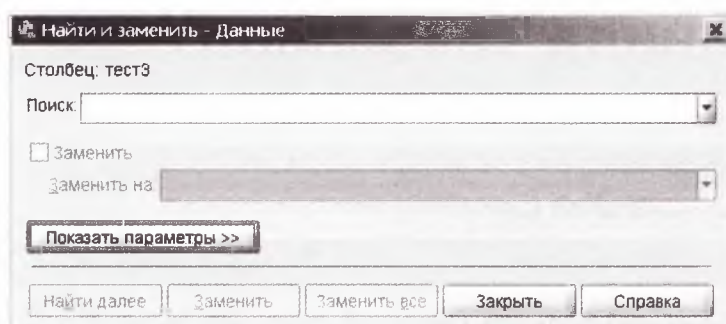


Рис. 3.5. Диалоговое окно поиска

Шаг 11

Попробуем, к примеру, найти в файле ex01.sav ученика с номером 93.

1. На вкладке Данные щелкните на имени переменной №, затем в меню Правка выберите команду Найти или щелкните на кнопке Поиск панели инструментов.
2. В поле Поиск открывшегося диалогового окна Найти и заменить — Данные введите число 93. Щелкните на кнопке Найти далее, чтобы найти ученика с номером 93.

После завершения поиска щелкните на кнопке Отмена, чтобы закрыть диалоговое окно.

В случае если вы обнаружите недопустимые или неверные значения какой-либо из переменных, можете применить приведенный алгоритм с установкой флажка Заменить и указанием нужного значения в окне Заменить на для исправления ошибки.

Пример файла данных

Файл данных ex01.sav содержит информацию об учащихся трех классов. Переменные для этого файла перечислены в табл. 3.2.

Таблица 3.2. Переменные файла данных ex01.sav

Переменная	Описание
№	Идентификационный номер ученика
пол	Пол ученика (1 — ЖЕН, 2 — МУЖ)
класс	Класс, в котором учится школьник (1 — «А», 2 — «Б», 3 — «В»)
пред	Предполагаемый для поступления вуз (1 — гуманитарный, 2 — экономический, 3 — технический, 4 — естественнонаучный)
хобби	Внешшкольное увлечение ученика (1 — спорт, 2 — компьютеры, 3 — искусство)
тест1	Показатель теста «Счет в уме»
тест2	Показатель теста «Числовые ряды»
тест3	Показатель теста «Словари»

продолжение ↗

Таблица 3.2 (продолжение)

Переменная	Описание
тест4	Показатель теста «Осведомленность»
тест5	Показатель теста «Кратковременная память»
отметка1	Средний балл отметок за 10-й класс
отметка2	Средний балл отметок за 11-й класс

Вкладки Данные и Переменные окна редактора данных для файла ex01.sav показаны на рис. 3.6 и 3.7 соответственно.

	№	пол	класс	вуз	хобби	тест1	тест2	тест3	тест4	тест5	отметка1	отметка2
1	1	МУЖ	Б	ЕСТ_Н	искусс	6	7	13	10	14	3,90	4,20
2	2	МУЖ	А	ЕСТ_Н	спорт	8	9	10	11	11	3,55	3,95
3	3	МУЖ	В	ТЕХН	компь	10	6	10	8	9	3,75	4,65
4	4	ЖЕН	В	ГУМ	компь	13	9	10	12	6	3,85	3,95
5	5	МУЖ	Б	ТЕХН	искусс	12	8	12	18	12	4,20	3,90
6	6	ЖЕН	В	ЭКОН	искусс	12	15	17	11	11	4,25	4,25
7	7	ЖЕН	В	ЭКОН	искусс	6	7	11	16	13	4,45	4,35
8	8	ЖЕН	А	ГУМ	компь	13	11	10	10	10	3,80	3,90
9	9	ЖЕН	Б	ЕСТ_Н	искусс	9	12	14	9	15	3,90	4,00
10	10	ЖЕН	В	ЭКОН	искусс	5	9	13	13	12	4,25	3,75

Рис. 3.6. Фрагмент данных файла ex01.sav

	Имя	Тип	Ширина	Десятичн	Метка	Значения	Пропущенные	Ширина	Выравнивание
1	№	Числовая	3	0		Нет	Нет	4	По правому
2	пол	Числовая	3	0		{1, ЖЕН}	Нет	4	По правому
3	класс	Числовая	3	0		{1, А}	Нет	5	По правому
4	вуз	Числовая	3	0		{1, ГУМ}	Нет	4	По правому
5	хобби	Числовая	3	0		{1, спорт}	Нет	5	По правому
6	тест1	Числовая	3	0	счет в уме	Нет	Нет	5	По правому
7	тест2	Числовая	3	0	числовые ряды	Нет	Нет	5	По правому
8	тест3	Числовая	3	0	словарь	Нет	Нет	5	По правому
9	тест4	Числовая	3	0	осведомленность	Нет	Нет	5	По правому
10	тест5	Числовая	3	0	кратковременна	Нет	Нет	5	По правому
11	отметка1	Числовая	3	2		Нет	Нет	8	По правому
12	отметка2	Числовая	3	2		Нет	Нет	8	По правому

Рис. 3.7. Переменные файла ex01.sav

4 Управление данными

53	Знакомство с возможностями управления данными
55	Получение информации о файле
56	Обработка пропущенных значений
57	Преобразование данных
61	Выбор наблюдений для анализа
64	Перекодировка в новую переменную
68	Перекодирование существующей переменной
70	Сортировка наблюдений
71	Объединение данных разных файлов
75	Агрегирование данных
77	Реструктурирование данных:
	▶ Преобразование групп переменных в группы значений
	▶ Преобразование групп значений в группы переменных

Материал этой главы посвящен организации эффективной работы с исходными данными. Описанные здесь операции весьма полезны в большинстве случаев обработки и анализа данных, так как практически всегда существует необходимость в предварительной подготовке и преобразовании исходных данных. Поэтому изложенные рекомендации по форматированию данных помогут вам работать с программой гораздо свободнее.

Знакомство с возможностями управления данными

В процессе работы вам могут понадобиться преобразованные данные, являющиеся результатом некоторых действий над исходными данными файла. К примеру, для вашего исследования может представлять интерес среднее значение или сумма баллов по нескольким тестам для каждого учащегося, ранг каждого ученика по успеваемости и т. п. Иногда желательно упорядочить данные файла по какому-либо признаку, например по результатам выполнения некоего задания. Нередко возникает необходимость обработки не всех данных файла, а лишь их подмножества, выделяемого по определенным критериям (например, по полу, классу, успеваемости и пр.). Существует и обратная задача: если данные хранятся в нескольких небольших файлах, может возникнуть потребность в их объединении для последующего анализа.

Перечисленные проблемы указывают на то, что для регулярной аналитической работы недостаточно умения вводить данные и применять к ним статистические процедуры. Возникает задача эффективного управления. Способы решения этой задачи бывают

весьма нетривиальными, и исчерпывающий рассказ о них не вполне соответствовал бы теме книги. Тем не менее представленного в этой главе материала вполне достаточно для того, чтобы научиться свободно манипулировать данными.

Несмотря на то что навыки управления данными приходят с опытом и требуют некоторого терпения, обязательно освоите их. Это придаст процессу исследования гибкость, простоту и легкость. Тогда выполнение статистических процедур, казавшихся сложными и громоздкими, станет для вас интуитивно понятным.

Мы рассмотрим следующие основные команды управления данными:

- ▶ команда Информация о файле данных позволяет получить сведения о переменных как открытого, так и любого внешнего файла данных SPSS: имена, метки имен и значений;
- ▶ команда Преобразовать ▶ Заменить пропущенные значения, как ясно из ее названия, работает с отсутствующими значениями переменных;
- ▶ команда Преобразовать ▶ Вычислить переменную позволяет путем вычислений создавать новые переменные на основе существующих;
- ▶ команда Преобразовать ▶ Ранжировать наблюдения позволяет создать новую переменную путем ранжирования значений существующей переменной;
- ▶ с помощью команды Данные ▶ Отобрать наблюдения можно выбрать подмножество наблюдений для дальнейшего анализа;
- ▶ команды Преобразование ▶ Перекодировать в другие переменные и Преобразование ▶ Перекодировать в те же переменные предназначены для изменения способа кодирования переменных, например уменьшения количества возможных значений;
- ▶ команда Данные ▶ Сортировать наблюдения позволяет упорядочить объекты в соответствии с назначенными критериями;
- ▶ команды подменю Слить файлы меню Данные используются для добавления в файл новых переменных или наблюдений из другого файла;
- ▶ команда Агрегировать данные меню Данные позволяет создавать такие значения переменных, каждое из которых представляет собой результат объединения группы исходных значений, например их среднее;
- ▶ команды подменю Реструктурировать меню Данные позволяют производить сложные манипуляции со структурой файла данных, например преобразовывать набор переменных в группы значений одной переменной или, наоборот, группы значений одной переменной — в набор переменных.

Все главы книги, начиная с этой, построены таким образом, что они практически не связаны друг с другом. После вводной части приведено описание пошаговых процедур, которые позволяют открывать файлы с данными и получать доступ к диалоговым окнам нужных статистических операций. Затем следуют описания этих диалоговых окон, а заключительные части глав посвящены интерпретации выводимых данных и их распечатке. Как правило, действия, связанные с вызовом статистической операции, приводятся на шаге 4, а на шаге 5 рассказывается о том, как задать параметры операции и завершить анализ данных. Иногда имеется несколько альтернативных способов выполнения шагов 4 или 5, в этом случае они обозначаются 4а, 4б, 5а, 5б и т. д.

Структура данной главы несколько отличается от описанной, поскольку в ней речь идет не о статистических операциях, а о средствах организации данных. После выполнения трех подготовительных шагов будет детально рассмотрена работа с каждой из ко-

манд управления данными. Главу завершит обзор необходимых завершающих действий с программой. В отличие от глав, посвященных анализу данных, здесь вы не найдете заключительного раздела, относящегося к выводу результатов.

Прежде чем двигаться дальше, выполните три подготовительных шага.

ШАГ 1

Подготовьте новый или существующий файл данных. Все необходимые для этого действия описаны в главе 3.

ШАГ 2

Запустите программу SPSS при помощи значка на Рабочем столе или выбрав в главном меню Windows команду Пуск ► Программы ► IBM SPSS Statistics ► IBM SPSS Statistics 20 и в открывшемся после запуска программы диалоговом окне выбора режима работы щелкните на кнопке Отмена.

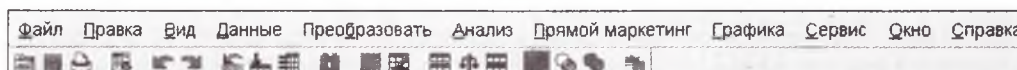
После выполнения этого шага на экране появится окно редактора данных SPSS.

ШАГ 3

Откройте файл данных, с которым вы намерены работать (в нашем случае это файл ex01.sav). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Данные или щелкните на кнопке Открыть данные панели инструментов.
2. В открывшемся диалоговом окне дважды щелкните на имени ex01.sav или введите его с клавиатуры и щелкните на кнопке OK.

Независимо от того, открыта программа SPSS только что или какие-то процедуры уже выполнялись, в верхней части главного окна должна присутствовать строка меню (она показана ниже). Пока строка меню на экране, доступны все команды анализа данных. При этом окно с данными видеть не обязательно.



При определенных исходных настройках программы SPSS одновременно с открытием файла данных открывается окно вывода. Чтобы вернуться к главному окну и стандартной строке меню, щелкните на кнопке сворачивания или восстановления текущего окна.

Получение информации о файле

Команда Информация о файле данных позволяет получить сведения о переменных файла данных SPSS. Это, пожалуй, самая простая из рассматриваемых команд. Она позволяет быстро получать список имен переменных, их меток, меток их значений не только для открытого (рабочего) файла данных, но и для любого другого (внешнего) файла данных SPSS.

При помощи шага 4 вы получите информацию об открытом (рабочем) файле ex01.sav, а выполнение шага 4а обеспечит вас сведениями о файле TestIQ.sav.

ШАГ 4

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

В меню Файл выберите команду Информация о файле данных ► Рабочий файл.

В результате выполнения этой команды откроется окно вывода, содержащее две таблицы. В первой таблице перечислены все переменные файла с указанием их характеристик, а во второй — значения и метки номинативных переменных.

ШАГ 5

Находясь в окне вывода, полученном в результате выполнения шага 4, сделайте следующее:

1. В меню Файл выберите команду Информация о файле данных ► Внешний файл.
2. В открывшемся окне выбора файла найдите файл TestIQ и дважды щелкните на его имени.

После выполнения этого шага к содержимому окна вывода добавится информация о файле данных TestIQ.sav, содержащая две таблицы. В первой таблице приведены общие сведения о файле, во второй — список переменных и их метки. Таблицы со значениями и метками номинативных переменных нет, так как файл их не содержит.

После просмотра закройте окно вывода без сохранения, чтобы оно не мешало дальнейшей работе.

Обработка пропущенных значений

В процессе работы с программой SPSS вы нередко будете сталкиваться с проблемой отсутствующих данных. Обратимся к переменным из примера ex01.sav. Вполне вероятна ситуация, когда кто-либо из учеников отсутствовал при проведении тестирования, или не ответил на вопрос о внешкольном увлечении, или не определился с перспективой поступления в вуз. Подобные случаи приводят к тому, что в данных рабочего файла появляются *пропущенные значения*. Пропущенные значения не только мешают осмысливать данные, но и могут оказывать нежелательное влияние на результаты анализа. Некоторые статистические процедуры игнорируют объекты (строки), в которых содержится хотя бы одно пропущенное значение. Если, к примеру, из 35 наблюдений 13 имеют пропущенные значения разных переменных, то анализу будут подлежать немногим более 60 % данных файла, что, несомненно, исказит результаты.

Большинство статистических методов SPSS позволяют учитывать пропуски в данных двумя принципиально различными способами: *построчно* (listwise) и *попарно* (pairwise). При построчном учете пропусков SPSS перед выполнением операции проверяет строки (объекты) на наличие пропущенных значений и в случае обнаружения последних исключает соответствующие строки из анализа целиком. Этот способ позволяет получить наиболее корректные статистические результаты, однако потери данных при этом максимальны. При попарном учете пропусков обработка выполняется без дополнительных проверок и в процессе вычислений не выполняются только те операции, которые требуют наличия пропущенного значения. Таким образом, в анализе участвуют все введенные данные, но результаты анализа содержат погрешности.

Мы рекомендуем вам по возможности решать проблему пропущенных значений на этапе ввода и кодирования данных, а не полагаться на то, что SPSS сделает это за вас. В любом случае, чем больше пропусков в исходных данных, тем менее точны и корректны результаты анализа.

Для номинальной переменной проблема пропущенных значений решается легко: вы можете просто ввести для нее еще одну градацию, которая соответствует пропуску в данных. Для количественной переменной (метрической или порядковой), имеющей множе-

ство возможных значений, в SPSS предусмотрены специальные процедуры заполнения пропусков: в меню Преобразовать есть команда Заменить пропущенные значения. При всем соблазне ее использовать следует помнить, что результаты обработки данных с заменой пропусков фиктивными значениями, например средними, вряд ли могут вызвать доверие. Поэтому лучше на месте пропуска честно оставлять пустую ячейку. А вопрос о построчном или попарном учете пропусков решать отдельно для каждого конкретного метода анализа данных.

В справочной системе SPSS часто используются два термина: *системные (физически) пропущенные значения* (system missing values) и *пользовательские (логически) пропущенные значения* (user missing values). Под физически пропущенными значениями понимаются значения, не введенные в компьютер. В редакторе данных пустые ячейки, не содержащие значений, помечены точкой. Логически пропущенные значения — это специальные значения переменной, отражающие невозможность адекватного кодирования некоторой ситуации. Если, например, 1, 2 и 3 — тестовые оценки испытуемого, 8 означает, что тест не завершен, а 9 фиксирует неявку испытуемого, то значения 8 и 9 относятся к логически пропущенным, поскольку их нельзя интерпретировать как результаты теста.

Преобразование данных

Вычисления

Операция Вычислить позволяет с помощью вычислений создавать новые переменные на основе существующих данных файла. Так, среди данных файла ex01.sav исследователю может понадобиться средний балл результатов тестирования (тест1, ..., тест5) каждого учащегося. Для этого необходимо создать новую переменную, например, с именем тест_ср.



После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics). При активном окне редактора данных в меню Преобразовать выберите команду Вычислить переменную. На экране появится диалоговое окно Вычислить переменную (рис. 4.1).

Диалоговое окно Вычислить переменную содержит в левой части уже ставший привычным список доступных переменных. Над списком, в левом верхнем углу окна, имеется поле Вычисляемая переменная, в которое необходимо ввести имя создаваемой, или вычисляемой, переменной. В поле Числовое выражение вводится выражение, с помощью которого вычисляется новая переменная. Вы можете вводить в это поле символы с клавиатуры, через буфер обмена, а также пользоваться перечисленными далее вспомогательными элементами интерфейса окна, позволяющими формировать выражение.

На панели калькулятора содержатся кнопки цифр и десятичного разделителя, а также разнообразных операций. Щелчок на каждой из этих кнопок приводит к появлению соответствующего символа (или нескольких символов) в поле Числовое выражение. Разумеется, при составлении выражения вы можете обойтись без панели калькулятора, поскольку все представленные там символы можно ввести с клавиатуры. Кнопки основных операций перечислены в табл. 4.1.

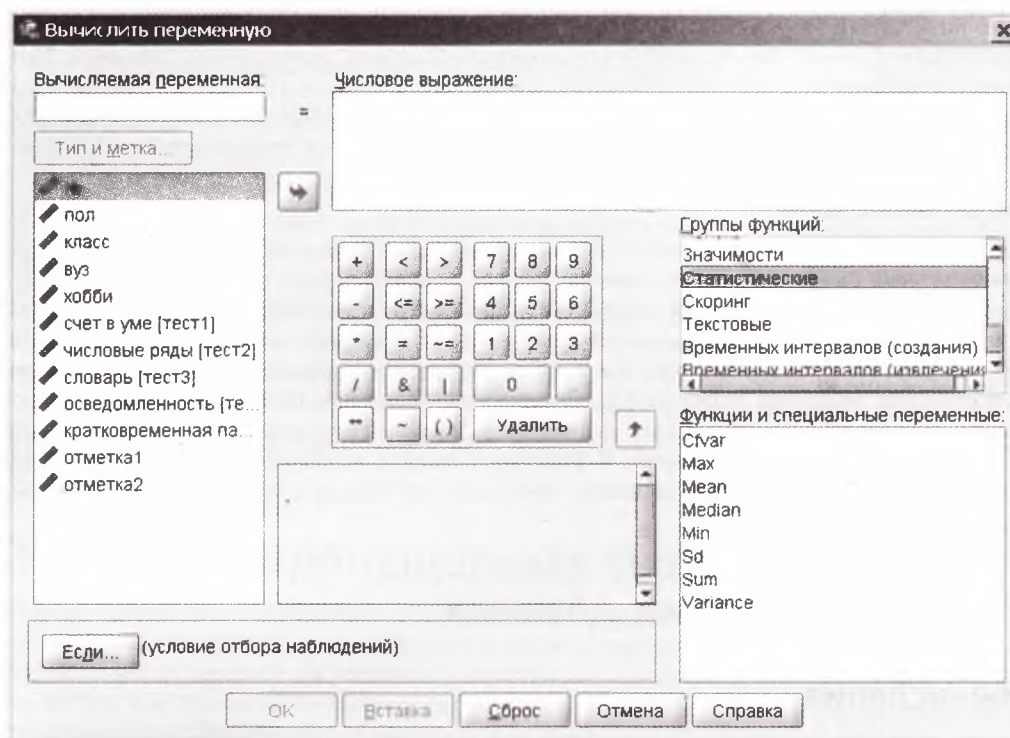


Рис. 4.1. Диалоговое окно Вычислить переменную

Таблица 4.1. Кнопки основных операций, доступных при создании переменных

Арифметические операции	Операции сравнения	Логические операции
+ сложение	< меньше	& И
- вычитание	> больше	ИЛИ
* умножение	<= меньше или равно	~ НЕ
/ деление	>= больше или равно	
** возведение в степень	= равно	
() очередность операций	!= не равно	

В списке Группы функций перечислены названия 20 групп функций, причем назначение большинства из них весьма туманно. Далее мы приводим описания нескольких групп функций, которые кажутся нам наиболее употребительными.

В списке Функции и специальные переменные перечислены операции, входящие в состав данной группы функций. Например, на рис. 4.1 выделена группа функций Статистические и в окне ниже перечислены доступные статистические функции. Для того чтобы

воспользоваться одной из них, необходимо ее выделить и при помощи стрелки, направленной вверх, перенести в окно Числовое выражение.

При щелчке на кнопке Если открывается диалоговое окно выбора наблюдений. Это окно используется в том случае, когда необходимо произвести вычисления только для части наблюдений (строк). Функция кнопки Если аналогична команде Отобразить наблюдения, описанной далее в этой главе.

Ниже перечислены основные группы функций, доступные при создании переменных:

- ▶ **Арифметические** — выполняют арифметические и алгебраические операции. Например, ABS(числвыр) возвращает абсолютное значение числового выражения в скобках (числвыр).
- ▶ **Пропущенных значений** — выполняют операции с пропущенными значениями. Например, NVALID(переменная[...]) возвращает количество действительных значений (не пропущенных) из списка переменных, имена которых заданы в квадратных скобках.
- ▶ **Статистические** — выполняют статистические операции. Например, MEAN(числвыр, числвыр[...]) возвращает среднее значение числовых выражений (числвыр, числвыр [...]), включающих в себя имена переменных, список которых указан в скобках.
- ▶ **Случайных чисел** — выполняют операции, генерирующие в качестве значений переменной случайные числа с заданными свойствами распределения. Например, Rv.Normal(срдн,стдоткл) возвращает массив нормально распределенных случайных чисел с заданным средним значением срдн и стандартным отклонением, равным stddev. Если Случнорм — целевая переменная, выражение NORMAL(0,3) создаст переменную Случнорм, значениями которой будут случайные числа, распределенные нормально с нулевым средним значением и стандартным отклонением, равным 3.

В следующем примере создадим новую переменную тест_ср. Она получается вычислением среднего для значений переменных тест1, ..., тест5.



Для выполнения следующих действий диалоговое окно Вычислить переменную должно быть открыто (см. рис. 4.2) в результате выполнения шага 4.

1. В поле Вычисляемая переменная введите имя тест_ср.
2. В окне Группы функций найдите и выделите щелчком мыши Статистические, а в окне Функции и специальные переменные выделите Mean (Среднее).
3. При помощи стрелки, направленной вверх, или двойным щелчком мыши на пункте Mean введите функцию Mean (?.?) в поле Числовое выражение.
4. В выражении MEAN(?.?) в поле Числовое выражение сотрите символы вопросов и запятую в скобках, оставив курсор между скобками.
5. Щелкните сначала на имени тест1, чтобы выделить его, а затем на кнопке с направленной вправо стрелкой (можно также дважды щелкнуть на имени тест1), чтобы ввести его в скобки в поле Числовое выражение.
6. Щелкните на кнопке , (или введите запятую с клавиатуры) и дважды щелкните на переменной тест2.
7. Повторите те же действия для переменных тест3, тест4 и тест5. В итоге должно получиться выражение: MEAN(тест1,тест2,тест3,тест4,тест5). Если это не так, внесите изменения. Всё выражение в поле Числовое выражение, в том числе имена переменных, можно целиком вводить с клавиатуры.
8. Щелкните на кнопке ОК.
9. В меню Файл выберите команду Сохранить.

В результате выполнения процедуры будет создана новая переменная `тест_ср`, которая разместится в крайнем правом столбце файла данных. Если вам удобнее держать эту переменную ближе к началу файла, можете воспользоваться операциями вырезания и вставки в нужную позицию.

После завершения процедуры программа возвращает вас в окно редактора данных, чтобы вы могли просмотреть содержимое новой переменной.

В более сложных вычислениях с привлечением разнообразных арифметических и других операций вам придется следить за приоритетами операций и при необходимости регулировать очередность вычислений с помощью скобок.

Ранжирование

Рассмотренная команда **Вычислить** позволяет создать новую переменную, значения которой вычисляются при помощи выражения, содержащего другие переменные. Команда **Преобразовать ▸ Ранжировать наблюдения** тоже позволяет создать новую переменную, значения которой — ранговые места наблюдений по заданной переменной. Эта процедура применяется, когда необходимо перейти от исходных значений переменной к рангам.

Команда ранжирования (перехода к рангам) выполняется при помощи диалогового окна, показанного на рис. 4.2. Для перехода к рангам достаточно задать имя переменной в поле **Переменные:** и щелкнуть на кнопке **ОК**. Появится новая переменная с заданным по умолчанию именем, содержащая ранги значений исходной переменной.

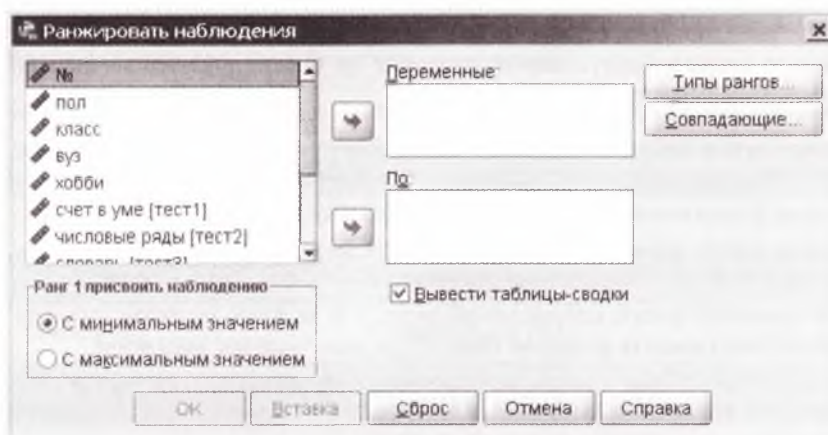


Рис. 4.2. Диалоговое окно Ранжировать наблюдения

Предположим, исследователь решил ранжировать всех учащихся по успеваемости за 11-й класс. Для этого ему необходимо перейти от исходных значений переменной `отметка2` к новой переменной, содержащей ранги учащихся по этой переменной.

Рассмотрим процедуру ранжирования на примере.

В результате выполнения этого шага в конце списка появится новая переменная с именем `Ротметка`, присвоенным по умолчанию. Это имя можно оставить или поменять на другое.

ШАГ 46

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

Для ранжирования переменной отметка2 выполните следующие действия:

1. В меню Преобразовать выберите команду Ранжировать наблюдения. На экране появится диалоговое окно, представленное на рис. 4.2.
2. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на кнопке со стрелкой.
3. Флажок Вывести таблицы-сводки можно снять, так как никаких результатов обработки, кроме новой переменной, получать не планируется.
4. По умолчанию переключатель Ранг 1 присвоить наблюдению установлен в положение С минимальным значением. При желании вы можете изменить порядок ранжирования на обратный, установив этот переключатель в положение С максимальным значением.
5. При щелчке на кнопке Типы рангов вам будет предложен дополнительный набор экстравагантных способов ранжирования, а при щелчке на кнопке Совпадающие — дополнительный набор способов обработки одинаковых (совпадающих) рангов. По умолчанию принято обычное присвоение одинаковым значениям переменной одного и того же среднего ранга (Средний). Для ранжирования по общепринятым правилам оставьте все те параметры, которые установлены по умолчанию, и щелкните на кнопке ОК.

Выбор наблюдений для анализа

В этом разделе мы отойдем от привычных вводных слов и сразу перейдем к пошаговым процедурам.



После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

В меню Данные выберите команду Отобрать наблюдения. На экране появится диалоговое окно, показанное на рис. 4.3.

Команда Отобрать наблюдения позволяет пользователю выбирать для обработки не все, а часть данных, удовлетворяющих заданным условиям. Поскольку необходимость в этом возникает довольно часто, команда Отобрать наблюдения является одной из самых востребованных при проведении исследований. Так, исследователю, использующему файл ex01.sav, могут понадобиться статистические сведения, касающиеся учащихся конкретного класса или определенного пола, и т. д. Это означает, что ему нужно указать программе, какие данные следует выделить для обработки. Именно для этого предназначена команда Отобрать наблюдения. После проведения анализа выбранной подгруппы вы можете вернуться к полному набору данных, установив в окне Отобрать наблюдения переключатель Выбрать в положение Все наблюдения. Если нужно создать файл данных, содержащий только выборку, можно удалить все исключенные объекты, а затем сохранить полученный файл с помощью команды Сохранить меню Файл. Номера выбранных наблюдений, в отличие от исключенных, после выполнения команды Отобрать наблюдения остаются незачеркнутыми, что позволяет легко находить их визуально. На

рис. 4.4 показан фрагмент окна редактора данных (вкладка Данные), на котором видны выбранные (девушки) и исключенные (юноши) наблюдения.

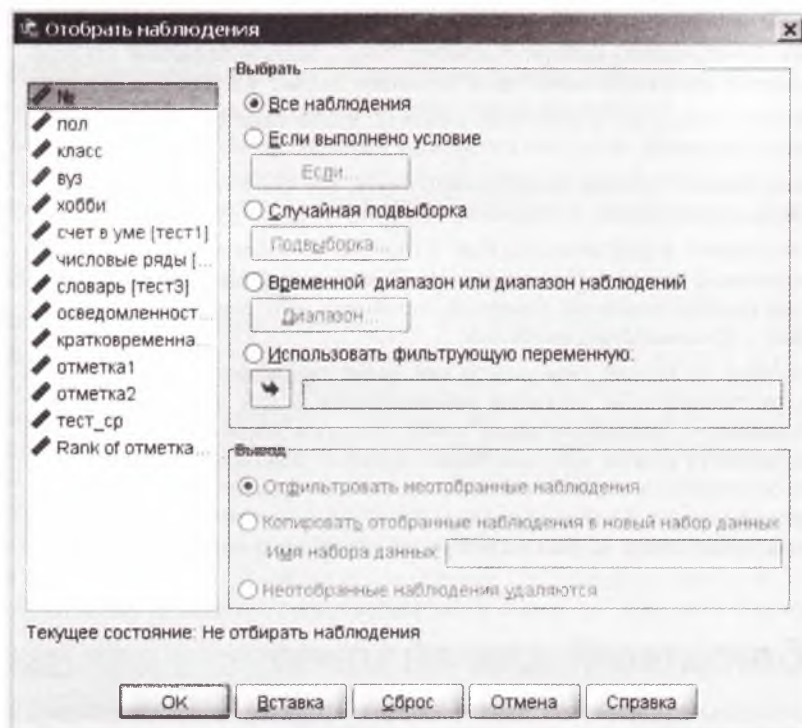


Рис. 4.3. Диалоговое окно Отобрать наблюдения

	№	пол	класс	вуз	хобби	тест1	тест2	те
1	1	МУЖ	Б	ЕСТ_Н	искус...	6	7	
2	2	МУЖ	А	ЕСТ_Н	спорт	8	9	
3	3	МУЖ	В	ТЕХН	компь...	10	6	
4	4	ЖЕН	В	ГУМ	компь...	13	9	
5	5	МУЖ	Б	ТЕХН	искус...	12	8	
6	6	ЖЕН	В	ЭКОН	искус...	12	15	

Рис. 4.4. Данные после отбора наблюдений для анализа

При выборе группы наблюдений SPSS создает специальную переменную с именем filter_\$, значение которой равно 1 для выбранных наблюдений и 0 для исключенных. Вы можете использовать эту переменную (например, для проведения сравнений между группами) или просто удалить ее.

В диалоговом окне Отобрать наблюдения имеется два переключателя. Переключатель Выбрать можно установить в 5 положений, из которых в данный момент нас интересуют только два:

- Все наблюдения фактически, предназначено для отмены операции, поскольку при этом выбранными полагаются все объекты файла;

- Если выполнено условие, позволяет задать условие отбора наблюдений с помощью кнопки Если. При щелчке на этой кнопке появляется диалоговое окно, очень похожее на окно, показанное на рис. 4.2. Чтобы задать условие отбора, вы можете использовать клавиатуру, панель калькулятора, список функций, а также буфер обмена. Описание функций основных кнопок приведено в табл. 4.1. Далее приведены примеры некоторых условий отбора.

пол = 1

Отбираются только девушки.

класс = 3

Отбираются ученики класса «В».

класс <= 3

Отбираются ученики классов «А» и «Б».

вуз >= 2 & вуз <= 3

вуз > 1 & вуз < 4

При задании любого из этих условий отбираются ученики, выбравшие экономические и технические вузы.

Чтобы ввести переменную в поле задания условия отбора, достаточно щелкнуть сначала на ней, а затем на кнопке со стрелкой. Знаки операций можно вводить как с клавиатуры, так и с панели калькулятора. В первом из примеров мы выберем для исследования всех учащихся-девушек, а во втором — учащихся, выбирающих экономические и технические вузы.



Сначала займемся девушками. При открытом окне Отобрать наблюдения (см. рис. 4.3) выполните следующие действия:

1. Переключатель Выбрать установите в положение Если выполнено условие и щелкните на кнопке Если.
2. В открывшемся окне сначала щелкните на имени переменной пол, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы ввести переменную пол в условие отбора. Далее на панели калькулятора щелкните на кнопке =, а затем — на кнопке 1, чтобы закончить составление условия отбора.
3. После создания условия отбора щелкните на кнопке Продолжить, чтобы закрыть первое диалоговое окно, и на кнопке ОК, чтобы закрыть второе диалоговое окно и вернуться в окно редактора данных.

После выполнения этого шага при любой обработке будут учитываться только данные для девушек. Чтобы сделать доступными все данные, достаточно в окне Отобрать наблюдения установить переключатель Выбрать в положение Все наблюдения.



Теперь выберем для исследований школьников, ориентированных на поступление в экономические и технические вузы. После выполнения шага 4в окно Отобрать наблюдения должно выглядеть так, как показано на рис. 4.3.

1. Переключатель Выбрать установите в положение Если выполнено условие и щелкните на кнопке Если.



2. В открывшемся окне сначала щелкните на имени переменной `вуз`, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы ввести переменную `вуз` в условие отбора. Далее на панели калькулятора щелкните на кнопке `>=`, введите число 2 и щелкните на кнопке `&`. Снова щелкните на имени переменной `вуз`, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы вторично ввести переменную `вуз` в условие отбора. На панели калькулятора щелкните на кнопке `<=` и введите число 3, закончив составление условия отбора.
3. После создания условия отбора щелкните на кнопке Продолжить, чтобы закрыть первое диалоговое окно, и на кнопке ОК, чтобы закрыть второе диалоговое окно и вернуться в окно редактора данных.

Перекодировка в новую переменную

Команда Преобразование ► Перекодировать в другие переменные создает новую переменную, однако ее значения определяются не как результат вычислений или ранжирования, а на основе замены множества значений существующей переменной небольшим числом категорий. Как правило, эта процедура применяется, когда необходимо разделить выборку наблюдений на подгруппы по диапазонам значений некоторой количественной переменной. При этом возможен учет значений одной или нескольких переменных.

Предположим, что на примере данных из файла `ex01.sav` требуется исходя из значений переменной `отметка2` (успеваемость в 11-м классе) создать две новые переменные: `Гр_усп1` и `Гр_усп2`. Переменная `Гр_усп1` должна делить всех учащихся на две группы: 1 — отметка ниже «4»; 2 — отметка «4» и выше. При помощи новой переменной `Гр_усп2` требуется разделить учащихся уже на 4 группы — в зависимости от успеваемости с учетом пола: выделить среди юношей и девушек тех, у кого средний балл отметки ниже «4» или «4» и выше. Таким образом, новая переменная `Гр_усп2` может принимать 4 значения: 1 — девушки, отметка `< «4»`; 2 — девушки, отметка «4» и выше; 3 — юноши, отметка `< «4»`; 4 — юноши, отметка «4» и выше.

Команда перекодирования в другие переменные выполняется с помощью двух диалоговых окон. Одно из них показано на рис. 4.5. В нем вы можете задать входную и выходную переменные.

Как видно на рис. 4.5, окно Перекодировать в другие переменные имеет вполне обычный вид: в его левой части находится список доступных переменных, а внизу — пять стандартных кнопок (см. главу 2).

Сначала создадим переменную `Гр_усп1`, разделяющую учащихся по успеваемости на 2 группы. Первое, что необходимо сделать в рассматриваемом нами примере, — выделить переменную `отметка2` и переместить ее в список Входная переменная ► Выходная переменная. Затем в поле Имя области Выходная переменная следует ввести имя новой переменной (в данном случае `Гр_усп1`). Щелчок на кнопке Изменить приведет к появлению переменной `Гр_усп1` в предыдущем списке: его содержимое будет иметь вид `отметка2 ► Гр_усп1`.

Щелчок на кнопке Старые и новые значения вызовет диалоговое окно, показанное на рис. 4.6. В нем можно задать значения новой переменной, которые будут соответствовать отдельным значениям или диапазонам значений старой переменной.

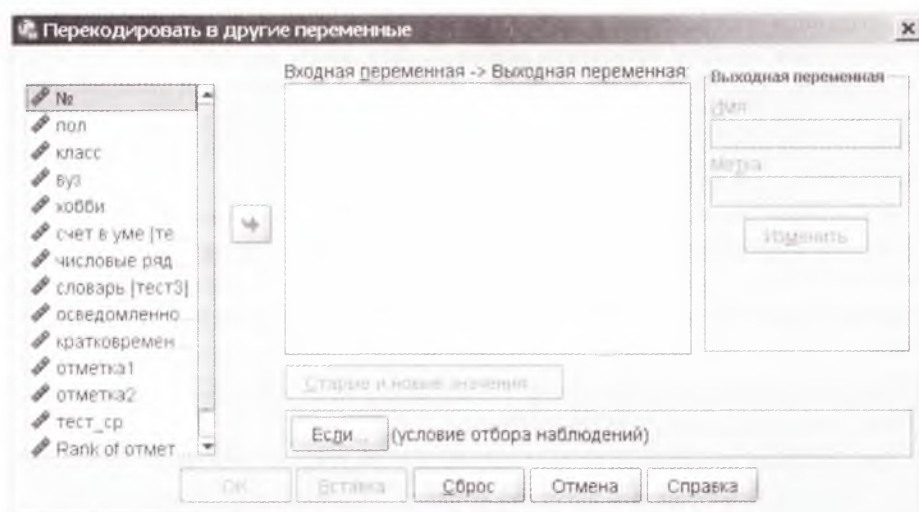


Рис. 4.5. Диалоговое окно Перекодировать в другие переменные

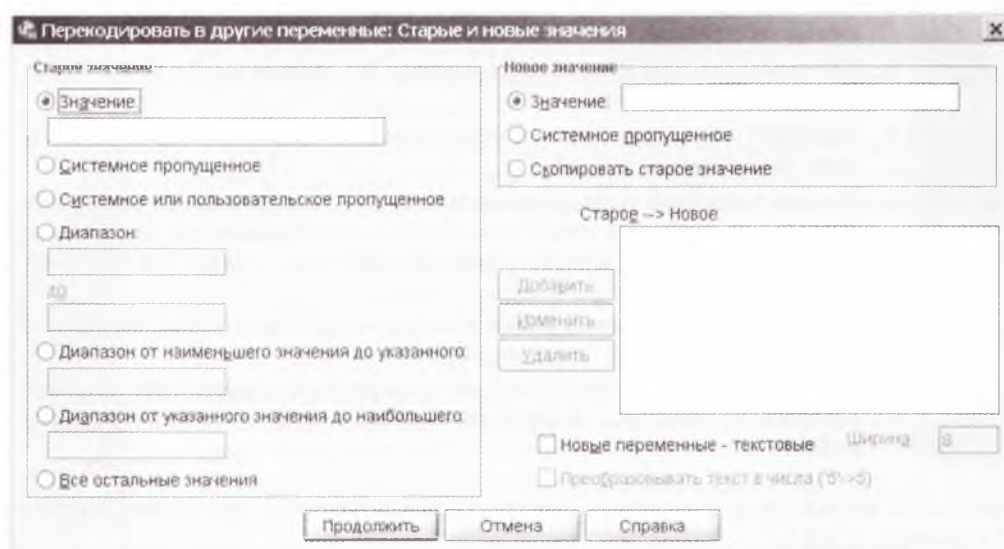


Рис. 4.6. Диалоговое окно Перекодировать в другие переменные: Старые и новые значения

Семь положений переключателя Старое значение в левой части окна позволяют устанавливать различные способы задания заменяемых значений исходной переменной.

Теперь требуется задать соответствие между диапазонами значений переменной *отметка2* и значениями новой переменной *Гр_усп1*: 1 — отметка «4» и выше, 2 — отметка до «4». Для того чтобы настроить окно на работу с диапазоном от заданного до наибольшего значения, установите переключатель Старое значение в положение Диапазон от указанного значения до наибольшего. Введите в поле, расположенное под ним, цифру 4, а в поле Значение переключателя Новое значение — цифру 1. После щелчка на кнопке Добавить

вы увидите созданное соответствие в поле Старое ► Новое: 4 through HIGHEST ► 1. Всем остальным значениям следует присвоить цифру 2. Для этого достаточно переключатель Старое значение установить в положение Все остальные значения и в поле Значение ввести новое значение 2. После щелчка на кнопке Добавить вы увидите последнее соответствие. Теперь щелкните сначала на кнопке Продолжить, затем в окне Перекодировать в другие переменные — на кнопке ОК.

Новыми переменными вы можете управлять так же, как и переменными, созданными обычным способом: назначать метки, вручную изменять значения, менять расположение в файле данных и т. п.

Обобщим процесс создания переменной Гр_усп1 в форме уже привычных нам пошаговых процедур.

ШАГ 4д

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

Для создания переменной Гр_усп1 выполните следующие действия:

1. В меню Преобразовать выберите команду Перекодировать ► Перекодировать в другие переменные. На экране появится диалоговое окно, показанное на рис. 4.5.
2. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на кнопке со стрелкой.
3. Для перехода в поле Имя нажмите клавишу Tab, введите имя Гр_усп1 и щелкните на кнопке Изменить.
4. Щелкните на кнопке Старые и новые значения, чтобы открыть диалоговое окно, показанное на рис. 4.6.
5. В левой части окна установите переключатель Старое значение в положение Диапазон от указанного значения до наибольшего. В поле под ним введите цифру 4, а в поле Значение в правом верхнем углу — цифру 1 и щелкните на кнопке Добавить.
6. Установите переключатель Старое значение в положение Все остальные значения и в поле Значение введите цифру 2.
7. Щелкните на кнопке Продолжить, чтобы вернуться в предыдущее диалоговое окно Перекодировать в другие переменные, в котором щелкните на кнопке ОК.

В результате выполнения этого шага в окне редактора данных появится новая переменная с именем Гр_усп1.

Теперь создадим новую переменную Гр_усп2, которая позволит разделить учащихся на 4 группы по успеваемости в зависимости от пола. В данном случае новая переменная, как и ранее Гр_усп1, создается исходя из значений переменной отметка2. Но при этом необходимо учесть значения переменной пол. Для этого в диалоговом окне Перекодировать в другие переменные (см. рис. 4.5) необходимо воспользоваться кнопкой Если. Функция этой кнопки аналогична команде Отбор наблюдений при установке переключателя Выбрать в положение Если выполнено условие (см. рис. 4.3). При помощи этой кнопки сначала выбираются девушки (пол = 1), и выполняются действия, аналогичные созданию переменной Гр_усп1. Затем при помощи кнопки Если выбираются юноши (пол = 2), и снова повторяются те же действия, но соответствующим группам присваиваются значения 3 и 4.

ШАГ 4е

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav – Редактор данных IBM SPSS Statistics).

Для создания переменной Гр_усп2 выполните следующие действия:

1. В меню Преобразовать выберите команду Перекодировать в другие переменные. На экране появится диалоговое окно, представленное на рис. 4.5. Если вы уже успели поработать с этой процедурой, нажмите кнопку Сброс.
2. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на кнопке со стрелкой.
3. Для перехода в поле Имя нажмите клавишу Tab, введите имя Гр_усп2 и щелкните на кнопке Изменить.
4. Щелкните на кнопке Если, чтобы открыть диалоговое окно Перекодировать в другие переменные: Отбор наблюдений.
5. В левой части окна выделите переменную пол, при помощи стрелки перенесите ее в правое окно и присвойте ей 1 (пол=1). Нажмите кнопку Продолжить.
6. Щелкните на кнопке Старые и новые значения, чтобы открыть диалоговое окно, показанное на рис. 4.6.
7. В левой части окна установите переключатель Старое значение в положение Диапазон от указанного значения до наибольшего. В поле под ним введите цифру 4, а в поле Значение в правом верхнем углу – цифру 1 и щелкните на кнопке Добавить.
8. Установите переключатель Старое значение в положение Все остальные значения и в поле Значение введите цифру 2. Щелкните на кнопках Добавить и Продолжить, чтобы вернуться в диалоговое окно Перекодировать в другие переменные.
9. Щелкните на кнопке Если, чтобы открыть диалоговое окно Перекодировать в другие переменные: Отбор наблюдений. В правом окне поменяйте значение переменной пол=2. Нажмите кнопку Продолжить и щелкните на кнопке Старые и новые значения.
10. В правой части в окне Старая->Новая выделите курсором строку 4 through HIGHEST (от 4 до наибольшего). В поле Значение в правом верхнем углу замените цифру 1 цифрой 3 и щелкните на кнопке Изменить.
11. Выделите строку ELSE->2. В поле Значение в правом верхнем углу замените цифру 2 цифрой 4 и щелкните на кнопке Изменить.
12. Щелкните на кнопке Продолжить, чтобы вернуться в предыдущее диалоговое окно Перекодировать в другие переменные, в котором щелкните на кнопке ОК.

В результате выполнения шага 4е в окне редактора данных появится новая переменная с именем Гр_усп2, созданная исходя из значений двух переменных: отметка2 и пол.

Сходным образом можно создавать новые переменные с учетом значений более двух исходных переменных. Кроме того, можно создавать сразу несколько новых переменных, если алгоритмы их получения из старых переменных совпадают.

Перекодирование существующей переменной

Иногда может возникнуть необходимость изменить кодирование какой-либо переменной. Это может быть обусловлено двумя причинами:

- ▶ Вы работаете одновременно с несколькими файлами данных, которые создавались разными людьми и содержат одни и те же переменные, но закодированные по-разному.
- ▶ В процессе исследования вам стало ясно, что существующее кодирование какой-либо из переменных следует изменить.

Обратимся к файлу ex01.sav. Как вы помните, в переменной пол значение 1 соответствует девушкам, а значение 2 — юношам. Вы можете столкнуться с тем, что у вас появится другой файл, в котором имеется переменная пол, но автор файла значением 1 закодировал юношей, а значением 2 — девушек. Чтобы избежать путаницы, вам придется перекодировать значения переменной пол в одном из двух файлов. Может оказаться и так, что значения, представляющие внешкольные увлечения учащихся, удобно расположить в порядке убывания их распространенности среди школьников, и если вы не подумали об этом сразу, то вам также придется заняться перекодированием.

Нередко при первоначальном кодировании вводится избыточное число уровней переменной. Как правило, эта избыточность проявляется на этапе проведения исследования. Типичной является следующая ситуация. Пусть имеется переменная вуз, отражающая выбираемый учащимся профиль образования. Первоначально задается 4 уровня этой переменной: 1 — гуманитарный, 2 — экономический, 3 — технический, 4 — естественнонаучный. При помощи процедуры, которая будет описана далее, можно объединить значения 1 и 2 в группу гуманитарных вузов, а значения 4 и 3 — в группу физико-технических вузов и присвоить этим группам значения 1 и 2 соответственно.

Перекодирование существующей переменной выполняется с помощью команды Перекодировать в те же переменные меню Преобразовать. После выбора этой команды на экране появится диалоговое окно, показанное на рис. 4.7.

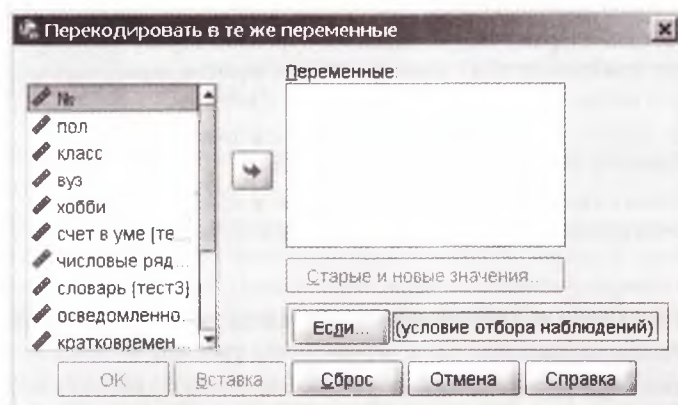


Рис. 4.7. Диалоговое окно Перекодировать в те же переменные

Операция перекодирования в те же переменные, как и предыдущая, управляется двумя диалоговыми окнами (второе показано на рис. 4.8), причем можно работать одновременно

но с несколькими переменными, однако для этого вам придется постоянно переключаться между диалоговыми окнами.

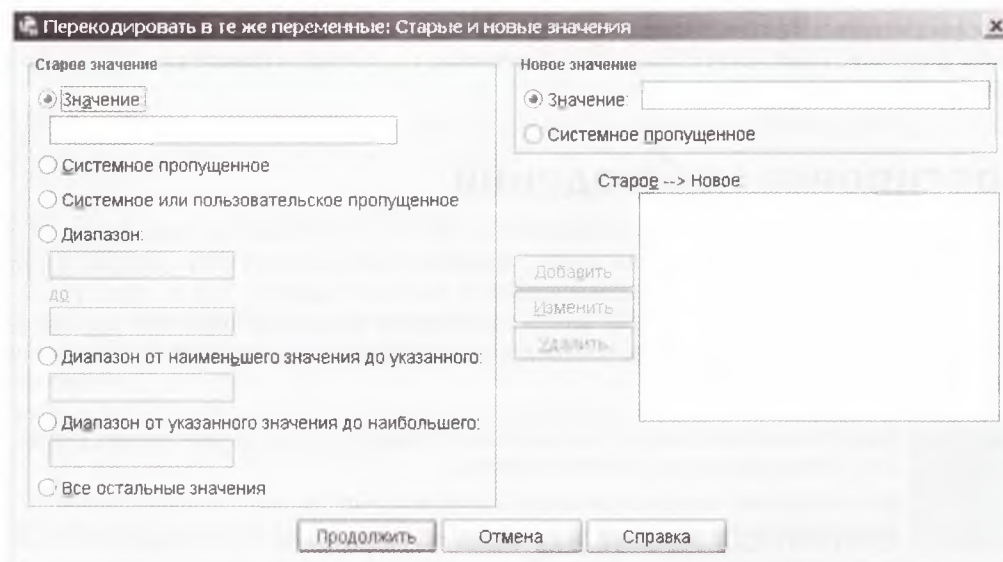


Рис. 4.8. Диалоговое окно Перекодировать в те же переменные: Старые и новые значения

Первым действием, которое необходимо выполнить в окне Перекодировать в те же переменные, является заполнение списка перекодировуемых переменных Переменные. Имена переменных содержатся в списке слева и выбираются при помощи кнопки со стрелкой. При щелчке на кнопке Старые и новые значения открывается диалоговое окно, показанное на рис. 4.8.

Как и в предыдущей процедуре, вы задаете одно или несколько старых и новых значений и помещаете их в списки Старое значение и Новое значение. В приведенном далее примере мы объединим 4 градации переменной вуз в две градации.

- ШАГ 4ж** После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).
- Для перекодирования переменной вуз выполните следующие действия:
1. В меню Преобразовать выберите команду Перекодировать в те же переменные. На экране появится диалоговое окно, показанное на рис. 4.7.
 2. Щелкните сначала на переменной вуз, чтобы выделить ее, затем на кнопке со стрелкой, чтобы переместить ее в список Переменные, и, наконец, — на кнопке Старые и новые значения, чтобы открыть диалоговое окно, показанное на рис. 4.8.
 3. Переключатель Старое значение установите в положение Диапазон, в верхнем поле введите значение 1, нажмите клавишу Tab, в нижнем поле введите значение 2, дважды нажмите клавишу Tab, чтобы перейти в поле Значение переключателя Новое значение. Введите там цифру 1 и щелкните на кнопке Добавить.
 4. Повторите предыдущее действие для старых значений 3, 4 и нового значения 2, затем щелкните на кнопке Продолжить, чтобы вернуться в исходное диалоговое окно, в котором щелкните на кнопке ОК.

Обратите внимание на то, что процедура перекодирования никак не влияет на метки значений. Другими словами, после выполнения перекодирования необходимо заново задать метки всем значениям, изменившим свой смысл. Действия с метками значений были рассмотрены в главе 3, и при необходимости вы можете обратиться к соответствующим разделам. Если хотите сохранить изменения, сделанные в файле данных, в меню Файл выберите команду Сохранить.

Сортировка наблюдений

Команда Сортировать наблюдения предназначена для реорганизации данных файла. Эта операция очень распространена, поскольку позволяет расположить информацию в том порядке, в котором это удобно исследователю в текущий момент. Так, к примеру, исследователь мог бы с помощью этой команды изменить исходный порядок следования данных, чтобы учащиеся были перечислены по классам или по убыванию успеваемости (значению переменной отметка2).

ШАГ 4з

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

В меню Данные выберите команду Сортировать наблюдения. На экране появится диалоговое окно, показанное на рис. 4.9.

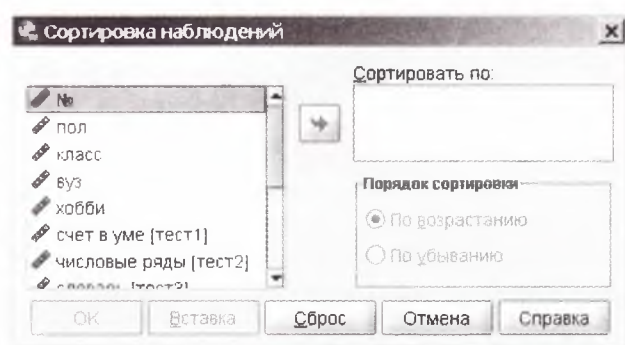


Рис. 4.9. Диалоговое окно Сортировка наблюдений

Параметры команды Сортировать наблюдения, в отличие от параметров предыдущих команд, настраиваются в одном небольшом диалоговом окне. Вы выбираете переменные в исходном списке и, устанавливая переключатель Порядок сортировки в положение По возрастанию или По убыванию, указываете порядок сортировки. Для текстовых переменных сортировка по возрастанию означает сортировку в алфавитном порядке.

При определении параметров сортировки можно задавать сразу несколько переменных. Это означает, что сначала данные будут отсортированы по значению первой выбранной переменной, затем наблюдения, имеющие одинаковые значения первой переменной, отсортируются по значению второй переменной и т. д. Например, если в список Сортировать по ввести переменные класс и пол, то все учащиеся будут перечислены по классам, а внутри каждого класса сначала будут указаны девушки, затем юноши.

В первом из двух приведенных далее примеров данные файла ex01.sav упорядочиваются по убыванию значения переменной отметка2; во втором примере реализована сортировка учащихся по классам и полу.

ШАГ 5з

При открытом диалоговом окне Сортировка наблюдений, показанном на рис. 4.9, щелкните сначала на переменной *отметка2*, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить ее в список Сортировать по. Далее установите переключатель Порядок сортировки в положение По убыванию и щелкните на кнопке ОК.

ШАГ 5и

При открытом диалоговом окне Сортировка наблюдений, показанном на рис. 4.9, выполните следующие действия:

1. Щелкните сначала на переменной *класс*, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить ее в список Сортировать по.
2. Повторите то же действие для переменной *пол* и щелкните на кнопке ОК.

Обратите внимание на то, что по умолчанию выполняется сортировка по возрастанию, поэтому никаких дополнительных действий для задания этого способа сортировки не требуется.

Исходный порядок следования учащихся можно восстановить, назначив сортировку по переменной № (по возрастанию).

Объединение данных разных файлов

Работа с данными нескольких файлов одновременно иногда способна сбить с толку даже опытных пользователей. Нередко файлы данных создаются при помощи разного программного обеспечения и имеют разные форматы. Это порождает различные проблемы их совместного использования, из которых мы рассмотрим наиболее типичную: дополнение *рабочего* файла SPSS содержимым *внешнего* файла. При этом существует две возможности дополнения данных рабочего файла SPSS:

- из внешнего файла Excel;
- из внешнего файла SPSS.

Мы рассмотрим оба варианта, но сначала сформулируем общие рекомендации:

- Если вы намерены добавлять переменные, убедитесь, что порядок следования наблюдений в рабочем и внешнем файлах одинаков.
- Если вы намерены добавлять объекты, убедитесь, что порядок следования переменных в рабочем и внешнем файлах одинаков.
- Настройте форматы всех переменных рабочего файла данных, чтобы они соответствовали данным внешнего файла (если добавляемая переменная содержит буквенные символы, то либо замените буквы числами во внешнем файле, либо измените тип переменной на строковую в рабочем файле).
- Перед добавлением данных создайте резервную копию рабочего файла, чтобы к ней можно было вернуться в случае неудачного переноса данных.

Следует иметь в виду, что чем больше соответствия между структурами рабочего и внешнего файлов, тем меньше вероятность ошибок при слиянии данных.

Если вы выполнили указанные рекомендации, то перенос данных из одного (внешнего) в другой (рабочий) файл не составит труда:

1. Откройте рабочий и внешний файлы с данными.
2. Во внешнем файле выделите необходимый для переноса блок данных.
3. Скопируйте его в буфер обмена.

4. Разверните свернутое окно SPSS с рабочим файлом на вкладке Данные.
5. Установите фокус ввода в левую верхнюю ячейку того места, куда вы намереваетесь поместить переносимые данные.
6. В меню Правка выберите команду Вставка.

После этого таблица редактора данных будет дополнена новыми данными. Проверьте корректность переноса данных. Если допущена ошибка, в меню Правка рабочего файла выберите команду Отменить. После возврата в исходное состояние проверьте свои действия и повторите попытку.

Указанная последовательность действий требует большого внимания и определенной сноровки и применима для небольших фрагментов данных. К счастью, существует более надежный способ объединения данных — *слияние* рабочего и внешнего файлов.

Слияние файлов допустимо, когда и рабочий, и внешний файлы созданы при помощи редактора данных SPSS и имеют одинаковые имена переменных (если мы добавляем объекты) или одинаковые количество и порядок следования наблюдений (если добавляем переменные). Таким образом, если внешним является файл Excel, то перед слиянием необходимо на его основе создать внешний файл SPSS (см. главу 3). Структура создаваемого файла должна быть максимально согласована со структурой рабочего файла. Для этого дополнительно к выполнению общих рекомендаций, изложенных ранее, нужно настроить форматы всех переменных файлов так, чтобы они были одинаковыми, и проверить идентичность имен переменных.

Хотя эти рекомендации не являются обязательными, все же следует иметь в виду, что чем больше соответствия между структурами файлов, тем меньше вероятность возникновения ошибок при слиянии. Особенное внимание следует уделять идентичности имен переменных, их форматам и расположению в файлах данных. Обратите внимание также на то, что при добавлении переменных одинаковый порядок следования наблюдений в файлах обязателен.

Добавление наблюдений

После завершения шага 3, описанного ранее в этой главе, у вас уже должен быть открыт файл ex01.sav.

Шаг 4. Для добавления наблюдений в меню Данные выберите команду Слить файлы ► Добавить наблюдения. На экране появится диалоговое окно, показанное на рис. 4.10.

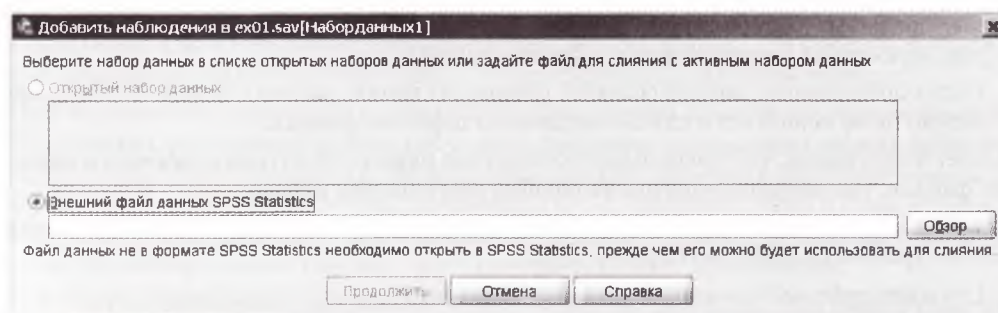


Рис. 4.10. Диалоговое окно Добавить наблюдения

С помощью диалогового окна **Добавить наблюдения** можно выбрать внешний файл данных, предназначенный для слияния с открытым (рабочим) файлом. Если нужный внешний файл есть в списке **Открытый набор данных**, то для доступа к нему достаточно щелкнуть на его имени.

Предположим, создан файл `ex01a.sav`, в котором содержатся данные о 10 новых учащихся с теми же переменными, что и в файле `ex01.sav`. В данном случае `ex01.sav` является рабочим файлом, а `ex01a.sav` — внешним. После задания имени файла щелкните на кнопке **Открыть**, а затем — **Продолжить**.

После того как внешний файл задан, на экране появляется диалоговое окно, показанное на рис. 4.11. Как видите, в заголовке окна отображено полное имя внешнего файла.

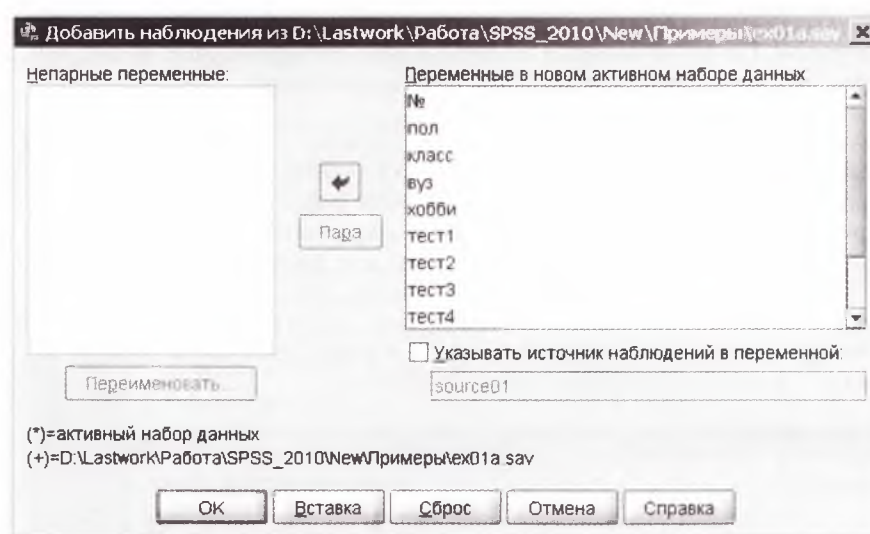


Рис. 4.11. Диалоговое окно **Добавить наблюдения**

Список **Непарные переменные** в данном случае пуст (как и следовало ожидать). Если же файлы содержат непарные переменные, все они попадают в этот список: переменные рабочего файла помечаются знаком * (звездочка), а переменные внешнего файла — знаком + (плюс). После того как вы убедитесь, что структуры рабочего и внешнего файлов совпадают, необходимо проверить на идентичность структуру каждой пары переменных. Перед щелчком на кнопке **ОК** можете при необходимости выполнить два дополнительных действия.

- ▶ Если вы не хотите, чтобы какая-либо из парных переменных присутствовала в новом файле, выделите ее щелчком в списке **Переменные в новом активном наборе данных** и перенесите ее в левое окно при помощи кнопки со стрелкой.
- ▶ По умолчанию все непарные переменные исключаются из нового рабочего файла. Если же вы хотите включить какую-либо из непарных переменных в рабочий файл, выделите ее в списке **Непарные переменные** и переместите в список **Переменные в новом активном наборе данных** щелчком на кнопке со стрелкой. Переменная войдет в рабочий файл, а ее отсутствующие значения будут восприниматься как пропущенные.



ТИП 3» Чтобы выполнить слияние файлов `ex01.sav` и `ex01a.sav`, в диалоговом окне, показанном на рис. 4.10, задайте имя файла `ex01a.sav`. Потом щелкните на кнопке **Продолжить** и в открывшемся диалоговом окне (см. рис. 4.11) — на кнопке **ОК**.

После выполнения этого шага в редакторе данных оказывается открытым новый рабочий файл с добавленными объектами, необходимо дать ему новое имя и сохранить. Содержимое старого рабочего файла и внешнего файла при этом останется прежним.

Добавление переменных

После завершения предыдущей пошаговой процедуры необходимо заново открыть файл ex01.sav.

ШАГ 4л

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

В меню Данные окна редактора данных выберите команду Слить файлы ► Добавить переменные. На экране появится диалоговое окно, показанное ранее на рис. 4.10, только заголовок окна будет другим: Добавить переменные.

Слияние с добавлением переменных очень похоже на слияние с добавлением наблюдений. Начало выполнения обеих пошаговых процедур одинаково: на экране появляется диалоговое окно, в котором вы указываете имя внешнего файла и щелкаете на кнопке Продолжить. При добавлении переменных после этого открывается диалоговое окно, показанное на рис. 4.12. В его заголовке отражено полное имя внешнего файла. Обратите внимание на то, что для демонстрации данной процедуры мы создали еще один специальный файл ex01b.sav, который содержит переменные идентификационного номера и класса для тех же наблюдений (учащихся), что и в файле ex01.sav, а также новую переменную тестб — результат дополнительного тестирования.

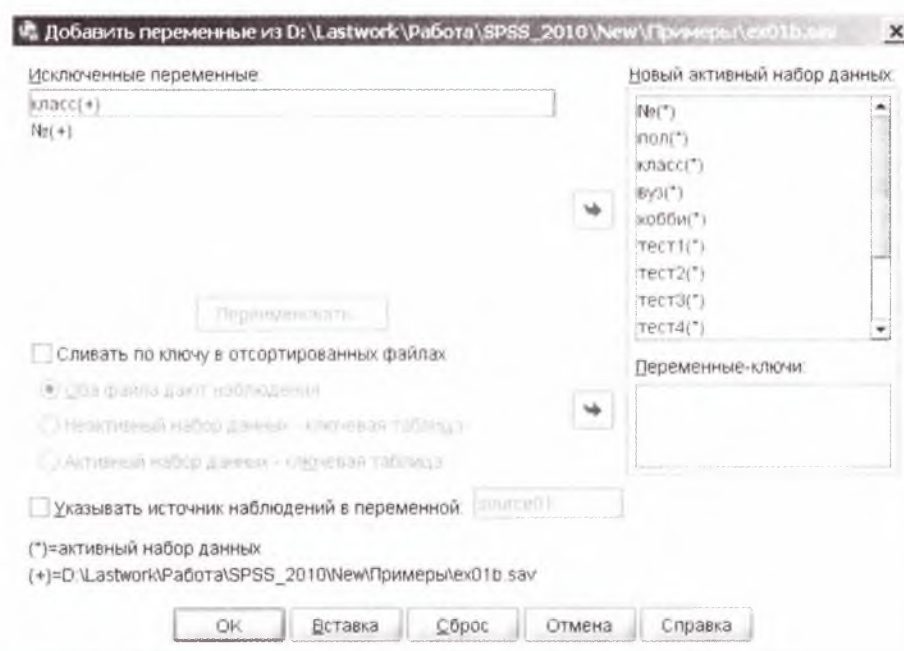


Рис. 4.12. Диалоговое окно Добавить переменные

Парные переменные попадают в список Исключенные переменные и помечаются символом + (плюс), указывающим на то, что они принадлежат внешнему файлу. Переменные, включенные в новый рабочий файл, перечисляются в списке Новый активный набор данных: переменные старого рабочего файла помечаются символом * (звездочка), а новые переменные — символом + (плюс).

Следующим и очень важным действием является задание ключевой переменной. Необходимо, чтобы порядок следования значений в ключевых переменных обоих файлов был одинаковым. В нашем примере в роли ключевой переменной рекомендуется использовать переменную № (идентификационный номер), поскольку ее значения уникальны для каждого наблюдения. Значения переменной № должны быть упорядочены, например, по возрасту. После того как вы убедитесь, что в обоих файлах ключевые переменные отсортированы одинаково, установите флажок Слить по ключу в отсортированных файлах, выделите ключевую переменную в списке Исключенные переменные и щелкните сначала на нижней кнопке со стрелкой, чтобы перенести ее в поле Переменные-ключи, а потом на кнопке ОК.

Далее приведен пример слияния рабочего файла ex01.sav с внешним файлом ex01b.sav, где в качестве ключевой переменной выступает переменная №, упорядоченная по возрасту.

ШАГ 5л

В диалоговом окне Добавить переменные, открытом после выполнения шага 4л, выполните следующие действия:

1. Задайте имя файла ex01b.sav, а потом щелкните на кнопке Продолжить. На экране появится диалоговое окно, показанное на рис. 4.12.
2. Установите флажок Слить по ключу в отсортированных файлах, в списке Исключенные переменные щелкните на переменной №, чтобы выделить ее, и щелкните на нижней кнопке со стрелкой. Имя № появится в списке Переменные-ключи.
3. Щелкните на кнопке ОК.
4. На экране появится предупреждение Слияние по ключу будет невозможно, если данные не были отсортированы по ключевой переменной в порядке возрастания. Щелкните на кнопке ОК.

Если у вас возникнут проблемы при выполнении этого примера, проверьте, действительно ли переменная № упорядочена по возрасту в обоих файлах.

После выполнения этого шага в редакторе данных оказывается открытым новый рабочий файл с добавленными объектами, необходимо дать ему новое имя и сохранить. Содержимое старого рабочего файла и внешнего файла при этом останется прежним.

Агрегирование данных

Агрегирование данных позволяет создавать такие значения переменных, которые представляют собой результат объединения группы исходных значений, например среднее. В процессе агрегирования задается группирующая переменная (например, номер респондента), каждое значение которой неоднократно встречается в исходных данных. Затем для каждого значения группирующей переменной вычисляются новые значения агрегируемых переменных, исходя из их исходных значений по заданной функции (например, среднее).

После выполнения этого шага в редакторе данных оказывается открытым новый рабочий файл с добавленными объектами, необходимо дать ему новое имя и сохранить. Содержимое старого рабочего файла и внешнего файла при этом останется прежним.

Добавление переменных

После завершения предыдущей пошаговой процедуры необходимо заново открыть файл ex01.sav.

ШАГ 4л

После выполнения шага 3 должно быть открыто окно редактора данных (ex01.sav — Редактор данных IBM SPSS Statistics).

В меню Данные окна редактора данных выберите команду Слить файлы ► Добавить переменные. На экране появится диалоговое окно, показанное ранее на рис. 4.10, только заголовок окна будет другим: Добавить переменные.

Слияние с добавлением переменных очень похоже на слияние с добавлением наблюдений. Начало выполнения обеих пошаговых процедур одинаково: на экране появляется диалоговое окно, в котором вы указываете имя внешнего файла и щелкаете на кнопке Продолжить. При добавлении переменных после этого открывается диалоговое окно, показанное на рис. 4.12. В его заголовке отражено полное имя внешнего файла. Обратите внимание на то, что для демонстрации данной процедуры мы создали еще один специальный файл ex01b.sav, который содержит переменные идентификационного номера и класса для тех же наблюдений (учащихся), что и в файле ex01.sav, а также новую переменную тестб — результат дополнительного тестирования.

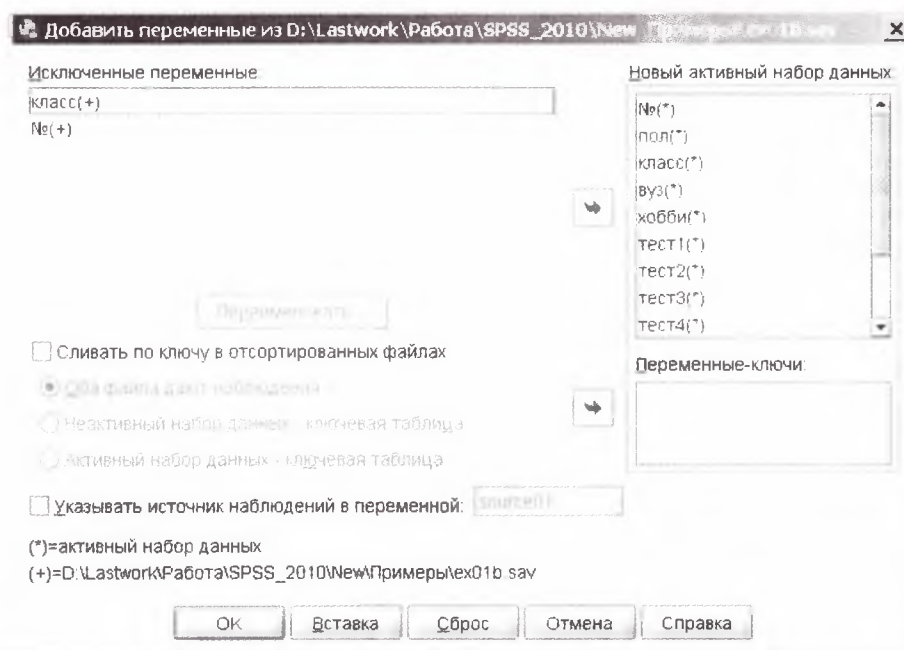


Рис. 4.12. Диалоговое окно Добавить переменные

Парные переменные попадают в список Исключенные переменные и помечаются символом + (плюс), указывающим на то, что они принадлежат внешнему файлу. Переменные, включенные в новый рабочий файл, перечисляются в списке Новый активный набор данных: переменные старого рабочего файла помечаются символом * (звездочка), а новые переменные — символом + (плюс).

Следующим и очень важным действием является задание ключевой переменной. Необходимо, чтобы порядок следования значений в ключевых переменных обоих файлов был одинаковым. В нашем примере в роли ключевой переменной рекомендуется использовать переменную № (идентификационный номер), поскольку ее значения уникальны для каждого наблюдения. Значения переменной № должны быть упорядочены, например, по возрасту. После того как вы убедитесь, что в обоих файлах ключевые переменные отсортированы одинаково, установите флажок Слить по ключу в отсортированных файлах, выделите ключевую переменную в списке Исключенные переменные и щелкните сначала на нижней кнопке со стрелкой, чтобы перенести ее в поле Переменные-ключи, а потом на кнопке ОК.

Далее приведен пример слияния рабочего файла ex01.sav с внешним файлом ex01b.sav, где в качестве ключевой переменной выступает переменная №, упорядоченная по возрасту.

ШАГ 5л

В диалоговом окне Добавить переменные, открытом после выполнения шага 4л, выполните следующие действия:

1. Задайте имя файла ex01b.sav, а потом щелкните на кнопке Продолжить. На экране появится диалоговое окно, показанное на рис. 4.12.
2. Установите флажок Слить по ключу в отсортированных файлах, в списке Исключенные переменные щелкните на переменной №, чтобы выделить ее, и щелкните на нижней кнопке со стрелкой. Имя № появится в списке Переменные-ключи.
3. Щелкните на кнопке ОК.
4. На экране появится предупреждение Слияние по ключу будет невозможно, если данные не были отсортированы по ключевой переменной в порядке возрастания. Щелкните на кнопке ОК.

Если у вас возникнут проблемы при выполнении этого примера, проверьте, действительно ли переменная № упорядочена по возрасту в обоих файлах.

После выполнения этого шага в редакторе данных оказывается открытым новый рабочий файл с добавленными объектами, необходимо дать ему новое имя и сохранить. Содержимое старого рабочего файла и внешнего файла при этом останется прежним.

Агрегирование данных

Агрегирование данных позволяет создавать такие значения переменных, которые представляют собой результат объединения группы исходных значений, например среднее. В процессе агрегирования задается группирующая переменная (например, номер респондента), каждое значение которой неоднократно встречается в исходных данных. Затем для каждого значения группирующей переменной вычисляются новые значения агрегируемых переменных, исходя из их исходных значений по заданной функции (например, среднее).

В качестве примера будем использовать файл `ex021.sav` (см. описание в главе 14). В нем значения переменных НАЧАЛО, СЕРЕДИНА, КОНЕЦ представлены для каждого респондента дважды (в две строки) в соответствии с переменной Отсрочка: без отсрочки (0) и с отсрочкой (1). Соответственно, номер каждого респондента (переменная N) повторяется дважды. Предположим, исследователь желает агрегировать значения переменных НАЧАЛО, СЕРЕДИНА, КОНЕЦ так, чтобы для респондента с неповторяющимся номером N каждое значение этих переменных представляло собой среднее двух исходных значений (с отсрочкой и без нее).

После выполнения шага 2 на экране должно быть окно редактора данных SPSS.

ШАГ 3м

Откройте файл данных, с которым вы намерены работать (в нашем случае `ex021.sav`). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Данные или щелкните на кнопке Открыть данные панели инструментов.
2. В открывшемся диалоговом окне дважды щелкните на имени `ex021.sav` или введите его с клавиатуры и щелкните на кнопке ОК.

ШАГ 4м

В меню Данные выберите команду Агрегировать данные, чтобы открыть окно, показанное на рис. 4.13.

Поле Группирующие переменные предназначено для помещения переменной, повторяющимся значениям которой должна соответствовать группа агрегируемых значений. А переменные, для которых агрегируются значения, переносятся в поле Итоги для переменной(ых). Для нашего примера группирующей переменной выступает номер респондента (N), а агрегируемые переменные — НАЧАЛО, СЕРЕДИНА, КОНЕЦ. В связи с тем, что для переменных Инт и Знач значения повторяющихся N идентичны, результаты агрегирования для них при вычислении среднего будут равны исходным значениям. Поэтому они тоже могут быть помещены в список Итоги для переменной(ых).

Группа переключателей Сохранить управляет судьбой агрегированных переменных. В нашем случае целесообразно создание нового набора данных, который после проверки результата агрегирования может быть сохранен.

Кнопка Функция позволяет задавать другие способы агрегирования, нежели принятое по умолчанию вычисление среднего, а кнопка Имя и Метка позволяет задавать агрегируемым переменным имена, отличные от присваиваемых по умолчанию.

На следующем шаге создадим новый набор данных с именем `ex021agg.sav`, содержащий агрегированные переменные НАЧАЛО, СЕРЕДИНА, КОНЕЦ: каждое новое их значение будет равно среднему двух исходных значений (с отсрочкой и без). Также включим в этот набор значения переменных Инт и Знач.

ШАГ 5м

1. В окне Агрегирование данных щелкните на переменной N, а затем на кнопке со стрелкой, чтобы перенести ее в поле Группирующие переменные.
2. Перенесите переменные Инт, Знач, НАЧАЛО, СЕРЕДИНА, КОНЕЦ в поле Итоги для переменной(ых). Для этого нажмите на клавиатуре клавишу Ctrl и, удерживая ее, щелкните последовательно на именах этих переменных для их выделения. Затем щелкните на соответствующей кнопке со стрелкой.

3. Установите переключатель Сохранить в положение Создать новый набор данных, содержащий только агрегированные переменные и задайте его имя ex021agg.
4. Для выполнения команды щелкните по кнопке ОК.

В результате выполнения последнего шага создан новый набор данных ex021agg.sav, в котором группирующая переменная N содержит неповторяющиеся номера, а каждое значение переменных НАЧАЛО, СЕРЕДИНА и КОНЕЦ представляет собой среднее двух исходных значений соответствующих переменных (с отсрочкой и без нее). Далее этот набор данных может быть сохранен как новый файл данных.

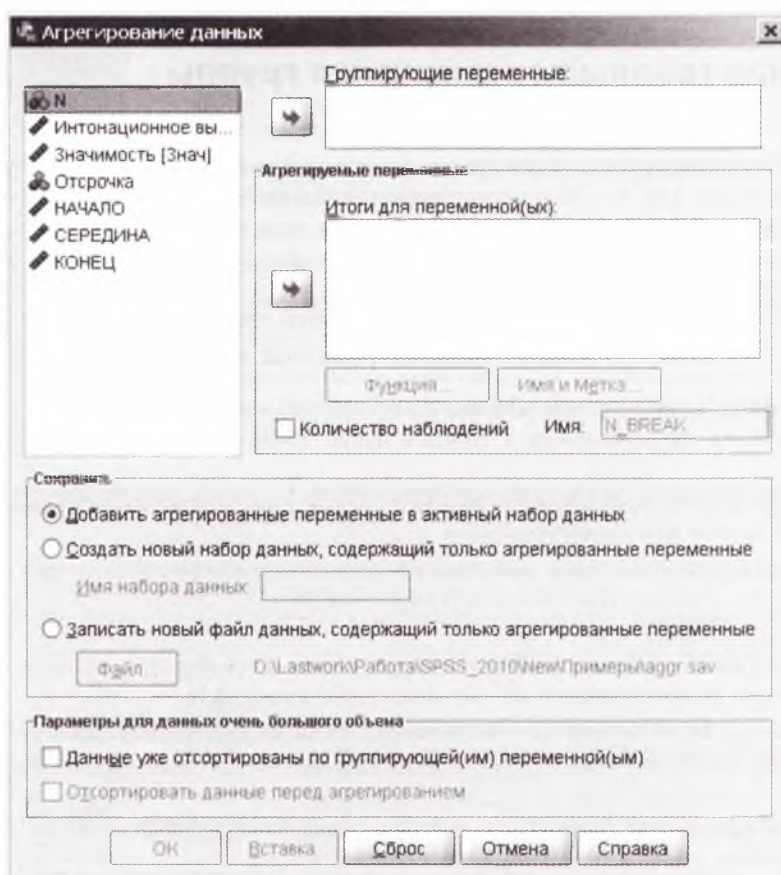


Рис. 4.13. Диалоговое окно Агрегирование данных

Реструктурирование данных

Команды Реструктурировать меню Данные позволяют производить весьма сложные преобразования структуры данных, например преобразовывать набор переменных в группы значений одной переменной или, наоборот, группы значений одной переменной — в на-

бор переменных. Чаще всего необходимость подобных преобразований возникает в случаях, когда на выборке произведено *несколько повторных измерений* одной переменной или группы переменных. Повторные измерения одной переменной (численностью P повторов) могут быть представлены в файле как последовательность P переменных для выборки численностью N (N строк). Те же данные могут быть представлены как P групп значений одной переменной. Тогда группы будет различать группирующая переменная, а количество строк будет $P \cdot N$. Подобным образом соотносятся файлы данных примеров ex020.sav, ex021.sav, ex022.sav (их описание приведено в главе 14). Далее будут приведены пошаговые инструкции для преобразования сначала 3 групп по 2 переменные в 3 переменные (с двумя группами значений), а затем — 6 групп значений одной переменной в группу из 3 переменных.

Преобразование группы переменных в группы значений

Для примера преобразования группы переменных в группы значений воспользуемся файлом ex020. В этом файле для $N = 20$ респондентов приведены 6 измерений продуктивности воспроизведения слов (НАЧАЛО1, ..., КОНЕЦ2): для трех частей ряда (начало, середина, конец), без отсрочки (индекс 1) и с отсрочкой (индекс 2). Преобразуем этот набор из шести переменных в 3 переменные в соответствии с частями ряда, а для различения групп в соответствии с отсрочкой, введем группирующую переменную Отсрочка. После выполнения шага 2 на экране должно быть окно редактора данных SPSS.

ШАГ 3н

Откройте файл данных, с которым вы намерены работать (в нашем случае ex020.sav). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Данные или щелкните на кнопке Открыть данные панели инструментов.
2. В открывшемся диалоговом окне дважды щелкните на имени ex020.sav или введите его с клавиатуры и щелкните на кнопке ОК.

ШАГ 4н

1. В меню Данные выберите команду Реструктурировать, чтобы открыть окно Конструктор реструктуризации данных, показанное на рис. 4.14.

Переключатель Что вы хотите сделать?, установленный по умолчанию, соответствует задаче преобразования групп переменных в группы наблюдений (значений).

2. Щелкните на кнопке Далее, чтобы перейти к окну Конструктор реструктуризации данных — Шаг 2 из 7 (рис. 4.15).

Под группой в данном случае подразумевается набор переменных, который после реструктуризации будет представлен одной переменной. В нашем случае таких групп 3 в соответствии с частями ряда слов, по 2 переменные в группе (с отсрочкой и без нее). Следовательно, необходимо задать 3 группы.

3. Установите переключатель Сколько групп переменных вы хотите реструктурировать? в положение Более одной и в поле Сколько? введите 3.
4. Щелкните на кнопке Далее, чтобы открыть окно Конструктор реструктуризации данных — Шаг 3 из 7, показанное на рис. 4.16.

Идентификатор групп наблюдений — это номер группы, одинаковый для значений в одной группе. В данном случае для нумерации групп можно использовать переменную N — номер респондента, но мы оставим установку по умолчанию для контроля результата.

Область Транспонируемые переменные предназначена для установления соответствия между исходными и итоговыми именами переменных. В итоге планируется на основе исходных шести создать 3 переменные (НАЧАЛО, СЕРЕДИНА, КОНЕЦ). Следовательно, каждой из трех новых переменных надо сопоставить две исходные переменные.

5. В окне Конструктор реструктуризации данных — Шаг 3 из 7 (см. рис. 4.16) в поле Целевая переменная замените trans1 на Начало.
6. Удерживая нажатой клавишу Ctrl на клавиатуре, щелкните сначала на переменной НАЧАЛО1, затем на переменной НАЧАЛО2, чтобы выделить их, и при помощи кнопки со стрелкой перенесите эти переменные в список Транспонируемые переменные.
7. В раскрывающемся списке Целевая переменная выберите trans2. Замените имя trans2 на Середина.
8. Повторите пункт 6 для переменных СЕРЕДИНА1 и СЕРЕДИНА2.
9. Повторите пункты 7 и 8 для целевой переменной Конец.
10. Перенесите переменные N, Инт и Знач в поле Фиксированные переменные.
11. Щелкните на кнопке Далее, чтобы перейти к следующему окну.
12. В появляющихся далее окнах последовательно щелкайте на кнопке Далее, а в последнем — на кнопке Готово для завершения процедуры реструктурирования.

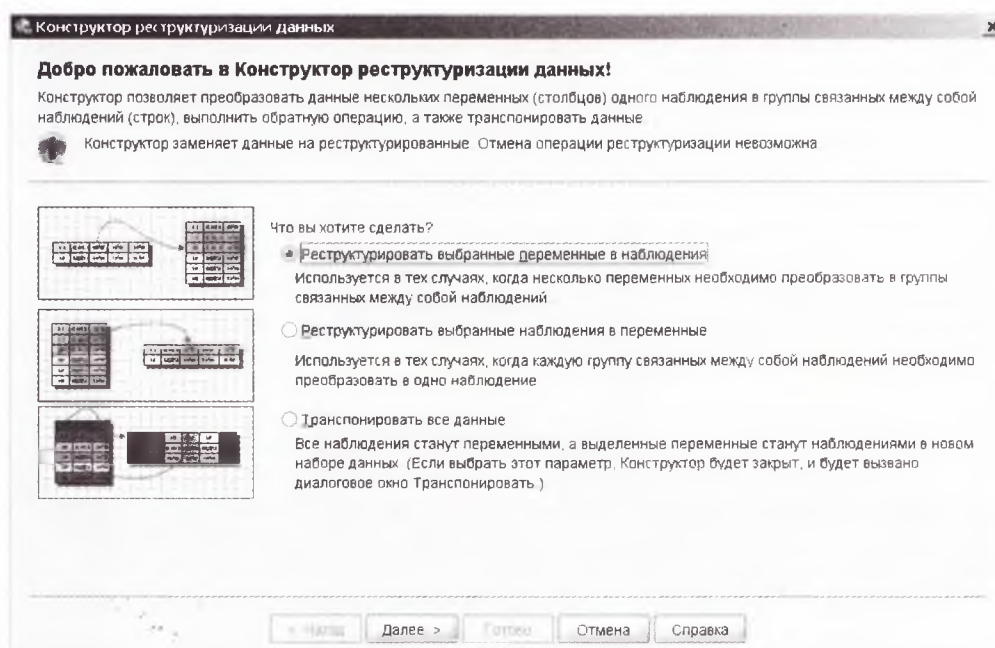


Рис. 4.14. Диалоговое окно Конструктор реструктуризации данных

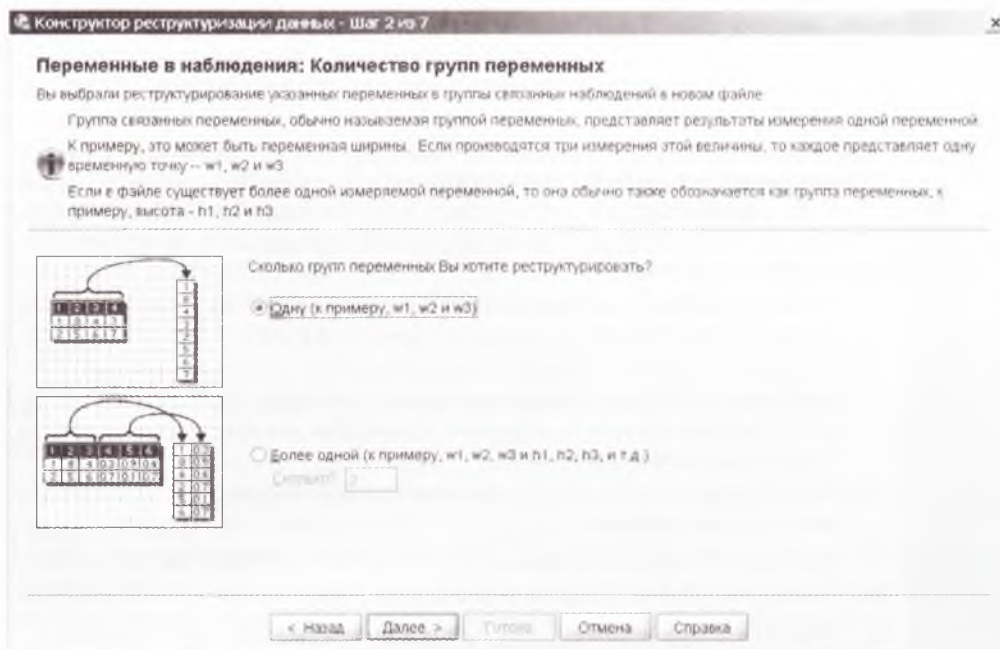


Рис. 4.15. Диалоговое окно Конструктор реструктуризации данных — Шаг 2 из 7

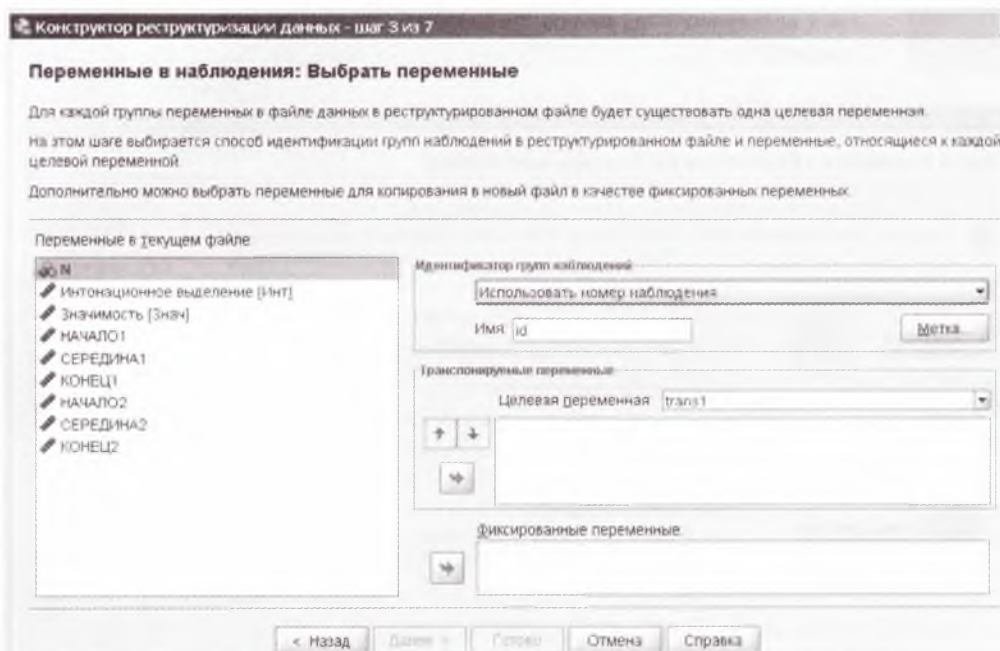


Рис. 4.16. Диалоговое окно Конструктор реструктуризации данных — Шаг 3 из 7

В итоге будет выполнено реструктурирование данных файла: вместо исходных шести переменных создано 3 новых (НАЧАЛО, СЕРЕДИНА, КОНЕЦ), а количество строк файла удвоится. Кроме того, появятся две новые переменные: `id` и `Индекс1`. Переменная `id` дублирует значения переменной `N` и нумерует группы наблюдений, а переменная `Индекс1` различает два наблюдения внутри групп в соответствии с исходными переменными, входящими в группу (1 — без отсрочки, 2 — с отсрочкой). Имя переменной `Индекс1` может быть заменено на другое в соответствии со смыслом группировки, например, в данном случае на имя `Отсрочка`.

Важно отметить, что результату реструктурирования приписывается имя исходного файла. Поэтому, чтобы сохранить исходный файл, результирующему файлу следует присвоить новое имя.

Преобразование групп значений в группы переменных

Для примера преобразования групп значений переменной в группу переменных воспользуемся файлом `ex021.sav`. В этом файле для $N = 20$ респондентов приведено 3 измерения продуктивности воспроизведения слов для трех частей ряда (НАЧАЛО, СЕРЕДИНА, КОНЕЦ). Для каждого респондента значения этих переменных представлены дважды в соответствии с группирующей переменной `Отсрочка` (1 — без отсрочки, 2 — с отсрочкой). Преобразуем этот набор из трех переменных в 6 переменных в соответствии с тремя частями ряда и отсрочкой (по 2 переменные для каждой части ряда).

После выполнения шага 2 на экране должно быть окно редактора данных SPSS.



Откройте файл данных, с которым вы намерены работать (в нашем случае `ex021.sav`). Если он расположен в текущей папке, выполните следующие действия:

1. Выберите в меню **Файл** команду **Открыть** ► **Данные** или щелкните на кнопке **Открыть данные** панели инструментов.
2. В открывшемся диалоговом окне дважды щелкните на имени `ex021.sav` или введите его с клавиатуры и щелкните на кнопке **OK**.



1. В меню **Данные** выберите команду **Реструктурировать**, чтобы открыть окно Конструктор реструктуризации данных, показанное на рис. 4.14.
2. Установите переключатель **Что вы хотите сделать?** в положение **Реструктурировать выбранные наблюдения в переменные** и щелкните на кнопке **Далее**, чтобы перейти к окну Конструктор реструктуризации данных — Шаг 2 из 5 (рис. 4.17).

В окне **Переменные для идентификации** (см. рис. 4.17) необходимо задать переменную для идентификации групп наблюдений. В файле `ex021.sav` это переменная `N`. Индексная переменная является группирующей внутри групп наблюдений и задает индексы для новых переменных. В нашем примере это переменная `Отсрочка`.

3. Выделите переменную `N` и перенесите ее в поле **Переменные для идентификации** при помощи кнопки со стрелкой.
4. Выделите переменную `Отсрочка` и перенесите ее в поле **Индексные переменные** при помощи кнопки со стрелкой.

5. Щелкните на кнопке Далее, чтобы перейти к следующему окну.
6. В появляющихся далее окнах последовательно щелкайте на кнопке Далее, а в последнем — на кнопке Готово для завершения процедуры реструктурирования.

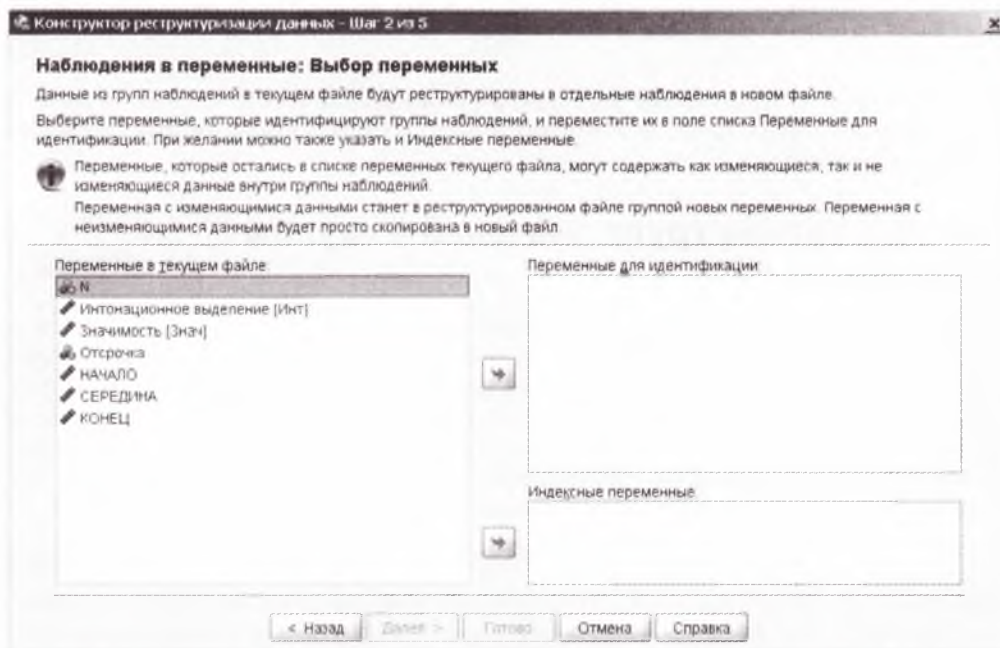


Рис. 4.17. Диалоговое окно Конструктор реструктуризации данных — Шаг 2 из 5

В итоге реструктурирования будет получен новый набор данных, содержащий вместо трех исходных переменных (НАЧАЛО, СЕРЕДИНА, КОНЕЦ) 6 переменных (НАЧАЛО.1, НАЧАЛО.2, ..., КОНЕЦ.2) в соответствии с двумя значениями группирующей переменной Отсрочка. Те переменные, значения которых дублируются внутри групп (N, Инт, Знач), войдут в новый набор данных, а индексная переменная (Отсрочка) в нем отсутствует. Чтобы сохранить исходный файл, результирующему файлу следует присвоить новое имя.

5 Диаграммы

83	Графика в программе SPSS
84	Настройка диаграмм
85	Команды построения диаграмм
94	Редактирование диаграмм
95	Выход из программы

Программа IBM SPSS Statistics обладает обширным арсеналом мощных и эффективных средств построения диаграмм. Разумеется, в книгу такого объема, как эта, бессмысленно даже пытаться вместить описания всех этих средств. Мы коснемся лишь основных аспектов работы с графическими объектами.

Графика в программе SPSS

Как правило, диаграмма зависит от конкретной статистической процедуры, следовательно, рассматривать все тонкости построения диаграмм имеет смысл лишь в контексте обработки данных. Поэтому вся специфическая информация, относящаяся к конкретным диаграммам, излагается по мере необходимости в главах книги, посвященных статистическому анализу. А здесь собраны лишь общие сведения, касающиеся диаграмм вообще, и по большей части эти сведения связаны с редактированием уже созданных диаграмм.

Далее перечислены далеко не все, но наиболее часто используемые виды графиков:

- **Столбики** (столбиковые диаграммы). Эти графики применяются для отображения итогов для значений (категорий) переменной. Столбики соответствуют категориям переменной. Высота каждого столбика пропорциональна вычисляемому итогу. В качестве итога по умолчанию вычисляется количество наблюдений для данной категории, то есть строится диаграмма распределения частот. Например, с помощью столбиковой диаграммы удобно представить распределение учащихся по трем классам или по их ориентации на поступление в вузы четырех типов и т. д. В качестве итога можно задать статистику для другой переменной, например среднее. Так, задав для переменной класс в качестве итога среднее для переменной отметка, можно получить диаграмму средней отметки для разных классов. Столбиковые диаграммы делятся на простые, кластерные и состыкованные.
- **Гистограммы** внешне напоминают столбиковые диаграммы, однако, как правило, иллюстрируют распределение наблюдений по диапазонам значений непрерывной переменной (имеющей большое количество возможных значений). С помощью гистограммы было бы удобно представить распределение учащихся по диапазонам значений успеваемости (отметки) или диапазонам тестовых значений.
- **Линии** (линейные диаграммы) отображают то же, что и столбики, но в виде ломаной линии.

- ▶ **Круг** (круговые диаграммы), как и столбиковые, зачастую применяются для иллюстрации распределения наблюдений в различных категориях.
- ▶ **Ящики** (коробчатые диаграммы) основаны на процентилях и являются прекрасным средством отображения различия между группами с учетом распределения данных.
- ▶ **Рассеяния/точки** (диаграммы рассеяния) часто используются для отображения характера взаимосвязи между переменными.

В меню Графика окна редактора данных имеется три подменю (SPSS 20): Конструктор диаграмм, Панель выбора диаграмм и Устаревшие диалоговые окна. Последнее подменю содержит список стандартных графиков, которые мы перечислили ранее. Меню Конструктор диаграмм и Панель выбора диаграмм предоставляют дополнительную очень удобную возможность построения графиков в диалоговом режиме. В этих меню представлен широкий спектр команд, позволяющих строить диаграммы в интерактивном режиме: добавлять переменные, изменять категории данных и т. п. Другими словами, работая исключительно с графиками, вы можете добиться результата и без обращения к вычислительным статистическим процедурам. Иногда такая возможность может оказаться полезной для пользователя, однако до этого необходимо освоить стандартные приемы работы с SPSS в отношении как статистики, так и графики. Поэтому описание процедур конструктора и выбора диаграмм выходит за рамки темы этой книги.

Настройка диаграмм

Установленные по умолчанию настройки для диаграмм SPSS далеко не всегда соответствуют требованиям пользователя программы. Например, вас может не устраивать установленный цвет, или вообще то, что диаграммы цветные. Проблемой может оказаться и некорректное отображение русскоязычных наименований переменных и их меток, что было характерно для ранних версий SPSS. Средства редактирования диаграмм позволяют решить большую часть этих и подобных проблем для каждого графика по отдельности. Но есть более радикальный способ — изменение общих настроек диаграмм, которые станут в дальнейшем настройками программы SPSS по умолчанию.

Для изменения общих настроек диаграмм необходимо в строке меню редактора данных открыть меню Правка и выбрать команду Параметры. В открывшемся диалоговом окне следует перейти на вкладку Диаграммы. Здесь можно задать по своему вкусу основные свойства диаграмм, которые в дальнейшем будут применяться программой по умолчанию ко всем графикам.

Рассмотрим решение наиболее часто встречающейся задачи настройки диаграмм: установки отображения графиков в черно-белом цвете.

В программе SPSS по умолчанию установлены цветные графики: Параметры цикла стиля: цикл только по цветам. При этом однотипные (повторяющиеся) элементы, относящиеся к разным категориям, будут различаться цветом. Для установки черно-белых графиков на вкладке Диаграммы окна Параметры в раскрывающемся меню Параметры цикла стиля достаточно выбрать Цикл только по узорам. В этом случае отображаемые на диаграмме элементы будут различаться узором, а не цветом (сплошные и пунктирные линии, столбики со штриховкой и без нее и т. д.).

Можно также задать стиль графиков в соответствии с рекомендациями Американской психологической ассоциации (APA). Для этого в разделе Шаблон диаграмм следует установить переключатель в положение Использовать шаблон диаграмм и указать путь к ме-

сту, где этот шаблон расположен. Если установлена версия SPSS 20, это путь C:\Program Files\IBM\SPSS\Statistics\20\Looks\APA_Styles.sgt. Этот шаблон под именем APA_Styles.sgt находится также в папке файлов примеров.

Команды построения диаграмм

Большинство статистических процедур, описанных в последующих главах книги, имеют встроенные средства построения графиков. Тем не менее программа SPSS предоставляет возможность разделения задач построения графиков и статистического анализа. Вы можете с равным успехом отредактировать как готовую диаграмму, сформированную статистической процедурой, так и диаграмму, созданную при помощи меню Диаграммы. Примеры и рекомендации, которые вы здесь найдете, позволят вам эффективно работать с диаграммами.

Основные типы команд и диалоговых окон мы рассмотрим на примере построения наиболее часто используемых столбиковых диаграмм. Будут представлены пошаговые процедуры построения столбиковых диаграмм распределения частот (простые (рис. 5.1) и кластеризованные (рис. 5.4)), а также средних значений (простые (рис. 5.6) и кластеризованные (рис. 5.9)).

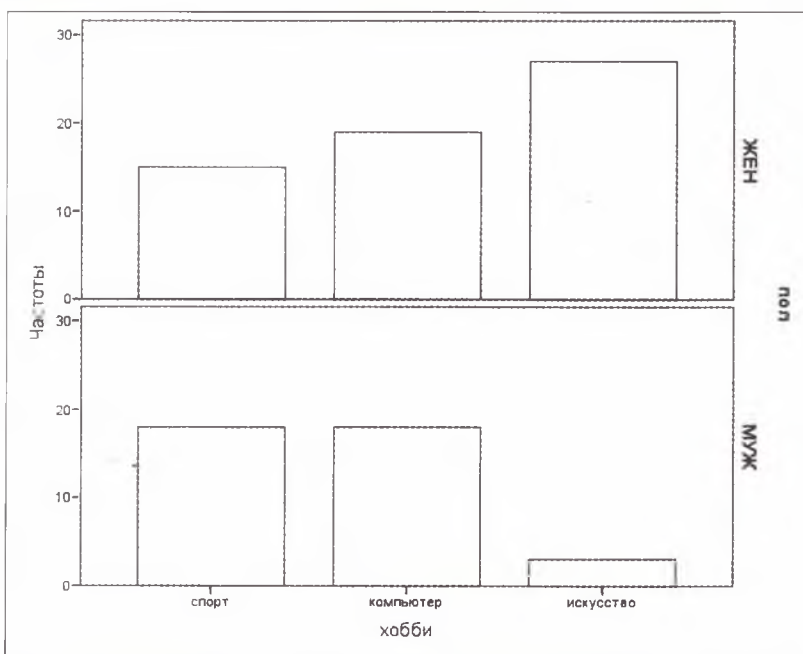


Рис. 5.1. Простая столбиковая диаграмма распределения частот

Для построения диаграмм будем использовать уже знакомый вам файл ex01.sav. Однако сначала следует выполнить три подготовительных шага. Эти шаги (1–3) позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex01.sav.

Сначала приведем пошаговые процедуры построения простой столбиковой диаграммы распределения частот (см. рис. 5.1).

ШАГ 4 В меню Графика выберите команду Устаревшие диалоговые окна ► Столбики. На экране появится диалоговое окно, показанное на рис. 5.2.

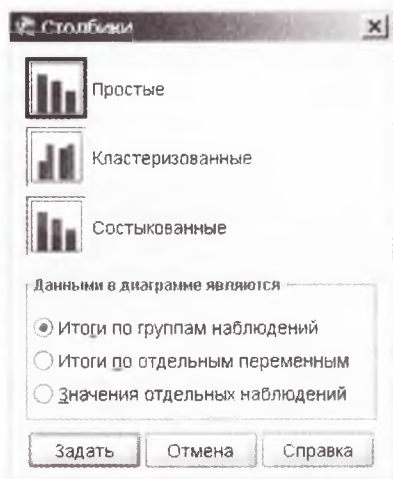


Рис. 5.2. Диалоговое окно задания типа столбиковой диаграммы

С помощью элементов интерфейса верхней части окна вы можете выбрать тип диаграммы — Простые, Кластеризованные и Состыкованные, а в нижней части указать, какие данные вы намерены использовать, установив переключатель Данными в диаграмме являются в нужное положение. Типы диаграмм будут меняться в зависимости от вида диалогового окна; для того чтобы указать нужный тип, щелкните на значке диаграммы слева от ее названия. Варианты используемых данных для большинства типов диаграмм одни и те же:

- положение Итоги по группам наблюдений означает, что задействуется одна переменная, а столбцы диаграммы отразят количество или процент наблюдений каждой градации этой переменной. Например, для переменной класс диаграмма будет состоять из трех столбцов, соответствующих трем классам учащихся; для переменной пол — из двух столбцов, отображающих количество (%) учащихся женского и мужского пола; для переменной вуз — из четырех столбцов, представляющих количество (%) учащихся в каждой из четырех групп предпочтительных вузов;
- положение Итоги по отдельным переменным указывает на то, что диаграмма будет содержать несколько столбцов, каждый из которых будет соответствовать среднему значению одной из переменных. Как правило, при построении этого типа диаграмм используют переменные, связанные между собой по смыслу. В файле ex01.sav такими переменными являются переменные тест1, ..., тест5: диаграмма, построенная для них, отразит средние баллы всех учащихся для каждого из пяти тестов;
- положение Значения отдельных наблюдений предназначено для данных с относительно небольшим количеством объектов. Если взять данные из файла ex01.sav и

построить подобную диаграмму для какой-либо переменной, например `test1`, то на экране появились бы 100 столбцов, отображающие баллы каждого из учащихся по первому тесту.

ШАГ 5

После выполнения шага 4 должно быть открыто диалоговое окно, изображенное на рис. 5.2.

Оставив выделенным по умолчанию тип диаграммы Простые, а переключатель Данными в диаграмме являются в положении Итоги по группам наблюдения, щелкните на кнопке Задать, чтобы открыть диалоговое окно, фрагмент которого показан на рис. 5.3.

Столбик представляет

☒ N наблюдений ☐ % наблюдений

☐ Накопленное N ☐ Накопленный %

☐ Другую статистику (например, среднее)

Переменная:

Изменить статистику

Категориальная ось:

Панель по

Строки

Столбцы

☐ Выводить переменные (не выводить пустые строки)

☐ Выводить переменные (не выводить пустые столбцы)

Рис. 5.3. Фрагмент диалогового окна Простые столбики: Итожащие функции по группам наблюдений

В верхней части окна находится переключатель Столбики представляют, который можно установить в одно из пяти положений. Чаще других используются положения N наблюдений и % наблюдений — каждый столбец или сегмент графика представляет количество (процент) наблюдений в каждой из категорий. При выборе положения Другую статистику (например, среднее) становится доступным поле Переменная, находящееся ниже. Если установить переключатель в это положение и ввести в поле имя количественной переменной, то столбики будут отражать средние этой переменной по каждой из категорий.

Поле Категориальная ось служит для задания имени переменной, для категорий которой будут строиться столбики, в нашем примере это переменная Хобби. Поля Строки и Столбцы в области Панель по позволяют задавать имена переменных, для категорий которых будут строиться отдельные графики. В нашем примере такой переменной является пол.

Продолжим построение столбиковой диаграммы для переменной хобби с учетом значений переменной пол, показанной на рис. 5.1. После выполнения шага 5 на экране должно появиться окно, фрагмент которого показан на рис. 5.3.

ШАГ 6

1. В списке переменных слева щелчком выделите переменную хобби и при помощи стрелки перенесите ее в поле Категориальная ось.
2. В списке переменных слева щелчком выделите переменную пол и при помощи стрелки перенесите ее в поле Строки области Панель по.
3. Оставив переключатель Столбики установленным в установленном по умолчанию положении N наблюдений, щелкните на кнопке ОК.

В результате выполнения шагов 4–6 появится окно вывода со столбиковой диаграммой, изображенной на рис. 5.1.

В следующем примере мы снова построим столбиковую диаграмму для переменной хобби с учетом переменной пол, однако для этого воспользуемся кластеризованной диаграммой и представим столбиками не N , а процент наблюдений. В результате должна получиться диаграмма, изображенная на рис. 5.4.

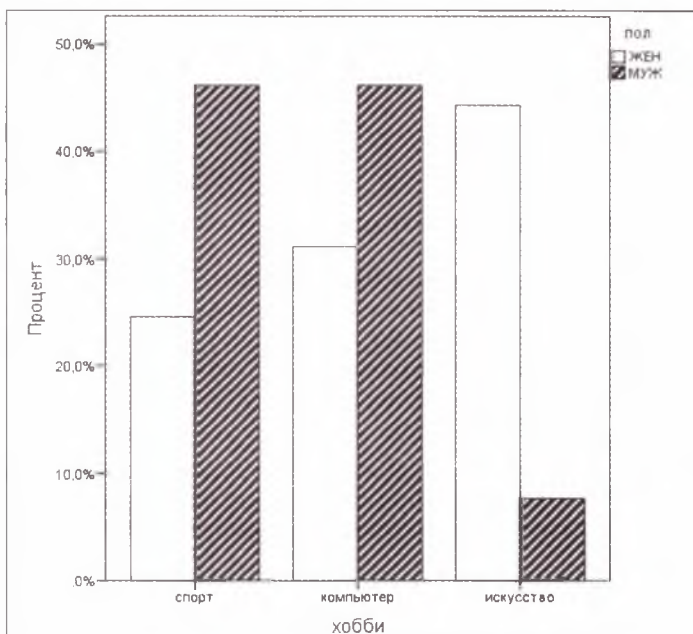


Рис. 5.4. Кластеризованная столбиковая диаграмма распределения частот (в процентах)

Выполните шаг 4. После этого должно быть открыто диалоговое окно, изображенное на рис. 5.2.

ШАГ 5а

- Щелчком мыши выделите тип диаграммы Кластеризованные и щелкните на кнопке Задать, чтобы открыть диалоговое окно, фрагмент которого показан на рис. 5.5.

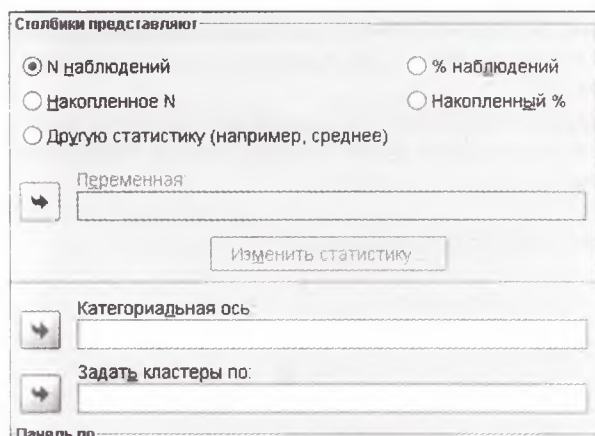


Рис. 5.5. Фрагмент диалогового окна Кластеризованные столбики: Итожащие функции по группам наблюдений

Единственное отличие диалогового окна на рис. 5.5 (для кластеризованных столбиков) от изображенного на рис. 5.3 (для простых столбиков) — наличие поля Задать кластеры по. В нем следует указать имя переменной, для значений которой будут строиться сопряженные столбики в каждой категории основной переменной. В нашем примере кластеры задаются по переменной пол.



1. В списке переменных слева щелчком выделите переменную хобби и при помощи стрелки перенесите ее в поле Категориальная ось.
2. В списке переменных слева щелчком выделите переменную пол и при помощи стрелки перенесите ее в поле Задать кластеры по.
3. Переключатель Столбики представляют установите в положение % наблюдений и щелкните на кнопке ОК.

В результате выполнения последовательности шагов 4, 5а и 6а появится окно вывода с кластеризованной столбиковой диаграммой (см. рис. 5.4). Обратите внимание на то, что за 100 % принимается количество наблюдений отдельно для муж и жен, то есть внутри каждой категории переменной пол, по которой были заданы кластеры.

В следующих двух примерах мы рассмотрим построение и редактирование столбиковых диаграмм для средних значений (рис. 5.6 и 5.9). Сначала приведем пошаговые инструкции построения простой столбиковой диаграммы средних значений переменной отметка2 для категорий переменной хобби с учетом категорий переменной пол (рис. 5.6).

Выполните шаг 4. После этого должно быть открыто диалоговое окно, изображенное на рис. 5.2.



1. Щелчком мыши выделите тип диаграммы Простые и оставьте переключатель Данными в диаграмме являются установленным в положение Итоги по группам наблюдений. Щелкните на кнопке Задать, чтобы открыть диалоговое окно Простые столбики: Итожащие функции по группам наблюдений, фрагмент которого показан на рис. 5.3. Если вы уже поработали с этим окном, щелкните на кнопке Сброс.

2. В качестве категориальной оси задайте переменную Хобби. Для этого щелчком мыши выделите ее в списке слева и при помощи стрелки перенесите в поле Категориальная ось.
3. Задайте отдельные графики для категорий переменной пол. Для этого щелчком мыши выделите ее в списке слева и при помощи стрелки перенесите в поле Строки области Панель по.

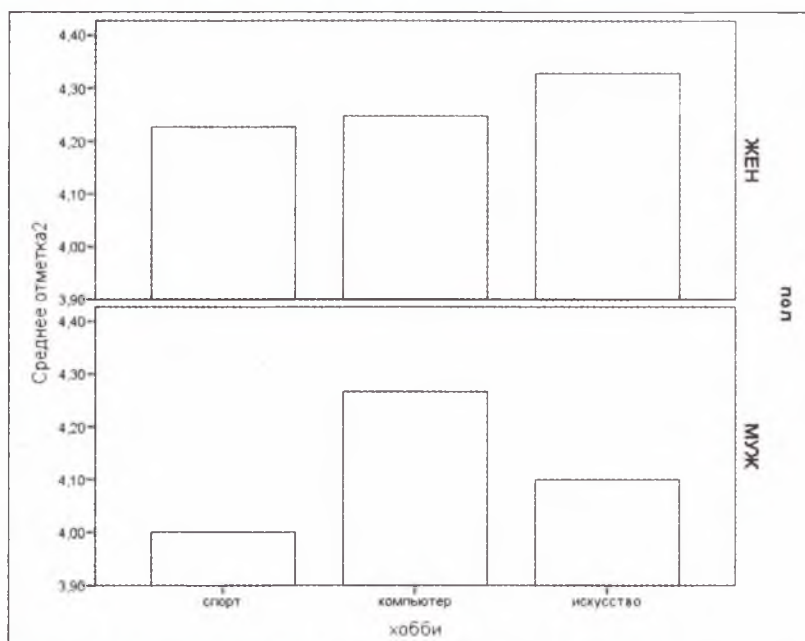


Рис. 5.6. Простая столбиковая диаграмма средних значений

До этого момента шаги построения графика средних совпадают с построением графика распределения частот. Для установки средних необходимо переключатель Столбики представить установить в положение Другую статистику (например, среднее). Тогда станет доступным окно для указания количественной переменной, по значениям которой будут вычисляться средние. После задания такой переменной станет доступной кнопка Изменить статистику, при щелчке на которой появляется окно, показанное на рис. 5.7. При помощи этого окна вместо среднего можно задать другую статистику, например медиану, если переменная порядковая (ранговая).

ШАГ 66

После выполнения шага 56 должно быть открыто окно, фрагмент которого показан на рис. 5.3. В этом окне заданы хобби в поле Категориальная ось и пол в поле Строки.

1. Переключатель Столбики представьте установить в положение Другую статистику (например, среднее). Станет доступным поле Переменная, находящееся под переключателем.
2. В списке слева выделите щелчком мыши переменную отсчета2 и при помощи стрелки перенесите ее в поле Переменная.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода с диаграммой.

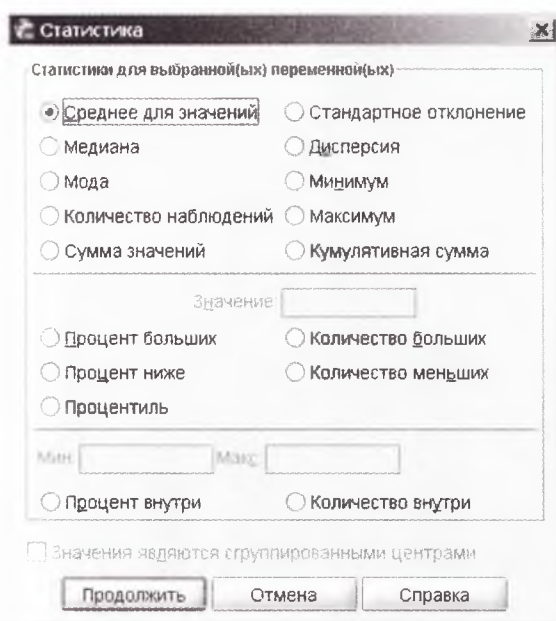


Рис. 5.7. Переключатели выбора статистики для диаграммы

Диаграмма, полученная в результате выполнения шага 6б, отличается от изображенной на рис. 5.6 тем, что вертикальная ось средних начинается со значения 0. В результате различия между средними визуально едва заметны. Исправить положение и сделать диаграмму средних более наглядной может изменение точки начала отсчета вертикальной оси, например, на значение 3,9. Следующий шаг демонстрирует некоторые возможности редактирования диаграммы.

ШАГ 7б

После выполнения шага 6б должно быть открыто окно вывода с диаграммой средних значений.

1. Двойным щелчком мыши на диаграмме откройте окно Редактор диаграмм, фрагмент которого показан на рис. 5.11.
2. В окне редактора диаграмм щелкните на вертикальной оси графика, чтобы выделить ее. Затем сделайте на этой оси двойной щелчок, чтобы открыть окно Свойства (рис. 5.8).
3. В диалоговом окне Свойства на вкладке Ось в поле Минимум задайте значение 3,9 и щелкните на кнопке Применить, чтобы изменения вступили в силу.
4. Закройте окно Редактор диаграмм, выбрав в его командной строке Файл ► Закроить.

В результате выполнения последнего шага в окне вывода появится диаграмма средних значений, аналогичная изображенной на рис. 5.6.

В следующем примере мы построим и отредактируем кластеризованную столбиковую диаграмму средних значений для нескольких переменных. Такие диаграммы можно строить для повторных измерений или для других родственных переменных, представленных в едином масштабе (в одной шкале). В нашем случае такими переменными являются показатели интеллекта тест1, тест2, ..., тест5. На рис. 5.9 изображена кластери-

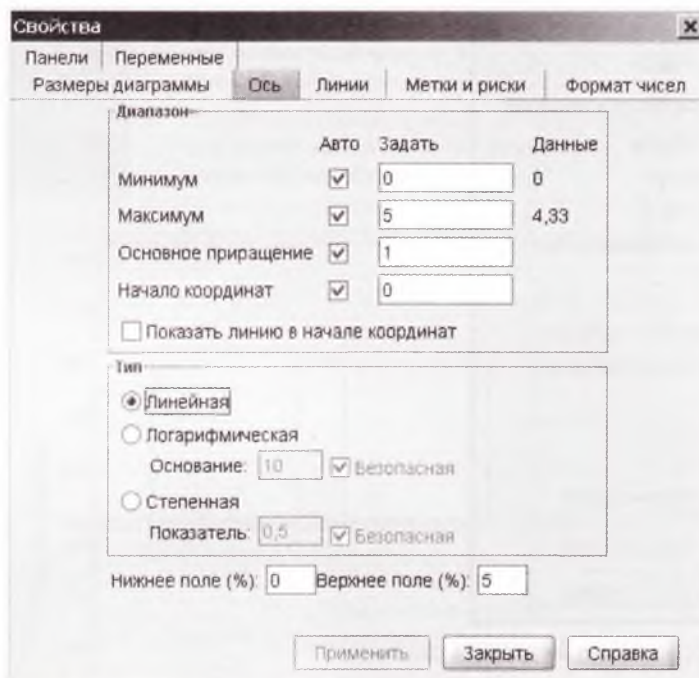


Рис. 5.8. Диалоговое окно редактора диаграмм Свойства

зованная диаграмма средних значений этих переменных для разных категорий переменных пол. Диаграмма будет построена в результате выполнения шагов 5в и 6в.

ШАГ 5в

Выполните шаг 4. После этого должно быть открыто диалоговое окно Столбики, изображенное на рис. 5.2.

1. Щелчком мыши выделите тип диаграммы Кластеризованные и установите переключатель Данными в диаграмме являются в положение Итоги по отдельным переменным. Щелкните на кнопке Задать, чтобы открыть диалоговое окно Кластеризованные столбики: Итожащие функции по переменным, фрагмент которого показан на рис. 5.10.
2. В качестве категориальной оси задайте переменную пол. Для этого щелчком мыши выделите ее в списке слева и при помощи стрелки перенесите в поле Категориальная ось.
3. Задайте переменные, по которым будут вычисляться средние значения. Для этого щелчком мыши выделите переменную тест1 в списке слева и при помощи стрелки перенесите ее в поле Столбики представляют. Повторите эту последовательность для переменных тест2, тест3, тест4 и тест5.
4. Щелкните на кнопке ОК, чтобы открыть окно вывода с диаграммой.

Кластеризованная диаграмма, полученная в результате выполнения шага 5в, отличается от изображенной на рис. 5.9 тем, что вертикальная ось средних начинается со значения 0. Сделать диаграмму средних более наглядной можно так же, как и в предыдущем примере, изменив точку отсчета вертикальной оси с 0 на 9.

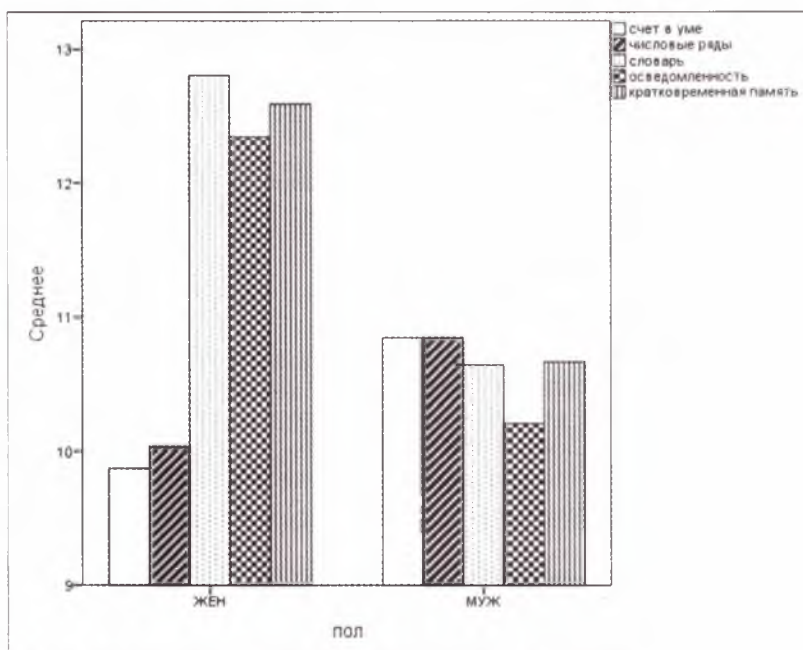


Рис. 5.9. Кластеризованная диаграмма средних значений

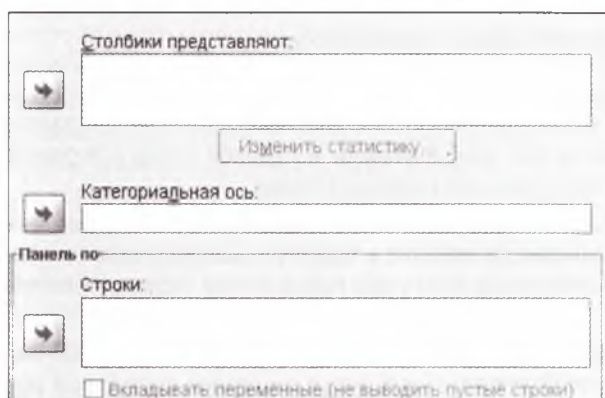


Рис. 5.10. Фрагмент диалогового окна Кластеризованные столбики: Итожащие функции по переменным

ШАГ 6в

После выполнения шага 5в должно быть открыто окно вывода с кластеризованной диаграммой средних значений.

1. Двойным щелчком мыши на диаграмме откройте окно Редактор диаграмм, фрагмент которого показан на рис. 5.11.
2. В окне редактора диаграмм щелкните на вертикальной оси графика, чтобы выделить ее. Затем сделайте на этой оси двойной щелчок, чтобы открыть окно Свойства, аналогичное показанному на рис. 5.8.



3. В диалоговом окне Свойства на вкладке Ось задайте в поле Минимум значение 9 и щелкните на кнопке Применить, чтобы изменения вступили в силу.
4. Закройте окно Редактор диаграмм, выбрав в его командной строке Файл ► Закреть.

В результате выполнения последнего шага в окне вывода появится кластеризованная диаграмма средних значений, идентичная изображенной на рис. 5.9.

Редактирование диаграмм

После создания диаграммы можно воспользоваться большим набором команд для ее редактирования. С некоторыми из них вы познакомились при построении столбиковых диаграмм в предыдущем разделе. Для редактирования диаграммы необходимо в окне вывода дважды щелкнуть мышью на выбранном графическом объекте. При этом на экране появится окно Редактор диаграмм, содержащее строку меню с набором команд, и панель инструментов, на которой находятся кнопки, соответствующие наиболее часто используемым командам (рис. 5.11).

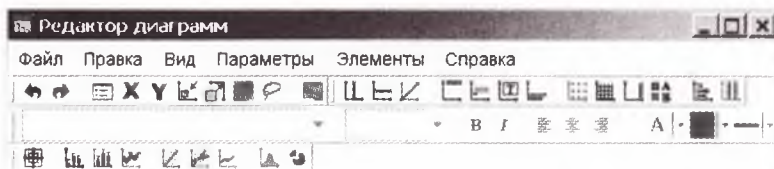


Рис. 5.11. Редактор диаграмм: строка меню и панель инструментов

Команды редактирования могут применяться к диаграмме только в том случае, если она отображена на экране в режиме редактирования. Как только вы дважды щелкнете на диаграмме, SPSS откроет новое окно с ее изображением и строкой меню в верхней части — это означает, что диаграмма доступна для редактирования.

Начиная с SPSS 13.0, графический редактор существенно изменился: теперь можно обойтись без команд строки меню и панели управления редактора. Для редактирования любого элемента графика достаточно навести на него указатель мыши и щелкнуть левой кнопкой для его выделения.

Если выделенный элемент диаграммы — текст, который вы хотите изменить, достаточно еще раз щелкнуть на нем правой кнопкой мыши. После этого он станет доступен в режиме обычного текстового редактора.

Если выделенный элемент является графическим (линия, маркер, заливка и т. п.), то двойной щелчок на нем левой кнопкой мыши открывает диалоговое окно для его редактирования. Большим удобством стало то, что для каждого элемента диаграммы открывается диалоговое окно, предоставляющее широкие возможности редактирования именно этого элемента.

Рассмотрим пример редактирования гистограммы из главы 6 (рис. 6.7, с. 104). Если навести указатель на столбцы гистограммы, выделить их щелчком левой кнопки мыши и затем дважды быстро щелкнуть левой кнопкой, появится диалоговое окно, фрагмент которого изображен на рис. 5.12.

Вкладки этого окна зависят от возможностей редактирования конкретного элемента. В данном случае окно открывается на вкладке Параметры гистограммы. Если переключи-

чатель Ось X установить в положение **Задается пользователем**, вместо параметров гистограммы по умолчанию можно задать **Размер интервала** или **Количество интервалов**.

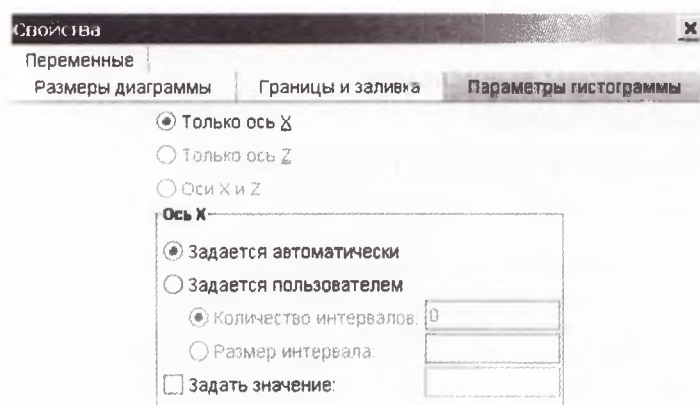


Рис. 5.12. Фрагмент диалогового окна редактирования гистограммы

После внесения изменений становится доступной кнопка **Применить**. При нажатии на нее изменения данного элемента вступают в силу.

Щелчок правой кнопкой мыши открывает общее диалоговое окно, форма которого не зависит от того, в каком месте диаграммы находится указатель мыши. Это окно содержит команды, соответствующие панели инструментов редактирования диаграмм с набором команд, которые доступны для данного графика. Некоторые команды при этом могут быть недоступны (затенены) — в зависимости от возможностей редактирования элемента, на котором в момент щелчка находился указатель.

Заканчивая обзор средств редактирования диаграмм, отметим, что возможности графического редактора SPSS настолько широки, что в рамках данной книги не представляется возможным описать даже часть их. Да и в этом нет необходимости, поскольку редактор обладает высокоразвитым интуитивным интерфейсом. Поэтому изложенных сведений вполне достаточно для того, чтобы без предварительной подготовки приступить к редактированию графиков. Если по мере редактирования вы почувствуете, что лучший вариант уже был раньше, то всегда сможете вернуться и найти его при помощи двух кнопок:

- ▶  Отменить действие пользователя;
- ▶  Вернуть действие пользователя.

После окончания редактирования графика достаточно щелкнуть на кнопке **Заккрыть** или выбрать в командной строке **Файл** ▶ **Заккрыть**, чтобы закрыть редактор диаграмм. В окне вывода вы увидите окончательный вариант диаграммы.

Выход из программы

Для дальнейшего использования полученного результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню **Файл** команду **Выход**.

6 Частоты

97	Пошаговые алгоритмы вычислений
102	Представление результатов
105	Завершение анализа и выход из программы

В этой главе рассматриваются частоты, их графическое представление (столбиковые и круговые диаграммы), гистограммы и проценти́ли. Частоты — одна из команд, позволяющих непосредственно работать с диаграммами, не прибегая к командам меню Графика. Некоторые детали, касающиеся редактирования диаграмм, были рассмотрены в главе 5. *Столбиковые и круговые диаграммы* отображают количество объектов (*частоту*) в различных категориях дискретной переменной (имеющей небольшое число возможных значений). Однако команда Частоты позволяет работать не только с дискретными, но и с непрерывными переменными, которые могут принимать множество значений.

Далее приведены краткие определения понятий, о которых пойдет речь в этой главе:

- ▶ **Частоты.** Команда Частоты является одной из самых простых и часто используемых команд SPSS. Ее действие сводится к подсчету количества объектов в каждой категории переменной. Это называется распределением частот по категориям переменной. Если мы анализируем переменную пол, программа подсчитает распределение численности девушек и юношей среди учащихся; если используется переменная класс, получим распределение численности учащихся по классам. Выводимый результат для каждой категории включает в себя метку значения переменной, само значение переменной, частоту, процент и накопленный процент от общей частоты.
- ▶ **Столбиковые диаграммы.** Обязательное условие для применимости столбиковой диаграммы — дискретность отображаемых данных. Такими свойствами обладают переменные пол, класс, хобби и т. п. Каждая из этих переменных делит все объекты файла данных на категории. Например, переменная вуз разбивает всех учащихся на 4 группы по выбору профиля дальнейшего образования. Столбиковыми диаграммами можно отображать также непрерывные переменные. При этом к категории относятся данные, попадающие в определенный диапазон (интервал). Примерами непрерывных переменных являются средний балл отметки каждого учащегося, время финиша в гоночных состязаниях, вес пациента при медосмотре и т. п.
- ▶ **Гистограммы** используются для отображения распределения частот непрерывных переменных. Особенностью гистограмм, отличающей их от столбиковых диаграмм, является то, что каждый столбец представляет собой не отдельное значение, а диапазон значений переменной. Гистограммы могут использоваться и для дискретных данных в случаях, если число значений переменной слишком велико, чтобы отображать каждое из них отдельным столбцом диаграммы. Весь диапазон изменчивости переменной разделяют на интервалы, а затем по ним строят гистограмму. Именно таким методом следовало бы отображать распределение частот для переменных от-метка¹ или отметка².

- *Процентиль* показывает, какой процент распределения лежит ниже заданной величины. Например, если говорят, что процентиль значения 111 равен 75, это означает, что 75 % всех значений переменной меньше, чем 111, а 25 % — больше, чем 111.

Пошаговые алгоритмы вычислений

Для иллюстрации вычислений при помощи команды Частоты будем использовать уже знакомый вам файл ex01.sav (мы не приводим примеры круговых диаграмм ввиду их исключительной простоты). Количество объектов в файле равно 100. Однако сначала следует выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex01.sav.

ШАГ 4

В меню Анализ выберите команду Описательные статистики ► Частоты. На экране появится диалоговое окно, показанное на рис. 6.1.

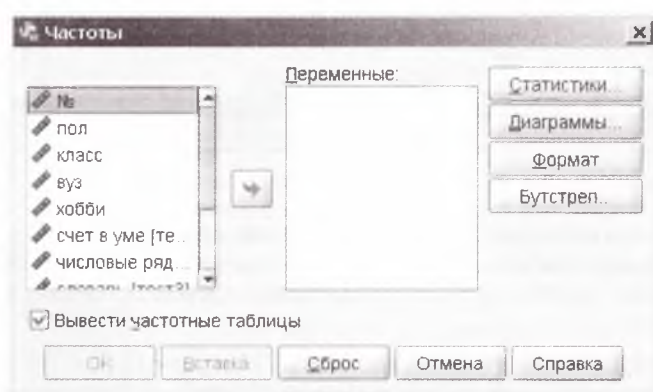


Рис. 6.1. Диалоговое окно Частоты

Частоты

Диалоговое окно Частоты типично для большинства статистических операций SPSS. В левой части окна расположен список всех доступных переменных. В нем необходимо выбрать те переменные, для которых надо вычислить распределение частот.

Для этого щелчком выделите нужную переменную, а затем щелчком на кнопке с направленной вправо стрелкой переместите ее в поле Переменные — целевой список. Если нужная переменная не видна в исходном списке, воспользуйтесь полосой прокрутки.

Для того чтобы удалить переменную из целевого списка, достаточно выделить ее, а затем воспользоваться кнопкой с направленной влево стрелкой, которая появляется на месте предыдущей кнопки при выделении любого пункта в списке Переменные. В этом случае переменная вновь переместится в исходный список. Чтобы полностью очистить список Переменные, щелкните на кнопке Сброс.

ШАГ 5

На этом шаге займемся вычислением частот для переменных пол, класс и вуз. При открытом диалоговом окне Частоты выполните следующие действия:

1. Щелкните сначала на переменной пол, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы переместить переменную в поле Переменные.
2. Повторите предыдущее действие для переменных класс и вуз.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода, показанное на рис. 6.2.

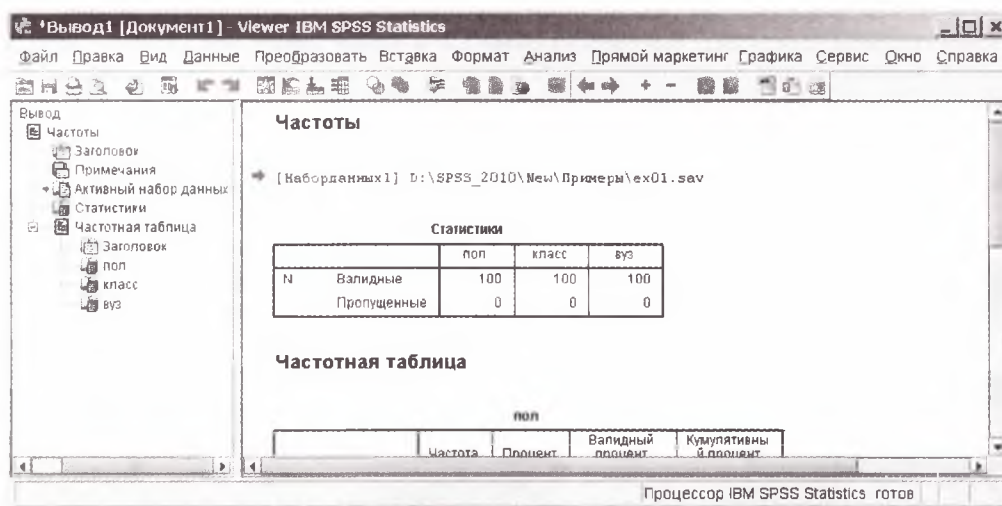


Рис. 6.2. Окно вывода программы SPSS

Переместив переменные в целевой список, вы завершаете выполнение операции щелчком на кнопке ОК. После этого программа SPSS формирует окно вывода с результатами выполнения команды. Выполнение практически всех команд статистических операций завершается именно так.

Окно с результатами работы программы имеет заголовок Вывод1 [Документ1]. Для просмотра результатов при необходимости можно воспользоваться полосой прокрутки в правой части окна. Структура выводимых результатов программы и основные приемы их редактирования описаны в разделе «Окно вывода и его редактирование» главы 2. Поскольку в заголовке окна видна строка меню, можно снова применить команду Частоты к выбранному подмножеству данных, для чего достаточно выбрать команду Анализ ► Описательные статистики ► Частоты и в открывшемся диалоговом окне сразу щелкнуть на кнопке ОК (ранее выбранные переменные окажутся в целевом списке автоматически), или предварительно изменить список анализируемых переменных.

Столбиковые диаграммы

Для того чтобы создать столбиковую диаграмму для дискретных данных, необходимо сначала выполнить все действия шага 5, кроме последнего, а затем в диалоговом окне Частоты щелкнуть на кнопке Диаграммы. На экране появится диалоговое окно Частоты: Диаграммы, показанное на рис. 6.3.

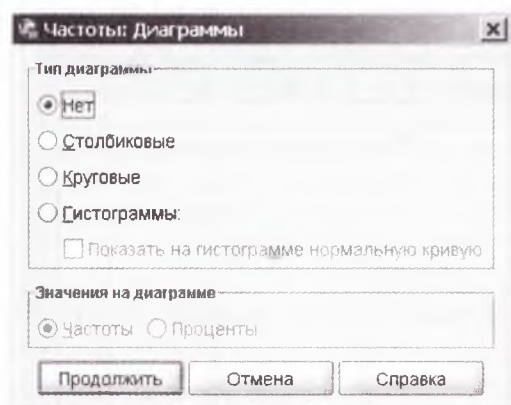


Рис. 6.3. Диалоговое окно Частоты: Диаграммы

Как видите, диалоговое окно Частоты: Диаграммы позволяет выбрать тип диаграммы с помощью переключателя. В данном примере вам понадобится установить переключатель Тип диаграммы в положение Столбиковые. В зависимости от величины, которую вы хотите использовать для отображения частот, установите переключатель Значения на диаграмме в положение Частоты или Проценты. Щелкните на кнопке Продолжить. После этого вы вернетесь в диалоговое окно Частоты и сможете завершить операцию щелчком на кнопке ОК.

ШАГ 5а

На этом шаге мы выполним построение столбиковых диаграмм для частот тех же трех переменных, пол, класс и вуз, которые использовались ранее. После выполнения шага 4 у вас должно быть открыто диалоговое окно Частоты, показанное на рис. 6.1. Если вы уже поработали с этим окном, очистите его нажатием кнопки Сброс. Для разнообразия будем заполнять целевой список двойными щелчками, а не щелчками на кнопке со стрелкой, как в предыдущем примере.

1. Дважды щелкните на переменной пол. Ее имя появится в поле Переменные. Сделайте то же самое для переменных класс и вуз и щелкните на кнопке Диаграммы, чтобы открыть диалоговое окно Частоты: Диаграммы.
2. Установите переключатель Тип диаграммы в положение Столбиковые. Если вы предпочитаете вместо частот выводить на диаграмме их доли от общей частоты в процентах, установите переключатель Значения на диаграмме в положение Проценты и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Частоты.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения этого шага программа сгенерирует три диаграммы, соответствующие выбранным переменным. При помощи полосы прокрутки можно просмотреть созданные диаграммы в окне вывода. Если возникнет необходимость отредактировать какую-либо из диаграмм, дважды щелкните на ней и обратитесь к материалу главы 5 для дальнейших действий. На рис. 6.4 показано окно вывода с диаграммой частот для переменной вуз.

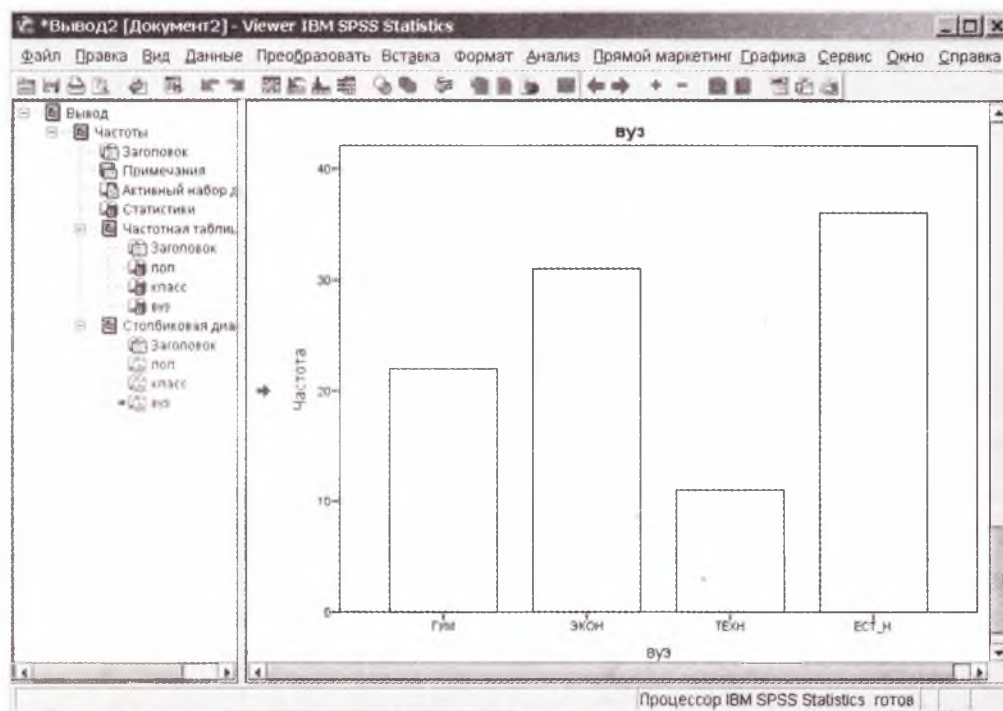


Рис. 6.4. Пример столбчатой диаграммы

Гистограммы

Процесс построения гистограмм аналогичен процессу построения столбчатых диаграмм. Единственным различием, которое следует всегда иметь в виду, является тот факт, что гистограммы предназначены для отображения распределения непрерывных переменных. Таким образом, для переменных пол, класс и вуз гистограммы непригодны. В то же время для переменной отметка2, отражающей среднюю отметку для каждого учащегося, гистограмма является весьма удобным средством описания распределения частот. Далее приведен пример построения гистограммы для переменной отметка2.

ШАГ 56

После выполнения шага 4 у вас должно быть открыто диалоговое окно Частоты, показанное на рис. 6.1. При необходимости повторите шаги 1–4. Если в целевом списке диалогового окна Частоты после работы над материалом предыдущего раздела остались переменные, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы переместить переменную в поле Переменные, и, наконец, на кнопке Диаграммы, чтобы открыть диалоговое окно Частоты: Диаграммы.
2. Установите переключатель Тип диаграммы в положение Гистограммы, если необходимо, установите флажок Показывать на гистограмме нормальную кривую и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Частоты.
3. В диалоговом окне Частоты снимите флажок Вывести частотные таблицы и щелкните на кнопке ОК.

Обратите особое внимание на необходимость снять флажок Вывести частотные таблицы. Этот флажок отвечает за наличие в выводимых данных команды Частоты всей неграфической информации. Если оставить его установленным при использовании непрерывных переменных, размер таблицы частот будет определяться всем множеством различных значений переменной, что редко оказывается полезным для исследователя. Поэтому при анализе непрерывных данных флажок Вывести частотные таблицы лучше снимать. Установленный флажок Показывать на гистограмме нормальную кривую указывает на то, что поверх гистограммы будет нанесена нормальная кривая.

Описательные статистики и процентили

Если вернуться к диалоговому окну Частоты и щелкнуть в нем на кнопке Статистики, на экране появится диалоговое окно Частоты: Статистики, показанное на рис. 6.5. В этом окне представлены такие показатели, как процентили и описательные статистики (подробно описательные статистики рассматриваются в главе 7). Из трех приведенных далее примеров вы узнаете, каким образом создать гистограмму, получить описательные статистики и вычислить процентили с заданным шагом или заданной величины. Во всех трех примерах используется переменная отметка2, а начинаются они с диалогового окна, полученного после выполнения шага 4 (см. рис. 6.1).

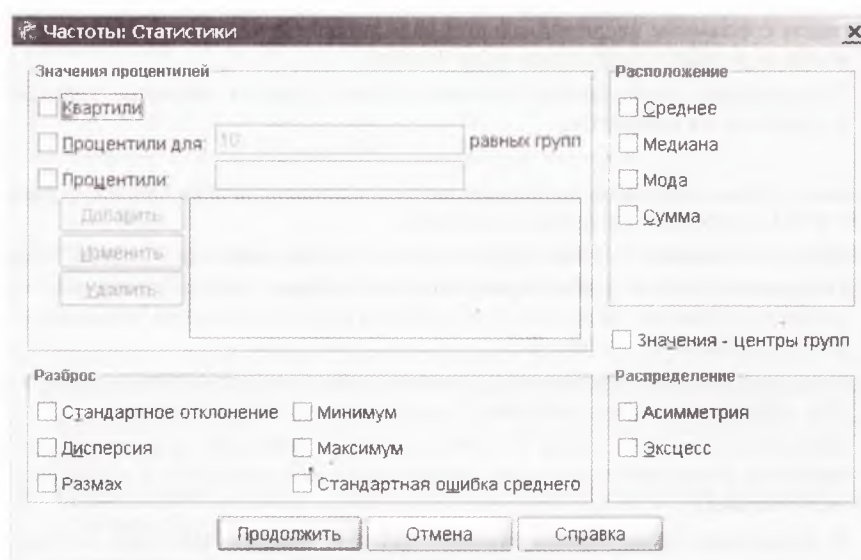


Рис. 6.5. Диалоговое окно Частоты: Статистики

ШАГ 5в

На этом шаге мы построим гистограмму распределения частот и вычислим для него среднее значение, стандартное отклонение, асимметрию и эксцесс распределения:

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, затем — на кнопке со стрелкой, чтобы переместить переменную в поле Переменные, и, наконец, на кнопке Диаграммы, чтобы открыть диалоговое окно Частоты: Диаграммы.



2. Установите переключатель Тип диаграммы в положение Гистограммы и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Частоты.
3. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Частоты: Статистики, показанное на рис. 6.5.
4. В области Расположение установите флажок Среднее, в области Распределение — флажки Асимметрия и Эксцесс, в области Разброс — флажок Стандартное отклонение и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Частоты.
5. В диалоговом окне Частоты снимите флажок Вывести частотные таблицы и щелкните на кнопке ОК.

ШАГ 5г



Чтобы вычислить процентиля переменной отметка2 с шагом 10, необходимо разделить шкалу процентов на 10 равных промежутков с границами, равными 10, 20, 30 и т. д.:

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, затем — на кнопке со стрелкой, чтобы переместить переменную в поле Переменные, и, наконец, на кнопке Статистики, чтобы открыть диалоговое окно Частоты: Статистики, показанное на рис. 6.5.
2. В области Значения процентов установите флажок Процентиля для, в поле рядом с флажком введите значение 10 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Частоты.
3. В диалоговом окне Частоты снимите флажок Вывести частотные таблицы и щелкните на кнопке ОК.

ШАГ 5д



Наконец, чтобы вычислить набор произвольных процентов (в данном случае 5, 70 и 95), выполните следующие действия:

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, затем — на кнопке со стрелкой, чтобы переместить переменную в поле Переменные, и, наконец, на кнопке Статистики, чтобы открыть диалоговое окно Частоты: Статистики, показанное на рис. 6.5.
2. В области Значения процентов установите флажок Процентиля, в поле рядом с флажком введите значение 5 и щелкните на кнопке Добавить.
3. Введите в то же поле числа 70 и 95, по очереди добавляя их в список процентов, и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Частоты.
4. В диалоговом окне Частоты снимите флажок Вывести частотные таблицы и щелкните на кнопке ОК.

Обратите внимание на то, что квартили (25-й, 50-й и 75-й процентиля) можно вычислить одним действием, установив флажок Квартили в области Значения процентов диалогового окна Частоты: Статистики (см. рис. 6.5).

Представление результатов

Несмотря на то что выводимые программой SPSS результаты занимают не очень много места, как правило, их можно представить в еще более компактной форме (см. раздел «Окно вывода и его редактирование» в главе 2).

На возможные небольшие различия между данными, которые вы видите на экране своего компьютера, и иллюстрациями этой книги можно не обращать внимания.

Частоты

Фрагмент окна вывода после выполнения шага 5 приведен на рис. 6.6.

пол					
		Частота	Процент	Валидный процент	Кумулятивный процент
Валидные	ЖЕН	61	61,0	61,0	61,0
	МУЖ	39	39,0	39,0	100,0
	Итого	100	100,0	100,0	

класс					
		Частота	Процент	Валидный процент	Кумулятивный процент
Валидные	А	33	33,0	33,0	33,0
	Б	35	35,0	35,0	68,0
	В	32	32,0	32,0	100,0
	Итого	100	100,0	100,0	

вуз					
		Частота	Процент	Валидный процент	Кумулятивный процент
Валидные	ГУМ	22	22,0	22,0	22,0
	ЭКОН	31	31,0	31,0	53,0
	ТЕХН	11	11,0	11,0	64,0
	ЕСТ_Н	36	36,0	36,0	100,0
	Итого	100	100,0	100,0	

Рис. 6.6. Фрагмент окна вывода после выполнения шага 5

Далее дана трактовка терминов, используемых программой в окне вывода:

- ▶ Частота — количество объектов, соответствующих каждой категории (градации) переменной;
- ▶ Процент — процент от общей численности (с учетом пропусков). Если бы в файле были пропущенные значения, их процент был бы указан в предпоследней строке;
- ▶ Валидный процент — процент значений для каждой категории за вычетом пропущенных значений;
- ▶ Кумулятивный процент — накопленный процент величины Валидный процент к данному значению переменной;
- ▶ Валидные — список значений переменной;
- ▶ Итого — сумма значений в столбце.

Гистограммы

На рис. 6.7 приведены результаты выполнения шага 5б. По горизонтальной оси отложены значения переменной с шагом, равным 0,125 (ширина каждого столбца тоже равна 0,125). Каждое указанное значение соответствует середине столбца. На вертикальной оси гистограммы отложены частоты диапазонов. Справа от гистограммы помещены вычисленные параметры: Среднее значение и Стандартное отклонение распределения, а также общее число объектов (N). Кроме того, на гистограмму нанесена нормальная

кривая, поскольку при выполнении процедуры был установлен флажок Показывать на гистограмме нормальную кривую. Гистограмма без труда может быть сделана более наглядной. В частности, для данной гистограммы желательно укрупнение интервалов и, соответственно, уменьшение количества столбцов. По вопросам редактирования гистограмм обратитесь к главе 5 (см. рис. 5.12).

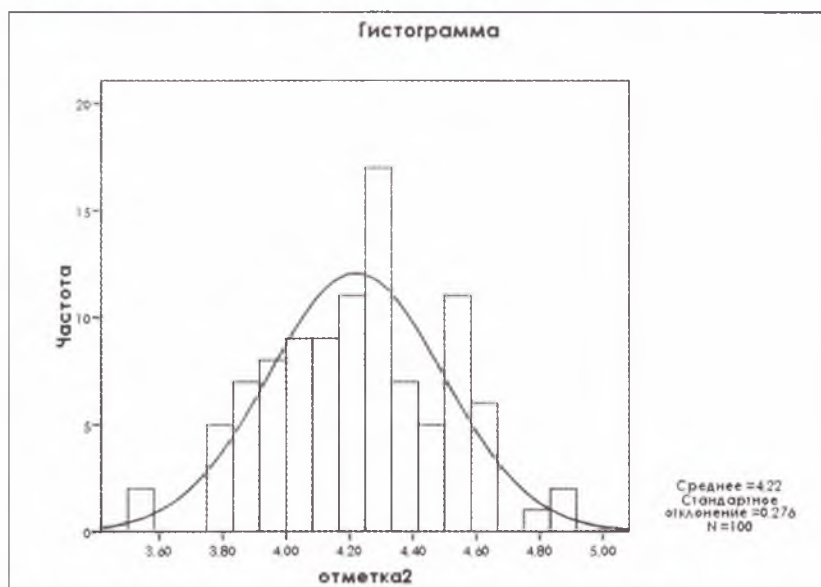


Рис. 6.7. Фрагмент окна вывода после выполнения шага 5б

Описательные статистики и процентиля

Результаты выполнения шагов 5г и 5д приведены на рис. 6.8.

Статистики		
отметка2		
N	Валидные	100
	Пропущенные	0
Среднее		4,2205
Стд. отклонение		,27589
Асимметрия		-,007
Стд. ошибка асимметрии		,241
Экцесс		-,357
Стд. ошибка эксцесса		,478

Статистики		
отметка2		
N	Валидные	100
	Пропущенные	0
Процентили	5	3,8000
	70	4,3500
	95	4,6500

Рис. 6.8. Фрагменты окна вывода после выполнения шагов 5г и 5д

Описательные статистики рассматриваются подробнее в главе 7, поэтому мы не станем обсуждать их здесь. Обратите внимание: если в вывод результатов включены Асимметрия и Экссесс, для них по умолчанию вычисляются стандартные ошибки (Стд. ошибка).

Что касается процентилей, то их можно трактовать следующим образом: 5 % значений переменной отметка² не превышают 3,8 (5 % учащихся имеют отметку не выше 3,8), 70 % значений не превышают 4,35 (30 % учащихся имеют отметку выше 4,35) и т. д.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» в главе 2). При необходимости отредактируйте диаграммы (см. главу 5).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

108	Пошаговый алгоритм вычислений
111	Представление результатов
112	Завершение анализа и выход из программы

Описательные статистики (descriptive statistics) — это различные вычисляемые показатели, характеризующие распределение значений переменной. Эти показатели условно можно разбить на несколько групп. Первая группа — меры центральной тенденции, вокруг которых «группируются» данные: среднее значение, медиана и мода. Вторая группа характеризует изменчивость значений переменной относительно среднего: стандартное отклонение и дисперсия. Диапазон изменчивости характеризуется минимумом, максимумом и размахом. Асимметрия и эксцесс представляют меру отклонения формы распределения от нормального вида. Кроме того, существуют величины, выражающие погрешности некоторых статистик: стандартная ошибка среднего, стандартная ошибка асимметрии и стандартная ошибка эксцесса. Последние два показателя вычисляются программой вместе с асимметрией и эксцессом по умолчанию. При помощи команды *Описательные* можно вычислить любую из указанных величин.

Меры центральной тенденции

Существует три основные меры центральной тенденции распределения:

- ▶ *среднее значение* (mean) равно сумме всех значений распределения, деленной на их количество. Для распределения [3 5 7 5 6 8 9] среднее значение равно $(3 + 5 + 7 + 5 + 6 + 8 + 9)/7 = 6,14$;
- ▶ *медиана* (median) определяется как значение, находящееся в середине распределения, полученного из исходного путем упорядочивания по возрастанию. Для распределения [3 5 7 5 6 8 9] медиана равна 6, поскольку значение 6, находится в центре последовательности [3 5 5 6 7 8 9];
- ▶ *мода* (mode) равна наиболее часто встречающемуся значению. В распределении [3 5 7 5 6 8 9] мода равна 5, поскольку число 5 встречается в нем дважды.

Меры изменчивости

Выделяют две величины, характеризующие изменчивость, или разброс, значений распределения относительно среднего:

- ▶ *дисперсия* (variance) равна сумме квадратов отклонений каждого значения от среднего, деленной на $N - 1$, где N — число значений в распределении. Для распределения [3 5 7 5 6 8 9] дисперсия равна $((3 - 6,14)^2 + (5 - 6,14)^2 + (7 - 6,14)^2 + (5 - 6,14)^2 + (6 - 6,14)^2 + (8 - 6,14)^2 + (9 - 6,14)^2)/6 = 4,1429$;
- ▶ *стандартное отклонение* (standard deviation) равно квадратному корню из дисперсии. Для распределения [3 5 7 5 6 8 9] стандартное отклонение равно 2,0354.

Стандартное отклонение является довольно наглядной и информативной для исследователя характеристикой распределения, а дисперсия, как правило, используется как вспомогательная величина в статистических вычислениях.

Характеристики диапазона распределения

Дополнительными мерами изменчивости являются 4 простые характеристики, отражающие границы распределения и его размах:

- ▶ **минимум** (minimum) равен наименьшему из значений распределения. Для распределения [3 5 7 5 6 8 9] минимум равен 3;
- ▶ **максимум** (maximum) равен наибольшему из значений распределения. Для распределения [3 5 7 5 6 8 9] максимум равен 9;
- ▶ **размах** (range) составляет разность между максимумом и минимумом распределения. Для распределения [3 5 7 5 6 8 9] размах равен $9 - 3 = 6$;
- ▶ **сумма** (sum) равна сумме всех значений распределения. Для распределения [3 5 7 5 6 8 9] сумма равна $3 + 5 + 7 + 5 + 6 + 8 + 9 = 43$.

Характеристики формы распределения

Для отражения близости формы распределения к нормальному виду существует две основные характеристики:

- ▶ **эксцесс** (kurtosis) является мерой сглаженности (остро- или плосковершинности) распределения. Если значение эксцесса близко к 0, это означает, что форма распределения близка к нормальному виду. Положительный эксцесс указывает на плосковершинное распределение, у которого максимум вероятности выражен не столь ярко, как у нормального. Значения эксцесса, превышающие 5,0, говорят о том, что по краям распределения находится больше значений, чем вокруг среднего. Отрицательный эксцесс, напротив, характеризует островершинное распределение, график которого более вытянут по вертикальной оси, чем график нормального распределения. Считается, что распределение с эксцессом в диапазоне от -1 до $+1$ примерно соответствует нормальному виду. В большинстве случаев вполне допустимо считать нормальным распределение с эксцессом, по модулю не превосходящим 2;
- ▶ **асимметрия** (skewness) показывает, в какую сторону относительно среднего сдвинуто большинство значений распределения. Нулевое значение асимметрии означает симметричность распределения относительно среднего значения, положительная асимметрия указывает на сдвиг распределения в сторону меньших значений, а отрицательная — в сторону больших значений. В большинстве случаев за нормальное принимается распределение с асимметрией, лежащей в пределах от -1 до $+1$. В исследованиях, не требующих высокой точности результатов, нормальным считают распределение с асимметрией, по модулю не превосходящей 2.

Стандартная ошибка

Стандартная ошибка (standard error) является характеристикой точности, или стабильности, величины, для которой она вычисляется. В контексте программы SPSS стандартная ошибка используется для среднего значения, асимметрии и эксцесса. Ее смысл заключается в следующем. Можно взять некоторое количество случайно выбранных

значений генеральной совокупности, составить выборку и вычислить для нее среднее значение. Повторив эту операцию несколько раз, вы получите набор средних значений выборок, которые также представляют собой некоторое распределение. Стандартное отклонение этого распределения и будет являться стандартной ошибкой для среднего значения генеральной совокупности. Аналогичным способом вычисляются стандартные ошибки для асимметрии и эксцесса. Чем меньше значение стандартной ошибки, тем выше стабильность величины, для которой она вычисляется.

Пошаговый алгоритм вычислений

Для применения команды Описательные мы обратимся к файлу ex01.sav. Количество объектов, или значений каждой переменной, в этом файле равно 100, поэтому при вычислении характеристик распределения для различных переменных программа будет считать N равным 100. Сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex01.sav.

ШАГ 4

В меню Анализ выберите команду Описательные статистики ► Описательные. На экране появится диалоговое окно Описательные статистики (рис. 7.1).

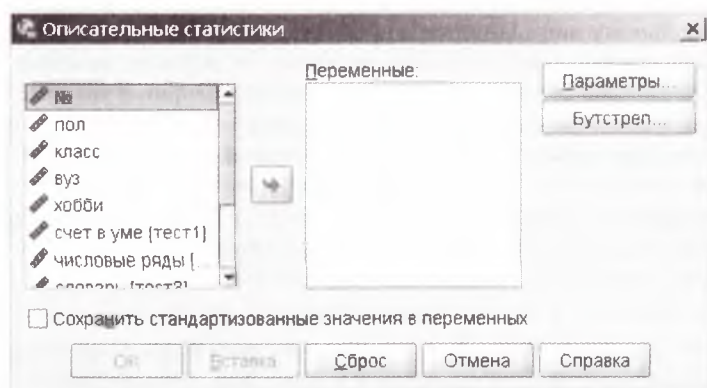


Рис. 7.1. Диалоговое окно Описательные статистики

В диалоговом окне Описательные статистики необходимо задать переменные, для которых будут вычислены описательные статистики. В левой части окна находится список всех доступных переменных текущего файла данных.

Для задания переменной щелчком выделите нужную переменную, а затем щелчком на кнопке с направленной вправо стрелкой переместите ее в поле Переменные. Если нужная переменная не видна в исходном списке, воспользуйтесь полосой прокрутки.

Для того чтобы удалить переменную из целевого списка, достаточно выделить ее, а затем воспользоваться кнопкой с направленной влево стрелкой, которая появляется на месте предыдущей кнопки при выделении любого пункта в списке Переменные. В этом

случае переменная вновь переместится в исходный список. Чтобы полностью очистить список Переменные, щелкните на кнопке Сброс.

Как вы могли заметить, помимо кнопок и списков диалоговое окно Описательные статистики содержит флажок Сохранить стандартизованные значения в переменных. Если этот флажок установлен, программа произведет z-преобразование (стандартизацию) всех выбранных переменных, создав таким образом новые переменные. Исходные переменные при этом останутся неизменными, а новым переменным будут присвоены старые имена, но начинающиеся с буквы z. Под стандартизованными (или z-преобразованными) значениями переменной понимается такое ее распределение, среднее значение которого равно 0, а стандартное отклонение — 1.

ШАГ 5

Для того чтобы создать таблицу описательных статистик, предлагаемую программой по умолчанию и включающую в себя среднее значение, стандартное отклонение, максимум и минимум, выполните следующие действия:

1. Щелкните сначала на переменной отметка1, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в поле Переменные.
2. Повторите те же действия для переменной отметка2.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Если вам понадобится вычислить дополнительные характеристики, не вычисляемые программой по умолчанию, перед щелчком на кнопке ОК нужно щелкнуть на кнопке Параметры, расположенной в верхнем правом углу диалогового окна Описательные статистики. Откроется диалоговое окно Описательные статистики: Параметры (рис. 7.2), в котором с помощью флажков можно задать все упоминавшиеся ранее характеристики, за исключением двух: медианы и моды. Эти характеристики доступны только через команды Частоты (глава 6) и Средние (глава 10). Кроме того, вы не увидите в диалоговом окне флажков, соответствующих стандартным ошибкам асимметрии и эксцесса, поскольку

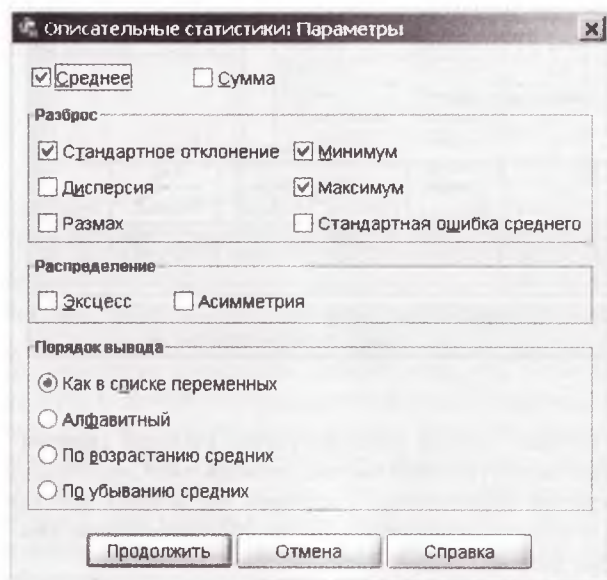


Рис. 7.2. Диалоговое окно Описательные статистики: Параметры

они всегда вычисляются автоматически. Установите флажки, соответствующие нужным характеристикам, и щелкните на кнопке Продолжить. При желании можно также установить в одно из четырех положений переключатель Порядок вывода. Положение Как в списке переменных установлено по умолчанию и означает, что в выводимых результатах переменные будут перечислены в том же порядке, в котором они представлены в файле данных. Положение Алфавитный указывает на то, что перечисление переменных будет происходить в алфавитном порядке. Положения По возрастанию средних и По убыванию средних позволяют отсортировать переменные в окне вывода по их вычисленным средним значениям.

Далее показано, каким образом можно организовать подсчет дополнительных статистик для переменных отметка1 и отметка2.

ШАГ 5а

После выполнения шага 4 у вас должно быть открыто окно Описательные статистики, приведенное на рис. 7.1.

1. Щелкните сначала на переменной отметка1, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы переместить переменную в поле Переменные.
2. Повторите те же действия для переменной отметка2 и щелкните на кнопке Параметры, чтобы открыть диалоговое окно Описательные статистики: Параметры.
3. Установите флажки для всех характеристик, которые следует вычислить, и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Описательные статистики.
4. Щелкните на кнопке ОК, чтобы открыть окно вывода, как на рис. 7.3.

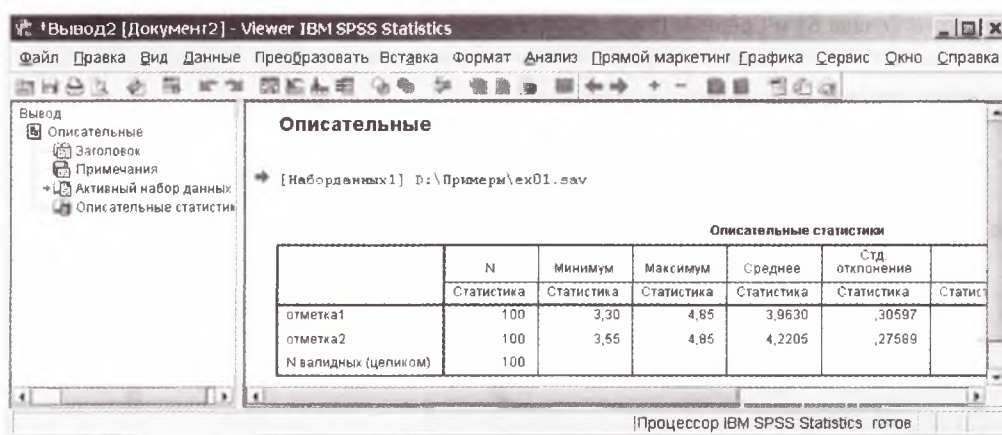


Рис. 7.3. Окно вывода SPSS

Для того чтобы полностью развернуть окно вывода, воспользуйтесь кнопкой разворачивания. Чтобы просматривать результаты внутри окна вывода, используйте вертикальную и горизонтальную полосы прокрутки. Как видите, в верхней части окна вывода имеется строка меню. Это означает, что можно выполнить и другие статистические процедуры, не обращаясь к окну редактора данных.

Представление результатов

Обратите внимание на то, что выводимые данные (см. рис. 7.3) не уместились на одной странице и целиком не видны на экране. Такая ситуация встречается довольно часто, и приходится пользоваться обеими полосами прокрутки, чтобы увидеть результаты анализа. Для быстрой навигации внутри окна вывода можно использовать иерархическую структуру в виде дерева объектов в левой части окна. Однако размер таблицы может быть таков, что при печати она не поместится на странице целиком.

Для решения этой проблемы рекомендуется воспользоваться средствами редактирования таблиц (см. раздел «Окно вывода и его редактирование» главы 2). Например, таблицу, приведенную на рис. 7.3, можно транспонировать (поменять местами строки и столбцы). Для этого необходимо войти в режим редактирования таблицы (двойным щелчком мыши на ней или через контекстное меню), а затем в появившемся меню Мобильная таблица выбрать пункт Транспонировать строки и столбцы. Иногда нужно изменить ширину того или иного столбца для лучшего отображения его содержимого. Чтобы сделанные изменения вступили в силу, следует просто выйти из режима редактирования таблиц.

На рис. 7.4 приведен один из возможных примеров улучшения вида окна вывода (см. рис. 7.3) после транспонирования таблицы в режиме ее редактирования. Как видите, таблица целиком поместилась на странице. В результат обработки включены Число значений (N), Минимум, Максимум, Среднее значение, Стандартное отклонение, Асимметрия и Эксцесс. Последние две величины приведены вместе со стандартными ошибками.

Описательные статистики		отметка1	отметка2	N валидных (целиком)
N	Статистика	100	100	100
Минимум	Статистика	3,30	3,55	
Максимум	Статистика	4,85	4,85	
Среднее	Статистика	3,9630	4,2205	
Стд. отклонение	Статистика	,30597	,27589	
Асимметрия	Статистика	,317	-,007	
	Стд. ошибка	,241	,241	
Эксцесс	Статистика	-,066	-,357	
	Стд. ошибка	,478	,478	

Рис. 7.4. Окно вывода после выполнения шага 5 и транспонирования таблицы

Приведенные результаты говорят о том, что в отношении рассматриваемых переменных доступны любые методы статистического анализа, так как значения асимметрии и эксцесса по модулю не превышают 1 для всех переменных. С терминологией мы познакомились ранее. Единственным исключением является слово «целиком» в последнем столбце таблицы. Оно указывает на то, что анализ проводился без учета тех объектов (строк), для которых пропущено хотя бы одно значение. Поскольку таких объектов в файле ex01.sav нет, число включенных в анализ объектов (100) равно числу исходных объектов.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» в главе 2).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (см. раздел «Сохранение, экспорт, перенос и печать результатов» в главе 2).

Для выхода из программы выберите в меню Файл команду Выход.

8 Таблицы сопряженности и критерий хи-квадрат

113	Таблицы сопряженности
114	Критерий независимости хи-квадрат
115	Пошаговый алгоритм вычислений
121	Представление результатов
123	Терминология, используемая при выводе
124	Завершение анализа и выход из программы

Таблицы сопряженности служат для описания связи двух или более номинальных (категориальных) переменных. Примерами номинальных переменных являются пол (женский, мужской), класс (А, Б, В), местность (город, пригород, село), ответ (да, нет) и т. д. Таблицы сопряженности неприменимы к непрерывным переменным, однако последние можно разбить на интервалы. Так, возраст человека, который следует считать непрерывным из-за большого количества его возможных значений, можно разделить на интервалы от 0 до 19 лет, от 20 до 39 лет, от 40 до 59 лет и т. д. Иногда такое разделение непрерывной переменной и ее последующее представление с помощью таблиц сопряженности делает результаты более наглядными. С точки же зрения статистического анализа переход от непрерывных (количественных) переменных к номинальным нежелателен, так как при этом теряется существенная часть информации о различии объектов. Например, два человека в возрасте 39 и 40 лет попадают в соседние возрастные категории и с точки зрения анализа ничем не отличаются от пары людей в возрасте 20 и 59 лет.

Для работы с таблицами сопряженности в программе SPSS используется команда Таблицы сопряженности.

Таблицы сопряженности

Обратимся к файлу ex01.sav. С помощью команды Частоты мы можем узнать, что среди школьников 39 юношей и 61 девушка, что 33 из них увлекаются спортом, 37 — компьютером и 30 — искусством. Однако команда Частоты не позволяет ответить на вопрос, сколько девушек увлекаются спортом или сколько юношей — искусством. Для этого в SPSS существует команда Таблицы сопряженности. Вполне логично, что для ответа на наш вопрос необходимо «сопрячь», или «пересечь», подмножество учащихся определенного пола с подмножеством учащихся с определенным увлечением. Такое сопряжение удобно представить в виде таблицы, строки которой соответствуют полу, столбцы — увлечению. Тогда в ячейке, находящейся, например, на пересечении строки «мужской» и столбца «искусство», мы увидим количество (частоту) юношей, которые увлекаются искусством. Поскольку в файле данных существуют 2 градации пола и 3 градации внешних увлечений (хобби), то наша перекрестная таблица будет состоять из $2 \cdot 3 = 6$

ячеек. Можно составлять и более сложные таблицы сопряженности, включающие в себя три и более переменные, но эта операция имеет смысл лишь для больших объемов данных, иначе частоты большинства ячеек будут малыми или нулевыми. Рассмотрим, что произойдет, если для данных файла ex01.sav создать таблицу сопряженности пол × хобби × класс × вуз. Эта таблица будет содержать $2 \cdot 3 \cdot 3 \cdot 4 = 72$ ячейки при том, что число объектов составляет лишь 100. Большинство ячеек этой таблицы сопряженности будут иметь значения от 0 до 1–2, что делает ее малоинформативной и статистически недостоверной. При задании этих четырех номинальных переменных программа SPSS вместо четырехмерной таблицы сопряженности построит 12 двумерных таблиц размерностью 2×3 , «вложенных» в одну таблицу.

Критерий независимости хи-квадрат

Помимо частот (или *наблюдаемых* величин) SPSS может вычислять *ожидаемые* значения для каждой ячейки таблицы. Ожидаемое значение вычисляется в предположении, что две номинальные переменные независимы друг от друга. Рассмотрим простой пример. Пусть в комнате находятся 100 человек, из которых 30 являются мужчинами, а 70 — женщинами. Если известно, что из этих 100 человек 10 увлекаются искусством, в случае, если увлечение не зависит от пола, следует ожидать, что из 10 увлекающихся искусством 3 являются мужчинами, а 7 — женщинами. Сопоставляя эти ожидаемые частоты с наблюдаемыми частотами, мы можем судить о том, действительно ли два номинальных признака независимы. Чем больше расхождение наблюдаемых и ожидаемых частот, тем сильнее эти два признака связаны друг с другом. Целью применения критерия независимости χ^2 и является установление степени соответствия между наблюдаемыми и ожидаемыми значениями ячеек.

В основе критерия независимости лежит вычисление величины χ^2 , определяемой как сумма отношений суммы квадратов отклонений наблюдаемой величины f_0 от ожидаемой величины f_e к ожидаемой величине в каждой ячейке:

$$\chi^2 = \sum (f_0 - f_e)^2 / f_e.$$

Как можно видеть из формулы, при больших отклонениях f_0 от f_e величина χ^2 также становится большой. Вместе с χ^2 вычисляется p -уровень значимости. При $p > 0,05$ считается, что различия между наблюдаемыми и ожидаемыми значениями незначительны. В противном случае предположение о независимости двух номинальных переменных отклоняется и делается вывод о том, что две классификации (переменные) зависят друг от друга. Более подробные описания величин, связанных с критерием χ^2 , вы можете найти в разделе «Представление результатов» этой главы.

Иногда данные не соответствуют требованиям критерия χ^2 , например, ожидаемая частота в некоторых ячейках таблицы сопряженности менее 5. Тогда можно воспользоваться одним из двух точных методов расчета значимости: методом статистических испытаний «Монте-Карло» или точными методами. Метод Точные по определению дает более точные результаты. Однако в некоторых случаях его расчет слишком трудоемок даже для компьютера, тогда используется метод Монте-Карло.

Зачастую исследователи воспринимают величину χ^2 ошибочно как величину *силы связи* между переменными. Однако это не так, поскольку χ^2 в значительной степени определяется количеством ячеек таблицы сопряженности и размером выборки. Таким образом, сравнение двух значений χ^2 , полученных при разных условиях, становится бессмыслен-

ным. По этой причине К. Пирсон предложил коэффициент ϕ , получаемый извлечением квадратного корня из отношения χ^2 к размеру выборки N . Целью введения новой величины было получение наглядной интерпретации связи между переменными в виде коэффициента, лежащего в пределах от 0 до 1 и принимающего нулевое значение для независимых переменных и единичное значение для строго связанных переменных. Однако цель не была достигнута полностью: если одна из переменных таблицы сопряженности имеет более двух градаций, значение ϕ (фи) может превышать 1. Крамеру удалось исправить последний недостаток путем введения коэффициента $V = \chi^2 / [N(k - 1)]$, где k — наименьшее из чисел градаций двух переменных. Этот коэффициент всегда принимает значения от 0 до 1 и служит характеристикой силы связи между переменными.

Пошаговый алгоритм вычислений

Для применения команды Таблицы сопряженности мы снова обратимся к файлу ex01.sav. Количество объектов, или наборов значений, в этом файле равно 100, поэтому программа будет считать N равным 100. Мы создадим несколько таблиц сопряженности, а также применим критерий χ^2 для определения зависимости переменных пол и хобби, но сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex01.sav.

ШАГ 4

В меню Анализ выберите команду Описательные статистики ► Таблицы сопряженности. На экране появится диалоговое окно Таблицы сопряженности (рис. 8.1).

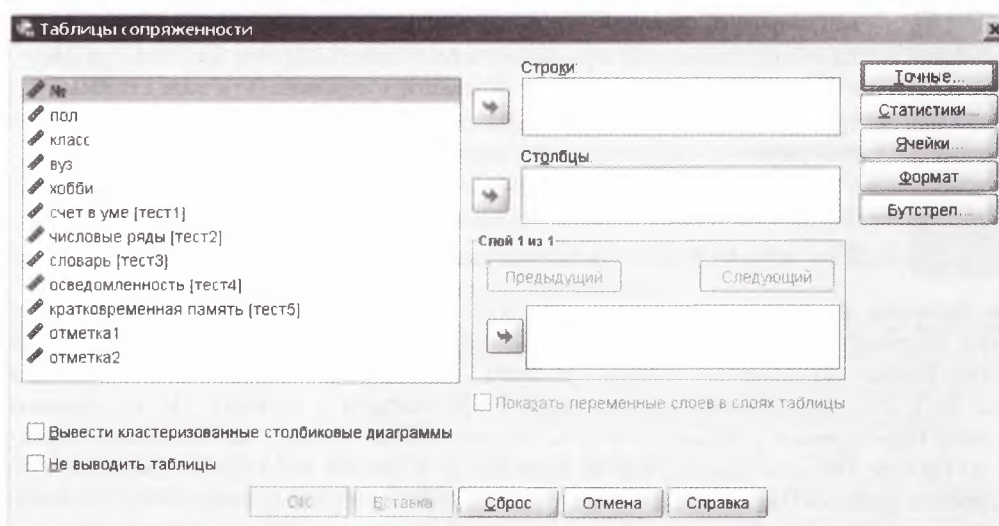


Рис. 8.1. Диалоговое окно Таблицы сопряженности

С помощью диалогового окна Таблицы сопряженности можно задать конфигурацию будущей таблицы сопряженности и столбиковую диаграмму для ее иллюстрации. Для этого требуется выбрать переменные в исходном списке и переместить их в целевые списки Строки и Столбцы, используя верхнюю и среднюю кнопки со стрелкой. Это позволит определить, значения какой переменной будут составлять строки таблицы, а какой — столбцы. Поместив в поле Строки переменную пол, в поле Столбцы — переменную хобби и щелкнув затем на кнопке ОК, вы создадите таблицу с двумя строками и тремя столбцами. Для создания столбиковой диаграммы следует установить флажок Вывести кластеризованные столбиковые диаграммы.

Область Слой 1 из 1 диалогового окна позволяет построить таблицу сопряженности для трех и более переменных. Если бы мы захотели подключить к таблице, содержащей переменные пол и хобби, еще и переменную класс, последнюю следовало бы поместить в нижнее поле при помощи нижней кнопки со стрелкой. В результате вы увидели бы на экране таблицу, разбитую на три части, каждая из которых представляла бы собой таблицу сопряженности пол × хобби для соответствующего значения переменной класс (1, 2 и 3). Кнопки Предыдущий и Следующий предназначены для случаев, когда необходимо построить таблицы сопряженности с более чем одной дополнительной переменной. Например, если бы вам понадобилось составить таблицы сопряженности с переменными пол, хобби, класс и вуз, в нижнем списке нужно было бы указать две переменные, класс и вуз, а в итоге программой была бы сгенерирована общая таблица сопряженности, включающая в себя 3 таблицы пол × хобби для каждой градации переменной класс и 4 таблицы пол × хобби для каждого значения переменной вуз.

В следующем примере мы создадим таблицу сопряженности пол × хобби × класс и проиллюстрируем ее столбиковыми диаграммами.

ШАГ 5

При открытом диалоговом окне Таблицы сопряженности (см. рис. 8.1) выполните следующие действия:

1. Щелкните сначала на переменной пол, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Строки.
2. Щелкните сначала на переменной хобби, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Столбцы.
3. Щелкните сначала на переменной класс, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле для дополнительных переменных в области Слой 1 из 1.
4. Установите флажок Вывести кластеризованные столбиковые диаграммы.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Как правило, исследователи редко ограничиваются вычислением частот, предпочитая дополнять процедуру заданием разнообразных параметров. При щелчке на кнопке Ячейки открывается диалоговое окно Таблицы сопряженности: Вывод в ячейках (рис. 8.2). Это окно позволяет управлять информацией в ячейках. По умолчанию флажок Наблюдаемые в области Частоты установлен, поскольку наблюдаемые частоты являются главной вычисляемой величиной в любой таблице сопряженности. Установка флажка Ожидаемые позволит вычислять в ячейках и ожидаемые частоты. Остальные параметры задают в зависимости от целей и предпочтений конкретного исследователя.

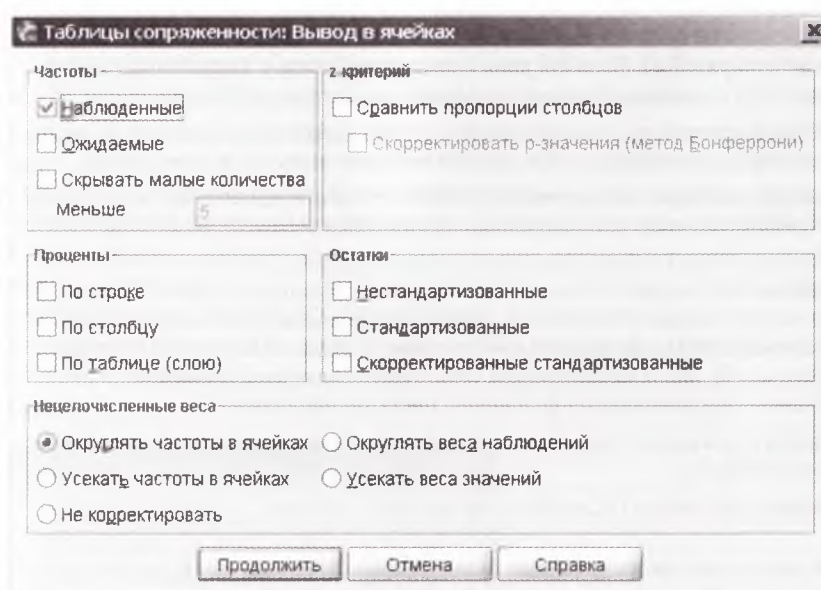


Рис. 8.2. Диалоговое окно Таблицы сопряженности: Вывод в ячейках

Далее описано назначение некоторых флажков, управляющих содержимым ячеек в диалоговом окне Таблицы сопряженности: Вывод в ячейках:

- ▶ При установке флажка Наблюдаемые в области Частоты отображается реальное количество объектов, соответствующее каждой ячейке.
- ▶ При установке флажка Ожидаемые в области Частоты отображается значение ожидаемой частоты для каждой ячейки, вычисленное в предположении, что соответствующие переменные являются независимыми.
- ▶ При установке флажка По строке в области Проценты отображается частота ячейки в процентах от суммарной частоты строки, в которой она находится.
- ▶ При установке флажка По столбцу в области Проценты отображается частота ячейки в процентах от суммарной частоты столбца, в котором она находится.
- ▶ При установке флажка По таблице (слою) в области Проценты отображается частота ячейки в процентах от суммарной частоты всей таблицы сопряженности.
- ▶ При установке флажка Нестандартизованные в области Остатки отображается разность между наблюдаемой и ожидаемой частотами.
- ▶ При установке флажка Сравнить пропорции столбцов в области z-критерий появляется индикация столбцов, которые статистически достоверно различаются или не различаются. Дополнительный флажок Скорректировать p-значения (метод Бонферрони) вводит поправку для p -уровня, если количество сравниваемых столбцов более 2.

В следующем примере мы создадим таблицу сопряженности пол $\times$ хобби, в ячейки которой включим наблюдаемые и ожидаемые частоты, процент по строке и нестандартизированный остаток. Также добавим индикацию статистической значимости различия столбцов.

ШАГ 5а

После выполнения шага 4 у вас должно быть открыто диалоговое окно Таблицы сопряженности (рис. 8.1). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной пол, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Строки.
2. Щелкните сначала на переменной хобби, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Столбцы.
3. Щелкните на кнопке Ячейки, чтобы открыть диалоговое окно Таблицы сопряженности: Вывод в ячейках, и установите флажки: в области Частоты — Ожидаемые, в области Проценты — По строке, в области Остатки — Нестандартизованные. В области z-критерий установите флажки Сравнить пропорции столбцов и Скорректировать p-значения (метод Бонферрони).
4. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Таблицы сопряженности.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Обратите внимание на то, что мы не устанавливали флажок Наблюдаемые, поскольку он уже установлен программой по умолчанию.

В приведенных примерах мы научились создавать таблицы сопряженности и заполнять их ячейки различными показателями. Как правило, на практике при создании таблиц сопряженности применяется критерий χ^2 . Для настройки параметров анализа необходимо в диалоговом окне Таблицы сопряженности: Статистики щелкнуть на кнопке Статистики. Программа откроет диалоговое окно Таблицы сопряженности: Статистики (рис. 8.3). В этом окне с помощью флажков можно выбрать желаемые статистические критерии и меры независимости и согласия распределений. Как и в ходе работы в окне Вывод в ячейках, вам необходимо установить нужные флажки, а затем вернуться в диалоговое окно Таблицы

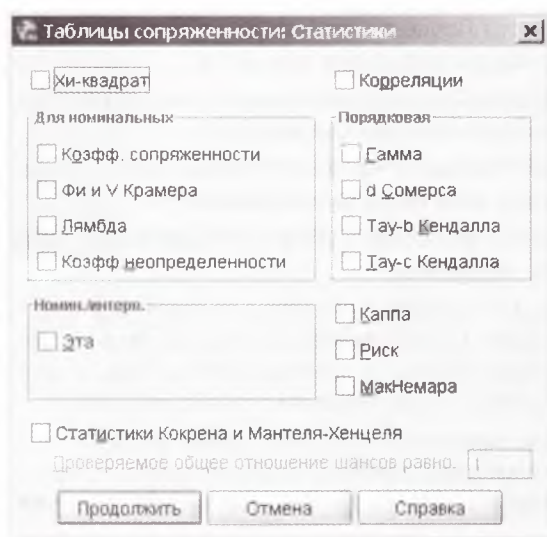


Рис. 8.3. Диалоговое окно Таблицы сопряженности: Статистики

сопряженности для завершения анализа. Описания остальных критериев и показателей, флажки для которых есть в диалоговом окне Таблицы сопряженности: Статистики, можно найти в справочной системе или документации к программе SPSS for Windows.

Если в диалоговом окне Таблицы сопряженности щелкнуть на кнопке Точные, появляется окно, позволяющее выбрать один из точных методов расчета значимости (рис. 8.4). По умолчанию переключатель в этом окне установлен в положении Только асимптотически, что соответствует определению приближительной значимости традиционным способом. Если данные не соответствуют ограничениям критерия χ^2 , при помощи переключателя можно воспользоваться одним из двух точных методов расчета значимости: методом статистических испытаний Монте-Карло или собственно точным критерием. Метод Точные по определению дает более точные результаты. Однако в некоторых случаях его расчет слишком трудоемок даже для компьютера, тогда используется метод Монте-Карло.

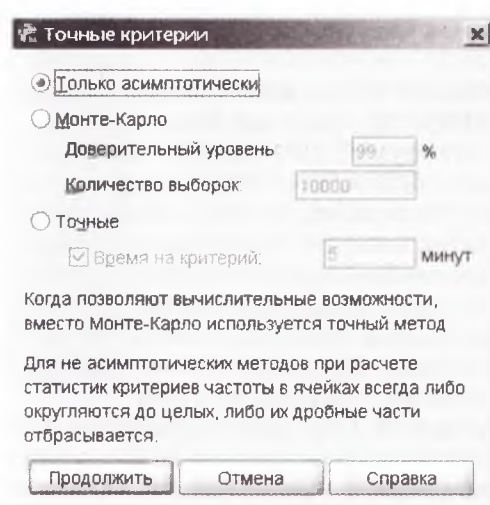


Рис. 8.4. Диалоговое окно Точные критерии

В следующем примере мы создадим таблицу сопряженности пол $\times$ хобби, в ячейки которой включим наблюдаемые частоты, проценты по строке, индикацию статистической значимости различия столбцов, применим критерий χ^2 и точный критерий, вычислим коэффициенты Фи и V Крамера. Анализ корреляций между переменными в данном случае не имеет смысла, так как переменные не являются количественными.

ШАГ 56

После выполнения шага 4 у вас должно быть открыто диалоговое окно Таблицы сопряженности, представленное на рис. 8.1. Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной пол, чтобы выделить ее, а затем — на верхней кнопке со стрелкой, чтобы переместить переменную в поле Строки.
2. Щелкните сначала на переменной хобби, чтобы выделить ее, а затем — на средней кнопке со стрелкой, чтобы переместить переменную в поле Столбцы.

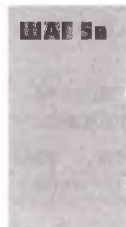


3. Щелкните на кнопке Ячейки, чтобы открыть диалоговое окно Таблицы сопряженности: Вывод в ячейках, и установите флажки По строке, Сравнить пропорции столбцов, Скорректировать р-значения (метод Бонферрони).
4. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Таблицы сопряженности.
5. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Таблицы сопряженности: Статистики, и установите флажки Хи-квадрат и Фи и V Крамера.
6. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Таблицы сопряженности.
7. Щелкните по кнопке Точные, чтобы открыть окно Точные критерии, и установите переключатель в положение Точные. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Таблицы сопряженности.
8. Установите флажок Вывести кластеризованные столбиковые диаграммы.
9. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Зачастую приходится строить таблицы сопряженности и применять критерий χ^2 не для всей выборки объектов, а только для их части. Например, может возникнуть необходимость в создании таблицы пол $\times$ хобби только для двух из трех градаций переменной хобби: спорт и искусство. Это делается следующим образом. После выбора переменных для построения таблицы сопряженности, определения набора величин, вычисляемых для каждой ячейки, и задания желаемых статистик выберите в меню Данные команду Отобразить наблюдения. На экране появится одноименное диалоговое окно (см. раздел «Выбор наблюдений для анализа» в главе 4), в котором переключатель Выбрать следует установить в положение Если выполнено условие, а затем щелкнуть на расположенной рядом кнопке Если. Откроется диалоговое окно Отобразить наблюдения: Условие. Элементы интерфейса этого окна позволяют выполнять операции, о которых шла речь в главе 4. В данном случае нас будет интересовать выбор уровней 1 и 3 переменной хобби. Для исключения уровня 2 из анализа выделите переменную хобби в исходном списке, введите ее в условие отбора, щелкнув на кнопке с направленной вправо стрелкой, на панели калькулятора щелкните на кнопке $\sim =$ (не равно) и введите в условие отбора число 2. После этого щелкните сначала на кнопке Продолжить, чтобы вернуться в диалоговое окно Отобразить наблюдения, а затем на кнопке ОК, чтобы вернуться в диалоговое окно Таблицы сопряженности. Наконец, щелчок на кнопке ОК в окне Таблицы сопряженности приведет к построению таблицы сопряженности с двумя выбранными градациями переменной хобби и двумя градациями переменной пол.

Для того чтобы в дальнейшем анализе участвовали все наблюдения, не забудьте в диалоговом окне Отобразить наблюдения вернуть переключатель в положение Все наблюдения.

Реализуем описанный алгоритм в виде пошаговых инструкций.



После выполнения шага 4 у вас должно быть открыто диалоговое окно Таблицы сопряженности (см. рис. 8.1). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной пол, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Строки.
2. Щелкните на переменной хобби, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Столбцы.



3. Щелкните на кнопке Ячейки, чтобы открыть диалоговое окно Таблицы сопряженности: Вывод в ячейках, и установите флажки По строке и Сравнить пропорции столбцов.
4. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Таблицы сопряженности.
5. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Таблицы сопряженности: Статистики, и установите флажки Хи-квадрат, Фи и V Крамера.
6. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Таблицы сопряженности.
7. В меню Данные выберите команду Отобразить наблюдения, чтобы открыть одноименное диалоговое окно, установите переключатель в положение Если выполнено условие и щелкните на расположенной рядом кнопке Если, открыв этим диалоговое окно Отобразить наблюдения: Условие.
8. В списке доступных переменных щелкните сначала на переменной хобби, чтобы выделить ее, а затем на кнопке с направленной вправо стрелкой, чтобы ввести переменную в условие отбора. Дополните условие отбора оператором $\sim =$ (не равно) и числом 2, используя панель калькулятора и клавиатуру, и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Отобразить наблюдения.
9. Щелкните на кнопке ОК, чтобы вернуться в диалоговое окно Таблицы сопряженности.
10. Щелкните на кнопке ОК, чтобы открыть окно вывода.
11. В меню Данные выберите команду Отобразить наблюдения, чтобы открыть одноименное диалоговое окно, установите переключатель в положение Все наблюдения и щелкните на кнопке ОК.

После выполнения шага 5, 5а, 5б или 5в программа автоматически активизирует окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

На рис. 8.5 и 8.6 приведены фрагменты выводимых результатов, сгенерированные программой после выполнения шага 5б.

Первое, на что необходимо обратить внимание при интерпретации выводимых результатов, — это соотношение частот (процентов) в строках таблицы сопряженности. Беглый взгляд на первую таблицу говорит о том, что эти величины заметно различаются. Подстрочные буквы указывают, какие ячейки в строке различаются статистически значимо, а какие — нет. Буква *a* в первых двух ячейках свидетельствует о том, что они не различаются статистически достоверно ($p > 0,05$), а буква *b* в третьей ячейке указывает на то, что частота в ней отличается статистически достоверно от частот как в первой, так и во второй ячейках ($p \leq 0,05$). В соответствии с установленным флажком Скорректировать *p*-значения (метод Бонферрони) введена поправка, необходимая при сравнении более чем двух ячеек. Различие частот в ячейках свидетельствует о том, что переменные

пол и хобби связаны друг с другом. Это подтверждается содержанием второй таблицы: относительно большим значением критерия χ^2 (15,405) и малым значением p -уровня ($p < 0,001$), то есть высокой статистической значимостью. Столбчатая диаграмма демонстрирует, что связь пол и хобби обусловлена прежде всего различным отношением юношей и девушек к выбору искусства в качестве хобби. Отметим, что результат применения точного критерия Точная значимость (2-стор.) идентичен результату применения традиционного критерия χ^2 (Асимпт. значимость (2-стор.)) в связи с тем, что не нарушены условия применения критерия χ^2 : нет ячеек с частотой менее 5. Если бы условия применения этого критерия были нарушены (свыше 25 % ячеек с частотой менее 5), следовало бы ориентироваться на критерий Точная значимость (2-стор.).

О величине связи переменных можно судить по симметричным мерам — значениям показателей Фи и V Крамера, которые аналогичны коэффициенту корреляции (см. рис. 8.5). Величина 0,392 свидетельствует об умеренной связи между переменными.

Таблица сопряженности пол * хобби

			хобби			Итого
			спорт	компьютер	искусство	
пол	ЖЕН	Частота	15 _a	19 _a	27 _b	61
		% в пол	24,6%	31,1%	44,3%	100,0%
	МУЖ	Частота	18 _a	18 _a	3 _b	39
		% в пол	46,2%	46,2%	7,7%	100,0%
Итого		Частота	33	37	30	100
		% в пол	33,0%	37,0%	30,0%	100,0%

Каждая подстрочная буква обозначает набор хобби категорий, для которых пропорции столбцов значимо не различаются между собой на уровне 0,5

Критерии хи-квадрат

	Значение	ст. св.	Асимпт. значимость (2-стор.)	Точная значимость (2-стор.)	Точная значимость (1-стор.)	Вероятность в точке
Хи-квадрат Пирсона	15,405 ^a	2	,000	,000		
Отношение правдоподобия	17,504	2	,000	,000		
Точный критерий Фишера	16,702			,000		
Линейно-линейная связь	12,652 ^b	1	,000	,000	,000	,000
Кол-во валидных наблюдений	100					

a. В 0 (0%) ячейках ожидаемая частота меньше 5. Минимальная ожидаемая частота равна 11,70

b. Стандартизованная статистика равна -3,557.

Симметричные меры

		Значение	Прибл. значимость	Точная знч
Номинальная по номинальной	Фи	,392	,000	,000
	V Крамера	,392	,000	,000
Кол-во валидных наблюдений		100		

Рис. 8.5. Фрагменты окна вывода после выполнения шага 5б: таблицы

Для таблиц 2×2 при небольшом количестве объектов (до 40–50) следует применять критерий χ^2 с поправкой на непрерывность. А при нарушении условий его применения (малая частота в ячейках) — Точный критерий Фишера (2-стор.).

При получении таблиц сопряженности большой размерности следует помнить о проблеме малых значений частот. Малым считается значение ожидаемой частоты, меньшее 5. Предполагается, что, если свыше 25 % ячеек таблицы сопряженности имеют малые ожидаемые значения частот, вероятность ошибки χ^2 -критерия очень высока. В этом случае следует пользоваться точным критерием.

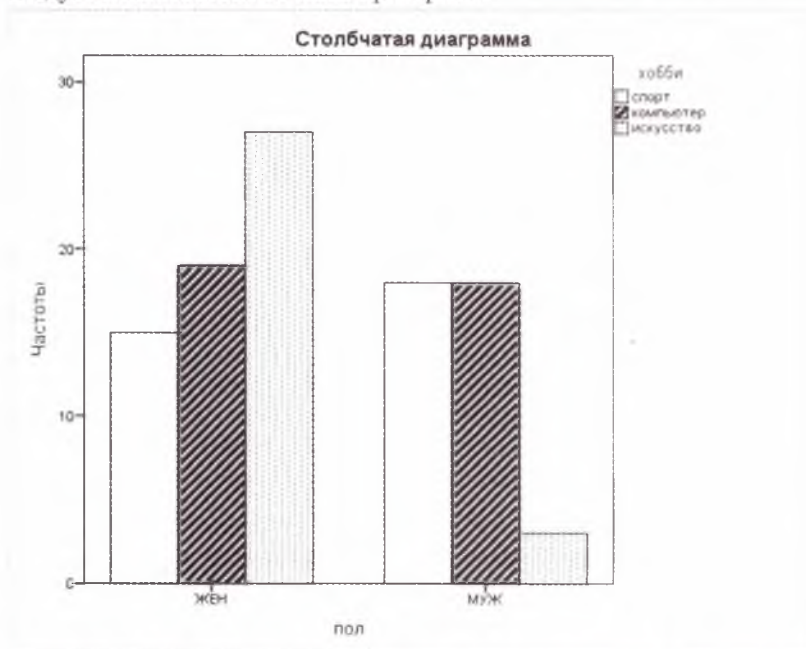


Рис. 8.6. Фрагмент окна вывода после выполнения шага 5б: диаграмма

Терминология, используемая при выводе

Трактовка терминов, используемых программой в окне вывода:

Хи-квадрат Пирсона и Поправка на непрерывность — два варианта критерия χ^2 . Второй вариант применяется для таблиц сопряженности размерностью 2×2 . При больших значениях N эти коэффициенты практически равны. Формула для хи-квадрата Пирсона выглядит следующим образом:

$$\chi^2 = \sum [(f_o - f_e)^2 / f_e].$$

Точный критерий Фишера — результаты расчета точной величины значимости по методу Фишера. Используется в тех случаях, когда применение критерия χ^2 к таблице 2×2 не является корректным, например при минимальной ожидаемой частоте в некоторых ячейках менее 5.

Значение — для критерия χ^2 значение тем больше, чем больше зависимость между переменными (как в нашем примере). Значения, близкие к 0, свидетельствуют о независимости переменных.

Ст. св. — степени свободы, произведение количеств градаций переменных, уменьшенных на 1. Для данной таблицы сопряженности число степеней свободы составляет $(2 - 1)(3 - 1) = 2$.

Асимпт. значимость — асимптотическая значимость, вероятность случайности связи или p -уровень значимости, то есть вероятность того, что связь является случайной. Чем меньше эта величина, тем выше статистическая значимость (достоверность) связи. Величина $p \leq 0,05$ свидетельствует о статистически значимом результате, который достоин содержательной интерпретации. Асимптотическая значимость определяется по традиционному критерию χ^2 .

Точная значимость — p -уровень значимости, вычисляемый точным методом, принимается во внимание, когда условия применения традиционного критерия нарушены (свыше 25 % ячеек таблицы сопряженности имеют частоту менее 5).

Точный критерий Фишера — вариант точного метода, который применяется для таблиц сопряженности размерностью 2×2 .

Линейно-линейная связь — статистический критерий, определяющий степень корреляции между переменными. Чаще всего результаты этого теста являются бессмысленными, поскольку порядок уровней переменных не имеет логической интерпретации. К примеру, нет смысла говорить об упорядоченной структуре хобби. Тем не менее, если бы одной из переменных был уровень дохода, упорядоченный по возрастанию, количественная оценка корреляции несла бы вполне определенную смысловую нагрузку.

Номинальная по номинальной — меры связи для двух номинальных переменных.

Фи — коэффициент, являющийся мерой связи двух переменных, аналог корреляции Пирсона. Значение $\phi = 0,392$ показывает умеренную связь между двумя переменными.

V Крамера — как и коэффициент Фи, этот коэффициент является мерой связи между двумя переменными, однако отличается тем, что всегда принимает значения от 0 до 1 и более приемлем для таблиц с $df > 2$.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» в главе 2). При необходимости отредактируйте диаграммы (см. главу 5).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы в меню Файл выберите команду Выход.

9 Корреляции

125	Понятие корреляции
126	Дополнительные сведения
128	Пошаговые алгоритмы вычислений
132	Представление результатов
133	Завершение анализа и выход из программы

Для вычисления корреляций между данными в программе SPSS используются команды подменю Корреляции меню Анализ. Корреляция представляет собой величину, заключенную в пределах от -1 до $+1$, и обозначается буквой r . Понятия «*корреляция*» и «*двумерная корреляция*» часто употребляются как синонимы, последнее означает «корреляция между двумя переменными» и подчеркивает, что рассматривается именно двумерное соотношение. Основным коэффициент корреляции r Пирсона предназначен для оценки связи между двумя переменными, измеренными с использованием метрической шкалы, распределение которых соответствует нормальному. Несмотря на то что величина r рассчитывается в предположении, что значения обеих переменных распределены по нормальному закону, формула для ее вычисления дает довольно точные результаты и в случаях аномальных распределений, а также когда одна из переменных является дискретной. Для распределений, не являющихся нормальными, предпочтительнее пользоваться ранговыми коэффициентами корреляции Спирмена или Кендалла. Команды подменю Корреляции позволяют вычислить как коэффициент Пирсона, так и коэффициенты Спирмена и Кендалла. Существуют и другие коэффициенты корреляции, применяющиеся для самых разных типов данных, однако их описание выходит за рамки темы этой книги.

Понятие корреляции

Корреляция, или *коэффициент корреляции*, — это статистический показатель вероятностной связи между двумя переменными, измеренными в единицах количественной шкалы. В отличие от функциональной связи, при которой каждому значению одной переменной соответствует строго определенное значение другой переменной, вероятностная связь характеризуется тем, что каждому значению одной переменной соответствует множество значений другой переменной. Примером вероятностной связи является связь между ростом и весом людей. Ясно, что один и тот же рост может быть у людей разного веса, как и наоборот. Значение коэффициента корреляции меняется от -1 до 1 . Крайние значения соответствуют линейной функциональной связи между двумя переменными, 0 — отсутствию связи.

Наглядное представление о связи двух переменных дает график двумерного рассеяния — соответствующая команда Рассеяние/точки имеется в меню Графика. На таком графике каждый объект представляет собой точку, координаты которой заданы значениями

двух переменных. Таким образом, множество объектов представляет собой на графике множество точек. По конфигурации этого множества точек можно судить о характере связи между двумя переменными.

Строгая положительная корреляция определяется значением $r = 1$. Термин «строгая» означает, что значения одной переменной однозначно определяются значениями другой переменной, а термин «положительная» — что с возрастанием значений одной переменной значения другой переменной также возрастают.

Строгая корреляция является математической абстракцией и почти не встречается в реальных исследованиях. Примером строгой корреляции является соответствие между временем пути и пройденным расстоянием при неизменной скорости.

Положительная корреляция соответствует значениям $0 < r < 1$. Положительную корреляцию следует интерпретировать следующим образом: если значения одной переменной возрастают, то значения другой имеют *тенденцию* к возрастанию. Чем коэффициент корреляции ближе к 1, тем сильнее эта тенденция, и наоборот, с приближением коэффициента корреляции к 0 тенденция ослабевает.

Примером значительной положительной корреляции служит зависимость между ростом и весом человека. Считается, что в этом случае коэффициент корреляции $r = 0,83$. Слабая положительная корреляция ($r = 0,12$) наблюдается между способностью человека к сочувствию и реальной помощью, которую он оказывает нуждающимся людям.

Отсутствие корреляции определяется значением $r = 0$. Нулевой коэффициент корреляции говорит о том, что значения переменных никак не связаны друг с другом. Примером пары величин с нулевой корреляцией является рост человека и результат его IQ-теста.

Отрицательная корреляция соответствует значениям $-1 < r < 0$. Если значения одной переменной возрастают, то значения другой имеют *тенденцию* к убыванию. Чем коэффициент корреляции ближе к -1 , тем сильнее эта тенденция, и наоборот, с приближением коэффициента корреляции к 0 тенденция ослабевает.

Слабая отрицательная корреляция ($r = -0,13$) наблюдается между агрессивностью человека по отношению к своему другу и помощью, которую он ему оказывает. Чем агрессивней человек, тем помощь меньше, однако зависимость выражена слабо. Примером значительной отрицательной корреляции ($r = -0,73$) служит зависимость между нервной возбудимостью человека и его эмоциональной уравновешенностью. Чем выше оказывается результат его теста на возбудимость, тем более низкий результат имеет его тест на уравновешенность.

Строгая отрицательная корреляция определяется значением $r = -1$. Она, так же как и строгая положительная корреляция, является абстракцией и не находит отражения в практических исследованиях. Пример, иллюстрирующий строгую отрицательную корреляцию, можно взять из школьного учебника физики: при равномерном движении расстояние равно произведению времени на скорость. При заданном расстоянии время и скорость являются обратно пропорциональными величинами: чтобы пройти путь за половину времени, необходимо идти вдвое быстрее.

Дополнительные сведения

Линейная и криволинейная корреляции

Основной коэффициент корреляции r Пирсона является мерой прямолинейной связи между переменными: его значения достигают максимума, когда точки на графике дву-

мерного рассеяния лежат на одной прямой линии. В реальной жизни отношения между переменными часто оказываются не только вероятностными, но и непрямолинейными, монотонными или немонотонными. Если связь нелинейная, но монотонная, вместо r Пирсона следует использовать ранговые корреляции Спирмена или Кендалла.

Нередко связь между двумя переменными является не только нелинейной, но и немонотонной. В качестве примера рассмотрим такие два фактора, как нервное возбуждение перед экзаменом и успешность его сдачи. Исследования показывают, что студенты, испытывающие умеренное нервное возбуждение, имеют наилучшие результаты на экзаменах, в то время как очень спокойные или очень нервные студенты сдают экзамены значительно хуже. Если по оси абсцисс отложить степень нервного возбуждения, а по оси ординат — результаты сдачи экзаменов, график зависимости между ними примет вид, близкий к перевернутой букве U. При этом любой коэффициент корреляции, вычисленный для этих величин, окажется весьма низким. Это объясняется тем, что для немонотонных отношений нужны другие методы оценки корреляции. Частично мы коснемся этих методов в главах 15 и 16, посвященных видам регрессионного анализа.

Прежде чем оценивать корреляцию двух переменных, рекомендуется построить график зависимости между ними — график двумерного рассеяния. Если график демонстрирует монотонность связи, для вычисления корреляции можно использовать команды подменю Корреляции.

Ранговые корреляции

Как уже отмечалось, необходимость в применении ранговых корреляций возникает в двух случаях: когда распределение хотя бы одной из двух переменных не соответствует нормальному и когда связь между переменными является нелинейной (но монотонной). В этих случаях вместо корреляции r Пирсона можно выбрать ранговые корреляции: r Спирмена либо τ (тау) Кендалла. Ранговыми они являются потому, что программа предварительно ранжирует переменные, между которыми они вычисляются.

Корреляцию r Спирмена программа SPSS вычисляет следующим образом. Сначала переменные переводятся в ранги, а затем к рангам применяется формула r Пирсона. Таким образом, r Спирмена интерпретируется по аналогии с r Пирсона. Иначе дело обстоит с корреляцией τ Кендалла, которая имеет вероятностную природу.

Рассмотрим принцип вычисления τ Кендалла на примере. Предположим, оценивается связь между ростом и весом в группе людей, предварительно ранжированных по этим переменным. Тогда при сравнении любых двух человек из этой группы возможны две ситуации: однонаправленное изменение переменных (совпадение), когда и рост, и вес одного больше, чем другого, и разнонаправленное изменение (инверсия), когда рост второго больше, а вес — меньше, чем первого. Перебрав все пары испытуемых, можно оценить вероятность совпадений (P) и вероятность инверсий (Q). Корреляция Кендалла — это разность вероятностей совпадений и инверсий: $\tau = P - Q$. По значению корреляции Кендалла можно всегда вычислить вероятность совпадений ($P = (1 + \tau)/2$) и инверсий ($Q = (1 - \tau)/2$). Например, если корреляция между ростом и весом $\tau = 0,5$, то вероятность совпадений (чем больше рост, тем больше вес) $P = 0,75$, а вероятность инверсий (чем больше рост, тем меньше вес) $Q = 0,25$. Таким образом, важным преимуществом корреляции τ Кендалла является ее отчетливая вероятностная интерпретация.

Значимость

Как и большинство статистических процедур, команды подменю Корреляции наряду с описательными статистиками (корреляциями в данном случае) вычисляют их уровень значимости. Напомним, что уровень значимости является мерой статистической достоверности результата вычислений, в данном случае корреляции, и служит основанием для интерпретации. Если исследование показало, что уровень значимости корреляции не превышает 0,05 ($p \leq 0,05$), это означает, что корреляция является случайной с вероятностью не более 5 %. Обычно это становится основанием для вывода о статистической достоверности корреляции. В противном случае ($p > 0,05$) связь признается статистически недостоверной и не подлежит содержательной интерпретации.

SPSS позволяет определять два теста значимости: односторонний и двухсторонний. Обычно используется двухсторонний тест значимости. Но если вы заранее знаете направление корреляции — положительное или отрицательное — и вас интересует только одно из них, можно использовать односторонний тест значимости. Однако такая ситуация встречается редко, а если и встречается, то правомерность односторонней проверки с трудом поддается обоснованию.

Частная корреляция

Понятие *частной корреляции* связано с ковариацией. Здесь мы упоминаем частную корреляцию лишь как одну из команд подменю Корреляции. Суть частной корреляции заключается в следующем. Если две переменные коррелируют, всегда можно предположить, что эта корреляция обусловлена влиянием третьей переменной как общей причины совместной изменчивости первых двух переменных. Для проверки этого предположения достаточно исключить влияние этой третьей переменной и вычислить корреляцию двух переменных без учета влияния третьей переменной (при фиксированных ее значениях). Корреляция, вычисленная таким образом, и называется частной. Например, при исследовании связи между скоростью чтения и зрелостью моральных суждений у детей разного возраста наверняка будет обнаружена корреляция этих двух переменных. Ответ на вопрос, связаны ли они непосредственно или связь обусловлена возрастом, позволяет дать частная корреляция. Если при фиксированных значениях возраста частная корреляция скорости чтения и зрелости моральных суждений приближается к нулю, можно заключить, что связь между этими переменными обусловлена возрастом.

Пошаговые алгоритмы вычислений

В пошаговых процедурах приведены несколько примеров, иллюстрирующих применение команд подменю Корреляции. В качестве файла данных используется файл `ex01.sav`. Сначала выполняются три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае `ex01.sav`). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом `ex01.sav`.

ШАГ 4

В меню Анализ выберите команду Корреляции ► Парные. На экране появится диалоговое окно Парные корреляции (рис. 9.1).

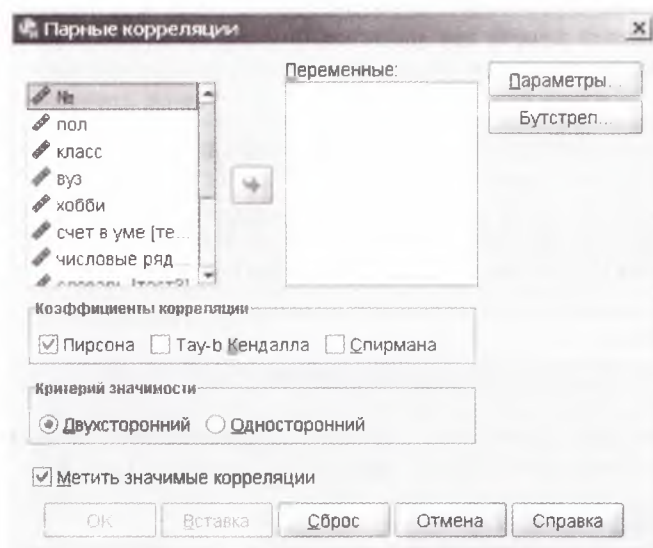


Рис. 9.1. Диалоговое окно Парные корреляции

Окно Парные корреляции позволяет настраивать параметры вычисления корреляций. В левом списке содержатся имена всех переменных, имеющих числовой тип. Строчные переменные для команды Корреляции ► Парные недоступны.

Чтобы выбрать переменные для вычисления корреляции, их требуется переместить в поле Переменные при помощи кнопки со стрелкой. Если несколько нужных переменных располагаются в исходном списке друг за другом, вы можете, наведя указатель мыши на верхнюю из них и нажав кнопку мыши, переместить указатель на нижнюю переменную и отпустить кнопку мыши, тем самым выделив сразу несколько переменных.

В области Коэффициенты корреляции по умолчанию установлен флажок Пирсона. Если требуется вычислить ранговые корреляции, то следует установить флажок Спирмена и (или) Тау-б Кендалла. Можете установить все три флажка, чтобы иметь возможность сравнивать три коэффициента корреляции для различных распределений данных.

Переключатель Критерий значимости по умолчанию установлен в положение Двухсторонний. Если вы заранее уверены в направлении (знаке) корреляции, то можете установить его в положение Односторонний.

Флажок Метить значимые корреляции по умолчанию установлен. Это означает, что корреляции, вычисленные с уровнем значимости от 0,01 до 0,05, будут помечены одной звездочкой (*), а с уровнем значимости от 0 до 0,01 — двумя звездочками (**). Вне зависимости от значимости в вывод включаются коэффициенты корреляции и *p*-уровни, вычисленные с точностью до трех знаков после запятой, а также количество объектов, участвовавших в процедуре.

В примерах, приведенных в этой главе, мы будем пользоваться коэффициентом корреляции Пирсона, двусторонним критерием значимости, а также помечать звездочками значимые корреляции. Если в ваших исследованиях понадобится использовать иную конфигурацию параметров, вы сами сможете легко их настроить, установив нужные флажки и переключатели. Отправной точкой при выполнении всех примеров служит диалоговое окно Парные корреляции.

ШАГ 5

На этом шаге мы создадим корреляционную матрицу для переменных тест1, ..., тест5 с вычислением корреляций Пирсона. После выполнения шага 4 должно быть открыто диалоговое окно Парные корреляции (см. рис 9.1).

1. Щелкните на переменной счет в уме [тест1], чтобы выделить ее. Удерживая клавишу Shift клавиатуры, щелкните на переменной кратковременная память [тест5] для выделения группы переменных тест1, ..., тест5.
2. Нажатием стрелки между окнами перенесите выделенную группу переменных в окно Переменные.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Кнопка Параметры позволяет задать дополнительные параметры корреляции. При щелчке на этой кнопке открывается диалоговое окно Парные корреляции: Параметры (рис. 9.2).

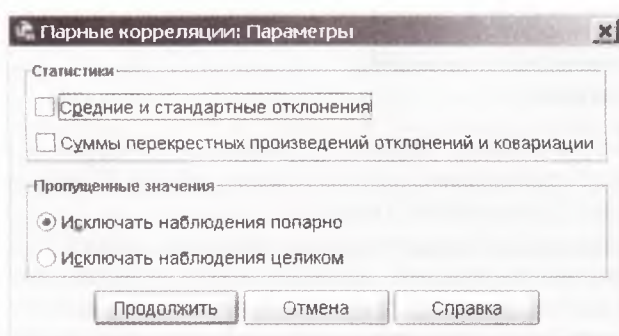


Рис. 9.2. Диалоговое окно Парные корреляции: Параметры

В области Статистики есть два флажка, управляющие отображением статистических величин: Средние и стандартные отклонения и Суммы перекрестных произведений отклонений и ковариации. Переключатель Пропущенные значения можно установить в одно из двух положений и тем самым выбрать способ исключения объектов, содержащих пропущенные значения. Установка в положение Исключать наблюдения попарно означает, что, если при вычислении корреляции между парой переменных для какого-нибудь объекта обнаружится отсутствующее значение, объект будет исключен из вычисления, но только для этой пары переменных. В результате может оказаться, что для разных пар переменных коэффициенты корреляции будут вычислены с разным количеством объектов. При установке переключателя в положение Исключать наблюдения целиком программа перед началом вычислительного процесса исключит из рассмотрения все объекты, содержащие хотя бы одно отсутствующее значение. В любом случае разрешение проблемы отсутствующих значений лучше выполнить до начала анализа.

В приведенном примере программа генерирует квадратную корреляционную матрицу. Нередко исследователю необходимо вычислить не все корреляции, а только их часть.

Например, можно представить ситуацию, когда возникает необходимость найти коэффициенты корреляции между одной переменной и 12 другими переменными, то есть вычислить неквадратную корреляционную матрицу. Возможности команд подменю Корреляция не позволяют этого делать. Конечно, можно сначала получить квадратную корреляционную матрицу (13×13), затем убрать лишние строки и столбцы при помощи редактора таблиц (см. главу 2, раздел «Окно вывода и его редактирование»). Но проще поступить по-другому — создать командный файл.

Для создания командного файла можно открыть окно *редактора синтаксиса*, выбрав команду Создать ► Синтаксис в меню Файл. Тогда придется вручную вводить все необходимые команды. Но в нашем случае проще открыть редактор синтаксиса нажатием кнопки Вставка диалогового окна Парные корреляции (см. рис. 9.1). Если переменные для анализа заданы, при нажатии на кнопку Вставка открывается окно редактора синтаксиса (рис. 9.3). Остается лишь внести необходимые изменения в уже существующий командный файл синтаксиса.

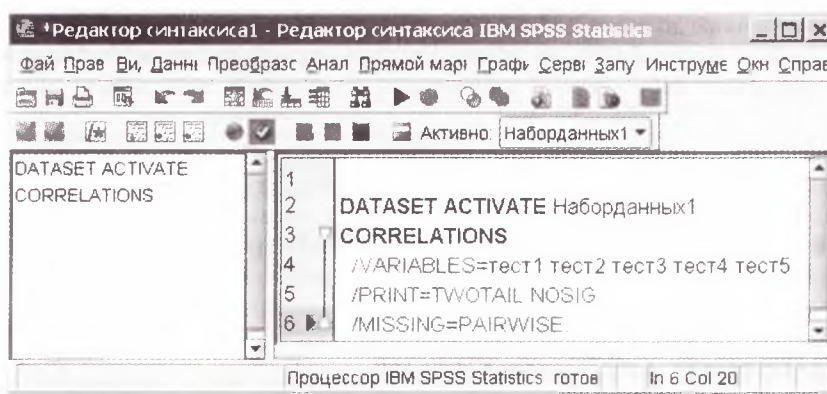


Рис. 9.3. Окно редактора синтаксиса

Чтобы выполнить введенную команду, ее нужно выделить и щелкнуть на кнопке со стрелкой Запустить выделенный фрагмент панели инструментов. Второй вариант запуска — выбрать в меню команду Запуск ► Все. Имейте в виду, что любой пропущенный знак, включая завершающую точку, или неверно написанное слово приведут к выдаче программой сообщения об ошибке.

ШАГ 5а

На этом шаге мы создадим корреляционную матрицу Пирсона размером 1×5 , предназначенную для оценки степени зависимости переменной отметка2 от переменных тест1, ..., тест5. После выполнения шага 4 должно быть открыто диалоговое окно Парные корреляции (см. рис 9.1). Если вы успели с ним поработать, очистите его щелчком на кнопке Сброс.

1. Щелкните на переменной счет в уме [тест1], чтобы выделить ее. Удерживая клавишу Shift клавиатуры, щелкните на переменной кратковременная память [тест5] для выделения группы переменных тест1, ..., тест5.
2. Нажатием стрелки между окнами перенесите выделенную группу переменных в окно Переменные.
3. Нажмите кнопку Вставка — откроется окно редактора синтаксиса (см. рис. 9.3).



4. В строке, где перечислены переменные, между именами тест5 и отметка2 введите с клавиатуры команду WITH так, чтобы до и после нее оставались пробелы:

```
CORRELATIONS
/VARIABLES=тест1 тест2 тест3 тест4 тест5 WITH отметка2
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE
```

5. Выберите в командной строке команду Запуск ► Все или выделите всю команду левой клавишей мыши и щелкните на кнопке Запустить выделенный фрагмент панели инструментов. Команда будет выполнена и откроется окно вывода.

В приведенной здесь конструкции можно произвольно изменять имена переменных. Допускается любое количество переменных как до, так и после ключевого слова WITH. Переменные, перечисленные до слова WITH, составят строки новой корреляционной матрицы, а перечисленные после него — столбцы. Если на экране появится сообщение об ошибке, то, скорее всего, либо неверно введено имя переменной, либо забыта точка в конце команды.

После выполнения шага 5 или 5а программа автоматически открывает окно вывода. Для просмотра результатов в нем при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Фрагмент выводимых данных, сгенерированных программой после выполнения шага 5, приведен на рис. 9.4.

Обратите внимание на структуру представленной на рисунке таблицы. Верхнее значение в каждой ячейке таблицы является коэффициентом корреляции между двумя соответствующими переменными, вычисленным с точностью до трех знаков после запятой. Следующее значение показывает уровень значимости этого коэффициента корреляции. Нижнее значение равно количеству объектов, участвовавших в вычислении коэффициента корреляции. При наличии пропусков в анализируемых данных это число в разных ячейках может быть разным. Надписи под таблицей поясняют смысл символов * и **, используемых в качестве меток для некоторых коэффициентов корреляции, и содержат информацию об уровне значимости.

Как легко заметить, на главной диагонали таблицы лежат единичные значения коэффициентов корреляции для всех переменных. Смысл этого результата прост: каждая переменная имеет строгую положительную корреляцию сама с собой.

После выполнения шага 5а с использованием редактора синтаксиса будет получена таблица результатов, представленная на рис. 9.5.

Значимая положительная корреляция в этой таблице наблюдается, в частности, между переменными кратковременная память (тест5) и отметка2 ($r = 0,294$, $p = 0,003$). Это означает, что чем лучше кратковременная память, тем выше средняя отметка за выпускной класс.

Корреляции		счет в уме	числовые ряды	словарь	осведомленность	кратковременная память
счет в уме	Корреляция Пирсона	1	,436**	-,051	-,134	,017
	Знач. (2-сторон)		,000	,617	,165	,867
	N	100	100	100	100	100
числовые ряды	Корреляция Пирсона	,436**	1	-,196	-,153	,015
	Знач. (2-сторон)	,000		,050	,127	,886
	N	100	100	100	100	100
словарь	Корреляция Пирсона	-,051	-,196	1	,441**	,483**
	Знач. (2-сторон)	,617	,050		,000	,000
	N	100	100	100	100	100
осведомленность	Корреляция Пирсона	-,134	-,153	,441**	1	,475**
	Знач. (2-сторон)	,165	,127	,000		,000
	N	100	100	100	100	100
кратковременная память	Корреляция Пирсона	,017	,015	,483**	,475**	1
	Знач. (2-сторон)	,867	,886	,000	,000	
	N	100	100	100	100	100

** Корреляция значима на уровне 0.01 (2-сторон).

Рис. 9.4. Фрагмент окна вывода после выполнения шага 5

Корреляции		отметка2
счет в уме	Корреляция Пирсона	,080
	Знач. (2-сторон)	,430
	N	100
числовые ряды	Корреляция Пирсона	,114
	Знач. (2-сторон)	,258
	N	100
словарь	Корреляция Пирсона	,040
	Знач. (2-сторон)	,692
	N	100
осведомленность	Корреляция Пирсона	,257**
	Знач. (2-сторон)	,010
	N	100
кратковременная память	Корреляция Пирсона	,294**
	Знач. (2-сторон)	,003
	N	100

** Корреляция значима на уровне 0.01 (2-сторон).

Рис. 9.5. Фрагмент окна вывода после выполнения шага 5а

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» в главе 2).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.srv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

10 Средние значения

134	Пошаговый алгоритм вычислений
138	Представление результатов
139	Завершение анализа и выход из программы

С помощью команды Таблицы сопряженности, описанной в главе 8, вычисляются частоты по значениям неколичественных (номинальных) переменных. Таблицы сопряженности позволяют сравнивать частоты для разных подгрупп, которые соответствуют значениям номинативной переменной. Например, из таблицы сопряженности пол × хобби, видно, что 15 девушек увлекаются спортом, а 27 — искусством, и т. п. Команда Средние предназначена для сравнения подгрупп наблюдений по таким показателям количественных переменных, как средние, медианы и пр. При этом предполагается, что среди данных имеются не только количественные переменные, для которых вычисляются средние, но и номинальные переменные, разделяющие объекты на подгруппы. Команда Средние может быть рассмотрена на примере данных файла ex01.sav. Так, при помощи этой команды можно сравнить средние значения успеваемости (отметка1, отметка2) юношей и девушек (пол), учащихся разных классов (класс) и т. д. Результаты вычислений представляются в виде таблиц, похожих на таблицы сопряженности, полученные с использованием команды Таблицы сопряженности. Различие заключается в том, что для каждой подгруппы вычисляются не только частота, но и другие статистики, например средние.

Команда Средние является одной из самых простых и часто используемых в SPSS. Для выбранных подгрупп она по умолчанию подсчитывает средние значения, стандартные отклонения и частоты. Кроме того, с помощью кнопки Параметры можно задать вывод других описательных статистик для групп наблюдений, а также результатов однофакторного дисперсионного анализа. В этой главе дисперсионный анализ упоминается без детального описания, поскольку этой теме целиком посвящены главы 13–16.

Пошаговый алгоритм вычислений

В пошаговых процедурах используется файл ex01.sav с входящими в него переменными класс, пол и отметка2. Сначала выполняются три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex01.sav.

ШАГ 4

В меню Анализ выберите команду Сравнение средних ► Средние. На экране появится диалоговое окно Средние (рис. 10.1).

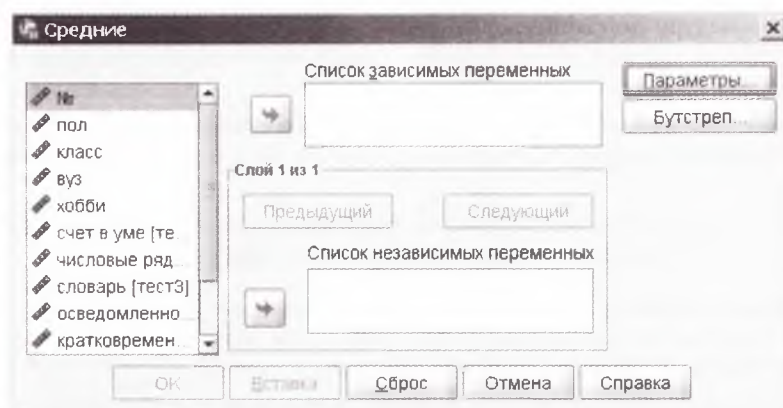


Рис. 10.1. Диалоговое окно Средние

В диалоговом окне Средние вам необходимо задать переменные, которые будут участвовать в процедуре. Список зависимых переменных в верхней части окна предназначен для количественных переменных. В качестве зависимых переменных могут выступать переменные отметка1, отметка2, тест1 и т. д.

Список независимых переменных служит для задания неколичественных (номинальных) переменных, градации которых определяют сравниваемые подгруппы объектов (пол, класс, вуз и т. п.). На основе этих переменных строится таблица сопряженности. Если включить в верхний список только переменную отметка2, а в нижний — только переменную пол, в результате выполнения команды Средние (с установками по умолчанию) мы увидим средние значения, стандартные отклонения и N отметок для юношей и девушек.

Кнопка Параметры позволяет задавать другие статистики для сравнения подвыборок наблюдений: медиану, асимметрию, эксцесс и пр.

Как правило, исследователи не ограничиваются единственной независимой переменной. Типичным является построение таблицы сопряженности с участием двух переменных. Например, таблица пол $\times$ класс позволяет получить средние значения отметок для юношей и девушек каждого из трех классов. Чтобы задать несколько переменных в Списке независимых переменных, используйте кнопки Предыдущий и Следующий, расположенные в области Слой 1 из 1. Как вы увидите в приведенных далее пошаговых процедурах, все действия в окне Средние просты и понятны.

ШАГ 5

После выполнения шага 4 должно быть открыто диалоговое окно Средние (см. рис. 10.1). Вычислим средние значения отметок для учащихся каждого из трех классов:

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список зависимых переменных.
2. Щелкните сначала на переменной класс, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в Список независимых переменных.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Можно включить в процедуру дополнительную независимую переменную, например пол, и тем самым построить таблицу сопряженности класс $\times$ пол, содержащую средние значения и стандартные отклонения.

Количество зависимых переменных может быть любым. Программа просто строит в таблице столько столбцов, сколько переменных указано в Списке зависимых переменных.

В следующем примере вычислим средние значения и стандартные отклонения для переменных `отметка1` и `отметка2` с использованием таблицы сопряженности класс $\times$ пол.

ШАГ 5a

После выполнения шага 4 должно быть открыто диалоговое окно Средние (см. рис. 10.1). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной `отметка1`, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список зависимых переменных.
2. Повторите предыдущее действие для переменной `отметка2`.
3. Щелкните сначала на переменной `класс`, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в Список независимых переменных.
4. Щелчком на кнопке Следующий освободите Список независимых переменных — метка Слой 1 из 1 изменится на Слой 2 из 2. После этого щелкните сначала на переменной `пол`, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в Список независимых переменных.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода результатов.

Для того чтобы выполнить однофакторный дисперсионный анализ или включить в ячейки дополнительную информацию, можно использовать кнопку Параметры в диалоговом окне Средние. После щелчка на этой кнопке откроется диалоговое окно Средние: Параметры (рис. 10.2).

С помощью диалогового окна Средние: Параметры можно задать дополнительные параметры вывода для команды Средние. Например, помимо величин (среднего значения, стандартного отклонения и числа наблюдений), вычисляемых по умолчанию, можно указать любую совокупность показателей, перечисленных в списке Статистики. Для этого следует выделить нужный пункт списка, а затем щелчком на кнопке со стрелкой добавить его в поле Статистики в ячейках.

Как уже упоминалось, команда Средние позволяет выполнять однофакторный дисперсионный анализ (Analysis Of Variances, ANOVA). Для этого в области Статистики для первого слоя нужно установить флажок Таблица дисперсионного анализа и эта. В процессе группировки зависимой переменной `отметка2` по значениям независимой переменной `класс` программа путем однофакторного дисперсионного анализа сравнит три средних значения для переменной `класс`.

В следующем примере мы вычислим средние значения и медианы для каждой из шести подгрупп, образующихся при сопряжении переменных `класс` и `пол`. Кроме того, выполним дисперсионный анализ с зависимой переменной `отметка2` и независимой переменной `класс`.

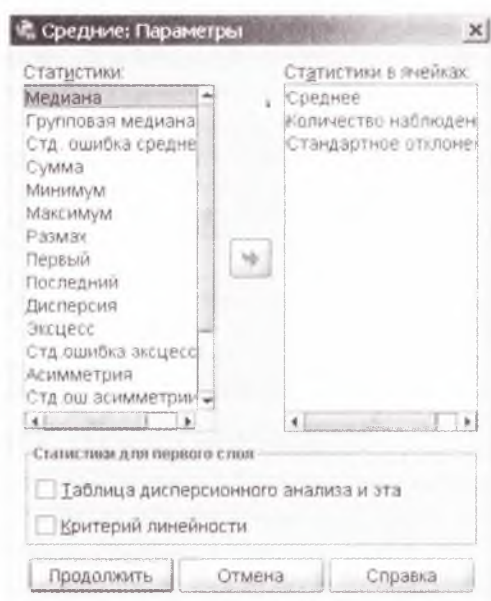


Рис. 10.2. Диалоговое окно Средние: Параметры

ШАГ 56

После выполнения шага 4 должно быть открыто диалоговое окно Средние (см. рис. 10.1). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список зависимых переменных.
2. Щелкните сначала на переменной класс, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в Список независимых переменных.
3. Щелчком на кнопке Следующий освободите Список независимых переменных — метка Слои 1 из 1 изменится на Слои 2 из 2. После этого щелкните сначала на переменной пол, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в Список независимых переменных.
4. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Средние: Параметры, и установите флажок Таблица дисперсионного анализа и эта.
5. Выделите в окне слева статистику Медиана и перенесите ее в окно Статистики в ячейках, нажав кнопку со стрелкой между этими окнами.
6. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Средние.
7. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5, 5а, или 5б программа автоматически активизирует окно вывода результатов. Для просмотра результатов вы при необходимости можете воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на

стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

На рис. 10.3 приведены фрагменты выводимых данных, сгенерированных программой после выполнения шага 5б.

В первой таблице приводятся средние значения, частоты (N) и стандартные отклонения и медианы зависимой переменной *отметка2* для каждой из подгрупп. Медианы и средние значения мало отличаются друг от друга, так как распределение переменной близко к нормальному. Во второй и третьей таблицах содержатся результаты дисперсионного анализа. Обратите внимание на то, что дисперсионный анализ выполнен только для средних значений первого слоя, который определяется градациями переменной *класс*. Для сравнения средних, соответствующих градациям переменной *пол*, следовало бы указать ее первой в Списке независимых переменных. Более подробное описание однофакторного дисперсионного анализа приводится в главе 13.

Вычисленные средние значения (4,096, 4,167 и 4,408) различаются на уровне значимости $p < 0,001$. Это свидетельствует о статистически достоверной зависимости успеваемости учащихся (*отметка2*) от класса, в котором они учатся. Коэффициент Эта подобен корреляции и оценивает связь между двумя переменными: количественной и номинальной.

Отчет

отметка2

класс	пол	Среднее	N	Стд Отклонение	Медиана
А	ЖЕН	4,1286	14	,21901	4,1250
	МУЖ	4,0711	19	,26579	4,0500
	Итого	4,0955	33	,24506	4,1000
Б	ЖЕН	4,2048	21	,20670	4,2500
	МУЖ	4,1107	14	,22717	4,1750
	Итого	4,1671	35	,21691	4,2000
В	ЖЕН	4,4173	26	,27711	4,5000
	МУЖ	4,3667	6	,26204	4,3500
	Итого	4,4078	32	,27093	4,4500
Итого	ЖЕН	4,2779	61	,26856	4,2500
	МУЖ	4,1308	39	,26622	4,2000
	Итого	4,2205	100	,27589	4,2000

Таблица ANOVA

		Сумма квадратов	ст.св	Средний квадрат	F	Знач
отметка2 * класс	Между группами	1,738	2	,869	14,544	,000
	В группах	5,797	97	,060		
	Итого	7,535	99			

Меры связи

	Эта	Эта квадрат
отметка2 * класс	,480	,231

Рис. 10.3. Фрагменты окна вывода после выполнения шага 5б

Коэффициент Эта квадрат — мера влияния независимой переменной на дисперсию зависимой переменной. Величина 0,231 свидетельствует о том, что 23,1 % дисперсии зависимой переменной объясняются влиянием независимой переменной.

Отметим, что команда Средние не позволяет определить статистическую значимость при попарном соотнесении средних значений. Так, в данном случае по результатам статистической проверки можно сделать вывод только о том, что классы различаются по успеваемости, без конкретизации, как эти классы отличаются друг от друга. Попарно сравнить средние значения можно при помощи *t*-критерия Стьюдента (см. главу 11), а лучше — при помощи множественных сравнений в рамках однофакторного дисперсионного анализа (см. главу 13).

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» в главе 2).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

11 Сравнение двух средних и t -критерий

141	Уровень значимости
141	Пошаговые алгоритмы вычислений
146	Представление результатов
148	Завершение анализа и выход из программы

Различные варианты обработки данных с применением t -критерия позволяют сделать вывод о различии двух средних значений. Например, в случае применения t -критерия для независимых выборок проверяется достоверность различия двух выборок по количественной переменной, измеренной у представителей этих двух выборок. Для этих выборок вычисляются средние значения количественной переменной, затем по t -критерию определяется статистическая значимость различия средних. Применение t -критерия, по-видимому, самый распространенный метод статистического вывода, так как позволяет ответить на простой вопрос: «Насколько существенны различия между двумя выборками по данной количественной переменной?» Основное требование к данным для применения этого критерия — представление переменных, по которым сравниваются выборки, в метрических единицах измерения. Если есть сомнения в соответствии этому требованию, нужно воспользоваться непараметрическими методами, изложенными в главе 12.

SPSS позволяет применять 3 варианта t -критерия: *t -критерий для независимых выборок*, *t -критерий для парных выборок*, *одновыборочный t -критерий*:

- ▶ t -критерий для независимых выборок предназначен для сравнения средних значений двух выборок. Для сравниваемых выборок должны быть определены значения одной и той же переменной. С помощью t -критерия для независимых выборок можно сравнить успеваемость студентов и студенток, степень удовлетворенности жизнью холостяков и женатых, средний рост футболистов двух команд и пр. Обязательным условием для проведения этого t -критерия является независимость выборок.
- ▶ t -критерий для парных, или зависимых, выборок позволяет сравнить средние значения двух измерений одного признака для одной и той же выборки, например результаты первого и последнего экзаменов группы студентов или значения показателя до и после воздействия на группу. Обязательным условием применения t -критерия для зависимых выборок является наличие повторного измерения для одной выборки.
- ▶ одновыборочный t -критерий позволяет сравнить среднее значение этой выборки с некоторой эталонной величиной. Например, отличается ли среднее значение некоторого теста для данной выборки от нормативной величины, отличается ли время, показанное бегунами во время соревнования, от 17 минут и т. д.

Уровень значимости

Результат сравнения средних значений с применением t -критерия оценивается по уровню значимости. Напомним, что *уровень значимости* (p -уровень) является мерой статистической достоверности результата вычислений, в данном случае — различий средних, и служит основанием для интерпретации. Если исследование показало, что p -уровень значимости различий не превышает 0,05, это означает, что с вероятностью не более 5 % различия являются случайными. Обычно это становится основанием для вывода о статистической достоверности различий. В противном случае ($p > 0,05$) различие признается статистически недостоверным и не подлежит содержательной интерпретации.

SPSS позволяет определять два уровня значимости: односторонний и двусторонний. Обычно используется двусторонний уровень значимости. Но если вы заранее знаете направление различий и вас интересует только одно направление, то можно использовать односторонний критерий значимости. Однако такая ситуация встречается редко, а если и встречается, то правомерность односторонней проверки с трудом поддается обоснованию.

Существенным различием между двумя типами уровней значимости является то, что значение двустороннего p -уровня оказывается вдвое большим, чем одностороннего. По умолчанию SPSS вычисляет двусторонний уровень значимости, но вы легко можете получить односторонний уровень значимости, разделив вычисленное программой значение на 2.

Пошаговые алгоритмы вычислений

Для применения t -критерия мы воспользуемся переменными из файла ex01.sav, но сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex01.sav.

Применение t -критерия для независимых выборок

ШАГ 4

В меню Анализ выберите команду Сравнение средних ► Т-критерий для независимых выборок. На экране появится диалоговое окно Т-критерий для независимых выборок (рис. 11.1).

В этом диалоговом окне вы можете задать параметры для применения t -критерия. Левая часть окна содержит список всех доступных переменных файла, а правая часть — поле Проверяемые переменные, в которое необходимо поместить имена переменных для применения t -критерия. Список может содержать любое количество переменных, а сами переменные должны быть метрического типа (переменные отметка1, отметка2, тест1 и т. п.). В поле Группировать по указывается имя переменной, значениям (градациям) которой соответствуют две независимые выборки для t -критерия. Как правило, группирующая переменная дискретна и имеет две градации. В файле ex01.sav примером двухуровневой переменной является переменная пол. Группирующая переменная может иметь более двух уровней и даже быть непрерывного типа, но в обоих случаях необходи-

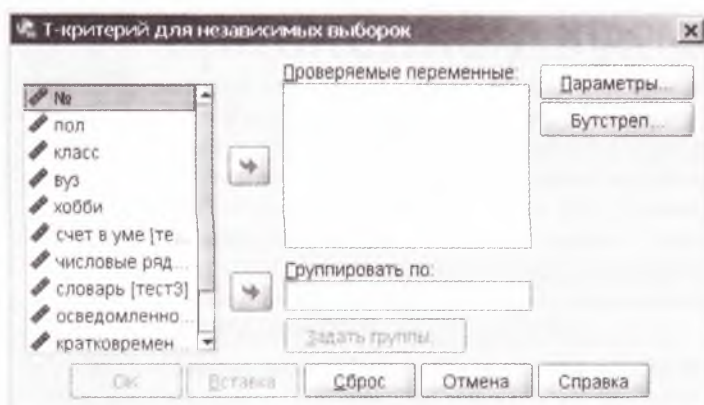


Рис. 11.1. Диалоговое окно Т-критерий для независимых выборок

мо разбить множество значений переменной на две категории или задать два значения из существующих. Так, мы можем взять переменную класс в качестве группирующей переменной и указать, что в одну выборку следует включать все наблюдения, имеющие значение 1, а в другую — наблюдения, имеющие значение 3. Это позволит сравнить средние значения для двух из трех классов.

Итак, после того как имя группирующей переменной задано, необходимо задать ее категории. Это делается щелчком на кнопке Задать группы. Обратите внимание на то, что данный шаг необходим даже для двухуровневых переменных. На экране появится диалоговое окно Задать группы (рис. 11.2). Если переключатель установлен в положение Пороговое значение, в соответствующем окне следует задать значение переменной, разбивающее весь ее диапазон на два непересекающихся интервала.

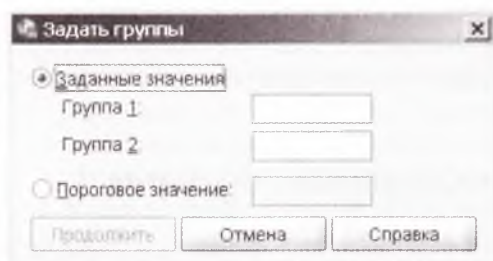


Рис. 11.2. Диалоговое окно Задать группы

ШАГ 5

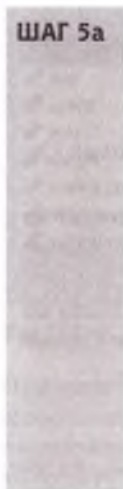
На этом шаге мы применим t -критерий для сравнения юношей и девушек (значений переменной пол) по переменной отметка2. После выполнения шага 4 на экране должно быть открыто окно Т-критерий для независимых выборок (см. рис. 11.1).

1. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Проверяемые переменные.
2. Щелкните сначала на переменной пол, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Группировать по.



3. Щелкните на кнопке Задать группы, чтобы открыть одноименное диалоговое окно, показанное на рис. 11.2.
4. Введите число 1 в поле Группа 1, нажмите клавишу Tab, введите число 2 в поле Группа 2 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Т-критерий для независимых выборок.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В следующем примере в качестве группирующей будет использована переменная отметка2. Это количественная переменная — средняя отметка за 10-й класс. С ее помощью разобьем выборку на 2 части по средней отметке: до 4 и от 4 и выше. В качестве проверяемой переменной используем кратковременную память (тест5).

**ШАГ 5a**

После выполнения шага 4 должно быть открыто диалоговое окно Т-критерий для независимых выборок (см. рис. 11.1). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной кратковременная память [тест5], чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Проверяемые переменные.
2. Щелкните сначала на переменной отметка2, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Группировать по.
3. Щелкните на кнопке Задать группы, чтобы открыть одноименное диалоговое окно (см. рис. 11.2).
4. Установите переключатель в положение Пороговое значение, введите в расположенное рядом поле значение 4 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Т-критерий для независимых выборок.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода результатов.

Применение t -критерия для парных выборок

Теперь рассмотрим два примера применения t -критерия для парных (зависимых) выборок. После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню.

**ШАГ 4a**

В меню Анализ выберите команду Сравнение средних ► Т-критерий для парных выборок. На экране появится диалоговое окно Т-критерий для парных выборок (рис. 11.3).

Применение t -критерия для зависимых выборок несколько проще, чем для независимых. Процедура управляется единственным диалоговым окном и не требует указания группирующей переменной. В левой части окна находится уже знакомый нам список доступных переменных файла, в котором выбираются переменные для применения t -критерия. Необычным здесь является то, что нужно указывать пары переменных — отсюда название целевого списка: Парные переменные.

Сначала нужно щелкнуть на имени первой переменной, затем на имени второй и после этого при помощи кнопки со стрелкой занести их в список. Количество выбираемых пар может быть любым.

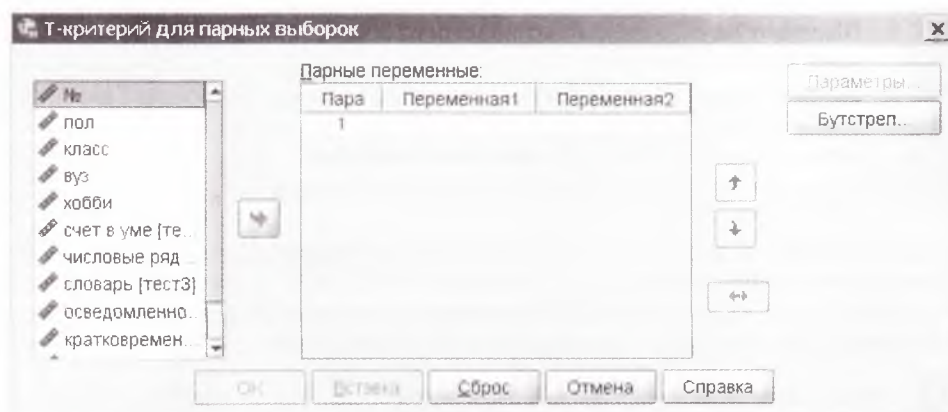


Рис. 11.3. Диалоговое окно Т-критерий для парных выборок

ШАГ 56

Чтобы сравнить отметки учащихся в 10-х и 11-х классах (отметка1 и отметка2), выполните следующие действия:

1. Щелкните на переменной отметка1, чтобы выделить ее, затем щелчком на стрелке перенесите ее в правое окно. Ее имя появится в области Парные переменные в строке Пара: 1 рядом с меткой Переменная 1.
2. Щелкните на переменной отметка2, чтобы выделить ее, затем щелчком на стрелке перенесите ее в правое окно. Ее имя появится в области Парные переменные в строке Пара: 1 рядом с меткой Переменная 2.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Чтобы применить t -критерий одновременно для нескольких пар переменных, их все нужно переместить в поле Парные переменные. В следующем примере результаты первого теста (тест1) сравниваются с результатами остальных четырех тестов.

ШАГ 5в

После выполнения шага 4а должно быть открыто диалоговое Т-критерий для парных выборок (см. рис. 11.3). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Нажмите на клавиатуре клавишу Ctrl и, удерживая ее, щелкните сначала на переменной счет в уме [тест1], затем — на переменной числовые ряды [тест2], чтобы они были выделены в левом списке переменных.
2. Щелкните на кнопке со стрелкой, чтобы переместить выбранную пару переменных в поле Парные переменные.
3. Повторите два предыдущих действия, только вместо переменной тест2 во втором действии последовательно используйте переменные тест3, тест4 и тест5.
4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Применение t -критерия для одной выборки

Иногда бывает необходимо сравнить среднее значение распределения с какой-либо фиксированной величиной. Представим такую ситуацию. Исследователь решил прове-

ритель, отличаются ли данные его выборки от нормативных показателей. Предположим, нормативный показатель по выбранной переменной равен 10. Для того чтобы проверить результат выборки на соответствие норме, нужно вычислить среднее значение для выборки и сравнить его с числом 10. Подобные сравнения в программе SPSS реализуются при помощи одновыборочного t -критерия — для одной выборки.

ШАГ 4б В меню Анализ выберите команду Сравнение средних ► Одновыборочный Т-критерий. На экране появится диалоговое окно Одновыборочный Т-критерий (рис. 11.4).

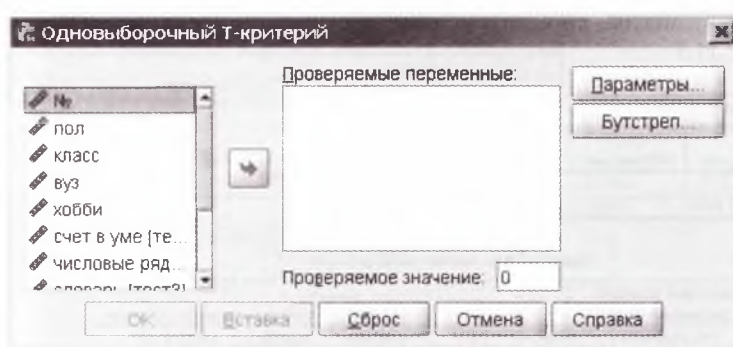


Рис. 11.4. Диалоговое окно Одновыборочный Т-критерий

Применение t -критерия для одной выборки является самым простым с точки зрения как вычислений, так и управления. Все, что нужно сделать, — выбрать одну или несколько переменных для анализа и, поместив ее (их) в поле Проверяемые переменные, задать константу для сравнения в поле Проверяемое значение, после чего щелкнуть на кнопке ОК. Если в поле Проверяемые переменные будут помещены несколько переменных, то их средние значения будут сравниваться с одной и той же величиной, указанной в поле Проверяемое значение. Если нужно выполнить сравнение с различными константами, то придется применить t -критерий для каждой из констант.

ШАГ 5г На этом шаге средние значения переменных числовые ряды [тест2] и словарь [тест3] сравниваются с числом 10 как нормативным для теста:

1. Щелкните сначала на переменной тест2, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в поле Проверяемые переменные.
2. Повторите предыдущее действие для переменной тест3.
3. Нажмите клавишу Tab, чтобы перевести фокус ввода в поле Проверяемое значение, и введите число 10.
4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5, 5а, 5б, 5в или 5г программа автоматически перейдет в окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

В этом разделе мы приводим результаты применения всех трех видов *t*-критерия. После каждого из результатов следуют краткие пояснения, раскрывающие их смысл, а в конце раздела помещен список терминов использовавшихся в окне вывода. Визуальная форма вывода немного изменена, что никак не влияет на его содержимое.

Результаты применения *t*-критерия для независимых выборок

Фрагменты выводимых результатов, сгенерированных программой после выполнения шага 5, приведены на рис. 11.5.

Групповые статистики				
поп	N	Среднее	Стд отклонение	Стд. ошибка среднего
отметка2 ЖЕН	61	4,2779	,26856	,03439
МУЖ	39	4,1308	,26622	,04263

Критерий для независимых выборок								
		Критерий равенства дисперсий Левина		t-критерий равенства средних				
		F	Знач.	t	ст.св.	Значимость (2-сторонняя)	Разность средних	Стд. ошибка разности
отметка2	Предполагается равенство дисперсий	,060	,807	2,681	98	,009	,14710	,05488
	Равенство дисперсий не предполагается			2,686	81,647	,009	,14710	,05477

Рис. 11.5. Фрагменты окна вывода после выполнения шага 5

Из результатов следует, что выборка из 39 юношей имеет средний балл 4,13, выборка из 61 девушки — средний балл 4,28. Различия статистически достоверны на высоком уровне значимости ($p = 0,009$). Критерий равенства дисперсий Левина указывает на то, что дисперсии двух распределений статистически значимо не различаются ($p = 0,807$), следовательно, применение *t*-критерия корректно.

Таблица результатов может быть велика и не помещаться на странице отчета. При желании можно удалить лишние столбцы или строки при помощи редактора таблиц. Для входа в него достаточно дважды щелкнуть на таблице левой кнопкой мыши. Более подробно редактирование таблиц рассмотрено в разделе «Окно вывода и его редактирование» главы 2.

Результаты применения *t*-критерия для зависимых выборок

Фрагменты выводимых результатов, сгенерированных программой после выполнения шага 5б, приведены на рис. 11.6.

Как видно из результатов, для выборки объемом $N = 100$ среднее значение переменной отметка2 (4,22) оказалось статистически значимо выше среднего значения переменной

отметка1 (3,96) с уровнем значимости $p < 0,001$. Кроме того, между переменными отметка1 и отметка2 существует значительная корреляция ($r = 0,434$, $p < 0,001$), свидетельствующая о том, что данные переменные действительно можно считать зависимыми выборками.

Статистики парных выборок

		Среднее	N	Стд. отклонение	Стд. ошибка среднего
Пара 1	отметка1	3,9630	100	,30597	,03060
	отметка2	4,2205	100	,27589	,02759

Корреляции парных выборок

		N	Корреляция	Знач.
Пара 1	отметка1 & отметка2	100	,434	,000

Критерий парных выборок

	Парные разности			t	ст. св.	Значимость (2-сторонняя)
	Среднее	95% доверительный интервал разности средних				
		Нижняя граница	Верхняя граница			
отметка1 - отметка2	-,25750	-,31913	-,19587	-8,290	99	,000

Рис. 11.6. Фрагменты окна вывода после выполнения шага 5б

Результаты применения *t*-критерия для одной выборки

Фрагменты выводимых результатов, сгенерированных программой после выполнения шага 5г, приведены на рис. 11.7.

Статистики для одновыборочного <i>t</i> -критерия				
	N	Среднее	Стд. отклонение	Стд. ошибка среднего
числовые ряды	100	10,35	2,768	,277
словарь	100	11,96	2,857	,286

Одновыборочный <i>t</i> -критерий						
	Проверяемое значение = 10					
	t	ст. св.	Значимость (2-сторонняя)	Разность средних	95% доверительный интервал разности средних	
					Нижняя граница	Верхняя граница
числовые ряды	1,264	99	,209	,350	-,20	,90
словарь	6,861	99	,000	1,960	1,39	2,53

Рис. 11.7. Фрагменты окна вывода после выполнения шага 5г

Из таблиц видно, что среднее значение переменной тест2 (числовые ряды) составляет 10,35 и статистически достоверно не отличается от 10 ($p > 0,1$). Среднее значение переменной тест3 (словарь) равно 11,96 и статистически достоверно отличается от 10 ($p < 0,001$).

Трактовка терминов

Трактовка терминов, используемых программой в окне вывода результатов:

- ▶ Стандартная ошибка — отношение стандартного отклонения к квадратному корню из размера выборки N . Является мерой стабильности среднего значения.
- ▶ F -критерий — величина, характеризующая соотношение дисперсий двух распределений.
- ▶ Значимость — значимость, или p -уровень значимости. При сравнении дисперсии двух распределений в зависимости от того, равны они или не равны, применяются различные виды статистических приближений. Значение $p > 0,05$ указывает на то, что дисперсии можно считать неразличающимися.
- ▶ t (t -критерий) — t -критерий определяется как отношение разности средних значений к стандартному отклонению.
- ▶ ст. св. — число степеней свободы, для t -критерия с независимыми выборками при равенстве дисперсий число степеней свободы равно разности числа объектов и числа групп ($100 - 2 = 98$), а при различии дисперсий применяется более сложная формула, дающая дробное значение 81,65. Для зависимых выборок и для одной выборки число степеней свободы для t -критерия определяется как $100 - 1 = 99$.
- ▶ Значимость (2-сторонняя) — по отношению к t -критерию двусторонняя значимость означает вероятность того, что разность между средними значениями является случайной, а по отношению к коэффициенту корреляции — вероятность того, что связь между двумя переменными является случайной.
- ▶ Стд. отклонение — стандартное отклонение. Для t -критерия с зависимыми выборками это стандартное отклонение разности между значениями повторных измерений.
- ▶ Корреляция — мера связи двух переменных, а для зависимых выборок — мера связи парных переменных. Численно определяется коэффициентом корреляции; в данном примере использовался коэффициент Пирсона. Более подробно корреляции описаны в главе 9.
- ▶ 95% доверительный интервал — в случае t -критерия термин «доверительный интервал» относится к разности между средними значениями выборок.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» в главе 2).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

12 Непараметрические критерии

149	Параметрические и непараметрические критерии
150	Пошаговые алгоритмы и результаты вычислений
166	Завершение анализа и выход из программы

Параметрические и непараметрические критерии

Перед тем как ввести термин «непараметрический критерий», необходимо уточнить, что это такое. *Параметрический критерий* — это метод статистического вывода, который применяется в отношении *параметров* генеральной совокупности. Главным условием для параметрических методов является нормальность распределения переменных и, как следствие, правомерность применения таких статистик, как среднее значение и стандартное отклонение. Несмотря на то что некоторые параметрические методы позволяют анализировать данные, распределенные по другим законам (например, биномиальному или Пуассона), непараметрические методы в этом смысле гораздо функциональнее, поскольку вообще не связывают анализ с каким-либо законом распределения.

Таким образом, непараметрические методы позволяют исследовать данные без каких-либо допущений о характере распределения переменных, в том числе при нарушении требования нормальности распределения. Так как эти методы предназначены для номинативных и ранговых переменных, в отношении которых недопустимо применение арифметических операций, они основаны на различных дополнительных вычислениях, среди которых можно отметить:

- ▶ ранжирование переменных;
- ▶ подсчет числа значений одного распределения, которые превышают значения другого распределения;
- ▶ применение весовых сравнений;
- ▶ определение степени отклонения распределения от случайного или биномиального распределения;
- ▶ проверка нормальности выборочного распределения;
- ▶ сравнение частот;
- ▶ сравнение групп путем вычисления частот значений, лежащих выше или ниже главной медианы.

Помимо прочего непараметрические критерии позволяют вычислять статистические показатели для одной выборки и сравнивать две выборки между собой. Несмотря на кажущуюся сложность, непараметрические методы в большинстве своем очень просты для понимания и применения.

Для использования непараметрических методов задействуем уже знакомый нам файл ex01.sav и некоторые специально подготовленные примеры. Описание переменных файла ex01.sav можно найти в главе 3 (количество объектов $N = 100$). Ссылки на имена файлов с дополнительными примерами будем давать по мере изложения материала.

По структуре эта глава несколько отличается от других. Из-за обилия методов пошаговые алгоритмы в этой главе объединены с описанием и интерпретацией результатов. После начальных трех шагов представлены два характерных для большинства непараметрических методов диалоговых окна. Далее даны описания восьми наиболее часто используемых непараметрических методов. Помимо общего описания каждого метода показаны пример его применения в виде шагов 4 и 5, 4а и 5а, 4б и 5б и т. д., программный вывод результатов, его интерпретация и терминология. Указанные восемь методов перечислены далее:

1. Сравнение двух независимых выборок (*критерий Манна — Уитни*) позволяет установить различия между двумя независимыми выборками по уровню выраженности порядковой переменной.
2. Сравнение двух связанных (зависимых) выборок может проводиться по двум критериям. *Критерий знаков* основан на подсчете числа отрицательных и положительных разностей между повторными измерениями; *критерий Уилкоксона* в дополнение к знакам разностей учитывает их величину.
3. *Критерий серий* определяет, является ли последовательность бинарных величин (событий) случайной или упорядоченной.
4. *Биномиальный критерий* определяет, отличается ли распределение дихотомической величины от заданного соотношения.
5. *Критерий Колмогорова — Смирнова* для одной выборки определяет отличие распределения переменной от нормального (равномерного, Пуассона и т. д.).
6. *Критерий хи-квадрат* для одной выборки определяет степень отличия наблюдаемого распределения частот по градациям переменной от ожидаемого распределения.
7. Сравнение K независимых выборок (*критерий H Краскела — Уоллеса*) позволяет установить степень различия между тремя и более независимыми выборками по уровню выраженности порядковой переменной.
8. Сравнение K связанных (зависимых) выборок (*критерий Фридмана*) позволяет установить степень различия между тремя и более зависимыми выборками по уровню выраженности порядковой переменной.

Пошаговые алгоритмы и результаты вычислений

Для выбора непараметрических методов будем использовать *традиционные диалоговые меню* (Legacy Dialogs — в англоязычном интерфейсе, в русскоязычном интерфейсе они почему-то названы «Устаревшие диалоговые окна») в группе анализа Непараметрические.

Перед применением непараметрических критериев необходимо выполнить 3 подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые

инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

Некоторые диалоговые окна непараметрических критериев

Применение непараметрических методов иногда требует определять сравниваемые группы по градациям группирующей переменной. Для этого в соответствующих диалоговых окнах предусмотрена кнопка Задать группы, при щелчке на которой открывается характерное диалоговое окно, представленное на рис. 12.1. Если группирующая переменная имеет уровни 1 и 2, то в поле Группа 1 следует ввести значение 1, в поле Группа 2 — значение 2, а затем щелкнуть на кнопке Продолжить.

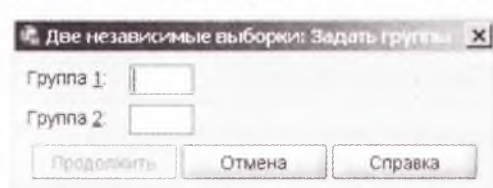


Рис. 12.1. Диалоговое окно определения групп

Еще одно новое для нас диалоговое окно появляется при щелчке на кнопке Параметры в основном диалоговом окне команды применения непараметрических методов (рис. 12.2). В области Статистики имеются флажки Описательные и Квартили, позволяющие включить в выводимые данные описательные статистики (среднее значение, стандартное отклонение, минимум, максимум и объем выборки) и квартили.

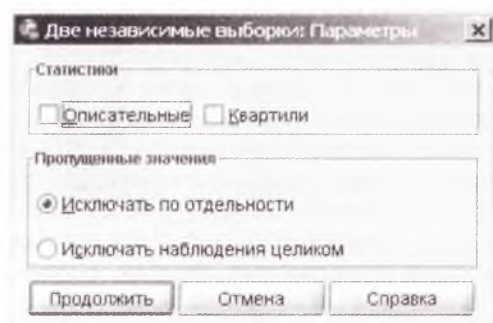


Рис. 12.2. Диалоговое окно настройки параметров непараметрических тестов

Процедуры непараметрических методов имеют дополнительную возможность расчета *точного значения* p -уровня значимости, если диалоговое окно метода содержит кнопку Точные. Если в диалоговом окне метода щелкнуть на кнопке Точные, появляется окно Точные критерии, позволяющее выбрать один из точных методов расчета значимости (рис. 12.3). По умолчанию переключатель в этом окне установлен в положении Только асимптотически, что соответствует определению приближительной значимости традиционным способом. Если данные не соответствуют требованиям применяемого метода, при помощи переключателя можно воспользоваться одним из двух точных методов расчета

значимости: методом статистических испытаний Монте-Карло или собственно точным критерием. Метод Точные по определению дает более точные результаты. Однако в некоторых случаях его расчет слишком трудоемок даже для компьютера, тогда используется метод Монте-Карло.

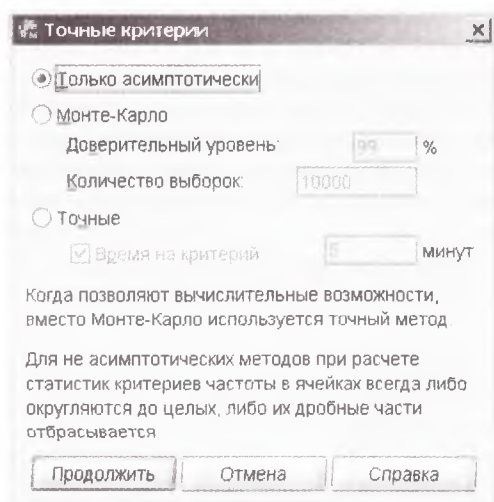


Рис. 12.3. Диалоговое окно Точные критерии

Сравнение двух независимых выборок

Критерий Манна — Уитни (Mann — Whitney), или *U-критерий*, по назначению аналогичен *t*-критерию для независимых выборок. Разница заключается в том, что *t*-критерии ориентированы на нормальные и близкие к ним распределения, а критерий Манна — Уитни — на распределения, отличные от нормальных. В частном случае критерий Манна — Уитни можно применять и для нормально распределенных данных, однако он менее чувствителен к различиям (является менее мощным), чем *t*-критерий. Далее в примере мы попытаемся выяснить, различаются ли юноши и девушки по успеваемости в выпускном классе. Напомним, что пол учащихся определяется значением переменной *пол*, а успеваемость — переменной *отметка2*. При реализации метода программа сначала ранжирует все объекты без учета принадлежности к сравниваемым группам, а затем вычисляет средние ранги для каждой из двух групп. Чем выше средний ранг группы, тем выше ее успеваемость. После нахождения средних рангов определяется *p*-уровень. Корректность применения этого критерия сомнительна, если переменная имеет небольшое число возможных значений, например 3-балльная шкала. В этом случае следует воспользоваться точными критериями. Диалоговое окно *U*-критерия представлено на рис. 12.4.

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы сравним успеваемость юношей и девушек в выпускном классе.

ШАГ 4

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Для двух независимых выборок, чтобы открыть диалоговое окно Критерии для двух независимых выборок (см. рис. 12.4).

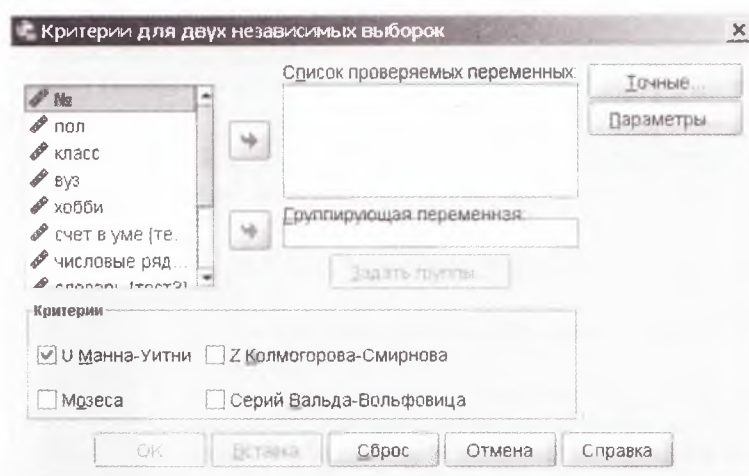


Рис. 12.4. Диалоговое окно Критерии для двух независимых выборок

ШАГ 5

Для применения метода выполните следующую последовательность действий:

1. Щелкните сначала на переменной *отметка2*, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список проверяемых переменных.
2. Щелкните сначала на переменной *пол*, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Группирующая переменная.
3. Щелкните на кнопке *Задать группы*, чтобы открыть диалоговое окно, показанное на рис. 12.1.
4. В поле *Группа 1* введите значение 1, нажмите клавишу Tab, чтобы переместить фокус ввода в поле *Группа 2*, введите значение 2 и щелкните на кнопке *Продолжить*, чтобы вернуться в диалоговое окно Критерий для двух независимых выборок.
5. Щелкните на кнопке *ОК*, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.5.

Средний ранг для девушек равен 56,21, для юношей — 41,56. Это значит, что у девушек успеваемость выше, чем у юношей. Статистика U Манна — Уитни равна 841. Значение Z является нормализованным, связанным с уровнем значимости $p = 0,014$. Поскольку величина уровня значимости (Асимпт. знч (двухсторонняя)) меньше 0,05, мы можем быть уверены в статистической достоверности вывода о том, что успеваемость девушек действительно выше успеваемости юношей.

Сравнение двух связанных (зависимых) выборок

Основные методы, которые используются для сравнения двух зависимых выборок, — это *критерий знаков* и *критерий Уилкоксона* (Wilcoxon test). Диалоговое окно Критерии для двух связанных выборок представлено на рис. 12.6.

Ранги				
	пол	N	Средний ранг	Сумма рангов
отметка2	ЖЕН	61	56,21	3429,00
	МУЖ	39	41,56	1621,00
	Всего	100		

Статистика критерия ^a		отметка2
Статистика U	Манна-Уитни	841,000
Статистика W	Уилкоксона	1621,000
Z		-2,469
Асимпт. знач. (двухсторонняя)		,014

a. Группирующая переменная: пол

Рис. 12.5. Фрагменты окна вывода после выполнения шага 5

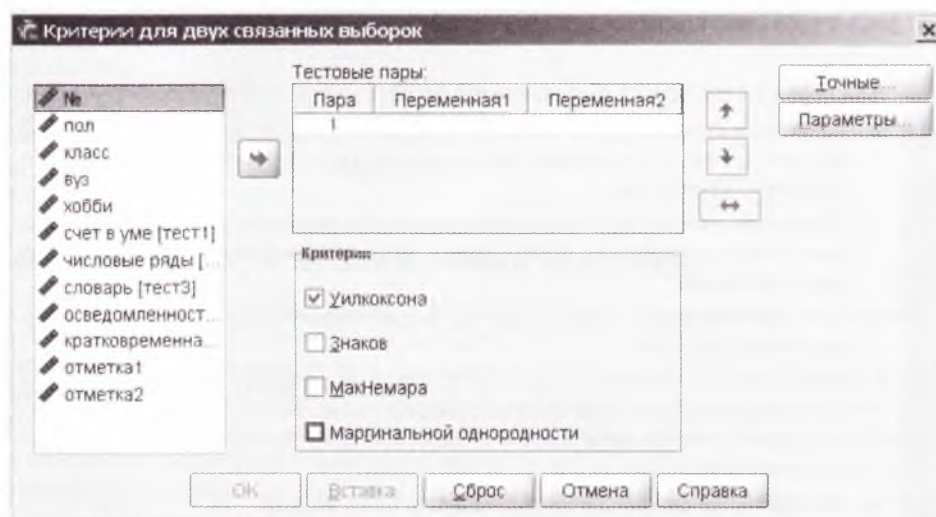


Рис. 12.6. Диалоговое окно Критерии для двух связанных выборок

Критерий знаков

Критерий знаков позволяет сравнить два измерения переменной на одной выборке (например, «до» и «после») по уровню ее выраженности путем сопоставления количества положительных и отрицательных разностей (сдвигов) значений. Для того чтобы продемонстрировать применение критерия, сравним результаты учащихся по второму (тест2) и четвертому (тест4) тестам. Для каждого объекта сначала определяется знак разности значений. Так, для первого объекта значение тест2 равно 7, а значение тест4 — 10. Сравнение этих значений даст отрицательный знак ($7 - 10 = -3 < 0$). Для четвертого объекта значения переменных тест2 и тест4 составляют, соответственно, 9 и 6, и знак разности будет положительным ($9 - 6 = 3 > 0$). Подсчитывается число положительных, отрицательных и нулевых разностей, а затем вычисляются нормализованное z-значение и p-уровень значимости.

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы сравним результаты учащихся по второму (тест2) и четвертому (тест4) тестам.

ШАГ 4а

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Для двух независимых выборок, чтобы открыть диалоговое окно Критерии для двух связанных выборок (см. рис. 12.6).

ШАГ 5а

Для применения критерия знаков выполните следующую последовательность действий:

1. В области Критерии установите флажок Знаков и снимите флажок Уилкоксона.
2. Щелкните на переменной числовые ряды [тест2], чтобы выделить ее, затем щелчком на стрелке перенесите переменную в правое окно. Ее имя появится в области Тестовые пары в строке Пара: 1 рядом с меткой Переменная 1.
3. Щелкните на переменной осведомленность [тест4], чтобы выделить ее, затем щелчком по стрелке перенесите переменную в правое окно. Ее имя появится в области Тестовые пары в строке Пара: 1 рядом с меткой Переменная 2.
4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.7.

Частоты		N
осведомленность - числовые ряды	Отрицательные разности	39
	Положительные разности	57
	Связи ^а	4
	Всего	100

а. осведомленность < числовые ряды

б. осведомленность > числовые ряды

с. осведомленность = числовые ряды

Статистики критерия ^а	
	осведомл енность - числовые ряды
Z	-1,735
Асимпт. знч. (двухсторонняя)	,083

а. Критерий знаков

Рис. 12.7. Фрагменты окна вывода после выполнения шага 5а

Полученные результаты говорят о том, что в 39 случаях значения переменной тест2 оказались меньшими, чем значения переменной тест4, в 57 случаях значения переменной тест2 превысили значения переменной тест4 и 4 раза было установлено равенство значений обеих переменных. Стандартизованное значение (Z) составляет $-1,735$, а уровень значимости $p = 0,083$. Это означает, что различия между результатами тестов тест4 и тест2 статистически недостоверны. Обратите внимание: поскольку переменные тест4

и тест2 являются метрическими, к ним предпочтительней применить t -критерий для парных выборок. Он показал бы, что средние значения тест4 и тест2 различаются с уровнем значимости $p = 0,01$. Таким образом, можно на практике убедиться в том, что статистические возможности t -критерия в отношении переменных значительно выше, чем возможности критерия знаков.

Критерий Уилкоксона

Недостатком критерия знаков является то, что он никак не учитывает величину разности двух значений. Так, для него нет разницы между результатами сравнения пар (10; 0) и (6; 5): в обоих случаях знак разности будет положительным. Однако очевидно, что абсолютное значение разности также характеризует соотношение распределений. Для того чтобы учесть это, применяется критерий Уилкоксона. Этот критерий основан на подсчете абсолютных разностей между парами значений с последующим их ранжированием. Затем вычисляются средние значения рангов для положительных и отрицательных разностей (сдвигов). Уровень значимости подсчитывается на основе стандартизованного значения. Корректность применения этого критерия сомнительна, если переменная имеет небольшое число возможных значений, например 3-балльная шкала. В этом случае следует воспользоваться точными критериями. Основное диалоговое окно критерия Уилкоксона то же, что и для критерия знаков (см. рис. 12.6).

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере, как и в предыдущем, мы сравним результаты учащих по второму (тест2) и четвертому (тест4) тестам. Однако на этот раз вместо критерия знаков воспользуемся критерием Уилкоксона.

ШАГ 4Б

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Для двух независимых выборок, чтобы открыть диалоговое окно Критерии для двух связанных выборок (см. рис. 12.6). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

ШАГ 5Б

Для применения критерия Уилкоксона выполните следующую последовательность действий:

1. Щелкните на переменной числовые ряды [тест2], чтобы выделить ее, затем щелчком на стрелке перенесите переменную в правое окно. Ее имя появится в области Тестовые пары в строке Пара: 1 рядом с меткой Переменная 1.
2. Щелкните на переменной осведомленность [тест4], чтобы выделить ее, затем щелчком на стрелке перенесите переменную в правое окно. Ее имя появится в области Тестовые пары в строке Пара: 1 рядом с меткой Переменная 2.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Обратите внимание на то, что мы не устанавливали флажок Уилкоксона, поскольку он установлен по умолчанию. Результаты работы программы показаны на рис. 12.8.

Результаты применения критерия Уилкоксона и критерия знаков очень похожи. Частота каждого из трех исходов N осталась неизменной. Информация о каждом из исходов (кроме равенства) теперь включает в себя также среднее и суммарное значения для соответствующих рангов. Визуальный анализ исходных данных говорит о том, что значения теста 4 (осведомленность) в целом несколько превышают значения теста 2 (числовые ряды). Это демонстрирует и величина $Z = -2,493$, которая значительно превосходит

Ранги		N	Средний ранг	Сумма рангов
осведомленность	Отрицательные ранги	39 ^a	42,26	1648,00
числовые ряды	Положительные ранги	57 ^b	52,77	3008,00
	Связи	4 ^c		
	Всего	100		

a осведомленность < числовые ряды

b осведомленность > числовые ряды

c осведомленность = числовые ряды

Статистика критерия^a

	осведомленность-числовые ряды
Z	-2,493 ^a
Асимпт. знч. (двухсторонняя)	,013

a. Используются отрицательные ранги

Рис. 12.8. Фрагменты окна вывода после выполнения шага 5б

по модулю соответствующее значение, полученное ранее для критерия знаков. Уровень значимости $p = 0,013$, что говорит о статистической достоверности различий. Таким образом, мы убеждаемся в том, что критерий Уилкоксона является более чувствительным к различиям (более мощным), чем критерий знаков. Тем не менее он оказывается несколько хуже t -критерия, обеспечивающего уровень значимости 0,01, что подтверждает предпочтительность последнего для анализа метрических данных.

Критерий серий

Как следует из названия, *критерий серий* применяется для анализа последовательности объектов (явлений, событий), упорядоченных во времени или в порядке возрастания (убывания) значений измеренного признака. Кроме того, критерий требует представления последовательности в виде бинарной переменной, то есть как чередования событий 0 и 1. Математическая идея критерия основана на подсчете числа серий в упорядоченной последовательности событий двух типов, например 0 и 1. *Серия* — это последовательность однотипных событий, непосредственно перед и после которой произошли события другого типа. Гипотеза о случайном распределении событий 1 среди событий 0 может быть отклонена, если количество серий либо слишком мало (однотипные события имеют тенденцию к группированию), либо слишком велико (события 0 и 1 имеют тенденцию к чередованию).

Основное диалоговое окно критерия серий представлено на рис. 12.9.

После завершения шага 3 на экране должно присутствовать окно редактора данных. Рассмотрим применение критерия серий на примере проверки гипотезы о неслучайном чередовании юношей и девушек (переменная пол) в списке испытуемых в файле ex01.sav. Для задания точки раздела значений переменной на две категории предназначены флажок и поле Задаваемое. В данном случае в одну группу надо включить значения переменной пол, равные 1, а в другую группу — значения, равные 2.

ШАГ 4в

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Серии, чтобы открыть диалоговое окно Критерий серий (см. рис. 12.9).

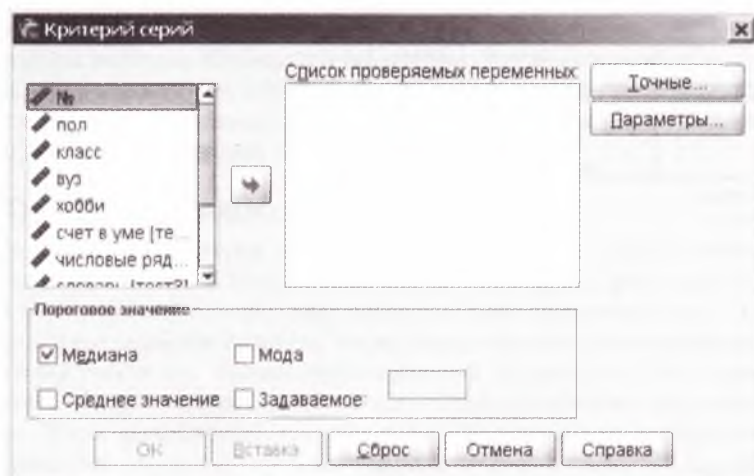


Рис. 12.9. Диалоговое окно Критерий серий

ШАГ 5в

Для применения критерия выполните следующую последовательность действий:

1. Щелкните сначала на переменной **пол**, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в поле **Список проверяемых переменных**.
2. В области **Пороговое значение** установите флажок **Задаваемое**, введите в расположенное рядом поле значение **2** и снимите флажок **Медиана**.
3. Щелкните на кнопке **ОК**, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.10.

Критерий серий					
	Проверяемое значение ^а	Всего наблюдений	Число серий	Z	Асимпт. знач. (двухсторонняя)
пол	2,00	100	49	,089	,929

а. Заданный пользователем.

Рис. 12.10. Фрагмент окна вывода после выполнения шага 5в

Количество серий равно 49. В результаты включено значение точки деления, введенное в поле **Задаваемое**. Величина *Z* и соответствующая значимость зависят от количества серий. Количество серий преобразуется к *z*-значению, для которого и определяется *p*-уровень. Большое значение *p*-уровня (0,929) свидетельствует о том, что чередование юношей и девушек в файле **ex01.sav** является случайным. Статистически значимый результат свидетельствовал бы о том, что чередование юношей и девушек в файле является неслучайным. Если бы при этом число серий было слишком велико, это свидетельствовало бы о том, что после юноши с высокой долей вероятности следует девушка (и наоборот). При малом значении числа серий можно было бы сделать вывод о том, что более вероятно группирование испытуемых в списке по половому признаку (после юноши чаще следует юноша, а после девушки — девушка).

Биномиальный критерий

Назначение *биномиального критерия* — определение вероятности того, что наблюдаемое распределение не отличается от ожидаемого (заданного) биномиального распределения. Свойством биномиального распределения является заранее заданное соотношение вероятностей двух взаимоисключающих событий (обычно равновероятное). Например, при многократном подбрасывании «правильной» монеты вероятности выпадения «орлов» и «решек» подчиняется биномиальному распределению.

В качестве примера мы снова исследуем распределение юношей и девушек в файле ex01.sav. Зная заранее, что здесь собраны данные о 61 ученице и 39 учениках, мы сможем проверить, отличается ли статистически достоверно это распределение (наблюдаемое) от ожидаемого (теоретического) равновероятного соотношения. Основное диалоговое окно биномиального критерия представлено на рис. 12.11.

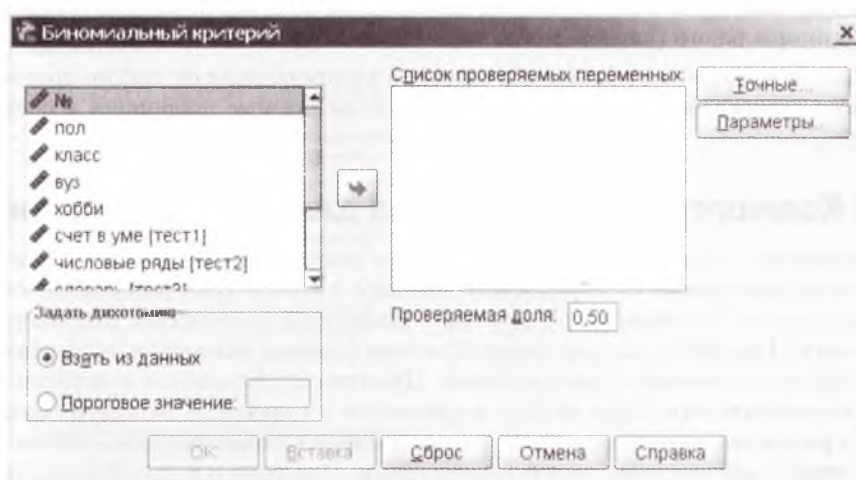


Рис. 12.11. Диалоговое окно биномиального критерия

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы попытаемся выяснить, является ли распределение юношей и девушек в выборке биномиальным.

ШАГ 4г

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Биномиальный, чтобы открыть диалоговое окно Биномиальный критерий (см. рис. 12.11).

ШАГ 5г

Для применения биномиального критерия выполните следующую последовательность действий:

1. Щелкните сначала на переменной *пол*, чтобы выделить ее, а затем — на кнопке со стрелкой, чтобы переместить переменную в Список проверяемых переменных. Обратите внимание на то, что ожидаемое соотношение, заданное по умолчанию (Проверяемая доля), равно 0,5.
2. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.12.

Биномиальный критерий

		Категория	N	Наблюденная доля	Проверяемая доля	Асимпт. змч (двухсторонняя)
пол	Группа 1	МУЖ	39	,39	,50	,035 ^a
	Группа 2	ЖЕН	61	,61		
	Всего		100	1,00		

a. Используется Z аппроксимация.

Рис. 12.12. Фрагмент окна вывода после выполнения шага 5г

Ожидаемая пропорция для биномиального теста равна 0,5 для обеих групп. Наблюдаемая пропорция для каждой из групп определяется как отношение размера группы (N) к размеру выборки (100). Как можно видеть, наблюдаемые пропорции значительно отличаются от 0,5 и составляют 0,39 для юношей и 0,61 для девушек. Уровень значимости, равный 0,035, свидетельствует о статистически достоверном отличии исследуемого распределения от биномиального (равновероятного).

При желании можно проверять отличие наблюдаемого распределения от любого другого биномиального распределения, задавая необходимые ожидаемые пропорции в поле Проверяемая доля.

Критерий Колмогорова — Смирнова для одной выборки

Критерий Колмогорова — Смирнова для одной выборки позволяет определить, отличается ли заданное распределение от нормального (эксцесс и асимметрия распределения равны 0), равномерного (значения распределены с одинаковой плотностью, например, как у целых чисел от 1 до 1000), распределения Пуассона (среднее значение и дисперсия равны λ ; при больших значениях λ распределение Пуассона приближается к нормальному) или экспоненциального. Суть метода заключается в сравнении эмпирического (наблюдаемого) распределения накопленных частот выборки с теоретическим (ожидаемым) распределением накопленных частот (нормальным, Пуассона и т. д.). В качестве примера исследуем распределение значений переменной `отметка1` в файле `ex01.sav` на соответствие нормальному распределению. Основное диалоговое окно критерия Колмогорова — Смирнова для одной выборки представлено на рис. 12.13.

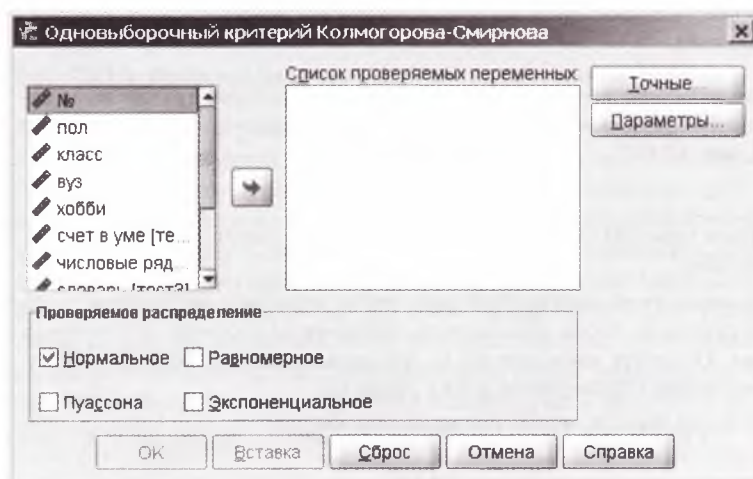


Рис. 12.13. Диалоговое окно Одновыборочный критерий Колмогорова — Смирнова

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы сравним распределение переменной `отметка1` с нормальным распределением.

ШАГ 4д

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Одновыборочный Колмогорова — Смирнова, чтобы открыть диалоговое окно Одновыборочный критерий Колмогорова — Смирнова (см. рис. 12.13).

ШАГ 5д

Для применения критерия выполните следующую последовательность действий:

1. Щелкните сначала на переменной `отметка1`, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в Список проверяемых переменных.
2. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.14.

Одновыборочный критерий Колмогорова-Смирнова		
		отметка1
N		100
Нормальные параметры	Среднее	3,9630
	Стд. отклонение	,30597
Разности экстремумов	Модуль	,072
	Положительные	,072
	Отрицательные	-,051
Статистика Z Колмогорова-Смирнова		,716
Асимпт. знач. (двухсторонняя)		,685

а. Сравнение с нормальным распределением
б. Оценивается по данным

Рис. 12.14. Фрагмент окна вывода после выполнения шага 5д

В строке Разности экстремумов приведены Модуль, а также Положительные и Отрицательные отклонения исследуемого распределения от теоретического (в данном случае нормального). Строка Статистика Z Колмогорова — Смирнова содержит z-значение, уровень значимости которого равен 0,685 (последняя строка). Это означает, что распределение значений переменной `отметка1` статистически не отличается от нормального ($p > 0,05$).

Критерий хи-квадрат для одной выборки

Рассматриваемый в данном разделе критерий χ^2 для одной выборки несколько отличается от рассмотренного ранее критерия χ^2 для таблиц сопряженности (см. главу 8). В данном случае в качестве ожидаемого (теоретического) распределения обычно выступает равномерное распределение объектов по градациям переменной, в отношении которой применяется критерий. Далее будет приведен пример применения критерия χ^2 к переменной `вуз` из файла `ex01.sav`. Поскольку число объектов (N) в файле `ex01.sav` равно 100, а переменная `вуз` имеет 4 градации, ожидаемые частоты для каждой градации равны $100/4 = 25$. Применение рассматриваемого критерия допускает задание не только равномерного ожидаемого распределения, но и любого другого. Например, можно проверить гипотезу о том, что количества учащихся, предпочитающих 4 категории специализаций, соотно-

20-20-30-30. Для этого переключатель Ожидаемые значения (рис. 12.15) следует установить в положение Значения, а затем при помощи поля и кнопки Добавить последовательно ввести в список значения 20, 20, 30, 30. После этих действий ожидаемые частоты изменятся в соответствии с заданными пропорциями. Основное диалоговое окно критерия χ^2 для одной выборки представлено на рис. 12.15.

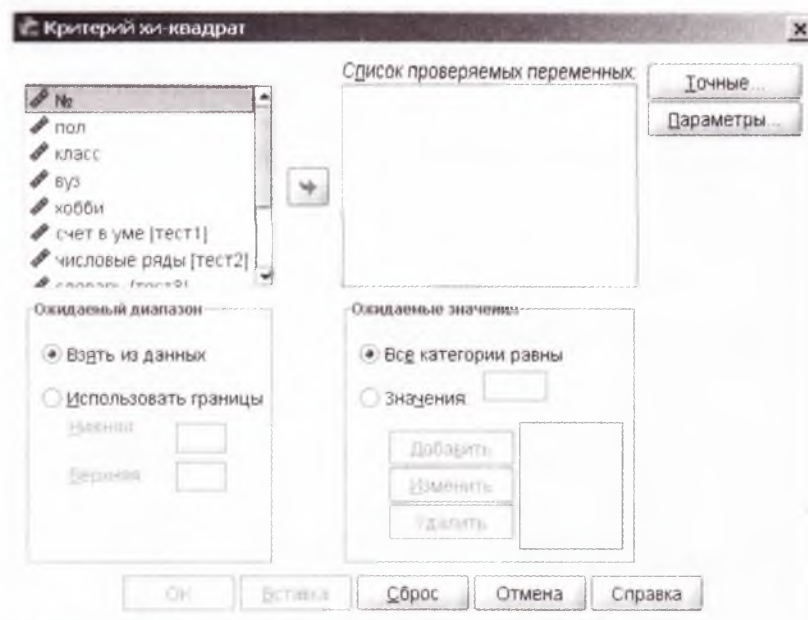


Рис. 12.15. Диалоговое окно Критерий хи-квадрат

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы применим критерий χ^2 к переменной вуз.

ШАГ 4е В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Хи-квадрат, чтобы открыть диалоговое окно Критерий хи-квадрат (см. рис. 12.15).

ШАГ 5е Для применения критерия выполните следующую последовательность действий:

1. Щелкните сначала на переменной вуз, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в Список проверяемых переменных.
2. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.16.

Первая из таблиц демонстрирует заметные различия между наблюдаемыми и ожидаемыми частотами. Остаток — это разность между наблюдаемыми и ожидаемыми частотами. Число степеней свободы (ст. св.) определяется как число значений (градаций) переменной, уменьшенное на 1. Уровень значимости ($p = 0,002$) свидетельствует о статистически достоверном отличии наблюдаемого распределения предпочтений от равномерного распределения.

вуз			
	Наблюде- нное N	Ожидаемое N	Остаток
ГУМ	22	25,0	-3,0
ЭКОН	31	25,0	6,0
ТЕХН	11	25,0	-14,0
ЕСТ_Н	36	25,0	11,0
Всего	100		

Статистика критерия	
	вуз
Хи-квадрат ^a	14,480
ст. св.	3
Асимпт. знач.	,002

a. Частоты, меньшие 5, ожидалось в 0 ячеек (,0%).
Минимальная ожидаемая частота равна 25,0

Рис. 12.16. Фрагменты окна вывода после выполнения шага 5е

Сравнение K независимых выборок и критерий Краскела — Уоллеса

Для сравнения более двух независимых выборок по уровню выраженности переменной применяются несколько критериев: *H-критерий Краскела — Уоллеса*, *критерий медианы*, *критерий Джонкира — Терпстра*. Среди них наибольшей чувствительностью к различиям обладает *H-критерий Краскела — Уоллеса*. Этот критерий является непараметрическим аналогом однофакторного дисперсионного анализа, отличаясь от него в двух отношениях. Во-первых, критерий Краскела — Уоллеса основан не на сравнении средних значений и дисперсий переменных, а на сравнении средних рангов. Во-вторых, вместо вычисления *F-критерия* на основе сравнения средних рангов с ожидаемыми значениями вычисляется критерий хи-квадрат. Для нормальных распределений однофакторный дисперсионный анализ обеспечивает более точные результаты, чем критерий Краскела — Уоллеса, однако применение последнего рекомендуется для распределений, отличающихся от нормального.

H-критерий Краскела — Уоллеса сходен с *U-критерием Манна — Уитни*. Как и последний, он оценивает степень пересечения (совпадения) нескольких рядов значений измеренного признака. Чем меньше совпадений, тем больше различаются ряды, соответствующие сравниваемым выборкам. Основная идея *H-критерия Краскела — Уоллеса* основана на представлении всех значений сравниваемых выборок в виде одной общей последовательности упорядоченных (ранжированных) значений с последующим вычислением среднего ранга для каждой из выборок. Если выполняется статистическая гипотеза об отсутствии различий, можно ожидать, что все средние ранги примерно равны и близки к общему среднему рангу.

Диалоговое окно непараметрических критериев для нескольких независимых выборок представлено на рис. 12.17.

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы проведем сравнение трех групп учащихся, различающихся внешними увлечениями (переменная хобби) и успеваемостью в выпускном классе (переменная отметка2).

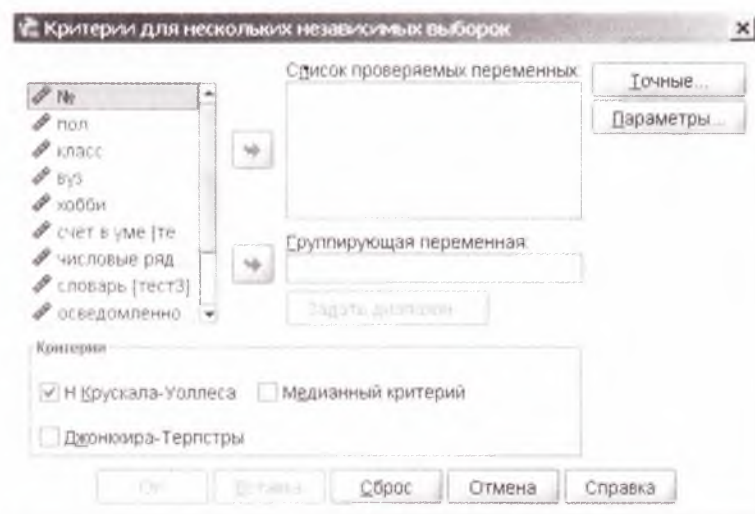


Рис. 12.17. Диалоговое окно Критерии для нескольких независимых выборок

ШАГ 4ж

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Для K независимых выборок, чтобы открыть диалоговое окно Критерии для нескольких независимых выборок (см. рис. 12.17).

ШАГ 5ж

Для применения критерия выполните следующую последовательность действий:

1. Щелкните сначала на переменной *отметка2*, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список проверяемых переменных.
2. Щелкните сначала на переменной *хобби*, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Группирующая переменная.
3. Щелкните на кнопке Задать диапазон, чтобы открыть одноименное диалоговое окно.
4. В поле Минимум введите значение 1, нажмите клавишу Tab, чтобы переместить фокус ввода в поле Максимум, введите значение 3 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Критерии для нескольких независимых выборок.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Обратите внимание на то, что флажок Н Краскала — Уоллеса установлен по умолчанию, поэтому мы ничего не меняли для выбора критерия.

Результаты работы программы показаны на рис. 12.18.

В первой таблице для каждой группы представлены ее численность и средний ранг. Во второй таблице указаны значение критерия χ^2 , число степеней свободы и уровень статистической значимости. Результаты обработки показывают статистически достоверную связь внешкольных увлечений учащихся с успеваемостью в выпускном классе.

Ранги		
хобби	N	Средний ранг
отметка2 спорт	33	38,18
компьютер	37	54,28
искусство	30	59,38
Всего	100	

Статистики критерия^{a,b}

	отметка2
Хи-квадрат	9,437
ст.св	2
Асимпт. знач.	,009

a. Критерий Краскела-Уоллеса

b. Группирующая переменная: хобби

Рис. 12.18. Фрагменты окна вывода после выполнения шага 5ж

Сравнение нескольких зависимых выборок и критерий Фридмана

Критерий Фридмана является непараметрическим аналогом однофакторного дисперсионного анализа для повторных измерений. Он позволяет проверять гипотезы о различии более двух зависимых выборок (повторных измерений) по уровню выраженности изучаемой переменной. Критерий Фридмана может быть более эффективен, чем его метрический аналог — однофакторный дисперсионный анализ — в случаях повторных измерений изучаемого признака на небольших выборках и при отклонении распределения от нормального.

Критерий Фридмана весьма сходен с критерием Краскела — Уоллеса и основан на ранжировании ряда повторных измерений для каждого объекта выборки. Затем вычисляется сумма рангов для каждого из условий (повторных измерений). Если выполняется статистическая гипотеза об отсутствии различий между повторными измерениями, можно ожидать примерного равенства сумм рангов для этих условий. Чем больше различаются зависимые выборки по изучаемому признаку, тем больше эмпирическое значение вычисляемого значения критерия χ^2 , по которому определяется p -уровень значимости.

Диалоговое окно команды для применения критерия Фридмана представлено на рис. 12.19.

После завершения шага 3 на экране должно присутствовать окно редактора данных. В этом примере мы сравним результаты тестов тест1, тест2, тест3, тест4 и тест5 для всех учащихся.

ШАГ 4з

В меню Анализ выберите команду Непараметрические критерии ► Устаревшие диалоговые окна ► Для К связанных выборок, чтобы открыть диалоговое окно Критерии для нескольких связанных выборок (см. рис. 12.19).

ШАГ 5з

Для применения критерия выполните последовательность действий:

1. Щелкните сначала на переменной тест1, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в поле Проверяемые переменные.
2. Повторите предыдущее действие для переменных тест2, тест3, тест4 и тест5.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Результаты работы программы показаны на рис. 12.20.

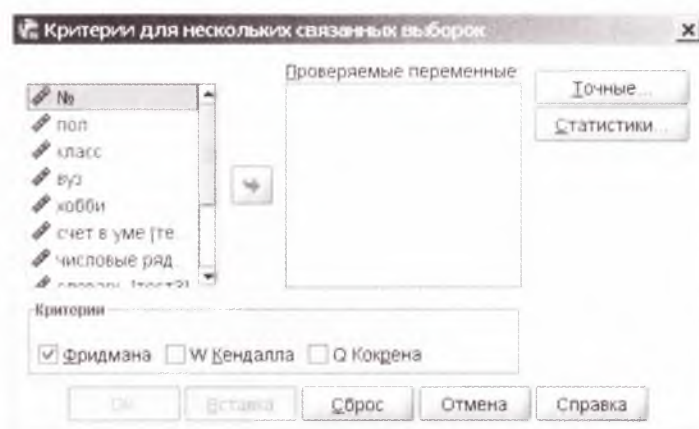


Рис. 12.19. Диалоговое окно критерия Фридмана

Ранги	
	Средний ранг
счет в уме	2,50
числовые ряды	2,85
словарь	3,41
осведомленность	3,08
кратковременная память	3,37

Статистика критерия ^a	
N	100
Хи-квадрат	29,696
ст.св.	4
Асимпт. знч.	,000

a. Критерий Фридмана

Рис. 12.20. Фрагменты окна вывода после выполнения шага 5з

Средние ранги определяются следующим образом: сначала для каждого наблюдения значения сравниваемых переменных ранжируются (по строке). Затем для каждой из сравниваемых переменных вычисляется средний ранг по всем объектам. Определяемый по критерию χ^2 уровень значимости Асимпт. знч. $< 0,001$. Он свидетельствует о статистически значимой разнице между пятью результатами тестирования. Различаться может любая пара переменных, и без попарного сравнения невозможно выяснить, какие именно пары вносят значимый вклад в статистическую достоверность результата.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

13 Однофакторный дисперсионный анализ

169	Пошаговые алгоритмы вычислений
174	Представление результатов
177	Терминология
177	Завершение анализа и выход из программы

Однофакторный дисперсионный анализ в SPSS реализуется с помощью команды Однофакторный дисперсионный анализ. Команды подменю Общая линейная модель, описываемые в главах 14 и 15, также отчасти позволяют проводить подобный анализ, однако их возможности несколько более специфичны, чем у команды Однофакторный дисперсионный анализ.

Дисперсионный анализ (Analysis Of Variances, ANOVA — общепринятое обозначение метода) — это процедура сравнения средних значений выборок, на основании которой можно сделать вывод о соотношении средних значений генеральных совокупностей. Ближайшим и более простым аналогом ANOVA является t -критерий, применение которого рассматривалось в главе 11. В отличие от t -критерия дисперсионный анализ предназначен для сравнения не двух, а нескольких выборок. Слово «дисперсионный» в названии указывает на то, что в процессе анализа сопоставляются компоненты дисперсии изучаемой переменной. Общая изменчивость переменной раскладывается на две составляющие: межгрупповую (факторную), обусловленную различием групп (средних значений), и внутригрупповую (ошибки), обусловленную случайными (неучтенными) причинами. Чем больше частное от деления межгрупповой изменчивости на внутригрупповую (F -отношение), тем больше различаются средние значения сравниваемых выборок и тем выше статистическая значимость этого различия.

Итак, название указывает на то, что вывод о различии средних значений делается на основе анализа компонентов дисперсии. А что означает слово «однофакторный»? Применяя команду Однофакторный дисперсионный анализ, вы увидите, что в ней можно задать единственную зависимую переменную (при этом она обязательно должна быть количественного, а точнее, метрического типа) и единственную независимую переменную (всегда номинальную, имеющую несколько градаций). Различные модели дисперсионного анализа, описанные в двух следующих главах, допускают наличие нескольких независимых переменных. Многомерный дисперсионный анализ, с которым вы столкнетесь в главах 15 и 16, позволяет анализировать как множество независимых, так и множество зависимых переменных.

Итак, при однофакторном дисперсионном анализе сравниваются между собой средние значения каждой выборки и вычисляется общий уровень значимости различий. Обратите внимание на то, что вывод по результатам ANOVA касается общего различия всех сравниваемых средних без конкретизации того, какие именно выборки различаются,

а какие нет. Для идентификации пар выборок, отличающихся друг от друга средними значениями, используются апостериорные критерии парных сравнений (Post Hoc), а для более сложных сопоставлений — метод контрастов (Contrasts).

Пошаговые алгоритмы вычислений

Для проведения однофакторного дисперсионного анализа будем использовать файл ex01.sav. В роли зависимой переменной выступит переменная тест1, а независимая переменная класс разделит объекты на 3 выборки, средние значения которых мы будем сравнивать, но сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex01.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

Однофакторный дисперсионный анализ

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню.

ШАГ 4. В меню Анализ выберите команду Сравнение средних ► Однофакторный дисперсионный анализ. На экране появится диалоговое окно Однофакторный дисперсионный анализ (рис. 13.1).

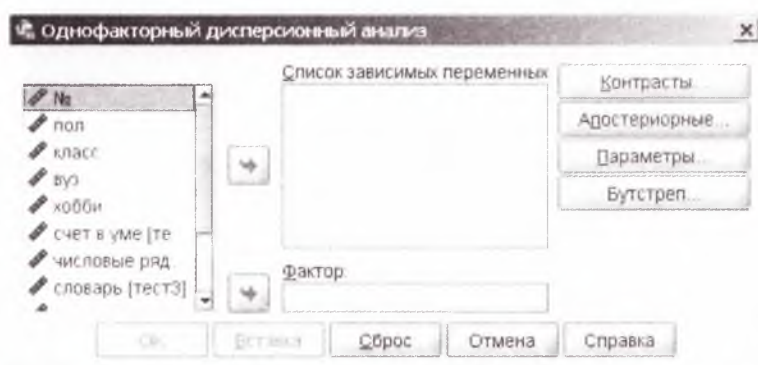


Рис. 13.1. Диалоговое окно Однофакторный дисперсионный анализ

Структура диалогового окна Однофакторный дисперсионный анализ вполне типична для большинства диалоговых окон SPSS. Слева мы видим список переменных текущего файла данных. В правой части окна расположены кнопки Контрасты, Апостериорные и Параметры, которые мы будем использовать при обработке. Список зависимых переменных предназначен для задания одной или нескольких зависимых переменных (в нашем примере будет использоваться единственная зависимая переменная тест1). Зависимые переменные должны быть метрического типа. Если в списке указаны несколько зависимых переменных, SPSS выполнит анализ для каждой из них. Под окном Список зависимых переменных находится поле Фактор, в котором нужно указать единственную независимую переменную, имеющую несколько градаций (в нашем случае хобби). Таким

образом, мы сравним результаты первого теста (счет в уме) для трех групп учащихся, различающихся внешкольными увлечениями.

На шаге 5 мы будем сравнивать между собой средние значения переменной тест1 для каждой из выборок по уровням переменной хобби.

ШАГ 5

После выполнения шага 4 должно быть открыто диалоговое окно Однофакторный дисперсионный анализ (см. рис. 13.1). При необходимости повторите шаг 4 и выполните следующие действия:

1. Щелкните сначала на переменной тест1, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список зависимых переменных.
2. Щелкните сначала на переменной хобби, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Фактор.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Существует два дополнительных действия, которые иногда желательно выполнять в процессе анализа. Приведенная последовательность инструкций позволяет получить результаты сравнения средних значений выборок, однако ни сами средние значения, ни результаты парного сравнения выборок в выводимых данных отображены не будут. Первая проблема решается при помощи кнопки Параметры. Диалоговое окно Однофакторный дисперсионный анализ: Параметры, появляющееся после щелчка на этой кнопке, показано на рис. 13.2. Установка в области Статистики флажка Описательные приведет к включению в выводимые данные всех средних значений, стандартных отклонений, стандартных ошибок, границ доверительных интервалов в 95 %, а также минимумов и максимумов выборок. Флажок Проверка однородности дисперсии позволяет вывести информацию о степени пригодности данных к дисперсионному анализу, а с помощью флажка График средних можно построить диаграмму, на которой будут изображены средние значения для каждой выборки.

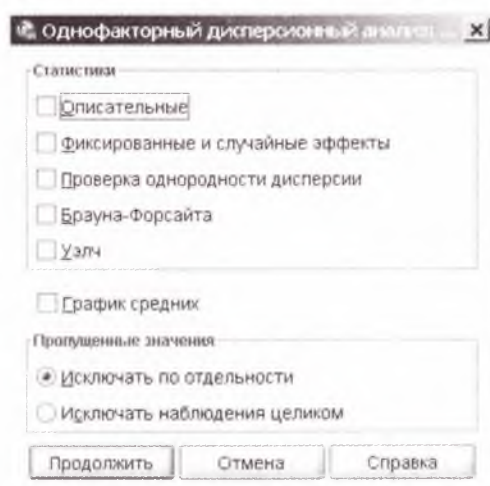


Рис. 13.2. Диалоговое окно Однофакторный дисперсионный анализ: Параметры

Парные сравнения

Нередко нас могут заинтересовать результаты парных сравнений уровней независимой переменной, и для этой цели в диалоговом окне Однофакторный дисперсионный анализ предусмотрена специальная кнопка Апостериорные. Термин «апостериорные» означает, что эта процедура проводится после установления статистически достоверного результата однофакторного дисперсионного анализа. Если результаты дисперсионного анализа оказались статистически недостоверными, применение процедуры парных сравнений некорректно. При щелчке на кнопке Апостериорные открывается диалоговое окно Однофакторный дисперсионный анализ: Апостериорные множественные сравнения (рис. 13.3). В этом диалоговом окне с помощью флажков можно задать 14 критериев для выборок с одинаковой дисперсией и 4 критерия для выборок с разной дисперсией.

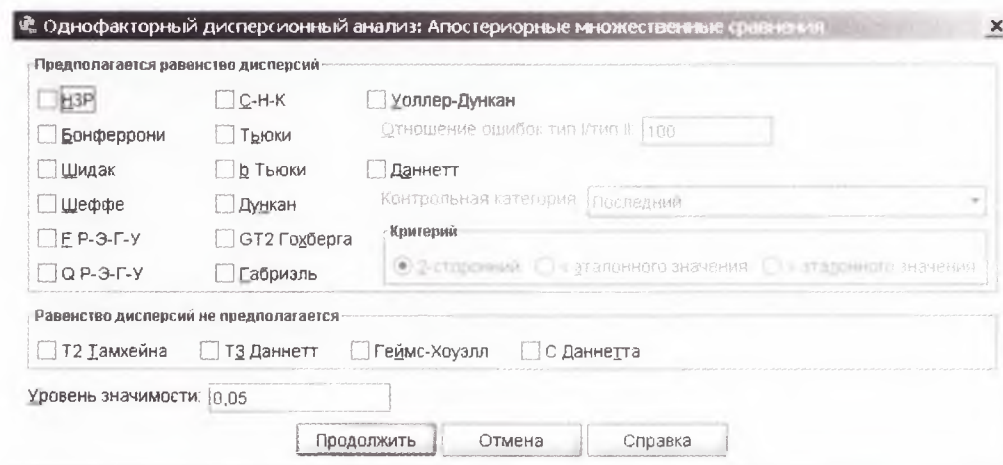


Рис. 13.3. Диалоговое окно Однофакторный дисперсионный анализ: Апостериорные множественные сравнения

Большинство из тестов используются очень редко, поэтому далее приведены описания только нескольких наиболее популярных:

- ▶ НЗР (Наименьшая значимая разность) — этот критерий представляет собой совокупность t -критериев для всех возможных пар градаций фактора. Критерий наименьшей значимой разности является одним из самых либеральных, поскольку наиболее подвержен ошибкам. Например, если независимая переменная имеет 5 уровней, то будет проведено 10 сравнений. При уровнях значимости каждого из сравнений, равных 0,05, существует почти 40 %-ная вероятность того, что хотя бы в одном из тестов значимый результат получен случайно.
- ▶ Бонферрони — этот критерий схож с критерием наименьшей значимой разности, однако лишен недостатка, связанного с повторными проверками: в нем уровень значимости делится на число сравнений. Таким образом, критерий Бонферрони является наиболее консервативным.
- ▶ Шеффе — критерий менее консервативный, чем критерий Бонферрони (используется F -критерий вместо t -критерия).

- Тьюки — критерий Тьюки использует статистику Стьюдента для определения различий между группами. Этот критерий часто применяется в случаях, когда исследуемый фактор имеет большое количество уровней.

Самым консервативными из предложенных является критерий Бонферрони, за ним следует критерий Шеффе. Часто используется также критерий Тьюки, называемый еще критерием подлинной значимости (Honestly Significant Difference, HSD). Последний задает наименьшую величину разности средних значений выборок, которую можно считать значимой. Например, если $HSD = 2,5$, а для двух выборок получены средние значения 3,7 и 6,3, то разность между ними, равная 2,6, согласно Тьюки, является значимой, поскольку она превышает величину HSD. При использовании критерия Тьюки программа SPSS также включает в вывод дополнительную статистическую информацию.

Необходимо отметить, что все описанные критерии (на самом деле большинство критериев парных сравнений) применяются в предположении, что дисперсии всех ячеек равны. Исключение составляют 4 критерия, флажки для которых помещены в отдельной области Равенство дисперсий не предполагается в нижней части диалогового окна, — они применяются в случаях, когда дисперсии ячеек разные.

В следующем примере выполним однофакторный дисперсионный анализ, а в выводимые результаты включим описательные статистики и критерий однородности дисперсии. Для парных сравнений воспользуемся критерием Шеффе. Как и в предыдущем примере, зависимой переменной будет переменная тест1, а независимой — переменная хобби.

ШАГ 5а

После выполнения шага 4 должно быть открыто диалоговое окно Однофакторный дисперсионный анализ (см. рис. 13.1). При необходимости повторите шаг 4 и выполните следующие действия:

1. Щелкните сначала на переменной тест1, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список зависимых переменных.
2. Щелкните сначала на переменной хобби, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Фактор.
3. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Однофакторный дисперсионный анализ: Параметры, показанное на рис. 13.2.
4. Установите флажки Описательные, Проверка однородности дисперсии и График средних, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Однофакторный дисперсионный анализ.
5. Щелкните на кнопке Апостериорные, чтобы открыть диалоговое окно Однофакторный дисперсионный анализ: Апостериорные множественные сравнения, показанное на рис. 13.3.
6. Установите флажок Шеффе и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Однофакторный дисперсионный анализ.
7. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Контрасты

Верхняя кнопка в правой части окна Однофакторный дисперсионный анализ имеет название Контрасты. Она предназначена для вызова диалогового окна Однофакторный дис-

персионный анализ: Контрасты, представленного на рис. 13.4. Это окно позволяет осуществлять различные сравнения выборок по градациям независимой переменной. Так, вы можете сравнивать одну градацию с другой, одну градацию со всеми остальными или разбить все градации на 2 группы и затем сравнить их между собой. Сравнение сводится к применению модифицированного варианта t -критерия. Отметим, что применение метода контрастов не требует предварительного получения статистически достоверного результата анализа ANOVA, в отличие от процедуры апостериорного множественного парного сравнения.

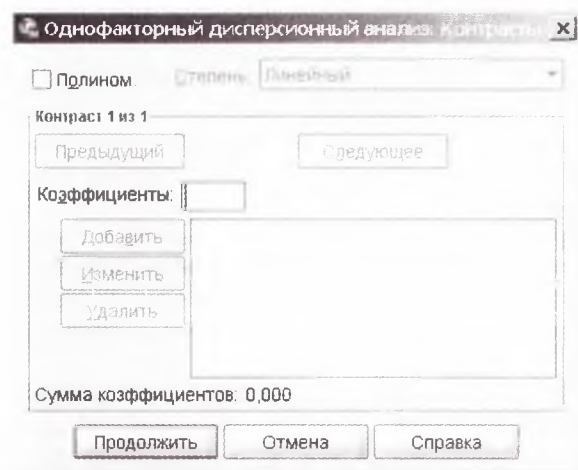


Рис. 13.4. Диалоговое окно Однофакторный дисперсионный анализ: Контрасты

В нашем примере независимая переменная хобби имеет 3 градации: 1 — спорт, 2 — компьютер, 3 — искусство. Для задания контраста предназначены поле и список Коэффициенты. Заполнение списка происходит следующим образом. Каждой градации фактора вы должны сопоставить число, определяющее его роль в контрасте: отрицательное число соответствует одной группе, положительное число — другой группе, а ноль означает, что градация в сравнениях не задействована. При этом абсолютные величины коэффициентов неважны, но последние должны вводиться в порядке следования градаций и в сумме давать нулевое значение. Например, если вам необходимо сравнить увлекающихся спортом с остальными учащимися, вы можете закодировать градации последовательностью -2 , 1 и 1 , а если вы хотите сравнить увлекающихся спортом только с теми, кто увлекается искусством, необходимо задать последовательность 1 , 0 и -1 . Не забывайте, что сумма всех коэффициентов обязательно должна равняться 0 .

Помещение чисел в список осуществляется путем ввода значения в поле справа от названия списка и щелчка на кнопке Добавить. Если нужно создать несколько «контрастов», т. е. разбиений на группы, щелкните на кнопке Следующее справа от метки Контраст 1 из 1 и повторите описанную процедуру.

Следующий пример иллюстрирует применение контрастов: для сравнения учащихся, увлекающихся компьютером, с теми, кто имеет другие увлечения, а также для сравнения увлекающихся спортом с остальными учащимися.

ШАГ 56

После выполнения шага 4 должно быть открыто диалоговое окно Однофакторный дисперсионный анализ, показанное на рис. 13.1. Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия.

1. Щелкните сначала на переменной тест1, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в Список зависимых переменных.
2. Щелкните сначала на переменной хобби, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Фактор.
3. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Однофакторный дисперсионный анализ: Параметры (см. рис. 13.2).
4. Установите флажки Описательные, Проверка однородности дисперсии и График средних а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Однофакторный дисперсионный анализ.
5. Щелкните на кнопке Апостериорные, чтобы открыть диалоговое окно Однофакторный дисперсионный анализ: Апостериорные множественные сравнения (см. рис. 13.3).
6. Установите флажок НЗР и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Однофакторный дисперсионный анализ.
7. Щелкните на кнопке Контрасты, чтобы открыть диалоговое окно Однофакторный дисперсионный анализ: Контрасты (см. рис. 13.4).
8. Нажмите клавишу Tab, чтобы перевести указатель в поле Коэффициенты, введите число 1 и щелкните на кнопке Добавить, задав первое число в списке.
9. Повторите предыдущее действие сначала для чисел -2 и 1, затем, щелкнув на кнопке Следующее, — для чисел -2, 1, 1, после чего щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Однофакторный дисперсионный анализ.
10. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически активизирует окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

В этом разделе мы приводим некоторые результаты процедур, описываемых в этой главе.

Однофакторный дисперсионный анализ

Фрагмент выводимых данных, сгенерированных программой после выполнения шага 5а, относящийся к однофакторному дисперсионному анализу, приведен на рис. 13.5.

Самым важным в этой таблице является уровень значимости $p = 0,002$. Он указывает на то, что разность между средними значениями переменной тест1 для трех групп статистически достоверна.

Однофакторный дисперсионный анализ

счет в уме

	Сумма квадратов	Ст. св.	Средний квадрат	F	Знач.
Между группами	82,969	2	41,484	6,556	,002
Внутри групп	613,781	97	6,328		
Итого	696,750	99			

Рис. 13.5. Фрагмент окна вывода после выполнения шага 5а

Описательные статистики

Фрагмент выводимых данных, сгенерированных программой после выполнения шага 5б, относящийся к описательным статистикам, приведен на рис. 13.6.

Описательные статистики

счет в уме

	N	Среднее	Стд. отклонение	Стд. ошибка	95% доверительный интервал для среднего		Минимум	Максимум
					Нижняя граница	Верхняя граница		
спорт	33	9,67	2,582	,449	8,75	10,58	4	14
компьютер	37	11,43	2,102	,346	10,73	12,13	7	17
искусство	30	9,43	2,885	,527	8,36	10,51	5	15
Итого	100	10,25	2,653	,265	9,72	10,78	4	17

Рис. 13.6. Фрагмент окна вывода после выполнения шага 5б, относящийся к описательным статистикам

Апостериорные парные сравнения

Фрагмент выводимых данных, сгенерированных программой после выполнения шага 5а, относящийся к парным сравнениям, приведен на рис. 13.7.

Средние значения переменной тест1 для каждой из трех выборок были перечислены в таблице, приведенной на рис. 13.6, здесь же даны разности между этими значениями. Знаком звездочки на рис. 13.7 помечены те пары выборок, для которых разность средних значений статистически достоверна, то есть со значением уровня значимости 0,05 и меньше. Проанализировав полученные данные, можно сделать вывод о том, что результаты теста 1 для тех, кто увлекается компьютером, статистически более значимы, чем для тех, кто увлекается спортом и искусством. Те же, кто увлекается спортом и искусством, по результатам теста 1 статистически достоверно не различаются.

Множественные сравнения

Зависимая переменная: счет в уме
Шеффе

(I) хобби	(J) хобби	Разность средних (I-J)	Стд. ошибка	Знач.	95% доверительный интервал	
					Нижняя граница	Верхняя граница
спорт	компьютер	-1,766*	,602	,016	-3,26	-,27
	искусство	,233	,635	,935	-1,34	1,81
компьютер	спорт	1,766*	,602	,016	,27	3,26
	искусство	1,999*	,618	,007	,46	3,54
искусство	спорт	-,233	,635	,935	-1,81	1,34
	компьютер	-1,999*	,618	,007	-3,54	-,46

* Разность средних значима на уровне .05

Рис. 13.7. Фрагмент окна вывода после выполнения шага 5а, относящийся к парным сравнениям

Критерий Ливиня

Фрагмент выводимых результатов, сгенерированных программой после выполнения шага 5б (п. 4), относящийся к критерию однородности дисперсии Ливиня, приведен на рис. 13.8.

Критерий однородности дисперсий

счет в уме

Статистика Ливиня	Ст. св. 1	Ст. св. 2	Знач.
1,858	2	97	,161

Рис. 13.8. Фрагмент окна вывода после выполнения шага 5б, относящийся к критерию Ливиня

Критерий однородности дисперсии Ливиня со значимостью 0,161 показал, что дисперсии для каждой из групп статистически достоверно не различаются. Следовательно, результаты ANOVA могут быть признаны корректными. Если бы результат применения критерия Ливиня оказался статистически достоверным, это послужило бы основанием для сомнения в корректности применения ANOVA.

Контрасты

Фрагменты выводимых результатов, сгенерированных программой после выполнения шага 5б, относящиеся к применению метода контрастов, приведены на рис. 13.9.

Коэффициенты контраста

Контраст	хобби		
	спорт	компьютер	искусство
1	1	-2	1
2	-2	1	1

Критерии контраста

			Значение контраста	Стд. ошибка	t	Ст. св.	Знач. (2-сторон.)
Контраст							
счет в уме	Предполагается равенство дисперсий	1	-3,76	1,042	-3,611	97	,000
		2	1,53	1,072	1,430	97	,156
	Равенство дисперсий не предполагается	1	-3,76	,978	-3,848	89,220	,000
		2	1,53	1,098	1,396	61,897	,168

Рис. 13.9. Фрагменты окна вывода после выполнения шага 5б, относящиеся к применению метода контрастов

Первая из таблиц на рисунке содержит коэффициенты, введенные при группировании уровней. Для каждого контраста были применены по два t -критерия, один из которых проводился с допущением о равных дисперсиях, а другой — с допущением о неравных дисперсиях. Первый контраст оказался статистически достоверным: те, кто увлекается компьютером, имеют статистически достоверно более высокий результат по переменной тест1. Второй контраст не достигает статистической значимости: учащиеся, увлекаю-

ющиеся спортом, статистически достоверно не отличаются от других учащихся по переменной тест1.

Терминология

- ▶ Сумма квадратов в строке Между группами означает сумму квадратов разностей между общим средним значением и средними значениями каждой группы, умноженными на весовые коэффициенты, равные числу объектов в группе, а в строке Внутри групп — сумму квадратов разностей среднего значения каждой группы и каждого значения этой группы.
- ▶ Ст. св. в строке Между группами означает межгрупповое число степеней свободы, равное числу групп, уменьшенному на 1, в строке Внутри групп — внутригрупповое число степеней свободы, равное разности между числом объектов и числом групп.
- ▶ Средний квадрат — отношение суммы квадратов к числу степеней свободы.
- ▶ F — F -критерий, отношение среднего квадрата между группами к среднему квадрату внутри группы.
- ▶ Знч. — статистическая значимость, вероятность того, что наблюдаемые различия случайны. Значимость $p = 0,002$ свидетельствует о статистически достоверных различиях.
- ▶ N — число объектов для каждой из градаций переменной хобби.
- ▶ Стд. отклонение — стандартное отклонение, мера разброса значений распределения относительно среднего.
- ▶ Стд. ошибка — стандартная ошибка среднего, отношение стандартного отклонения к квадратному корню из числа объектов.
- ▶ 95 % доверительный интервал для среднего — при большом числе выборок из генеральной совокупности 95 % средних значений этих выборок попадут в интервал, определяемый указанными в таблице границами.
- ▶ Минимум — наименьшее из наблюдаемых значений для группы.
- ▶ Максимум — наибольшее из наблюдаемых значений для группы.
- ▶ Значение контраста — это значение не представляет интереса для исследователя, поскольку является всего лишь произведением разности средних с весовым коэффициентом.
- ▶ t — значение t -критерия, отношение величины различия средних к стандартной ошибке.
- ▶ Знч. (2-сторон) — двусторонний уровень значимости. Вероятность того, что отличие от нуля является случайным.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2).

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

14 Многофакторный дисперсионный анализ

180	Файлы данных для группы методов Общая линейная модель
181	Дисперсионный анализ с двумя факторами
182	Дисперсионный анализ с тремя и более факторами
183	Влияние ковариат
183	Пошаговые алгоритмы вычислений
188	Представление результатов
192	Терминология, используемая при выводе
192	Завершение анализа и выход из программы

В этой главе речь пойдет о дисперсионном анализе (ANOVA) с двумя и более факторами. В нем участвует единственная зависимая переменная, которая, как и при однофакторном дисперсионном анализе, должна быть метрической. Главное отличие многофакторного анализа от однофакторного заключается в том, что в нем участвуют не одна, а несколько независимых переменных, каждая из которых должна быть номинальной, то есть иметь несколько градаций, или уровней. Многофакторный дисперсионный анализ реализуется в SPSS с помощью команды *Общая линейная модель* ► ОЛМ-одномерная.

Дисперсионному анализу посвящены главы 13–16. Мы настоятельно рекомендуем читать эти главы по порядку, поскольку в них в развитии описывается одна и та же статистическая концепция. Представление об однофакторном анализе является основой для понимания дисперсионного анализа с двумя и более факторами, который, в свою очередь, позволяет перейти к более сложным моделям дисперсионного анализа. Отметим, что дисперсионный анализ является очень сложным разделом статистики, поэтому в этой книге мы приводим лишь самые необходимые сведения, оставляя для самостоятельного изучения все специфические вопросы.

Программа SPSS предназначена для того, чтобы сделать работу исследователя как можно более простой и комфортной. Чтобы выполнить дисперсионный анализ с одним, двумя и даже семью факторами, достаточно сделать лишь нескольких щелчков на нужных кнопках. Компьютер выполнит за вас всю вычислительную работу, и вы получите результат в наглядной и удобной форме, но, к сожалению, нередко неясный по своему содержанию. Простота и удобство вычислительного процесса подчас приводят к тому, что исследователи расслабляются и забывают простую истину: успех исследования определяется его продуманностью. Если однофакторный дисперсионный анализ не представляет сложности для понимания, то анализ с двумя и более факторами требует очень внимательного отношения и определенных навыков в интерпретации результатов. Так, для интерпретации результатов трехфакторного дисперсионного анализа, как правило, необходимы немалый опыт и умение свободно обращаться со статистическими величинами. Что же касается четырехфакторного анализа, то его интерпретация чрезвычайно сложна даже для специалиста высокой квалификации.

Многофакторный дисперсионный анализ отличается от однофакторного прежде всего тем, что появляется новая *проблема взаимодействия факторов*. Решение этой проблемы принципиально не зависит от числа факторов. С увеличением числа факторов существенно возрастает лишь сложность интерпретации взаимодействий.

В этой главе выполним двух- и трехфакторный дисперсионный анализ с учетом влияния ковариаты. Мы также воспользуемся графическими средствами интерпретации средних значений и в конце главы, как всегда, проведем разбор полученных результатов.

Файлы данных для группы методов

Общая линейная модель

В главах 14–16 рассматриваются сложные варианты дисперсионного анализа (ANOVA), которые реализуются тремя командами в группе методов Общая линейная модель: ОЛМ-одномерная, ОЛМ-многомерная и ОЛМ-повторные измерения. Для примеров используются три файла, содержащие одни и те же данные, полученные в ходе одного и того же исследования, различающиеся только структурой: ex020.sav, ex021.sav и ex022.sav. Каждый из этих файлов может быть реструктурирован в два других при помощи команды Реструктурировать, как описано в разделе «Реструктурирование данных» главы 4.

Все три файла данных содержат гипотетические результаты эксперимента по изучению эффективности запоминания слов в зависимости от частоты их встречаемости и от интонации, с которой они предъявлялись (зачитывались). Ряды из 24 не связанных по смыслу слов одинаковой длины зачитывали 20 испытуемым. Сразу после предъявления испытуемых просили воспроизвести эти слова. Подсчитывали количество правильно воспроизведенных слов из начала ряда — первые 8 слов, из середины ряда и из конца ряда — 8 завершающих слов. Через 2 дня испытуемым предлагали вспомнить предъявляемые слова и снова подсчитывали количество правильно воспроизведенных слов из начала, середины и конца ряда.

Все три файла идентичны по следующим переменным:

- ▶ N — номер испытуемого.
- ▶ Номинальная переменная Инт соответствует делению испытуемых на две группы: первой (Инт = 1) все слова читались с одинаковой интонацией; второй (Инт = 2) середина ряда интонационно выделялась.
- ▶ Количественная переменная Знач отражает эмоциональную значимость предъявляемого ряда слов: ряды различались по этой переменной.

Различия между файлами:

- ▶ Файл ex020.sav включает в себя 6 переменных продуктивности воспроизведения слов из 3 частей ряда без отсрочки (Начало1, Середина1, Конец1) и с отсрочкой в 2 дня (Начало2, Середина2, Конец2). Соответственно, число строк этого файла соответствует количеству испытуемых ($N = 20$).
- ▶ Файл ex021.sav включает в себя 3 переменных продуктивности воспроизведения слов из 3 частей ряда (Начало, Середина, Конец). Дополнительно введена номинальная переменная Отсрочка, удваивающая количество строк файла: Отсрочка = 1 — воспроизведение без отсрочки, Отсрочка = 2 — воспроизведение с отсрочкой. Таким

образом, число строк файла равно 40 ($2N$), а значения переменных *N*, *инт* и *знач* повторяются дважды (для двух значений переменной *Отсрочка*).

- Файл *ex022.sav* включает в себя *1 переменную продуктивности воспроизведения слов* Слова. Для их различения по части ряда введена переменная *Ч_ряда*: 1 — начало, 2 — середина, 3 — конец. А по отсрочке воспроизведения их делит переменная *Отсрочка*: 1 — без отсрочки, 2 — с отсрочкой. Таким образом, число строк этого файла равно 120 ($N \cdot 2 \cdot 3$), а значения переменных *N*, *инт* и *знач* повторяются 6 раз (для 2 значений переменной *Отсрочка* и 3 значений переменной *Ч_ряда*).

Таким образом, данные каждого из трех файлов позволяют проверить одни и те же гипотезы о влиянии интонации и значимости слов на продуктивность их воспроизведения в зависимости от части ряда и от отсрочки. Но в зависимости от структуры данных для этого необходимо применить три разных варианта ANOVA:

- *ex022.sav* — ОЛМ-одномерный (многофакторный ANOVA), рассматриваемый в этой главе;
- *ex021.sav* — ОЛМ-многомерный (многомерный ANOVA, глава 15);
- *ex020.sav* — ОЛМ-повторные измерения (ANOVA с повторными измерениями, глава 16).

Дисперсионный анализ с двумя факторами

Как было показано в предыдущей главе, дисперсионный анализ (ANOVA) определяет статистическую достоверность различия между выборками путем сравнения их средних значений. Чтобы получить представление о двухфакторном ANOVA, сначала немного вернемся назад и обобщим наши знания об однофакторном ANOVA. Мы сравнивали три класса (переменная *класс*) по уровню выраженности переменной *тест1*. Однофакторность анализа заключалась в том, что деление на группы производилось по градациям одной независимой переменной (*класс*). Для однофакторного ANOVA в SPSS существует специальная упрощенная команда *Однофакторный дисперсионный анализ*. Команды подменю *Общая линейная модель* позволяют выполнять одно-, двух-, трехфакторный и т. д. анализ, однако обращение с ними несколько сложнее.

В этой главе мы будем использовать файл *ex022.sav*. В качестве зависимой переменной возьмем переменную *Слова*, а роль независимых будут играть переменные *Инт* и *Ч_ряда*. Попытаемся определить степень влияния переменных *Инт*, *Ч_ряда* и их взаимодействия *Инт* × *Ч_ряда* на распределение значений переменной *Слова*. Такая схема анализа может быть лаконично обозначена как ANOVA 2 × 3 (*Интонация* × *Часть ряда*). Исследование позволит получить ответы на перечисленные далее вопросы:

- Существует ли главный эффект фактора *Инт*, то есть существует ли значимое различие в продуктивности воспроизведения всего ряда из 24 слов в зависимости от интонационного выделения середины ряда и какова степень этого различия?
- Существует ли главный эффект фактора *Ч_ряда*, то есть существует ли значимое различие в продуктивности воспроизведения трех частей ряда (начала, середины и конца) и какова степень этого различия?
- Существует ли взаимодействие переменных *Инт* и *Ч_ряда*, то есть зависит ли влияние одной из этих переменных от уровней (значений, градаций) другой?

Таким образом, двухфакторный дисперсионный анализ позволяет проверить три гипотезы: две о главном эффекте и одну о взаимодействии факторов. Ответы на два первых вопроса можно было бы получить, дважды применив однофакторный дисперсионный анализ. Специфика многофакторного анализа проявляется в содержании третьего вопроса, который касается взаимодействия факторов. Взаимодействие двух факторов означает, что влияние одного из них проявляется по-разному на разных уровнях другого фактора. Весьма полезным для интерпретации взаимодействий является построение диаграмм средних значений для каждой ячейки таблицы сопряженности независимых переменных. Мы займемся построением диаграмм в конце этой главы. На диаграммах будут представлены все значения вне зависимости от степени их значимости, что позволит нам более ясно представлять, почему воздействие одних переменных оказывается значимым, а других — незначимым. Более детально проблема взаимодействия будет рассмотрена на примере при обсуждении результатов.

SPSS позволяет включить в окно вывода средние значения всех выборок, соответствующих всем возможным сочетаниям градаций факторов (в данном случае $2 \cdot 3 = 6$), а также вычислить F -величины и соответствующие p -уровни. По этим характеристикам мы сможем судить о степени влияния каждой из независимых переменных на распределение зависимой переменной и о взаимодействии независимых переменных.

Дисперсионный анализ с тремя и более факторами

Как уже отмечалось, дисперсионный анализ с тремя и более факторами принципиально не отличается от двухфакторного анализа. При количестве факторов более двух возрастает лишь сложность интерпретации взаимодействий факторов. Эта сложность обусловлена появлением большого числа взаимодействий, часть из которых с трудом поддаются интерпретации. Рассмотрим эту проблему на примере.

Предположим, изучается влияние на переменную Слова трех факторов: Инт, Ч_ряда и Отсрочка. Такая схема анализа может быть обозначена как ANOVA $2 \times 3 \times 2$ (Интонация $\times$ Часть ряда $\times$ Отсрочка). Применение трехфакторного дисперсионного анализа позволило бы получить ответы на следующие вопросы:

- ▶ Существует ли главный эффект фактора Инт, то есть существует ли значимое различие в продуктивности воспроизведения всего ряда из 24 слов в зависимости от интонационного выделения середины ряда и какова степень этого различия?
- ▶ Существует ли главный эффект фактора Ч_ряда, то есть существует ли значимое различие в продуктивности воспроизведения трех частей ряда (начала, середины и конца) и какова степень этого различия?
- ▶ Существует ли главный эффект фактора Отсрочка, то есть существует ли значимое различие в продуктивности воспроизведения всего ряда в зависимости от отсрочки?
- ▶ Существует ли взаимодействие переменных Инт и Ч_ряда, то есть зависит ли влияние одной из этих переменных от уровней (значений, градаций) другой?
- ▶ Существует ли взаимодействие переменных Инт и Отсрочка, то есть зависит ли влияние одной из этих переменных от градаций другой?
- ▶ Существует ли взаимодействие переменных Ч_ряда и Отсрочка, то есть зависит ли влияние одной из этих переменных от градаций другой?

- Существует ли взаимодействие переменных Инт, Ч_ряда и Отсрочка, то есть зависит ли взаимодействие двух из этих переменных от градаций третьей?

Таким образом, трехфакторный дисперсионный анализ предполагает проверку уже семи гипотез. Три из них касаются взаимодействия двух факторов и одна — взаимодействия трех факторов. Если добавить еще один, четвертый фактор, появится взаимодействие 4 факторов, а число проверяемых гипотез возрастет до 15. Для 5 факторов число проверяемых гипотез составит уже 31! Общее число проверяемых гипотез равно $2^P - 1$, где P — число факторов. При увеличении числа факторов быстро возрастает не только количество проверяемых гипотез, но и сложность интерпретации взаимодействий. Если интерпретация взаимодействий двух факторов обычно не составляет труда, то взаимодействия трех факторов обязательно требуют построения графиков средних значений, а интерпретация взаимодействий более высокого порядка вряд ли вообще возможна.

Влияние ковариат

Ковариаты используются для исключения влияния количественной переменной на зависимую переменную. Ковариату проще всего представить как переменную, значительно коррелирующую с зависимой переменной и позволяющую уменьшить ее дисперсию. За счет включения в анализ ковариаты дисперсия зависимой переменной уменьшается, что позволяет сделать более очевидным влияние анализируемых факторов.

В нашем исследовании в качестве ковариаты будет использоваться переменная *Знач.* Эта переменная (эмоциональная значимость предъявляемого ряда слов) в существенной степени коррелирует с продуктивностью воспроизведения этого ряда (*Слова*). Если мы хотим проследить влияние факторов *Инт* и *Ч_ряда* на зависимую переменную *Слова* без учета влияния фактора *Знач* (исключив его влияние), последний необходимо включить в анализ в качестве ковариаты. Появление ковариаты не влияет на описательные статистики, однако может изменить сумму квадратов (как правило, в меньшую сторону) и значение *F*-критерия (как в меньшую, так и, возможно, в большую сторону).

Пошаговые алгоритмы вычислений

Для проведения многофакторного дисперсионного анализа сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае — файл *ex022.sav*). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом *ex022*.

ШАГ 4

В меню *Анализ* выберите команду *Общая линейная модель* ► *ОЛМ-одномерная*. На экране появится диалоговое окно *ОЛМ-одномерная* (рис. 14.1).

Как вы, вероятно, обратили внимание, в меню *Общая линейная модель* помимо команды *ОЛМ-одномерная* находится еще три команды: *ОЛМ-многомерная*, *ОЛМ-повторные измерения* и *Компоненты дисперсии*. Команда *ОЛМ-многомерная* используется для многомер-

ного дисперсионного и ковариационного анализов и рассмотрена в главе 15. Команда ОЛМ-повторные измерения может применяться как для однофакторного, так и для многофакторного анализов, ее описание приведено в главе 16. Команда Компоненты дисперсии обладает несколько большей гибкостью по сравнению с предыдущими, однако ее сложность не позволяет рассказать о ней в рамках данной книги.

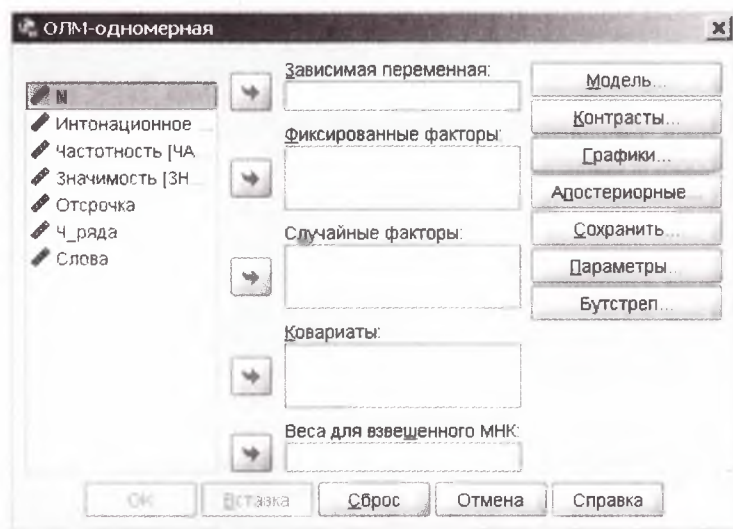


Рис. 14.1. Диалоговое окно ОЛМ-одномерная

В левой части диалогового окна ОЛМ-одномерная расположен список всех доступных переменных файла данных, а справа находятся поля и кнопки, с помощью которых задаются основные установки дисперсионного анализа. Поле Зависимая переменная предназначено для указания единственной зависимой переменной анализа, поле Фиксированные факторы — для имен номинальных независимых переменных, или факторов, поле Ковариаты — для указания имен ковариат как дополнительных количественных факторов. Кнопки Контрасты и Апостериорные действуют идентично своим аналогам из диалогового окна Однофакторный дисперсионный анализ, поэтому при необходимости вернитесь к главе 13. Кнопка Модель позволяет задать пользовательскую модель анализа — на ваш выбор.

Особое значение имеет кнопка Графики, позволяющая строить графики средних значений. Для построения графиков в программе SPSS предусмотрено отдельное меню ОЛМ-одномерная: Графики профилей, так как графики средних значений необходимы для интерпретации взаимодействий факторов.

Двухфакторный дисперсионный анализ

После задания зависимой переменной Слова необходимо задать независимые переменные, по очереди помещая их в поле Фиксированные факторы.

Когда имена всех переменных, участвующих в анализе, определены, можно перейти к заданию параметров выполняемых действий. Для этого щелкните на кнопке Параметры. На экране появится диалоговое окно ОЛМ-одномерная: Параметры (рис. 14.2). Вас будет интересовать область Вывести, позволяющая задать список включаемых в вывод

величин. Так, флажок *Описательные статистики* позволяет вывести для каждой ячейки таблицы сопряженности средние значения, стандартные отклонения и размеры выборок по всем градациям всех факторов. Нередко устанавливается флажок *Оценки размера эффекта*, включающий в выводимые данные величину эффекта (η^2) каждой из независимых переменных, а также их взаимодействия.

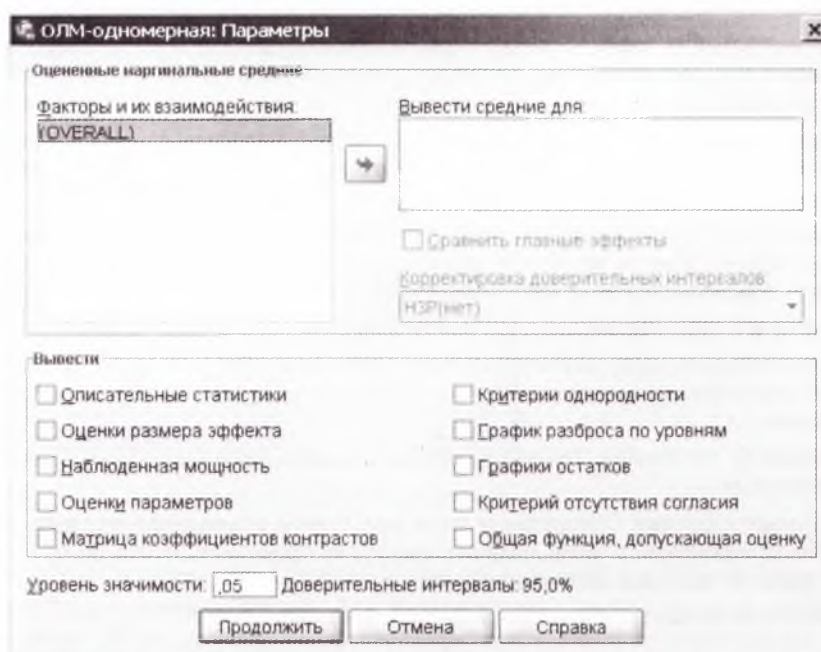


Рис. 14.2. Диалоговое окно ОЛМ-одномерная: Параметры

Следующий пример представляет собой практическую реализацию описанных действий.

ШАГ 5

После выполнения шага 4 должно быть открыто диалоговое окно ОЛМ-одномерная (см. рис. 14.1).

1. Щелкните сначала на переменной *Слова*, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле *Зависимая переменная*.
2. Щелкните сначала на переменной *Интонационное выделение [Инт]*, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле *Фиксированные факторы*.
3. Повторите предыдущее действие для переменной *Часть ряда [Ч_ряда]*.
4. Щелкните на кнопке *Параметры*, чтобы открыть диалоговое окно ОЛМ-одномерная: Параметры.
5. Установите флажки *Описательные статистики*, *Оценки размера эффекта* и *Критерии однородности*, а затем щелкните на кнопке *Продолжить*, чтобы вернуться в диалоговое окно ОЛМ-одномерная.
6. Щелкните на кнопке *ОК*, чтобы открыть окно вывода.

Влияние ковариаты

В следующем пошаговом алгоритме мы реализуем двухфакторный дисперсионный анализ с учетом ковариаты. Все переменные останутся теми же, что и на шаге 5, но дополнительно в качестве ковариаты включим переменную *Знач.* Это позволит сравнить результаты дисперсионного анализа с учетом и без учета влияния ковариаты.

ШАГ 5a

После выполнения шага 4 должно быть открыто диалоговое окно ОЛМ-одномерная (см. рис. 14.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной *Слова*, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.
2. Щелкните сначала на переменной *Интонационное выделение [Инт]*, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Фиксированные факторы.
3. Повторите предыдущее действие для переменной *Часть ряда [Ч_ряда]*.
4. Щелкните сначала на переменной *Знач*, чтобы выделить ее, а затем на второй снизу кнопке со стрелкой, чтобы переместить переменную в поле Ковариаты.
5. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-одномерная.
6. Установите флажки *Описательные статистики*, *Оценки размера эффекта* и *Критерии однородности*, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-одномерная.
7. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Результаты выполнения шага 5a будут отличаться от результатов выполнения шага 5 только учетом ковариаты *Знач*.

Графические средства интерпретации взаимодействий

В большинстве случаев для интерпретации взаимодействия факторов необходимо построение графиков средних значений. Для этого служит кнопка *Графики*, открывающая диалоговое окно ОЛМ-одномерная: *Графики профилей* (рис. 14.3). В этом окне можно задать параметры сразу нескольких графиков средних значений. Каждый график представляет собой ломаную линию (профиль), соединяющую точки — средние значения для групп. Горизонтальной оси соответствуют уровни одного из факторов, а средние значения откладываются по вертикальной оси.

Вид графиков задается тремя параметрами. В поле *Горизонтальная ось* вводится имя фактора, градациям которого будет соответствовать горизонтальная ось графика или графиков. В поле *Отдельные линии* указывается фактор, градациям которого будут соответствовать разные линии на графике. Поле *Отдельные графики* служит для задания имени фактора, градациям которого будут соответствовать разные графики. Щелчок на кнопке *Добавить* подтверждает правильность введенных параметров и позволяет перейти к заданию графиков другого типа.

При проведении двухфакторного дисперсионного анализа, как в нашем примере, достаточно построить один график. Для этого надо сначала задать имя фактора для ок

Горизонтальная ось. Обычно это фактор, который имеет больше градаций, — в нашем случае Ч_ряда. Затем необходимо задать имя фактора для окна Отдельные линии — в нашем случае это Инт. В результате будет построен график с двумя ломаными линиями, каждая из которых соединяет три точки. Вертикальная ось этого графика будет соответствовать средним значениям, горизонтальная ось — переменной Ч_ряда, а две отдельные линии — переменной Инт.

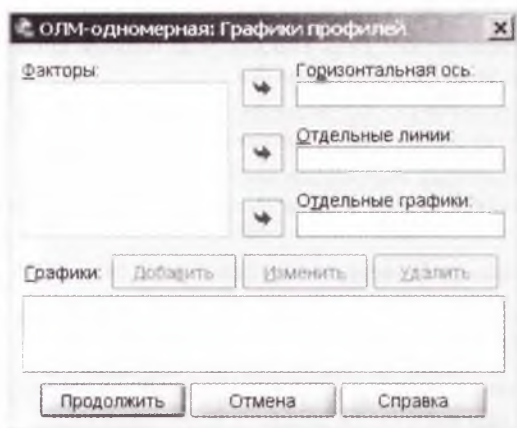


Рис. 14.3. Диалоговое окно ОЛМ-одномерная: Графики профилей

При трехфакторном анализе подобный график позволит интерпретировать только двухфакторное взаимодействие. Для интерпретации трехфакторного взаимодействия необходимо градациям третьего фактора поставить в соответствие разные Отдельные линии. Например, если бы третьим фактором в нашем примере был фактор Отсрочка, то задание этого фактора как имени для разных графиков привело бы к построению двух графиков — по одному для двух условий (есть, нет), каждый из которых соответствовал бы двухфакторному взаимодействию переменных Инт и Ч_ряда.

Пошаговый алгоритм, приведенный далее, повторяет действия предшествующего шага, но добавляет в окно вывода график средних значений, позволяющий интерпретировать взаимодействие факторов Инт и Ч_ряда. А шаг 6 демонстрирует построение альтернативного графика: столбиковой диаграммы средних значений.

ШАГ 56

После выполнения шага 4 должно быть открыто диалоговое окно ОЛМ-одномерная (см. рис. 14.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Задайте переменные для анализа. Для этого щелкните на переменной Слова, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная; щелкните на переменной Инт, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Фиксированные факторы. Повторите то же действие для переменной Ч_ряда. Щелкните на переменной Знач, чтобы выделить ее, а затем на второй снизу кнопке со стрелкой, чтобы переместить переменную в поле Ковариаты.

- 
2. Задайте параметры результатов. Для этого щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-одномерная: Параметры, и установите флажки Описательные статистики, Критерии однородности, Оценки размера эффекта, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-одномерная.
 3. Щелкните на кнопке Графики, чтобы открыть диалоговое окно ОЛМ-одномерная: Графики профилей (см. рис. 14.3).
 4. Выделите в списке Факторы переменную Ч_ряда и щелчком на верхней кнопке со стрелкой перенесите ее в поле Горизонтальная ось.
 5. Выделите в списке Факторы переменную Инт и щелчком на второй сверху кнопке со стрелкой перенесите ее в поле Отдельные линии.
 6. Подтвердите правильность введенных параметров щелчком на кнопке Добавить — в нижнем поле появится строка Ч_ряда*Инт, обозначающая тип графика средних.
 7. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-одномерная.
 8. Щелкните на кнопке ОК, чтобы открыть окно вывода.

При желании можно построить альтернативные графики, например столбиковые, воспользовавшись меню Графика в командной строке.

ШАГ 6

На экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex022.

- 
1. В командной строке выберите Графика ► Устаревшие диалоговые окна ► Столбики.
 2. В открывшемся окне Столбики (см. рис. 5.2) выделите значок Кластеризованные и щелкните на кнопке Задать.
 3. В открывшемся окне Кластеризованные столбики: Итожащие функции по группам наблюдений установите переключатель в положение Другую статистику (например, среднее) и в поле Переменное перенесите из левого поля переменную Слова. В поле Категориальная ось перенесите переменную Интонация, а в поле Задать кластеры по — переменную Ч_ряда.
 4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шагов 5, 5а, 5б и 6 программа автоматически перейдет в окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Двухфакторный дисперсионный анализ

Фрагменты выводимых данных, сгенерированных программой после выполнения шага 5, приведены на рис. 14.4.

- 
2. Задайте параметры результатов. Для этого щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-одномерная: Параметры, и установите флажки Описательные статистики, Критерии однородности, Оценки размера эффекта, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-одномерная.
 3. Щелкните на кнопке Графики, чтобы открыть диалоговое окно ОЛМ-одномерная: Графики профилей (см. рис. 14.3).
 4. Выделите в списке Факторы переменную Ч_ряда и щелчком на верхней кнопке со стрелкой перенесите ее в поле Горизонтальная ось.
 5. Выделите в списке Факторы переменную Инт и щелчком на второй сверху кнопке со стрелкой перенесите ее в поле Отдельные линии.
 6. Подтвердите правильность введенных параметров щелчком на кнопке Добавить — в нижнем поле появится строка Ч_ряда*Инт, обозначающая тип графика средних.
 7. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-одномерная.
 8. Щелкните на кнопке ОК, чтобы открыть окно вывода.

При желании можно построить альтернативные графики, например столбиковые, воспользовавшись меню Графика в командной строке.

ШАГ 6

На экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex022.

- 
1. В командной строке выберите Графика ► Устаревшие диалоговые окна ► Столбики.
 2. В открывшемся окне Столбики (см. рис. 5.2) выделите значок Кластеризованные и щелкните на кнопке Задать.
 3. В открывшемся окне Кластеризованные столбики: Итожащие функции по группам наблюдений установите переключатель в положение Другую статистику (например, среднее) и в поле Переменное перенесите из левого поля переменную Слова. В поле Категориальная ось перенесите переменную Интонация, а в поле Задать кластеры по — переменную Ч_ряда.
 4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шагов 5, 5а, 5б и 6 программа автоматически перейдет в окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Двухфакторный дисперсионный анализ

Фрагменты выводимых данных, сгенерированных программой после выполнения шага 5, приведены на рис. 14.4.

Описательные статистики

Зависимая переменная: Слова

Интонационное выделение	Часть ряда	Среднее	Стд. Отклонение	N
нет	Начало	3,50	1,573	20
	Середина	3,10	1,021	20
	Конец	3,75	1,585	20
	Всего	3,45	1,419	60
есть	Начало	2,65	1,424	20
	Середина	4,30	1,261	20
	Конец	2,80	,834	20
	Всего	3,25	1,398	60
Всего	Начало	3,08	1,542	40
	Середина	3,70	1,285	40
	Конец	3,28	1,339	40
	Всего	3,35	1,406	120

Оценка эффектов межгрупповых факторов

Зависимая переменная: Слова

Источник	Сумма квадратов типа III	ст. св.	Средний квадрат	F	Знач.	Частная Эта в Квадрате
Скорректированная модель	39,800 ^a	5	7,760	4,502	,001	,165
Свободный член	1346,700	1	1346,700	781,292	,000	,873
ИНТ	1,200	1	1,200	,696	,406	,006
Ч_ряда	8,150	2	4,075	2,364	,099	,040
ИНТ * Ч_ряда	29,450	2	14,725	8,543	,000	,130
Ошибка	196,500	114	1,724			
Всего	1582,000	120				
Скорректированный итог	235,300	119				

^a R квадрат = ,165 (Скорректированный R квадрат = ,128)

Рис. 14.4. Фрагменты окна вывода после выполнения шага 5

В таблице Описательные статистики содержатся характеристики всех выборок, образованных по каждой из независимых переменных в отдельности (2 выборки по переменной Инт и 3 выборки по переменной Ч_ряда), а также при пересечении градаций независимых переменных ($3 \cdot 2 = 6$ выборок). Таблица Оценка эффектов межгрупповых факторов содержит результаты проверки трех основных гипотез двухфакторного дисперсионного анализа:

- Переменная Ч_ряда не оказывает статистически достоверного влияния на распределение зависимой переменной Слова (средние значения для начала, середины и конца ряда составили, соответственно, 3,08, 3,70 и 3,28, $F = 2,364$, $p = 0,099$).
- Переменная Инт не оказывает статистически значимого влияния на распределение зависимой переменной Слова (средние значения для групп «нет» и «есть» составили, соответственно, 3,45 и 3,25, $F = 0,696$, $p = 0,406$).
- Обнаружено статистически достоверное взаимодействие на высоком уровне статистической значимости между независимыми переменными Ч_ряда и Инт ($F = 8,543$, $p < 0,001$).

Взаимодействие независимых переменных более подробно обсуждается в последнем разделе этой главы, посвященном использованию графиков для интерпретации взаимодействий.

Влияние ковариаты

Фрагмент выводимых данных, сгенерированных программой после выполнения шага 5а, приведен на рис. 14.5.

Оценка эффектов межгрупповых факторов

Зависимая переменная Слова

Источник	Сумма квадратов типа III	ст.св	Средний квадрат	F	Знач.	Частная Эта в Квадрате
Скорректированная модель	90,862 ^a	6	15,144	11,848	,000	,386
Свободный член	25,047	1	25,047	19,595	,000	,148
ЗНАЧ	52,062	1	52,062	40,730	,000	,265
ИНТ	5,662	1	5,662	4,429	,038	,038
Ч_ряда	8,150	2	4,075	3,188	,045	,053
ИНТ * Ч_ряда	29,450	2	14,725	11,520	,000	,169
Ошибка	144,438	113	1,278			
Всего	1582,000	120				
Скорректированный итог	235,300	119				

^a R квадрат = ,386 (Скорректированный R квадрат = ,354)

Рис. 14.5. Фрагмент окна вывода после выполнения шага 5а

Включение ковариаты ЗНАЧ не меняет описательные статистики, но вносит некоторые изменения в результаты проверки гипотез дисперсионного анализа, как видно из таблицы Оценка эффектов межгрупповых факторов.

Ковариата оказывает влияние на суммы квадратов и, как следствие, на число степеней свободы (в некоторых случаях), средние квадраты, величины F и частной эта в квадрате (η^2), а также значения p -уровней значимости. Если это влияние существенно, то, как правило, значения среднего квадрата и F -критерия увеличиваются, а соответствующие значения p -уровней уменьшаются.

Обратите внимание на то, что ковариата ЗНАЧ оказывает значительное влияние на разброс зависимой переменной Слова: значение η^2 составляет 0,265, то есть 26,5 % дисперсии переменной Слова обусловлено влиянием ковариаты. Дисперсия скорректированной модели представляет собой сумму всех сумм квадратов дисперсий, обусловленных влияниями независимых переменных и их взаимодействий.

Сравните между собой результаты дисперсионного анализа с ковариатой и без ковариаты. Благодаря тому, что ковариата «съемает» часть дисперсии зависимой переменной, для главных эффектов и взаимодействия факторов F -критерий увеличивается, а p -уровень уменьшается. В результате главные эффекты факторов становятся статистически достоверными. Разумеется, подобный результат проявляется не всегда: ковариата способна не только повысить, но и понизить статистическую значимость эффектов факторов и их взаимодействий — заранее это непредсказуемо.

Таким образом, двухфакторный дисперсионный анализ с зависимой переменной Слова, независимыми переменными Инт и Ч_ряда и ковариатой ЗНАЧ дал следующие результаты:

- ▶ Ковариата ЗНАЧ оказывает статистически достоверное влияние на зависимую переменную Слова ($F = 40,73$, $p < 0,001$).
- ▶ Переменная Инт оказывает статистически достоверное влияние на распределение зависимой переменной Слова ($F = 4,429$, $p = 0,038$).

- Переменная Ч_ряда оказывает статистически значимое влияние на распределение зависимой переменной Слова ($F = 3,188, p = 0,045$).
- Обнаружено статистически достоверное взаимодействие между независимыми переменными Ч_ряда и Инт ($F = 11,52, p < 0,001$).

Отметим, что не очень удобно интерпретировать результаты по средним значениям из таблицы Описательные статистики. Для этого лучше использовать графики средних значений, полученные после выполнения шага 5б или 6.

Использование графиков для интерпретации взаимодействий

График средних значений, который генерируется программой после выполнения шага 5б, приведен на рис. 14.6.

График позволяет без труда интерпретировать взаимодействие факторов: влияние фактора Инт на переменную отметка2 проявляется по-разному при разных градациях переменной Ч_ряда: в случае наличия интонационного выделения (Инт: есть) воспроизведение слов из середины ряда существенно более продуктивно, чем при его отсутствии (Инт: нет).

Вообще говоря, взаимодействие факторов на графике выглядит как заметное различие формы профилей соответствующих линий на рис. 14.6 или соответствующих кластеров столбиков на рис. 14.7. Чем ближе по форме профили линий (кластеров), тем меньше взаимодействие факторов. Но интерпретации подлежат только те графики, которые соответствуют статистически достоверному взаимодействию факторов.

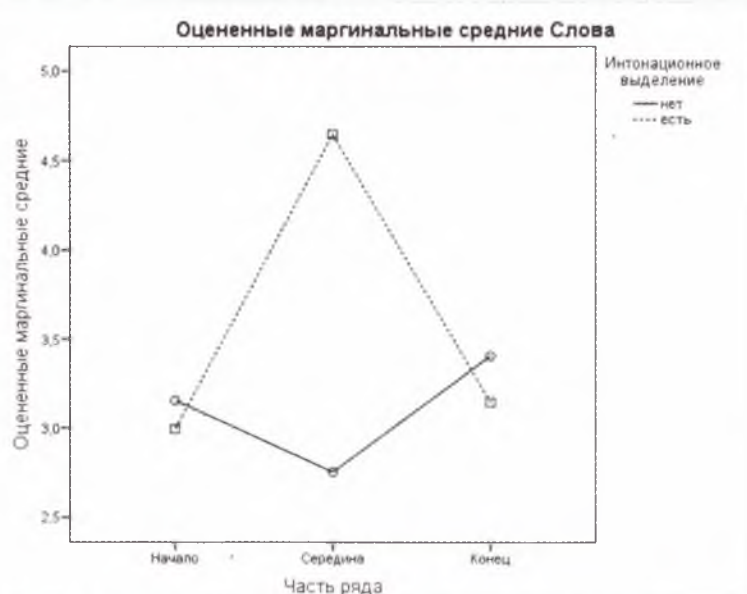


Рис. 14.6. График средних значений, генерируемый после выполнения шага 5б

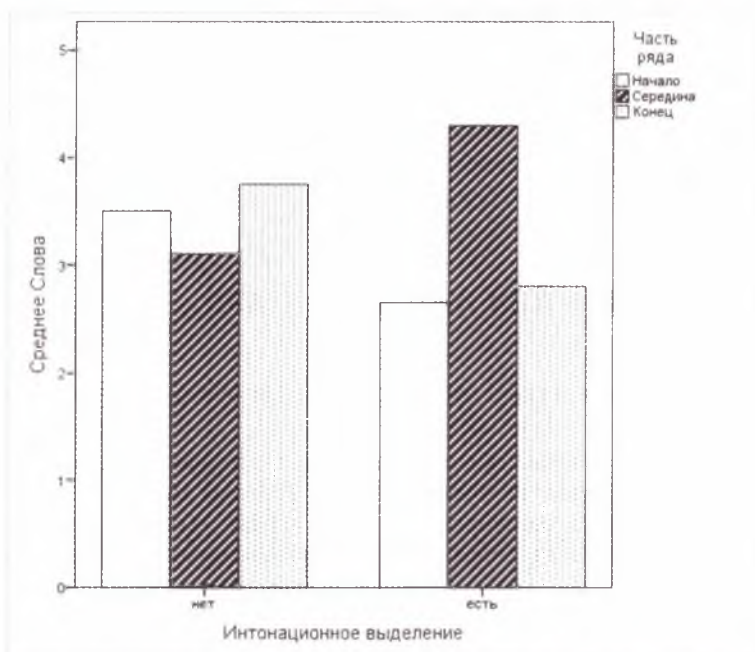


Рис. 14.7. График средних значений, генерируемый после выполнения шага 6

Если бы проводился трехфакторный анализ, то интерпретация трехфакторного взаимодействия требовала бы построения нескольких графиков. Так, при добавлении фактора Отсрочка потребовалось бы построение трех графиков двухфакторного взаимодействия: по одному для каждой градации фактора Отсрочка. Попробуйте самостоятельно провести трехфакторный дисперсионный анализ, добавив при выполнении инструкций шага 5б переменную Отсрочка в качестве третьего фактора и при задании параметров графиков средних значений в диалоговом окне указав эту переменную в поле Отдельные графики.

Терминология, используемая при выводе

Далее дана трактовка новых терминов, используемых программой в окне вывода (трактовка остальных терминов ANOVA приведена в конце главы 13):

- Частная Эта в квадрате — оценка величины эффекта (η^2), по смыслу совпадает с RSQ или R^2 . Под эффектом понимается вклад независимой переменной или взаимодействия переменных в разброс значений зависимой переменной (доля от полной ее дисперсии, принимаемой за 1).

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его

редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы в меню Файл выберите команду Выход.

15 Многомерный дисперсионный анализ

195	Пошаговые алгоритмы вычислений
202	Представление результатов
207	Завершение анализа и выход из программы

В этой главе мы рассмотрим методы обработки данных, которые содержат несколько зависимых переменных: многомерный дисперсионный анализ (Multivariate Analysis Of Variances, MANOVA) и многомерный ковариационный анализ (Multivariate Analysis Of Covariance, MANCOVA). Команды многомерного анализа, входящие в подменю Общая линейная модель, относятся к наиболее сложным командам в SPSS. Помимо команды ОЛМ-многомерная, предназначенной для проведения анализов MANOVA и MANCOVA, к командам многомерного анализа относится и команда ОЛМ-повторные измерения, позволяющая провести анализ MANOVA с повторными измерениями. Первый тип многомерного анализа будет рассмотрен в данной главе, второй — в главе 16.

Ввиду значительной сложности мы упомянем лишь наиболее понятные и широко используемые параметры команды ОЛМ-многомерная. Фактически, критерии MANOVA и MANCOVA являются расширением дисперсионного анализа (ANOVA), рассмотренного нами в главах 13 и 14. Если вы не знакомы с дисперсионным анализом, рекомендуем, прежде чем читать дальше, обратиться к указанным главам.

Как упоминалось в предыдущих главах, t -критерий для двух выборок позволяет выяснить, существуют ли различия между двумя средними значениями для этих выборок. Эту простейшую ситуацию (единственная независимая переменная с двумя градациями и одна зависимая переменная метрического типа) можно последовательно усложнить тремя способами:

- ▶ ввести в рассмотрение независимую переменную с более чем двумя градациями — в такой ситуации применяется однофакторный дисперсионный анализ;
- ▶ ввести не одну, а несколько независимых переменных — для этого предназначен многофакторный дисперсионный анализ;
- ▶ задействовать ковариаты.

Во всех трех случаях зависимая переменная остается единственной и имеет метрический тип. Тем не менее существуют задачи, в которых требуется учитывать не одну, а несколько зависимых переменных. В этой главе мы займемся рассмотрением проблемы проведения анализа с участием более чем одной зависимой переменной, при этом не станем усложнять требования к независимым переменным.

В этой главе для примера используется файл данных ex021.sav. Подробное описание переменных файла приводится в разделе «Файлы данных для группы методов Общая линейная модель» главы 14. В данном файле, как и в двух других (ex020.sav и ex022.sav), содержатся данные исследования влияния интонационного выделения средней части

ряда из 24 предъявляемых слов на продуктивность их воспроизведения в зависимости от части ряда, эмоциональной значимости слов и отсрочки воспроизведения. В файле ex021.sav зависимыми переменными являются: количество воспроизведенных слов в начале ряда (НАЧАЛО), в середине ряда (СЕРЕДИНА) и в конце ряда (КОНЕЦ). Независимые переменные (факторы): Инт (интонационное выделение: 0 — нет, 1 — есть); Отсрочка (0 — нет, 1 — есть); Значимость (эмоциональная значимость ряда слов, количественная переменная).

Представим, что необходимо сравнить продуктивность воспроизведения при интонационном выделении и без него (переменная Отсрочка) одновременно по всем трем показателям (переменные НАЧАЛО, СЕРЕДИНА, КОНЕЦ). В подобной ситуации одним из возможных подходов является троекратное применение t -критерия или однофакторного дисперсионного анализа (эти методы эквивалентны, поскольку $t^2 = F$). Очевидными достоинствами такого решения являются простота и ясность, однако нельзя не заметить и двух недостатков: во-первых, при неоднократном применении статистического критерия (в данном случае троекратном) увеличивается вероятность ошибки, то есть вероятность случайности общего результата исследования; во-вторых, если между зависимыми переменными имеется некоторая корреляция (а в рассматриваемом случае она есть), то результат, полученный в отношении каждой из переменных в отдельности, не способен отразить этот важный факт.

Описанные недостатки привели к усовершенствованию как t -критерия, так и дисперсионного анализа: первый был расширен с помощью критерия Хотеллинга (Hotelling), а вместо второго стал использоваться многомерный дисперсионный анализ (MANOVA). Оба типа анализа реализуются командой ОЛМ-многомерная. Кроме того, эта команда позволяет проводить многомерный ковариационный анализ (MANCOVA), учитывающий влияние ковариат. Особенностью всех типов многомерного анализа является то, что они обрабатывают все зависимые переменные одновременно. В примерах этой главы показано, каким образом исследовать структуру изменений зависимых переменных путем применения серий одномерных F -критериев или серий множественных сравнений постфактум.

Применяя MANOVA в отношении множества зависимых переменных, следует помнить, что линейная функциональная связь между ними недопустима. Иными словами, следует избегать применения анализа MANOVA к тем зависимым переменным, корреляция между которыми близка к 1.

Как и в одномерном дисперсионном анализе, в MANOVA для определения значимости различий между группами используется F -критерий. Отметим, что многомерный дисперсионный анализ, как и одномерный, позволяет оценить влияние не только отдельных независимых переменных (главных эффектов), но и их взаимодействий. Поскольку мы рассматриваем наличие нескольких зависимых переменных, F -статистика носит многомерный характер и для ее формирования используется матричная алгебра.

Для многомерного анализа необходимо иметь как минимум две зависимые переменные (иначе анализ не является многомерным) и как минимум одну независимую переменную. Теоретически количество зависимых и независимых переменных не ограничено, однако на практике объем выборки диктует необходимость существенного ограничения их числа. Использование ковариат, разумеется, не обязательно.

Пошаговые алгоритмы вычислений

Для проведения многомерного дисперсионного анализа сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, за-

пустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex021.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения – в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex021.sav.

ШАГ 4 В меню Анализ выберите команду **Общая линейная модель** ► **ОЛМ-многомерная**. Откроется диалоговое окно ОЛМ-многомерная (рис. 15.1).

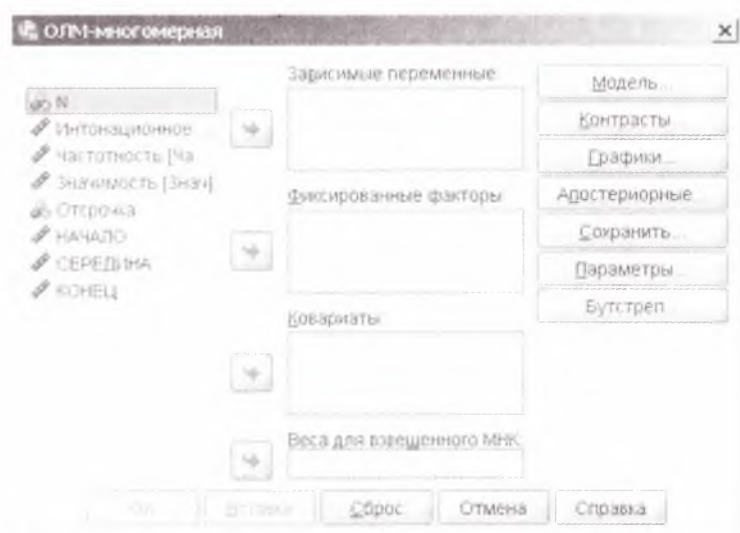


Рис. 15.1. Диалоговое окно ОЛМ-многомерная

Окно ОЛМ-многомерная позволяет задавать переменные для многомерного анализа, а также управлять диалоговыми окнами, предназначенными для определения параметров анализа. Те переменные, которые будут использоваться в качестве зависимых, необходимо поместить в поле Зависимые переменные, а независимые переменные – в поле Фиксированные факторы.

Как и прежде, для перемещения переменных из исходного списка в целевой используется одна из кнопок со стрелками. Мы будем использовать в примерах анализа 3 зависимые и 2 независимые переменные из файла ex021.sav. При желании в анализ могут быть введены одна или несколько ковариат с помощью списка Ковариаты.

В правой части диалогового окна расположены 7 кнопок, предназначенных для настройки параметров анализа. Использование кнопок Контрасты, Сохранить и Бутстреп не рассматривается, поскольку настройка соответствующих параметров требует от исследователя очень высокой квалификации.

В большинстве случаев требуется многомерный анализ как для каждой из независимых переменных, так и для всех вариантов их взаимодействий (полнофакторная модель). Полнофакторная модель является моделью анализа, используемой по умолчанию. В то же время иногда полезно исключить из рассмотрения часть взаимодействий или от-

дельные факторы. Для управления моделью анализа используется кнопка Модель. При щелчке на ней открывается диалоговое окно ОЛМ-многомерная: Модель (рис. 15.2).

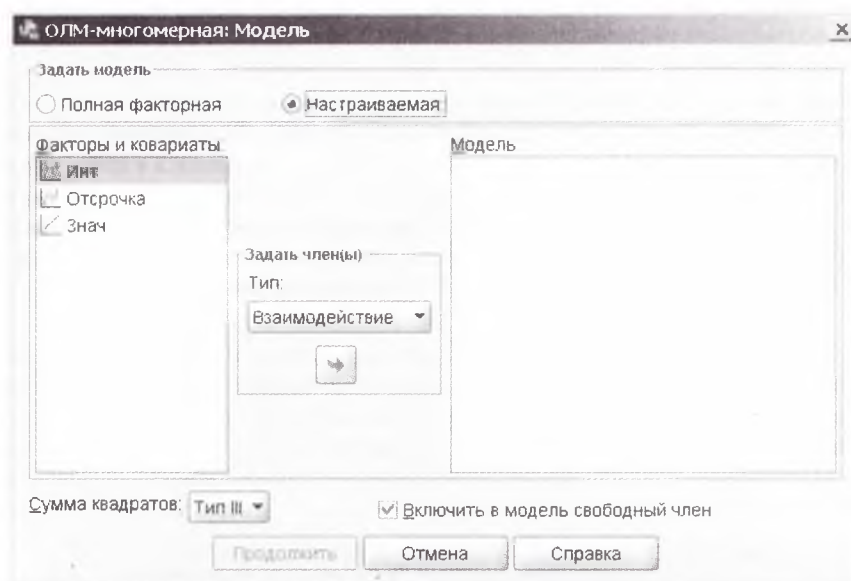


Рис. 15.2. Диалоговое окно ОЛМ-многомерная: Модель

По умолчанию переключатель Задать модель установлен в положение Полная факторная, однако вы можете установить его в положение Настраиваемая, а затем с помощью кнопки с направленной вправо стрелкой переместить в поле Модель только те переменные из списка Факторы и ковариаты, которые хотите использовать. При этом в раскрывающемся списке Тип (область Задать член(ы)) для каждого сочетания факторов вы можете указать тип проверяемых гипотез (Взаимодействие, Главные эффекты, Все 2-факторные и т. д.).

С помощью раскрывающегося списка Сумма квадратов можно также задать вариант расчета суммы квадратов. По умолчанию выбран вариант Тип III, подходящий для большинства ситуаций. Если исходные данные содержат пропущенные значения, иногда выбирают пункт Тип IV.

При интерпретации результатов многомерного анализа зачастую удобно иметь перед глазами графическое представление средних значений зависимых переменных, определяемых различными комбинациями градаций независимых переменных (факторов). Для построения подобных графиков используется кнопка Графики, назначение которой идентично назначению одноименной кнопки в окне ОЛМ-одномерная (см. главу 14). При щелчке на ней открывается диалоговое окно ОЛМ-многомерная: Графики профилей, аналогичное изображенному на рис. 14.3 (глава 14). Данное окно позволяет определить, каким образом отображать тот или иной фактор. Для каждой зависимой переменной в выводимые данные включаются отдельные графики. Обратите внимание на то, что при выполнении анализа с участием ковариат вместо фактических данных отображаются приближения средних значений, учитывающие наличие ковариат. В подобных случаях может быть полезным получение изображений без использования ковариат.

Вид графиков задается тремя параметрами. В поле Горизонтальная ось вводится имя фактора, градациям которого будет соответствовать горизонтальная ось графика или

графиков. В поле Отдельные линии указывается фактор, градациям которого будут соответствовать разные линии на графике. Поле Отдельные графики служит для задания имени фактора, градациям которого будут соответствовать разные графики. Щелчок на кнопке Добавить подтверждает правильность введенных параметров и позволяет перейти к заданию графиков другого типа.

Графическая интерпретация весьма полезна для оценки влияния независимых переменных и их взаимодействий. Для более детального анализа требуются множественные (парные) сравнения постфактум. Множественные сравнения проводятся в случае установления статистически достоверного влияния независимой переменной и позволяют определять, между какими именно градациями независимой переменной имеются различия. Подобные сравнения проводятся для каждой из зависимых переменных и каждого выбранного фактора. Множественные сравнения неприменимы при использовании ковариат.

Для проведения парных сравнений применяется кнопка Апостериорные. При щелчке на ней открывается окно ОЛМ-многомерный: Апостериорные множественные сравнения для наблюдаемых средних (рис. 15.3). Это окно идентично диалоговому окну Однофакторный дисперсионный анализ: Апостериорные множественные сравнения, которое подробно описано в главе 13. Отличается окно для многомерного анализа только наличием списков Фактор(ы) и Апостериорные критерии для.

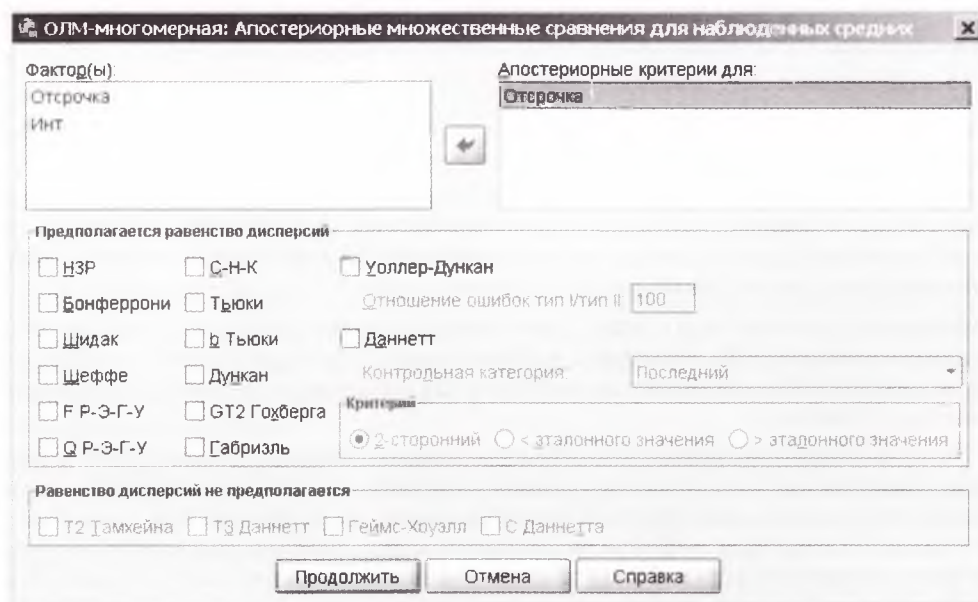


Рис. 15.3. Диалоговое окно ОЛМ-многомерная: Апостериорные множественные сравнения для наблюдаемых средних

В список Апостериорные критерии для следует включить факторы, для которых будут проводиться сравнения. Все доступные факторы перечислены в списке Фактор(ы). За описанием наиболее популярных критериев парных сравнений и результатов их применения обратитесь к главе 13.

При щелчке на кнопке Параметры открывается диалоговое окно ОЛМ-многомерная: Параметры (рис. 15.4). По умолчанию все флажки в этом окне сняты.

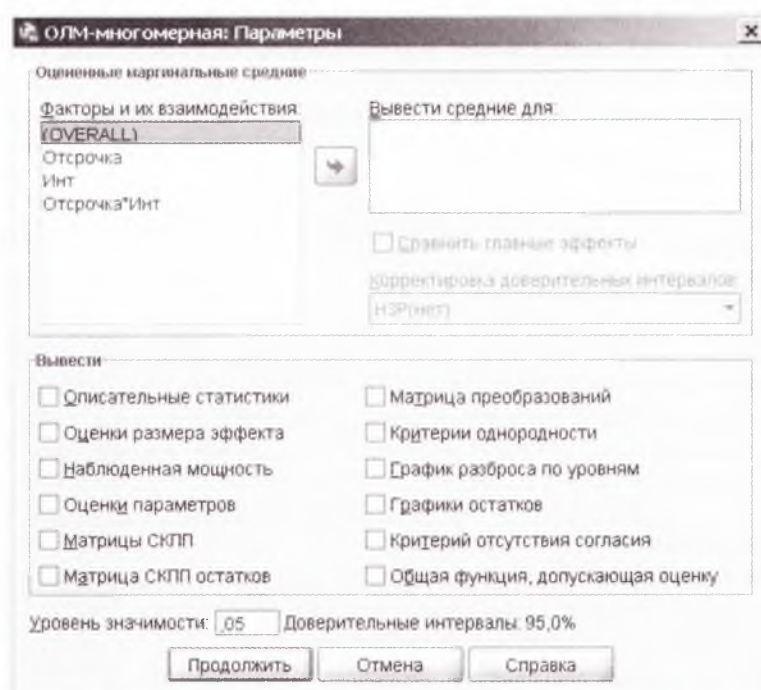


Рис. 15.4. Диалоговое окно ОЛМ-многомерная: Параметры

Если выделить нужные пункты в списке Факторы и их взаимодействия и переместить в поле Вывести средние для, то в выводимые данные будут включены средние значения каждой из зависимых переменных для всех уровней указанных факторов. Если анализ проводится без ковариат, включаемые средние значения будут фактическими средними значениями данных; в противном случае средние значения будут подсчитаны с учетом влияния ковариат. Если установить флажок Сравнить главные эффекты, может быть проведена серия апостериорных парных сравнений средних значений ячеек для каждого из факторов: при помощи раскрывающегося списка Корректировка доверительных интервалов вы можете задать либо критерий Бонферрони, либо более консервативный критерий Шидака. Другие критерии задаются при помощи описанного ранее диалогового окна ОЛМ-многомерный: Апостериорные множественные сравнения для наблюдаемых средних при условии, что анализ проводится без ковариат.

Мы опустим описания большей части флажков и упомянем лишь те, которые используются наиболее часто:

- ▶ Описательные статистики — вычисляются средние значения и стандартные отклонения каждой зависимой переменной для каждой ячейки.
- ▶ Оценки размера эффекта — вычисляется значение η^2 , характеризующее величину воздействия фактора на зависимую переменную и показывающее, какая доля общей дисперсии обусловлена данным фактором.

- Оценки параметров — вычисляются приближения и соответствующие значимости для всех факторов и ковариат модели. Это особенно удобно в случаях, когда в модель включена хотя бы одна ковариата.
- Критерии однородности — дисперсии для всех ячеек проверяются на равенство (однородность).

Поле Уровень значимости позволяет задать величину α для анализа. По умолчанию задан критический уровень значимости 0.05. Если вы измените это значение, SPSS пересчитает соответствующие доверительные интервалы.

В рассматриваемом далее примере проводится многомерный дисперсионный анализ (MANOVA) с тремя зависимыми переменными НАЧАЛО, СЕРЕДИНА, КОНЕЦ и двумя независимыми переменными (факторами) Инт и Отсрочка.

ШАГ 5

После выполнения предыдущего шага у вас должно быть открыто диалоговое окно ОЛМ-многомерная (см. рис. 15.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной НАЧАЛО, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной КОНЕЦ. В результате окажутся выделенными переменные НАЧАЛО, СЕРЕДИНА, КОНЕЦ.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Зависимые переменные.
3. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Фиксированные факторы.
4. Повторите предыдущее действие для переменной Отсрочка.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В следующем примере мы включим в анализ ковариату Знач.

ШАГ 5а

После выполнения шага 4 у вас должно быть открыто диалоговое окно ОЛМ-многомерная (см. рис. 15.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной НАЧАЛО, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной КОНЕЦ. В результате окажутся выделенными переменные НАЧАЛО, СЕРЕДИНА, КОНЕЦ.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Зависимые переменные.
3. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Фиксированные факторы.
4. Повторите предыдущее действие для переменной Отсрочка.
5. Щелкните сначала на переменной Знач, чтобы выделить ее, а затем на третьей сверху кнопке со стрелкой, чтобы переместить переменную в поле Ковариаты.
6. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В следующем примере демонстрируется использование нескольких описанных ранее параметров, генерирующих полезную статистическую и графическую информацию. Об-

ратите внимание на то, что мы не включаем в данный анализ ковариату. Напоминаем, что для анализа с ковариатами парные сравнения неприменимы, поскольку все графики будут соответствовать не фактическим исходным данным, а данным, скорректированным влияниями ковариат. Кроме того, парные сравнения имеют смысл лишь для факторов, имеющих более двух уровней, но этой причине в данном примере парные сравнения не применяются. Если же у вас какой-либо фактор имеет более двух уровней, воспользуйтесь кнопкой Апостериорные. Она действует практически идентично своему аналогу из диалогового окна Однофакторный дисперсионный анализ, поэтому при необходимости обратитесь к главе 13. Различие лишь в том, что в данном случае следует указать каждый фактор, для которого необходимы парные сравнения.

ШАГ 5Б

После выполнения шага 4 у вас должно быть открыто диалоговое окно ОЛМ-многомерная (см. рис. 15.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной НАЧАЛО, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной КОНЕЦ. В результате окажутся выделенными переменные НАЧАЛО, СЕРЕДИНА, КОНЕЦ.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Зависимые переменные.
3. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Фиксированные факторы.
4. Повторите предыдущее действие для переменной Отсрочка.
5. Щелкните на кнопке Графики, чтобы открыть диалоговое окно ОЛМ-многомерная: Графики профилей (см. рис. 14.3).
6. Щелкните сначала на переменной Отсрочка, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Горизонтальная ось.
7. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Отдельные линии.
8. Щелкните сначала на кнопке Добавить, чтобы зафиксировать заданные параметры построения диаграмм, а затем на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-многомерная.
9. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-многомерная: Параметры (см. рис. 15.4). В левом окне будут перечислены факторы и их взаимодействие (Отсрочка*Инт).
10. Щелкните сначала в левом окне на строке Отсрочка*Инт, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить ее в поле Вывести средние для.
11. Установите флажки Описательные статистики, Оценки размера эффекта, Критерии однородности. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-многомерная.
12. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 (5а, 5б) программа автоматически активизирует окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикаль-

ной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

На представленных далее рисунках приведены наиболее важные фрагменты выводимых результатов, генерируемых программой при выполнении шагов 5 и 5б. Мы не приводим эти данные целиком, поскольку их объем составляет 7 страниц.

Межгрупповые факторы

В таблице, представленной на рис. 15.5, перечислены все независимые переменные со своими уровнями и объемами выборок (N).

Межгрупповые факторы			
		Метка значения	N
Интонационное выделение	1	нет	20
	2	есть	20
Отсрочка	1	нет	20
	2	есть	20

Рис. 15.5. Фрагмент окна вывода после выполнения шага 5 (Межгрупповые факторы)

Описательные статистики

Таблица описательных статистик, полученная в результате выполнения шага 5 (на рис. 15.6 представлен только ее фрагмент), содержит средние значения, стандартные отклонения и объемы выборок для всех элементов модели.

Описательные статистики					
	Интонационное выделение	Отсрочка	Среднее	Стд. Отклонение	N
НАЧАЛО	нет	нет	4,20	1,476	10
		есть	2,80	1,398	10
		Всего	3,50	1,573	20
	есть	нет	3,30	1,494	10
		есть	2,00	1,054	10
		Всего	2,65	1,424	20
	Всего	нет	3,75	1,517	20
		есть	2,40	1,273	20
		Всего	3,08	1,542	40

Рис. 15.6. Фрагмент окна вывода после выполнения шага 5 (Описательные статистики)

Проверка ковариационных матриц на равенство

Результаты проверки равенства ковариационных матриц, которые появляются, если установлен флажок Критерии однородности в окне ОЛМ-многомерная: Параметры, при-

ведены на рис. 15.7. Эти результаты говорят о том, что дисперсионно-ковариационные матрицы не различаются между собой (значение p -уровня превышает 0,05). Если бы результаты проверки были статистически достоверными, то мы не могли бы ручаться за достоверность многомерного анализа, одним из условий проведения которого является равенство дисперсионно-ковариационных матриц.

**Критерий Бокса
равенства
ковариационных
матриц**

М Бокса	19,003
F	,893
ст. св. 1	18
ст. св. 2	4579,730
Знач	,588

Рис. 15.7. Фрагмент окна вывода после выполнения шага 5 (Критерий Бокса равенства ковариационных матриц)

Многомерные критерии

Таблица, показанная на рис. 15.8, — результат выполнения шага 5 или 5б (без ковариаты). В ней содержится информация, которая является результатом проверки влияния факторов и их взаимодействий на зависимые переменные. Секция Свободный член характеризует остаточную дисперсию и не подлежит интерпретации. Как можно видеть, обнаружены статистически значимые главные эффекты факторов Инт и Отсрочка ($p < 0,05$). Взаимодействие этих факторов (Инт*Отсрочка) тоже статистически досто-

Многомерные критерии^а

Эффект		Значения	F	Ст. св. гипотезы	Ст. св. ошибки	Знач	Частная Эта в Квадрате
Свободный член	След Пиллая	,957	253,376 ^а	3,000	34,000	,000	,957
	Лямбда Уилкса	,043	253,376 ^а	3,000	34,000	,000	,957
	След Хотеллинга	22,357	253,376 ^а	3,000	34,000	,000	,957
	Наибольший корень Роя	22,357	253,376 ^а	3,000	34,000	,000	,957
Инт	След Пиллая	,768	37,415 ^а	3,000	34,000	,000	,768
	Лямбда Уилкса	,232	37,415 ^а	3,000	34,000	,000	,768
	След Хотеллинга	3,301	37,415 ^а	3,000	34,000	,000	,768
	Наибольший корень Роя	3,301	37,415 ^а	3,000	34,000	,000	,768
Отсрочка	След Пиллая	,425	8,371 ^а	3,000	34,000	,000	,425
	Лямбда Уилкса	,575	8,371 ^а	3,000	34,000	,000	,425
	След Хотеллинга	,739	8,371 ^а	3,000	34,000	,000	,425
	Наибольший корень Роя	,739	8,371 ^а	3,000	34,000	,000	,425
Инт * Отсрочка	След Пиллая	,378	6,895 ^а	3,000	34,000	,001	,378
	Лямбда Уилкса	,622	6,895 ^а	3,000	34,000	,001	,378
	След Хотеллинга	,608	6,895 ^а	3,000	34,000	,001	,378
	Наибольший корень Роя	,608	6,895 ^а	3,000	34,000	,001	,378

^а Точная статистика

^б План: Свободный член + Инт + Отсрочка + Инт * Отсрочка

Рис. 15.8. Фрагмент окна вывода после выполнения шага 5 (многомерные критерии)

верно. Если в анализе участвуют ковариаты, они также отображаются в этой таблице. Выполнив шаг 5а, вы можете убедиться, что ковариата знач, применяемая в нашем примере, оказывает значимое влияние на зависимые переменные и меняет статистическую значимость влияния факторов и их взаимодействия.

Далее дана трактовка терминов, используемых программой в окне вывода:

- ▶ Значения — в этом столбце представлены значения различных критериев для проверки влияния независимых переменных. Наиболее надежным из всех критериев считается тест Пилая (Pillai).
- ▶ F — оценка F -критерия.
- ▶ Ст. св. гипотезы — степень свободы гипотезы. Произведение числа зависимых переменных и уровней каждой независимой переменной, уменьшенных на 1. Для данного примера

$$3 \cdot (2 - 1)(2 - 1) = 3 \cdot 1 \cdot 1 = 3.$$

- ▶ Ст. св. ошибки — величина, вычисляемая различными способами в зависимости от критерия.
- ▶ Знч. — уровень значимости для соответствующего F -критерия.

Критерий равенства дисперсий

Критерий равенства дисперсий (критерий Ливиня) позволяет проверить допущение о том, что дисперсии всех зависимых переменных одинаковы (рис. 15.9).

Критерий Ливиня проверки равенства дисперсий.

	F	ст. св. 1	ст. св. 2	Знч.
НАЧАЛО	,471	3	36	,704
СЕРЕДИНА	1,818	3	36	,161
КОНЕЦ	1,390	3	36	,262

Рис. 15.9. Фрагмент окна вывода после выполнения шага 5 (критерий Ливиня проверки равенства дисперсий)

Большие значения p -уровня для каждой из зависимых переменных не дают поводов для излишнего беспокойства относительно корректности дисперсионного анализа.

Одномерные критерии

В дополнение к многомерным критериям к каждой из зависимых переменных в отдельности применяется одномерный F -критерий. Несмотря на то что многомерный критерий, как было сказано ранее, является более совершенным, чем серия одномерных критериев, последние необходимы для интерпретации статистически достоверных результатов применения многомерного критерия. Таким образом, если многомерный критерий показывает статистически значимый результат, то для детальной интерпретации используются результаты вычисления одномерных критериев.

В нашем случае статистически достоверными являются результаты применения одномерных критериев для главных эффектов и взаимодействия факторов Инт и Отсрочка. На рис. 15.10 приведен фрагмент таблицы с результатами одномерного анализа. Как видно по этому фрагменту, влияние фактора Инт статистически достоверно в отношении двух из трех зависимых переменных: СЕРЕДИНА и КОНЕЦ, — а влияние фактора Отсрочка

статистически достоверно в отношении каждой из трех переменных. Взаимодействие факторов (Инт*Отсрочка) статистически достоверно в отношении переменной КОНЕЦ. Для содержательной интерпретации одномерных критериев в нашем случае можно обратиться к описательным статистикам (средним), а лучше — к графикам средних значений. Если бы факторы имели более двух градаций, перед интерпретацией необходимо было бы применить метод апостериорных парных сравнений. Если анализ включает в себя ковариаты, для каждой зависимой переменной одномерный критерий будет учитывать и их влияние.

Оценка эффектов межгрупповых факторов							
Источник	Зависимая переменная	Сумма квадратов типа III	ст. св.	Средний квадрат	F	Знач.	Частная Eta в квадрате
Скорректированная модель	НАЧАЛО	25,475 ^a	3	8,492	4,542	,008	,275
	СЕРЕДИНА	33,800 ^b	3	11,267	13,255	,000	,525
	КОНЕЦ	34,275 ^c	3	11,425	11,521	,000	,490
Свободный член	НАЧАЛО	378,225	1	378,225	202,319	,000	,849
	СЕРЕДИНА	547,600	1	547,600	644,235	,000	,947
	КОНЕЦ	429,025	1	429,025	432,630	,000	,923
Инт	НАЧАЛО	7,225	1	7,225	3,865	,057	,097
	СЕРЕДИНА	14,400	1	14,400	16,941	,000	,320
	КОНЕЦ	9,025	1	9,025	9,101	,005	,202
Отсрочка	НАЧАЛО	18,225	1	18,225	9,749	,004	,213
	СЕРЕДИНА	16,900	1	16,900	19,882	,000	,356
	КОНЕЦ	21,025	1	21,025	21,202	,000	,371
Инт * Отсрочка	НАЧАЛО	,025	1	,025	,013	,909	,000
	СЕРЕДИНА	2,500	1	2,500	2,941	,095	,076
	КОНЕЦ	4,225	1	4,225	4,261	,046	,106

Рис. 15.10. Фрагмент окна вывода после выполнения шага 5 (одномерные критерии)

Оценки средних значений для уровней факторов

Таблица, приведенная на рис. 15.11, получена после выполнения шага 5б. Она содержит средние значения каждой зависимой переменной для каждой градации факторов. По этим данным можно дать подробную содержательную интерпретацию статистически достоверных результатов. Так, статистически достоверное влияние фактора Инт на совокупность зависимых переменных было детализировано результатами применения одномерных критериев: фактор Отсрочка значимо влияет на переменные СЕРЕДИНА и КОНЕЦ. Судя по средним значениям для этих переменных, интонационное выделение середины ряда слов приводит к тому, что эта часть ряда запоминается лучше, а конец ряда — хуже, чем без интонационного выделения. Влияние фактора Отсрочка проявляется в том, что отсроченное воспроизведение менее продуктивно для каждой из трех частей ряда.

Статистически достоверное взаимодействие факторов легче всего интерпретировать по графикам средних, которые получены после выполнения шага 5б. На рис. 15.12 приведен один из этих графиков для переменной КОНЕЦ. Как показывает график, взаимодействие факторов проявляется в том, что спад продуктивности отсроченного воспроизведения конца ряда слов проявляется по-разному в зависимости от интонационного выделения: при интонационном выделении спад меньше, а при его отсутствии спад выражен сильнее. Более подробно интерпретация графиков объяснялась в разделе «Представление результатов» главы 14.

Интонационное выделение * Отсрочка						
Зависимая переменная	Интонационное выделение	Отсрочка	Среднее	Стд. Ошибка	95% доверительный интервал	
					Нижняя граница	Верхняя граница
НАЧАЛО	нет	нет	4,200	,432	3,323	5,077
		есть	2,800	,432	1,923	3,677
	есть	нет	3,300	,432	2,423	4,177
		есть	2,000	,432	1,123	2,877
СЕРЕДИНА	нет	нет	3,500	,292	2,909	4,091
		есть	2,700	,292	2,109	3,291
	есть	нет	5,200	,292	4,609	5,791
		есть	3,400	,292	2,809	3,991
КОНЕЦ	нет	нет	4,800	,315	4,161	5,439
		есть	2,700	,315	2,061	3,339
	есть	нет	3,200	,315	2,561	3,839
		есть	2,400	,315	1,761	3,039

Рис. 15.11. Фрагмент окна вывода после выполнения шага 5б (оценки средних значений для уровней факторов)

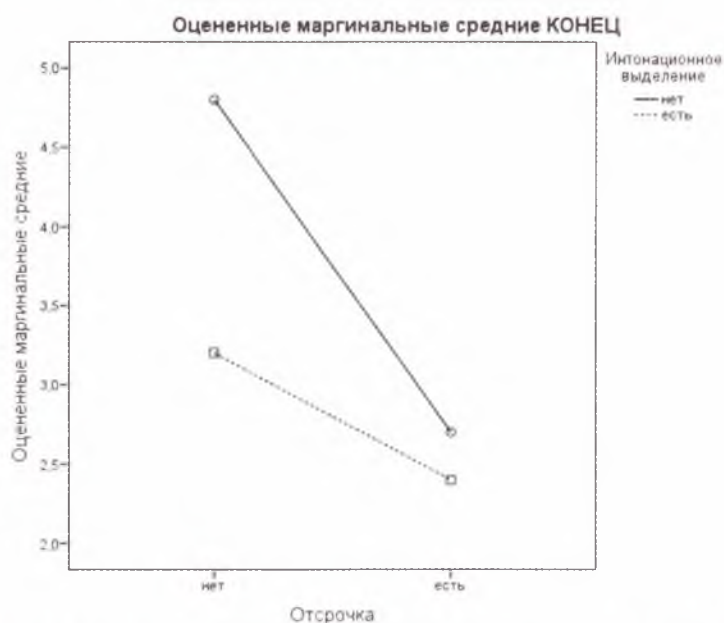


Рис. 15.12. Фрагмент окна вывода после выполнения шага 5б (график средних)

При желании можно построить альтернативные графики, например столбиковые, воспользовавшись меню в командной строке Графика ► Устаревшие диалоговые окна ► Столбики (см. шаг 6 и рис. 14.7 в главе 14).

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

16 Дисперсионный анализ с повторными измерениями

210	Пошаговые алгоритмы вычислений
216	Представление результатов
220	Завершение анализа и выход из программы

В предыдущей главе рассматривалась ситуация, когда необходимо было анализировать влияние на совокупность зависимых переменных нескольких независимых переменных, каждая из которых могла иметь два или более уровня. В подобных случаях проводятся межгрупповые сравнения, где группам соответствуют уровни факторов и их сочетания. Такие факторы, разным уровням которых соответствуют разные группы объектов, называются *межгрупповыми*. Однако на практике нередко используются и *внутригрупповые факторы*, разным уровням которых соответствует одна и та же группа объектов. В этом случае говорят о внутригрупповой модели, с помощью которой исследуется одна и та же группа объектов для нескольких уровней независимой переменной.

Наиболее распространенным примером применения внутригрупповой модели является исследование эффекта какого-либо воздействия. В этом случае обычно используются две выборки: на одну из них воздействие оказывается (экспериментальная выборка), а на другую — нет (контрольная выборка). Зависимые переменные, в отношении которых ожидается эффект воздействия, измеряются для обеих выборок дважды: до воздействия (предварительное тестирование) и после воздействия (итоговое тестирование).

Такая схема исследования предполагает изучения влияния на зависимые переменные двух факторов: межгруппового (два уровня: контрольная и экспериментальная выборки) и внутригруппового (два уровня: до и после воздействия). Обратите внимание на то, что внутригрупповому фактору в данном случае соответствуют два измерения одной и той же переменной (или группы переменных). В общем случае внутригрупповой фактор может иметь и большее число уровней (повторных измерений). Отметим, что наибольший интерес для исследователя будет представлять гипотеза о взаимодействии межгруппового и внутригруппового факторов. Если она подтвердится, можно будет утверждать, что обнаружены различия между контрольной и экспериментальной группами в динамике изменения зависимой переменной.

В примерах этой главы мы будем использовать файл `ex020.sav`, содержащий те же гипотетические результаты эксперимента по изучению эффективности запоминания слов, что и в файлах `ex021.sav` и `ex022.sav`. Но в файле `ex021.sav` (глава 14) продуктивность запоминания представлена одной (зависимой) переменной Слова, в файле `ex022.sav` (глава 15) — тремя переменными в соответствии с частями ряда (НАЧАЛО, СЕРЕДИНА, КОНЕЦ),

а в файле ex020.sav — шестью переменными в соответствии с тремя частями ряда слов и повтором их воспроизведения. Более подробное описание особенностей представления данных в соответствии с вариантами дисперсионного анализа приведено в разделе «Файлы данных для группы методов Общая линейная модель» (глава 14).

План исследования в соответствии со структурой данных в файле ex020.sav может быть представлен как комбинация двух внутригрупповых факторов (часть ряда и отсрочка), одного межгруппового фактора (интонационное выделение) и одной ковариаты (значимость ряда слов). Первый внутригрупповой фактор соответствует части ряда слов и имеет 3 уровня (начало, середина и конец). Второй внутригрупповой фактор — отсрочка воспроизведения — имеет 2 уровня (без отсрочки, с отсрочкой в 2 дня). Нетрудно заметить, что существует 6 комбинаций уровней этих двух факторов, каждой из которых соответствует одна из переменных (табл. 16.1).

Таблица 16.1. Сочетания комбинаций части ряда слов и отсрочки их воспроизведения

Часть ряда	Отсрочка воспроизведения	
	Нет	Да
Начало	начало1	начало2
Середина	середина1	середина2
Конец	конец1	конец2

Переменные, перечисленные в таблице, по сути представляют собой 6 повторных измерений одной и той же характеристики — продуктивности воспроизведения 8 предъявленных ранее слов. Поэтому для анализа таких данных используется модель многомерного дисперсионного анализа с повторными измерениями.

Данные файла ex020.sav могут быть обработаны и иначе, в предположении наличия трех разных зависимых переменных (начало, середина, конец), измеренных дважды: без отсрочки и после нее. В этом случае в рассмотрение следует ввести лишь один внутригрупповой фактор — отсрочка (2 уровня). Напомним, что при предъявлении слов для одной выборки середина ряда интонационно выделялась, для другой — нет. Следовательно, такой план исследования предполагает проведение двухфакторного дисперсионного анализа (межгрупповой фактор — интонационное выделение, внутригрупповой фактор — отсрочка) в отношении трех зависимых переменных (начало, середина, конец).

Модель многомерного ANOVA с повторными измерениями имеет много общего с рассмотренными ранее вариантами дисперсионного анализа без повторных измерений, поэтому материал этой главы в значительной степени опирается на материал трех предыдущих. Если вы пропустили главы 13–15, настоятельно рекомендуем вернуться к ним, прежде чем приступать к интерпретации результатов приводимых здесь примеров.

Следует иметь в виду, что ANOVA с повторными измерениями — самый сложный метод дисперсионного анализа. Поэтому в некоторых случаях, например при большом количестве повторов, целесообразнее реструктурировать данные так, чтобы можно было воспользоваться более простыми вариантами анализа, как это описано в разделе «Файлы данных для группы методов Общая линейная модель» (глава 14).

Пошаговые алгоритмы вычислений

При проведении многомерного дисперсионного анализа с повторными измерениями сначала необходимо выполнить 3 подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае ex020.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом ex020.sav.

ШАГ 4

В меню Анализ выберите команду Общая линейная модель ► ОЛМ-повторные измерения. На экране появится диалоговое окно ОЛМ-повторные измерения: Задать факторы (рис. 16.1).

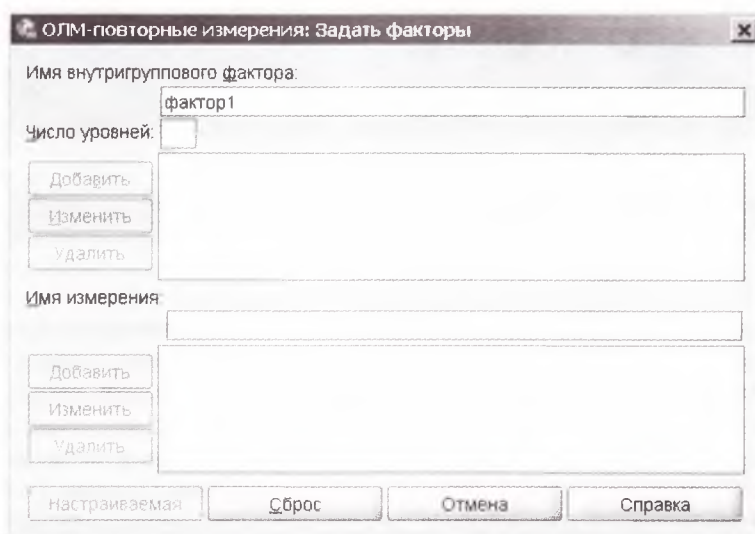


Рис. 16.1. Диалоговое окно ОЛМ-повторные измерения: Задать факторы

В этом окне вам предстоит задать имена внутригрупповых факторов. Обратите внимание на то, что имена, которые вы будете вводить в поле Имя внутригруппового фактора, не относятся к существующим переменным из файла данных, а определяют виртуальные переменные для команды ОЛМ-повторные измерения. В нашем примере присвоим этим переменным (внутригрупповым факторам) имена Ч_ряда (3 уровня: начало, середина, конец) и Отсрочка (2 уровня: нет, есть).

Для задания имени внутригрупповой переменной сначала необходимо ввести его в поле Имя внутригруппового фактора, затем задать число уровней переменной в поле Число уровней и щелкнуть на кнопке Добавить.

Довольно часто изучают динамику не одного показателя, а целой группы переменных, например, при оценке влияния внутригрупповых факторов (динамики) на несколько зависимых переменных. Для задания имен нескольких зависимых переменных, каждой из которых соответствует несколько повторов измерений, предназначено поле Имя измерения. Так, результаты эксперимента из файла ex020.sav могут быть обработаны

в предположении наличия трех *разных* зависимых переменных (запоминание слов): начало, середина и конец. Каждая из них измерена дважды: непосредственно после предъявления слов и с отсрочкой в два дня. В соответствии с этим планом будет выполнен анализ влияния только одного внутригруппового фактора (отсрочка, 2 уровня) на 3 зависимые переменные. Для этого в поле Имя измерения необходимо последовательно ввести новые имена переменных, подтверждая задание каждого имени нажатием кнопки Добавить.

После того как имена внутригрупповых факторов и переменных определены, чтобы завершить задание параметров анализа, щелкните на кнопке Настраиваемая. На экране появится диалоговое окно ОЛМ-повторные измерения (рис. 16.2).

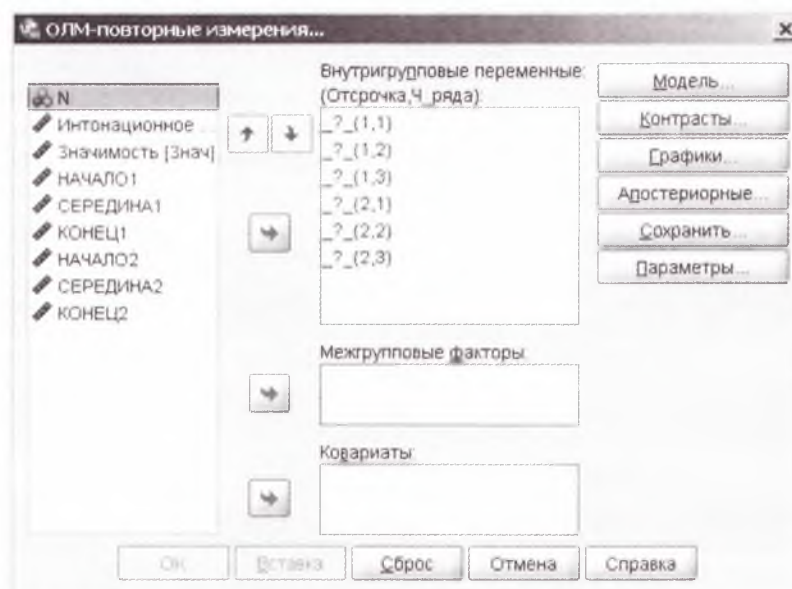


Рис. 16.2. Диалоговое окно ОЛМ-повторные измерения (два внутригрупповых фактора)

В левой части окна находится список переменных файла ex02.sav. В списке Внутригрупповые переменные вы можете указать соответствие между переменными файла и каждым уровнем внутригруппового фактора.

В нашем случае в анализ включены два внутригрупповых фактора, один из которых имеет 2, а другой — 3 уровня; таким образом, каждой из 6 комбинаций уровней необходимо сопоставить переменную из исходного списка. Пары чисел, указанные в скобках в списке Внутригрупповые переменные, определяют, к какой ячейке модели относятся переменные файла, которые далее будут обозначаться как внутригрупповые переменные (табл. 16.2).

Обратите внимание на то, что при движении от начала к концу списка переменных файла (первый столбец табл. 16.2) градации фактора «отсрочка» меняются реже, чем градации фактора «часть ряда». Поэтому в качестве первого фактора лучше выбирать тот, который имеет меньше уровней. Несмотря на то что эта иллюстрация относится к случаю исследования двух факторов с небольшим числом уровней, она с легкостью может быть обобщена для любого числа факторов и их уровней.

Таблица 16.2. Соответствие переменных файла и ячеек модели с двумя факторами

Переменная	Ячейка	Отсрочка	Часть ряда
начало1	(1,1)	Нет — 1	Начало — 1
середина1	(1,2)	Нет — 1	Середина — 2
конец1	(1,3)	Нет — 1	Конец — 3
начало2	(2,1)	Есть — 2	Начало — 1
середина2	(2,2)	Есть — 2	Середина — 2
конец2	(2,3)	Есть — 2	Конец — 3

Альтернативная схема анализа проще: она предполагает наличие одного внутригруппового фактора Отсрочка, но трех разных зависимых переменных (начало, середина, конец). Соответствие переменных файла ячейкам модели показано в табл. 16.3.

Таблица 16.3. Соответствие переменных файла и ячеек модели с одним фактором

Переменная	Ячейка	Отсрочка
начало1	(1, начало)	Нет — 1
начало2	(2, начало)	Есть — 2
середина1	(1, середина)	Нет — 1
середина2	(2, середина)	Есть — 2
конец1	(1, конец)	Нет — 1
конец2	(2, конец)	Есть — 2

Для того чтобы ввести какую-либо переменную в поле Внутригрупповые переменные, достаточно выделить ее имя в списке слева и щелкнуть на верхней кнопке с направленной вправо стрелкой.

По умолчанию SPSS сопоставляет первую введенную переменную первой ячейке модели (в нашем случае ячейке 1,1), вторую переменную — второй ячейке (1,2) и т. д. Если же вы хотите нарушить данный порядок, воспользуйтесь кнопками с направленными вверх и вниз стрелками, чтобы переместить введенную переменную в нужное место списка ячеек.

После того как все внутригрупповые зависимые переменные анализа определены, вы можете указать все межгрупповые переменные (факторы) и ковариаты. Для этого переместите нужные переменные из исходного списка в поля Межгрупповые факторы и Ковариаты.

В правой части диалогового окна ОЛМ-повторные измерения расположены 6 кнопок: Модель, Контрасты, Графики, Апостериорные, Сохранить и Параметры. Щелчок на кнопке Параметры открывает диалоговое окно, идентичное окну ОЛМ-многомерная: Параметры,

с которым мы познакомились в главе 15. Кнопка Диаграммы работает так же, как ее аналог из глав 14 и 15, но позволяет задавать диаграммы не только для межгрупповых, но и для внутригрупповых факторов. Функции кнопки Апостериорные также остались прежними, единственным отличием является то, что парные апостериорные сравнения применимы только к межгрупповым факторам. За описанием кнопок Контрасты и Сохранение обратитесь к руководству пользователя.

При щелчке на кнопке Модель открывается диалоговое окно ОЛМ-повторные измерения: Модель, представленное на рис. 16.3. В большинстве случаев исследователи предпочитают пользоваться полной факторной моделью и оставляют переключатель Задать модель установленным в положение Полная факторная. Тем не менее иногда требуется задать набор главных эффектов и взаимодействий вручную. Для этого следует установить переключатель Задать модель в положение Настраиваемая, а затем переместить желаемые эффекты и взаимодействия из левых списков в правые. Заметьте, что вы не сможете переместить межгрупповые факторы во внутригрупповую секцию модели и наоборот, так как это некорректно. После того как желаемая модель определена, щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.

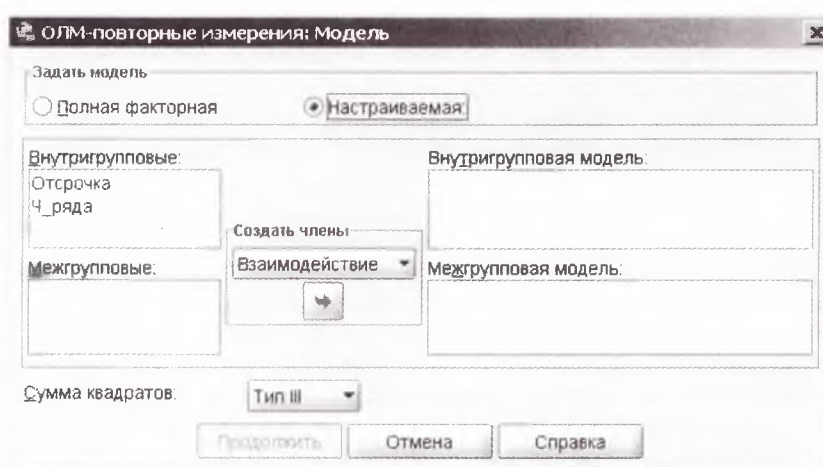


Рис. 16.3. Диалоговое окно ОЛМ-повторные измерения: Модель

Далее будут приведены два примера использования команды ОЛМ-повторные измерения. В первом примере мы проведем дисперсионный анализ с двумя внутригрупповыми факторами: Часть ряда (3 градации) и Отсрочка воспроизведения (2 градации). С каждой из 6 ячеек модели будет связана одна из переменных: начало1, середина1, конец1, начало2, середина2, конец2.

ШАГ 5

После выполнения предыдущего шага должно быть открыто диалоговое окно ОЛМ-повторные измерения: Задать факторы (см. рис. 16.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. С помощью клавиши Delete или Backspace удалите имя фактор1 из поля Имя внутригруппового фактор, введите туда имя Отсрочка, в поле Число уровней, введите число 2 и щелкните на кнопке Добавить.

2. В поле Имя внутригруппового фактора введите имя Ч_ряда, в поле Число уровней введите число 3 и щелкните сначала на кнопке Добавить, а потом — на кнопке Настраиваемая, чтобы открыть диалоговое окно ОЛМ-повторные измерения (см. рис. 16.2).
3. Наведите указатель мыши на переменную начало1, нажмите кнопку мыши и, не отпуская ее, переместите указатель на переменную конец2 и отпустите кнопку. В результате окажутся выделенными шесть переменных: начало1, середина1, конец1, начало2, середина2, конец2.
4. Щелкните на верхней кнопке с направленной вправо стрелкой, чтобы переместить выделенные переменные в поле Внутригрупповые переменные.
5. Выделите в левом окне переменную Инт и нажатием средней кнопки со стрелкой перенесите ее в поле Межгрупповые факторы.
6. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-повторные измерения: Параметры.
7. Установите флажки Описательные статистики и Оценки размера эффекта, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.
8. Щелкните на кнопке Графики, чтобы открыть диалоговое окно ОЛМ-повторные измерения: Графики профилей, аналогичное показанному на рис. 14.3.
9. Щелкните сначала на переменной Ч_ряда, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Горизонтальная ось.
10. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Отдельные линии.
11. Щелкните сначала на кнопке Добавить, чтобы зафиксировать заданные параметры построения диаграмм, а затем на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.
12. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Следующий пример реализует альтернативную схему анализа тех же данных. В нем проводится смешанный анализ влияния одного внутригруппового фактора Отсрочка и одного межгруппового фактора интонационного выделения (Инт) на три зависимые переменные (начало, середина, конец).

ШАГ 5

После выполнения предыдущего шага должно быть открыто диалоговое окно ОЛМ-повторные измерения: Задать факторы (см. рис. 16.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. С помощью клавиши Delete или Backspace удалите имя фактор1 из поля Имя внутригруппового фактор, введите туда имя Отсрочка, в поле Число уровней, введите число 2 и щелкните на кнопке Добавить.
2. В поле Имя внутригруппового фактора введите имя Ч_ряда, в поле Число уровней введите число 3 и щелкните сначала на кнопке Добавить, а потом на кнопке Настраиваемая, чтобы открыть диалоговое окно ОЛМ-повторные измерения (см. рис. 16.2).

3. Наведите указатель мыши на переменную начало1, нажмите кнопку мыши и, не отпуская ее, переместите указатель на переменную конец2 и отпустите кнопку. В результате окажутся выделенными шесть переменных: начало1, середина1, конец1, начало2, середина2, конец2.
4. Щелкните на верхней кнопке с направленной вправо стрелкой, чтобы переместить выделенные переменные в поле Внутригрупповые переменные.
5. Выделите в левом окне переменную Инт и нажатием средней кнопки со стрелкой перенесите ее в поле Межгрупповые факторы.
6. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-повторные измерения: Параметры.
7. Установите флажки Описательные статистики и Оценки размера эффекта, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.
8. Щелкните на кнопке Графики, чтобы открыть диалоговое окно ОЛМ-повторные измерения: Графики профилей, аналогичное показанному на рис. 14.3.
9. Щелкните сначала на переменной Ч_ряда, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Горизонтальная ось.
10. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Отдельные линии.
11. Щелкните сначала на кнопке Добавить, чтобы зафиксировать заданные параметры построения диаграмм, а затем на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.
12. Щелкните на кнопке ОК, чтобы открыть окно вывода.

ШАГ 5A

После выполнения предыдущего шага должно быть открыто диалоговое окно ОЛМ-повторные измерения: Задать факторы (см. рис. 16.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. С помощью клавиши Delete или Backspace удалите имя фактор1 из поля Имя внутригруппового фактора, введите туда имя Отсрочка, а в поле Число уровней, введите число 2 и щелкните на кнопке Добавить.
2. В окно Имя измерения введите слово начало и щелкните на кнопке Добавить.
3. Повторите предыдущее действие со словами середина и конец. Щелкните на кнопке Настраиваемая, чтобы открыть диалоговое окно ОЛМ-повторные измерения (см. рис. 16.2).
4. Наведите указатель мыши на переменную начало1, нажатием левой кнопки мыши выделите ее, нажмите на клавиатуре клавишу Ctrl, не отпуская ее, переместите указатель мыши на переменную начало2 и выделите ее нажатием левой кнопки мыши. Отпустите клавишу Ctrl. В результате окажутся выделенными две переменные: начало1 и начало2.
5. Щелкните на верхней кнопке с направленной вправо стрелкой, чтобы переместить выделенные переменные в поле Внутригрупповые переменные. Они займут первые две позиции: (1, начало), (2, начало).

6. Повторите пункты 4 и 5 в отношении пар переменных середина1, середина2 и конец1, конец2 так, чтобы они заняли следующие пары позиций в правом поле.
7. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на средней кнопке с направленной вправо стрелкой, чтобы переместить переменную в поле Межгрупповые факторы.
8. Щелкните на кнопке Графики, чтобы открыть диалоговое окно ОЛМ-повторные измерения: Графики профилей, аналогичное показанному на рис. 14.3.
9. Щелкните сначала на переменной Отсрочка, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Горизонтальная ось.
10. Щелкните сначала на переменной Инт, чтобы выделить ее, а затем на средней кнопке со стрелкой, чтобы переместить переменную в поле Отдельные линии.
11. Щелкните сначала на кнопке Добавить, чтобы зафиксировать заданные параметры построения диаграмм, а затем на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.
12. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно ОЛМ-повторные измерения: Параметры.
13. Установите флажки Описательные статистики, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно ОЛМ-повторные измерения.
14. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически откроет окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Большая часть результатов, генерируемых командой ОЛМ-повторные измерения, аналогична результатам многомерного дисперсионного анализа (глава 15), поэтому далее будут представлены лишь те фрагменты, которые являются специфичными для внутригрупповых моделей. Также не приводятся некоторые результаты, которые являются избыточными и практически не требуют интерпретации.

Многомерные критерии

Таблица Многомерные критерии (рис. 16.4), получена после выполнения шага 5. Из нее удалена часть строк, не представляющих интереса для исследователя (по поводу редактирования окна вывода обращайтесь к соответствующему разделу главы 2).

Как видно из таблицы, взаимодействие факторов Часть ряда и Интонационное выделение (Ч_ряда*Инт) статистически достоверно. Это значит, что влияние интонационного выделения на воспроизведение слов проявляется по-разному в зависимости от части

ряда. Более детальную интерпретацию позволяет сделать график средних значений (рис. 16.5).

Многомерные критерии							
Эффект		Значения	F	Ст. св. гипотезы	Ст. св. ошибки	Знач	Частная Эта в Квадрате
Отсрочка * Инт	След Пиллая	,014	,253	1,000	18,000	,621	,014
	Лямбда Уилкса	,986	,253	1,000	18,000	,621	,014
	След Хотеллинга	,014	,253	1,000	18,000	,621	,014
	Наибольший корень Роя	,014	,253	1,000	18,000	,621	,014
Ч_ряда * Инт	След Пиллая	,829	41,271	2,000	17,000	,000	,829
	Лямбда Уилкса	,171	41,271	2,000	17,000	,000	,829
	След Хотеллинга	4,855	41,271	2,000	17,000	,000	,829
	Наибольший корень Роя	4,855	41,271	2,000	17,000	,000	,829
Отсрочка * Ч_ряда	След Пиллая	,022	,193	2,000	17,000	,826	,022
	Лямбда Уилкса	,978	,193	2,000	17,000	,826	,022
	След Хотеллинга	,023	,193	2,000	17,000	,826	,022
	Наибольший корень Роя	,023	,193	2,000	17,000	,826	,022
Отсрочка * Ч_ряда * Инт	След Пиллая	,622	13,998	2,000	17,000	,000	,622
	Лямбда Уилкса	,378	13,998	2,000	17,000	,000	,622
	След Хотеллинга	1,647	13,998	2,000	17,000	,000	,622
	Наибольший корень Роя	1,647	13,998	2,000	17,000	,000	,622

Рис. 16.4. Фрагмент окна вывода после выполнения шага 5: Многомерные критерии

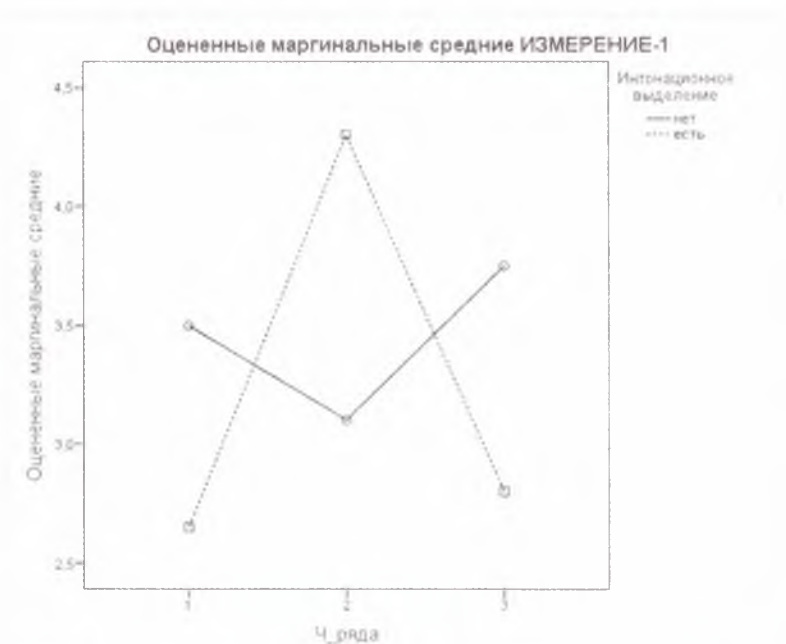


Рис. 16.5. Фрагмент окна вывода после выполнения шага 5 (график средних значений)

Отметим, что тот же результат был получен с применением более простого метода при условии реструктурирования данных (глава 14).

Проверка внутригрупповых эффектов

Таблица, показанная на рис. 16.6, получена после выполнения шага 5а.

Многомерный ^{а, с}						
Внутригрупповой эффект		Значения	F	Ст. св. гипотезы	Ст. св. ошибки	Знач.
Отсрочка	След Пиплая	,861	33,101 ^а	3,000	16,000	,000
	Лямбда Уилкса	,139	33,101 ^а	3,000	16,000	,000
	След Хотеллинга	6,206	33,101 ^а	3,000	16,000	,000
	Наибольший корень Роя	6,206	33,101 ^а	3,000	16,000	,000
Отсрочка * Инт	След Пиплая	,632	9,151 ^а	3,000	16,000	,001
	Лямбда Уилкса	,368	9,151 ^а	3,000	16,000	,001
	След Хотеллинга	1,716	9,151 ^а	3,000	16,000	,001
	Наибольший корень Роя	1,716	9,151 ^а	3,000	16,000	,001

а. Точная статистика

б. План. Свободный член + Инт

в. Пределах объектов Design: Отсрочка

с. Критерии основаны на усредненных переменных

Рис. 16.6. Фрагмент окна вывода после выполнения шага 5а (проверка внутригрупповых эффектов: многомерные критерии)

Полученные результаты во многом дублируют результаты предшествующего применения многомерных критериев. Таблица содержит главный эффект внутригруппового фактора (Отсрочка), а также эффект его взаимодействия с межгрупповым фактором (Отсрочка*Инт). Все эффекты оказываются статистически достоверными. Обратите внимание на то, что данные результаты относятся к совокупности всех трех зависимых переменных. На это указывает подзаголовок таблицы «Многомерный».

Если многомерные критерии внутригрупповых эффектов дают статистически достоверные результаты, то для детальной интерпретации этих результатов следует обратиться к таблице внутригрупповых эффектов «Одномерные критерии», которая приведена на рис. 16.7 в отредактированном виде.

Одномерные критерии							
Источник	Измерение		Сумма квадратов типа III	ст. св.	Средний квадрат	F	Энц.
Отсрочка * Инт	Начало	Предполагаемая сферичность	,025	1	,025	,086	,773
		Гринхауз-Гайссер	,025	1,000	,025	,086	,773
		Юнга-Фельдта	,025	1,000	,025	,086	,773
		Ограниченный снизу	,025	1,000	,025	,086	,773
	Середина	Предполагаемая сферичность	2,500	1	2,500	8,03	,011
		Гринхауз-Гайссер	2,500	1,000	2,500	8,03	,011
		Юнга-Фельдта	2,500	1,000	2,500	8,03	,011
		Ограниченный снизу	2,500	1,000	2,500	8,03	,011
	Конец	Предполагаемая сферичность	4,225	1	4,225	8,22	,010
		Гринхауз-Гайссер	4,225	1,000	4,225	8,22	,010
		Юнга-Фельдта	4,225	1,000	4,225	8,22	,010
		Ограниченный снизу	4,225	1,000	4,225	8,22	,010

Рис. 16.7. Фрагмент окна вывода после выполнения шага 5а (проверка внутригрупповых эффектов: одномерные критерии)

Одномерные критерии показывают, что взаимодействие факторов Инт и Отсрочка является статистически достоверным в отношении переменных середина и конец. Для интерпретации взаимодействия следует обратиться к графикам средних значений (рис. 16.8). Судя по графику для переменной середина, взаимодействие факторов Инт и Отсрочка обусловлено тем, что на графике наблюдается более крутой наклон линии для случая интонационного выделения. Иначе говоря, на продуктивность воспроизведения середины ряда слов отсрочка сильнее влияет в том случае, когда слова были интонационно выделены.

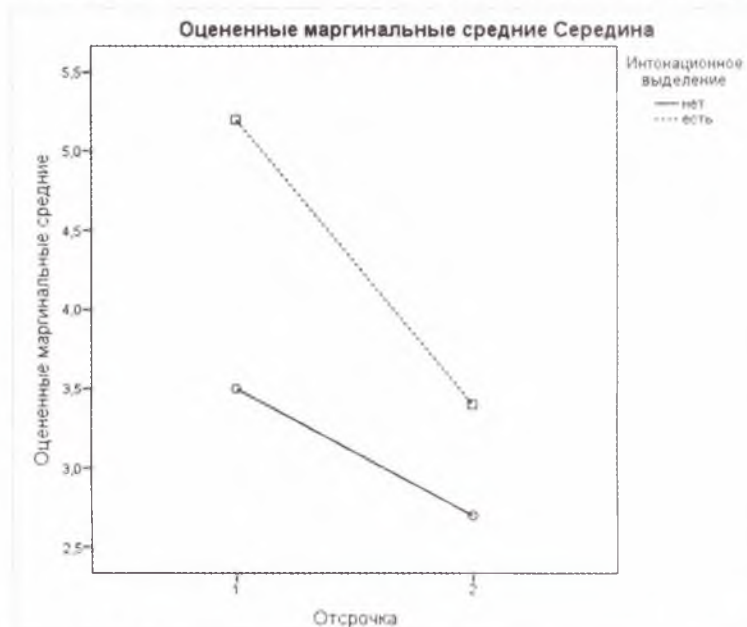


Рис. 16.8. Фрагмент окна вывода после выполнения шага 5а (график средних)

Оценка эффектов межгрупповых факторов

Таблица, показанная на рис. 16.9, получена после выполнения шага 5а.

Оценка эффектов межгрупповых факторов						
Преобразуемая переменная Середина						
Источник	Измерение	Сумма квадратов типа III	ст. сб.	Средний квадрат	F	Знач.
Свободный член	Начало	378,225	1	378,225	109,719	,000
	Середина	547,600	1	547,600	394,272	,000
	Конец	429,025	1	429,025	291,964	,000
Инт	Начало	7,225	1	7,225	2,096	,165
	Середина	14,400	1	14,400	10,368	,005
	Конец	9,025	1	9,025	6,142	,023
Ошибка	Начало	62,050	18	3,447		
	Середина	25,000	18	1,389		
	Конец	26,450	18	1,469		

Рис. 16.9. Фрагмент окна вывода после выполнения шага 5а (оценка эффектов межгрупповых факторов)

Межгрупповые эффекты фактора Инт демонстрируют статистически достоверные результаты в отношении переменных середина и конец. Это значит, что влияние интонационного выделения на эти переменные проявляется независимо от отсрочки.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

17 Простая линейная регрессия

221	Простая регрессия
223	Оценка криволинейности
225	Пошаговые алгоритмы вычислений
228	Представление результатов
230	Терминология, используемая при выводе
231	Завершение анализа и выход из программы

Команда Линейная регрессия позволяет выполнять как простой, так и множественный регрессионный анализ. Простой регрессионный анализ обсуждается в этой главе, а множественный — в следующей. Такое разбиение одного статистического раздела было сделано нами в целях упрощения изложения материала, касающегося основ регрессионного анализа. Если вы не знакомы с множественным регрессионным анализом, эта глава послужит введением. Мы рассмотрим такие понятия, как прогнозируемые значения зависимой переменной и уравнение регрессии, покажем связь между простой регрессией и корреляцией двух переменных, рассмотрим влияние одной переменной на дисперсию другой, а также ознакомимся с оценкой криволинейности связи двух переменных.

Простая регрессия

Зачастую, имея информацию об одной характеристике объекта, мы пытаемся на основе этой информации делать выводы о другой его характеристике, связанной с исходной. Например, если мы знаем рост человека, то могли бы попытаться оценить его возможный вес. Например, зная, что человек имеет рост 214 см, с определенной вероятностью мы можем прогнозировать, что его вес превысит 91 кг. Можно привести множество пар связанных между собой величин: коэффициент интеллекта (IQ) и академическая успеваемость, число сокращений мышц ног в секунду и скорость бега, калорийность пищи и вес, симпатия к человеку и желание оказать ему помощь и т. д. Каждый день мы оцениваем связи различных факторов между собой. Именно на проведение таких оценок и нацелена простая регрессия. Этот метод не может дать абсолютно достоверного результата, однако позволяет ответить на вопрос о связи переменных, а также по заданному значению одной переменной рассчитать наиболее вероятное значение другой переменной.

Для иллюстрации практического применения простой регрессии воспользуемся новым файлом данных с именем exam.sav. Он содержит 10-балльную оценку нервной возбудимости (тревожности) 36 студентов и количество решенных ими зачетных тестовых заданий (из 20 возможных). Гипотеза о линейности отношений этих двух переменных говорит о том, что чем выше нервная возбудимость студента, тем выше его резуль-

тативность (например, потому что спокойных студентов меньше волнуют их знания, а тревожные студенты проводят больше времени за подготовкой к зачету). В качестве зависимой переменной (критерия) выступает переменная *тест*, а в качестве независимой переменной (предиктора) — переменная *трев*. Мы будем пытаться прогнозировать значения переменной *тест* по известным значениям переменной *трев*, используя *уравнение регрессии*. Уравнение регрессии формируется на основе общего уравнения, связывающего фактическую успеваемость студента и нервную возбудимость:

$$\text{тест}_{\text{истина}} = \text{константа} + \text{коэффициент} \cdot \text{трев} + \text{остаток},$$

где $\text{тест}_{\text{истина}}$ — переменная, отражающая реальный результат выполнения тестового задания; *константа* — некоторая константа; *коэффициент* — регрессионный коэффициент при значении оценки тревожности; *остаток* — остаток. В результате применения линейного регрессионного анализа константа оказалась равной 9,3114, а коэффициент регрессии 0,6751. Соответственно, уравнение для прогноза результата зачетного тестирования выглядит следующим образом:

$$\text{тест}_{\text{прогноз}} = 9,3114 + 0,6751 \cdot \text{трев};$$

$$\text{тест}_{\text{истина}} = \text{тест}_{\text{прогноз}} + \text{остаток}.$$

Прогнозируемое значение будет отличаться от истинного значения. Разница отражает тот факт, что рассчитываемые результаты экзамена никогда не бывают абсолютно точными и, чтобы получить истинный результат, необходимо ввести в уравнение член, равный разности прогнозируемого и реального значений. Этот член и называют *остатком*:

$$\text{реальное значение} = \text{прогнозируемое значение} + \text{остаток}.$$

Можно предположить, что помимо нервной возбудимости на результаты зачетного тестирования влияют и другие факторы. Регрессионный анализ, учитывающий влияние нескольких факторов, называется *множественным* и рассматривается в следующей главе.

Далее перечислены величины, которые вычисляются при проведении регрессионного анализа с помощью команд подменю Линейная регрессия:

- Коэффициент, характеризующий связь между значениями зависимой и независимой переменных, обозначается буквой *R* и представляет собой уже знакомый нам по предыдущим главам коэффициент корреляции. Использование прописной буквы вместо строчной в окне вывода результатов регрессионного анализа объясняется тем, что в общем случае речь идет о *множественных* корреляциях с участием нескольких независимых переменных. Корреляция подробно рассмотрена в главе 9.
- Для рассчитанного значения множественного коэффициента корреляции *R* программа определяет уровень значимости *p*. Как и ранее, величина $p < 0,05$ свидетельствует о значимой корреляции переменных. При $p > 0,05$ вероятность случайности результата считается слишком высокой, и в этом случае говорят, что связь между переменными слабая или не обнаружена.
- Значение R^2 является обычным квадратом величины *R*, однако этот коэффициент имеет собственный смысл. Коэффициент R^2 характеризует долю дисперсии одной переменной, обусловленной воздействием другой переменной. Так, для переменных *трев* и *тест* значение $R = 0,546$, а $R^2 = 0,298$. Это означает, что 29,8 % дисперсии переменной *тест* объясняется влиянием независимой переменной *трев*.
- Константу и коэффициент уравнения регрессии часто называют *B*-величинами. *B*-величины характеризуют связь между значениями переменных *трев* и *тест*.

Оценка криволинейности

Выдвинутая в предыдущем разделе гипотеза о том, что увеличение нервной возбудимости перед экзаменом всегда улучшает результат студента, очевидно, вызывает сомнения. Более разумными кажутся следующие соображения. При низкой возбудимости результаты экзамена должны быть низкими, поскольку излишнее спокойствие снижает потребность студента в подготовке. С увеличением возбудимости результат до определенного момента должен улучшаться, однако слишком возбудимые студенты вряд ли способны сконцентрироваться и показать хороший результат. Поэтому наилучших показателей должны добиваться те студенты, чей уровень возбудимости является промежуточным. При регрессионном анализе вне зависимости от того, является он множественным или простым, соотношение между зависимой и независимой переменными считается линейным. В приведенном ранее примере мы видим значительную корреляцию между переменными трев и тест ($R = 0,546$, $p < 0,001$), однако возможная ошибка прогноза велика (только 29,8 % дисперсии переменной тест объясняется влиянием переменной трев). Можно предположить, что если изменить вид общего уравнения (например, включить в него квадрат переменной трев), прогнозируемые значения будут ближе к реальным. График, приведенный на рис. 17.1, может быть получен при обращении Графика ► Устаревшие диалоговые окна Рассеяния/точки. Он наглядно демонстрирует характер отношений между переменными трев и тест.

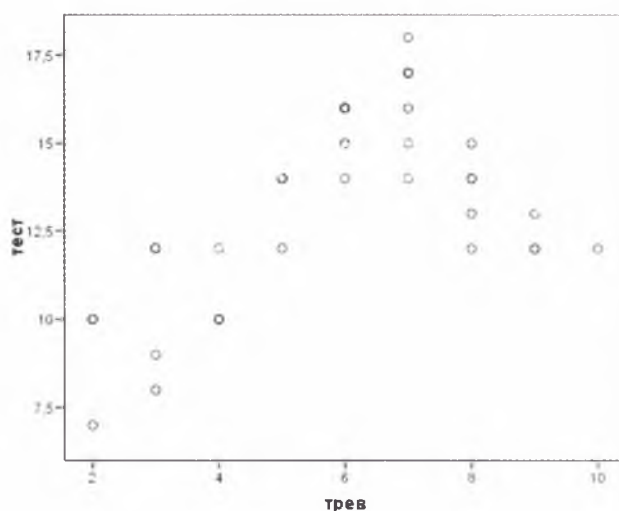


Рис. 17.1. Диаграмма рассеяния, демонстрирующая криволинейную зависимость

По вертикальной оси диаграммы отложены значения переменной тест, по горизонтальной — значения переменной трев. Очевидно, что зависимость между переменными не является линейной: у нее имеется выраженный максимум, а при движении к краям диаграммы наблюдается убывание значений переменной тест. Нельзя не отметить, что только путем графической интерпретации невозможно достоверно установить характер отношений между переменными. Необходимым условием является применение статистических критериев.

Для того чтобы статистически оценить криволинейность, в подменю Регрессия предусмотрена команда Подгонка кривых. В ее диалоговом окне необходимо задать зависимую переменную (тест), независимую переменную (трев) и установить флажки Линейная и Квадратичная. После щелчка на кнопке ОК появятся результаты, приведенные ниже (рис. 17.2), а также диаграмма, показанная на рис. 17.3.

Зависимая переменная: тест

Уравнение	Сводка модели					Оценки параметров		
	R квадрат	F	ст. св1	ст. св2	Знач	Константа	b1	b2
Линейная	,298	14,466	1	34	,001	9,311	,675	
Квадратичная	,675	34,283	2	33	,000	,161	4,490	,338

Независимой переменной является трев

Рис. 17.2. Таблица результатов оценки криволинейности

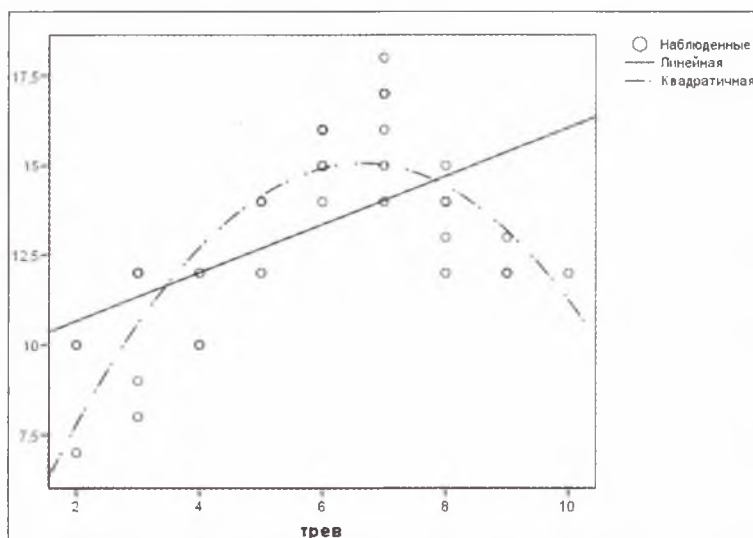


Рис. 17.3. Диаграмма, демонстрирующая линейную и криволинейную тенденции

По диаграмме можно оценить, насколько близка к линейной или квадратичной зависимость между значениями переменных. В результаты включены значения коэффициентов B регрессии (Константа b_0 , b_1 , b_2), поэтому несложно составить линейное и квадратичное уравнения регрессии для прогнозируемых значений. На диаграмму помимо линейной и квадратичной зависимостей нанесен разброс данных файла. Обратите внимание на сходство диаграмм, представленных на рис. 17.1 и 17.3, а также на то, что константа и коэффициенты уравнений указаны в последних трех столбцах выводимых результатов (см. рис. 17.2).

Для линейной регрессии уравнение имеет вид

$$\text{тест}_{\text{прогноз}} = 9,3114 + 0,6751 \cdot \text{трев}.$$

А уравнение для квадратичной регрессии выглядит следующим образом:

$$\text{тест}_{\text{прогноз}} = 0,1615 + 4,4896 \cdot \text{трев} - 0,3381 \cdot (\text{трев})^2.$$

Как видно из рис. 17.2, в случае линейной регрессии величина R^2 (столбец R квадрат в таблице выводимых результатов) равна 0,298, то есть 29,8 % дисперсии переменной

тест обусловлено воздействием со стороны переменной трев. В то же время для квадратичной регрессии, которая учитывает и линейную, и криволинейную связи, $R^2 = 0,675$, то есть она обуславливает 67,5 % дисперсии переменной тест. Малый p -уровень для обоих уравнений свидетельствует об очень высокой статистической достоверности полученных результатов. Очевидно, что квадратичная регрессия описывает отношения между переменными тест и трев более адекватно, чем линейная.

Пошаговые алгоритмы вычислений

Сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае exam.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом exam.sav.

Простой регрессионный анализ

ШАГ 4

В меню Анализ выберите команду Регрессия ► Линейная. На экране появится диалоговое окно Линейная регрессия (рис. 17.4).

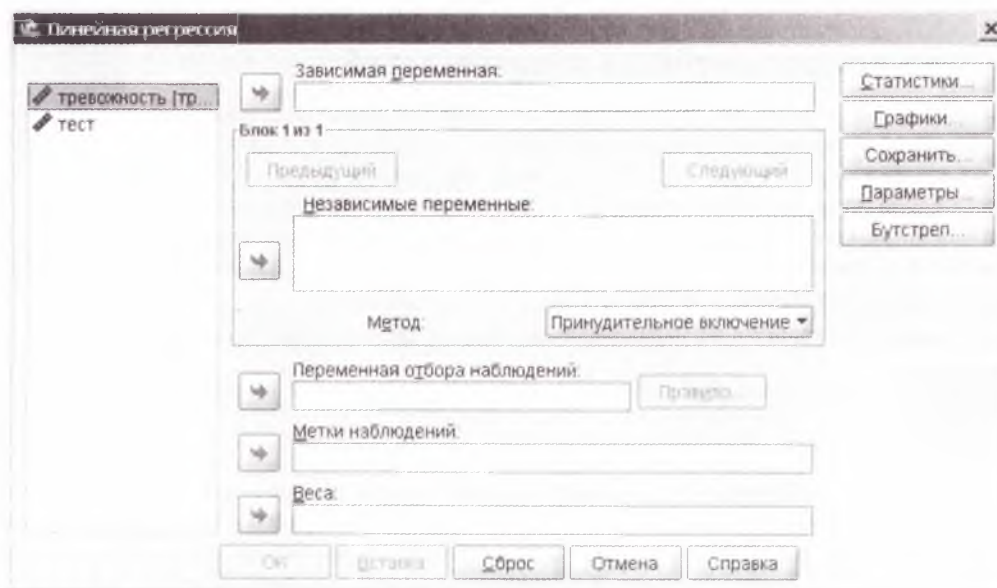


Рис. 17.4. Диалоговое окно Линейная регрессия

Диалоговое окно Линейная регрессия позволяет проводить разные виды регрессионного анализа. Поскольку в данном случае выполняется простой регрессионный анализ, некоторые из элементов интерфейса этого окна мы рассматривать не будем.

Поскольку в файле `exam.sav` содержатся только 2 переменные, в списке, находящемся в левой части диалогового окна, лишь 2 имени — `трев` и `тест`. Позже нами будет создана еще одна переменная с именем `трев2`. Для того чтобы провести регрессионный анализ, достаточно выполнить несколько простых действий: поместить переменную `тест` в поле Зависимая переменная, переменную `трев` — в поле Независимые переменные и щелкнуть на кнопке ОК. В результате программа вычислит величины R , R^2 , F , соответствующие им p -уровни, B -коэффициенты (коэффициенты и константы уравнения регрессии), а также стандартизированные β -коэффициенты, характеризующие степень зависимости между значениями исследуемых переменных.

В следующем примере приведем последовательность действий, необходимую для выполнения простого регрессионного анализа с участием независимой переменной `трев` и зависимой переменной `тест`.

ШАГ 5

После выполнения шага 4 должно быть открыто диалоговое окно Линейная регрессия (см. рис. 17.4).

1. Щелкните сначала на переменной `тест`, чтобы выделить ее, а затем — на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.
2. Щелкните сначала на переменной `трев`, чтобы выделить ее, а затем — на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Независимые переменные.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Анализ криволинейных зависимостей

Рассмотрим две пошаговые процедуры, относящиеся к анализу криволинейных зависимостей. Первая позволяет получить результаты, приведенные в начале этой главы (таблицу, приведенную на рис. 17.2, и диаграмму, показанную на рис. 17.3), вторая создает специальную квадратичную переменную, которую при желании можно использовать в дальнейших вычислениях. Несмотря на то что криволинейные зависимости рассматриваются в контексте простого регрессионного анализа, все сказанное в этой главе справедливо и для множественного регрессионного анализа.

После завершения шага 3 на экране должно присутствовать окно редактора данных.

ШАГ 4а

В меню Анализ выберите команду Регрессия ► Подгонка кривых. На экране появится диалоговое окно Подгонка кривых (рис. 17.5).

Первоначальные действия в окне Подгонка кривых те же, что и в окне Линейная регрессия: поместите переменную `тест` в поле Зависимые, а переменную `трев` в поле Переменная в области Независимая. Обратите внимание на то, что по умолчанию в диалоговом окне установлены три флажка: Включить в уравнение константу, Графики моделей, позволяющий получить диаграмму (см. рис. 17.3), и Линейная, обеспечивающий проверку гипотезы о прямолинейности и включение в диаграмму прямой линии.

Как можно видеть, в области Модели помимо флажка Линейная имеется множество других флажков, позволяющих проводить проверки на соответствие изучаемого распределения самым разным видам кривых и включать эти кривые в диаграмму. В следующем примере продемонстрируем проверку зависимости на квадратичность.

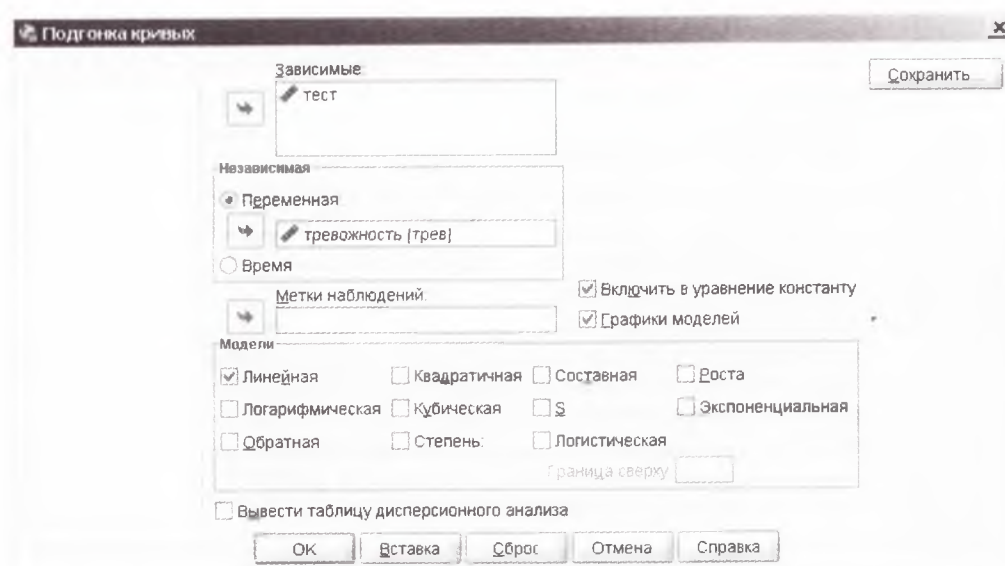


Рис. 17.5. Диалоговое окно Подгонка кривых

ШАГ 5а

Далее перечислены действия, позволяющие провести проверку криволинейности. После выполнения шага 4а должно быть открыто диалоговое окно Подгонка кривых (см. рис. 17.5).

1. Щелкните сначала на переменной тест, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимые.
2. Щелкните сначала на переменной трев, чтобы выделить ее, затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Переменная в области Независимая. Установите флажок Квадратичная.
3. Щелкните на кнопке ОК, чтобы открыть окно вывода.

При составлении квадратного уравнения регрессии SPSS вычисляет новую переменную, значения которой равны квадратам соответствующих значений переменной трев. Квадратичная зависимость графически имеет вид параболы — кривой с двумя ветвями, которые могут быть направлены вверх, если коэффициент при квадратном члене положительный, или вниз, если коэффициент при квадратном члене отрицательный. Поскольку квадратное уравнение имеет не только квадратный, но и положительный линейный член, изображаемый в виде возрастающей наклонной прямой, конец левой ветви параболы расположен ниже, чем конец правой ветви.

Для того чтобы включить в анализ регрессии квадрат переменной трев, нам необходимо вручную создать переменную трев2, содержащую квадраты значений переменной трев, а затем указать ее в качестве независимой переменной. Процесс создания новых переменных описан в разделе «Преобразование данных» главы 4. Здесь мы приведем все необходимые инструкции, однако для более детального изучения вопроса следует обратиться к упомянутому разделу, в котором, в частности, рассказывается, что создание новой переменной начинается с диалогового окна Вычислить переменную.

ШАГ 56

На этом шаге мы создадим новую переменную `трев2` и включим ее в последующий регрессионный анализ в качестве независимой. Перед выполнением следующих инструкций у вас должно быть открыто окно редактора данных:

1. В меню Преобразовать окна редактора данных SPSS выберите команду Вычислить переменную. На экране появится одноименное диалоговое окно.
2. В поле Вычисляемая переменная введите имя `трев2`.
3. Щелкните сначала на переменной `трев`, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы ввести переменную в условие отбора, составляемое в поле Числовое выражение.
4. Следом за именем переменной в поле Числовое выражение нажмите кнопку с изображением двух звездочек (**) или введите две звездочки с клавиатуры, затем введите цифру 2 и щелкните на кнопке ОК, чтобы вернуться в окно редактора данных.
5. В меню Анализ выберите команду Регрессия ► Линейная регрессия, чтобы открыть диалоговое окно Линейная регрессия (см. рис. 17.4). Если вы уже работали с этим окном, щелкните на кнопке Сброс.
6. Щелкните сначала на переменной `тест`, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.
7. Щелкните сначала на переменной `трев`, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Независимые переменные.
8. Повторите предыдущее действие для переменной `трев2`.
9. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 56 программа автоматически перейдет в окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Простой регрессионный анализ

Фрагменты выводимых данных, сгенерированные программой после выполнения шага 5, приведены на рис. 17.6.

Как можно видеть, между переменными `тест` и `трев` имеется значимая линейная связь. Это означает, что с увеличением нервной возбудимости (тревожности) студента количество выполненных им тестовых заданий на зачете также имеет тенденцию к увеличению.

Регрессионный анализ криволинейной зависимости

После выполнения шага 5а в результаты вывода войдут сводка модели с оценками параметров (см. рис. 17.2) и диаграмма зависимостей (см. рис. 17.3).



Фрагменты выводимых данных, сгенерированные программой после выполнения шага 5б, приведены на рис. 17.7.

Сводка для модели					
Модель	R	R квадрат	Скорректи- рованный R квадрат	Стд. ошибка оценки	
1	,546 ^a	,298	,278	2,306	

а. Предикторы: (константа) трев

Дисперсионный анализ ^b						
Модель		Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
1	Регрессия	76,901	1	76,901	14,466	,001 ^a
	Остаток	180,738	34	5,316		
	Итого	257,639	35			

а. Предикторы: (константа) трев

б. Зависимая переменная: тест

Коэффициенты ^a						
Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стд. ошибка	Бета		
1	(Константа)	9,311	1,118		8,327	,000
	трев	,675	,177	,546	3,803	,001

а. Зависимая переменная: тест

Рис. 17.6. Фрагмент окна вывода после выполнения шага 5

Сводка для модели					
Модель	R	R квадрат	Скорректированный R квадрат	Стд. ошибка оценки	
1	,922 ^a	,675	,655	1,593	

а Предикторы: (константа) трев2, трев

Дисперсионный анализ ^b						
Модель		Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
1	Регрессия	173,929	2	86,964	34,283	,000 ^a
	Остаток	83,710	33	2,537		
	Итого	257,639	35			

а Предикторы: (константа) трев2, трев

б Зависимая переменная тест

Коэффициенты ^a						
Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стд. ошибка	Бета		
1	(Константа)	,161	1,669		,097	,924
	трев	4,490	,629	3,633	7,139	,000
	трев2	-,338	,055	-,3148	-,6185	,000

а Зависимая переменная тест

Рис. 17.7. Фрагмент окна вывода после выполнения шага 5б

Термины, используемые в приведенных результатах, имеют тот же смысл, что и в предыдущем случае. Обратите внимание на значительный коэффициент корреляции зависимой переменной тест с независимыми переменными трев и трев2. Величина $R^2 = 0,675$ означает, что 67,5 % дисперсии переменной тест обусловлено влиянием переменных трев и трев2. Значения F -критерия и соответствующие значимости (для F и t) говорят о сильном воздействии на зависимую переменную как обеих независимых переменных, так и каждой переменной в отдельности. То, как влияют несколько независимых переменных, будет рассмотрено в следующей главе. Как вы можете видеть, значения β (бета) для нелинейных отношений не ограничиваются диапазоном от -1 до 1 . В столбце В последней таблицы приведены коэффициенты уравнения регрессии. Сравните их с коэффициентами уравнения для квадратичной регрессии, приведенного в середине главы.

Терминология, используемая при выводе

- ▶ R — поскольку в анализе участвовала единственная независимая переменная, ее значение равно коэффициенту корреляции (r) между переменными тест и трев.
- ▶ R квадрат — квадрат величины R (R^2), равный доле дисперсии переменной тест, обусловленной воздействием переменной трев.
- ▶ Скорректированный R квадрат — скорректированная величина R^2 . Величина R^2 , используемая в расчетах, на практике оказывается несколько завышенной. Скорректированная величина R^2 менее формальна и ближе к реальным результатам.
- ▶ Стд. ошибка оценки — в таблице Сводка для модели это стандартное отклонение оценок значений зависимой переменной тест.
- ▶ Регрессия — статистики, оценивающие долю дисперсии зависимой переменной, обусловленную влиянием независимых переменных.
- ▶ Остаток — статистики, оценивающие долю дисперсии зависимой переменной, не обусловленную влиянием независимых переменных.
- ▶ ст. св. — число степеней свободы, для регрессии оно равно числу независимых переменных, для остатка — разности размера выборки и числа степеней свободы регрессии, уменьшенной на единицу ($36 - 1 - 1 = 34$).
- ▶ Сумма квадратов, Средний квадрат, F — показатели, относящиеся к дисперсионному анализу (глава 13).
- ▶ Знч. — величина p -уровня значимости, вероятность случайности полученного результата.
- ▶ B — коэффициент и константа линейного уравнения регрессии:

$$\text{тест}_{\text{прогноз}} = 9,311 + 0,675(\text{трев}).$$
- ▶ Стд. ошибка — стандартная ошибка, в таблице Коэффициенты это характеристика стабильности коэффициента B , равная стандартному отклонению коэффициентов B , рассчитанных для большого числа выборок из генеральной совокупности.
- ▶ Бета — стандартизованный коэффициент регрессии (β). Представляет собой коэффициент B для стандартизованных значений переменной трев. Для линейных отношений ее значение всегда лежит в диапазоне от -1 до 1 , а для криволинейных отношений может выходить за границы этого диапазона.
- ▶ t — отношение коэффициента B к его стандартной ошибке.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

18 Множественный регрессионный анализ

232	Уравнение множественной регрессии
233	Коэффициенты регрессии
234	Коэффициент детерминации и пошаговые методы
235	Условия получения приемлемых результатов анализа
236	Пошаговые алгоритмы вычислений
242	Представление результатов
244	Терминология, используемая при выводе
244	Завершение анализа и выход из программы

Множественная регрессия является расширением простой линейной регрессии, описанной в главе 17. С помощью простой регрессии оценивалась степень влияния одной независимой переменной (предиктора) на зависимую переменную (критерий). В отличие от простой регрессии, множественная регрессия исследует влияние двух и более предикторов на критерий.

Анализ регрессии можно свести к геометрической интерпретации. Когда вычислена простая корреляция между двумя переменными, можно построить *линию регрессии* (линию наилучшего соответствия). Эта линия строится на основе уравнения регрессии, ее угол определяется коэффициентом при независимой переменной, а сдвиг по вертикальной оси — константой. Далее мы продемонстрируем множественную регрессию как последовательное усложнение простого регрессионного уравнения. Специально для этой главы нами создан файл данных `help.sav`. Этот файл содержит данные психологического исследования склонности людей оказывать помощь своим знакомым. Хотя данные являются вымышленными, результаты их обработки близки к результатам одного из реальных исследований.

Уравнение множественной регрессии

Итак, в этой главе мы используем новый набор данных `help.sav`. Переменная *помощь* представляет время (в секундах), потраченное человеком на оказание помощи своему партнеру, и ее значения имеют нормальное распределение (среднее равно 30, стандартное отклонение — 10). Переменная *симпатия* отражает оценку симпатии к партнеру в баллах от 1 до 20. На примере этих двух переменных мы продемонстрируем простую регрессию. В качестве зависимой выступит переменная *помощь*, а в качестве независимой — переменная *симпатия* (предполагается, что симпатия и сочувствие заставляют человека оказывать помощь, а не наоборот). Как показал анализ, коэффициент корреляции между переменными *помощь* и *симпатия* составляет 0,416 при значимости $p = 0,004$, что говорит о значительной связи между этими переменными. Константа и коэффициент регрессии

составили 14,739 и 1,547 соответственно. Таким образом, уравнение регрессии имеет следующий вид:

$$\text{помощь}_{\text{прогноз}} = 14.739 + 1.547 \cdot (\text{симпатия}).$$

Если для некоторого испытуемого значение переменной симпатия составит 16, то на основе регрессионного уравнения мы можем прогнозировать, что переменная помощь примет следующее значение:

$$14.739 + 1.547 \cdot 16 = 39.5.$$

Значение 16 выше средней симпатии, в результате прогнозируемое значение помощи превышает среднее ее значение почти на одно стандартное отклонение.

Аналогичные расчеты можно выполнять и при множественном регрессионном анализе. Различие заключается лишь в том, что при множественном анализе уравнение регрессии включает в себя более чем одну зависимую переменную.

Помимо переменной симпатия с переменной помощь коррелируют и другие переменные файла help.sav. В частности, это переменные агрессия (агрессивность человека по отношению к партнеру, измеренная в баллах от 1 до 20) и польза (самооценка собственной полезности в баллах от 1 до 20). Множественный регрессионный анализ показал следующие коэффициенты при каждой из переменных: $B_{(\text{симпатия})} = 1.0328$, $B_{(\text{агрессия})} = 1.1676$, $B_{(\text{польза})} = 1.2569$, $\text{константа} = -5.3147$. Уравнение регрессии для множественного анализа имеет следующий вид:

$$\begin{aligned} \text{помощь}_{\text{прогноз}} = & -5.3147 + 1.0328 \cdot (\text{симпатия}) + 1.1676 \cdot (\text{агрессия}) + \\ & + 1.2569 \cdot (\text{польза}). \end{aligned}$$

Возьмем объект с номером 7 и рассчитаем для него прогнозируемое значение переменной помощи:

$$\text{помощь}_{\text{прогноз}} = -5.3147 + 1.0328 \cdot 2 + 1.1676 \cdot 10 + 1.2569 \cdot 9 = 19.74.$$

Таким образом, человек, имеющий низкий показатель симпатии и средние показатели агрессивности и самооценки полезности, должен, согласно прогнозу, оказывать незначительную помощь. Фактическое значение переменной помощь для объекта 7 составило 21, что свидетельствует о высокой точности нашего прогноза.

Коэффициенты регрессии

Положительный коэффициент при независимой переменной говорит о том, что с ее возрастанием значение зависимой переменной также возрастает. Верно и противоположное утверждение: при отрицательном коэффициенте с возрастанием независимой переменной значение зависимой переменной убывает. Тем не менее в большинстве исследований соотношение коэффициентов не позволяет делать вывод о воздействии того или иного фактора на зависимую переменную, поскольку независимые переменные, как правило, измеряются в разных шкалах и имеют разный масштаб. Чтобы разрешить эту проблему, был введен коэффициент регрессии β , принимающий значения от -1 до 1 . Он отражает *частную корреляцию* независимой и зависимой переменных. Под частной корреляцией понимается воздействие, которое оказывает на зависимую переменную одна из независимых переменных при фиксированных значениях других независимых переменных (с учетом влияния последних). Чем больше данная независимая переменная коррелирует с другими независимыми переменными, тем меньше абсолютное значение ее коэффи-

циента β . Так, в обсуждаемом примере коэффициент β для переменной агрессия — это ее корреляция с зависимой переменной *после* исключения влияния переменных симпатия и польза. Для простой регрессии с одной независимой переменной коэффициент β равен значению парной корреляции зависимой и независимой переменных. Таким образом, коэффициент β является универсальной мерой влияния независимой переменной, его часто называют *стандартным коэффициентом регрессии*. И именно стандартные коэффициенты регрессии позволяют сравнивать независимые переменные по их влиянию на оценку зависимой переменной.

Обратимся к записанному ранее уравнению регрессии. Как можно видеть, все 3 независимые переменные входят в него с положительными коэффициентами. Такой результат вполне логичен для переменных симпатия и польза, однако вызывает недоумение то, что рост агрессивности субъекта влечет за собой увеличение его стремления оказывать помощь. Причина появления такого странного результата — хороший повод для дискуссии. Каждый раз, когда вы будете попадать в подобные ситуации, проверяйте правильность кодирования и ввода данных. Помните, что программа способна лишь генерировать результаты анализа, но в отношении их практической интерпретации она беспомощна.

Коэффициент детерминации и пошаговые методы

Коэффициент R является мерой связи всей совокупности независимых переменных и зависимой переменной. Часто его называют *коэффициентом множественной корреляции*. Значение R^2 равно доле дисперсии зависимой переменной, обусловленной влиянием независимых переменных, и называется *коэффициентом детерминации*. Для регрессионного анализа с тремя независимыми переменными, речь о котором шла ранее, значение R составило 0,571, а R^2 — 0,326. Это означает, что 32,6 % дисперсии переменной *помощь* определяется совокупным воздействием переменных *агрессия*, *симпатия* и *польза*. Дополнительная информация о коэффициентах множественной корреляции и детерминации приведена в разделе «Представление результатов».

Множественный регрессионный анализ позволяет использовать любое количество предикторов, но присутствие большого числа независимых переменных не всегда удобно. Было бы предпочтительно иметь в качестве предикторов как можно больше переменных, оказывающих значимое влияние на критерий, и как можно меньше переменных, не оказывающих такого влияния. В процедуру множественной регрессии SPSS включены методы, позволяющие производить пошаговый отбор в регрессионное уравнение только значимых независимых переменных. Одним из них является метод Включение, суть которого заключается в следующем. Сначала процедура вычисляет, какая из независимых переменных имеет наибольший коэффициент корреляции с зависимой переменной, а затем составляет уравнение регрессии с участием этой переменной. Далее из числа оставшихся предикторов выбирается тот, который имеет наибольший коэффициент β при условии, что β является значимым. Выбранный предиктор также включается в уравнение регрессии. Процесс продолжается до тех пор, пока не будут выбраны все предикторы, оказывающие значимое воздействие на зависимую переменную (имеющие статистически достоверные коэффициенты β). По умолчанию SPSS продолжает выбирать независимые переменные до тех пор, пока уровень значимости (p) коэффициентов

составили 14,739 и 1,547 соответственно. Таким образом, уравнение регрессии имеет следующий вид:

$$\text{помощь}_{\text{прогноз}} = 14,739 + 1,547 \cdot (\text{симпатия}).$$

Если для некоторого испытуемого значение переменной симпатия составит 16, то на основе регрессионного уравнения мы можем прогнозировать, что переменная помощь примет следующее значение:

$$14,739 + 1,547 \cdot 16 = 39,5.$$

Значение 16 выше средней симпатии, в результате прогнозируемое значение помощи превышает среднее ее значение почти на одно стандартное отклонение.

Аналогичные расчеты можно выполнять и при множественном регрессионном анализе. Различие заключается лишь в том, что при множественном анализе уравнение регрессии включает в себя более чем одну зависимую переменную.

Помимо переменной симпатия с переменной помощь коррелируют и другие переменные файла help.sav. В частности, это переменные агрессия (агрессивность человека по отношению к партнеру, измеренная в баллах от 1 до 20) и польза (самооценка собственной полезности в баллах от 1 до 20). Множественный регрессионный анализ показал следующие коэффициенты при каждой из переменных: $B_{(\text{симпатия})} = 1,0328$, $B_{(\text{агрессия})} = 1,1676$, $B_{(\text{польза})} = 1,2569$, константа = -5,3147. Уравнение регрессии для множественного анализа имеет следующий вид:

$$\begin{aligned} \text{помощь}_{\text{прогноз}} = & -5,3147 + 1,0328 \cdot (\text{симпатия}) + 1,1676 \cdot (\text{агрессия}) + \\ & + 1,2569 \cdot (\text{польза}). \end{aligned}$$

Возьмем объект с номером 7 и рассчитаем для него прогнозируемое значение переменной помощь:

$$\text{помощь}_{\text{прогноз}} = -5,3147 + 1,0328 \cdot 2 + 1,1676 \cdot 10 + 1,2569 \cdot 9 = 19,74.$$

Таким образом, человек, имеющий низкий показатель симпатии и средние показатели агрессивности и самооценки полезности, должен, согласно прогнозу, оказывать незначительную помощь. Фактическое значение переменной помощь для объекта 7 составило 21, что свидетельствует о высокой точности нашего прогноза.

Коэффициенты регрессии

Положительный коэффициент при независимой переменной говорит о том, что с ее возрастанием значение зависимой переменной также возрастает. Верно и противоположное утверждение: при отрицательном коэффициенте с возрастанием независимой переменной значение зависимой переменной убывает. Тем не менее в большинстве исследований соотношение коэффициентов не позволяет делать вывод о воздействии того или иного фактора на зависимую переменную, поскольку независимые переменные, как правило, измеряются в разных шкалах и имеют разный масштаб. Чтобы разрешить эту проблему, был введен коэффициент регрессии β , принимающий значения от -1 до 1. Он отражает *частную корреляцию* независимой и зависимой переменных. Под частной корреляцией понимается воздействие, которое оказывает на зависимую переменную одна из независимых переменных при фиксированных значениях других независимых переменных (с учетом влияния последних). Чем больше данная независимая переменная коррелирует с другими независимыми переменными, тем меньше абсолютное значение ее коэффи-

циента β . Так, в обсуждаемом примере коэффициент β для переменной агрессия — это ее корреляция с зависимой переменной *после* исключения влияния переменных симпатия и польза. Для простой регрессии с одной независимой переменной коэффициент β равен значению парной корреляции зависимой и независимой переменных. Таким образом, коэффициент β является универсальной мерой влияния независимой переменной, его часто называют *стандартным коэффициентом регрессии*. И именно стандартные коэффициенты регрессии позволяют сравнивать независимые переменные по их влиянию на оценку зависимой переменной.

Обратимся к записанному ранее уравнению регрессии. Как можно видеть, все 3 независимые переменные входят в него с положительными коэффициентами. Такой результат вполне логичен для переменных симпатия и польза, однако вызывает недоумение то, что рост агрессивности субъекта влечет за собой увеличение его стремления оказывать помощь. Причина появления такого странного результата — хороший повод для дискуссии. Каждый раз, когда вы будете попадать в подобные ситуации, проверяйте правильность кодирования и ввода данных. Помните, что программа способна лишь генерировать результаты анализа, но в отношении их практической интерпретации она беспомощна.

Коэффициент детерминации и пошаговые методы

Коэффициент R является мерой связи всей совокупности независимых переменных и зависимой переменной. Часто его называют *коэффициентом множественной корреляции*. Значение R^2 равно доле дисперсии зависимой переменной, обусловленной влиянием независимых переменных, и называется *коэффициентом детерминации*. Для регрессионного анализа с тремя независимыми переменными, речь о котором шла ранее, значение R составило 0,571, а R^2 — 0,326. Это означает, что 32,6 % дисперсии переменной помощь определяется совокупным воздействием переменных агрессия, симпатия и польза. Дополнительная информация о коэффициентах множественной корреляции и детерминации приведена в разделе «Представление результатов».

Множественный регрессионный анализ позволяет использовать любое количество предикторов, но присутствие большого числа независимых переменных не всегда удобно. Было бы предпочтительно иметь в качестве предикторов как можно больше переменных, оказывающих значимое влияние на критерий, и как можно меньше переменных, не оказывающих такого влияния. В процедуру множественной регрессии SPSS включены методы, позволяющие производить пошаговый отбор в регрессионное уравнение только значимых независимых переменных. Одним из них является метод Включение, суть которого заключается в следующем. Сначала процедура вычисляет, какая из независимых переменных имеет наибольший коэффициент корреляции с зависимой переменной, а затем составляет уравнение регрессии с участием этой переменной. Далее из числа оставшихся предикторов выбирается тот, который имеет наибольший коэффициент β при условии, что β является значимым. Выбранный предиктор также включается в уравнение регрессии. Процесс продолжается до тех пор, пока не будут выбраны все предикторы, оказывающие значимое воздействие на зависимую переменную (имеющие статистически достоверные коэффициенты β). По умолчанию SPSS продолжает выбирать независимые переменные до тех пор, пока уровень значимости (p) коэффициентов

β не превысит значения 0,05. Разумеется, при желании вы можете изменить величину порогового уровня значимости.

Условия получения приемлемых результатов анализа

Регрессия, как и корреляция, анализирует линейные зависимости. В предыдущей главе мы рассмотрели процедуру оценки криволинейных зависимостей в контексте простого регрессионного анализа. Большая часть того, что было справедливо для простой регрессии, оказывается применимым и для множественного анализа. Если теория или статистический расчет показывает, что между критерием и одним или несколькими предикторами существует криволинейная зависимость, то можно ввести в качестве дополнительного предиктора, к примеру, квадрат какой-либо из независимых переменных. О том, что для этого нужно сделать, подробно рассказано в предыдущей главе.

В заключение обзора множественной регрессии рассмотрим основные условия, выполнение которых способствует получению действительно ценных и концептуально осмысленных результатов анализа:

- ▶ Исследование должно быть продумано по форме и исполнению. Анализ регрессии для не связанных по смыслу величин приводит к бесполезным результатам. К сожалению, описание условий, которые делают исследование продуманным по форме и исполнению, выходит за рамки темы данной книги.
- ▶ Для того чтобы существующие корреляции были признаны значимыми, необходимо иметь достаточные размеры выборок. Трудно указать точные границы достаточно большой выборки, однако, как правило, проблемы со значимостью начинают возникать при $N < 50$. Разумеется, чем большее число переменных вы привлекаете для анализа, тем больший размер выборки требуется для получения значимых результатов.
- ▶ Данные должны быть корректными и не должны содержать ошибок.
- ▶ Распределение значений предикторов должно быть близким к нормальному. Желательно, чтобы значения асимметрий и эксцессов по модулю не превосходили 1. Тем не менее можно получить весьма точные результаты, если это требование не выполняется строго для каждого из предикторов и даже если в анализ входит дискретная переменная с небольшим числом значений. Нормальность распределения зависимой переменной также желательна, однако допустимы как отклонения от нормальности, так и использование дискретных переменных с малым числом значений.
- ▶ Наиболее жестким требованием является запрет на использование зависимых переменных, корреляции между которыми близки к 1 (–1). Для проверки этого требования можно использовать статистики коллинеарности.

Множественный регрессионный анализ является весьма сложной процедурой. И нескольких страниц краткого обзора явно недостаточно для его изучения и исчерпывающего понимания. Поэтому желательно, чтобы перед проведением множественного регрессионного анализа у вас были более глубокие знания о нем.

После беглого обзора простого и множественного регрессионного анализа мы можем приступить к его практической реализации с помощью программы SPSS, а также к интерпретации результатов.

Пошаговые алгоритмы вычислений

Файл данных, который мы будем использовать в приводимых далее примерах, называется `help.sav`. Число объектов в этом файле 46, то есть $N = 81$. Далее перечислены переменные файла, которые мы будем использовать:

- ▶ **помощь** — зависимая переменная, интерпретируемая как время (в секундах) оказания помощи партнеру (среднее — 30, стандартное отклонение — 10);
- ▶ **симпатия** — оценка своей симпатии к партнеру, нуждающемуся в помощи (по 20-балльной шкале);
- ▶ **агрессия** — оценка своей агрессивности к партнеру (по 20-балльной шкале);
- ▶ **польза** — самооценка пользы от оказываемой помощи (по 20-балльной шкале);
- ▶ **проблема** — оценка серьезности проблемы своего партнера (по 20-балльной шкале);
- ▶ **эмпатия** — оценка эмпатии (склонности к сопереживанию) как результат тестирования (по 10-балльной шкале).

Для проведения множественного регрессионного анализа сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае `help.sav`). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом `help.sav`.

ШАГ 4

В меню Анализ выберите команду Регрессия ▶ Линейная. На экране появится диалоговое окно Линейная регрессия (рис. 18.1).

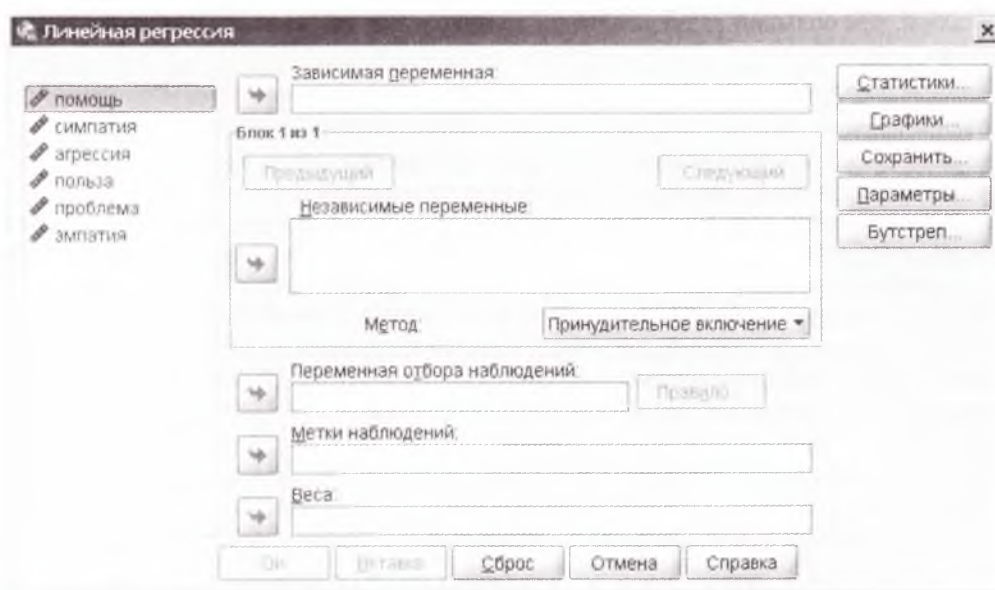


Рис. 18.1. Диалоговое окно Линейная регрессия

В верхней части диалогового окна расположено поле Зависимая переменная, предназначенное для указания единственной переменной-критерия. Ниже — поле Независимые переменные, заполняемое одной или несколькими независимыми переменными, количество которых теоретически не ограничено. С помощью кнопки Следующий, расположенной справа от названия области Блок 1 из 1, вы можете задать не один, а несколько наборов предикторов и тем самым провести одновременно несколько вариантов регрессионного анализа. Как только вы определите все независимые переменные и установите необходимые параметры анализа, щелчок на кнопке Следующий приведет к очистке поля Независимые переменные, которое будет готово к принятию нового набора предикторов. После того как вы зададите все блоки независимых переменных и щелкнете на кнопке ОК, SPSS выполнит все варианты регрессионного анализа и сгенерирует результаты для каждого из них. Обратите внимание на то, что во всех вариантах анализа, проводимых одновременно, должна быть общая зависимая переменная, иначе процедуру регрессии каждый раз придется запускать заново.

Важным элементом диалогового окна Линейная регрессия является раскрывающийся список Метод. Пункты этого списка определяют алгоритмы включения независимых переменных в уравнение регрессии:

- ▶ Принудительное включение — метод, применяющийся по умолчанию. Все независимые переменные включаются в уравнение независимо от степени их корреляции с переменной-критерием.
- ▶ Включение — пошаговое включение переменных с проверкой на значимость их частной корреляции с критерием. В результате в уравнение включаются все переменные, имеющие значимую частную корреляцию с переменной-критерием. Включение производится в порядке возрастания r -уровня.
- ▶ Исключение — пошаговый метод, сначала включающий в уравнение регрессии все независимые переменные, а затем поочередно удаляющий те из них, чья корреляция с критерием имеет уровень значимости выше заданного порогового значения. Как правило, пороговым значением является $p = 0,1$.
- ▶ Шаговый отбор — комбинация пошаговых методов включения и исключения. Основной идеей является изменение доли влияния независимой переменной на критерий при появлении в уравнении других независимых переменных. Если влияние какой-либо из включенных переменных становится слишком слабым, она исключается из уравнения. Этот метод при регрессионном анализе используется наиболее часто.
- ▶ Блочное исключение — это метод принудительного удаления переменных. Он требует предварительного задания метода Включение в качестве предыдущего блока, например Блок 1 из 1. При задании следующего блока, в данном случае Блок 2 из 2, вы сможете ввести в список Независимые переменные те независимые переменные, которые хотите исключить из уравнения регрессии. При выполнении команды получите результат со всеми заданными переменными, а затем — результат с удаленными переменными. Если в анализе участвуют несколько блоков, можно задавать операцию удаления после каждого из них.

Поле и кнопка Переменная отбора наблюдений дадут возможность выбрать группирующую переменную для задания подгруппы наблюдений, в отношении которой будет проводиться анализ.

Кнопка Графики используется для задания графического отображения остатков. Кнопка Статистики открывает диалоговое окно Линейная регрессия: Статистики (рис. 18.2). По умолчанию в окне установлены два флажка. Флажок Оценки включает в вывод коэффициенты B , стандартные коэффициенты регрессии β , а также соответствующие стандартные ошибки, t -критерии и уровни значимости. Флажок Согласие модели генерирует значения множественного коэффициента корреляции R , R^2 , таблицу дисперсионного анализа, соответствующие F -величины и уровни их значимости. Таким образом, указанные флажки отвечают за генерацию основных элементов регрессионного анализа.

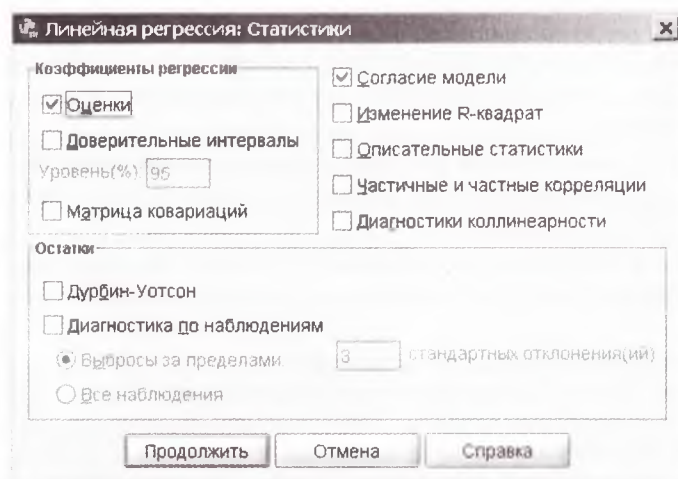


Рис. 18.2. Диалоговое окно Линейная регрессия: Статистики

Далее пояснено назначение наиболее важных флажков диалогового окна Линейная регрессия: Статистики:

- ▶ Доверительные интервалы — включает в вывод для коэффициентов B доверительный интервал 95 %;
- ▶ Матрица ковариаций — генерирует таблицу, под главной диагональю которой расположены ковариации, на главной диагонали — дисперсии, а над главной диагональю — корреляции;
- ▶ Изменение R-квадрат — для методов Включение и Шаговый отбор указывает изменение коэффициента R^2 при введении новых переменных в уравнение регрессии;
- ▶ Описательные статистики — включает в себя средние значения переменных, стандартные отклонения, а также корреляционную матрицу;
- ▶ Диагностики коллинеарности — устанавливает наличие коллинеарностей (корреляций, близких к 1) между переменными.

Щелчок на кнопке Сохранить в диалоговом окне Линейная регрессия приводит к открытию диалогового окна Линейная регрессия: Сохранить (рис. 18.3). В нем множество флажков, названия которых, как правило, представляют загадку даже для людей, искушенных в математике. Однако некоторые весьма полезны.

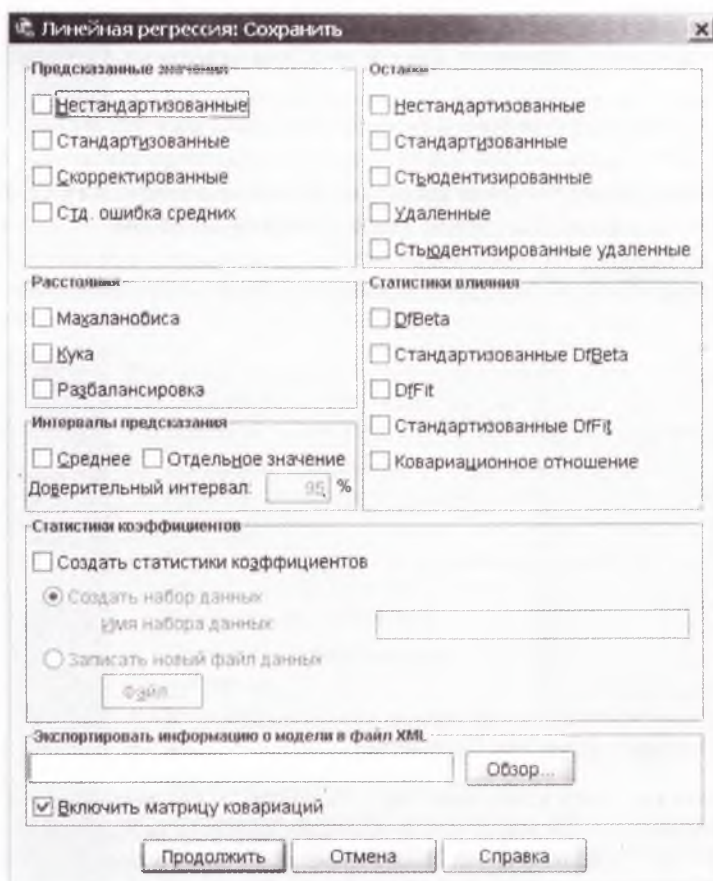


Рис. 18.3. Диалоговое окно Линейная регрессия: Сохранить

Данное окно позволяет создать в файле данных новые переменные, содержащие значения, соответствующие установленным флажкам:

- В области Предсказанные значения 4 флажка. Флажок Нестандартизованные генерирует прогнозируемые значения, которые бывает полезно сравнить с фактическими значениями для оценки адекватности уравнения регрессии. Кроме того, он позволяет получать прогнозы (оценки) зависимой переменной для тех объектов, для которых ее истинные значения неизвестны. Флажок Стандартизованные позволяет рассчитывать стандартизированные прогнозируемые значения (в z-значениях).
- В область Остатки включены 5 флажков, позволяющих задавать сохраняемые значения остатков.
- Флажки в области Статистики влияния позволяют исключать из выборки те или иные объекты. Так, если в команде спортсменов-бегунов один пробегает дистанцию гораздо хуже или гораздо лучше других, его результаты значительно искажают статистические показатели всей команды. Иногда подобные значения (выбросы) желательно исключать из анализа. К сожалению, подробное изложение этой процедуры выходит за пределы темы данной книги.

- С помощью поля и флажков из области Интервалы предсказания можно изменять число процентов в доверительном интервале для средних или отдельных значений (по умолчанию 95 %).
- Флажки в области Расстояния предоставляют 3 способа измерения расстояния между объектами (см. главу 22).

Наконец, рассмотрим диалоговое окно Линейная регрессия: Параметры (рис. 18.4). Это окно открывается при щелчке на кнопке Параметры в окне Линейная регрессия.

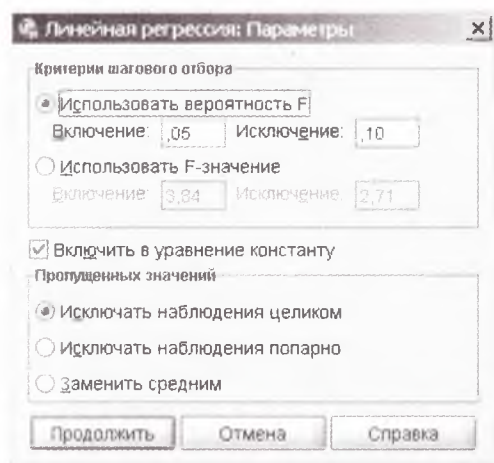


Рис. 18.4. Диалоговое окно Линейная регрессия: Параметры

Флажок Включить в уравнение константу установлен по умолчанию, и без веских на то причин снимать его не рекомендуется. Переключатель Критерии шагового отбора позволяет управлять выполнением методов Включение, Исключение и Шаговый отбор. В поле Включение вы можете указать пороговую величину значимости для включения переменных в уравнение регрессии (по умолчанию 0,05), а в поле Исключение — для исключения переменных (по умолчанию 0,1). Переключатель Пропущенных значений позволяет выбрать способ обработки отсутствующих значений (см. раздел «Обработка пропущенных значений» в главе 4).

Далее приведены два примера (шаги 5 и 5а), в первом из которых выполняется множественный регрессионный анализ с участием зависимой переменной помощь и пяти предикторов: симпатия, проблема, эмпатия, польза и агрессия. Для составления уравнения регрессии воспользуемся установленным по умолчанию методом Принудительное включение. Во втором примере мы используем некоторые из описанных ранее параметров и пошаговый метод Шаговый отбор.

ШАГ 5

После выполнения шага 4 у вас должно быть открыто диалоговое окно Линейная регрессия (см. рис. 18.1).

1. Щелкните сначала на переменной помощь, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.
2. Щелкните сначала на переменной симпатия, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Независимые переменные.

- 
3. Повторите предыдущее действие для переменных проблема, эмпатия, польза и агрессия.
 4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В результате программа сгенерирует данные, показывающие, какая из независимых переменных оказывает наибольшее влияние на зависимую переменную.

В следующем примере мы проведем регрессионный анализ с участием тех же переменных, что и в предыдущем, однако будем использовать метод Шаговый отбор, включим в результат статистики для коэффициентов B , описательные статистики и характеристики модели.

ШАГ 5а

После выполнения шага 4 должно быть открыто диалоговое окно Линейная регрессия (см. рис. 18.1). Если вы уже успели поработать с этим окном, очистите его щелчком на кнопке Сброс и выполните следующие действия:

1. Щелкните сначала на переменной помощь, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.
2. Щелкните сначала на переменной симпатия, чтобы выделить ее, а затем на второй сверху кнопке со стрелкой, чтобы переместить переменную в поле Независимые переменные.
3. Повторите предыдущее действие для переменных проблема, эмпатия, польза и агрессия.
4. В раскрывающемся списке Метод выберите пункт Шаговый отбор.
5. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Линейная регрессия: Статистики (см. рис. 18.2).
6. Установите флажок Описательные статистики и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Линейная регрессия.
7. Щелкните на кнопке Сохранить, чтобы открыть диалоговое окно Линейная регрессия: Сохранить (см. рис. 18.3).
8. Установите флажок Нестандартизованные и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Линейная регрессия.
9. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В результате выполнения приведенных инструкций будут сгенерированы данные, позволяющие судить о том, какая из независимых переменных оказывает наибольшее влияние на критерий. При составлении уравнения регрессии сначала в него включаются переменные, чья частная корреляция (β) с зависимой переменной имеет уровень значимости не выше 0,05. Если затем обнаружится, что какие-либо из включенных переменных обнаруживают новый уровень значимости, превышающий 0,1, они исключаются из уравнения. Кроме того, в результате выполнения процедуры будет создана переменная для хранения прогнозируемых значений переменной помощь, рассчитанных по составленному уравнению регрессии. В окне вывода вы также сможете найти корреляционную матрицу для всех переменных и описательные статистики.

После выполнения шагов 5 и 5а программа автоматически активизирует окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку

меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Множественный регрессионный анализ

Представленные на рис. 18.5 данные генерируются программой при выполнении шага 5. В таблицах представлены не все, а только окончательные результаты. Для удобства восприятия визуальная форма некоторых таблиц немного изменена без ущерба для их содержания.

В уравнение регрессии включены все пять предикторов. Коэффициент множественной корреляции R отражает связь зависимой переменной помощь с совокупностью независимых переменных и равен 0,598. Значение R^2 составляет 0,358 и показывает, что 35,8 % дисперсии переменной помощь обусловлены влиянием предикторов. Стандартные коэффициенты регрессии β отражают относительную степень влияния каждого из предикторов, но ни один из них не достигает статистической значимости ($p > 0,05$). Следовательно, вклад предикторов в оценку зависимой переменной не может быть интерпретирован и результат имеет сомнительную ценность.

Сводка для модели

Модель	R	R квадрат	Скорректированный R квадрат	Стд ошибка оценки
1	,598 ^a	,358	,277	8,627

a. Предикторы: (константа) эмпатия, проблема, польза, симпатия, агрессия

Дисперсионный анализ^b

Модель		Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
1	Регрессия	1657,676	5	331,535	4,454	,003 ^a
	Остаток	2977,128	40	74,428		
	Итого	4634,804	45			

a. Предикторы: (константа) эмпатия, проблема, польза, симпатия, агрессия

b. Зависимая переменная: помощь

Коэффициенты^a

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стд. ошибка	Бета		
1	(Константа)	-8,709	8,475		-1,028	,310
	симпатия	,844	,527	,227	1,601	,117
	агрессия	,832	,619	,192	1,344	,186
	польза	1,029	,626	,226	1,643	,108
	проблема	,572	,506	,162	1,131	,265
	эмпатия	,978	1,198	,116	,816	,419

a. Зависимая переменная: помощь

Рис. 18.5. Фрагменты окна вывода после выполнения шага 5

Изменение показателей при пошаговом добавлении переменных

Представленные на рис. 18.6 результаты генерируются программой при выполнении шага 5а. Как и в предыдущем случае, мы максимально адаптировали их для лучшего восприятия. Полученные данные показывают, каким образом изменялись характеристики регрессионного анализа при поэтапном включении новых предикторов в уравнение регрессии.

С возрастанием числа предикторов значения R , R^2 и исправленное значение R^2 возрастают, в то время как значение стандартной ошибки убывает. Та же тенденция наблюдается и в отношении числа степеней свободы, суммы квадратов и среднего квадратов.

Как можно видеть, в результате применения пошагового метода из пяти предикторов в уравнение регрессии включены лишь три (модель 3): симпатия, агрессия и польза. Коэффициент множественной корреляции R отражает связь зависимой переменной помощи с совокупностью независимых переменных и равен 0,571. Значение R^2 составляет 0,326 и показывает, что 32,6 % дисперсии переменной помощи обусловлены влиянием предикторов. Стандартные коэффициенты регрессии β являются статистически достоверными, что позволяет интерпретировать относительную степень влияния каждого из предикторов; для переменной симпатия $\beta = 0,278$, а для переменных агрессия и польза $\beta = 0,276$ и $\beta = 0,269$ соответственно. Каждая из независимых переменных вносит примерно одинаковый вклад в оценку зависимой переменной и коррелирует с ней положительно.

Сводка для модели^d

Модель	R	R квадрат	Скорректированный R квадрат	Стд ошибка оценки
1	,416	,173	,154	9,335
2	,508	,258	,224	8,941
3	,571	,326	,278	8,622

^d Зависимая переменная: помощь

Дисперсионный анализ^d

Модель		Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
3	Регрессия	1512,813	3	504,271	6,784	,001 ^c
	Остаток	3121,991	42	74,333		
	Итого	4634,804	45			

^c Предикторы: (константа) симпатия, польза, агрессия

^d Зависимая переменная: помощь

Коэффициенты

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стд. ошибка	Бета		
3	(Константа)	-5,315	8,075		-,658	,514
	симпатия	1,033	,500	,278	2,065	,045
	польза	1,267	,603	,276	2,083	,043
	агрессия	1,168	,567	,269	2,059	,046

Рис. 18.6. Фрагменты окна вывода после выполнения шага 5а

Терминология, используемая при выводе

Далее дана трактовка терминов, используемых программой в окне вывода:

- ▶ Вероятность F-включения — максимальный уровень значимости переменных, *вводимых* в уравнение регрессии, в данном случае равный $p = 0,050$.
- ▶ R — коэффициент множественной корреляции, отражающий связь совокупности предикторов симпатия, агрессия и польза с критерием помощь.
- ▶ R квадрат — коэффициент детерминации (R^2), равный доле дисперсии зависимой переменной помощь, обусловленной влиянием независимых переменных симпатия, агрессия и польза.
- ▶ Скорректированный R квадрат — исправленная величина R^2 . Величина R^2 , используемая в расчетах, на практике оказывается несколько завышенной. Исправленная величина R^2 ближе к реальным результатам.
- ▶ Стд. ошибка оценки — в таблице Сводка для модели стандартное отклонение ожидаемого значения переменной помощь. Как видно из приводимых данных, с добавлением каждой новой независимой переменной в уравнение регрессии эта величина уменьшается.
- ▶ Регрессия — статистика, отражающая влияние предикторов на зависимую переменную.
- ▶ Остаток — статистика, отражающая внешнее (не обусловленное предикторами) влияние на независимую переменную.
- ▶ B — нестандартизированные коэффициенты и константа уравнения регрессии, связывающего критерий и предикторы:

$$\text{помощь}_{\text{прогноз}} = -5,3147 + 1,0328 (\text{симпатия}) + 1,1676 (\text{агрессия}) + 1,2569 (\text{польза}).$$
- ▶ Стд. ошибка — в таблице Коэффициенты является мерой стабильности коэффициентов B и равна стандартному отклонению их значений, рассчитанных для большого числа выборок.
- ▶ Бета — стандартизованный коэффициент регрессии (β), представляющий собой коэффициенты B для независимых переменных, представленных в z-шкале. Для линейных взаимодействий β по абсолютному значению не превосходит 1, для криволинейных взаимодействий это условие не является обязательным.
- ▶ t — отношение коэффициента B к своей стандартной ошибке.
- ▶ Бета включения — значения коэффициента β для переменных, не включенных в уравнение регрессии в предположении, что они в него включены.
- ▶ Частная корреляция — коэффициенты частной корреляции для переменных, входящих в уравнение регрессии. Наличие в этом уравнении нескольких коррелирующих переменных взаимно снижает их частную корреляцию.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его

редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

19 Анализ надежности

246	Коэффициент альфа
247	Надежность половинного расщепления
247	Пошаговые алгоритмы вычислений
252	Представление результатов
256	Завершение анализа и выход из программы

Анализ надежности, рассматриваемый в этой главе, имеет отношение к надежности измерений в психологии и других социальных науках. Этот анализ еще известен как анализ заданий (пунктов теста) и применяется для отбора наиболее пригодных вопросов или заданий измерительной методики (вопросника, анкеты, теста). Обычно при разработке методики сначала составляют ее предварительный вариант, включающий в себя избыточное количество заданий (пунктов), который опробуется на достаточно представительной выборке респондентов. Затем проводят анализ надежности, который позволяет при помощи многочисленных критериев исключить неподходящие задания.

Мы коснемся двух показателей надежности: *коэффициента Кронбаха* (Chronbach), обозначаемого буквой α (альфа), и *надежности половинного расщепления* (split-half reliability). Существуют и другие типы надежности, однако ограничимся лишь их упоминанием в разделе пошаговых процедур, а основное внимание сосредоточим на наиболее употребительном коэффициенте α . После описаний двух видов надежности приведем примеры, иллюстрирующие их применение.

Коэффициент альфа

Коэффициент α является мерой внутренней согласованности, или однородности, измерительной шкалы. Как правило, α лежит в пределах от 0 до 1, однако может принимать и отрицательные значения. Последние свидетельствуют о том, что часть элементов, или пунктов, шкалы измеряет противоположные величины. Чем ближе коэффициент α к 1, тем выше внутренняя согласованность системы заданий. Коэффициент α можно рассматривать как показатель корреляции между измеренными признаками для выборок равного объема. Если, к примеру, составить 10 тестов с сотней случайных вопросов из банка данных объемом 1000 вопросов, коэффициент α будет характеризовать корреляцию между этими тестами. Формула для вычисления коэффициента α выглядит следующим образом:

$$\alpha = \frac{kr}{1 + (k-1)r},$$

где k — число пунктов шкалы; r — средний коэффициент корреляции между каждым пунктом и суммой остальных пунктов.

Из приведенной формулы следует, что коэффициент α возрастает как с увеличением числа пунктов шкалы, так и с усилением корреляции между пунктами.

Применение альфы Кронбаха основано на модели, предполагающей наличие большей дисперсии у более надежного теста: чем надежнее тест, тем больше чувствительность (различительная способность) пунктов теста.

Если в методике применяются задания дихотомического типа (да — нет, правильно — неправильно), формула для вычисления коэффициента α Кронбаха (Chronbach) становится идентична формуле Кюдера — Ричардсона (Kuder & Richardson).

Надежность половинного расщепления

Надежность половинного расщепления вычисляется в случае, когда число элементов достаточно велико для того, чтобы можно было разделить шкалу пополам. Рассмотрим тест 16 личностных свойств Раймонда Кеттелла. Он состоит из 187 вопросов, из которых от 10 до 14 вопросов определяют одно из 16 свойств. Разумеется, было бы бессмысленно делить на две половины все 187 вопросов, поскольку в каждой половине оказалось бы множество вопросов, относящихся к разным свойствам, а следовательно, несопоставимых друг с другом. Более разумным подходом было бы разделить пополам совокупность вопросов, измеряющих одно свойство: например, из 14 вопросов, определяющих тревожность, в случайном порядке выбрать 7 и сформировать первую группу, а из оставшихся — вторую группу. В качестве показателя надежности используется коэффициент корреляции Спирмена — Брауна между двумя образованными половинами.

Пошаговые алгоритмы вычислений

Для проведения анализа надежности мы будем использовать файл данных TestA.sav, содержащий 12 переменных для 120 объектов (респондентов). Переменные имеют имена a1, ..., a12 и соответствуют пунктам разрабатываемой семантической шкалы для оценки степени альтруизма. Каждая переменная — это оценка респондентом другого человека по наличию (1 — да) или отсутствию (0 — нет) у него одной из 12 характеристик. Список этих характеристик (прилагательных) вы можете найти в разделе «Представление результатов».

Чтобы приступить к анализу надежности, сначала нужно выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае TestA.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом TestA.sav.

ШАГ 4

В меню Анализ выберите команду Шкалирование ► Анализ надежности. Откроется диалоговое окно Анализ надежности (рис. 19.1).

Параметры команды анализа надежности настраиваются при помощи двух диалоговых окон: основного, изображенного на рис. 19.1, и Анализ надежности: Статистики (рис. 19.2), открывающегося после щелчка на кнопке Статистики.

Основное окно команды Анализ надежности почти не отличается от диалоговых окон большинства команд статистических процедур SPSS. Слева расположен список доступных переменных текущего файла данных, а внизу находятся 5 стандартных кнопок. Поле Пункты предназначено для ввода имен переменных, включаемых в анализ надежности.

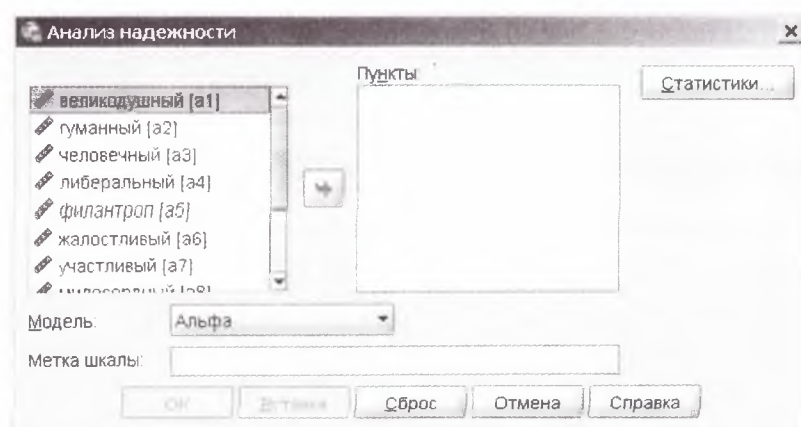


Рис. 19.1. Диалоговое окно Анализ надежности

Раскрывающийся список Модель позволяет задать модель анализа надежности, по умолчанию выбран пункт Альфа. Далее приводятся описания остальных пунктов списка:

- ▶ Расщепления пополам — программа разбивает переменные на 2 группы в порядке их перечисления (число переменных в группах одинаково, если общее их число четно, в противном случае первая группа содержит на одну переменную больше) и сравнивает эти группы между собой.
- ▶ Гутмана — анализ основан на операции вычисления нижних границ.
- ▶ Параллельная — анализ проводится в предположении, что все переменные имеют равные дисперсии.
- ▶ Строго параллельная — анализ проводится в предположении, что все переменные имеют равные средние значения, дисперсии фактических значений и дисперсии ошибок.

Поле Метка шкалы позволяет назначить имя каждому заданному набору пунктов и варианту анализа. Если оставить его пустым, то по умолчанию в качестве метки шкалы будет использовано имя Шкальные: Все переменные.

Как уже упоминалось, при щелчке на кнопке Статистики открывается диалоговое окно Анализ надежности: Статистики. В нем вы можете задать параметры команды Анализ надежности. Далее будут приведены описания всех флажков данного диалогового окна, за исключением тех, которые находятся в области Таблица дисперсионного анализа. Дисперсионный анализ (ANOVA) и критерий χ^2 рассматривались нами ранее, кроме того, эти виды анализа весьма редко используются в контексте анализа надежности. Остальные флажки перечислены в порядке их размещения в областях Описательные для, Итоговые статистики, Межпунктовые. Кроме того, последними даны описания флажков Т-квадрат Хотеллинга и Критерий аддитивности Тьюки.

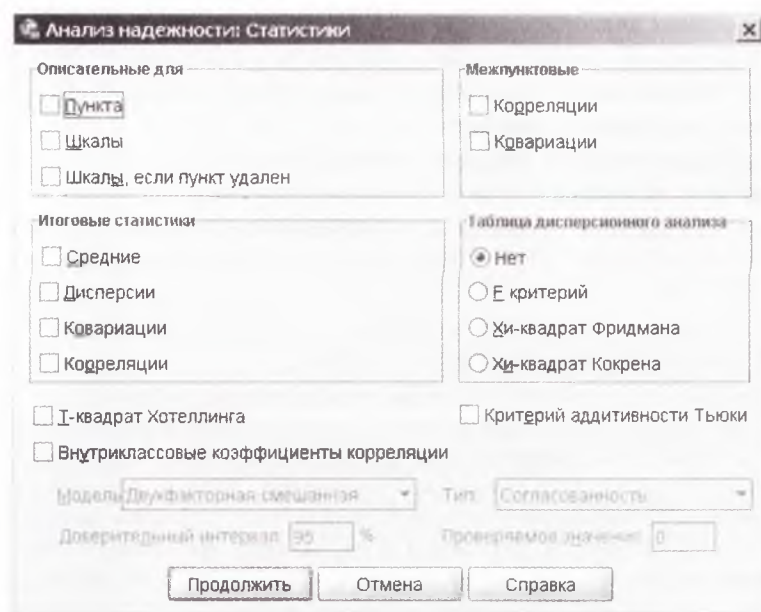


Рис. 19.2. Диалоговое окно Анализ надежности: Статистики

- ▶ Пункта — вычисление средних значений и стандартных отклонений переменных, участвующих в анализе.
- ▶ Шкалы — включение в вывод среднего значения, дисперсии, стандартного отклонения и размера выборки для суммы всех переменных шкалы. Под суммой переменных понимается сумма значений всех 12 переменных (пунктов) для каждого объекта. Таким образом, обеспечивается вывод статистик для переменной, полученной в результате суммирования.
- ▶ Шкалы, если пункт удален — включение в вывод для каждой переменной значения коэффициента α , подсчитанного для всей шкалы в предположении, что данная переменная исключена. В большинстве случаев этот флажок следует устанавливать.
- ▶ Средние — вычисление средних значений по всем наблюдениям для каждой переменной. В результат включаются все средние значения, минимальное и максимальное из этих значений, размах, отношение максимального среднего к минимальному, а также дисперсия средних значений. Несмотря на некоторую сложность, этот флажок позволяет получить важные для анализа надежности результаты.
- ▶ Дисперсии — этот флажок почти аналогичен флажку Средние, за тем исключением, что вычисления выполняются не для средних значений, а для дисперсий.
- ▶ Ковариации в группе Итоговые статистики — вычисление ковариаций между каждой переменной и суммой всех остальных переменных. В выводимые результаты также включаются средние значения полученных ковариаций, их минимум, максимум, размах, отношение максимума к минимуму и дисперсия.
- ▶ Корреляции в группе Итоговые статистики — этот флажок отличается от флажка Ковариации лишь тем, что вычисляются не ковариации, а корреляции. Среднее корреляции

ляций для всех переменных является значением r в формуле вычисления коэффициента α .

- ▶ Корреляции в группе Межпунктовые — включение в вывод корреляционной матрицы для всех переменных, участвующих в анализе.
- ▶ Ковариации в группе Межпунктовые — создание ковариационной матрицы для всех переменных, участвующих в анализе.
- ▶ Т-квадрат Хотеллинга — применение множественного сравнения по t -критерию для проверки достоверности различий средних значений всех переменных, участвующих в анализе.
- ▶ Критерий аддитивности Тьюки — применение критерия для проверки линейности зависимости между переменными, участвующими в анализе.

Далее приведены 3 примера, иллюстрирующие применение критерия надежности. На шаге 5 проводится анализ надежности с параметрами, позволяющими оценить вклад каждого пункта во внутреннюю согласованность шкалы. На шаге 5а выполняется анализ надежности с применением некоторых из описанных параметров. Данные, полученные в результате выполнения шага 5, показывают, что для 12 переменных надежность весьма невелика. Поэтому из списка на шаге 5а удаляются 3 переменные, нарушающие внутреннюю согласованность шкалы, тем самым повышается ее надежность. Наконец, на шаге 5б выполняется анализ надежности половинного расщепления с 9 из 12 исходных переменных. Результаты всех трех вариантов анализа включены в раздел «Представление результатов».

В следующем примере проводится анализ надежности с включением всех 12 пунктов.

ШАГ 5

После выполнения шага 4 должно быть открыто диалоговое окно Анализ надежности (см. рис. 19.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Перенесите группу переменных $a_1, \dots, a_{12}$ из левого в правое поле. Для этого щелкните сначала на переменной a_1 , чтобы выделить ее. Затем нажмите клавишу Shift и, удерживая ее, щелкните на переменной a_{12} , а затем — на кнопке со стрелкой, чтобы переместить переменные в поле Пункты.
2. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Анализ надежности: Статистики (см. рис. 19.2).
3. В области Описательные для установите флажок Шкалы, если пункт удален. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Анализ надежности.
4. Щелкните на кнопке ОК, чтобы открыть окно вывода.

При выполнении этого шага все переменные попадают в поле Пункты в том порядке, в котором они были в исходном файле. В вывод вместе с именами переменных включаются метки, представляющие собой соответствующие вопросы теста, а также коэффициент α и объем выборки N . Кроме того, в окне вывода будут представлены результаты расчета шкалы альфа для каждого пункта в предположении, что он из шкалы удален.

Далее приведен пример анализа надежности с участием 9 переменных, для которых предыдущий анализ выявил наибольшую надежность ($a_1, a_2, a_3, a_4, a_7, a_8, a_9, a_{10}, a_{12}$). Кроме того, мы зададим некоторые дополнительные параметры и снова включим в выводимые результаты метки переменных.

ШАГ 5а

После выполнения шага 4 должно быть открыто диалоговое окно Анализ надежности (см. рис. 19.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной a_1 , чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в поле Пункты.
2. Повторите предыдущее действие для переменных $a_2, \dots, a_4, a_7, \dots, a_{10}$ и a_{12} . Поскольку переменные $a_1, \dots, a_4$ в исходном списке расположены последовательно, для их одновременного выделения можно щелкнуть на переменной a_1 , нажать клавишу Shift и щелкнуть на переменной a_4 . То же касается переменных $a_7, \dots, a_{10}$.
3. В поле Метка шкалы введите текст, соответствующий применяемому методу и участвующим в анализе переменным: Альфа Кронбаха, пункты a_1 - a_4, a_7 - a_{10}, a_{12} .
4. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Анализ надежности: Статистики (см. рис. 19.2).
5. В области Описательные для установите флажки Шкалы и Шкалы, если пункт удален, в области Итоговые статистики — Средние, Дисперсии и Корреляции, в области Межпунктовые — Корреляции. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Анализ надежности.
6. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В следующем примере мы выполним ту же процедуру, что и на шаге 5а, однако вместо коэффициента α будем рассчитывать надежность половинного расщепления.

ШАГ 5б

После выполнения шага 4 должно быть открыто диалоговое окно Анализ надежности (см. рис. 19.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной a_1 , чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в список Пункты.
2. Повторите предыдущее действие для переменных $a_2, \dots, a_4, a_7, \dots, a_{10}$ и a_{12} . Поскольку переменные $a_1, \dots, a_4$ в исходном списке расположены последовательно, для их одновременного выделения можно щелкнуть на переменной a_1 , нажать клавишу Shift и щелкнуть на переменной a_4 . То же касается переменных $a_7, \dots, a_{10}$.
3. Раскройте список Модель и выберите пункт Расщепления пополам.
4. В поле Метка шкалы введите текст, соответствующий применяемому методу и участвующим в анализе переменным: Половинного расщепления, пункты a_1 - a_4, a_7 - a_{10}, a_{12} .
5. Щелкните на кнопке Статистики, чтобы открыть диалоговое окно Анализ надежности: Статистики (см. рис. 19.2).
6. В области Описательные для установите флажок Шкалы, в области Итоговые статистики — Средние, Дисперсии и Корреляции, в области Межпунктовые — Корреляции. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Анализ надежности: Статистики.
7. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шагов 5, 5а и 5б программа автоматически переходит в окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикаль-

ной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

В этом разделе представлены фрагменты выводимых данных, генерируемые программой при выполнении шагов 5, 5а и 5б. Каждый блок результатов озаглавлен в соответствии с введенными в поле Метка шкалы данными, например, после выполнения шага 5: Шкальные: Все переменные.

Коэффициент альфа с параметрами по умолчанию

Первая из таблиц, Статистики пригодности, (рис. 19.3) содержит величину Альфа Кронбаха для шкалы, содержащей все 12 переменных. Во второй таблице, Статистики соотношения пункта с суммарным баллом, для каждой переменной показаны статистические характеристики, позволяющие сравнить ее с совокупностью остальных переменных.

Шкальные: ВСЕ ПЕРЕМЕННЫЕ

Статистики пригодности	
Альфа Кронбаха	Количество пунктов
,696	12

Статистики соотношения пункта с суммарным баллом				
	Среднее шкалы при удалении пункта	Дисперсия шкалы при удалении пункта	Корреляция пункта с суммарным баллом	Альфа Кронбаха при удалении пункта
великодушный	5,68	6,389	,643	,629
гуманный	5,73	6,588	,546	,644
человечный	5,68	6,893	,425	,663
либеральный	5,73	7,125	,324	,678
филантроп	5,82	8,470	-,164	,745
жалостливый	5,81	7,971	,008	,723
участливый	5,76	6,924	,404	,666
милосердный	5,64	6,736	,501	,652
заботливый	5,72	6,759	,474	,655
благосклонный	5,75	6,727	,486	,653
ласковый	5,82	8,235	-,084	,735
добродушный	5,73	6,601	,540	,645

Рис. 19.3. Фрагмент окна вывода после выполнения шага 5

Далее дана трактовка терминов, используемых программой в окне вывода и относящихся к вычислению коэффициента α с параметрами по умолчанию:

- Среднее шкалы при удалении пункта — для каждой переменной вычисляется сумма остальных 11 переменных по всем 120 объектам файла.
- Дисперсия шкалы при удалении пункта — дисперсия суммы всех переменных с той переменной, которая удалена.

- Корреляция пунктов с суммарным баллом — коэффициент корреляции переменной с суммой остальных 11 переменных.
- Альфа Кронбаха при удалении пункта — коэффициент α , полученный при удалении соответствующей переменной из шкалы.

Из рис. 19.3 видно, что для трех переменных значения в столбце Корреляция пункта с суммарным баллом оказались весьма низкими, а для переменных a5 и a11 даже отрицательными. Для увеличения коэффициента α (указанного в таблице Статистики пригодности) необходимо изъять из рассмотрения такие переменные, после удаления которых величина Альфа Кронбаха больше, чем величина α для всей шкалы. Анализ, проведенный после исключения этих трех переменных, действительно позволил найти новое, более высокое значение коэффициента α . Таким образом, удалось выявить, что наибольшее значение α принимает в случае удаления из шкалы переменных с номерами 5, 6, и 11. Обратите внимание на то, что подобные переменные программа не исключает автоматически: существуют причины, по которым иногда требуется сохранять их в шкале.

Коэффициент альфа после исключения переменных и задания дополнительных параметров

Приведенные на рис. 19.4 результаты формируются после удаления из шкалы трех упомянутых переменных (шаг 5а). Таблица Статистики пригодности содержит величину Аль-

Шкальные: Альфа Кронбаха, пункты a1-a4, a7-a10, a12

Статистики пригодности		
Альфа Кронбаха	Альфа Кронбаха, основанная на стандартизованных пунктах	Количество пунктов
,825	,825	9

Сводка статистик пункта						
	Среднее	Минимум	Максимум	Размах	Максимум / Минимум	Дисперсия
Средние пунктов	,547	,500	,617	,117	1,233	,002
Дисперсии пунктов	,248	,238	,262	,014	1,058	,000
Межпунктовые корреляции	,344	,134	,632	,499	4,733	,013

Статистики соотношения пункта с суммарным баллом					
	Среднее шкалы при удалении пункта	Дисперсия шкалы при удалении пункта	Корреляция пункта с суммарным баллом	Квадрат коэффициента множественной корреляции	Альфа Кронбаха при удалении пункта
великодушный	4,34	6,462	,668	,549	,791
гуманный	4,39	6,492	,645	,487	,793
человечный	4,34	6,950	,458	,331	,815
либеральный	4,39	7,081	,397	,283	,822
участливый	4,43	6,986	,434	,259	,818
милосердный	4,31	6,837	,515	,331	,809
заботливый	4,38	6,894	,474	,383	,814
благосклонный	4,42	6,716	,546	,362	,805
добродушный	4,40	6,545	,621	,550	,796

Рис. 19.4. Фрагмент окна вывода после выполнения шага 5а

фа Кронбаха для шкалы, содержащей 9 переменных. Вторая таблица содержит основные статистические показатели для шкалы: средние, дисперсии и корреляции. Таблица Статистики соотношения пункта с суммарным баллом по содержанию аналогична одноименной таблице, приведенной на рис. 19.3, однако здесь вместо 12 исходных переменных содержатся лишь 9. Как легко заметить, значение коэффициента α увеличилось и составляет 0,825. Данное значение α больше нельзя увеличить, удалив какую-либо из 9 оставшихся переменных, так как альфы Кронбаха при дальнейшем исключении любого из оставшихся пунктов получаются меньше, чем α , полученная по всем девяти пунктам. Таким образом, оставшиеся 9 пунктов образуют оптимальный состав шкалы.

Далее дана трактовка новых терминов, используемых программой в окне вывода и относящихся к вычислению коэффициента α после исключения нескольких переменных и задания дополнительных параметров:

- ▶ Средние пунктов — описательная информация о средних значениях девяти пунктов шкалы. Среднее полученных средних значений равно 0,547, наименьшее из 9 средних равно 0,5 и т. д.
- ▶ Дисперсии пунктов — статистика, аналогичная средним элементов, но полученная для их дисперсий.
- ▶ Межпунктовые корреляции — описательная статистика, характеризующая корреляцию между каждой переменной шкалы и суммой остальных переменных. Для каждой переменной вычисляется коэффициент корреляции, а в окне вывода приводятся среднее значение, минимум и т. д. полученных коэффициентов. Среднее значение коэффициентов корреляции является величиной r в формуле расчета α .
- ▶ Квадрат коэффициента множественной корреляции — представленные в этой секции величины вычислены с помощью уравнения множественной регрессии, позволяющего получать прогнозируемый коэффициент корреляции данной переменной с остальными переменными, участвующими в анализе.
- ▶ Альфа Кронбаха — коэффициент α , вычисляемый по формуле, приведенной в начале главы. Число элементов шкалы (k) в данном случае равно девяти. Поскольку α зависит от числа элементов шкалы, нельзя четко определить критерий приемлемости полученного значения α . Для большинства случаев можно придерживаться следующих рекомендаций по оценке внутренней согласованности шкалы:
 - ▶ $\alpha > 0,9$ — отличная;
 - ▶ $\alpha > 0,8$ — хорошая;
 - ▶ $\alpha > 0,7$ — приемлемая;
 - ▶ $\alpha > 0,6$ — сомнительная;
 - ▶ $\alpha > 0,5$ — малопригодная;
 - ▶ $\alpha < 0,5$ — недопустимая.
- ▶ Альфа Кронбаха, основанная на стандартизованных пунктах — коэффициент α , вычисляемый в предположении, что элементы шкалы заранее стандартизованы. Как можно видеть, значения обычного и стандартизованного коэффициентов α в данном случае одинаковы.

Надежность половинного расщепления

Результаты анализа надежности половинного расщепления с использованием тех же девяти переменных, что и в предыдущем случае, приведены на рис. 19.5. Как правило, такое количество переменных слишком мало для подобного анализа, тем не менее результаты вполне пригодны для интерпретации. Большая часть полученной информации по смыслу аналогична результатам, полученным для коэффициента α . Мы приводим лишь основные результаты анализа.

Обратите внимание на то, что значения α , вычисленные для каждой половины шкалы, ниже, чем для шкалы в целом. Это отражает тот факт, что при уменьшении числа элементов шкалы ее внутренняя согласованность снижается.

Далее дана трактовка нескольких новых терминов, относящихся к вычислению надежности половинного расщепления:

- ▶ Корреляция между формами — приближенное значение надежности шкалы, рассчитанное в предположении, что она содержит 5 элементов.
- ▶ Коэффициент Спирмена — Брауна: Равная длина — результат гипотетического анализа надежности половинного расщепления шкалы, содержащей 10 элементов.
- ▶ Коэффициент Спирмена — Брауна: Неравная длина — результат гипотетического анализа надежности половинного расщепления шкалы, содержащей не 10, а 9 элементов, что приводит к необходимости формирования двух неравных групп.
- ▶ Коэффициент половинного расщепления Гутмана — значение надежности, полученное вычислением нижних пределов.

Шкальные: Половинное расщепление, пункты a1-a4, a7-a10, a12

Статистики пригодности			
Альфа Кронбаха	Часть 1	Значение	,714
		Количество пунктов	5 ^a
	Часть 2	Значение	,695
		Количество пунктов	4 ^b
Общее количество пунктов			9
Корреляция между формами			,696
Коэффициент Спирмена-Брауна	Равная длина		,821
	Неравная длина		,822
Коэффициент половинного расщепления Гутмана			,814

a. Пункты великодушный, гуманный, человечный, либеральный, участливый
b. Пункты милосердный, заботливый, благосклонный, добродушный

Рис. 19.5. Основные статистические показатели для шкалы при анализе надежности половинного расщепления

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

20 Факторный анализ

257	Вычисление корреляционной матрицы
258	Извлечение факторов
259	Выбор и вращение факторов
261	Интерпретация факторов
261	Пошаговые алгоритмы вычислений
267	Представление результатов
271	Терминология, используемая при выводе
272	Завершение анализа и выход из программы

С конца 1950-х годов факторный анализ приобрел значительную популярность в самых различных исследованиях. Во многом этому способствовала разработка Раймондом Кеттеллем знаменитого 16-факторного личностного опросника (16PF). Именно при помощи факторного анализа ему удалось свести около 4500 наименований личностных особенностей к 187 вопросам, которые, в свою очередь, позволяют измерить 16 различных свойств личности. Факторный анализ дает возможность количественно определить нечто, непосредственно не измеряемое, исходя из нескольких доступных измерению переменных. Например, характеристики «посещает развлекательные мероприятия», «много разговаривает», «охотно идет на контакт с любым незнакомым человеком» могут служить оценками качества «общительность», которое непосредственно не поддается количественному измерению. Факторный анализ позволяет установить для большого числа исходных признаков сравнительно узкий набор «свойств», характеризующих связь между группами этих признаков и называемых факторами. Процедура факторного анализа состоит из четырех основных стадий:

1. Вычисление корреляционной матрицы для всех переменных, участвующих в анализе.
2. Извлечение факторов.
3. Вращение факторов для создания упрощенной структуры.
4. Интерпретация факторов.

Первые 3 операции мы кратко рассмотрим в следующих нескольких разделах, последняя операция описывается лишь на концептуальном уровне.

Как и многие другие методы, рассматриваемые в этой книге, факторный анализ является весьма сложной процедурой. Несколько страниц краткого обзора явно недостаточно для его изучения и исчерпывающего понимания. Поэтому желательно, чтобы перед проведением факторного анализа у вас уже были более глубокие знания.

Вычисление корреляционной матрицы

Первая операция, которая производится при выполнении факторного анализа, — вычисление корреляционной матрицы для переменных, участвующих в анализе. Уже из

этого можно сделать вывод о том, что факторный анализ основан на взаимодействии переменных. Для проведения факторного анализа совсем необязательно специально строить корреляционную матрицу: при необходимости программа SPSS создаст ее сама на основе данных файла. Иногда не нужны даже исходные данные — достаточно иметь корреляционную матрицу, которая в этом случае вводится в командный файл SPSS. Однако подобный подход весьма сложен и в рамках этой книги не рассматривается. При необходимости вы можете обратиться к руководству пользователя программы SPSS.

Извлечение факторов

Извлечение факторов является следующим этапом факторного анализа. С математической точки зрения извлечение факторов имеет определенную аналогию с множественным регрессионным анализом. Если вы обратитесь к главе 18, то вспомните, что первым шагом множественного регрессионного анализа является выбор той независимой переменной, которая обуславливает наибольшую долю дисперсии зависимой переменной. Затем операция повторяется для оставшихся независимых переменных до тех пор, пока добавляемая доля дисперсии не перестанет быть значимой. В факторном анализе существует аналогичная процедура, и для ее описания достаточно лишь немного изменить соответствующий раздел главы 18.

Извлечение фактора начинается с подсчета суммарного разброса значений всех участвующих в анализе переменных (данная величина чем-то похожа на общую сумму квадратов). Для этого «суммарного разброса» непросто подобрать логическую интерпретацию, однако он является вполне строго определенной математической величиной. Первой задачей факторного анализа является выбор взаимодействующих переменных, чья взаимная корреляция обуславливает наибольшую долю общей дисперсии. Эти переменные образуют *первый фактор*. Затем первый фактор исключается и из оставшегося множества переменных снова выбираются те, чье взаимодействие определяет наибольшую долю оставшейся общей дисперсии. Эти переменные образуют *второй фактор*. Процедура извлечения факторов продолжается до тех пор, пока не будет исчерпана вся общая дисперсия переменных.

По умолчанию в процедуре факторного анализа каждая переменная имеет единичное значение *общности*. Этот показатель равен доле дисперсии переменной, обусловленной совокупным влиянием факторов. Общность можно сравнить с множественным коэффициентом корреляции R , принимающим значение 0 в случае, если факторы не влияют на переменную, и значение 1 в случае, если дисперсия переменной целиком определяется выделяемыми факторами. Перед началом извлечения факторов единичное значение общности установлено по умолчанию для всех переменных, участвующих в факторном анализе.

После того как процедура извлекает первый фактор, напротив его номера появляется его *собственное значение*, например, рядом с числом 1 появляется значение 5,13312. Собственное значение фактора пропорционально доле общей дисперсии, определяемой данным фактором (обратите внимание на то, что здесь речь идет о факторе, а не о переменной, как было в случае общности). Первое собственное значение всегда является наибольшим и превышает единицу, это определяется алгоритмом работы процедуры, согласно которому факторы извлекаются в порядке убывания их влияния на дисперсию переменных. Затем вычисляются процент дисперсии, обуславливаемый данным фактором и равный отношению собственного значения фактора к числу переменных, а также соответствующий кумулятивный (накопленный) процент. С извлечением каждого ново-

го фактора собственные значения уменьшаются, а кумулятивный процент приближается к 100. Обратите внимания на то, что в контексте процедуры извлечения факторов ни разу не прозвучал термин «значимость»: в отличие от регрессионного анализа здесь извлечение факторов происходит до тех пор, пока не будет исчерпана вся дисперсия переменных, независимо от того, является значимым влияние фактора на дисперсию или нет.

Выбор и вращение факторов

Извлеченные факторы, за очень редкими исключениями, не представляют интереса для исследователя. Если факторов окажется столько же, сколько исходных переменных, факторный анализ теряет смысл, поскольку его целью является сокращение исходного набора переменных.

Итак, нужно принять решение, какие из факторов следует оставить для дальнейшего анализа. Здесь рекомендуется руководствоваться в первую очередь здравым смыслом и оставлять те факторы, которые имеют понятную теоретическую или логическую интерпретацию. Однако не всегда представляется возможным заранее установить назначение каждого фактора, поэтому исследователи на первом этапе обычно используют формальные критерии. При выполнении факторного анализа с установками по умолчанию все факторы, чьи собственные значения превышают единицу, сохраняются для дальнейшего анализа. Поскольку число факторов равно числу переменных, лишь для небольшого количества факторов собственные значения оказываются больше единицы, а значит, выполнение команды с параметрами по умолчанию позволяет радикально сократить число факторов. Существуют и другие критерии выделения факторов (например, критерий «каменистой осыпи» Р. Кеттелла), кроме того, вы можете выбирать факторы, основываясь на известных вам особенностях конкретного файла данных. В любом случае окончательное решение о числе факторов обычно принимается после интерпретации факторов, следовательно, факторный анализ предполагает неоднократное выделение различного числа факторов. В разделе пошаговых процедур рассмотрены несколько вариантов выполнения факторного анализа, отличные от принятого по умолчанию.

Следующим шагом после выделения факторов является их вращение. Вращение требуется потому, что изначально структура факторов, будучи математически корректной, как правило, трудна для интерпретации. Целью вращения является получение простой структуры, которой соответствует большое значение нагрузки каждой переменной только по одному фактору и малое — по остальным факторам. *Нагрузка* отражает связь между переменной и фактором, являясь подобием коэффициента корреляции. Значение нагрузки лежит в пределах от -1 до 1 . Идеальная простая структура предполагает, что каждая переменная имеет нулевые значения нагрузок для всех факторов, кроме одного, для которого нагрузка этой переменной близка к 1 (-1). Графическая интерпретация процедуры вращения представлена на рис. 20.1. До вращения (вверху) точки, соответствующие переменным, расположены на удалении от осей факторов. После поворота осей (внизу) переменные оказываются вблизи осей, что соответствует максимальной нагрузке каждой переменной только по одному фактору. На практике строгая ориентация переменных вдоль осей факторов обычно не достигается, однако операция поворота позволяет приблизиться к желательному результату. После выполнения вращения факторов программа SPSS включает в окно вывода похожие графические изображения, для удобства снабдив их таблицей координат.

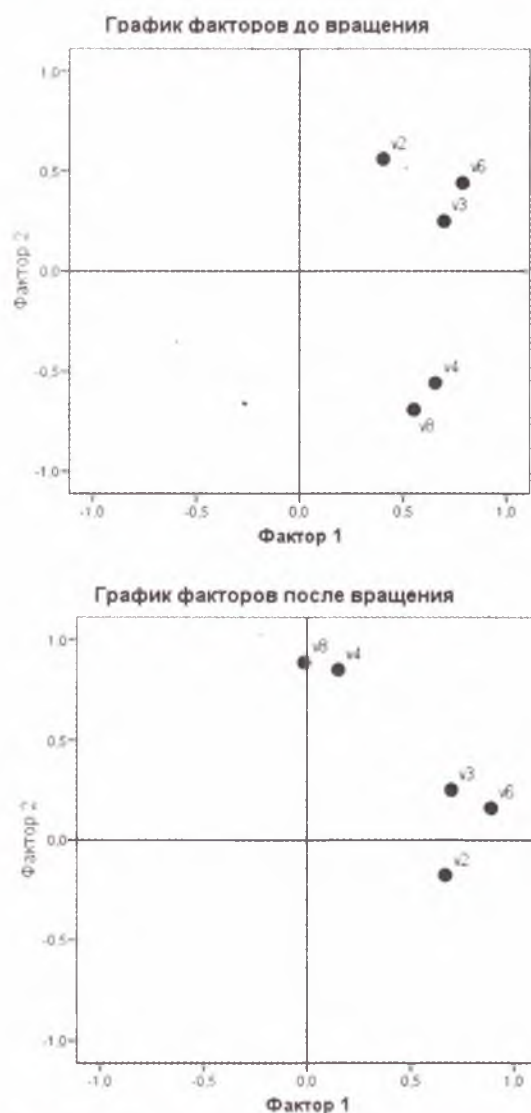


Рис. 20.1. Графическая интерпретация процедуры вращения факторов

Вращение факторов несколько не влияет на математическую строгость анализа: взаимное положение точек-переменных не меняется при повороте осей. Изначально процедура вращения выполнялась вручную, и исследователь самостоятельно задавал такое положение осей, при котором точки переменных максимально к ним приближались бы. Теперь SPSS позволяет выполнить несколько вариантов вращения, при которых оси поворачиваются так, чтобы получилась простая структура, удовлетворяющая тому или иному критерию. Наиболее популярным вариантом вращения является метод Варимакс. Вариант вращения Варимакс является ортогональным, поскольку при таком вращении сохраняется взаимное расположение осей под прямым углом. Иногда можно получить более предпочтительную простую структуру, если изменить угол между осями. Для это-

го предназначены варианты вращения Прямой облимин и Промакс. Наличие непрямого угла между осями факторов означает, что они не являются полностью независимыми друг от друга. Поскольку в реальных исследованиях факторы действительно не являются абсолютно независимыми, отклонения угла между осями от прямого при вращении вполне допустимо. Неортогональное вращение — довольно сложная процедура, и для ее применения необходимо четко представлять себе суть происходящего. В целом, аналогичную рекомендацию можно дать относительно факторного анализа вообще: не следует применять его без предварительного изучения соответствующих разделов статистики.

Интерпретация факторов

Итак, пусть в некоторой ситуации (близкой к идеальной) путем вращения мы добились того, что значение нагрузки для рассматриваемого фактора является большим (свыше 0,5), а для остальных факторов — малым (менее 0,2). Кроме того, мы четко представляем смысл нашего фактора, то есть то, что он измеряет. Разумеется, в большинстве исследований переменные могут взаимодействовать с «ненужным» фактором, а нередко таких факторов может быть несколько. Как правило, исследователь не ограничивается только числовыми результатами факторного анализа — необходимым условием успеха факторного анализа является понимание содержательной специфики конкретных данных и взаимосвязей между ними. Вы сможете убедиться в этом при чтении раздела «Представление результатов» этой главы.

Для факторного анализа мы будем использовать данные тестирования интеллекта 184 школьников. Тест включал в себя 11 субтестов (переменные и1, и2, ..., и11), наименования которых вы найдете в разделе «Представление результатов». Предполагалось, что эти 11 субтестов позволят измерить 3 и более обобщенные интеллектуальные характеристики: математические, вербальные и невербальные (образные). Факторный анализ должен был установить соотношение субтестов и факторов.

Файл данных, который мы будем использовать в примере, называется IQ.sav, число объектов (N) равно 184.

В разделе пошаговых процедур приведены два варианта шага 5. Шаг 5 иллюстрирует простейший вариант факторного анализа, в котором используются значения по умолчанию для всех параметров. Шаг 5а, напротив, включает в себя многие из действий, упомянутых ранее в этой главе.

Пошаговые алгоритмы вычислений

При проведении факторного анализа сначала выполняются три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае IQ.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом IQ.sav.

ШАГ 4

В меню Анализ выберите команду Снижение размерности ► Факторный анализ. На экране появится диалоговое окно Факторный анализ (рис. 20.2).

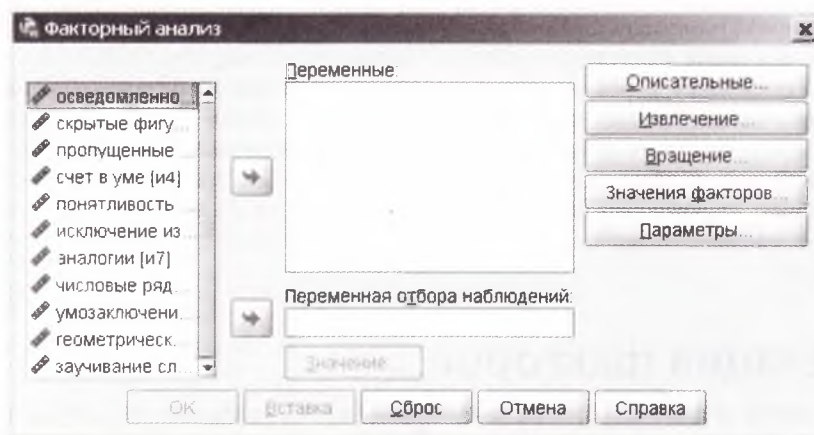


Рис. 20.2. Диалоговое окно Факторный анализ

Диалоговое окно Факторный анализ почти ничем не отличается от большинства диалоговых окон команд статистических операций SPSS: список доступных переменных находится в левой части, поле Переменные, предназначенное для указания имен переменных, участвующих в анализе, — в центре, а 5 кнопок, с помощью которых задаются параметры анализа, — в правой. Несмотря на то что существует множество математических нюансов, касающихся факторного анализа, для его проведения в простейшем случае достаточно всего лишь поместить в список Переменные несколько переменных, выбрать вариант вращения и щелкнуть на кнопке ОК. Далее будет приведен пример, в котором мы выберем 11 переменных и1, ..., и11 и проведем факторный анализ с параметрами по умолчанию.

Назначение поля Переменная отбора наблюдений примерно такое же, как у кнопки Если, упоминавшейся в главе 4. Оно позволяет указать имя переменной и ее уровень. Если, к примеру, поместить в поле Переменная отбора наблюдений имя пол, щелкнуть на кнопке Значение и ввести значение 1, то в факторном анализе будут участвовать только те объекты, для которых значения переменной пол равны 1.

Кнопки, расположенные в нижней части окна, позволяют управлять множеством разнообразных параметров анализа и могут использоваться независимо друг от друга в любом порядке.

Кнопка Описательные предназначена для открытия диалогового окна Факторный анализ: Описательные (рис. 20.3).

В этом окне находится две области: Статистики и Корреляционная матрица. Флажок Одномерные описательные позволяет получить в таблице имена переменных, их средние значения, стандартные отклонения и метки. Флажок Начальное решение по умолчанию установлен и отвечает за включение в выводимые данные имен переменных, начальных общностей (по умолчанию равных 1,0), факторов, собственных значений, а также общего и кумулятивного процента общей дисперсии для каждого фактора. В область Корреляционная матрица входят 7 флажков, управляющих выводом корреляционной матрицы; далее приведены описания четырех из них, которые используются чаще других:

- ▶ Коэффициенты — включает в матрицу коэффициенты корреляции, ради которых она обычно создается.
- ▶ Уровни значимости — в отдельной таблице приводятся значения p -уровней, соответствующих вычисленным коэффициентам корреляции.

- Детерминант — задает детерминант корреляционной матрицы, использующийся для критериев многомерной нормальности.
- КМО и критерий сферичности Барлетта — включает в матрицу два критерия: Бартлетта (проверяет коррелированность переменных) и адекватности выборки (КМО определяет применимость факторного анализа к выбранным переменным). По умолчанию эти тесты не проводятся, но они обеспечивают процедуру важной начальной статистической информацией. Более подробно данный параметр рассматривается в разделе «Представление результатов».

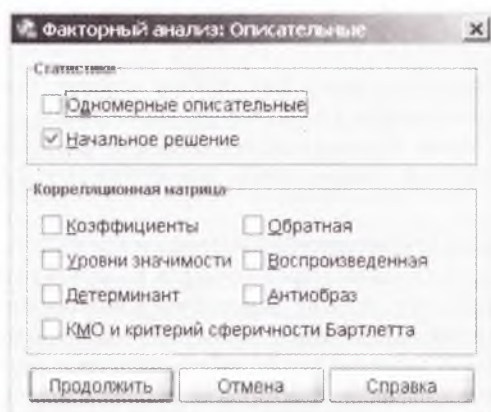


Рис. 20.3. Диалоговое окно Факторный анализ: Описательные

Щелчок на кнопке Извлечение в окне Факторный анализ приводит к открытию диалогового окна Факторный анализ: Выделение факторов (рис. 20.4), позволяющего выбрать метод извлечения, задать критерий числа факторов, сконфигурировать выводимые данные, связанные с процедурой извлечения, а также определить максимальное количество итераций, приводящих к получению результата.

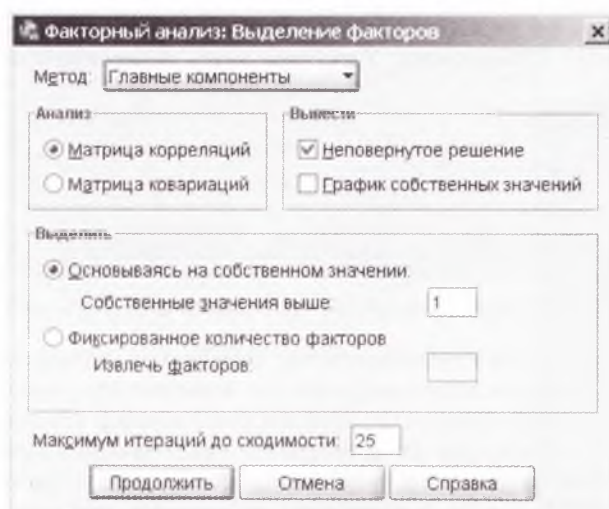


Рис. 20.4. Диалоговое окно Факторный анализ: Выделение факторов

Раскрывающийся список Метод содержит 7 пунктов. По умолчанию выбран пункт Главные компоненты; остальные пункты, соответствующие методам извлечения, перечислены далее:

- ▶ Невзвешенный МНК;
- ▶ Обобщенный МНК;
- ▶ Максимум правдоподобия;
- ▶ Факторизация главной оси;
- ▶ Альфа-факторизация;
- ▶ Анализ образов.

Формально, именно эти 6 перечисленных методов относятся к факторному анализу. Тем не менее мы ограничимся анализом главных компонент (АГК) как установленным по умолчанию и к тому же наиболее простым методом. Однако следует помнить, что АГК дает более грубое приближение к реальной структуре взаимосвязей, чем другие методы факторного анализа.

Устанавливая переключатель Анализ в положения Корреляционная матрица и Ковариационная матрица, можно выбрать начальные условия анализа. Переключатель Выделить позволяет задать критерий числа извлекаемых факторов: если он установлен в положение Основываясь на собственном значении, то можно воспользоваться единичным значением, установленным по умолчанию в поле Собственные значения выше, либо задать свое значение. Установка его в положение Фиксированное количество факторов позволяет извлечь количество факторов, указанное в поле Извлечь факторов. Область Вывести содержит два флажка: Неповоротное решение, установленный по умолчанию, и График собственных значений, часто используемый исследователями для определения числа факторов по критерию Р. Кеттелла. Если вы достаточно опытни в факторном анализе, то, вероятно, предпочтете установить оба указанных флажка. Наконец, при желании можно изменить значение в поле Максимум итераций до сходимости (по умолчанию 25).

Щелчок на кнопке Вращение в окне Факторный анализ открывает диалоговое окно Факторный анализ: Вращение (рис. 20.5). При выполнении факторного анализа установкой переключателя Метод в нужное положение можно выбрать один из трех ортогональных методов вращения: Варимакс, Эквимакс и Квартимакс. Установка этого переключателя

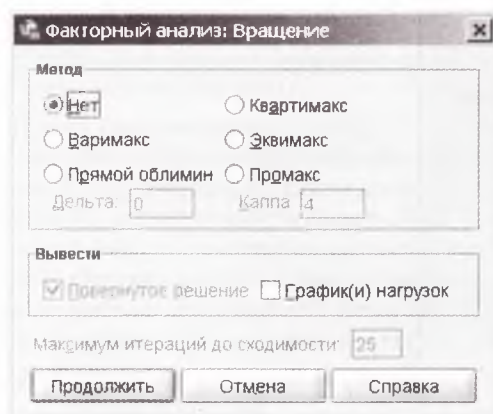


Рис. 20.5. Диалоговое окно Факторный анализ: Вращение

в положения Прямой облимин и Промакс позволяет выполнять неортогональное вращение факторов. Значения в полях Дельта и Каппа практически всегда рекомендуется оставлять установленными по умолчанию. Как говорилось ранее, настоятельно не рекомендуется использовать неортогональное вращение, если вы не обладаете исчерпывающими теоретическими знаниями в области факторного анализа.

Флажок Повернутое решение в области Вывести по умолчанию установлен, если выбран один из методов вращения. Он позволяет включить в выводимые данные наиболее важные результаты факторного анализа. Более подробно об этом рассказано в разделе «Представление результатов». Если вам необходим визуальный анализ структуры факторов после вращения, установите флажок График(и) нагрузок. Тогда по умолчанию в выводимые данные будет включено трехмерное изображение первых трех факторов, если их больше двух. Вы можете перемещать оси и менять точку наблюдения изображения, но, пожалуй, неизменным в большинстве случаев будет одно: вы не сможете уловить смысла того, что видите.

Вернемся к основному диалоговому окну команды Фактор и обратимся к кнопке Значения факторов. При щелчке на ней открывается диалоговое окно, позволяющее сохранить вычисленные оценки факторов как переменные. Мы не приводим иллюстрацию данного диалогового окна — вам достаточно знать, что для включения в выводимые данные матрицы факторных коэффициентов необходимо установить флажок Вывести матрицу коэффициентов значений факторов. Установка флажка Сохранить как переменные позволяет сохранить в файле данных вычисленные значения факторных оценок для наблюдений в качестве новых переменных.

Кнопка Параметры в окне Факторный анализ предназначена для вызова диалогового окна Факторный анализ: Параметры (рис. 20.6).

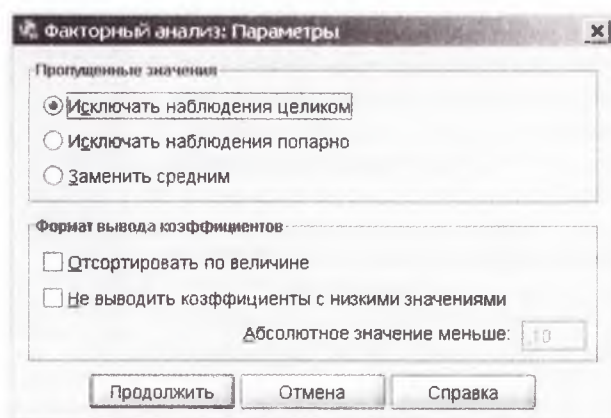


Рис. 20.6. Диалоговое окно Факторный анализ: Параметры

Элементы интерфейса, расположенные в области Формат вывода коэффициентов, управляют отображением матрицы факторных нагрузок после вращения. Часто полезно установить флажок Отсортировать по величине, который позволяет отсортировать переменные в соответствии с величиной их нагрузок по соответствующим факторам. Так, если 6 переменных максимально нагружают фактор 1, то эти переменные будут перечислены в порядке убывания их нагрузок в столбце с названием Фактор 1. Аналогичным образом будут выведены нагрузки для всех остальных факторов. Флажок Не выводить коэффици-

циенты с низкими значениями позволяет исключить из таблицы все значения нагрузок, меньшие заданного. Это делается для исключения из результатов тех нагрузок, которые не имеют существенного для интерпретации значения.

Переключатель Пропущенные значения позволяет выбрать режим обработки отсутствующих значений при вычислении корреляций (см. главу 4). Установка его по умолчанию в положение Исключать наблюдения целиком при вычислении корреляций исключает всю строку (наблюдение), если в ней пропущено хотя бы одно значение. Установка в положение Исключать наблюдения попарно обеспечивает исключение только пары значений (для двух переменных), если одно из них пропущено. Это приводит к меньшим потерям наблюдений, чем при установке по умолчанию, но может вызвать искажение результатов, если пропусков много. Установка переключателя в положение Заменить средним приводит к замене пропущенного значения переменной ее средним значением.

Далее приведены два варианта выполнения шага 5. В первом из них почти для всех параметров факторного анализа оставлены значения по умолчанию; второй вариант является более интересным и иллюстрирует применение некоторых из упоминавшихся параметров.

В следующем примере рассматривается применение факторного анализа к 11 переменным, и1, ..., и11, файла IQ.sav с параметрами по умолчанию и вращением по методу Варимакс.

ШАГ 5

После выполнения предыдущего шага у вас должно быть открыто диалоговое окно Факторный анализ (см. рис. 20.2). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной и1, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной и11. В результате окажутся выделенными все промежуточные переменные, начиная переменной и1 и заканчивая переменной и11.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
3. Щелкните на кнопке Вращение, чтобы открыть диалоговое окно Факторный анализ: Вращение (см. рис. 20.5).
4. Установите переключатель Метод в положение Варимакс и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Факторный анализ.
5. Щелкните на кнопке ОК, чтобы открыть окно вывода.

При выполнении этого шага проводится факторный анализ, включающий в себя следующие операции:

1. Вычисление корреляционной матрицы для 11 заданных переменных.
2. Извлечение 11 факторов методом главных компонентов.
3. Выбор для вращения всех факторов, чьи собственные значения не меньше 1.
4. Вращение факторов по методу Варимакс.
5. Вывод матрицы факторных нагрузок после вращения и других результатов.

В следующем примере проводится факторный анализ с участием тех же 11 переменных, что и в предыдущем случае, однако теперь задаются некоторые дополнительные параметры. Так, мы включим в вывод одномерные описательные статистики всех переменных, коэффициенты корреляции, а также применим критерии многомерной нормальности

и адекватности выборки. Для извлечения факторов будет использоваться метод главных компонент, а для отображения — график собственных значений. Вращение факторов будет производиться методом Варимакс. Наконец, мы отсортируем переменные по величине их нагрузок по факторам и отобразим те нагрузки, абсолютная величина которых не менее 0,3.

ШАГ 5а

После выполнения шага 4 у вас должно быть открыто диалоговое окно Факторный анализ (см. рис. 20.2). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной и1, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной и11. В результате окажутся выделенными все промежуточные переменные, начиная переменной и1 и заканчивая переменной и11.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
3. Щелкните на кнопке Описательные, чтобы открыть диалоговое окно Факторный анализ: Описательные (см. рис. 20.3).
4. В области Статистики установите флажок Одномерные описательные, в области Корреляционная матрица — флажки Коэффициенты и КМО и критерий сферичности Барлетта и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Факторный анализ.
5. Щелкните на кнопке Извлечение, чтобы открыть диалоговое окно Факторный анализ: Выделение факторов (см. рис. 20.4).
6. Установите флажок График собственных значений и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Факторный анализ.
7. Щелкните на кнопке Вращение, чтобы открыть диалоговое окно Факторный анализ: Вращение (см. рис. 20.5).
8. Установите переключатель Метод в положение Варимакс и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Факторный анализ.
9. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Факторный анализ: Параметры (см. рис. 20.6).
10. Установите флажки Отсортировать по величине и Не выводить коэффициенты с низкими значениями; в поле Абсолютное значение меньше вместо 0,10 задайте 0,4. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Факторный анализ.
11. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически активизирует окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Далее приведены данные, сгенерированные программой при выполнении шага 5а. Фрагмент, полученный благодаря установке флажка Критерии КМО и сферичности Барлетта в окне Факторный анализ: Описательные, показан на рис. 20.7.

Мера адекватности и критерий Бартлетта		
Мера выборочной адекватности Кайзера-Мейера-Олкина.		,736
Критерий сферичности Бартлетта	Прибл. хи-квадрат	689,447
	ст. св.	55
	Знач.	,000

Рис. 20.7. КМО и критерий сферичности Бартлетта

Величина КМО демонстрирует приемлемую адекватность выборки для факторного анализа. Критерий сферичности Бартлетта показывает статистически достоверный результат ($p < 0,05$): корреляции между переменными значимо отличаются от 0.

В первой из двух таблиц, показанных на рис. 20.8, перечислены имена переменных и общности. Столбцы второй таблицы содержат характеристики выделенных факторов: их порядковые номера (с 1-го по 3-й), суммы квадратов нагрузок, процент общей дисперсии, обусловленной фактором, и соответствующий кумулятивный (накопленный) процент (до и после вращения).

Общности		
	Начальные	Извлеченные
осведомленность	1,000	,578
скрытые фигуры	1,000	,644
пропущенные слова	1,000	,574
счет в уме	1,000	,718
понятливость	1,000	,615
исключение изображений	1,000	,493
анalogии	1,000	,695
числовые ряды	1,000	,629
умозаключения	1,000	,573
геометрическое сложение	1,000	,725
заучивание слов	1,000	,661

Метод выделения: Анализ главных компонент

Полная объясненная дисперсия						
Компонента	Суммы квадратов нагрузок извлечения			Суммы квадратов нагрузок вращения		
	Итого	% Дисперсии	Кумулятивный %	Итого	% Дисперсии	Кумулятивный %
1	3,916	35,598	35,598	2,666	24,236	24,236
2	1,788	16,258	51,856	2,143	19,480	43,716
3	1,199	10,904	62,760	2,095	19,044	62,760

Метод выделения: Анализ главных компонент

Рис. 20.8. Имена переменных, общности и характеристики факторов

Чем больше процент дисперсии, обусловленной фактором (см. рис. 20.8), тем больший вес имеет данный фактор. А чем больше кумулятивный процент, накопленный к последнему фактору, тем более состоятельным является факторное решение. Если этот накопленный процент менее 50 %, следует либо сократить количество переменных, либо уве-



личить количество факторов. В данном случае накопленный процент дисперсии вполне приемлем.

Как видите, в первом столбце первой таблицы перечислены метки, а не имена (и1, ..., и11) переменных. Если вы считаете, что метки занимают слишком много места в окне вывода, в меню командной строки Правка выберите команду Параметры, в открывшемся диалоговом окне перейдите на вкладку Метки в выводе. В области Метки в мобильных таблицах в раскрывающемся списке Переменные в метках вместо пункта Метки выберите пункт Имена и щелкните на кнопке ОК. После этого в таблицах окна вывода всегда будут представлены имена, а не метки переменных.

Диаграмма, приведенная на рис. 20.9, называется *графиком собственных значений*, или *диаграммой «каменистой осыпи»* (scree plot). Точками показаны собственные значения в пространстве двух координат. Этот тип диаграммы обычно используется при определении достаточного числа факторов перед вращением. При этом руководствуются следующим правилом: оставлять нужно лишь те факторы, которым соответствуют первые точки на графике до того, как кривая станет более пологой. Отметим, что по умолчанию SPSS вращает все факторы, чьи собственные значения превышают 1; в данном примере таких факторов три, а в соответствии с упомянутым правилом нужно было бы взять не три, а четыре фактора.

График собственных значений

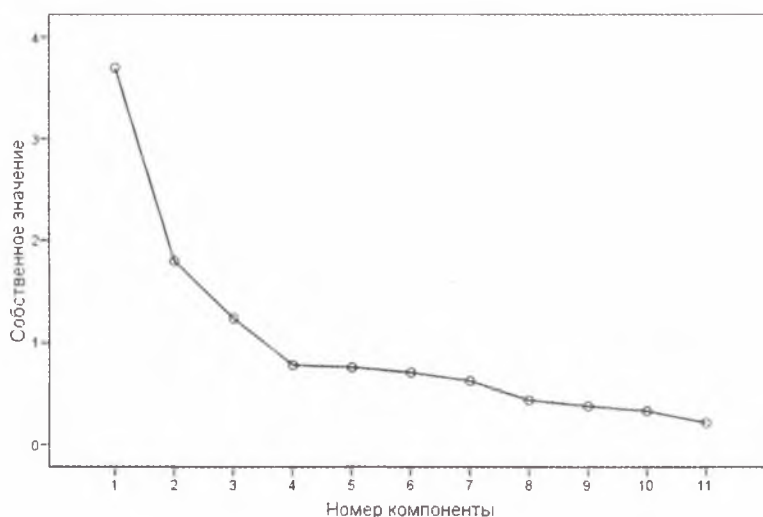


Рис. 20.9. График собственных значений

Далее SPSS включает в вывод исходную структуру факторных нагрузок (до вращения). Эти данные в большинстве случаев не представляют интереса, и мы не станем приводить их в этом разделе.

Преобразованная *матрица факторных нагрузок после вращения* изображена на рис. 20.10. Именно эта матрица является главным итогом факторного анализа и подлежит содержательной интерпретации.

Матрица повернутых компонент^а

	Компонента		
	1	2	3
анalogии	,816		
счет в уме	,778		
числовые ряды	,745		
умозаключения	,734		
заучивание слов		,810	
осведомленность		,755	
пропущенные слова		,660	
исключение изображений		,526	,464
геометрическое сложение			,759
понятливость			,757
скрытые фигуры			,737

Метод выделения: Анализ методом главных компонент

Метод вращения: Варимакс с нормализацией Кайзера

а. Вращение сошлось за 5 итераций.

Рис. 20.10. Матрица факторных нагрузок после вращения

Обратите внимание на то, что благодаря установке флажка Отсортировать по величине в окне Факторный анализ: Параметры факторные нагрузки отсортированы следующим образом:

- наибольшие значения нагрузок для каждого фактора сортируются в отдельных блоках;
- внутри каждого блока нагрузки факторов упорядочены по убыванию.

Кроме того, установка флажка Не выводить коэффициенты с низкими значениями и задание в поле Абсолютное значение меньше значения 0,4 позволили вывести только существенные для интерпретации величины факторных нагрузок.

Если бы описанный анализ провел реальный исследователь, он был бы отчасти удовлетворен полученными результатами. Очевидно, что первый из факторов соответствует предполагаемым математическим способностям, так как объединяет субтесты «счет в уме», «анalogии», «числовые ряды» и «умозаключения». Во второй фактор попали три субтеста, относящиеся к вербальным способностям: «заучивание слов», «осведомленность», «пропущенные слова», а в третий фактор — четыре субтеста, относящиеся к невербальным способностям: «скрытые фигуры», «геометрическое сложение», «понятливость» и «исключение изображений». К странностям результатов можно отнести распределение переменной «исключение изображений» между вторым и третьим фактором. Подобные отклонения обычно требуют отдельного изучения. В частности, можно увеличить число факторов или исключить неопределенные переменные и повторить анализ. Но лучшим способом проверки адекватности результатов факторного анализа является конфирматорный факторный анализ, описание которого приводится в главе 26 на этом же примере.

Целью приведенного примера было показать, каким образом факторный анализ группирует переменные, объединяя их по факторам. Каждый фактор интерпретируется как причина совместной изменчивости (корреляции) группы переменных. Отметим напоследок, что чаще всего исследователь не ограничивается однократной факторизацией

данных, а получает несколько вариантов решения с разными наборами переменных и разным числом факторов. Затем выбирается то решение, которое является наилучшим по признакам простоты структуры и концептуальной осмысленности. После получения приемлемого решения можно вычислить факторные оценки для объектов как новые переменные для дальнейшего анализа. Для этого в диалоговом окне Факторный анализ необходимо щелкнуть на кнопке Значения факторов и в открывшемся диалоговом окне Факторный анализ: Значения факторов установить флажок Сохранить как переменные. В итоге будут созданы новые переменные (по количеству факторов), которые можно использовать в дальнейшем анализе вместо исходных переменных.

Терминология, используемая при выводе

Далее дана трактовка терминов, используемых программой в окне вывода и относящихся к критериям КМО и Барлетта:

- ▶ КМО (мера выборочной адекватности Кайзера — Мейера — Олкина) — величина, характеризующая степень применимости факторного анализа к данной выборке:
 - ▶ более 0,9 — безусловная адекватность;
 - ▶ более 0,8 — высокая адекватность;
 - ▶ более 0,7 — приемлемая адекватность;
 - ▶ более 0,6 — удовлетворительная адекватность;
 - ▶ более 0,5 — низкая адекватность;
 - ▶ менее 0,5 — факторный анализ неприменим к выборке.
- ▶ Критерий сферичности Барлетта — критерий коррелированности переменных (нулевая гипотеза — о равенстве нулю корреляций между переменными). Значение *p*-уровня, меньшее 0,05, указывает на то, что данные вполне приемлемы для проведения факторного анализа, так как корреляции между переменными существенно отличаются от нуля.

Далее дана трактовка терминов, относящихся к общности и характеристикам факторов:

- ▶ Анализ главных компонент — метод извлечения факторов, используемый программой SPSS по умолчанию.
- ▶ Начальные — по умолчанию все переменные имеют исходное значение общности, равное 1.
- ▶ Извлеченные — значения общностей после извлечения факторов.
- ▶ Компонента — номер извлеченного фактора (компоненты).
- ▶ Суммы квадратов нагрузок извлечения — величины, характеризующие информативность каждого из факторов до вращения.
- ▶ Суммы квадратов нагрузок вращения — величины, характеризующие информативность каждого из факторов после вращения.
- ▶ Всего — суммы квадратов нагрузок для каждого из факторов.
- ▶ Процент дисперсии — процент дисперсии, обусловленный фактором, равный отношению суммы квадратов нагрузок данного фактора к количеству переменных (в данном случае 11).

- ▶ Кумулятивный процент — кумулятивный (накопленный) процент дисперсии.
- ▶ Матрица повернутых компонент — матрица факторных нагрузок после вращения, основной результат факторного анализа для содержательной интерпретации.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

21 Кластерный анализ

273	Сравнение кластерного и факторного анализа
275	Этапы кластерного анализа
276	Кластерный анализ матрицы различий (сходства)
278	Пошаговые алгоритмы вычислений
285	Представление результатов
288	Завершение анализа и выход из программы

Программа SPSS реализует три метода кластерного анализа: Двухэтапный кластерный анализ (TwoStep), Кластеризация K-средними (K-means) и Иерархическая кластеризация (Hierarchical).

Двухэтапный кластерный анализ позволяет выявить группы (кластеры) объектов по заданным переменным, если эти группы действительно существуют. При этом программа автоматически определяет количество существующих кластеров. Если однозначно определить это количество невозможно, все объекты помещаются в один.

Кластеризация K-средними разбивает по заданным переменным все множество объектов на заданное пользователем число кластеров так, чтобы средние значения для кластеров по каждой из переменных максимально различались.

Иерархическая кластеризация как наиболее гибкий из рассматриваемых методов позволяет детально исследовать структуру различий между объектами и выбрать наиболее оптимальное число кластеров. В силу этого иерархический кластерный анализ применяется наиболее часто и будет рассмотрен подробнее остальных.

Зачастую описывать новый статистический метод удобно путем его сравнения с другим методом. Именно таким образом мы и начнем рассмотрение иерархического кластерного анализа — сравнивая его с факторным анализом. При многочисленных общих чертах между названными статистическими методами существует немало различий. Если вы еще не знакомы с факторным анализом, настоятельно рекомендуем обратиться к началу главы 20, поскольку далее мы займемся сравнительной характеристикой факторного и кластерного анализов. Как обычно, после теоретической части рассматриваются примеры практической реализации статистических методов средствами SPSS, оформленные в виде пошаговых процедур.

Сравнение кластерного и факторного анализа

Основное сходство между кластерным и факторным анализом заключается в том, что оба они предназначены для перехода от исходной совокупности множества переменных (или объектов) к существенно меньшему числу факторов (кластеров). Тем не менее реализация статистических процедур и интерпретация результатов двух типов анализа различаются; пять основных различий приведены далее:

- ▶ Целью факторного анализа является замена большого числа исходных переменных меньшим числом факторов. Кластерный анализ, как правило, применяется для того, чтобы уменьшить число объектов путем их группировки. Другими словами, в ходе кластерного анализа переменные обычно не группируются, а выступают в качестве критериев для группировки объектов. Так, в примере факторного анализа (см. главу 20) 11 субтестов интеллекта (переменных) были сведены к 3 факторам, каждый из которых объединил несколько родственных исходных переменных. Кластерный анализ обычно применяется для выделения групп объектов на основе их сходства по измеренным признакам. Применительно к примеру с 11 субтестами интеллекта типичной задачей кластерного анализа была бы классификация учащихся (объектов) таким образом, чтобы по измеренным 11 показателям внутри каждой группы объекты были бы более похожи друг на друга, чем на объекты из других групп. Группы объектов, выделенные в результате кластерного анализа на основе заданной меры сходства между объектами, называются кластерами.
- ▶ Заявленные в предыдущем пункте различия между кластерным и факторным вариантами анализа со всей полнотой категоричности могут быть отнесены лишь к ранним версиям SPSS. Начиная с версии SPSS 10.0, программа позволяет с равным успехом проводить кластерный анализ не только объектов, но и переменных. В последнем случае кластерный анализ может выступать как более простой аналог факторного анализа. В разделе пошаговых процедур мы продемонстрируем оба варианта кластерного анализа.
- ▶ Действия, выполняемые в ходе статистических операций в каждом из вариантов анализа, принципиально различаются. В факторном анализе на каждом этапе извлечения фактора для каждой переменной подсчитывается доля дисперсии, которая обусловлена влиянием данного фактора. При кластерном анализе вычисляется расстояние между текущим объектом и всеми остальными объектами и кластер образует та пара, для которой расстояние оказалось наименьшим. Подобным образом каждый объект либо группируется с другим объектом, либо включается в состав существующего кластера. Процесс кластеризации конечен и продолжается до тех пор, пока все объекты не будут объединены в один кластер. Разумеется, подобный результат в общем случае не имеет смысла, и исследователь должен самостоятельно определить, в какой момент кластеризация должна быть прекращена.
- ▶ В контексте кластерного анализа особое место занимает один из его видов, называемый *иерархическим кластерным анализом*. В SPSS он реализуется с помощью команды Иерархическая кластеризация. Этот вид кластерного анализа чаще используется в биологии, экономике, социологии, политологии, нежели в психологии. Психологи обычно анализируют переменные с целью найти статистические связи между ними. Эти связи, как правило, указывают на сходство между теми или иными исследуемыми факторами. Деление выборки на группы в психологических анализах редко представляет интерес — в случаях, когда это оказывается необходимым, психологи отдают предпочтение дискриминантному, а не кластерному анализу (см. главу 22).
- ▶ Поскольку кластеризация переменных оказывается весьма доступной операцией, было бы интересно сравнить ее результаты с результатами более сложного факторного анализа. Как и в случае факторного анализа, выполнение кластерного анализа и его результаты зависят от ряда параметров: способа вычисления расстояния между объектами, кластеризации индивидуальных объектов и т. д.

Этапы кластерного анализа

Кластерный анализ выполняется в несколько этапов, дающих конечный результат. Сначала мы рассмотрим пример, созданный специально, чтобы показать суть кластерного анализа. Отметим, что кластерный анализ неприменим к файлам данных, использовавшимся ранее, поскольку при их составлении основное внимание было уделено смыслу и связям между переменными, а содержимое объектов (то есть информация, касающаяся субъектов) практически не играло роли. Для демонстрации кластерного анализа нами подготовлен файл `cars.sav`, содержащий гипотетические данные о 15 подержанных автомобилях разных марок, выставленных на продажу. Файл имеет структуру, подходящую для наглядной иллюстрации кластерного анализа.

Итак, выделяют несколько этапов кластерного анализа:

1. **Выбор переменных-критериев для кластеризации.** В нашем примере кластеризация будет осуществляться по следующим переменным: цена (стоимость в долларах), `t_сост` (экспертная оценка технического состояния по 10-балльной шкале), возраст (количество лет эксплуатации), пробег (количество километров, пройденных с начала эксплуатации).
2. **Выбор способа измерения расстояния между объектами, или кластерами** (изначально считается, что каждый объект соответствует одному кластеру). По умолчанию используется квадрат евклидова расстояния, согласно которому расстояние между объектами равно сумме квадратов разностей между значениями одноименных переменных объектов. Предположим, что марка автомобиля А имеет показатели технического состояния и возраста 5 и 6, а марка В — 7 и 4 соответственно. Тогда по этим двум переменным (координатам) расстояние между марками А и В вычисляется следующим образом: $(5 - 7)^2 + (6 - 4)^2 = 8$. При выполнении анализа сумма квадратов разностей вычисляется для всех переменных. Получаемые расстояния используются программой при формировании кластеров. Помимо евклидова существуют и другие виды расстояний, вычисляемые по другим формулам, однако мы не будем на них останавливаться. При необходимости обратитесь к руководству пользователя SPSS.

Относительно вычисления расстояния может возникнуть следующий вопрос: будет ли адекватным результат кластерного анализа в том случае, если переменные имеют различные шкалы измерения? Так, все переменные файла `cars.sav` имеют самые разные шкалы. Для решения проблемы шкалирования в SPSS используется стандартизация, в частности, ее простой метод — нормализация переменных, приводящая все переменные к стандартной z-шкале (среднее равно 0, стандартное отклонение — 1). При нормализации всех переменных при проведении кластерного анализа их веса становятся одинаковыми. В случае если все исходные данные имеют одну и ту же шкалу измерения либо веса переменных по смыслу должны быть разными, стандартизацию переменных проводить не нужно.

3. **Формирование кластеров.** Существует два основных метода формирования кластеров: *метод слияния* и *метод дробления*. В первом случае исходные кластеры увеличиваются путем объединения до тех пор, пока не будет сформирован единственный кластер, содержащий все данные. Метод дробления основан на обратной операции: сначала все данные объединяются в один кластер, который затем делится на части

до тех пор, пока не будет достигнут желаемый результат. По умолчанию программой SPSS используется метод слияния, и мы рассмотрим его в этой главе.

В методе слияния предусмотрены несколько способов объединения объектов. Способ, применяемый по умолчанию, называется *межгрупповым связыванием*, или *связыванием средних внутри групп*. SPSS вычисляет наименьшее среднее значение расстояния между всеми парами групп и объединяет две группы, оказавшиеся наиболее близкими. На первом шаге, когда все кластеры представляют собой одиночные объекты, данная операция сводится к обычному попарному сравнению расстояний между объектами. Термин «среднее значение» приобретает смысл лишь на втором этапе, когда сформированы кластеры, содержащие более одного объекта. Так, в нашем примере на начальном этапе имеется 15 кластеров (объектов). Сначала в кластер объединяются два объекта с наименьшим расстоянием друг от друга, затем подсчет расстояний повторяется и в кластер объединяется еще одна пара переменных. На втором этапе вы получите либо 13 свободных объектов и 1 кластер, объединяющий 2 объекта, либо 11 свободных объектов и 2 кластера по 2 объекта в каждом. В конечном счете все объекты окажутся в одном большом кластере. Существуют и другие методы объединения объектов, однако мы не станем рассматривать их здесь. При необходимости обратитесь к руководству пользователя SPSS.

4. **Интерпретация результатов.** Как и при проведении факторного анализа, желаемое число кластеров и оценка результатов анализа зависят от целей исследователя. Для рассматриваемого примера наиболее предпочтительным представляется количество кластеров, равное 3. Как показывает анализ, все марки можно разделить на 3 группы: первая группа имеет высокую стоимость (среднее значение — 15 230 долларов), небольшой срок эксплуатации (4 года) и средний пробег (85 400 км). Вторая группа имеет среднюю стоимость, небольшой пробег, наибольший возраст, но хорошее техническое состояние. Третья группа содержит недорогие модели с большим пробегом и невысоким рейтингом технического состояния.

Кластерный анализ матрицы различий (сходства)

Довольно часто исходной информацией для кластерного анализа являются не данные типа «объекты-переменные», а непосредственно данные о различии (сходстве) объектов. Например, респондент может оценивать различие (сходство) объектов, предъявляемых попарно, с перебором всех пар объектов. Или данными могут быть частоты совместной встречаемости каждой пары объектов. В этих случаях исходные данные представляют собой квадратную матрицу, как правило, симметричную относительно главной диагонали, каждый элемент которой — мера различия (или сходства) пары объектов, которым соответствуют строка и столбец матрицы.

К сожалению, кнопочный интерфейс программы SPSS не позволяет обрабатывать матрицы различий методами кластерного анализа (в многомерном шкалировании это возможно). Поэтому для обработки таких данных следует обратиться к командному языку Синтаксис (Syntax).

В SPSS предусмотрен формат данных для матрицы различия (сходства). При помощи синтаксиса такая матрица создается следующим образом¹. Для создания командного файла Синтаксис необходимо открыть окно *редактора синтаксиса*, выбрав в меню Файл команду Открыть ► Синтаксис. Предположим, необходимо создать матрицу различий 8×8 с именами объектов a b c d e f g h. Для этого в открытом окне Редактор синтаксиса следует ввести следующий текст (не важно, прописными или строчными буквами):

```
DATA LIST FREE /ROWTYPE_(a8) VARNAME_(a8) a b c d e f g h.
```

```
BEGIN DATA
```

```
PROX   A   0 6 2 7 18 0 1 5
PROX   B   6 0 14 3 8 4 19 14
PROX   C   2 14 0 30 12 4 6 8
PROX   D   7 3 30 0 1 16 3 4
PROX   E  18 8 12 1 0 19 13 14
PROX   F   0 4 4 16 19 0 5 1
PROX   G   1 19 6 3 13 5 0 16
PROX   H   5 14 8 4 14 1 16 0
```

```
END DATA.
```

```
EXECUTE.
```

```
VALUE LABELS ROWTYPE_ 'PROX' 'DISIMILARITY'.
```

Результатом выполнения этой команды будет создание матрицы различий 8×8 в окне редактора данных.

Матрица может содержать меры не различия, а сходства, например, если каждое значение матрицы — частота совместной встречаемости. Тогда последнюю строку следует заменить следующей:

```
VALUE LABELS ROWTYPE_ 'PROX' 'SIMILARITY'.
```

Чтобы выполнить введенную команду, ее нужно выделить и щелкнуть на кнопке Запустить выделенный фрагмент панели инструментов. Второй вариант запуска — выбрать в меню команду Запуск ► Всё. Имейте в виду, что любой пропущенный знак, включая завершающую точку, или неверно написанное слово приведут к выдаче программой сообщения об ошибке.

После выполнения команды в редакторе данных можно заменить полученную матрицу своими данными, следя за тем, чтобы количество строк равнялось количеству столбцов, в каждой строке переменной ROWTYPE_ стояло значение DISIMILARITY (или SIMILARITY, если данные — меры сходства), а в строках переменной VARNAME_ присутствовало имя объекта.

Для выполнения кластерного анализа методом межгруппового связывания с выводом таблицы шагов агломерации и дендрограммы следует выполнить такой синтаксис:

```
CLUSTER
```

```
  /MATRIX IN  (*)
```

```
  /METHOD BAVES
```

```
  /PRINT SCHEDULE
```

```
/PLOT DENDROGRAM.
```

¹ Решение проблемы заимствовано с сайта <http://www.spsstools.ru> (А. Балабанов Raynald's SPSS Tools по-русски).

Вся синтаксическая конструкция, создающая матрицу *различий* 8×8 (попарных различий 8 объектов) и выполняющая кластерный анализ указанным методом, сохранена в папке примеров Examples в файле Synt_Clust.sps¹. При желании вы можете открыть этот файл и задать его выполнение указанным ранее способом без предварительной редакции или отредактировав. Например, можете заменить указание матрицы различий на задание матрицы сходства. Для этого достаточно заменить в строке VALUE LABELS ROWTYPE_ слово DISIMILARITY на слово SIMILARITY.

Пошаговые алгоритмы вычислений

Структура раздела пошаговых процедур принципиально ничем не отличается от аналогичных разделов в других главах книги: шаги 1–4 являются подготовительными для выполнения анализа, шаги 6–7 требуются для печати результатов и завершения работы программы, а несколько вариантов (в данном случае два) шага 5 соответствуют разным способам выполнения статистической операции. Сначала продемонстрируем вариант использования кластерного анализа объектов с включением в вывод нескольких интересующих нас величин. Второй вариант шага 5 проиллюстрирует кластерный анализ переменных. В нем мы вновь обратимся к файлу IQ.sav, для того чтобы вы имели возможность сравнить кластерный анализ переменных с факторным анализом, описанным в предыдущей главе.

При проведении кластерного анализа сначала выполняются три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае cars.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом cars.sav.

ШАГ 4

В меню Анализ выберите команду Классификация ► Иерархическая кластеризация. Откроется диалоговое окно Иерархический кластерный анализ (рис. 21.1).

Дальнейшие действия в значительной степени зависят от того, какой тип кластеризации вы выберете. Для этой цели предусмотрены два положения переключателя Кластеризовать: Наблюдения и Переменные. Вначале в поле Переменные указываются имена тех переменных, значения которых будут использоваться при кластеризации. В нашем примере сюда следует поместить все переменные, кроме переменной марка, поскольку последняя представляет собой марку автомобиля. Далее следует задать способ идентификации объектов. Как правило, в роли идентификатора выступает переменная, содержащая уникальный номер объекта или его имя в виде строки. В данном случае мы будем использовать вполне подходящую для этого переменную марка. Имя идентифицирующей переменной указывается в поле Метить значениями.

Если переключатель Кластеризовать установить не в положение Наблюдения, а в положение Переменные, в поле Переменные потребуется указать кластеризуемые переменные, а поле Метить значениями останется пустым. По умолчанию флажки Статистики и Графики в области Вывести установлены, и в большинстве случаев нет необходимости их сбрасывать. В правой части диалогового окна расположены 4 кнопки, предназначенные для задания дополнительных параметров команды.

¹ Все файлы данных, включая файл Синтаксис, вы можете найти по адресу <http://www.piter.com/books/download/978591180318>.



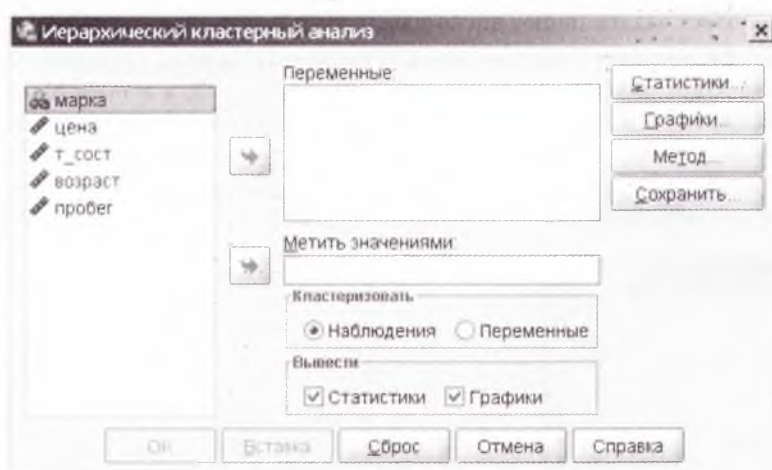


Рис. 21.1. Диалоговое окно Иерархический кластерный анализ

При щелчке на кнопке Статистики на экране появляется диалоговое окно Иерархический кластерный анализ: Статистики (рис. 21.2).

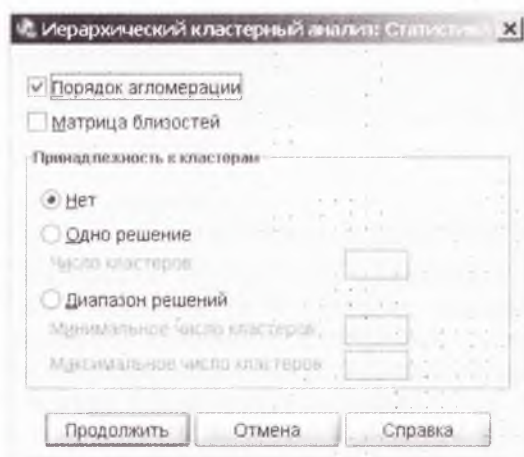


Рис. 21.2. Диалоговое окно Иерархический кластерный анализ: Статистики

Флажок Порядок агломерации по умолчанию установлен, обеспечивая включение в результаты кластерного анализа стандартного компонента окна вывода, который будет подробно рассмотрен в разделе «Представление результатов». Флажок Матрица близостей предназначен для отображения информации о расстояниях между объектами и кластерами. Использование матрицы удобно лишь для небольших файлов данных, поскольку с увеличением числа объектов размер матрицы резко возрастает, что делает ее громоздкой и неудобной для восприятия. Переключатель Принадлежность к кластерам можно установить в одно из трех положений:

- Нет — в выводимые результаты включаются все кластеры. Этот вариант установлен по умолчанию;

- Одно решение — позволяет задать точное число кластеров в решении;
- Диапазон решений — обеспечивает вывод нескольких решений с разным числом кластеров. Так, если указать в поле Минимальное число кластеров — 3, а в поле Максимальное число кластеров — 5, то в выводимые результаты будут включены все решения с 3, 4 и 5 кластерами.

Щелчок на кнопке Графики в окне Иерархический кластерный анализ открывает диалоговое окно Иерархический кластерный анализ: Графики (рис. 21.3).

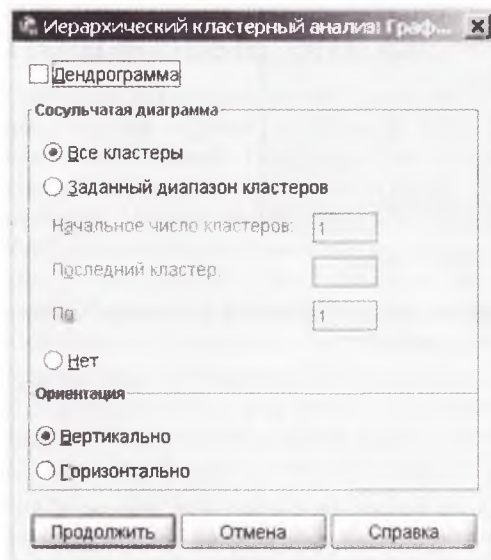


Рис. 21.3. Диалоговое окно Иерархический кластерный анализ: Графики

Флажок Дендрограмма позволяет включить в выводимые результаты ту же информацию, которая содержится на сосудчатой диаграмме, предлагаемой по умолчанию, а также относительную величину разности между переменными или кластерами на каждом шаге процесса. Поскольку сосудчатая диаграмма лишь дублирует информацию других разделов, а ее вид оставляет желать лучшего, мы откажемся от ее вывода. Для этого достаточно установить переключатель Сосудчатая диаграмма в положение Нет.

Щелчок на кнопке Метод в окне Иерархический кластерный анализ открывает диалоговое окно Иерархический кластерный анализ: Метод (рис. 21.4).

В этом окне для нас представляют интерес раскрывающиеся списки Метод, Интервальная и Стандартизация. Мы опишем лишь наиболее важные пункты этих списков и кратко упомянем остальные. Более детальные описания вы можете найти самостоятельно в руководстве пользователя SPSS.

В списке Метод наиболее часто используется пункт Межгрупповые связи. Суть метода заключается в том, что объединению на каждом шаге подвергаются кластеры или объекты, расстояние между которыми минимально. Более полное описание межгруппового связывания можно найти в разделе «Представление результатов». Помимо пункта Межгрупповые связи раскрывающийся список Метод содержит следующие пункты:

- Внутригрупповые связи;
- Ближайший сосед;

- Дальний сосед;
- Центроидная кластеризация;
- Медианная кластеризация;
- Метод Варда.

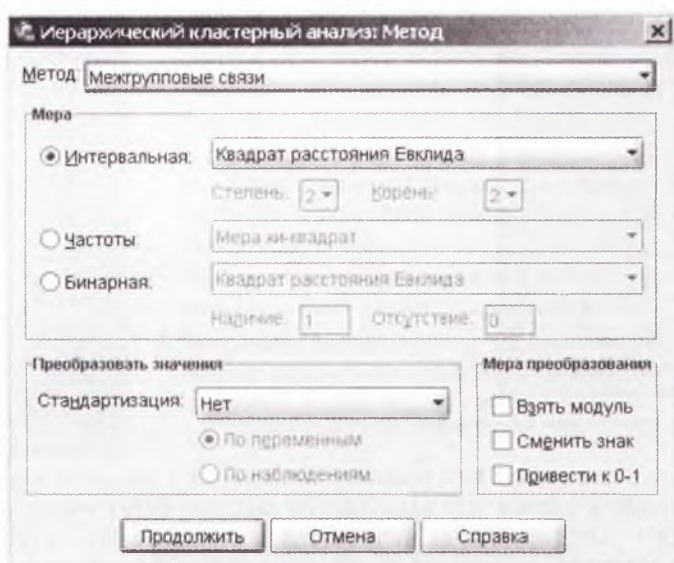


Рис. 21.4. Диалоговое окно Иерархический кластерный анализ: Метод

В раскрывающемся списке Интервальная по умолчанию выбран пункт Квадрат расстояния Евклида. Это означает, что расстояние между объектами вычисляется как разность квадратов соответствующих переменных этих объектов, участвующих в анализе. Остальные пункты списка Интервальная:

- Косинус — метод измерения близости, основанный на косинусах векторов значений;
- Корреляция Пирсона — метод измерения близости, основанный на корреляции векторов значений;
- Чебышев — вычисление расстояния как максимума абсолютного значения разности между элементами;
- Блок — определение меры расстояния по метрике города (см. главу 23);
- Минковского — определение меры расстояния Минковского (см. главу 23);
- Настроенная — позволяет задавать пользовательскую меру расстояния.

Процедура стандартизации выбирается в раскрывающемся списке Стандартизация. По умолчанию выбран пункт Нет, однако в случаях, когда переменные представлены в разных шкалах (единицах измерения), стандартизация необходима, и чаще всего выбирают пункт z-значения. Оставшиеся пункты, в которых допускается варьирование среднего значения или стандартного отклонения распределения, могут привести к другим результатам; выбор какого-либо из них определяется степенью его применимости к исследуемым данным и удобством для исследователя.

Переключатель Преобразовать значения по умолчанию установлен в положение По переменным, и в большинстве случаев менять его на альтернативное (По наблюдениям) нет необходимости.

В области Мера преобразования имеется три флажка, позволяющие изменить значения переменных: Взять модуль, Сменить знак и Привести к 0–1.

Последняя из четырех функциональных кнопок окна Иерархический кластерный анализ — кнопка Сохранить — открывает диалоговое окно Иерархический кластерный анализ: Сохранить (рис. 21.5). С помощью этого окна можно создавать новые переменные, значения которых будут указывать принадлежность наблюдений кластерам.

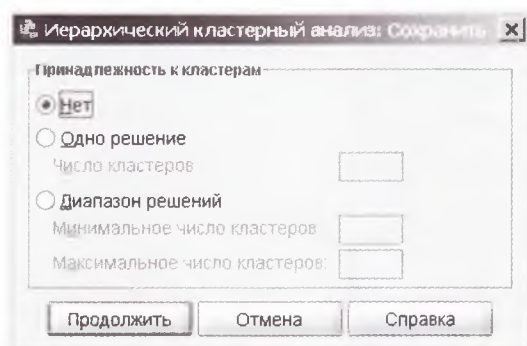
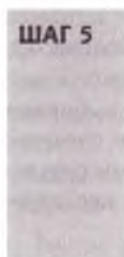


Рис. 21.5. Диалоговое окно Иерархический кластерный анализ: Сохранить

Если переключатель установлен в положение Нет, никакого сохранения в процессе анализа не производится. В противном случае при выполнении анализа будут созданы переменные, которые окажутся в конце файла. Если установить переключатель в положение Одно решение и указать в поле число кластеров 3, то получим новую переменную, значение которой равно 1, 2 или 3 в зависимости от того, какому кластеру будет принадлежать соответствующий объект в решении. Если же установить переключатель в положение Диапазон решений, в поле Минимальное число кластеров указать 3, а в поле Максимальное число кластеров — 5, это приведет к созданию трех новых переменных: первая будет принимать значения от 1 до 3, вторая — от 1 до 4, третья — от 1 до 5.

В следующем примере проводится кластерный анализ с несколькими дополнительными параметрами, обсуждавшимися ранее. В качестве идентификатора используется переменная марка. Остальные переменные файла задействуются для вычисления расстояния между объектами. Мы включим в выводимые результаты последовательность слияния и дендрограмму, но исключим диаграмму накопления. Значения всех переменных нормализуем для того, чтобы придать им равные веса и привести к одной шкале. В качестве расстояния между объектами зададим квадрат евклидова расстояния, а в качестве метода кластеризации — межгрупповое связывание. Кроме того, создадим новую переменную, в которой сохраним решение с тремя кластерами.



После выполнения шага 4 должно быть открыто диалоговое окно Иерархический кластерный анализ (см. рис. 21.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной марка, а затем — на нижней кнопке со стрелкой, чтобы переместить переменную в поле Метить значениями.
2. Щелкните сначала на переменной цена и, нажав на клавиатуре кнопку Shift, щелкните на переменной пробег. В результате окажутся выделенными все оставшиеся в списке переменные.

- 
3. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
 4. Щелкните на кнопке Графики, чтобы открыть диалоговое окно Иерархический кластерный анализ: Графики (см. рис. 21.3).
 5. Установите флажок Дендрограмма и переключатель Сосудчатая диаграмма в положение Нет. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Иерархический кластерный анализ.
 6. Щелкните на кнопке Метод, чтобы открыть диалоговое окно Иерархический кластерный анализ: Метод (см. рис. 21.4).
 7. В раскрывающемся списке Метод оставьте выбранным пункт Межгрупповые связи, в списке Стандартизация выберите пункт z-значения и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Иерархический кластерный анализ.
 8. Щелкните на кнопке Сохранить, чтобы открыть диалоговое окно Иерархический кластерный анализ: Сохранить (см. рис. 21.5).
 9. Установите переключатель в положение Одно решение, введите в расположенное рядом поле значение 3 и щелкните на кнопке Продолжить, чтобы вернуться в окно Иерархический кластерный анализ.
 10. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В следующем примере проводится кластерный анализ, в котором вместо объектов участвуют переменные. Мы используем данные файла IQ.sav, содержащего 11 переменных: и1, ..., и11. Обратите внимание на следующее обстоятельство. Обычно при группировании переменных исследователя интересует их взаимосвязь, а не их различие (сходство), как при группировании объектов. Исключением является случай, когда данные представляют собой оценки объектов экспертами, — тогда строки соответствуют экспертам, а столбцы — оцениваемым объектам. Поскольку в нашем примере интерес представляют именно взаимосвязи между переменными и мы хотим сравнить результаты с факторным анализом, то в качестве меры близости целесообразно выбрать корреляцию. При этом корреляции надо учитывать по абсолютной величине, так как большие (по модулю) отрицательные их величины так же свидетельствуют о связи, как и большие положительные. Все это необходимо иметь в виду, если речь идет о кластеризации переменных. Большинство остальных параметров команды оставим установленными по умолчанию; даже в стандартизации в данном случае нет необходимости, так как на величину корреляции не влияют единицы измерения переменных. Добавим лишь в выводимые результаты дендрограмму и исключим из них сосудчатую диаграмму. Обратите внимание: поскольку для выполнения анализа необходим файл IQ.sav, а не cars.sav, который использовался в предыдущем примере, мы начинаем с шага 3а.

Шаг 3а

Откройте файл данных, с которым вы намерены работать (в нашем случае IQ.sav). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Данные или щелкните на кнопке Открыть файл данных панели инструментов.
2. В открывшемся диалоговом окне дважды щелкните на имени IQ.sav или введите его с клавиатуры и щелкните на кнопке ОК.

ШАГ 4а

В меню Анализ выберите команду Классификация ► Иерархическая кластеризация. Откроется диалоговое окно Иерархический кластерный анализ (см. рис. 21.1).

ШАГ 5а

После выполнения предыдущего шага должно быть открыто диалоговое окно Иерархический кластерный анализ. Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Установите переключатель Кластеризовать в положение Переменные.
2. Нажмите кнопку мыши на переменной и1 и, не отпуская ее, перетащите указатель на переменную и11, затем отпустите кнопку мыши. В результате окажутся выделенными все 11 переменных: и1, ..., и11.
3. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
4. Щелкните на кнопке Графики, чтобы открыть диалоговое окно Иерархический кластерный анализ: Графики (см. рис. 21.3).
5. Установите флажок Дендрограмма и переключатель Сосудчатая диаграмма в положение Нет. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Иерархический кластерный анализ.
6. Щелкните на кнопке Метод, чтобы открыть диалоговое окно Иерархический кластерный анализ: Метод (см. рис. 21.4).
7. В раскрывающемся списке Интервальная выберите пункт Корреляция Пирсона, а в области Мера преобразований установите флажок Взять модуль. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Иерархический кластерный анализ.
8. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Эта процедура реализует кластерный анализ 11 показателей теста интеллекта. В выводимые данные включается такая информация о переменных, как число наблюдений, число пропущенных значений и т. п. Затем выводятся шаги агломерации (объединения в кластеры) и горизонтальная дендрограмма. Рекомендуем сравнить полученные результаты с результатами факторного анализа в главе 20.

После выполнения шага 5 (5а) программа автоматически откроет окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

В следующем примере кластерный анализ проводится с применением языка Синтаксис.

ШАГ 3б

Создайте файл Синтаксис, как это описано в разделе «Кластерный анализ матрицы различий (сходства)» в этой главе, или откройте файл примера Synt_Clust.sps. Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Синтаксис.
2. В открывшемся диалоговом окне дважды щелкните на имени Synt_Clust.sps. Откроется редактор синтаксиса с текстом команд.
3. Выберите в меню команду Запуск ► Все.

После выполнения шага 3б будет создан и открыт новый файл данных, содержащий матрицу 8×8 попарных различий восьми объектов. Кроме того, будет открыто окно вывода, содержащее результаты кластерного анализа этих восьми объектов методом средней связи (между группами). В окне редактора данных вы можете заменить представленную матрицу своими данными и выполнить кластерный анализ, отредактировав открытое окно редактора синтаксиса. Для этого надо просто удалить все строки с начала до строки CLUSTER и запустить команду Запуск ► Всё.

Представление результатов

В этом разделе представлены фрагменты выводимых данных, сгенерированные программой при выполнении шагов 5 и 5а.

После выполнения шага 5 в окне вывода появится показанная на рис. 21.6 таблица Шаги агломерации. В этой таблице вторая колонка Кластер объединен с содержит столбцы Кластер 1 и Кластер 2, которые соответствуют номерам кластеров, объединяемых на данном шаге. После объединения кластеру присваивается номер, соответствующий номеру в колонке Кластер 1. Так, на первом шаге объединяются объекты 5 и 14 и кластеру присваивается номер 5, далее на шаге 3 этот кластер объединяется с элементом 4 и новому кластеру присваивается номер 4 и т. д. Следующая колонка Коэффициенты содержит значение расстояния между кластерами, которые объединяются на данном шаге. Колонка Этап первого появления кластера показывает, на каком шаге до этого появлялись первый и второй объединяемые кластеры. Последняя колонка Следующий этап показывает, на каком шаге снова появится кластер, образованный на этом шаге.

Шаги агломерации						
Этап	Кластер объединен с		Коэффициенты	Этап первого появления кластера		Следующий этап
	Кластер 1	Кластер 2		Кластер 1	Кластер 2	
1	5	14	,439	0	0	3
2	2	12	,550	0	0	5
3	4	5	,671	0	1	6
4	7	13	,942	0	0	10
5	2	15	1,203	2	0	7
6	4	11	1,586	3	0	11
7	2	8	1,810	5	0	9
8	6	10	1,847	0	0	11
9	1	2	2,471	0	7	13
10	3	7	4,136	0	4	13
11	4	6	4,492	6	8	12
12	4	9	5,914	11	0	14
13	1	3	9,656	9	10	14
14	1	4	10,498	13	12	0

Рис. 21.6. Таблица шагов агломерации

Как видно из таблицы, на первом этапе в кластер объединяется пара объектов, расстояние между которыми является наименьшим. На втором этапе SPSS снова подсчитывает расстояния между объектами и объединяет в кластер пару наиболее близких объектов, в результате может получиться либо один кластер из трех объектов, либо

два кластера из двух объектов. Процесс слияния продолжается до тех пор, пока все объекты не попадут в один кластер. Попробуем объяснить результаты, полученные на этапах 1 и 13:

- На этапе 1 происходит объединение объектов 5 и 14. Расстояние между объектами равно 0,439. Ни один из двух объектов не принадлежит какому-либо кластеру, о чем свидетельствуют нули в столбцах Кластер 1 и Кластер 2 колонки Этап первого появления кластера. Следующим этапом для данного кластера, судя по столбцу Следующий этап, является этап 3, на котором к кластеру присоединяется объект 4.
- На этапе 13 происходит объединение кластеров, содержащих объекты 1 и 3. Объект 1 был объединен с кластером, содержащим объект 2, на этапе 9, а объект 3 — с объектами 7 и 13 на этапе 10. Расстояние между объединяемыми на этом этапе кластерами равно 9,656. Образованный на этом этапе кластер вновь появляется на следующем шаге.

По таблице шагов агломерации можно предварительно оценить число кластеров. Для этого необходимо проследить динамику увеличения расстояний по шагам кластеризации и определить шаг, на котором отмечается резкое возрастание расстояний. Оптимальному числу классов соответствует разность между числом объектов и порядковым номером шага, на котором было обнаружено резкое возрастание расстояний. В нашем примере это обнаруживается при переходе от шага 12 к шагу 13. Следовательно, оптимальное количество кластеров должно быть получено на шаге 12 или 13. Оно равно численности объектов минус номер шага, то есть $15 - 12 = 3$ кластера или $15 - 13 = 2$ кластера. Выбор того или иного решения зависит от содержательных соображений.

Дендрограмма представляет процесс кластеризации в форме древовидной структуры (рис. 21.7).

Дендрограмма не только позволяет перейти к любому объекту на любом уровне кластеризации, но и дает возможность судить о том, каково расстояние между кластерами или объектами на каждом из уровней. Числа от 0 до 25 являются условной шкалой этих расстояний; 0 соответствует наименьшему расстоянию на первом этапе, а 25 — наибольшему расстоянию на последнем этапе.

На дендрограмме любое решение характеризуется вертикальной линией, число точек пересечения которой с деревом соответствует количеству кластеров на текущем этапе. Так, для нашего примера эту линию следует расположить на уровне 15–20: между шагами 12 и 13 кластеризации. В этом случае получается 3 кластера. Для того чтобы установить состав каждого кластера, необходимо вернуться к корням дерева и выяснить соответствующие номера объектов. Напомним, что в результате выполнения шага 5 в файле `cars.sav` появилась новая переменная, определяющая принадлежность каждого объекта к одному из трех кластеров.

В заключение на рис. 21.8 приведем дендрограмму, полученную в результате выполнения шага 5а (кластеризация 11 переменных).

Дендрограмма показывает, что в результате кластеризации переменные группируются в три кластера, состав которых соответствует факторам, полученным в отношении тех же данных при факторном анализе (см. главу 20).

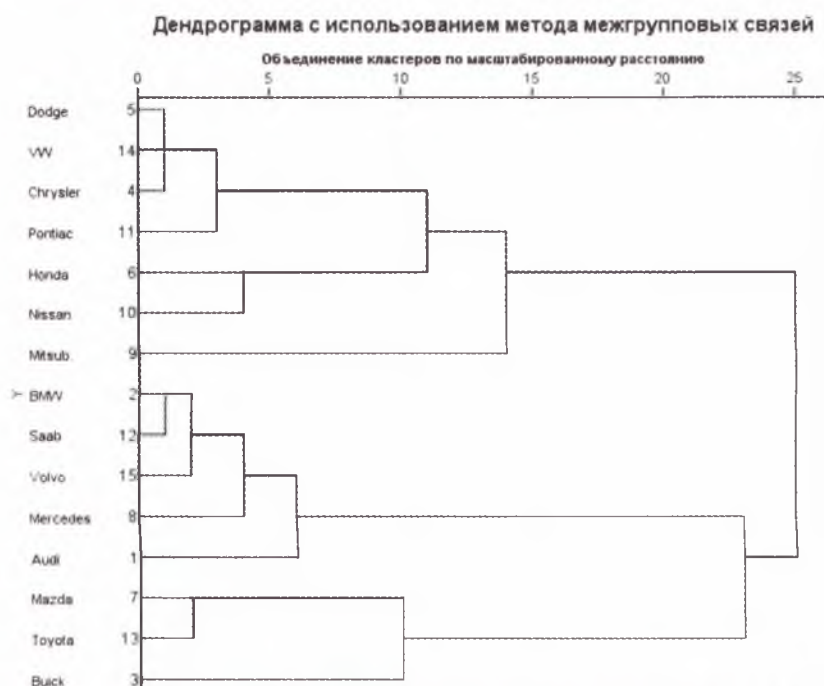


Рис. 21.7. Дендрограмма, полученная в результате выполнения шага 5



Рис. 21.8. Дендрограмма, полученная в результате выполнения шага 5a

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

22 Дискриминантный анализ

290	Этапы дискриминантного анализа
292	Пошаговые алгоритмы вычислений
297	Представление результатов
301	Терминология, используемая при выводе
303	Завершение анализа и выход из программы

Дискриминантный анализ позволяет предсказать принадлежность объектов к двум или более непересекающимся группам. Исходными данными для дискриминантного анализа является множество объектов, разделенных на группы так, что каждый объект может быть отнесен только к одной группе. Допускается при этом, что некоторые объекты не относятся ни к одной из групп (являются неизвестными). Для каждого из объектов имеются данные по ряду количественных переменных. Такие переменные называются дискриминантными переменными, или предикторами. Задачей дискриминантного анализа является определение:

- решающих правил, позволяющих по значениям дискриминантных переменных отнести каждый объект (в том числе неизвестный) к одной из известных групп;
- веса каждой дискриминантной переменной для разделения объектов на группы.

Существует множество ситуаций, в которых было бы весьма желательно вычислить вероятность того или иного исхода в зависимости от совокупности измеряемых переменных: например, подходит ли соискатель работы для той или иной должности, страдает психически больной человек шизофренией или психозом, вернется заключенный в тюрьму или к нормальной жизни после выхода на свободу, какие факторы влияют на увеличение риска пациента получить сердечный приступ и т. п. Все перечисленные ситуации обладают двумя общими чертами: во-первых, есть информация о принадлежности некоторых (не всех) субъектов к той или иной группе; во-вторых, о каждом субъекте имеется дополнительная информация для создания формулы, которая позволит спрогнозировать принадлежность субъекта к той или иной группе.

Дискриминантный анализ имеет определенное сходство с кластерным анализом, которое заключается в том, что в обоих случаях исследователь ставит перед собой цель разделить совокупность объектов (а не переменных) на несколько более мелких (значимых) групп. Тем не менее процесс классификации в двух видах анализа принципиально различен. В кластерном анализе объекты классифицируются на основе их различий без какой-либо предварительной информации о количестве и составе классов. В дискриминантном анализе изначально заданы количество и состав классов и основная задача заключается в определении того, насколько точно можно предсказать принадлежность объектов к классам при помощи данного набора предикторов.

Дискриминантный анализ представляет собой альтернативу множественного регрессионного анализа (см. главу 18) для случая, когда зависимая переменная является не количественной, а номинальной переменной. При этом дискриминантный анализ решает, по сути, те же задачи, что и множественный регрессионный анализ: предсказание зна-

чений зависимой переменной (в данном случае — категорий номинального признака) и определение того, какие независимые переменные лучше всего подходят для такого предсказания. Дискриминантный анализ основан на составлении уравнения регрессии (см. главы 17 и 18), использующего номинальную зависимую переменную (обратите внимание на то, что она не является количественной, как в случае регрессионного анализа). Уравнение регрессии составляется на основе тех объектов, групповая принадлежность которых известна, что позволяет максимально точно подобрать его коэффициенты. После того как уравнение регрессии получено, его можно использовать для группировки интересующих нас объектов в целях прогнозирования их принадлежности к какому-либо классу. Команда дискриминантного анализа весьма непростая и требует настройки множества параметров, описание большинства из которых лежит за рамками темы данной книги. При необходимости вы можете обратиться за дополнительной информацией к руководству пользователя SPSS.

Параметры дискриминантного анализа, как и параметры большинства сложных статистических операций, в основном определяются особенностями данных, а также задачами исследователя. Как всегда мы рассмотрим пример (на этот раз единственный) проведения дискриминантного анализа в разделе пошаговых процедур, а раздел «Представление результатов» посвятим интерпретации выводимых данных.

Для демонстрации дискриминантного анализа рассмотрим пример прогнозирования успешности обучения на основе предварительного тестирования. Файл `class.sav` содержит данные о 46 учащихся (объекты с 1-го по 46-й), юношах и девушках (переменная `пол`), закончивших курс обучения, оценки успешности обучения которых известны — для этого используется переменная `оценка` (1 — низкая, 2 — высокая). Кроме того, в файл включены данные предварительного тестирования этих учащихся до начала обучения (13 переменных):

- ▶ `и1`, ..., `и11` — 11 показателей теста интеллекта;
- ▶ `э_и` — показатель экстраверсии по тесту Г. Айзенка (H. Eysenck);
- ▶ `н` — показатель нейротизма по тесту Г. Айзенка.

Еще для 10 претендентов на курс обучения (объекты с 47-го по 56-й) известны лишь результаты их предварительного тестирования (13 перечисленных переменных). Значения переменной `оценка` для них, разумеется, неизвестны, и в файле данных им соответствуют пустые ячейки. В процессе дискриминантного анализа мы, в частности, попытаемся спрогнозировать успешность обучения этих 10 претендентов в предположении, что выборки закончивших обучение и претендентов идентичны.

Этапы дискриминантного анализа

Дискриминантный анализ состоит из трех основных этапов:

1. **Выбор переменных-предикторов.** Исследователь использует теоретические знания, практический опыт, догадки и т. п., для того чтобы составить список переменных, которые могут повлиять на результат группировки (переменную-критерий). В рассматриваемом файле помимо переменной-критерия (`оценка`) содержатся 13 переменных, характеризующих каждого учащегося. Это позволяет нам сделать все 13 переменных предикторами и включить их в уравнение регрессии. Если бы число переменных было велико (например, несколько сотен), было бы невозможно применить дискриминантный анализ ко всем им одновременно. Это обусловлено как

концептуальными причинами (возможная коллинеарность переменных, потеря степеней свободы и т. п.), так и практическими ограничениями (недостаточный объем оперативной памяти компьютера). Обычно на начальном этапе дискриминантного анализа для предикторов формируется корреляционная матрица. В данном контексте она имеет особый смысл, называется *общей внутригрупповой корреляционной матрицей* и содержит средние коэффициенты корреляции для двух или более корреляционных матриц (каждая для одной группы). Помимо общей внутригрупповой корреляционной матрицы можно также вычислить ковариационные матрицы для отдельных групп, для всей выборки либо общую внутригрупповую ковариационную матрицу. Нередко исследователи применяют серию *t*-критериев между двумя группами для каждой переменной либо однофакторный дисперсионный анализ, если количество групп оказывается больше двух. Поскольку целью дискриминантного анализа является составление наилучшего уравнения регрессии, дополнительный анализ исходных данных никогда не будет лишним. Так, в результате применения *t*-критериев для данных нашего примера были найдены значимые различия между двумя уровнями переменной оценка для 8 из 13 предикторов. Мы рассмотрим один из наиболее распространенных вариантов дискриминантного анализа, при проведении которого программа автоматически исключает несущественные для предсказания предикторы по критериям, которые устанавливает исследователь.

2. **Выбор параметров.** В этой главе будет продемонстрирован один из вариантов дискриминантного анализа. По умолчанию программа реализует метод, который основан на *принудительном включении* в регрессионное уравнение всех предикторов, указанных исследователем. В другом варианте используется метод Уилкса (Wilks), относящийся к категории *пошаговых методов* и основанный на минимизации коэффициента Уилкса (λ) после включения в уравнение регрессии каждого нового предиктора. Так же как и в случае множественного регрессионного анализа, существует критерий для включения предикторов в уравнение регрессии (по умолчанию таким критерием является $F > 3,84$) и критерий для исключения предикторов из уравнения регрессии (по умолчанию $F < 2,71$). Коэффициент λ представляет собой отношение внутригрупповой суммы квадратов к общей сумме квадратов и характеризует долю влияния предиктора на дисперсию критерия. Со значением λ связаны величины F и p , характеризующие его значимость. Более полное описание вы можете найти в разделе «Представление результатов».

Какой же из двух методов предпочтительнее? Как показывает практика, зачастую компьютер справляется с составлением уравнения регрессии лучше, чем исследователь, задающий список предикторов вручную. Однако встречаются ситуации, когда полезней ограничить самостоятельность компьютера. Например, если провести дискриминантный анализ для наших данных с включением всех переменных, то неверно классифицированы будут 5 объектов из 46. Той же точности прогноза можно достичь всего с 7 предикторами, если выбрать пошаговый метод с *отличающимися* от принятых по умолчанию установками (как указано далее в пошаговом алгоритме). В то же время, если использовать пошаговый метод с установками по умолчанию, оставляющий только 3 предиктора, количество неверно сгруппированных объектов увеличится до 9. Помимо рассмотренных программа SPSS располагает и другими методами выбора предикторов, однако их описание выходит за рамки темы данной книги. Рекомендуем при необходимости обращаться к руководству пользователя SPSS.

3. **Интерпретация результатов.** Целью дискриминантного анализа является составление уравнения регрессии с использованием выборки, для которой известны значения и предикторов, и критерия. Это уравнение позволяет по известным значениям предикторов определить неизвестные значения критерия для другой выборки. Разумеется, точность рассчитываемых значений критерия для второй выборки в общем случае не выше, чем для исходной. Так, в нашем примере регрессионное уравнение обеспечило около 90 % корректных результатов для той выборки, с помощью которой оно было создано. Соответственно, точность предсказания успешности обучения для 10 претендентов может достигать 90 % лишь в том случае, если выборка претендентов совершенно идентична тем 46 учащимся, данные для которых послужили основой прогноза.

Пошаговые алгоритмы вычислений

Перед началом дискриминантного анализа выполняются три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае class.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом class.sav.

ШАГ 4

В меню Анализ выберите команду Классификация ► Дискриминантный анализ. На экране появится диалоговое окно Дискриминантный анализ (рис. 22.1).

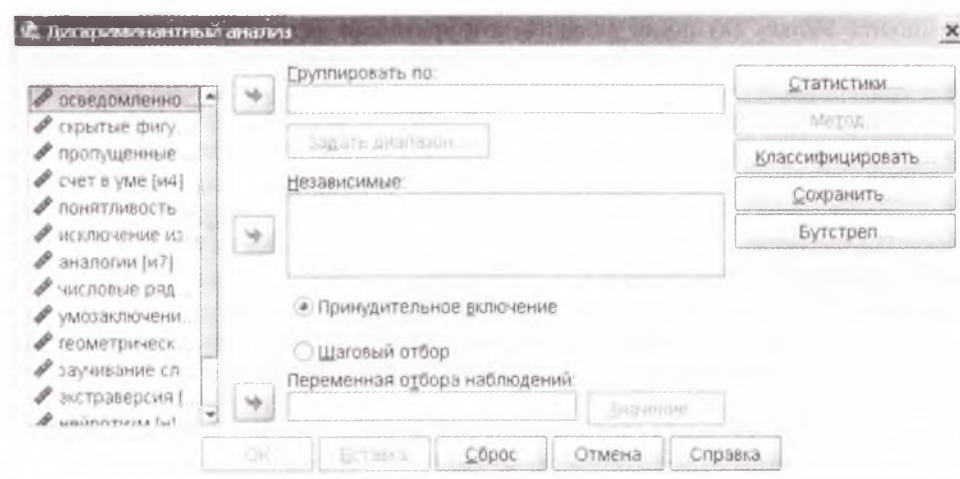


Рис. 22.1. Диалоговое окно Дискриминантный анализ

Помимо таких привычных элементов интерфейса, как список доступных переменных слева и 5 стандартных кнопок справа, диалоговое окно Дискриминантный анализ содержит 4 дополнительных элемента.

- Поле Группировать по предназначено для задания единственной зависимой переменной.
- В поле Независимые включается любое число переменных, участвующих в дискриминантном анализе в качестве предикторов.

- Положения переключателя Принудительное включение и Шаговый отбор позволяют выбрать один из двух вариантов дискриминантного анализа.
- Пять кнопок управления дополнительными параметрами в правой части диалогового окна предназначены для открытия дополнительных диалоговых окон.

Назначение поля Переменная отбора наблюдений примерно такое же, как у кнопки Если, упоминавшейся в главе 4. Оно позволяет указать имя переменной и ее уровень. Если, к примеру, поместить в поля Переменная отбора наблюдений имя пол, щелкнуть на кнопке Значение и ввести значение 1, в дискриминантном анализе будут участвовать только те объекты, для которых значения переменной пол равны 1.

Кнопка Сохранить позволяет сохранять в качестве новых переменных для каждого объекта (в том числе, неизвестного) следующие величины:

- прогнозируемый номер группы;
- оценки дискриминантных функций;
- вероятность принадлежности к каждой группе.

Важными элементами интерфейса диалогового окна Дискриминантный анализ являются элементы управления уровнями зависимой переменной. Хотя в нашем примере дискриминантный анализ используется для разделения объектов на две группы, количество групп может быть и больше. После ввода имени зависимой переменной в поле Группировать по щелчок на кнопке Задать диапазон позволяет открыть диалоговое окно Дискриминантный анализ: Задание диапазона (рис. 22.2). Поскольку переменная оценка, используемая в нашем примере как зависимая, имеет лишь два уровня (1 и 2), их следует указать в полях Минимум и Максимум, а затем щелкнуть на кнопке Продолжить. Если число уровней группирующей переменной больше двух, описанная операция позволит задать любой диапазон уровней.

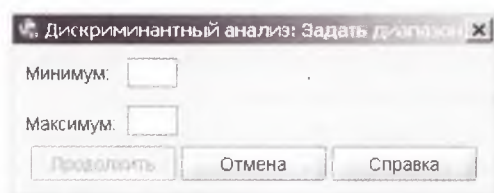


Рис. 22.2. Диалоговое окно Дискриминантный анализ: Задать диапазон

Щелчок на кнопке Статистики в окне Дискриминантный анализ открывает диалоговое окно Дискриминантный анализ: Статистики (рис. 22.3). Поскольку почти все элементы интерфейса этого окна активно используются исследователями, далее приведено описание каждого из них. Как правило, с помощью этих элементов исследователи выясняют наличие зависимостей между предикторами перед началом дискриминантного анализа.

Флажки в области Описательные:

- Средние — средние значения и стандартные отклонения для каждой переменной каждой группы (в данном случае двух групп, образованных уровнями переменной оценка) и выборки в целом;
- Однофакторный дисперсионный анализ — если число уровней зависимой переменной равно 2, то это t -критерии для сравнения между собой двух средних значений для групп по каждой переменной;
- М Бокса — критерий равенства дисперсионно-ковариационных матриц для уровней зависимой переменной.

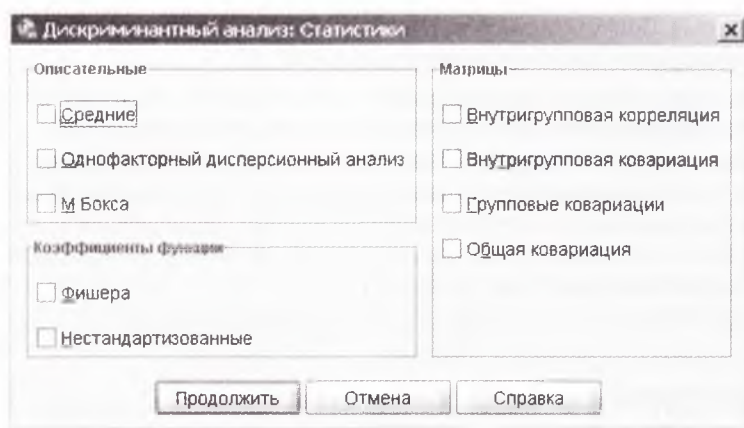


Рис. 22.3. Диалоговое окно Дискриминантный анализ: Статистики

Флажки в области Коэффициенты функции:

- ▶ Фишера — канонические коэффициенты дискриминантных функций;
- ▶ Нестандартизованные — нестандартизованные коэффициенты дискриминантных функций, вычисленные для исходных значений предикторов.

Флажки в области Матрицы:

- ▶ Внутригрупповая корреляция — матрица, состоящая из соответствующих средних значений корреляционных матриц для уровней зависимой переменной;
- ▶ Внутригрупповая ковариация — матрица, аналогичная предыдущей, с той разницей, что вместо корреляций используются ковариации;
- ▶ Групповые ковариации — отдельные ковариационные матрицы для каждого уровня зависимой переменной;
- ▶ Общая ковариация — ковариационная матрица для всей выборки.

Щелчок на кнопке Метод в окне Дискриминантный анализ приводит к появлению диалогового окна Дискриминантный анализ: Шаговый отбор (рис. 22.4). Обратите внимание на то, что кнопка Метод доступна лишь в том случае, если переключатель в окне Дискриминантный анализ установлен в положение Шаговый отбор.

В этом окне при помощи переключателя Метод можно выбрать один из пошаговых методов составления дискриминантного уравнения, в области Критерии задать критерии для включения в дискриминантное уравнение и исключения из него предикторов, а также включить в окно вывода желаемые величины при помощи флажков в области Вывести. Из положений переключателя выбора метода составления уравнения для нас представляет интерес лишь положение Лямбда Уилкса. Соответствующий метод основан на минимизации коэффициента Уилкса (λ) после включения в уравнение регрессии каждого нового предиктора. Более подробное описание приведено в разделе «Представление результатов».

По умолчанию критериями для включения и исключения предикторов являются пороговые значения F -критерия. Как правило, значению $F = 3,84$ соответствует уровень значимости $p \approx 0,05$, а значению $F = 2,71$ — $p \approx 0,1$. Исследователи нередко предпочитают использовать значения по умолчанию, нежели задавать собственные.

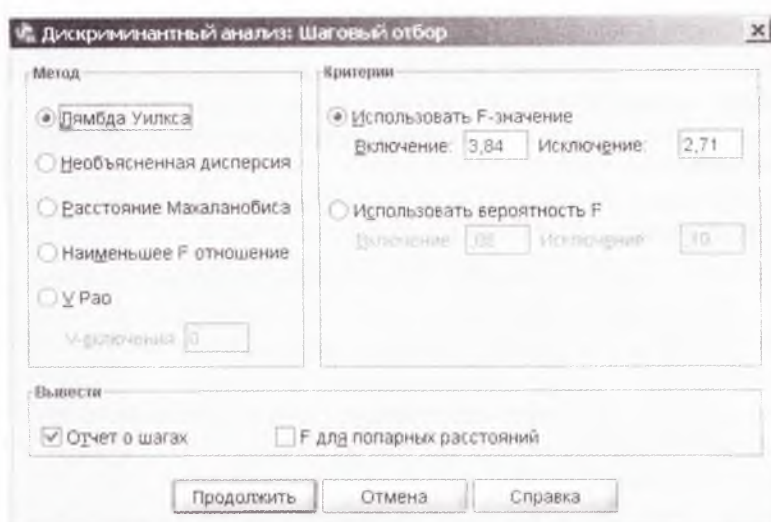


Рис. 22.4. Диалоговое окно Дискриминантный анализ: Шаговый отбор

Флажок Отчет о шагах в области Вывести управляет пошаговым выводом основных результатов дискриминантного анализа, поэтому по умолчанию он установлен. Флажок F для попарных расстояний относится только к методу Расстояние Махаланобиса.

Щелчок на кнопке Классифицировать в окне Дискриминантный анализ открывает диалоговое окно Дискриминантный анализ: Классификация (рис. 22.5). Большая часть элементов интерфейса этого окна активно используется исследователями и описана далее.

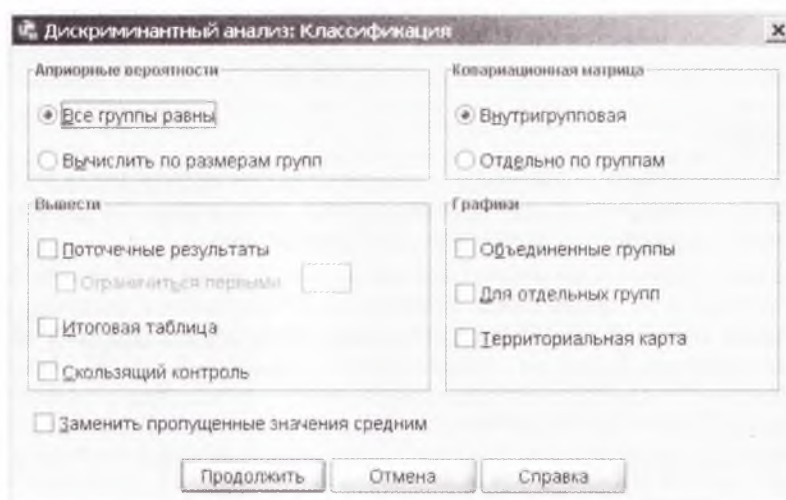


Рис. 22.5. Диалоговое окно Дискриминантный анализ: Классификация

Переключатель Априорные вероятности можно установить в одно из двух положений:

- ▶ Все группы равны — вероятности принадлежности объекта к каждой из групп полагаются равными;
- ▶ Вычислить по размерам групп — вероятности принадлежности объекта к каждой из групп пропорциональны размерам групп.

Назначение флажков в области Графики:

- ▶ Объединенные группы — в вывод включается гистограмма (если число групп равно 2) или диаграмма разброса (если число групп больше 2);
- ▶ Для отдельных групп — в вывод включаются несколько диаграмм, каждая из которых соответствует одной группе;
- ▶ Территориальная карта — используется только в случаях, когда зависимая переменная имеет три или более уровня.

Переключатель Ковариационная матрица позволяет выбрать ковариационную матрицу:

- ▶ Внутригрупповая — в это положение переключатель установлен по умолчанию, что позволяет классифицировать объекты при помощи общей внутригрупповой ковариационной матрицы;
- ▶ Отдельно по группам — объекты классифицируются при помощи отдельных ковариационных матриц для каждой группы.

Флажки в области Вывести относятся к отображению результатов анализа:

- ▶ Поточечные результаты — этот флажок полезно устанавливать, если файл данных не слишком велик: в результат включается список объектов и для каждого объекта указываются фактическая группа, прогнозируемая группа, вероятность попадания в прогнозируемую группу и значения дискриминантных функций (подобный список приведен в разделе «Представление результатов»);
- ▶ Итоговая таблица — этот флажок позволяет включить в результат число и процент корректных и некорректных классификаций для каждой группы (в разделе «Представление результатов» имеются соответствующие результаты);
- ▶ Скользящий контроль — при установке этого флажка каждый объект классифицируется с помощью функций, полученных по всем остальным наблюдениям, кроме данного. Этот метод также известен как U-метод.

Название флажка Заменить пропущенные значения средним вполне отражает его назначение: при его установке пропущенное значение переменной заменяется ее средним (только на этапе классификации).

В этой главе приведена единственная версия шага 5. Разумеется, можно выполнить дискриминантный анализ с параметрами по умолчанию, однако значительно интереснее сделать это с применением некоторых из описанных здесь параметров.

В следующем примере проводится дискриминантный анализ для зависимой переменной оценка, имеющей два уровня, и 13 предикторов. Предикторы добавляются в дискриминантное уравнение пошаговым методом (Уилкса) с установками, отличающимися от предлагаемых по умолчанию: для включения предикторов в уравнение $F = 1,125$, а для исключения — значение $F = 1$. Для анализа зависимости между предикторами вычисляются все описательные статистики. Кроме того, мы включаем в окно вывода нестандартные коэффициенты дискриминантного уравнения, результаты для каждого объекта и итоговую таблицу.

ШАГ 5

После выполнения предыдущего шага у вас должно быть открыто диалоговое окно Дискриминантный анализ (см. рис. 22.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной оценка, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Группировать по.



2. Щелкните на кнопке Задать диапазон, чтобы открыть диалоговое окно Дискриминантный анализ: Задать диапазон (см. рис. 22.2).
3. В поле Минимум введите значение 1, нажмите клавишу Tab, чтобы переместить фокус ввода в поле Максимум, введите значение 2 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Дискриминантный анализ.
4. Выделите переменную и1, нажмите на клавиатуре клавишу Shift и, не отпуская ее, щелкните левой кнопкой мыши на переменной n. В результате окажутся выделенными все переменные от и1 до n.
5. Щелкните на средней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Независимые, установите переключатель в положение Шаговый отбор и щелкните на кнопке Статистики, чтобы открыть диалоговое окно Дискриминантный анализ: Статистики (см. рис. 22.3).
6. Установите флажки Средние, Однофакторный дисперсионный анализ, М Бокса, Нестандартизованные и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Дискриминантный анализ.
7. Щелкните на кнопке Метод, чтобы открыть диалоговое окно Дискриминантный анализ: Шаговый отбор (см. рис. 22.4).
8. В поле Использовать F-значение: Включение введите значение 1,25, нажмите клавишу Tab, чтобы переместить фокус ввода в поле Исключение, введите значение 1 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Дискриминантный анализ.
9. Щелкните на кнопке Классифицировать, чтобы открыть диалоговое окно Дискриминантный анализ: Классификация (см. рис. 22.5).
10. Установите флажки Поточечные результаты, Итоговая таблица и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Дискриминантный анализ.
11. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически откроет окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Далее приведены результаты, сгенерированные программой при выполнении шага 5. Как всегда, мы постарались представить данные, выведенные SPSS, в максимально удобной и понятной форме. Из весьма объемных результатов вывода выбраны наиболее важные фрагменты, иллюстрирующие суть дискриминантного анализа. Для удобства представления результатов форматы выводимых таблиц, приведенных на рисунках, изменены без ущерба для их содержания.

Первой в ряду выводимых результатов стоит таблица Групповые статистики. В ней для каждой переменной показаны средние значения и стандартные отклонения на обоих уровнях (в обеих категориях) зависимой переменной, а также общие средние значения. За таблицей Групповые статистики следует таблица Критерии равенства групповых средних, фрагмент которой приведен на рис. 22.6. В ней содержатся коэффициенты Лямбда Уилк-

са, F -критерии (колонка F) и уровни значимости, характеризующие различия средних значений для групп по каждой из переменных (Знч.). Наиболее важная для исследователя информация относится к величинам F -критерия и уровням значимости, поскольку именно по ним можно судить, для каких переменных различие двух групп является значимым.

Критерий равенства групповых средних					
	Лямбда Уилкса	F	ст.св1	ст.св2	Знч.
осведомленность	,901	4,859	1	44	,033
скрытые фигуры	,850	7,746	1	44	,008
пропущенные слова	,931	3,271	1	44	,077
счет в уме	,806	10,563	1	44	,002

Рис. 22.6. Коэффициенты лямбда Уилкса, F -критерии и уровни значимости

Трактовка терминов приведена в конце главы.

На рис. 22.7 представлены фрагменты выводимых результатов, связанные с пошаговым дискриминантным анализом. В трех таблицах содержится статистическая информация о переменных, вошедших и не вошедших в дискриминантное уравнение, а также порядок включения и исключения переменных в процессе составления дискриминантного уравнения. Первая таблица иллюстрирует пошаговый процесс составления дискриминант-

Введенные/исключенные переменные				
Шаг	Введенные	Лямбда Уилкса		
		Статистика	Точное значение F	
			Статистика	Знч.
1	счет в уме	,806	10,562	,002
2	заучивание слов	,708	8,849	,001
3	скрытые фигуры	,639	7,897	,000
4	умозаключения	,606	6,675	,000
5	анalogии	,570	6,043	,000
6	понятливость	,539	5,566	,000
7	экстраверсия	,514	5,125	,000

Переменные в анализе				
Шаг		Толерантность	F исключения	Лямбда Уилкса
7	счет в уме	,677	6,720	,605
	заучивание слов	,932	2,073	,542
	скрытые фигуры	,889	3,012	,555
	умозаключения	,549	5,832	,593
	анalogии	,600	3,006	,555
	понятливость	,855	2,525	,549
	экстраверсия	,886	1,797	,539

Переменные, не включенные в анализ					
Шаг		Толерантность	Мин. Толерантность	F включения	Лямбда Уилкса
7	осведомленность	,768	,537	,064	,513
	пропущенные слова	,662	,509	,014	,514
	исключ. изобр.	,790	,542	,016	,514
	числовые ряды	,555	,522	,096	,513
	геометр. слож.	,519	,519	,000	,514
	нейротизм	,915	,546	1,123	,499

Рис. 22.7. Пошаговое включение и исключение переменных (фрагмент окна вывода)

ного уравнения. В него поочередно вводятся предикторы на основе заданного критерия включения (по умолчанию критерием является $F \geq 3,84$, в нашем случае $F \geq 1,25$), а из него исключаются те предикторы, которые удовлетворяют критерию исключения (по умолчанию таким критерием является $F \leq 2,71$, в нашем случае $F \leq 1$). Представленные таблицы являются завершающими фрагментами очень больших таблиц, которые генерирует программа в результате выполнения семи шагов заданного пошагового метода.

Обратите внимание на то, что все переменные, вошедшие в уравнение после шага 7, имели достаточный уровень толерантности (выше 0,1) и значения F -критерия, превышающие пороговое значение 1,125. Переменные, не попавшие в дискриминантное уравнение, также имеют достаточную толерантность, однако значения F -критерия у них оказались меньше 1,125. В окне вывода можно видеть, что шаги 1–7 приводят к последовательному введению в уравнение новых предикторов. Обращает на себя внимание тот факт, что в окончательный результат не вошли некоторые переменные, различия между группами для которых статистически достоверны (см. рис. 22.6). Напротив, в дискриминантное уравнение были включены переменные экстраверсия и умозаключения, различия между группами по которым статистически недостоверны. Это связано с тем, что при включении переменных в дискриминантное уравнение учитывается не только дискриминативная способность каждой переменной в отдельности, но и ее уникальный вклад в совокупности с остальными переменными.

На рис. 22.8–22.10 приведены очередные таблицы, содержащие основные показатели канонической дискриминантной функции. В таблице на рис. 22.8 значение 1 в графе Функция говорит о том, что в процессе дискриминантного анализа была получена одна дискриминантная функция. Если бы зависимая переменная имела не 2, а 3 уровня, то были бы составлены две дискриминантные функции. Чем больше значение Хи-квадрат, тем сильнее дискриминантная функция различает группы и тем лучше она соответствует своему назначению. О ее состоятельности свидетельствует статистическая значимость Знч., заметно меньшая 0,05.

Собственные значения и Лямбда Уилкса

Функция	Собственное значение	Объясняющая дисперсия	Кумулятивный %	Каноническая корреляция	Лямбда Уилкса	Хи-квадрат	ст. св.	Знч.
1	,944	100,0	100,0	,697	,514	26,926	7	,000

Рис. 22.8. Основные статистики канонической дискриминантной функции

На рис. 22.9 приведена таблица Коэффициенты канонической дискриминантной функции — список нестандартизованных коэффициентов и константа дискриминантного уравнения. Это уравнение подобно линейному уравнению множественной регрессии и применяется для предсказания. По этому уравнению подсчитывается значение функции для каждого из объектов.

Структурная матрица (рис. 22.10) содержит корреляции между дискриминантной функцией и каждой из переменных. Переменные упорядочены по абсолютной величине корреляций.

На рис. 22.11 приведена таблица нормированных коэффициентов канонической дискриминантной функции. Эти коэффициенты служат для определения относительного вклада каждой переменной в значение дискриминантной функции с учетом влияния

Коэффициенты канонической дискриминантной функции

Функция	скрытые фигуры	счет в уме	понятливость	анalogии	умозаключения	заучивание слов	экстраверсия	(Константа)
1	,162	,382	,214	,185	-,241	,157	,097	-9,865

Ненормированные коэффициенты

Рис. 22.9. Коэффициенты канонической дискриминантной функции

Структурная матрица

Функция	счет в уме	понятливость	скрытые фигуры	анalogии	заучивание слов	геометрическое сложение	отгадываемость ^а	числовые ряды ^а	пропущенные слова ^а	исключения из правил ^а	нейротизм ^а	экстраверсия	умозаключения
1	,504	,442	,432	,405	,328	,326	,290	,259	,258	,245	,181	,123	,063

Объединенные внутригрупповые корреляции между дискриминантными переменными и нормированными каноническими дискриминантными функциями

Переменные упорядочены по абсолютной величине корреляций внутри функции

^а Эта переменная не используется в анализе

Рис. 22.10. Структурная матрица

Нормированные коэффициенты канонической дискриминантной функции

	Функция 1
скрытые фигуры	,412
счет в уме	,676
понятливость	,387
анalogии	,501
умозаключения	-,706
заучивание слов	,338
экстраверсия	,324

Рис. 22.11. Нормированные коэффициенты дискриминантной функции

остальных переменных. Чем больше абсолютное значение коэффициента, тем больше относительный вклад данной переменной в значение дискриминантной функции, разделяющей классы.

Таблица, содержащая координаты центроидов групп, приведена на рис. 22.12. Центроид представляет собой значение функции, получаемое при подстановке в дискриминантное уравнение средних значений предикторов. Обратите внимание на то, что центроиды равны по абсолютной величине, но имеют разные знаки. Граничным значением для двух групп является ноль.

Таблица, приведенная на рис. 22.13, содержит информацию о фактической и прогнозируемой группах для каждого объекта, вероятности его принадлежности к группе, а также значения (баллы) дискриминантной функции. В целях экономии места мы привели значения лишь для некоторых объектов. Для объектов, отмеченных двумя звездочками

(**), фактическая и прогнозируемая группы не совпали. Всего таких объектов 5 из 46. В отношении последних 10 объектов, для которых принадлежность к группе не была известна, в таблице представлены результаты предсказания, полученные при помощи уравнения дискриминантной функции.

Функции в центроидах групп

	Функция
оценка успешности	1
низкая	-.950
высокая	.950

Ненормированные канонические дискриминантные функции вычислены в центроидах групп

Рис. 22.12. Координаты центроидов групп

Поточечные статистики

Номер наблюдения	Фактическая группа	Наивероятнейшая группа				Дискриминан- тные баллы
		Предсказанная группа	P(D>d G=g)		P(G=g D=d)	Функция 1
			p	ст. св.		
Исходные	1	1	,994	1	,861	-,958
2	2	2	,853	1	,896	1,135
6	2	1**	,507	1	,633	-,287
47	несгрупп	2	,997	1	,858	,946

** Неправильно классифицированное наблюдение

Рис. 22.13. Фрагмент таблицы Поточечные статистики

Как показывают результаты классификации (рис. 22.14) при данном наборе дискриминантных переменных точность классификации составляет 89,13 % (41 из 46 правильных предсказаний в отношении известных объектов).

Результаты классификации^а

	оценка успешности	Предсказанная принадлежность к группе		Итого
		низкая	высокая	
Исходные Частота	низкая	22	1	23
	высокая	4	19	23
	Несгруппированные наблюдения	7	3	10
%	низкая	95,7	4,3	100,0
	высокая	17,4	82,6	100,0
	Несгруппированные наблюдения	70,0	30,0	100,0

^а 89,1% исходных сгруппированных наблюдений классифицировано правильно.

Рис. 22.14. Результаты классификации

Терминология, используемая при выводе

Трактовка терминов, используемых программой в окне вывода и относящихся к средним значениям, λ Уилкса, F -критериям и уровням значимости (см. рис. 22.6):

- Среднее — средние значения для каждой категории (низкая, высокая) и для всех категорий Итого переменной оценка.

- ▶ Лямбда Уилкса — отношение внутригрупповой суммы квадратов к общей сумме квадратов (λ). Данный коэффициент характеризует долю дисперсии оценок дискриминантной функции, которая не обусловлена различиями между группами, принимает значение 1 в случае, если средние значения для всех групп оказываются равными, и уменьшается с ростом разностей средних значений.
- ▶ F — значения F -критерия, такие же как при однофакторном дисперсионном анализе и равные квадрату t -критериев двух выборок.
- ▶ Знч. — уровни значимости F -критериев, равные вероятности того, что соответствующие различия являются случайными.

Трактовка терминов, используемых программой в окне вывода и относящихся к пошаговому составлению дискриминантного уравнения (см. рис. 22.7):

- ▶ F включения — минимальное значение F , при котором предиктор включается в дискриминантное уравнение.
- ▶ F исключения — максимальное значение F , при котором предиктор исключается из дискриминантного уравнения.
- ▶ Толерантность — мера линейной зависимости между одним предиктором и всеми остальными. Если толерантность окажется меньше 0,001, SPSS воспримет такой результат как наличие значительной линейной зависимости и не включит соответствующий предиктор в дискриминантное уравнение.
- ▶ Знч. — значимости влияния данного предиктора на дисперсию зависимой переменной.

Трактовка терминов, используемых программой в окне вывода и относящихся к канонической дискриминантной функции, ее коэффициентам и центроидам групп (см. рис. 22.8–22.12):

- ▶ Коэффициенты канонической дискриминантной функции — список нестандартизованных коэффициентов и константа дискриминантного уравнения. Это уравнение подобно линейному уравнению множественной регрессии. По этому уравнению подсчитывается значение функции для каждого из объектов.
- ▶ Функция — значение 1 в этой ячейке говорит о том, что в процессе дискриминантного анализа была получена одна дискриминантная функция. Если бы зависимая переменная имела не 2, а 3 уровня, то были бы составлены две дискриминантные функции.
- ▶ Лямбда Уилкса — отношение внутригрупповой суммы квадратов к общей сумме квадратов (λ).
- ▶ Хи-квадрат — мера статистического отличия друг от друга двух уровней дискриминанта (χ^2). Чем больше данное значение, тем сильнее отличие и тем лучше дискриминантная функция соответствует своему назначению.
- ▶ Собственное значение — отношение межгрупповой суммы квадратов к внутригрупповой сумме квадратов. Чем больше данное значение, тем лучше подходит составленная функция для дискриминантного анализа.
- ▶ % объясненной дисперсии, Кумулятивный % — дискриминантная функция всегда вычисляется для дисперсии зависимой переменной, равной 100 %.

- ▶ Структурная матрица — структурная матрица содержит корреляции между значениями дискриминантной функции и каждой из переменных. Переменные упорядочены по абсолютной величине корреляций.
- ▶ Центроиды групп — средние значения дискриминантной функции для двух групп. Если точнее, центроид представляет собой значение функции, получаемое при подстановке в дискриминантное уравнение средних значений предикторов.

Трактовка терминов, используемых программой в таблице Поточечные статистики (см. рис. 22.13).

- ▶ Несгруппированный — объект, для которого заранее не известна принадлежность к группе.
- ▶ Фактическая группа — группа, к которой принадлежит данный объект.
- ▶ Предсказанная группа — группа, вычисленная для объекта с помощью уравнения дискриминантной функции.
- ▶ $P(D > d \mid G = g)$ — вероятность принадлежности объекта к группе (G) при данной величине дискриминантной функции (D).
- ▶ $P(G = g \mid D = d)$ — вероятность наблюдаемого значения дискриминантной функции (D), если задана принадлежность объекта к группе (G).
- ▶ Наивероятнейшая группа — группа, для которой прогнозируемая вероятность включения данного объекта максимальна.
- ▶ Вторая вероятнейшая группа — группа, имеющая вторую по значению вероятность (после прогнозируемой) включения данного объекта. Поскольку число групп в данном случае равно 2, то такая группа для каждого объекта определена заранее.
- ▶ Дискриминантные баллы — дискриминантные оценки, получаемые при подстановке значений переменных объекта в уравнение дискриминантной функции.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

23 Многомерное шкалирование

305	Квадратная асимметричная матрица различий
306	Квадратная симметричная матрица различий
306	Модель индивидуальных различий
307	Пошаговые алгоритмы вычислений
313	Представление результатов
317	Завершение анализа и выход из программы

Основное достоинство многомерного шкалирования — представление больших массивов данных о различии объектов в наглядном, доступном для интерпретации графическом виде. При многомерном шкалировании матрица различий между объектами (вычисленными, например, по их экспертным оценкам) представляется в виде одно-, двух- или трехмерного графического изображения взаимного расположения этих объектов. Доступно и более трех измерений, но эта возможность редко применяется на практике.

Основным преимуществом многомерного шкалирования является возможность очень наглядного визуального сравнения объектов анализа. Если две точки на изображении удалены друг от друга, то между соответствующими объектами имеется значительное расхождение; и наоборот, близость точек говорит о сходстве объектов.

Многомерное шкалирование имеет много общих черт с факторным анализом (см. главу 20). Как и при факторном анализе, создается система координат пространства, в котором определяется расположение точек. Как и при факторном анализе, происходит снижение размерности и упрощение данных. Однако при факторном анализе обычно используются коэффициенты корреляции, а при многомерном шкалировании — меры различия между объектами. Наконец, в факторном анализе наибольший интерес вызывают углы между точками, представляющими собой данные, а в многомерном шкалировании ключевой величиной является расстояние между этими точками.

Многомерное шкалирование имеет несколько общих черт также с кластерным анализом (см. главу 21). В обоих случаях анализируется расстояние между объектами, однако при кластерном анализе типичной является количественная процедура объединения объектов в группы (кластеры), а при многомерном шкалировании качественный анализ объектов проводится визуально с помощью диаграммы.

В этой книге рассматривается наиболее известная процедура многомерного шкалирования SPSS, имеющая название ALSCAL. Фактически ALSCAL не является одной программой, а представляет собой набор процедур, каждая из которых соответствует своему типу данных. Здесь мы проведем несколько вариантов анализа для различных типов данных, постаравшись по возможности коротко и ясно изложить смысл сопутствующих терминов. Если вы сталкиваетесь с многомерным шкалированием впервые, то рекомендуем обратиться к дополнительной литературе.

В первом примере обработаем гипотетическую социограмму для группы учащихся, при этом количественные оценки их взаимоотношений будут преобразованы в соответствующее графическое изображение. Во втором примере рассмотрим результаты тестирования учащихся по пяти показателям и покажем различия между ними графически на плоском изображении. Наконец, третий пример будет представлять собой небольшое исследование восприятия и понимания студентами пяти многомерных методов статистического анализа.

Квадратная асимметричная матрица различий

Представьте себе, что преподаватель решил создать идеальную психологическую обстановку в группе во время занятия, рассадив учащихся так, чтобы ни один из них не оказался рядом с тем, кто ему не нравится. Для этого каждому из 12 студентов было предложено оценить степень своей симпатии к однокурсникам по пятибалльной шкале (от 1 до 5, где 1 — максимум симпатии, а 5 — максимум антипатии). Результаты этого вымышленного опроса мы поместили в файл данных `mds1.sav`. Чтобы добиться желаемого результата, преподавателю необходимо максимально далеко рассадить негативно настроенных в отношении друг друга учащихся. Здесь весьма полезной окажется диаграмма, на которой удаленность точек будет соответствовать отношениям между учащимися. Для построения диаграммы воспользуемся средствами многомерного шкалирования.

Первое, что необходимо сделать для решения задачи, — создать квадратную (12×12) матрицу различий. Позже на основе этой матрицы будет построено двумерное изображение, иллюстрирующее взаимоотношения студентов. В ходе многомерного шкалирования исходная матрица 12×12 преобразуется в гораздо более простую матрицу 12×2 (где 2 — количество измерений или шкал), содержащую координаты точек для изображения. Исходную матрицу называют *квадратной асимметричной матрицей различий*. Поясним, что означают составляющие это определение термины.

- **Квадратная матрица** — это матрица, строки и столбцы которой представляют собой один и тот же набор объектов. В данном случае этим набором объектов является группа учащихся: каждый учащийся оценивает меру своей симпатии/антипатии к каждому из остальных. Если бы строки и столбцы матрицы представляли не одни и те же, а разные объекты, матрица называлась бы не квадратной, а прямоугольной. Прямоугольные матрицы в контексте данной книги не рассматриваются.
- **Асимметричная матрица** — это матрица, для которой отношение двух объектов друг к другу может быть разным. Так, например, симпатия Петра к Ирине не означает, что Ирине Петр тоже симпатичен. Визуально асимметричность матрицы выражается в том, что как минимум для одной пары ячеек, симметрично расположенных относительно главной диагонали матрицы, значения различны. Если же ни одного такого различия для матрицы не установлено, матрица называется симметричной. Примером матрицы, которая всегда симметрична, является корреляционная матрица.
- **Матрица различий** — матрица, данные которой представляют собой меру различия. В данном случае значения матрицы отражают степень отличия отношения одного студента к другому от идеального: чем больше значение, тем больше различие. Существуют также матрицы сходств, в которых значения отражают степень некоторого сходства. Если бы преподаватель вместо матрицы различий составил матрицу

сходств, то изображение, полученное в результате многомерного шкалирования, скорее внесло бы путаницу, чем помогло в решении задачи.

Квадратная симметричная матрица различий

Пусть наш гипотетический преподаватель решает другую задачу: ему нужно рассадить 12 учащихся в соответствии с результатами их тестирования по пяти показателям. Поскольку результаты тестирования не относятся к данным, характеризующим различия, необходимо сначала вычислить различия по имеющимся данным и таким образом свести задачу к предыдущей. Исходные данные для этой задачи естественно представить в виде прямоугольной матрицы 12×5 , в которой для каждого из 12 учащихся указаны результаты пяти тестов (файл `mds2.sav`). Затем по исходным данным строится квадратная (12×12) симметричная матрица различий между учащимися. Наконец, как и в предыдущем примере, SPSS создает матрицу координат 12×2 и визуально представляет ее в виде диаграммы.

Обратите внимание на два ключевых свойства матрицы различий: она является *квадратной* и *симметричной*. Несмотря на то что исходная матрица является прямоугольной, то есть ее строки (объекты) соответствуют учащимся, а столбцы (переменные) — тестам, в матрице различий как строки, так и столбцы соответствуют учащимся, следовательно, матрица является квадратной размером 12×12 . Кроме того, поскольку учащийся 1 отличается от учащегося 5 по результатам тестирования так же, как учащийся 5 от учащегося 1, матрица различий является симметричной.

Модель индивидуальных различий

В обоих рассмотренных примерах создавалась единственная матрица различий, на основе которой вычислялись координаты точек-объектов для визуального представления. Однако в некоторых случаях возможно координатное представление нескольких матриц различий, примером может служить ситуация, когда несколько экспертов оценивают один и тот же набор объектов. В таких случаях говорят о многомерном шкалировании индивидуальных различий.

Предположим, преподаватель предлагает студентам сравнить между собой 5 многомерных статистических методов: множественный регрессионный анализ (`mra`), факторный анализ (`fa`), кластерный анализ (`ka`), дискриминантный анализ (`da`) и многомерное шкалирование (`mds`). Каждому студенту предъявляются все возможные пары этих методов для оценки различия между ними (1 — максимально сходны, 5 — максимально различны). Затем по данным для 6 студентов составляются 6 матриц различий (5×5 каждая).

Поскольку SPSS предъявляет определенные требования к формату данных при использовании модели индивидуальных различий, мы создали специальный файл `mds3.sav`. Требования к формату файла, по сути, сводятся к наличию в этом файле нескольких квадратных матриц одинакового размера. Поскольку в данном примере мы имеем дело с 6 студентами, оценивающими 5 объектов (многомерных методов), в файле `mds3.sav` присутствуют 6 матриц размером 5×5 . Первые 5 строк файла соответствуют первой матрице, следующие 5 строк — второй матрице и т. д. Всего под матрицы отведено 30 строк.

Пошаговые алгоритмы вычислений

Для проведения многомерного шкалирования сначала необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае mds1.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом mds1.sav.

ШАГ 4

В меню Анализ выберите команду Шкалирование ► Многомерное шкалирование (ALSCAL). Откроется диалоговое окно Многомерное шкалирование (рис. 23.1).

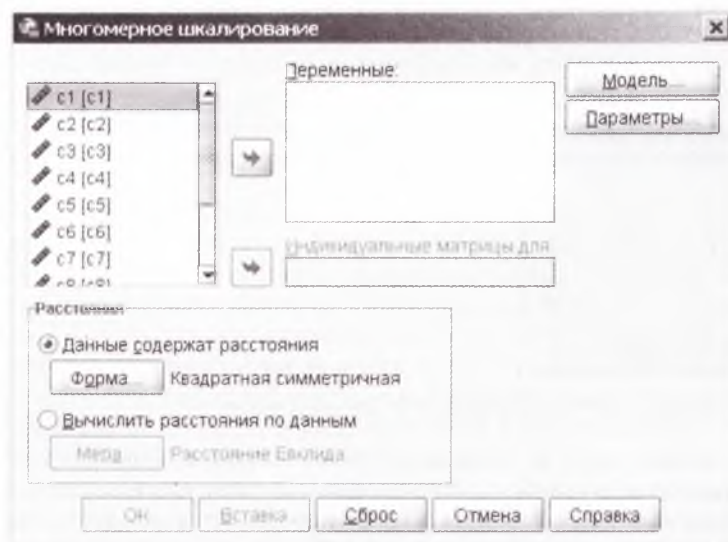


Рис. 23.1. Диалоговое окно Многомерное шкалирование

В диалоговом окне Многомерное шкалирование вы можете задать список переменных, которые будут участвовать в анализе. В SPSS существует два типа многомерного шкалирования ALSCAL: с заданной матрицей различий и с вычислением матрицы различий.

В том случае, когда матрица различий задана, вы при желании можете указать столбцы файла, которые будут участвовать в анализе. Обратите внимание на следующий нюанс: вы можете включить в процедуру многомерного шкалирования часть столбцов файла данных, однако не можете включить часть строк. Последнее достигается косвенно — при помощи описанной подробно в главе 4 команды Отобрать наблюдения, однако вы не можете исключить лишние объекты непосредственно. Если используется квадратная матрица, число выбранных переменных (столбцов) должно совпадать с числом объектов файла (строк).

В случае если вы не используете матрицу различий из файла данных, а хотите, чтобы программа SPSS создала ее самостоятельно, требуется указать переменные, для которых будут вычислены расстояния между объектами. В обоих случаях выбор переменных сводится к одной и той же процедуре: вы выделяете нужную переменную (или несколь-

ко переменных) в исходном списке в левой части диалогового окна и щелчком на верхней кнопке со стрелкой помещаете ее (их) в поле Переменные.

Если вы хотите, чтобы программа SPSS вычислила расстояния между объектами вашего файла, и одновременно намереваетесь использовать модель многомерного шкалирования для нескольких выборок, то можете задать имя нужной переменной в поле Индивидуальные матрицы для, однако эта процедура весьма трудоемка, и мы не будем на ней останавливаться.

После того как заданы переменные, которые должны быть включены в анализ, следует информировать программу SPSS о структуре вашего файла данных. Для этого предназначен переключатель Расстояния. Если файл уже содержит матрицу различий, этот переключатель следует оставить установленным в положение Данные содержат расстояния. В случае если матрица с данными является квадратной и симметричной (значения над ее главной диагональю равны значениям под главной диагональю), дополнительно определять тип данных не нужно. Если же вы намерены использовать асимметричную или прямоугольную матрицу, щелкните на кнопке Форма, и на экране появится диалоговое окно Многомерное шкалирование: Форма данных (рис. 23.2).

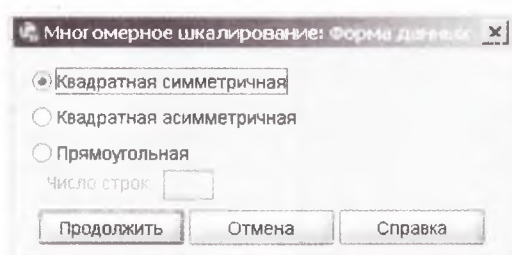


Рис. 23.2. Диалоговое окно Многомерное шкалирование: Форма данных

В этом окне вы можете выбрать один из возможных типов матрицы. Если матрица квадратная (число строк равно числу столбцов) и симметричная (значения над главной диагональю равны значениям под главной диагональю), то переключатель следует оставить установленным в положение Квадратная симметричная. Этот тип матрицы используется в примере с многомерными методами, поскольку сравнения одного метода с другим и наоборот дают одинаковый результат. Если матрица является квадратной, но не симметричной, переключатель следует установить в положение Квадратная асимметричная. Такой тип матрицы используется в примере с социограммой студентов. Наконец, если матрица не является квадратной, установите переключатель в положение Прямоугольная и введите число строк в расположенное рядом поле. Этот тип матрицы используется редко, например при шкалировании предпочтений, и в данной книге не описывается.

Если необходимо, чтобы программа SPSS создала матрицу различий по имеющимся данным, следует в окне Многомерное шкалирование установить переключатель Расстояния в положение Вычислить расстояния по данным. В ответ SPSS создаст матрицу различий для тех переменных, которые вы указали в списке Переменные. По умолчанию вычисляется евклидово расстояние (по формуле Пифагора $a^2 + b^2 = c^2$), и если такой способ вычисления вас устраивает, никаких дополнительных действий не требуется. Если же вы хотите применить иной способ вычисления расстояния, следует щелкнуть на кнопке Мера, открыв диалоговое окно Многомерное шкалирование: Создание меры по данным (рис. 23.3).

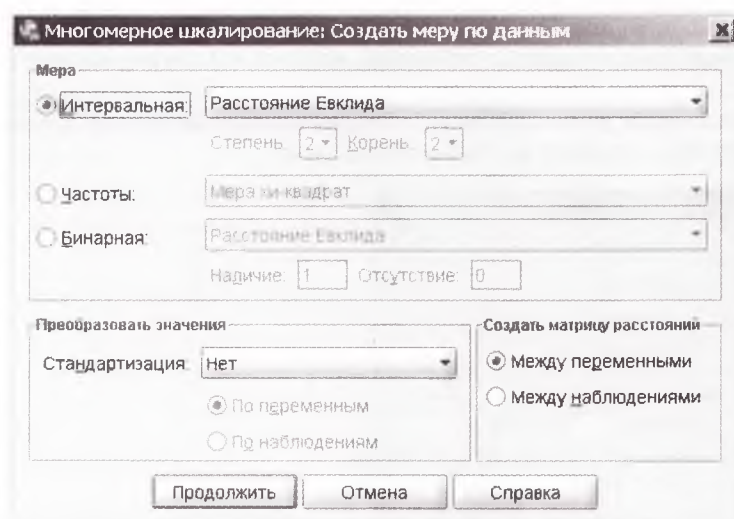


Рис. 23.3. Диалоговое окно Многомерное шкалирование: Создать меру по данным

Как видите, диалоговое окно содержит 3 области: Мера, Преобразовать значения и Создать матрицу расстояний. Помимо формулы для евклидова расстояния существует еще несколько довольно распространенных формул, например формула Минковского, которая напоминает формулу Пифагора, однако позволяет вместо квадратов использовать любую степень, например четвертую: $a^4 + b^4 = c^4$. Кроме этого полезным способом определения расстояния является метрика города. Для того чтобы понять ее суть, представьте себе городской район, в котором вы можете передвигаться только по улицам, огибая дома. Метрика города применяется при использовании разнородных по смыслу и (или) ранговых переменных.

Если все переменные рабочего файла имеют одну и ту же шкалу (например, целые числа от 1 до 5), вероятно, не придется вместо пункта Нет, выбранного в списке Стандартизация по умолчанию, выбирать что-то другое. Если же для переменных применяются различные шкалы, то может понадобиться какой-либо из вариантов стандартизации, предлагаемых SPSS. Чаще всего используется вариант z-значения. И наконец, по умолчанию SPSS вычисляет расстояния между переменными, однако при желании можно потребовать вычислить расстояние между объектами. Для этого переключатель Создание матрицы расстояний необходимо установить не в положение Между переменными, а в положение Между наблюдениями.

После задания режима вычисления расстояний можно выбрать желаемый тип модели многомерного шкалирования. Для этого в окне Многомерное шкалирование щелкните на кнопке Модель, и на экране появится диалоговое окно Многомерное шкалирование: Модель (рис. 23.4). Наибольший интерес в этом окне представляют переключатели Шкала измерения и Модель шкалирования. В отличие от факторного анализа (см. главу 20), при многомерном шкалировании программа не может автоматически определить количество шкал, которое следует использовать. Поскольку в подавляющем большинстве случаев требуется не более трех шкал, выбор нужного числа измерений не представляет трудностей: вы легко можете перебрать все варианты представления и остановиться на том, который наиболее удобен для зрительного восприятия. Чаще всего при многомерном шкалировании применяют модель евклидовых расстояний, в ней вам необходимо задать

либо одну матрицу различий, либо несколько идентичных матриц для анализа. Однако если матрицы различий получены от разных субъектов, которых вы хотите сравнить между собой, придется использовать модель индивидуальных различий. Примером служит ситуация с упоминавшимися ранее многомерными методами: каждый из шести студентов сравнивает 5 методов, оценивая различия между ними.

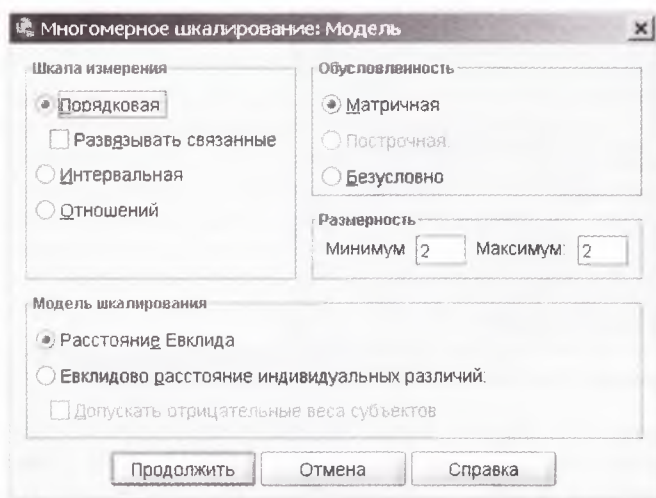


Рис. 23.4. Диалоговое окно Многомерное шкалирование: Модель

Переключатель Обусловленность позволяет выбрать, какие условия сравнения следует выполнять: если он установлен в положение Матричная, то все ячейки каждой матрицы сравниваются между собой (эта ситуация соответствует примеру с методами, когда сравниваются различия, оцененные одним студентом, но не сравниваются различия из разных таблиц). Установка переключателя в положение Построчная означает, что сравнения производятся между ячейками одной строки. Такую установку следует сделать в примере с социограммой, поскольку в нем каждая строка таблицы различий соответствует субъекту, оценивающему свое отношение к однокурсникам. В общем случае учащиеся могут использовать разные шкалы, однако сравнения в рамках одной строки имеют смысл. Наконец, установка переключателя в положение Безусловно означает, что в программе предполагаются сравнения всех ячеек по всем матрицам.

Переключатель Шкала измерения позволяет указать программе, в какой шкале интерпретировать величины матрицы различий. Как правило, между результатами установки в положения Порядковая, Интервальная и Отношений нет существенных различий.

Последним рассматриваемым элементом интерфейса диалогового окна Многомерное шкалирование является кнопка Параметры. При щелчке на ней открывается диалоговое окно Многомерное шкалирование: Параметры (рис. 23.5). Наибольший интерес здесь представляет область Вывести. Флажок Групповые графики практически всегда следует устанавливать, поскольку в противном случае вам придется рисовать изображение вручную. Флажок Матрица данных позволяет выводить на экран исходные данные. Эта возможность оказывается весьма полезной, если вы используете SPSS для создания матрицы различий. Флажок Сводка по модели и параметрам позволяет включить в выводимые данные тип и параметры выбранной модели. Остальные элементы

интерфейса диалогового окна Многомерное шкалирование: Параметры используются очень редко.

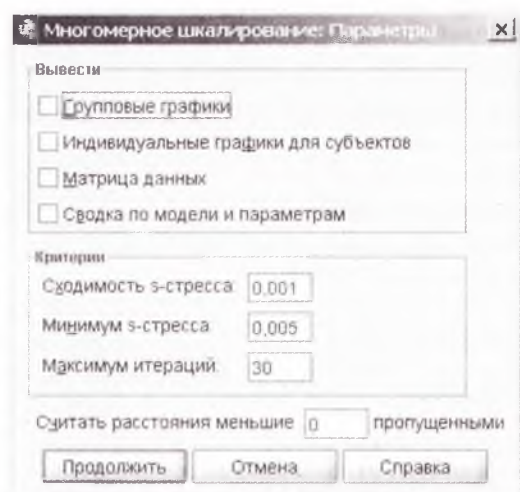


Рис. 23.5. Диалоговое окно Многомерное шкалирование: Параметры

Следующий пример иллюстрирует практическую реализацию двумерного шкалирования для заданной матрицы различий. Подробное описание задачи приведено в разделе «Квадратная асимметричная матрица различий».

ШАГ 5

После выполнения предыдущего шага у вас должно быть открыто диалоговое окно Многомерное шкалирование (см. рис. 23.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной c_1 , нажмите клавишу Shift и, не отпуская ее, щелкните на переменной c_{12} . В результате окажутся выделенными все промежуточные переменные, начиная с c_1 и заканчивая c_{12} .
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
3. Щелкните на кнопке Форма, чтобы открыть диалоговое окно Многомерное шкалирование: Форма данных (см. рис. 23.2).
4. Установите переключатель в положение Квадратная асимметричная и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
5. Щелкните на кнопке Модель, чтобы открыть диалоговое окно Многомерное шкалирование: Модель (см. рис. 23.4).
6. Установите переключатель Обусловленность в положение Построчная и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
7. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Многомерное шкалирование: Параметры (см. рис. 23.5).
8. Установите флажок Групповые графики и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
9. Щелкните на кнопке ОК, чтобы открыть окно вывода.

В следующем примере демонстрируется двумерное шкалирование квадратной симметричной матрицы различий, которую SPSS создает при задании переменных из файла данных. Данные матрицы различий имеют интервальный тип. Подробное описание задачи было приведено в разделе «Квадратная симметричная матрица различий». В этом примере используется файл данных mds2.sav.

ШАГ 3а

Откройте файл данных, с которым вы намерены работать (в нашем случае mds2.sav). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Данные.
2. В открывшемся диалоговом окне дважды щелкните на имени mds2.sav или введите его с клавиатуры и щелкните на кнопке ОК.

ШАГ 4а

В меню Анализ выберите команду Шкалирование ► Многомерное шкалирование (ALSCAL). Откроется диалоговое окно Многомерное шкалирование (см. рис. 23.1).

ШАГ 5а

После выполнения шага 4а у вас должно быть открыто диалоговое окно Многомерное шкалирование. Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной тест1, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной тест5. В результате окажутся выделенными все промежуточные переменные, начиная тест1 и заканчивая тест5.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
3. Установите переключатель Расстояния в положение Вычислить расстояния по данным и щелкните на кнопке Мера, чтобы открыть диалоговое окно Многомерное шкалирование: Создать меру по данным (см. рис. 23.3).
4. Установите переключатель Создать матрицу расстояний в положение Между наблюдениями и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
5. Щелкните на кнопке Модель, чтобы открыть диалоговое окно Многомерное шкалирование: Модель (см. рис. 23.4).
6. Установите переключатель Шкала измерения в положение Интервальная и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
7. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Многомерное шкалирование: Параметры (см. рис. 23.5).
8. Установите флажок Групповые графики и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
9. Щелкните на кнопке ОК, чтобы открыть окно вывода.

Последним мы рассмотрим пример двумерного шкалирования с использованием нескольких квадратных симметричных матриц и модели индивидуальных различий. Подробное описание задачи было приведено в разделе «Модель индивидуальных различий» этой главы. Обратите внимание на то, что для выполнения анализа используется файл mds3.sav.

ШАГ 36 Откройте файл данных, с которым вы намерены работать (в нашем случае `mds3.sav`). Если он расположен в текущей папке, то выполните следующие действия:

1. Выберите в меню Файл команду Открыть ► Данные.
2. В открывшемся диалоговом окне дважды щелкните на имени `mds3.sav` или введите его с клавиатуры и щелкните на кнопке ОК.

ШАГ 46 В меню Анализ выберите команду Шкалирование ► Многомерное шкалирование (ALSCAL). Откроется диалоговое окно Многомерное шкалирование (см. рис. 23.1).

ШАГ 56 После выполнения предыдущего шага у вас должно быть открыто диалоговое окно Многомерное шкалирование. Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните на переменной `mra`, нажмите клавишу Shift и, не отпуская ее, щелкните на переменной `mds`. В результате окажутся выделенными все промежуточные переменные, начиная `mra` и заканчивая `mds`.
2. Щелкните на верхней кнопке со стрелкой, чтобы переместить выделенные переменные в поле Переменные.
3. Щелкните на кнопке Модель, чтобы открыть диалоговое окно Многомерное шкалирование: Модель (см.рис. 23.4).
4. Установите переключатель Шкала измерения в положение Отношений, а переключатель Модель шкалирования — в положение Евклидово расстояние индивидуальных различий, после чего щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
5. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Многомерное шкалирование: Параметры (см. рис. 23.5).
6. Установите флажок Групповые графики и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Многомерное шкалирование.
7. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически откроет окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

В этом разделе приведены данные, сгенерированные командой Многомерное шкалирование (ALSCAL) при выполнении шагов 5, 5а и 5б.

Итерации

Значения, записанные в столбце S-stress (рис. 23.6), характеризуют отклонение результата от идеального (точно соответствующего матрице различий) на различных ите-

рациях применения модели. SPSS применяет заданную модель столько раз, сколько необходимо для получения достаточно низкого значения в столбце S-stress. Если число итераций оказывается больше 30, это, как правило, указывает на проблемы с исходными данными.

Iteration history for the 2 dimensional solution (in squared distances)		
Young's S-stress formula 1 is used.		
Iteration	S-stress	Improvement
1	.36850	
2	.34885	.01965
3	.33645	.01240
4	.32766	.00879
5	.32256	.00509
6	.32076	.00181
7	.32003	.00072
Iterations stopped because		
S-stress improvement is less than .001000		

Рис. 23.6. Фрагмент окна вывода после выполнения шага 5 (итерации)

Стрессы и квадраты коэффициентов корреляции

Для каждой строки асимметричной матрицы различий, для каждой матрицы модели индивидуальных различий, а также для всей модели при многомерном шкалировании вычисляются стресс и коэффициент R^2 (рис. 23.7). Стресс по своему смыслу схож со стрессом предыдущей модели, однако для его расчета используется другое уравнение, позволяющее упростить вычислительный процесс сравнения различий. Коэффициент R^2 (столбец RSQ) характеризует долю дисперсии в матрице различий, обусловленную данной моделью. Чем лучше модель, тем выше значение коэффициента R^2 . Если вы, к примеру, строите несколько графических изображений с разным числом шкал-координат, то значения стресса и R^2 могут служить критериями при выборе наиболее подходящей модели.

Stress and squared correlation (RSQ) in distances					
Matrix 1					
Stimulus	Stress	RSQ	Stimulus	Stress	RSQ
1	.225	.826	2	.293	.690
3	.222	.846	4	.360	.380
5	.239	.750	6	.188	.839
7	.206	.796	8	.256	.713
9	.249	.716	10	.344	.607
11	.250	.722	12	.162	.880
Averaged (rms) over stimuli					
Stress = .256		RSQ = .730			

Рис. 23.7. Фрагмент окна вывода после выполнения шага 5 (стрессы и квадраты коэффициентов корреляции)

Координаты стимулов

Для каждого шкалируемого объекта указываются его координаты по каждой шкале (рис. 23.8). Это сделано для того, чтобы вы могли на основе этих координат построить собственное графическое изображение или использовать координаты для дальнейшего анализа. В данном случае левый столбец соответствует координате x , а правый — координате y .

Stimulus Coordinates			
Stimulus Number	Stimulus Name	Dimension	
		1	2
1	C1	.9541	1.1290
2	C2	1.3036	.2898
3	C3	1.1314	.8383
4	C4	-.7832	.7595
5	C5	-1.6958	.2091
6	C6	-1.6310	-.0369
7	C7	-.3179	-1.5237
8	C8	.0372	-1.6126
9	C9	-1.0067	.7802
10	C10	.8591	.7622
11	C11	.8739	-.0226
12	C12	.2754	-1.5722

Рис. 23.8. Фрагмент окна вывода после выполнения шага 5 (координаты стимулов)

Субъективные веса

Если используется модель индивидуальных различий, то SPSS включает в окно вывода подсчитанные значения весов для каждого субъекта (рис. 23.9). В примере со сравнением многомерных методов каждый из 6 студентов по каждой из двух шкал имеет свой вес, который характеризует, какая доля дисперсии субъективных оценок приходится на соответствующую шкалу. Кроме того, в вывод включен общий вес для каждой из шкал. Величина Weirddness (Предсказуемость) характеризует разброс весов для каждого субъекта; для субъекта с номером 3 она имеет максимальное значение, равное 0,51.

Subject Weights			
Subject Number	Weirddness	Dimension	
		1	2
1	.1841	.7054	.6774
2	.1869	.7075	.6825
3	.5110	.8968	.2596
4	.1161	.7555	.6500
5	.1431	.7258	.6521
6	.2415	.8951	.4346
Overall importance of each dimension:		.6169	.3381

Рис. 23.9. Фрагмент окна вывода после выполнения шага 5б (субъективные веса)

Итоговая конфигурация объектов

Диаграмма, показанная на рис. 23.10, представляет собой итог применения модели многомерного шкалирования (в результате выполнения шага 5). Она отображает взаимоотношения 12 студентов таким образом, что чем больше различия между учащимися в исходной матрице, тем дальше они находятся друг от друга на диаграмме. На ней видно, что в исследуемой группе выделяются три относительно компактные подгруппы, самая крупная из которых состоит из пяти человек и располагается в правом верхнем углу диаграммы. Отношения внутри каждой из группировок характеризуются симпатией (точки расположены близко), чего не скажешь об отношениях между группами. В данном случае смысл каждой из шкал не имеет значения — главным является взаимное расположение точек.

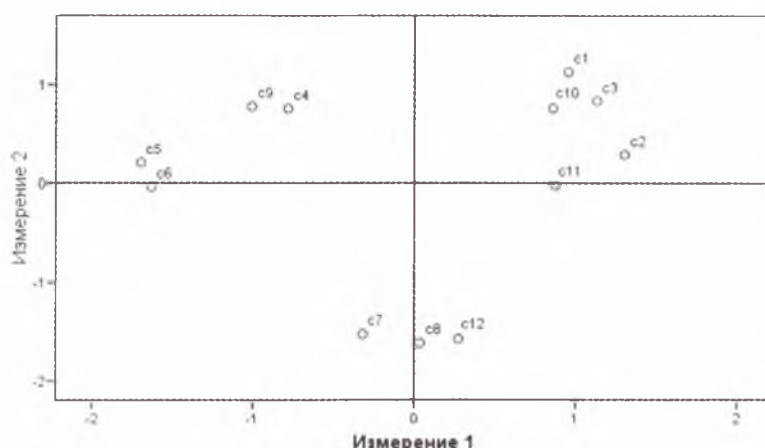


Рис. 23.10. Фрагмент окна вывода после выполнения шага 5 (конфигурация стимулов)

На диаграмме, приведенной на рис. 23.11, изображена итоговая конфигурация сравниваемых студентами статистических методов (результат выполнения шага 5б). В этом случае можно попытаться дать содержательную интерпретацию выделенных шкал как критериев, которыми пользуются студенты при сравнении методов.

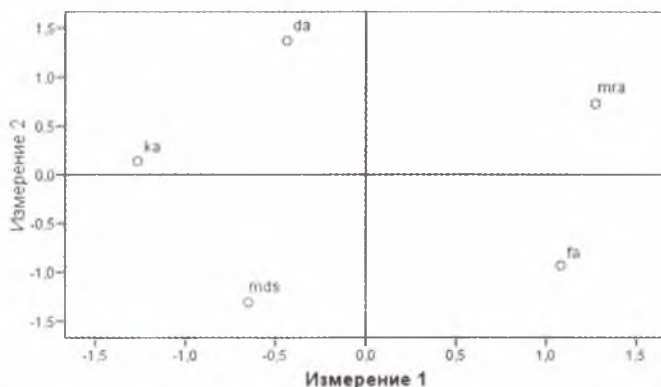


Рис. 23.11. Фрагмент окна вывода после выполнения шага 5б (итоговая диаграмма)

На правом полюсе шкалы 1 расположились множественный регрессионный (mra) и факторный анализ (fa), а на левом — кластерный анализ (ka) и многомерное шкалирование (mds). Таким образом, эта шкала различает методы по предмету анализа: правый полюс шкалы соответствует корреляциям (факторный и регрессионный анализ), левый — различиям и расстояниям (шкалирование и кластерный анализ). На положительном полюсе шкалы 2 оказались дискриминантный (da) и множественный регрессионный анализ (mra), а на отрицательном — многомерное шкалирование (mds) и факторный анализ (fa). По всей видимости, эта шкала разделяет методы по их назначению: положительный полюс соответствует методам предсказания, а отрицательный — структурным методам. О значении этих шкал как субъективных критериев, которыми пользовались студенты при сравнении методов, можно судить по субъективным весам (см. рис. 23.9).

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

24 Логистическая регрессия

319	Математическое описание логистической регрессии
320	Пошаговые алгоритмы вычислений
321	Представление результатов
327	Терминология, используемая при выводе
328	Завершение анализа и выход из программы

Логистическая регрессия представляет собой расширение множественной регрессии, рассмотренной в главе 18, и отличается от последней тем, что в качестве зависимой переменной используется не количественная, а дихотомическая переменная, имеющая лишь два возможных значения. Как правило, эти два значения символизируют принадлежность или непринадлежность объекта к какой-либо группе, ответ типа «да» или «нет» и т. п.

Поскольку логистическая и множественная регрессии основаны на сходных принципах, мы будем исходить из того, что вы уже с ними знакомы (см. главу 18). Кратко смысл регрессионного анализа можно свести к нахождению аналитического выражения, наиболее адекватно отражающего связь между зависимой переменной и множеством независимых переменных. Главным отличием логистической регрессии от множественной является толкование уравнения регрессии. Если множественная регрессия позволяет прогнозировать количественное значение зависимой переменной (критерия) на основе известных значений независимых переменных (предикторов), то логистическая регрессия прогнозирует вероятность некоторого события, находящуюся в пределах от 0 до 1. Кроме того, при помощи индикаторной схемы кодирования допускается использование в качестве предикторов категориальных (номинативных) переменных. Более детально особенности логистической регрессии рассмотрены в разделе «Представление результатов», сейчас же достаточно знать, что категориальный предиктор может быть представлен серией бинарных переменных — по одной на каждую категорию предиктора. Этим бинарным переменным присваивается значение 1 или 0 в зависимости от того, к какой категории относится объект.

Математически суть логистической регрессии излагается на примере переменных из файла helpLR.sav. Содержимое этого файла примерно то же, что и файла help.sav, который использовался в главе 18. При помощи файла help.sav прогнозировалось количественное значение переменной помощь в зависимости от ряда количественных предикторов (симпатия, проблема, эмпатия, польза и агрессия). В данном случае мы будем прогнозировать не величину помощи, а мнение партнера о том, полезна или нет оказанная ему помощь. Как нетрудно понять, речь идет о дихотомической зависимой переменной, поскольку предполагается лишь два варианта ответа: «да» или «нет». Помимо того что в файле helpLR.sav зависимая переменная польза представлена в дихотомическом виде, в него добавлена еще одна категориальная переменная условия. Эта переменная обозначает одно из трех условий, в которых моделировалось оказание помощи: 1 — на работе, 2 — на улице в людном месте, 3 — в лесу, при встрече один на один. Данные для файла helpLR.

sav подбирались искусственно, для того чтобы проиллюстрировать рассматриваемый тип статистического анализа.

Математическое описание логистической регрессии

С логистической регрессией связаны такие математические понятия, как вероятность, шанс и натуральный логарифм шанса. Вероятность — это ожидаемая относительная частота некоторого события. Шанс представляет собой отношение вероятности того, что событие произойдет, к вероятности того, что событие не произойдет. Так, если вероятность дождя равна 0,2, то вероятность отсутствия дождя равна 0,8, следовательно, шанс, что дождь все-таки прольется, равен $0,2/0,8 = 0,25$. Обратите внимание на то, что шанс, в отличие от вероятности, не ограничен максимальным единичным значением: если, к примеру, вероятность дождя составляет не 0,2, а 0,8, то получаем шанс $0,8/0,2 = 4$. Единичное значение шанса соответствует ситуации, когда вероятности события и его отсутствия равны.

Ключевым параметром логистической регрессии является *логит*. Логит равен натуральному логарифму шанса. Например, логит вероятности 20 % равен $-1,386$.

Уравнение регрессии, используемое в главе 18, имеет следующий вид:

$$\text{помощь} = B_0 + B_1 \cdot \text{симпатия} + B_2 \cdot \text{агрессия} + B_3 \cdot \text{польза}.$$

Согласно этому уравнению величина оказываемой помощи равна сумме константы (B_0) и значений трех переменных (отражающих симпатию, агрессивность и пользу), умноженных на соответствующие коэффициенты регрессии. Несмотря на то что в логистическом анализе оценивается полезность или бесполезность помощи, а не ее величина, уравнение логистической регрессии похоже на уравнение множественной регрессии:

$$\ln \left[\frac{P_{\text{помощь}}}{1 - P_{\text{помощь}}} \right] = B_0 + B_1 x_1 + B_2 x_2 + B_3 x_3,$$

где $P_{\text{помощь}}$ — вероятность оказания помощи; x_1 , x_2 и x_3 — значения переменных симпатия, агрессия и польза соответственно.

Можно избавиться от натурального логарифма в левой части и преобразовать уравнение к виду

$$\frac{P_{\text{помощь}}}{1 - P_{\text{помощь}}} = e^{B_0} + e^{B_1 x_1} + e^{B_2 x_2} + e^{B_3 x_3}.$$

Это уравнение, в свою очередь, можно преобразовать к следующему:

$$P_{\text{помощь}} = \frac{1}{1 + e^{-B_0} \cdot e^{-B_1 x_1} \cdot e^{-B_2 x_2} \cdot e^{-B_3 x_3}}.$$

Вероятно, подобные формулы не дают интуитивного представления о зависимости, которую они отражают, но на первых этапах это вполне нормально, поскольку интуиция всегда приходит с опытом. Тем не менее следует проявить максимум внимания к интерпретации модели логистической регрессии, поскольку выбор модели требует ее ясного теоретического понимания.

Поскольку данный раздел статистики весьма непрост для изучения, рекомендуем перед проведением логистического регрессионного анализа почитать дополнительную литературу.

Пошаговые алгоритмы вычислений

Перед началом логистического регрессионного анализа необходимо выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае helpLR.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом helpLR.sav.

ШАГ 4

В меню Анализ выберите команду Регрессия ► Логистическая. На экране появится диалоговое окно Логистическая регрессия (рис. 24.1).

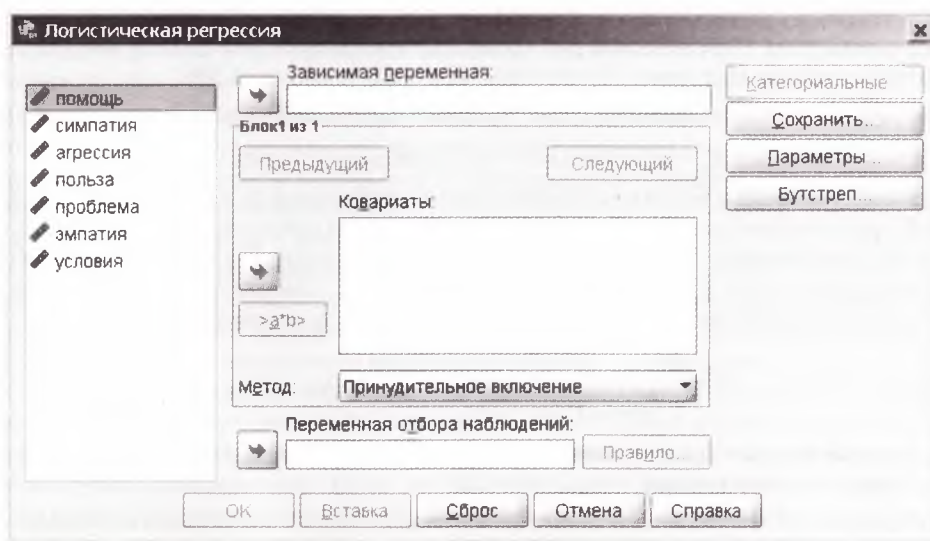


Рис. 24.1. Диалоговое окно Логистическая регрессия

В диалоговом окне Логистическая регрессия вводятся имена зависимой переменной и предикторов, а также задаются другие параметры анализа. Как обычно, в его левой части располагается список всех переменных файла данных. Справа от списка находится поле Зависимая переменная, в которое нужно поместить имя бинарной переменной, значения которой должны прогнозироваться в процессе анализа. В нашем примере в качестве зависимой будет использоваться переменная помощь, имеющая два уровня: 0 (нет) означает неоказанную помощь, а 1 (есть) — оказанную. Особенностью логистического анализа является то, что в процессе его проведения любые значения двух уровней зависимой переменной заменяются нулем и единицей. Это всегда происходит автоматически, и вам не нужно предпринимать никаких дополнительных действий для перекодирования переменной.

Назначение поля Переменная отбора наблюдений примерно такое же, как у кнопки Если, упоминавшейся в главе 4. Оно позволяет указать имя переменной и ее уровень для участия в анализе только части объектов.

В центре окна Логистическая регрессия находится поле Ковариаты, предназначенное для задания предикторов, что, как обычно, выполняется щелчком на кнопке со стрелкой.

Помимо задания обычных переменных вы можете указать способ их взаимодействия. Для этого следует выделить те переменные, которые участвуют во взаимодействии, а затем щелкнуть на кнопке $>a*b>$. Однако при этом следует помнить, что подобные действия ведут к не вполне очевидным последствиям и для успешной интерпретации результатов вы должны иметь ясное концептуальное представление о подобных действиях.

Раскрывающийся список Метод позволяет выбрать способ построения уравнения регрессии. Пункт Принудительное включение, выбранный по умолчанию, предполагает включение в регрессионное уравнение всех указанных предикторов независимо от того, имеют они значимое влияние на зависимую переменную или нет. Пункт Включение: ОП (ОП — отношение правдоподобия) предполагает пошаговое включение в уравнение предикторов, оказывающих наибольшее воздействие на зависимую переменную, до последнего предиктора, чье воздействие окажется значимым. Пункт Исключение: ОП предполагает обратный принцип: сначала в уравнение регрессии включаются все выбранные предикторы, а затем происходит пошаговое исключение тех предикторов, влияние которых на зависимую переменную оказывается недостаточным согласно установленному критерию. Исключение проводится до тех пор, пока в уравнении не останутся только те предикторы, которые удовлетворяют заданному критерию. Описанные три метода из представленных в списке Метод используются чаще других.

В случае если вы хотите применить к разным предикторам различные методы, следует разделить предикторы на несколько блоков с помощью кнопок Предыдущий и Следующий, а затем для каждого блока указать собственный метод. SPSS будет последовательно обрабатывать блоки указанными методами. В нашем примере мы не станем делить предикторы на блоки, а выберем для всех предикторов метод Включение: ОП.

Если в анализе используются категориальные предикторы, то после задания всех предикторов в поле Ковариаты следует воспользоваться кнопкой Категориальные. На экране появится диалоговое окно Логистическая регрессия: Задать категориальные переменные (рис. 24.2).

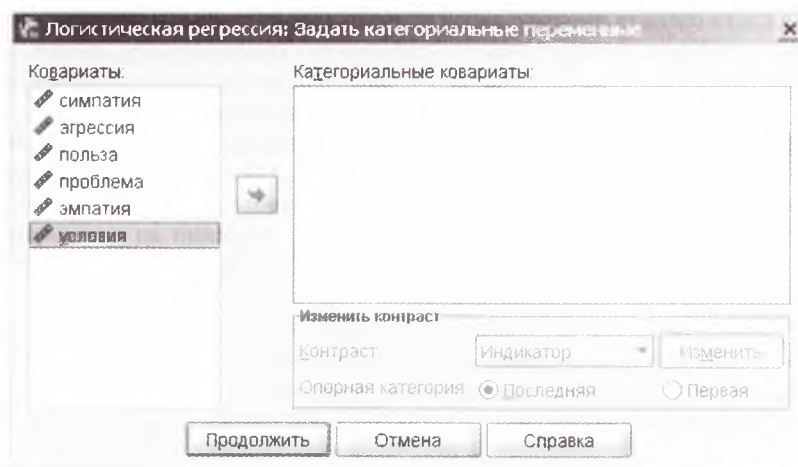


Рис. 24.2. Диалоговое окно Логистическая регрессия: Задать категориальные переменные

В списке Ковариаты вы можете видеть имена всех предикторов, указанных вами в одноименном списке диалогового окна Логистическая регрессия. В нашем примере переменная условие является не количественной, а категориальной, поэтому ее следует выделить и переместить из списка Ковариаты в поле Категориальные ковариаты. Допускается применение нескольких способов обработки категориальных ковариат, каждый из которых задается при помощи списка Изменить контраст, находящегося в правом нижнем углу диалогового окна.

Кнопка Параметры в диалоговом окне Логистическая регрессия позволяет задать несколько дополнительных параметров, иногда оказывающихся весьма полезными на практике. При щелчке на ней открывается диалоговое окно Логистическая регрессия: Параметры (рис. 24.3).

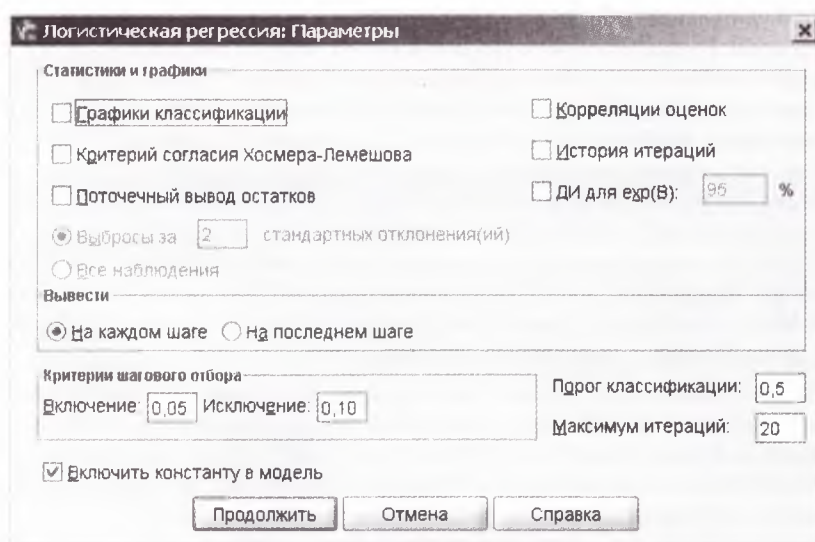


Рис. 24.3. Диалоговое окно Логистическая регрессия: Параметры

Флажок Графики классификации позволяет включить в выводимые результаты диаграмму, на которой для каждого объекта вместе с вероятностью прогноза указана его классификация, полученная при помощи регрессионного уравнения. Таким образом, диаграмма наглядно иллюстрирует, насколько адекватным оказалось составленное уравнение регрессии. Флажок Корреляции оценок указывает на необходимость построения корреляционной матрицы для всех переменных, включенных в регрессионное уравнение. В области Критерии шагового отбора имеются поля Включение и Исключение. В первом указывается значение вероятности, служащее критерием включения предиктора в уравнение регрессии, а во втором — значение вероятности для исключения из уравнения регрессии. По умолчанию в поле Включение указано значение 0,05, а в поле Исключение — значение 0,10. При желании исследователь может изменять их.

Кнопка Сохранить в диалоговом окне Логистическая регрессия позволяет сохранить вычисленные значения вероятностей и предсказанные значения зависимой переменной как новые переменные для наблюдений в файле данных. Зачастую это полезно для сравнения фактической и предсказанной группировок для отдельных объектов. В практи-

ческом плане интерес могут представлять предсказания для тех объектов, для которых фактические значения зависимой переменной неизвестны.

В следующем примере проводится логистический регрессионный анализ, в котором в качестве зависимой переменной выступает переменная помощь, а в качестве предикторов — переменные симпатия, агрессия, польза и условие.

ШАГ 5

После выполнения предыдущего шага у вас должно быть открыто диалоговое окно Логистическая регрессия (см. рис. 24.1). Если вы уже успели поработать с этим окном, щелкните на кнопке Сброс.

1. Щелкните сначала на переменной помощь, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.
2. Щелкните сначала на переменной симпатия, чтобы выделить ее, а затем на нижней кнопке со стрелкой, чтобы переместить переменную в поле Ковариаты.
3. Повторите предыдущее действие для переменных агрессия, польза и условие.
4. В раскрывающемся списке Метод выберите пункт Включение: ОП и щелкните на кнопке Категориальные, чтобы открыть диалоговое окно Логистическая регрессия: Задать категориальные переменные (см. рис. 24.2).
5. Щелкните сначала на переменной условие, чтобы выделить ее, а затем на кнопке со стрелкой, чтобы переместить переменную в список Категориальные ковариаты.
6. В раскрывающемся списке Изменить контраст выберите пункт Отклонения и для подтверждения выбора щелкните на кнопке Изменить. Щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Логистическая регрессия.
7. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Логистическая регрессия: Параметры (см. рис. 24.3).
8. Установите флажок Графики классификации а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Логистическая регрессия.
9. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически откроет окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

Далее приведены результаты, сгенерированные программой при выполнении шага 5.

Поскольку методы Включение: ОП и Исключение: ОП являются пошаговыми, при их использовании выводится несколько таблиц, характеризующих каждый шаг и позволяющих наглядно представить себе процесс формирования уравнения регрессии. Здесь мы ограничились лишь итоговой таблицей.

Кодировки категориальных переменных

Если в вашем анализе использовались категориальные ковариаты и контрасты, то в начале выводимых результатов будет таблица (рис. 24.4), показывающая, каким образом исходные переменные (строки таблицы) были преобразованы в серию новых переменных (столбцы таблицы). Так, к примеру, для переменной условие, имеющей 3 уровня, были сформированы 2 дополнительные переменные (1 и 2), представляющие собой серию контрастов между тремя условиями.

Кодировки категориальных переменных

		Частота	Кодирование параметра	
			(1)	(2)
условия	на работе	12	1,000	,000
	на улице	17	,000	1,000
	в лесу	17	-1,000	-1,000

Рис. 24.4. Кодировки категориальных переменных

Объединенные критерии для коэффициентов модели

Объединенные критерии для коэффициентов модели приведены на рис. 24.5.

Объединенные тесты для коэффициентов модели

		Хи-Квадрат	ст. св.	Знач.
Шаг 1	Шаг	15,920	2	,000
	Блок	15,920	2	,000
	Модель	15,920	2	,000
Шаг 2	Шаг	9,212	1	,002
	Блок	25,132	3	,000
	Модель	25,132	3	,000

Сводка для модели

Шаг	-2 Log Правдоподобие	R квадрат Кокса и Снеппа	R квадрат Нэйджелкерка
1	47,501 ^a	,293	,391
2	38,289 ^b	,421	,563

Рис. 24.5. Объединенные критерии для коэффициентов модели

Трактовка терминов дана в конце этой главы.

Классификационная таблица

В классификационной таблице сравниваются прогнозируемые значения зависимой переменной, рассчитанные по уравнению регрессии, и фактические наблюдаемые значения (рис. 24.6). При определении прогнозируемой величины SPSS вычисляет вероятность для каждого объекта и на основании этой вероятности присваивает объекту одно из двух значений дихотомической переменной. Если вероятность оказалась менее 0,5, помощь оценивается как бесполезная (значение переменной помощь равно 0), в противном случае — как полезная (значение переменной помощь равно 1). Как показывают данные правого столбца таблицы, для 78,3 % объектов результаты прогноза оказались верными.

Наблюдаемое			Предсказанное		Процент корректных
			помощь		
			нет	есть	
Шаг 1	помощь	нет	13	8	61,9
		есть	4	21	84,0
	Общий процент				73,9
Шаг 2	помощь	нет	17	4	91,0
		есть	6	19	76,0
	Общий процент				78,3

a. Разделяющее значение = ,500

Рис. 24.6. Классификационная таблица

Переменные в уравнении

Таблица, приведенная на рис. 24.7, демонстрирует эффекты включения переменных в уравнение на каждом шаге его построения. Строка Константа для каждого шага соответствует константе B_0 регрессионного уравнения.

Переменные в уравнении

		B	Стд. Ошибка	Вальд	ст. св.	Знач.	Exp(B)
Шаг 1 ^a	условия			11,413	2	,003	
	условия(1)	-,279	,497	,314	1	,575	,757
	условия(2)	-1,457	,495	8,659	1	,003	,233
	Константа	,279	,369	,570	1	,450	1,321
Шаг 2 ^b	агрессия	,652	,270	5,827	1	,016	1,920
	условия			9,842	2	,007	
	условия(1)	-,447	,559	,639	1	,424	,639
	условия(2)	-1,658	,611	7,371	1	,007	,190
	Константа	-6,210	2,735	5,156	1	,023	,002

a. Переменная(ые) включенная на шаге 1: условия.

b. Переменная(ые) включенная на шаге 2: агрессия.

Рис. 24.7. Переменные в уравнении регрессии

Трактовка терминов дана в конце этой главы.

Параметры модели при исключении переменных

Все переменные модели проверяются на предмет их возможного исключения, и результаты такой проверки включаются в таблицу (рис. 24.8). Уровень значимости для каждой переменной сравнивается со значимостью 0,1, указанной в поле Исключение диалогового окна Логистическая регрессия: Параметры. Если значимость переменной оказывается больше 0,1, переменная исключается.

Модель при исключении члена

Переменная	Log Правдоподобие Модели	Изменение в -2 Log Правдоподобие	ст. св.	Знач. Изменений
Шаг 1 условия	-31,711	15,920	2	,000
Шаг 2 агрессия	-23,750	9,212	1	,002
условия	-27,048	15,808	2	,000

Рис. 24.8. Результаты проверки переменных на возможность исключения из модели

ной предсказанной вероятности, а его высота — количеству объектов, для которых предсказана данная вероятность. В случае идеальной логистической регрессии все буквы *n* окажутся левее букв *e*, а разделять их будет вероятность 0,5. Как видно из диаграммы, некоторые столбики включают в себя обе буквы, что свидетельствует об ошибках предсказания (высота в два символа соответствует одному объекту). Символам *n* в правой части диаграммы и символам *e* в левой части диаграммы соответствуют неправильные предсказания относительно оказания помощи. О количестве правильных и неправильных предсказаний позволяет судить классификационная таблица (см. рис. 24.6).

Терминология, используемая при выводе

Далее дана трактовка терминов, используемых программой в окне вывода и относящихся к объединенным критериям для коэффициентов модели:

- ▶ Шаг — методу Включение: ОП понадобилось всего 2 шага, чтобы включить все нужные переменные в уравнение регрессии. На самом деле позже в модель вводятся и другие переменные, однако на данном этапе будем считать, что их всего две.
- ▶ Хи-квадрат — хи-квадрат шага, блока или модели. Это критерии статистической значимости воздействия на зависимую переменную всех предикторов заданной модели, блока или шага. На шаге 1 все три критерия χ^2 оказались равными: для модели и шага потому, что на шаге 1 они тождественны, а для блока и модели потому, что модель содержит единственный блок. Большие значения критерия χ^2 говорят о том, что первая включенная переменная оказывает существенное влияние на зависимую переменную. Критерии χ^2 шага 2 свидетельствуют о том, что включение второй переменной в модель значительно улучшает последнюю: обратите внимание на то, что критерий χ^2 модели равен сумме критериев χ^2 шага и модели на предыдущем шаге.
- ▶ -2 Log Правдоподобие — эта величина характеризует модель и показывает, насколько хорошо она соответствует исходным данным. Чем меньше это значение, тем более адекватна сформированная модель.
- ▶ R квадрат Кокса и Снелла, R квадрат Нэйджелкерка — приближения значения R^2 , показывающие долю влияния всех предикторов модели на дисперсию зависимой переменной.

Далее дана трактовка остальных терминов, используемых программой в окне вывода и относящихся к переменным в уравнении регрессии:

- ▶ *B* — коэффициенты регрессионного уравнения, отражающие влияние соответствующих предикторов на зависимую переменную. Так, переменная агрессия оказывает положительное влияние на вероятность оказания помощи.
- ▶ Стд. Ошибка — стандартная ошибка, мера изменчивости коэффициентов *B*.
- ▶ Вальд — критерий значимости Вальда коэффициента *B* для соответствующего предиктора. Чем больше его значение (вместе с числом степеней свободы), тем выше значимость.
- ▶ Знч. — значимость по критерию Вальда.
- ▶ $\text{Exp}(B)$ — величина (e^B), которая может использоваться для интерпретации результатов анализа наравне с коэффициентом *B* (вспомните о двух формах регрессионного уравнения, в одной из которых используются коэффициенты *B*, а в другой — e^B).

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.spv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» главы 2).

Для выхода из программы выберите в меню Файл команду Выход.

25 Логлинейный анализ таблиц сопряженности

329	Понятие логлинейной модели
330	Логлинейный метод подбора модели
331	Пошаговые алгоритмы вычислений
335	Представление результатов
339	Завершение анализа и выход из программы

Логлинейные модели применяются для анализа таблиц сопряженности нескольких категориальных переменных. Если анализируется сопряженность двух таких переменных (двумерная таблица сопряженности), вполне достаточно применения критерия χ^2 (см. главу 8). Однако зачастую данные содержат существенно большее число категориальных переменных, и тогда визуальный анализ таблиц сопряженности становится невозможным. Например, визуально интерпретировать таблицу сопряженности пол $\times$ класс $\times$ вуз $\times$ хобби практически невозможно. В подобных ситуациях на помощь исследователям приходят логлинейные модели, во многом сходные с регрессионными моделями и дисперсионным анализом.

В этой главе для иллюстрации иерархического логлинейного анализа мы используем файл `helpLLM.sav`. Этот файл содержит несколько модифицированных переменных файла `helpLR.sav`: переменная агрессия представлена в бинарном виде (1 — низкая, 2 — высокая), переменная симпатия имеет три категории (1 — слабая, 2 — средняя, 3 — сильная). Переменные помощь (2 градации) и условие (3 градации) оставлены без изменений. Таким образом, все переменные файла `helpLLM.sav` являются категориальными. В этой главе будет проводиться логлинейный анализ таблицы сопряженности агрессия $\times$ симпатия $\times$ условия $\times$ помощь.

Понятие логлинейной модели

Логлинейная модель, в сущности, представляет собой множественную регрессионную модель, в которой категориальные переменные и их взаимодействия выступают в качестве предикторов, а роль зависимой переменной играет натуральный логарифм частот категорий. Именно использование логарифмической меры обуславливает линейность модели. Уравнение взаимодействия переменных агрессия, симпатия и условия будет иметь вид

$$\ln(\text{частота}) = \mu + \lambda^A + \lambda^C + \lambda^Y + \lambda^{A \cdot C} + \lambda^{A \cdot Y} + \lambda^{C \cdot Y} + \lambda^{A \cdot C \cdot Y}.$$

В этом уравнении частота — это частота текущей ячейки частотной таблицы, μ — общее среднее воздействия, эквивалентное константе во множественном регрессионном анализе, а каждое значение λ представляет собой воздействие одной или более независимой переменной. Так, λ^A является воздействием переменной агрессия, λ^Y и λ^C — переменных

условия и симпатия соответственно, $\lambda^{A \times C}$ — взаимодействия переменных агрессия и симпатия, $\lambda^{A \times C \times Y}$ — взаимодействия переменных агрессия, симпатия и условия. Представленная здесь модель называется насыщенной, поскольку содержит все предикторы и их возможные взаимодействия. Однако насыщенная модель обычно не является оптимальной, так как редко все главные эффекты и взаимодействия оказываются значимыми. Как правило, существуют более предпочтительные альтернативы в виде ненасыщенных моделей, которые отражают лишь статистически значимые главные эффекты и взаимодействия переменных.

Подменю Логлинейный анализ содержит три команды:

- ▶ **Общий** — эта команда допускает вхождение в модель любых факторов и их взаимодействий и предполагает, что исследователь перед проведением анализа уже имеет гипотезы о составе модели;
- ▶ **Логит** — применение этой команды позволяет рассматривать дихотомические переменные как зависимые, а одну (или более) категориальную переменную — как независимую. При этом зависимая дихотомическая переменная используется не для прогнозирования частот категорий, а для разделения всех категорий на две группы. Смысл понятия «логит» раскрывается в главе 24;
- ▶ **Подбор модели** — эта команда позволяет из всех возможных ненасыщенных моделей подобрать ту, которая в наибольшей степени соответствует исходным данным. Как правило, подбор осуществляется автоматически. В результате выявляется совокупность значимых связей между категориальными переменными и вычисляются параметры μ и λ логлинейной модели.

В данной главе рассматривается последний из перечисленных методов, который основан на применении иерархической логлинейной модели.

Логлинейный метод подбора модели

Теоретически, из насыщенной модели можно удалить любые элементы, получив таким образом произвольную ненасыщенную модель. Далее можно проверить состоятельность этой модели и в случае несоответствия ее исходным данным перейти к анализу другой ненасыщенной модели. Однако большинство исследователей отдают предпочтение иерархическим логлинейным моделям, которые позволяют упорядочить процесс подбора окончательной состоятельной модели. Основной особенностью иерархических моделей является то, что присутствие какого-либо взаимодействия переменных означает присутствие всех взаимодействий, имеющих более низкий порядок, и главных эффектов этих переменных. Например, если в модели присутствует взаимодействие агрессия $\times$ симпатия, то в ней присутствуют главные эффекты переменных агрессия и симпатия; если в модели присутствует взаимодействие агрессия $\times$ симпатия $\times$ условия, то в ней также присутствуют взаимодействия агрессия $\times$ симпатия, агрессия $\times$ условия, симпатия $\times$ условия и т. д.

Существует три вспомогательных метода, которые предназначены для подбора адекватной модели. Все они оказываются весьма полезными и приводят к сходным результатам. Тем не менее окончательный выбор модели определяется не только статистическими результатами применения метода, но и пониманием исследователем особенностей исходных данных. Краткие описания методов приведены далее, более подробные — в разделе «Представление результатов»:

- Метод *исследования оценок параметров* предназначен для вычисления оценок параметров для насыщенной модели. Помимо обычных оценок параметров SPSS вычисляет стандартизованные оценки. Если значения последних невелики, то они не оказывают значимого влияния на модель и обычно исключаются.
- Методом *вычисления частичного критерия хи-квадрат* в дополнение к оценкам параметров модели SPSS вычисляется критерий χ^2 , характеризующий степень соответствия модели исходным данным. При помощи этого критерия проверяют, являются ли все однофакторные эффекты, а также эффекты более высоких порядков статистически значимыми (анализируются комбинации всех эффектов первого, второго и более высоких порядков). При этом отсутствие общей значимости эффектов второго порядка вовсе не означает, что все эффекты первого порядка не являются значимыми. Аналогично, из отсутствия общей значимости эффектов любого порядка не вытекает отсутствие значимости отдельных взаимодействий этого порядка. Вследствие этих двух особенностей в SPSS предусмотрена возможность раздельной проверки главных эффектов и эффектов взаимодействий.
- Суть метода *пошагового исключения* состоит в автоматической подгонке модели, этим он сходен с методом исключения предикторов из уравнения регрессии: из насыщенной модели постепенно исключаются те элементы (переменные и их взаимодействия), которые не оказывают значимого воздействия. Данный метод построения модели относится к иерархическому логлинейному моделированию. Так, если обнаружено статистически значимое взаимодействие четырех переменных, не проверяется (на предмет исключения из модели) взаимодействие трех из этих переменных, иначе модель не являлась бы иерархической по определению. Окончательный результат подгонки модели наиболее приемлем, если все оставшиеся в ней элементы оказываются статистически достоверными.

Иерархические логлинейные модели относятся к очень сложному разделу статистики, поэтому мы рекомендуем перед проведением логлинейного анализа обратиться к дополнительным источникам.

Пошаговые алгоритмы вычислений

До применения логлинейного анализа для подбора модели нужно выполнить три подготовительных шага. Эти шаги позволят подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 20 и открыть файл (в данном случае helpLLM.sav). Пошаговые инструкции для этого процесса приведены в главе 4 (с. 55), а подробные разъяснения — в главе 2.

После завершения шага 3 на экране должно присутствовать окно редактора данных со строкой меню и загруженным файлом helpLLM.sav.

ШАГ 4

В меню Анализ выберите команду Логлинейный ► Подбор модели. На экране появится диалоговое окно Логлинейный анализ: Подбор модели (рис. 25.1).

В этом диалоговом окне задаются переменные для анализа, а также выбирается метод построения модели. Слева находится список всех переменных файла данных. Для включения переменной в анализ достаточно выделить ее и переместить в поле Факторы, щелкнув на верхней кнопке со стрелкой. Данный метод применим только к категориальным переменным, диапазон каждой из которых задается кнопкой Задать диапазон.

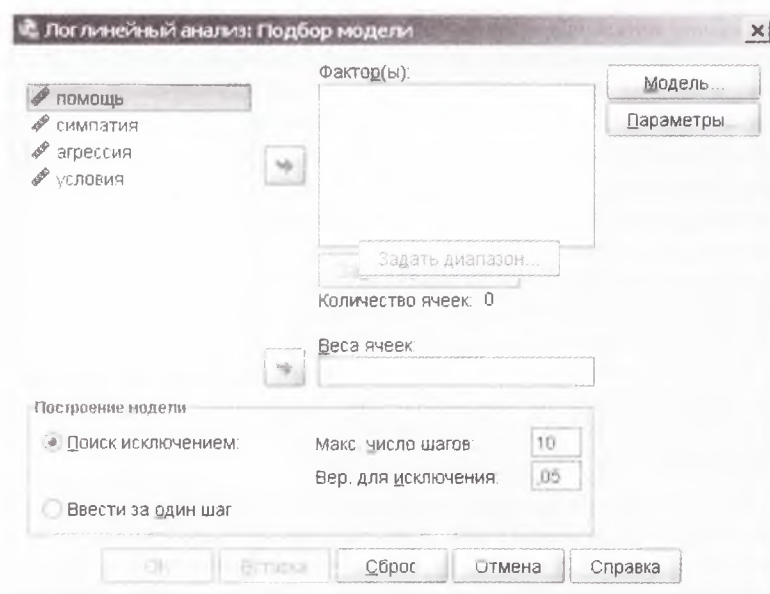


Рис. 25.1. Диалоговое окно Логлинейный анализ: Подбор модели

Эта кнопка открывает окно Логлинейный анализ: Задать диапазон (рис. 25.2). В нем необходимо задать минимальный и максимальный уровни каждой переменной в полях Минимум и Максимум.

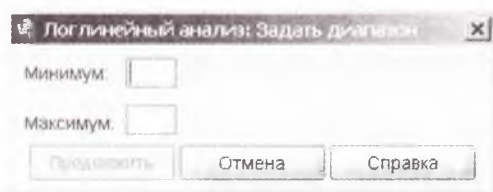


Рис. 25.2. Диалоговое окно Логлинейный анализ: Задать диапазон

Поле Веса ячеек диалогового окна Логлинейный анализ: Подбор модели используется в том случае, если ваша модель содержит структурные нули. Понятие структурных нулей в этой книге не рассматривается; при необходимости обратитесь к руководству пользователя SPSS.

Диалоговые окна задания переменных для не рассматриваемых в этой книге команд Логлинейный ► Общий и Логлинейный ► Логит отличаются тем, что позволяют задавать в качестве факторов не только категориальные, но и количественные переменные (ковариаты). Помимо этого при использовании логит-моделей необходимо задать одну или несколько бинарных зависимых переменных, каждая из которых используется в анализе для деления независимых переменных на две категории.

Рассмотрим переключатель Построение модели, расположенный в нижней части диалогового окна Логлинейный анализ: Подбор модели. Установка его в положение Поиск исключением определяет построение модели методом пошагового исключения: из насыщенной модели в результате серии шагов исключаются все элементы, не оказывающие

значимого влияния. Положение переключателя Ввести за один шаг позволяет получить модель со всеми элементами. Если используется метод пошагового исключения, в поле Вер. для исключения нужно указать величину значимости, которая будет служить критерием исключения элементов из модели. По умолчанию это значение $p = 0,05$, и нередко оно устраивает исследователей. Под *шагом* понимается удаление очередного элемента из модели: в поле Макс. число шагов вы можете задать максимальное количество шагов, для того чтобы процедура принудительно прекратила процесс исключения предикторов. По умолчанию максимальное число шагов равно 10.

После того как метод построения модели и переменные, входящие в ее состав, заданы, можно переходить к определению других параметров команды, используя кнопки Параметры и Модель.

При щелчке на кнопке Параметры открывается диалоговое окно Логлинейный анализ: Параметры (рис. 25.3). Оно позволяет настроить параметры генерируемых командой итоговых данных, а также задать некоторые технические аспекты ее выполнения (последние в этой книге не рассматриваются). Установка флажка Частоты включает в выводимые данные таблицу частот модели, а флажка Остатки — разности между фактическими и прогнозируемыми частотами, рассчитываемыми при помощи модели.

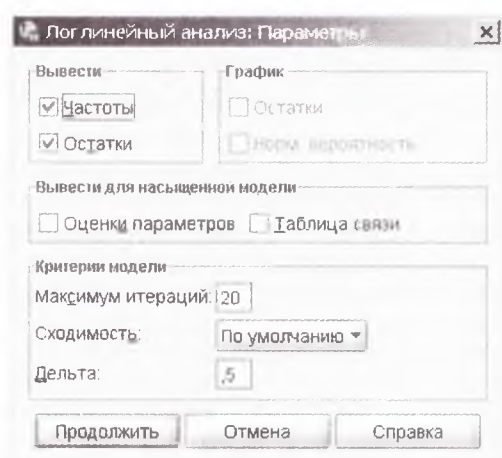


Рис. 25.3. Диалоговое окно Логлинейный анализ: Параметры

Нередко используются флажки Оценки параметров и Таблица связи. Для насыщенной модели эти флажки позволяют включить в выводимые данные оценки параметров, что может быть полезно при выборе модели, и таблицу ассоциаций, содержащую критерии χ^2 для всех главных эффектов и взаимодействий. Оба флажка при исследовании насыщенной модели рекомендуется оставлять установленными. Если же для анализа необходима ненасыщенная модель, включающая в себя лишь некоторые взаимодействия и факторы, то, щелкнув на кнопке Модель, следует открыть диалоговое окно Логлинейный анализ: Модель (рис. 25.4).

Если переключатель Задать модель установить в положение Настраиваемая, то можно выбирать переменные в списке Факторы. Раскрывающийся список в центре окна позволяет выбрать один из типов их взаимодействий: Главные эффекты, Взаимодействие, Все 2-факторные, Все 3-факторные, Все 4-факторные и Все 5-факторные. Поскольку команда Логлинейный ► Подбор модели строит иерархические модели, при выборе Взаимодействие все взаимодействия более низких порядков включаются в модель автоматически.

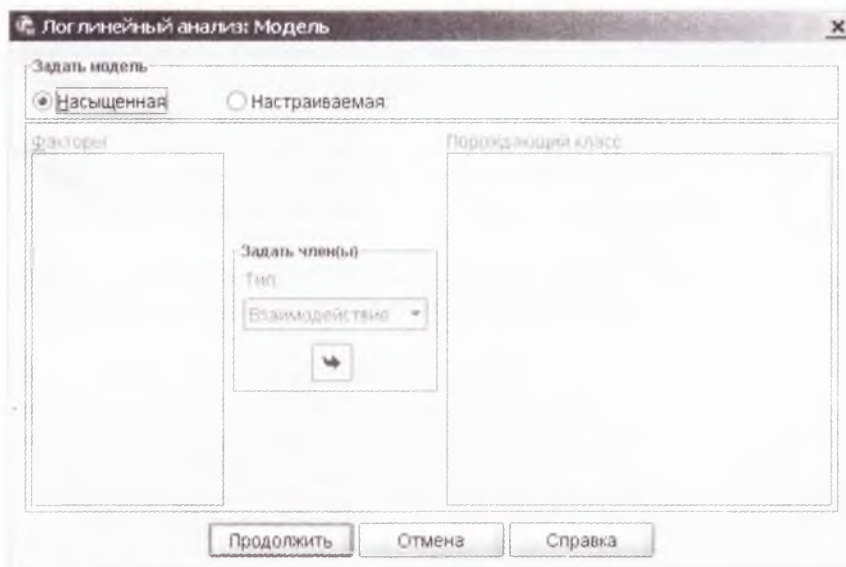


Рис. 25.4. Диалоговое окно Логлинейный анализ: Модель

Задание модели вручную требуется в тех случаях, когда есть предварительные гипотезы о составе ненасыщенной модели. Иногда модель задается для того, чтобы ограничить количество рассматриваемых взаимодействий, если взаимодействия высоких порядков вызывают трудности в интерпретации. Для настройки модели вручную следует установить переключатель *Настраиваемая*, тогда становится доступным включение в модель необходимых главных факторов и взаимодействий. Главные факторы включаются в модель следующим образом: имя соответствующего фактора выделяется в списке *Факторы* и с помощью кнопки с направленной вправо стрелкой перемещается в правый список. Если требуется включить в модель взаимодействие факторов, следует щелчками выделить в списке *Факторы* имена всех факторов, участвующих во взаимодействии, затем в раскрывающемся списке *Задать член(ы)*: *Тип* выбрать пункт *Взаимодействие* и, наконец, как и в предыдущем случае, щелкнуть на кнопке с направленной вправо стрелкой. Чтобы построить сразу несколько взаимодействий определенного порядка для выбранных факторов, нужно выполнить те же действия, что и для одиночного взаимодействия, но в раскрывающемся списке выбрать пункт в зависимости от желаемого результата.

В следующем примере проводится иерархический логлинейный анализ переменных агрессия, симпатия, условия и помощь, имеющих, соответственно, 2, 3, 3 и 2 уровня. Мы будем исследовать насыщенную модель, а в выводимые данные включим оценки параметров и критерии χ^2 для отдельных эффектов. Затем мы применим метод пошагового исключения, чтобы удалить из насыщенной модели те элементы, которые не оказывают значимого воздействия.

ШАГ 5

После выполнения предыдущего шага у вас должно быть открыто диалоговое окно *Логлинейный анализ: Подбор модели* (см. рис. 25.1). Если вы уже успели поработать с этим окном, щелкните на кнопке *Сброс*.

1. Щелкните сначала на переменной *помощь*, чтобы выделить ее, а затем на верхней кнопке со стрелкой, чтобы переместить переменную в список *Факторы*.



Щелкните на кнопке Задать диапазон, чтобы открыть диалоговое окно Логлинейный анализ: Задать диапазон (см. рис. 25.2).

2. В поле Минимум введите число 0, нажмите клавишу Tab, чтобы переместить фокус ввода в поле Максимум, введите число 1 и щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Логлинейный анализ: Подбор модели.
3. Повторите первые три действия для переменных симпатия, агрессия и условия, используя для задания значений переменных симпатия и условия числа 1 и 3, а для задания значений переменной агрессия числа 1 и 2.
4. Щелкните на кнопке Параметры, чтобы открыть диалоговое окно Логлинейный анализ: Параметры (см. рис. 25.3).
5. Установите флажки Оценки параметров и Таблица связи, а затем щелкните на кнопке Продолжить, чтобы вернуться в диалоговое окно Логлинейный анализ: Подбор модели.
6. Щелкните на кнопке ОК, чтобы открыть окно вывода.

После выполнения шага 5 программа автоматически откроет окно вывода. Для просмотра результатов при необходимости можно воспользоваться вертикальной и горизонтальной полосами прокрутки. Обратите внимание на стандартную строку меню в верхней части окна вывода: ее присутствие позволяет выполнять любые статистические операции, не переключаясь обратно в окно редактора данных.

Представление результатов

В этом разделе приводятся фрагменты данных, генерируемые программой при выполнении шага 5. Напоминаем, что, если выводимые данные занимают слишком много места, они представлены в окне вывода лишь частично. В этом случае в том месте окна вывода, в котором скрыты дополнительные данные, присутствует значок в виде красного треугольника. Чтобы увидеть такие результаты полностью, дважды щелкните на имеющемся фрагменте.

Проверка на равенство нулю взаимодействий порядка K и более высокого порядка

В таблице, приведенной на рис. 25.5, представлены результаты проверки эффекта от взаимодействий указанного и более высокого порядка и отдельно — только указанного порядка. Для проверки используются два варианта критерия χ^2 . Оба критерия оценивают, являются ли ожидаемые частоты в ячейках для соответствующей модели значимо отличающимися от наблюдаемых частот или нет. Если отличие значимо, гипотеза об отсутствии взаимодействия отвергается и делается вывод о связи соответствующих переменных.

Рассмотрим первую часть таблицы Эффекты K-го порядка и более высоких порядков. Последняя строка этой части таблицы соответствует четвертому порядку взаимодействия (значение в столбце K равно 4) и характеризует общий (суммарный) эффект от взаимодействий четырех и более переменных. Большой уровень значимости и низкое значение χ^2 говорят о том, что взаимодействие всех четырех переменных не оказывает существенного влияния на модель. То же касается взаимодействия третьего и более высоких порядков. Противоположный вывод можно сделать об остальных взаимодействиях: χ^2

возрастает, а значимость убывает с убыванием порядка. Первые две строки свидетельствуют о том, что среди всех взаимодействий низкого порядка имеются такие, которые оказывают на модель значимое влияние.

Эффекты К-го порядка и более высоких порядков

	К	Ст. св.	Отношение правдоподобия		Пирсона		Число итераций
			Хи-квадрат	Знач.	Хи-квадрат	Знач.	
Эффекты К-го порядка и более высоких порядков	1	35	58,726	,007	52,609	,028	0
	2	29	54,241	,003	47,940	,015	2
	3	16	21,784	,150	17,649	,345	4
	4	4	,397	,983	,212	,995	4
Эффекты К-го порядка ^б	1	6	4,486	,611	4,669	,587	0
	2	13	32,456	,002	30,291	,004	0
	3	12	21,387	,045	17,437	,134	0
	4	4	,397	,983	,212	,995	0

Число степеней свободы, использованное в данных критериях НЕ было скорректировано на наличие структурных или выборочных нулей. Критерии, основанные на данном числе степеней свободы, могут быть консервативными

а Результаты тестов эффектов К-го и более высоких порядков равны нулю

б Результаты тестов эффектов К-го порядка равны нулю

Рис. 25.5. Проверка эффекта от взаимодействий разных порядков

Вторая часть таблицы на рис. 25.5 Эффекты К-го порядка проверяет на равенство нулю только эффекты заданного порядка. Соответственно, здесь характеризуются взаимодействия только указанного порядка. Как можно видеть в столбце Отношение правдоподобия, значимое воздействие на модель оказывают взаимодействия второго и третьего порядков.

Далее дана трактовка терминов, используемых программой в окне вывода:

- К — порядок взаимодействий. В разделе для порядка К и выше: 4 — порядок 4 и выше, 3 — порядок 3 и выше и т. д. В разделе только для порядка К: 1 — порядок 1, 2 — порядок 2 и т. д.
- Ст. св. — число степеней свободы для соответствующих взаимодействий.
- Отношение правдоподобия, Хи-квадрат — критерий χ^2 для определения вероятности того, что общий эффект от соответствующих взаимодействий равен нулю (по замыслу его авторов, более «правдоподобный», чем критерий хи-квадрат Пирсона).
- Пирсона Хи-квадрат — критерий χ^2 для определения вероятности того, что общий эффект от взаимодействий равен нулю.
- Знач. — значимость, вероятность того, что общий эффект от соответствующих взаимодействий равен нулю. Чем меньше вероятность, тем большее влияние эффект оказывает на модель.
- Число итераций — число итераций для вычисления значений χ^2 .

Проверка частных взаимосвязей

На рис. 25.6 приведены критерии χ^2 для каждого из воздействий насыщенной модели. Воздействия с высоким значением χ^2 оказывают значимое влияние на модель, чем ниже значение χ^2 , тем ниже значимость. В данном примере главные эффекты переменных не являются статистически значимыми, а к значимым можно отнести двухфакторные взаимодействия помощь × условие и помощь × симпатия, а также взаимодействие трех переменных помощь × симпатия × агрессия. Поскольку эти частные критерии χ^2 не обяза-

тельно являются независимыми, их доля в итоговом значении χ^2 для насыщенной модели может отличаться от их доли в итоговом значении в ненасыщенной модели.

Частные взаимосвязи

Эффект	Ст. св.	Частный Хи-квадрат	Знач.	Число итераций
помощь*симпатия*агрессия	2	6,288	,043	3
помощь*симпатия*условия	4	,091	,999	5
помощь*агрессия*условия	2	,545	,762	5
симпатия*агрессия*условия	4	2,441	,655	6
помощь*симпатия	2	7,430	,024	3
помощь*агрессия	1	1,956	,162	4
симпатия*агрессия	2	1,944	,378	4
помощь*условия	2	16,593	,000	3
симпатия*условия	4	2,277	,685	3
агрессия*условия	2	1,230	,541	4
помощь	1	,348	,555	2
симпатия	2	1,605	,448	2
агрессия	1	1,398	,237	2
условия	2	1,134	,567	2

Рис. 25.6. Проверка частных взаимосвязей

Далее дана трактовка терминов, используемых программой в окне вывода:

- Ст. св. — число степеней свободы соответствующего эффекта.
- Частный Хи-квадрат — критерий χ^2 для отдельного эффекта.
- Знач. — статистическая значимость, вероятность того, что величина эффекта равна нулю. Чем меньше вероятность, тем большее влияние эффект оказывает на модель.
- Число итераций — число итераций для вычисления соответствующей связи.

Оценки параметров модели

Фрагмент результатов, представленный на рис. 25.7, содержит оценки параметров, представляющие собой значения λ из логлинейного уравнения для трехфакторного взаимодействия помощь $\times$ симпатия $\times$ агрессия и двухфакторного взаимодействия помощь $\times$ условия. По уровню статистической значимости первое из указанных взаимодействий приближается к статистически достоверному ($p < 0,1$), а второе статистически достоверно ($p < 0,05$). Поскольку сумма всех параметров должна быть равна нулю, на основе полученных данных можно вычислить другие. Так, в нижней строке и правом столбце табл. 25.1 показаны значения λ , вычисленные на основе представленных ранее данных для взаимодействия помощь $\times$ условия исходя из того, что сумма значений в строках и столбцах должна быть равна 0.

Оценки параметров

Эффект	Параметр	Оценка	Станд. ошибка	Z	Знач.	95% доверительный интервал	
						Нижняя граница	Верхняя граница
помощь*симпатия*агрессия	1	-,286	,258	-1,108	,268	-,791	,220
	2	,443	,236	1,881	,060	-,019	,905
помощь*условия	1	,037	,246	,150	,881	-,446	,520
	2	,483	,239	2,026	,043	,016	,951

Рис. 25.7. Оценки параметров модели

Таблица 25.1. Пример кодирования взаимодействия

Уровни переменной «помощь»	Уровни переменной «симпатия»		
	1	2	3
0	0,0369	0,4833	-0,4798
1	-0,0369	-0,4833	0,4798

Полученные оценки можно использовать для интерпретации распределения частот. Например, из табл. 25.1 (нижняя строка) видно, что при наибольшем уровне симпатии (3) помощь оказывают чаще. Но лучше для этих целей использовать таблицы сопряженности, воспользовавшись командой **Описательные статистики** ► **Таблицы сопряженности**.

Далее дана трактовка терминов, используемых программой в окне вывода и относящихся к оценкам параметров модели:

- Параметр — номер параметра. Нумерация параметров производится от наименьших значений независимых переменных к наибольшим (в качестве примера кодирования взаимодействий (см. табл. 25.1)).
- Оценка — оценка коэффициента λ в логлинейном уравнении.
- Станд. ошибка — стандартная ошибка, мера изменчивости оценки коэффициента.
- Z — z-значение коэффициента. Значения больше 1,96 (меньше -1,96) статистически достоверны (имеют уровень значимости меньше 0,05).
- Знач. — статистическая значимость оценки.
- 95 % доверительный интервал — границы интервала, в который фактическое (не приближенное) значение коэффициента попадает с вероятностью 95 %.

Обратное пошаговое исключение

Результаты обратного пошагового исключения являются, пожалуй, наиболее важными среди остальных, поскольку дают представление о наиболее состоятельной модели (рис. 25.8). В представленном фрагменте результатов содержатся критерии χ^2 , их значимости и все взаимодействия, участвующие в процессе пошагового исключения на шагах 5–7 (шаги 1–4 для экономии места опущены). Процесс пошагового исключения начинается с насыщенной модели. Сначала SPSS вычисляет значение χ^2 для модели, включающей в себя взаимодействия четвертого и более низких порядков (включая главные эффекты), это значение равно 0, поскольку изначально модель является насыщенной. Далее устанавливается, что при исключении взаимодействия четвертого порядка изменение величины χ^2 незначимо, и исключение производится. Затем анализируется модель, состоящая из взаимодействия третьего и более низких порядков, из модели исключаются незначимые взаимодействия. Весь процесс занимает 7 шагов, и в результате формируется иерархическая модель, в которую входят взаимодействия **помощь** × **симпатия** × **агрессия** и **помощь** × **условия**. Оба взаимодействия статистически достоверны. Значение критерия χ^2 для всей модели составляет 13,268, число степеней свободы равно 20. Следует отметить, что критерий χ^2 для всей модели оценивает расхождение между наблюдаемыми частотами в исходной таблице сопряженности и ожидаемыми частотами, вычисленными при помощи модели.

Пошаговый отчет						
Шаг ^b	Эффекты		Хи-квадрат ^a	Ст. св.	Знач.	Число итераций
5	Генерируемые классы ^c		помощь*симпатия*агрессия, помощь*условия, агрессия*условия	12,122	18	,841
	Эффект удаления	1	помощь*симпатия*агрессия	11,939	2	,003
		2	помощь*условия	16,440	2	,000
		3	агрессия*условия	1,146	2	,564
6	Генерируемые классы ^c		помощь*симпатия*агрессия, помощь*условия	13,268	20	,866
	Эффект удаления	1	помощь*симпатия*агрессия	11,944	2	,003
		2	помощь*условия	15,920	2	,000
7	Генерируемые классы ^c		помощь*симпатия*агрессия, помощь*условия	13,268	20	,866

a. 'Эффект удаления' - это изменение Хи-квадрат в результате удаления эффекта из модели

b. На каждом шаге удаляется эффект с наибольшим уровнем значимости отношения правдоподобия, при условии, что уровень значимости превышает ,050

c. Выводятся статистики наилучшей модели на каждом шаге после нулевого

Рис. 25.8. Некоторые результаты обратного пошагового исключения

Поскольку результат оказался незначимым, созданная модель адекватна исходным данным: ожидаемые в соответствии с моделью частоты вполне точно соответствуют наблюдаемым частотам. Для содержательной интерпретации модели следует обратиться к значимым компонентам модели, в данном случае к таблицам сопряженности помощь × симпатия × агрессия и помощь × условия.

Завершение анализа и выход из программы

Отредактируйте содержимое окна вывода в соответствии со своими предпочтениями: скройте лишнюю информацию, исправьте таблицы и пр. (см. раздел «Окно вывода и его редактирование» главы 2). За инструкциями по редактированию графиков обратитесь к главе 5.

Для дальнейшего использования результата все содержимое окна вывода или его фрагменты можно сохранить в файле *.srv, экспортировать в другой формат (например, Word), перенести в документ Word или вывести на печать (подробности см. в разделе «Сохранение, экспорт, перенос и печать результатов» в главе 2).

Для выхода из программы выберите в меню Файл команду Выход.

26 AMOS: Моделирование структурными уравнениями

341	SEM и другие методы
343	Особенности и этапы SEM
345	Формирование модели и ее идентификация
348	Оценка, проверка согласованности и коррекция модели
353	Требования к данным
356	Пошаговые алгоритмы вычислений ▶ Конфирматорный факторный анализ ▶ Структурная модель
379	Представление результатов ▶ Конфирматорный факторный анализ ▶ Структурная модель
384	Терминология, используемая при выводе
386	Завершение анализа и выход из программы

AMOS (Analysis of Moment Structures — анализ структур моментов), с одной стороны, является самостоятельной программой, которая может работать с данными других форматов, например Excel, с другой — при установке интегрируется в оболочку предварительно установленной программы IBM SPSS как один из ее модулей. При этом в меню Анализ появляется пункт IBM SPSS AMOS. Этот модуль реализует особую методологию анализа данных, которая называется SEM (Structural equation modeling — моделирование структурными уравнениями), также известную как структурное моделирование, анализ ковариационных структур, или каузальное моделирование. Этот подход основан на множестве хорошо известных методов, таких как корреляция, множественная регрессия, факторный анализ, общие линейные модели (дисперсионный и ковариационный анализ). Однако, являясь комбинацией и развитием перечисленных методов, SEM по своим возможностям анализа данных далеко выходит за их пределы. Еще одно принципиальное отличие SEM от предшествующих методов заключается в том, что его методология в большей степени основана на дедуктивной логике (конфирматорной), нежели на индуктивной (эксплораторной): она предполагает предварительное построение структурной модели направленных и ненаправленных связей между изучаемыми конструктами с целью ее дальнейшей проверки на соответствие эмпирическим данным и коррекции средствами SEM.

AMOS разработан таким образом, чтобы в максимальной степени облегчить применение SEM — метода, который часто воспринимается как чрезмерно сложный и недоступный для постижения. Исследователь, обладающий лишь поверхностной статистической

подготовкой, может создавать, модифицировать и просматривать модели SEM в виде наглядных диаграмм, используя исключительно графические инструменты AMOS. Сочетая в себе простой графический интерфейс с самыми современными компьютерными средствами SEM, AMOS позволяет получать результат в виде наглядной блок-схемы, полностью готовой к вставке в статью или диссертацию.

Следует, однако, предупредить начинающего пользователя AMOS, воодушевленного первыми опытами применения SEM. Несмотря на легкость и простоту создания моделей в AMOS, SEM — чрезвычайно сложный инструмент исследователя, и его *корректное* применение требует знания множества технических деталей, а главное — статистических и методологических основ, изложение которых выходит за рамки этой книги. Эта глава поможет получить первый опыт применения SEM в среде AMOS, но для дальнейшей работы с ним следует обратиться к дополнительной литературе, список которой приведен в конце книги. Наиболее известной является книга Р. Клайна [5], которая была одним из источников при написании главы.

SEM и другие методы

Структурное моделирование, точнее, моделирование структурными уравнениями — дедуктивный метод проверки априорных гипотез на их соответствие эмпирическим данным. Гипотезы формулируются в терминах направленных (причинно-следственных) и (или) ненаправленных (корреляционных) связей между переменными: явными (измеренными) и латентными (гипотетическими конструктами). Переменные и связи между ними графически изображаются в виде блок-схемы, которая при помощи компьютерной программы AMOS трансформируется в систему линейных регрессионных уравнений (допустима и обратная трансформация). Программное обеспечение позволяет оценить степень соответствия модели исходным данным и параметры модели, такие, например, как регрессионные и корреляционные коэффициенты. Модель признается состоятельной (подтверждающей исходные гипотезы), если она по принятым критериям соответствует исходным данным, а параметры модели являются статистически достоверными.

SEM находит все более широкое применение в зарубежной психологии: к 2003 году количество ссылок на его применение в APA PsycInfo сравнялось с частотой упоминания наиболее популярного до той поры статистического метода ANOVA (MANOVA); только в период с 1997 до 2004 года число публикаций, подготовленных с использованием этого метода, возросло почти в 2 раза. SEM в зарубежной литературе является общим понятием для тех более частных моделей, которые включают в себя направленные связи между переменными, непосредственно не измеримые конструкты (например, латентные переменные) и потенциальные ошибки измерения. Активно обсуждаются возможности применения SEM для проверки каузальных гипотез.

Особенности и преимущества SEM можно понять только в сопоставлении его с более традиционными статистическими методами, такими как корреляционный, регрессионный, факторный и дисперсионный анализ. Поэтому для продолжения знакомства с SEM рекомендуется обратиться к соответствующим главам этой книги.

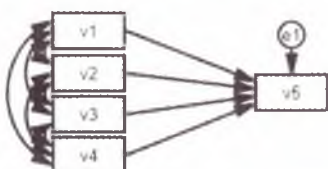
В табл. 26.1 изображены все графические элементы для любой модели SEM. На рис. 26.1 при помощи этих простых и немногочисленных элементов изображены структурные модели широко известных статистических методов анализа взаимосвязи пяти явных переменных (измерений). Рассмотрим специфику каждого из этих методов.

Таблица 26.1. Основные графические элементы SEM

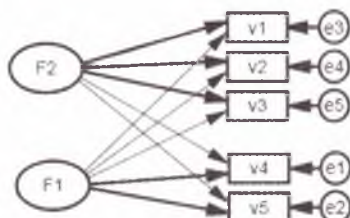
Изображение	Назначение
	Явная переменная (представленная в файле исходных данных), прямоугольный контур
	Латентная переменная (фактор), округлый контур
	Латентная переменная (ошибка), округлый контур
	Направленная (причинно-следственная) связь от независимой к зависимой переменной, односторонняя стрелка
	Ковариация, корреляция (ненаправленная связь), двусторонняя стрелка



Корреляционный анализ. Двусторонние стрелки — корреляции пар переменных



Множественная регрессия. Зависимая переменная (ЗП) — v_5 , независимые переменные (НП) — $v_1, \dots, v_4$. Односторонние стрелки — коэффициенты регрессии, оцениваемые с учетом взаимосвязи НП (двусторонние стрелки). Ошибка модели (e_1) определяется как остаточная доля дисперсии ЗП за вычетом влияния НП ($1 - R^2$)



Факторный анализ: взаимосвязь пяти переменных объясняется влиянием двух ортогональных (не коррелирующих) факторов (F_1, F_2). Выделенные стрелки — доминирующие факторные нагрузки (ФН), тонкие — малые ФН. $e_1, \dots, e_5$ — ошибки, связанные с ошибками измерения и спецификой каждой переменной. Дисперсия ошибки = 1 — общность, где общность — часть дисперсии переменной, объясняемая влиянием выделенных факторов

Рис. 26.1. Структурные модели традиционных методов анализа взаимосвязи пяти переменных

Корреляционный анализ множества переменных (см. главу 9) сводится к вычислению корреляционной матрицы размерностью $P \times P$ (P — количество переменных). Всего вычисляется $P(P - 1)/2$ корреляций, для каждой из которых определяется p -уровень значимости. Основные недостатки:

- ▶ множество статистических проверок требуют коррекции уровня значимости;
- ▶ невозможно судить о том, как на корреляцию данной пары переменных влияют остальные переменные.

В силу этих обстоятельств при большом P (около 10 и более) любые попытки содержательной интерпретации отдельно взятой корреляции весьма сомнительны.

Множественная регрессия (см. главу 18) корректно решает задачу исследования влияния данного набора НП на ЗП. Основными результатами являются:

- ▶ коэффициент множественной корреляции R и квадрат множественной корреляции (КМК) R^2 как показатель доли дисперсии ЗП, которая объясняется влиянием набора НП;
- ▶ регрессионные коэффициенты — стандартизованные (*бета*), отражающие относительный вклад каждой НП в дисперсию ЗП с учетом взаимосвязей между НП, и нестандартизованные, позволяющие оценить значения ЗП.

Основные недостатки метода:

- ▶ в модель можно включить только одну ЗП,
- ▶ невозможно учесть опосредованные связи (так называемые медиаторы и модуляторы).

Эти недостатки с успехом преодолевает так называемый анализ путей в рамках SEM.

Факторный анализ (эксплораторный, исследовательский — традиционный, см. главу 20) позволяет объяснить взаимосвязи между переменными влиянием факторов. Каждый фактор интерпретируется как причина совместной изменчивости группы переменных, имеющих по этому фактору максимальные по модулю факторные нагрузки. Соответственно, основным результатом факторного анализа являются факторные нагрузки переменных (их корреляции с факторами). Дополнительно для каждой переменной вычисляется общность как доля дисперсии переменной, объясняемая влиянием всех выделенных факторов. Традиционно факторная модель предполагает, что факторы ортогональны (не коррелируют). Основные недостатки факторного анализа — ряд неопределенностей:

- ▶ Достаточно ли велика факторная нагрузка переменной, чтобы причислить ее к данному фактору?
- ▶ Достаточно ли данного набора переменных с такими факторными нагрузками, чтобы судить о наличии фактора?
- ▶ Как изменится факторная структура, если учесть корреляцию между факторами?

Эти неопределенности снимаются применением так называемого *конфирматорного факторного анализа* в рамках SEM.

Особенности и этапы SEM

Объекты анализа

SEM, если судить по названию, представляет собой просто расширение множественной регрессии (МРА), только вместо одного линейного уравнения решается система из множества таких уравнений. Это близко к истине, но не совсем так. SEM относится к целому семейству взаимосвязанных процедур. Другие названия этого метода — анализ ковариационных структур, моделирование ковариационных структур — акцентируют внимание на том, что вместо корреляционной матрицы, как в множественной регрессии, исходной информацией для анализа является ковариационная матрица, каждый элемент которой является ковариацией двух переменных:

$$\text{COV}_{xy} = r_{xy} \sigma_x \sigma_y,$$

где r_{xy} — корреляция переменных x и y , σ_x , σ_y — стандартные отклонения этих переменных. Отметим, что на диагонали ковариационной матрицы располагаются дисперсии переменных как ковариации каждой переменной с самой собой. Таким образом, в SEM анализируются и оцениваются не только связи переменных, но и их дисперсии (изменчивость). Более того, в анализе принимают участие и средние значения переменных, что роднит SEM с вариантами дисперсионного анализа (ANOVA), а наряду с ковариационной структурой в модели представлена и структура средних, например оценки средних значений факторов (латентных переменных).

Модели

Вспомним, что специфика метода заключается в априорности структурных моделей, преобладании дедуктивного подхода (конфирматорного, подтверждающего). Применение SEM начинается с построения модели в соответствии с исходными гипотезами исследования, часто еще до сбора эмпирических данных. Затем модель проверяется на соответствие исходным данным. Одновременно производится оценка параметров моделей: регрессионных коэффициентов, ковариаций, дисперсий — с оценкой их статистической значимости. Однако обычно исходная модель не соответствует эмпирическим данным, что требует коррекции модели и, соответственно, гипотез, на которых она основана. В этом смысле выделяют 3 подхода — в зависимости от соотношения дедуктивной и индуктивной логики при использовании SEM [5]. Исключительно *конфирматорный подход* сводится к принятию или отклонению гипотез, на которых основана модель, по результатам ее проверки на соответствие исходным данным. В этом случае сначала формируется модель, затем разрабатываются измерения переменных модели, формируется выборка, а проверка модели является заключительным этапом исследования. Например, при разработке многошкального теста одна из шкал (латентных переменных) в ходе конфирматорного факторного анализа (разновидность SEM) может оказаться несостоятельной и потребует ее исключения. *Проверка альтернативных моделей* выполняется, когда возможна проверка взаимоисключающих моделей с использованием одних и тех же данных. Более правдоподобной признается та модель, которая лучше соответствует исходным данным. *Генерация модели* применяется, когда отсутствуют отчетливые априорные гипотезы о структуре взаимосвязей и исходная модель формируется на основе предварительных результатов применения к исходным данным более традиционных методов (корреляции, множественная регрессия, факторный анализ и пр.). Затем модель многократно проверяется при тех же исходных данных с пошаговым уточнением, пока не будут получены три ее свойства: теоретическая обоснованность, лаконичность, адекватность исходным данным. Таким образом, при генерации модели преобладает индуктивная логика, и этот подход к применению SEM можно отнести к *data mining* (обнаружение скрытых, неочевидных закономерностей в данных).

Этапы SEM

Большинство авторов (В. Вугне [3], R. Kline [5] и др.) выделяют 5 шагов, или этапов, применения SEM:

1. Формирование модели. Как правило, модель изображает графически априорные представления исследователя о структуре направленных и ненаправленных связей измеряемых переменных и латентных конструкторов. При этом решают, какие *параметры модели* (связи, дисперсии) оставлять свободными (будут оцениваться), а какие жестко зафиксировать (как 0 или 1).

2. Идентификация модели. На этом этапе определяется соотношение между тем, что именно будет оцениваться (свободные параметры), и исходной информацией, содержащейся в данных (ковариации, дисперсии). Понятно, что исходной информации должно быть достаточно, а свободных параметров — не слишком много.
3. Оценка модели. На этом шаге выбирается метод оценки параметров модели (преимущественно в зависимости от характера исходных данных).
4. Проверка согласованности модели. Модель проверяется на соответствие исходным данным с использованием многочисленных индексов соответствия, которые оценивают величину расхождения между исходными данными и тем, что предсказывает модель.
5. Коррекция модели. Изменение модели с целью достижения ее соответствия исходным данным производится одним из двух способов:
 - ▶ добавлением новых связей,
 - ▶ исключением незначимых связей.

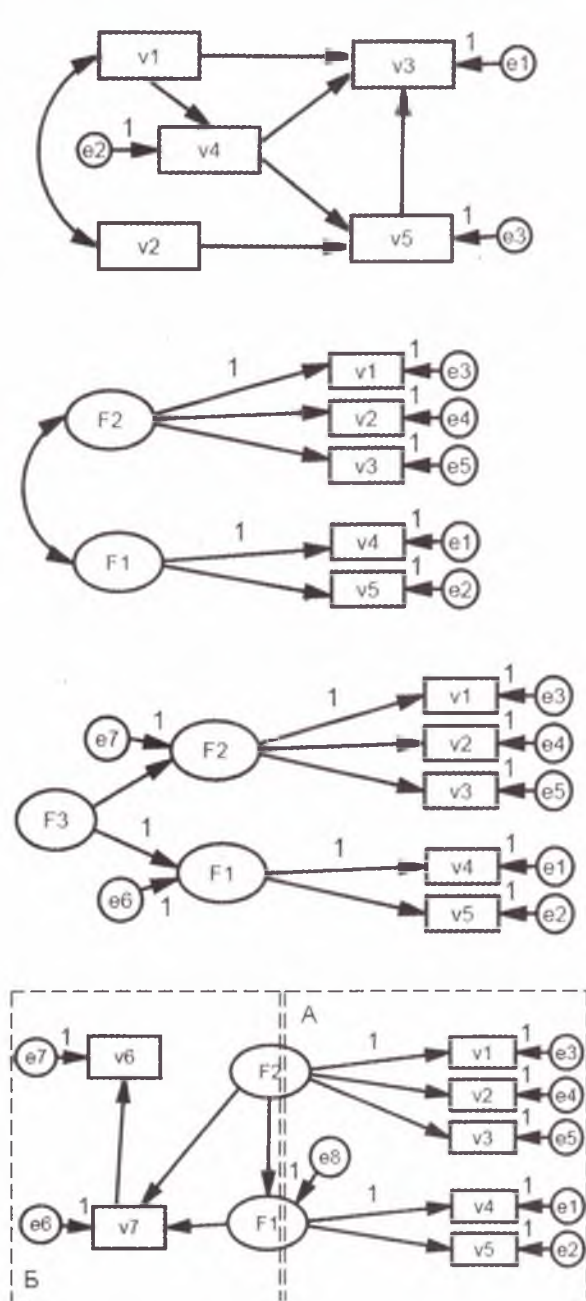
Формирование модели и ее идентификация

Типичные варианты моделей SEM

Модель путей (рис. 26.2, а) представляет собой гипотетическую систему направленных (причинно-следственных) и ненаправленных (ковариационных) связей между явными переменными. В модели путей, как правило, несколько зависимых (v_3, v_4, v_5) и несколько независимых переменных (v_1, v_2, v_4, v_5). Соответственно, допускается, чтобы некоторые переменные выступали одновременно в качестве зависимых от одних переменных и независимых — от других (v_4, v_5).

Конфирматорный факторный анализ, или *измерительная модель* (рис. 26.2, б и в), основан на предположении о том, что взаимосвязь между несколькими явными переменными обусловлена общей причиной их совместной изменчивости: фактором как латентной переменной. Каждая из таких явных переменных измерена, представлена в исходных данных и рассматривается в качестве *индикатора* некоторого гипотетического латентного конструкта, непосредственно не поддающегося измерению. Измерительная модель позволяет проверить предположение о том, что данного набора индикаторов достаточно для измерения латентного конструкта, и определить, какой именно вклад вносит каждый индикатор в его оценку. Корректировка модели заключается в исключении несостоятельных индикаторов (явных переменных) и (или) латентных переменных (факторов), и (или) включении индикаторов одних латентных переменных в состав индикаторов других латентных переменных. В последнем случае границы конструктов размываются: конструкты становятся пересекающимися. Измерительная модель позволяет также проверять гипотезы о состоятельности вторичных факторов (рис. 26.2, в).

Структурная модель (рис. 26.2, г) включает в себя измерительную составляющую и модель путей. Как правило, такие модели формируются пошагово. Сначала измерительные составляющие модели проверяются на соответствие исходным данным и в случае необходимости корректируются. Затем к ним добавляется структурная составляющая, после чего вся модель проверяется на соответствие исходным данным.



a — модель путей; v_1, v_2 — экзогенные переменные, v_3, v_4, v_5 — эндогенные переменные, e_1, e_2, e_3 — ошибки для эндогенных переменных

б — конфирматорный факторный анализ (измерительная модель): F_1, F_2 — латентные переменные, факторы; $v_1, \dots, v_5$ — явные переменные, индикаторы факторов; $e_1, \dots, e_5$ — ошибки явных переменных

в — конфирматорный факторный анализ с вторичным фактором (измерительная модель):

F_1, F_2 — первичные латентные переменные, факторы; F_3 — вторичная латентная переменная (фактор); $v_1, \dots, v_5$ — явные переменные; $e_1, \dots, e_7$ — ошибки эндогенных переменных

г — структурная модель: *A* — измерительная составляющая модели; *Б* — составляющая путей модели; F_1, F_2 — латентные переменные; $v_1, \dots, v_7$ — явные переменные; $e_1, \dots, e_8$ — ошибки эндогенных переменных

Рис. 26.2. Типичные виды моделей SEM

Переменные и параметры модели

В модели выделяют *экзогенные* переменные (v_1 и v_2 на рис. 26.2, *а*, или F_1 и F_2 на рис. 26.2, *б*), выступающие исключительно как независимые, и *эндогенные* переменные, к которым направлена хотя бы одна стрелка ($v_1, \dots, v_3$ на рис. 26.2, *а*, или $v_1, \dots, v_5$ на рис. 26.2, *б*). Такая классификация имеет принципиальное значение: причины изменчивости экзогенных переменных выходят за рамки модели, не учитываются в ней, следовательно, нет необходимости предполагать их ошибку. В то же время экзогенные переменные могут коррелировать между собой, и это учитывается в модели при помощи двунаправленной стрелки, обозначающей ковариацию, подлежащую оценке. В свою очередь, причины изменчивости эндогенной переменной в той или иной степени учитываются в модели, поэтому следует предполагать часть ее дисперсии, не учтенной в модели. Поэтому для эндогенных переменных обязательным является добавление ошибки ($e_1, \dots, e_3$ на рис. 26.2, *а*).

Параметры модели — это связи между переменными модели (регрессионные коэффициенты, ковариации, корреляции), дисперсии и средние переменных. При формировании модели необходимо различать 4 типа параметров модели и 3 способа их определения. Типы параметров:

- ▶ *связи* (направленные и ненаправленные) между явными и (или) латентными переменными;
- ▶ *ошибки явных переменных*, обусловленные их спецификой (uniquenesses), например $e_1, \dots, e_5$ на рис. 26.2, *б*;
- ▶ *ошибки латентных переменных*, обусловленные вмешивающимися, не учтенными в модели причинами (disturbances), например e_8 на рис. 26.2, *г*;
- ▶ *дисперсии* независимых латентных переменных.

Каждый элемент модели соотносится с параметрами, часть из которых оставляют свободными для последующего оценивания, а часть фиксируют (заранее задают их значение).

Все параметры, требующие конкретизации (оценивания или фиксирования), изображаются в модели в явном виде как связи между переменными или характеристики самих переменных (дисперсии, средние). При этом каждый параметр может быть определен одним из трех способов: как свободный, фиксированный или исключенный. *Свободный параметр* подлежит оценке в ходе вычислений по исходным данным. *Фиксированный параметр* обычно приравнивается к 1 см (рис. 26.2) и в процессе вычислений не оценивается. *Исключенный* или *ограниченный параметр* (constrained parameter) — обычно отсутствующая связь между переменными, своего рода фиксированный параметр, приравненный к нулю. Например, на рис. 26.2, *а*, исключен параметр связи между переменными v_2 и v_4 , следовательно, он равен 0.

Идентификация модели

Проблема идентификации модели касается соотношения между тем, что должно быть оценено (свободные параметры), и информацией, которая может быть для этого использована (корреляции, дисперсии и ковариации исходных переменных). Информация, содержащаяся в исходных данных, может быть определена количественно как число элементов ковариационной матрицы переменных, включая диагональ, на которой располагаются дисперсии переменных: $P(P + 1)/2$, где P — количество переменных в исходных данных. Таким образом, появляется возможность количественно оценить

уровень идентификации модели, сопоставив количество ее свободных параметров (T) и количество исходной информации для их оценки ($P(P + 1)/2$).

Выделяют три уровня идентификации модели:

1. *Недоопределенная модель* (under-identified model) характеризуется тем, что количество свободных параметров T превышает число элементов ковариационной матрицы ($P(P + 1)/2$) и оценка параметров невозможна.
2. *Строго определенная модель* (justidentified model) характеризуется равенством количества свободных параметров T и $P(P + 1)/2$, при этом все параметры модели оцениваются однозначно.
3. *Переопределенная модель* (over-identified model), в которой свободных параметров T меньше, $P(P + 1)/2$, допускает более одного способа определения свободных параметров.

Число степеней свободы модели (df) определяется как разность числа элементов ковариационной матрицы и количества свободных параметров модели, подлежащих оценке: $df = P(P + 1)/2 - T$. Очевидно, что оценка параметров модели, как, впрочем, и ее анализ в целом, возможны при условии df не меньше 0. Также очевидно, что модель будет в точности соответствовать данным, если $df = 0$, и этот случай вряд ли может представлять какой-то интерес. При одном и том же числе переменных модели чем выше df , тем более экономно и лаконично модель отражает данные и тем выше ее объяснительная ценность.

Следует отметить несколько проблем, связанных с идентификацией модели, которые довольно часто встречаются на практике:

- Во-первых, ковариационная (корреляционная) матрица должна быть положительно определена (positive definite), иначе анализ будет невозможен. Это означает отсутствие функциональных связей между переменными (корреляций, равных 1 или близких к этому значению).
- Во-вторых, модель должна быть рекурсивной, то есть не должна иметь обратных (реципрокных) направленных связей. При наличии таких связей между переменными анализ, как правило, невозможен. Подходы к анализу нерекурсивных моделей изложены Р. Клайном [5, с. 132].
- В-третьих, может отсутствовать сходимость процедуры оценки параметров, что обычно связано с неудачным фиксированием параметров. Например, в измерительной модели связь вторичного и одного из первичных факторов должна быть фиксированной — равной 1 (см. рис. 26.2, в). Если на самом деле эта связь близка к 0, то высока вероятность того, что анализ не будет выполнен. В этом случае решением проблемы может быть фиксирование альтернативного параметра.

Оценка, проверка согласованности и коррекция модели

Когда модель сформирована — заданы переменные из файла исходных данных, свободные и фиксированные параметры, можно запускать процедуру анализа модели средствами AMOS. В процессе анализа AMOS оценивает параметры модели и вычисляет показатели (индексы) ее соответствия исходным данным.

Оценка параметров модели

Оценка параметров модели может осуществляться в AMOS пятью методами, выбор которых доступен на вкладке Amos Graphics ► View ► Analysis Properties: Estimation (рис. 26.3). По умолчанию задан наиболее робастный метод Maximum likelihood (Максимального правдоподобия), и в подавляющем большинстве случаев нет необходимости его менять. Однако этот метод имеет ограничение: многомерную нормальность распределения переменных (см. раздел «Требования к данным»). Если это требование никак не удастся соблюсти, то в качестве альтернативы может быть применен метод Asymptotically distribution-free (Приблизительно свободный от распределения), который, однако, адекватен только для очень больших выборок (более 400–500 наблюдений). Для выборок меньшей численности рекомендуется применять методы Бутстреп (Bootstrap methods) [3], [5].

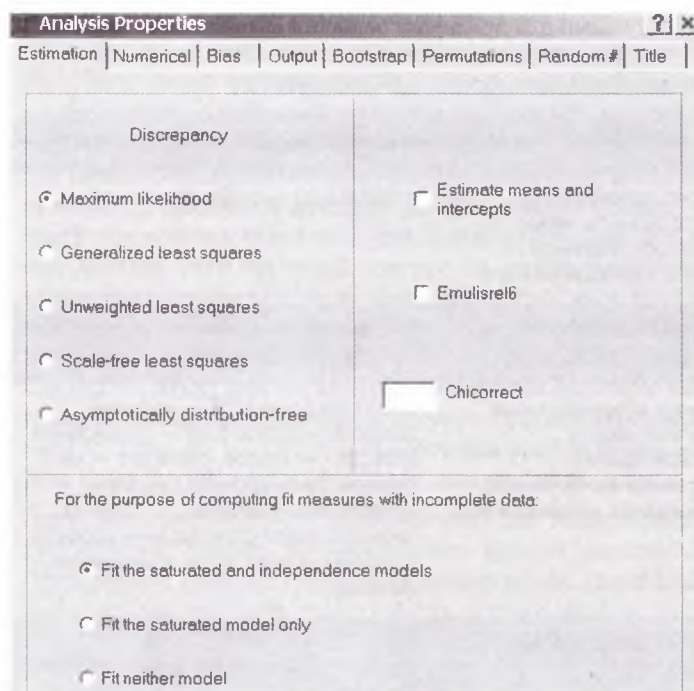


Рис. 26.3. Окно Analysis Properties: Estimation

Оценки параметров модели с их статистической значимостью доступны в окне вывода, в разделе Estimates (Оценки): Amos Graphics ► View ► Text Output ► Estimates.

Проверка согласованности модели

Одновременно с вычислением оценок параметров модели программа AMOS вычисляет различные критерии (индексы) согласия модели с исходными данными. Эти показатели оценивают *расхождение* (discrepancy) между исходной ковариационной матрицей и репродуцированной в соответствии с моделью. Всего программа вычисляет около 20 показателей согласованности, сгруппированных при выводе результатов в 10 таблиц.

Показатели согласованности вычисляются для трех моделей: Default model (Модель по умолчанию), то есть пользовательская модель; Saturated model (Насыщенная модель), в которой число оцениваемых (свободных) параметров максимально возможно и равно T — количеству элементов ковариационной матрицы исходных данных; Independence model (Независимая модель), в которой все корреляции между переменными модели равны 0. При оценке согласия модели исходным данным принимают во внимание только показатели для модели по умолчанию. Показатели для двух других моделей свидетельствуют о диапазоне возможных значений каждого показателя для данной модели: для насыщенной модели — минимальное значение, для независимой модели — максимальное значение.

Исследователь принимает решение о приемлемой согласованности модели с исходными данными на основе нескольких критериев согласия. Назначение и смысл каждого из них подробно обсуждаются в [5]. А здесь ограничимся краткой характеристикой наиболее популярных критериев и их граничными для принятия решения значениями (табл. 26.2).

Таблица 26.2. Некоторые критерии согласия модели с данными (по [5])

Критерий	Название	Интерпретация
CMIN, df , p	Критерий хи-квадрат, проверяет H_0 на предмет того, что разность между эмпирической и воспроизведенной по модели матрицами ковариаций равна 0	p более 0,05 — хорошее согласие
CMIN/df	Отношение хи-квадрат к числу степеней свободы	В пределах 2 — приемлемое согласие
RMSEA, LO90 HI90, PCLOSE	Root mean square error of approximation (RMSEA) — квадратный корень среднеквадратической ошибки аппроксимации; границы его 90 %-ного доверительного интервала, и оценка его «точности»	RMSEA не более 0,05 — хорошее согласие, не более 0,08 — приемлемое, 0,08...0,1 — слабое, более 0,1 — нет. Верхняя граница (HI90) не должна существенно превышать 0,1. Оценка точности: около 0,5 и выше — хорошее согласие, не ниже 0,1 — приемлемое согласие
GFI, NFI	Comparative Fit Index (сравнительный индекс согласия), Normed Fit Index (нормированный индекс согласия)	Оба — не менее 0,95 (хорошее согласие) или не менее 0,9 (приемлемое согласие)
GFI, AGFI	Goodness-of-Fit Index (критерий согласия), Adjusted Goodness-of-Fit Index (исправленный критерий согласия)	Оба — не менее 0,90 (хорошее согласие). Обычно AGFI < GFI

Следует отметить, что объем выборки заметно влияет на эти критерии: чем больше объем выборки, тем хуже показатели согласия. Поэтому для больших выборок (более 300–400 наблюдений) для принятия решения о согласии модели с данными могут применяться менее строгие пределы их значений. Решение о согласии модели с исходными данными обычно принимается на основе не менее трех (а лучше пяти) из этих критериев. Отметим, что традиционным является критерий хи-квадрат, а наиболее робастным, по-видимому, RMSEA.

Коррекция модели

Обычно исходная (априорная) модель по критериям согласия, мягко говоря, не вполне соответствует исходным данным. Говоря точнее, расхождение между исходной и воспроизведенной по модели ковариационными матрицами оказывается слишком велико, чтобы считать модель приемлемо соответствующей данным. Поэтому практически всегда возникает необходимость коррекции модели.

Очевидный путь коррекции модели — удаление из модели статистически недостоверных связей: удаляя ту или иную стрелку, мы переводим соответствующий свободный параметр в статус исключенного, то есть фиксируем его нулевое значение в модели. Статистическую значимость всех параметров (кроме фиксированных) содержит раздел вывода Estimates: Scalar Estimates (Оценки: Скалярные оценки). Если же интересует статистическая значимость фиксированного в модели параметра, следует сделать этот параметр свободным, то есть зафиксировать альтернативный параметр и вновь вычислить оценки.

Удаление какой-либо связи из модели (фиксирование нулевого значения соответствующего параметра) увеличивает число степеней свободы модели на 1 и делает ее более экономичной, лаконичной, повышая ее ценность в соответствии с общенаучным принципом экономии доводов (бритвы Оккама). Такая более лаконичная модель, полученная путем фиксирования нулевого значения одного из параметров (исключение связи), называется *вложенной моделью* (по отношению к исходной). Для оценки целесообразности исключения какой-либо связи можно использовать *критерий различий для вложенных моделей*. Этот критерий основан на оценке статистической достоверности различия критерия хи-квадрат для исходной и вложенной моделей. Сначала вычисляется хи-квадрат и df для исходной ($CMIN1$, $df1$) и вложенной ($CMIN2$, $df2$) моделей. Затем по таблице критических значений хи-квадрат определяется статистическая достоверность разности $CMIN2 - CMIN1$ для числа степеней свободы $df2 - df1$. Если результат статистически достоверен, значит, исходная модель лучше согласуется с данными, удаление связи ухудшает модель. Если же результат статистически недостоверен, это означает, что вложенная модель соответствует данным не хуже, чем исходная, и такая модификация целесообразна.

Дополнительно AMOS предоставляет два типа информации, весьма полезной для диагностики причин несоответствия модели данным: *стандартизованные остатки* (standardized residuals) и *индексы модификации* (modification indices). Отметим, что эта информация доступна только в том случае, если в данных нет пропущенных значений. Для того чтобы получить в окне вывода эти показатели, достаточно в окне Amos Graphics ► View ► Analysis Properties: Output (Параметры анализа: Вывод) (рис. 26.4) установить флажки Residual moments (Остатки моментов) и Modification indices (Индексы модификации).

Стандартизованные остатки в выводе представлены в виде матрицы размерностью $P \times P$ (P — количество явных переменных модели, содержащихся в данных) (рис. 26.5). Каждый элемент этой матрицы — стандартизованная разность между исходной ковариацией и воспроизведенной по модели.

Поскольку остатки стандартизованы, они являются аналогами z -значений, и это существенно облегчает их интерпретацию. Например, значение остатка более 2,58 статистически достоверно отличается от 0 на уровне 0,01. Если наблюдаются настолько значительные отклонения для ковариации пары переменных, это веский аргумент в пользу коррекции модели: либо заданием свободного параметра, связывающего одну или обе переменные с другими или между собой, либо удалением из модели одной из них.

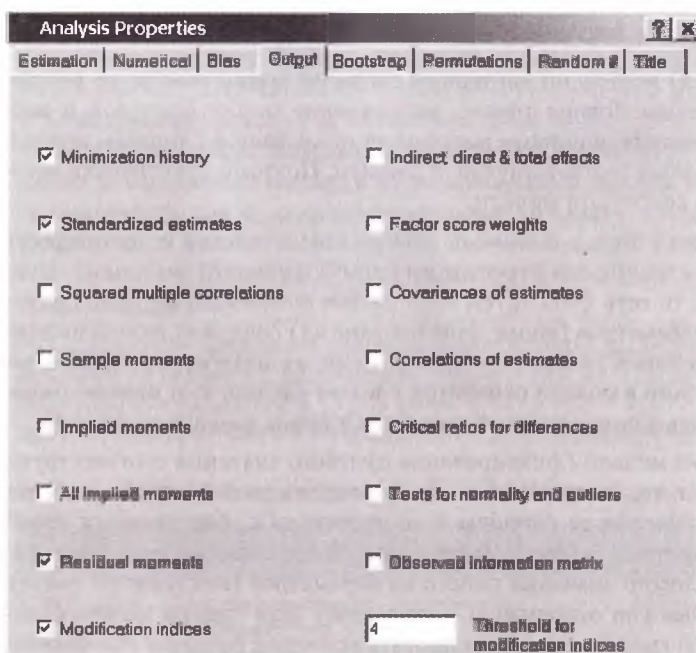


Рис. 26.4. Окно Analysis Properties: Output (Параметры анализа: Вывод)

Standardized Residual Covariances (Group number 1 - Default model)		тест5	тест4	тест3	тест2	тест1
тест5		,000				
тест4		,102	,000			
тест3		,050	-,162	,000		
тест2		1,236	-,466	-,846	,000	
тест1		,678	-,792	,032	,000	,000

Рис. 26.5. Фрагмент окна вывода Standardized Residual Covariances (Стандартизованные остатки ковариаций)

Индексы модификации (Modification indices, M.I.) доступны в одноименном разделе окна вывода (рис. 26.6). По смыслу они являются аналогами критерия хи-квадрат с одной степенью свободы и вычисляются для каждого исключенного параметра (фиксированного как 0). Значение M.I. показывает, насколько приблизительно уменьшится хи-квадрат модели при освобождении этого параметра (задании соответствующей связи). Для каждого M.I. указывается также величина ожидаемого изменения параметра (expected parameter change), обозначаемая Par Change и предсказывающая, насколько и в каком направлении отклонится от нуля соответствующий параметр, если он будет освобожден. Ориентируясь на величину M.I., исследователь может целенаправленно добавлять связи (освобождать параметры) для повышения согласия модели с данными. При этом, разумеется, модель теряет свою ценность: становится более громоздкой и менее экономичной.

ременную, имеющую 3 категории, как две бинарные фиктивные переменные (dummy variables)¹.

Нормальность распределения и выбросы. Большинство процедур SEM предполагают нормальность распределения переменных, точнее, их многомерную нормальность. Для проверки этого предположения в AMOS включена удобная процедура Tests for normality and outliers (Критерии нормальности и выбросы), которая доступна, если в данных нет пропусков. Соответствующий результат появляется в выводе, если в диалоговом окне Analysis Properties: Output (Параметры анализа: Вывод) (см. рис. 26.4) установлен флажок Tests for normality and outliers. В разделе Assessment of normality (Оценка нормальности) окна вывода (рис. 26.7) приводятся показатели асимметрии (skew), эксцесса (kurtosis), их критических коэффициентов (c.r. — critical ratio). Многомерный эксцесс и его критический коэффициент приведены в последней строке Multivariate (Многомерный). Наиболее важен критический коэффициент для многомерного эксцесса: его величина в пределах пяти свидетельствует о приблизительной многомерной нормальности распределения переменных. Если значение критического коэффициента для многомерного эксцесса превышает 5, то причиной отклонения от многомерной нормальности может являться ненормальность распределения отдельных переменных. Об этом можно судить по показателям асимметрии и эксцесса для каждой из переменных. Распределение переменной существенно отклоняется от нормального, если показатель асимметрии (skew) превышает 3 и (или) показатель эксцесса (kurtosis) превышает 7 [3], [5].

Assessment of normality (Group number 1)						
Variable	min	max	skew	c.r.	kurtosis	c.r.
test5	5,000	19,000	,031	,128	-,160	-,327
test4	4,000	21,000	,195	,796	,182	,372
test3	4,000	22,000	,257	1,051	1,162	2,372
test2	4,000	17,000	-,142	-,581	-,438	-,894
test1	4,000	17,000	-,074	-,304	-,439	-,896
Multivariate					2,337	1,397

Рис. 26.7. Фрагмент окна вывода Assessment of normality (Оценка нормальности)

Для поиска и исключения причин отклонения распределения переменных от многомерной нормальности можно использовать раздел окна вывода Observations farthest from the centroid (Mahalanobis distance) (Наблюдаемые удаления от центроида (Расстояния Махаланобиса)) (рис. 26.8). В этом окне наблюдения упорядочены по мере убывания величины квадрата их расстояния от центра многомерного распределения (центроида). В двух последних столбцах указаны вероятности в предположении многомерной нормальности: p1 — вероятность того, что любое случайное наблюдение превысит данную величину удаления от центроида, p2 — вероятность того, что максимальное значение расстояния от центроида отдельно взятого наблюдения превысит указанную величину. В случае многомерной нормальности малые значения p1 не редкость. А очень малое значение p2 свидетельствует о необычности наблюдения для многомерного нормального распределения. Таким образом, исключая из выборки наблюдения с очень малым значением p2 (выбросы), можно уменьшить отклонение распределения переменных от многомерной нормальности.

¹ Казалось бы, логичное замещение такой номинальной переменной тремя бинарными на самом деле некорректно, так как значения двух из них однозначно детерминируют значение третьей переменной, то есть неизбежна их коллинеарность.

Observations farthest from the centroid (Mahalanobis distance) (Group number 1)			
Observation number	Mahalanobis d-squared	p1	p2
95	16.702	.005	.400
39	13.635	.018	.542
17	13.510	.019	.297
60	12.088	.034	.434
56	11.933	.036	.287
37	11.490	.042	.252
55	10.630	.059	.381

Рис. 26.8. Фрагмент окна вывода Observations farthest from the centroid (Mahalanobis distance) (Наблюдаемые удаления от центраоида (Расстояния Махаланобиса))

Если по тем или иным причинам требование многомерной нормальности соблюсти не удастся, то можно применить специальный метод оценки Asymptotically distribution-free (Приблизительно свободный от распределения), установив в окне Analysis Properties (Параметры анализа) переключатель Discrepancy в соответствующее положение (см. рис. 26.3). Однако применять этот метод рекомендуется только для очень больших выборок (не менее 400–500 наблюдений).

Проблема коллинеарности

Коллинеарность является проблемой для всех многомерных методов анализа корреляционных или ковариационных матриц и обозначает, что связь между некоторыми переменными равна 1 или приближается к ней (по модулю)¹. В этом случае некоторые вычислительные процедуры с матрицами, например вычисление обратной матрицы, если и возможны, то производятся с ошибками. При наличии коллинеарности программа AMOS либо вообще не вычисляет оценки параметров, либо вычисляет их с ошибками: могут появляться оценки корреляций или стандартизированных коэффициентов, превышающие 1, или отрицательные оценки дисперсий. Однако ошибки оценок как последствия коллинеарности могут быть и не столь очевидными и заметными. Поэтому рекомендуется перед применением SEM выполнить диагностику коллинеарности.

Один из доступных способов диагностики коллинеарности — применение множественного регрессионного анализа (см. главу 18). Для этого анализ можно провести для каждой из переменных (как зависимой) с остальными (как независимыми). Показателем экстремально высокой многомерной коллинеарности в этом случае может служить КМК, превышающий 0.9. Можно ограничиться и однократным применением множественной регрессии в отношении одной из переменных, (как зависимой) с остальными (как независимыми). В этом случае в меню Линейная регрессия: Статистики (см. рис. 18.2) следует установить флажок Диагностики коллинеарности. При установке этого флажка для каждой независимой переменной вычисляется *толерантность* ($1 - \text{КМК этой переменной с остальными независимыми переменными}$). Показателями экстремальной многомерной коллинеарности будут являться значения КМК зависимой переменной, превышающие 0,9, и (или) толерантности хотя бы одной независимой переменной, меньшие 0,1.

Очевидный способ снижения или устранения коллинеарности — удаление переменной с максимальным значением КМК или с минимальным значением показателя толерантности.

¹ Корреляционная (ковариационная) матрица не является положительно определенной.

Пропущенные значения

Наличие пропусков в данных не является непреодолимым препятствием для применения SEM. Если в данных есть пропущенные значения, то для оценки параметров необходимо установить флажок Estimate means and intercepts (Оценка средних и свободных членов) в окне Analysis Properties (Параметры анализа) (см. рис. 26.3). Однако при установке этого флажка многие полезные показатели модели становятся недоступными. Решением проблемы может быть создание двух файлов данных: исходного (с пропусками) и дополнительного, без пропусков (с удаленными строками, или с восстановленными пропусками). Сначала модель анализируется и модифицируется с использованием файла данных без пропусков, затем проверяется на файле с пропусками с установленным флажком Estimate means and intercepts (Оценка средних и свободных членов).

Наиболее корректным и простым способом создания файла без пропущенных значений является удаление из исходного файла строк, содержащих пропущенные значения. Однако такой способ может привести к существенному сокращению численности выборки. Альтернативой является восстановление пропущенных значений либо заменой пропущенных значений средними для соответствующих переменных, либо применением специальных процедур в рамках анализа пропущенных значений. Первый способ более грубый, доступен в меню Преобразовать ► Заменить пропущенные значения (Метод — Среднее ряда). Второй способ более точный, доступен в меню Анализ ► Анализ пропущенных значений. При выборе метода восстановления по методу регрессии (Оценивание — Регрессия) и задании всех количественных переменных для восстановления программа восстанавливает каждый пропуск как оценку значения данной переменной (как зависимой) по значениям остальных переменных (как независимых).

Объем выборки

Требования к объему выборки для SEM строже, чем для традиционных многомерных методов, что аргументируется обычно зависимостью точности и стабильности оценок параметров от количества наблюдений. Однако общепринятые правила оценки достаточного объема выборки еще не сложились. Рекомендуемая минимальная численность наблюдений колеблется в диапазоне от 100 до 400, а наиболее часто встречающееся в исследованиях количество наблюдений обычно не менее 200.

Более определенные критерии предлагает Р. Клайн, основываясь на соотношении числа наблюдений N и количества оцениваемых параметров $T = P(P + 1)/2 - df$ [5], где P — количество явных переменных модели, df — число степеней свободы. В соответствии с его рекомендациями идеальной следует считать такую численность выборки N , которая в 20 раз превышает количество оцениваемых параметров T , 10-кратное превышение N над T считается приемлемым, а 5-кратное — явно недостаточным для уверенности в точности и стабильности оценок параметров. Еще строже требования к объему выборки при использовании методов оценок, свободных от предположения о многомерной нормальности.

Пошаговые алгоритмы вычислений

Как правило, модель формируется и корректируется исключительно в графическом виде в главном окне Amos Graphics (рис. 26.9).

Средства AMOS преобразуют эту графическую модель в исполняемую программу, включающую в себя систему линейных уравнений. Пользователь AMOS имеет возможность

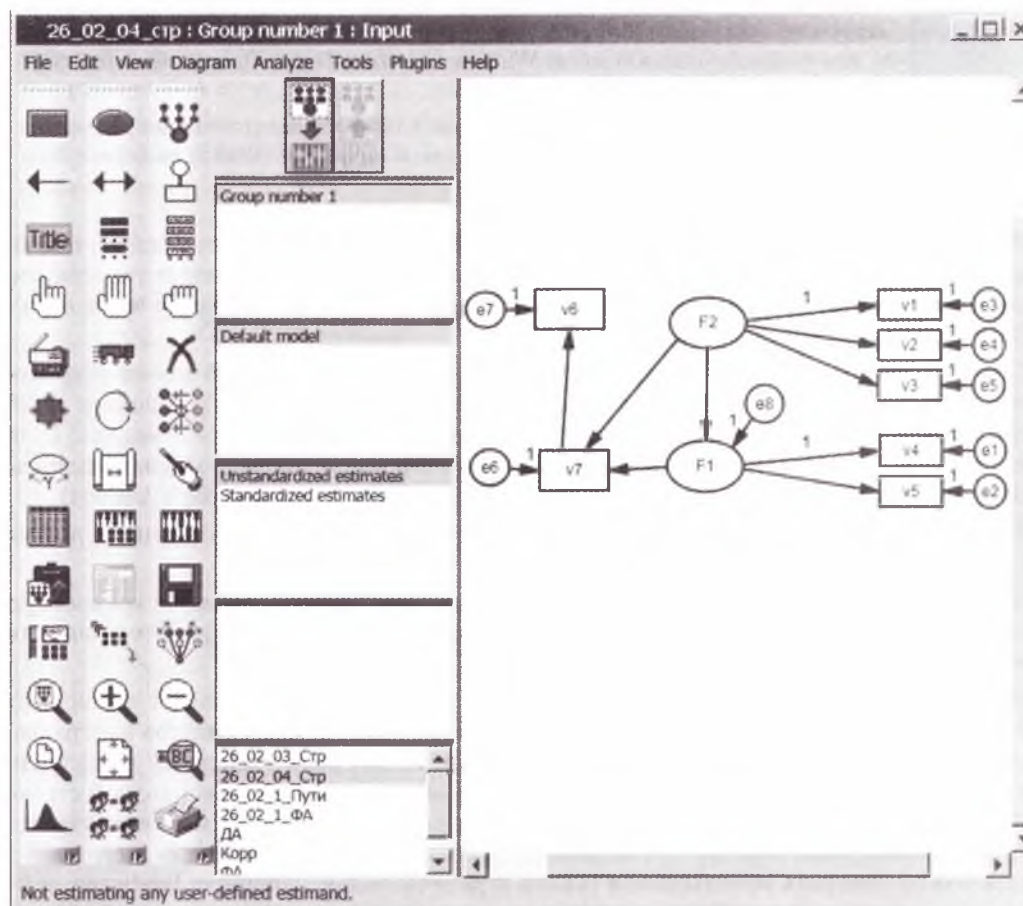


Рис. 26.9. Главное окно Amos Graphics

обратиться к этому тексту программы на любом этапе формирования и корректировки модели. Если модель изначально строится программными средствами, для этого необходимо вместо Amos Graphics в меню Пуск ► Программы ► IBM SPSS Statistics ► IBM SPSS Amos 20 выбрать Program Editor. Если же графическая модель уже готова и требуется ее редакция программными средствами, то достаточно в главном окне Amos Graphics в командной строке в меню Tools выбрать Write a Programm и в появившемся диалоговом окне нажать OK. Однако использование программных средств для работы с моделью — удел весьма продвинутых пользователей. Поэтому здесь мы ограничимся работой только с графическими средствами формирования и коррекции моделей, предоставляемыми интерфейсом модуля Amos Graphics.

До применения AMOS нужно выполнить два подготовительных шага. Эти шаги позволят запустить программу IBM SPSS Amos 20.

ШАГ 1

Подготовьте новый или существующий файл данных. Для конфирматорного факторного анализа это файл IQ.sav (шаги 3–10), а для структурной модели — файл ex01.sav (шаги 3а–10а). Все необходимые для этого действия описаны в главе 3.

ШАГ 2

Запустите программу IBM SPSS Amos 20 при помощи значка на Рабочем столе или выбрав в главном меню Windows команду Пуск ► Программы ► IBM SPSS Statistics ► IBM SPSS Amos 20 ► Amos Graphics.

Если вы уже работали с этой программой, то на Рабочем столе будет последняя из сохраненных диаграмм. В этом случае в меню File (Файл) выберите New (Новый).

После выполнения этого шага на экране появится главное окно Amos Graphics, подобное изображенному на рис. 26.9. Основные области главного окна: строка состояния, командная строка (вверху окна), панель инструментов с кнопками (левая часть окна), Рабочий стол (правая часть окна).

Строка состояния — верхняя строка главного окна, информирует об имени открытого файла с моделью, его состоянии (Input (Ввод) или Output (Вывод)) и номере группы (выборки, подвыборки), для которой строится и будет проверяться модель. Если вы впервые приступаете к построению модели, то строка состояния содержит надпись: Unnamed Project: Group number 1: Input (Безымянный проект: Группа номер 1: Ввод).

Командная строка подобно одноименной строке в SPSS Statistics открывает доступ к соответствующим меню.

Панель инструментов содержит 42 кнопки, каждая из которых реализует конкретную операцию. С большинством из них мы познакомимся в процессе выполнения пошаговых инструкций.

Рабочий стол занимает всю левую, основную часть окна и представляет собой лист бумаги, изначально пустой, на котором при помощи панелей инструментов выстраивается модель. Формат Рабочего стола, по умолчанию Portrait-Letter (Книжный — письменный), можно менять в любой момент построения модели, выбирая в командной строке View ► Interface Properties (Вид ► Параметры интерфейса). В открывшемся окне Interface Properties на вкладке Page Layout (Разметка страницы) в меню Paper Size (Размер страницы) можно изменить ориентацию и размер страницы, задав, например, Landscape — A4 (Альбомный, размер A4). Если модель ориентирована преимущественно горизонтально, как на рис. 26.9, стоит задать альбомную ориентацию страницы (Landscape), а если вертикально — книжную (Portrait).

Далее приведены пошаговые инструкции для формирования и проверки средствами Amos Graphics двух моделей: измерительной модели (конфирматорный факторный анализ, шаги 3–11) с использованием файла данных IQ.sav и структурной модели, включающей в себя измерительную и «путевую» составляющие, с использованием файла данных ex01.sav (структурная модель, шаги 3а–11а).

Конфирматорный факторный анализ

Формирование модели КФА

Для проведения конфирматорного факторного анализа (КФА) будем использовать тот же файл данных, IQ.sav, в отношении которого применялся традиционный факторный анализ в главе 20. Этот файл данных включает в себя результаты тестирования 184 школьников по 11 субтестам интеллекта. Предположим, эти 11 показателей интеллекта представляют собой измерения (индикаторы) трех способностей как латентных конструктов: вербальных способностей, логико-математических способностей и невербальных (образных) способностей. Это предположение может быть представлено в виде

исходной гипотетической модели — блок-схемы с использованием правил построения моделей в среде AMOS (рис. 26.10). Эта модель сохранена в наборе файлов примеров под именем КФА.amw.

Таким образом, на первом этапе КФА исходные предположения о структуре данных изображаются в виде блок-схемы, к построению которой средствами Amos Graphics мы и приступим.

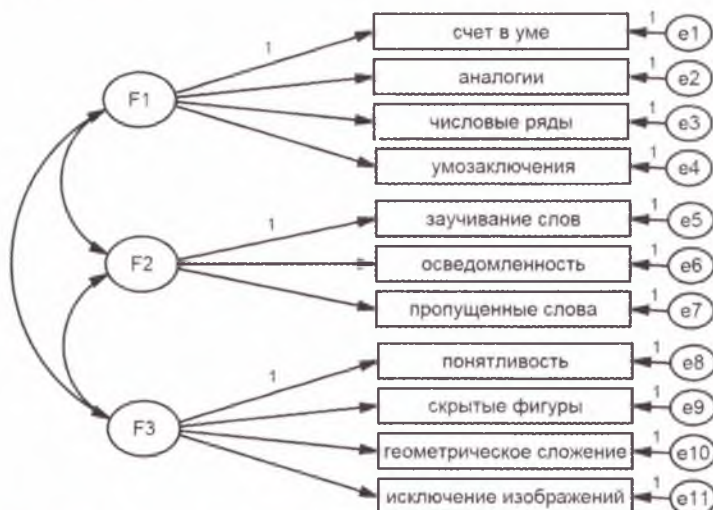


Рис. 26.10. Исходная модель структуры интеллекта

Модель состоит из трех частей, каждая из которых соответствует одному из факторов. Фактор — это независимая латентная переменная, соотносящаяся с набором явных переменных — его индикаторов, выступающих в качестве независимых переменных. В модели КФА факторы являются экзогенными переменными, а их индикаторы — эндогенными переменными. Переменная является эндогенной в силу того, что часть ее дисперсии обусловлена влиянием независимых переменных, включенных в модель. Остаточная дисперсия этой переменной обусловлена влиянием причин, внешних по отношению к модели (ошибкой модели, связанной со специфичностью данной переменной). Поэтому в SEM предполагается строгое правило построения моделей: если переменная эндогенная, то для нее необходимо в явном виде представить источник ее специфичности — ошибку, которую обычно обозначают буквой «е» (error — ошибка). Причины изменчивости экзогенных переменных выходят за рамки модели, поэтому в явном виде в ней не представлены. Если предполагается связь эндогенной переменной с другими переменными модели, то она всегда направленная, а связь между экзогенными переменными всегда ненаправленная. Еще одно правило касается задания регрессионных коэффициентов для индикаторов латентной переменной: один из этих коэффициентов как свободный параметр должен быть фиксирован как 1.

После выполнения шага 2 у вас должно быть открыто главное окно Amos Graphics, отличающееся от изображенного на рис. 26.9 только тем, что Рабочий стол представляет собой пустой лист, на котором предстоит построить модель.

В дидактических целях каждую из трех частей модели мы будем строить разными способами, чтобы познакомиться с инструментами построения модели (шаги 3–5).

ШАГ 3

1. На панели инструментов щелкните на значке, изображающем овал, Draw unobserved variables (Рисовать латентные переменные) и переведите указатель мыши на Рабочий стол. Рядом с указателем появится овал, а соответствующая кнопка будет выделена. В соответствии со схемой, изображенной на рис. 26.10, нарисуйте овал для переменной F1, удерживая нажатой левую клавишу мыши.
2. Щелкните на кнопке с изображением прямоугольника Draw observed variables (Рисовать явные переменные) и нарисуйте прямоугольник, соответствующий переменной счет в уме. Нарисуйте тем же способом прямоугольники, соответствующие трем остальным переменным-индикаторам фактора F1.
3. Щелкните на кнопке с овалом и нарисуйте справа от первого прямоугольника кружок, соответствующий ошибке e1. Для этого достаточно щелкнуть в соответствующем месте на Рабочем столе правой клавишей мыши (рядом с указателем должен быть виден овал). Повторите это действие для ошибок e2, e3 и e4.
4. Щелкните на кнопке с изображением однонаправленной стрелки и нарисуйте все направленные связи между изображенными переменными в соответствии со схемой (см. рис. 26.10). Начинайте рисовать каждую стрелку (нажатием левой клавиши мыши) внутри контура переменной и заканчивайте внутри контура другой переменной (отпуская левую клавишу мыши).

Теперь необходимо зафиксировать некоторые свободные параметры, задав их равными 1. Это касается регрессионного коэффициента для одного из индикаторов латентной переменной и регрессионных коэффициентов для ошибок.

5. Наведите указатель на стрелку от фактора к первому из его индикаторов (она станет красной) и сделайте двойной щелчок правой клавишей мыши. В открывшемся диалоговом окне Object Properties (Параметры объекта) перейдите на вкладку Parameters (Параметры) (рис. 26.11). В поле Regression weight (Регрессионный коэффициент) введите 1. Закрыв окно Object Properties, вы увидите 1 возле стрелки, что означает фиксирование свободного параметра на 1. Тем же образом задайте 1 для регрессионных весов (стрелок) от каждой ошибки к соответствующей явной переменной.

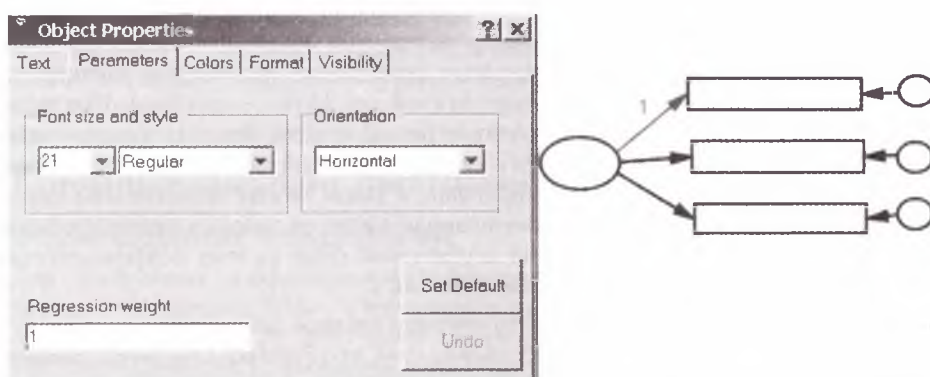


Рис. 26.11. Фиксирование свободного параметра в окне Object Properties: Parameters (Свойства объекта: Параметры)

В результате выполнения шага 3 у вас на Рабочем столе должны появиться контуры части модели, относящейся к фактору F1 (см. рис. 26.10). Как правило, эти контуры безупречны и требуют коррекции.

Коррекция элементов диаграммы. Если вы хотите удалить какой-то элемент диаграммы, щелкните на кнопке с крестом ✕ **Erase objects** (Удаление объектов). Указатель мыши на Рабочем столе превратится в крест, щелкните им на удаляемом элементе. Если вы хотите передвинуть элемент, щелкните на кнопке  **Move objects** (Перемещение объектов) и перетащите этот элемент, удерживая его левой клавишей мыши. Наконец, если вы хотите изменить размер и пропорции элемента, щелкните на кнопке  **Change the shape of objects** (Изменение формы объекта), затем наведите указатель на редактируемый овал или прямоугольник и меняйте его форму. Выравнивание стрелок производится щелчком «волшебной палочкой»  **Touch up a variable** (Выравнивание переменной) на каждой из переменных.

Сохранение диаграммы. Сохранять диаграмму можно на любом этапе ее построения, но лучше сделать это в самом начале. В дальнейшем можно создавать несколько ее версий в соответствии с проверяемыми гипотезами. В процессе построения одной модели и ее проверки AMOS создает множество файлов, размещаемых в той же папке, где сохранена модель. Поэтому стоит для сохранения моделей, относящихся к одному файлу данных, создавать отдельную папку. Как правило, в процессе моделирования создается несколько альтернативных моделей, поэтому следует к имени каждой из моделей добавлять индекс. Например, для данной модели целесообразно в папке Примеры_SPSS создать папку IQ, в которой и сохранить ее под указанным именем (КФА1): **File** ► **Save As** ► **Примеры_SPSS** ► **IQ** ► (Имя файла: КФА1) ► **Сохранить**.

Теперь приступим к построению второй части модели, относящейся к фактору F2. Воспользуемся новыми инструментами:

-  — **Duplicate objects** (Копирование объектов);
-  — **Select one objects at a time** (Выделение одного объекта);
-  — **Deselect all objects** (Отмена выделения объектов);
-  — **Select all objects** (Выделение всех объектов);
-  — **Add a unique variable to an existing variable** (Добавление специфической переменной к существующей переменной).



1. Щелкните на кнопке **Select one objects at a time** (Выделение одного объекта). Теперь щелкните поочередно на трех из явных переменных на Рабочем столе, чтобы выделить их (контуры выделенных объектов становятся синими).
2. Щелкните на кнопке **Duplicate objects** (Копирование объектов). Наведите указатель на одну из выделенных переменных и, удерживая нажатой левую клавишу мыши, перетащите копии выделенных переменных вниз, чтобы образовать группу явных переменных для фактора F2. Щелкните на кнопке **Deselect all objects** (Отмена выделения объектов).
3. Щелкните на кнопке **Add a unique variable to an existing variable** (Добавление специфической переменной к существующей переменной). Наведите указатель на первую из новых явных переменных и щелкните на ней. Появится кружок со стрелкой, соответствующий ошибке для этой переменной. Щелкните на этой переменной еще дважды, чтобы изменить расположение ошибки. Расставьте ошибки для остальных двух переменных, повторив последние действия.

4. Выделите при помощи кнопки Select one objects at a time (Выделение одного объекта) три новые ошибки для их коррекции. Откорректируйте расположение и размер ошибок так, чтобы все ошибки выглядели одинаково. Для этого используйте кнопки Move objects (Перемещение объектов) и Change the shape of objects (Изменение формы объекта). После коррекции отмените выделение объектов.
5. Щелкните на кнопке Draw unobserved variables (Рисовать латентные переменные) и нарисуйте овал, соответствующий фактору F1. Щелкните на кнопке со стрелкой и нарисуйте направленные связи (стрелки) от фактора к каждой из трех явных переменных.
6. Зафиксируйте регрессионный коэффициент первой из явных переменных фактора F2, задав его значение 1, как это было сделано в п. 6 на шаге 3.

В результате выполнения шага 4 на диаграмме появится контур раздела, соответствующий фактору F2 (см. рис. 26.10). Надо отметить, что тот же результат можно было бы получить проще, выделив весь раздел для фактора F1 при помощи кнопки , его копирования при помощи кнопки  и перетаскивания копии вниз.

На следующем шаге мы воспользуемся специальным инструментом для создания латентной переменной и ее индикаторов — кнопкой  Draw a latent variable or add an indicator to a latent variable (Рисовать латентную переменную или добавить индикатор для латентной переменной). Дополнительно будем использовать новый инструмент  Rotate the indicators of a latent variables (Поворот индикаторов латентных переменных) и кнопку  Draw covariances (Рисовать ковариации).

ШАГ 5

1. Щелкните на кнопке Draw a latent variable or add an indicator to a latent variable (Рисовать латентную переменную или добавить индикатор для латентной переменной). Наведите указатель на свободное место Рабочего стола под контуром фактора F2 и щелкните левой клавишей мыши: появится овал, соответствующий латентной переменной (фактору). Каждый последующий щелчок на этом овале будет добавлять один индикатор и его ошибку для данного фактора. Щелкните 4 раза, чтобы добавить 4 индикатора (явные переменные).
2. Щелкните на кнопке Rotate the indicators of a latent variables (Поворот индикаторов латентных переменных). Наведите указатель на новую латентную переменную и щелкните на ней один раз: ее индикаторы развернутся на 90° по часовой стрелке.
3. Щелкните на кнопке с двунаправленной стрелкой Draw covariances (Рисовать ковариации). Соедините все факторы между собой двунаправленными стрелками. Начинайте рисовать стрелки с самого нижнего фактора, проводя каждую стрелку от внутреннего контура одной и заканчивая ее во внутреннем контуре другой переменной.
4. Откорректируйте изображение диаграммы, используя уже знакомые кнопки выделения объектов, их перемещения, изменения формы, снятия выделения, выравнивания переменной.

В результате выполнения шагов 1–5 на Рабочем столе окна Amos Graphics должны быть контуры модели КФА1 для конфирматорного факторного анализа. При желании можно вместо книжной выбрать альбомную ориентацию диаграммы. Для этого в командной строке в меню View (Вид) следует выбрать Interface Properties (Параметры интерфейса)

и в окне Paper Size (Размер страницы) на вкладке Page Layout (Макет страницы) в раскрывающемся списке выбрать Landscape – A4 (Альбомная – формат A4), после чего нажать кнопку Apply (Применить). Для оптимального расположения диаграммы на странице можно воспользоваться кнопкой  Resize the path diagram to fit on a page (Изменить размер диаграммы под размер страницы).

На следующем шаге мы зададим имена переменных модели. Но для этого сначала необходимо задать файл данных, в котором эти переменные содержатся (IQ.sav).

ШАГ 6

1. Выберите в меню File (Файл) команду Data Files (Файлы данных) или щелкните на кнопке  панели инструментов. В открывшемся окне Data Files (Файлы данных) щелкните на кнопке File Name (Имя файла) и в окне Открыть найдите и задайте файл IQ.sav. После чего в окне Data Files (Файлы данных) нажмите OK.
2. Выберите в меню View (Вид) команду Variables in dataset (Переменные в наборе данных) или щелкните на кнопке  панели инструментов. Откроется окно со списком имен и меток переменных рабочего файла данных IQ.sav.
3. Ориентируясь на рис. 26.10, перетащите поочередно переменные из окна Variables in dataset (Переменные в наборе данных) в прямоугольные контуры для явных переменных. Для этого выделяйте переменную в списке левой клавишей мыши и, удерживая ее, перетаскивайте. После задания имен окно Variables in dataset (Переменные в наборе данных) можно закрыть щелчком на кнопке с крестиком в правом верхнем углу.
4. Подгоните контуры явных переменных под их названия. Для этого в меню Plugins (Надстройки) щелкните на пункте Resize Observed Variables (Изменить размер наблюдаемых переменных).
5. Задайте имена латентных переменных, включая ошибки. Для этого в меню Plugins (Надстройки) щелкните на пункте Name Unobserved Variables (Присвоить имена ненаблюдаемым переменным). Пустые контуры ненаблюдаемых переменных заполнятся именами, присваиваемыми программой по умолчанию: F — для латентных переменных, e — для ошибок.
6. Сохраните построенную модель в созданной ранее папке IQ под именем KFA1: File ► Save As ► Примеры_SPSS ► IQ ► (Имя файла: KFA1) ► Сохранить.

После выполнения шага 6 у вас должна быть построена модель, аналогичная изображенной на рис. 26.10, которая полностью готова для анализа.

На следующем шаге настроим параметры анализа, состав вывода и запустим процедуру оценивания модели.

ШАГ 7

1. Выберите в меню View (Вид) команду Analysis Properties (Параметры анализа) или щелкните на кнопке  панели инструментов. Одноименное окно откроется на вкладке Estimation (Оценка) (см. рис. 26.3). Поскольку в данных нет пропусков и предполагается приблизительная нормальность переменных, оставляем все параметры на этой вкладке установленными по умолчанию.
2. В окне Analysis Properties (Параметры анализа) перейдите на вкладку Output (Вывод) для настройки параметров вывода (см. рис. 26.4). Для вывода важных результатов анализа установите флажки Standardized estimates (Стандартизованные оценки), Residual moments (Ошибки моментов), Modification indices (Индексы модификации), Tests for normality and outliers (Критерии нормальности и выбросы). Закройте окно Analysis Properties (Параметры анализа).



3. Для запуска процедуры анализа и вычисления оценок выберите в меню Analyse (Анализ) команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов.

Если модель была задана без ошибок, то программа закончит вычисления, о чем будет свидетельствовать появление красной стрелки на кнопке  View the output path diagram (Обзор результирующей диаграммы путей), которая становится доступной. При щелчке на ней рядом с элементами диаграммы появляются числа — нестандартизованные оценки дисперсий, ковариаций и коэффициентов регрессии. Как правило, больший интерес представляют стандартизованные оценки, которые видны, если в окне между панелью инструментов и Рабочим столом вместо Unstandardized estimates (Нестандартизованные оценки) выделить строку Standardized estimates (Стандартизованные оценки). Однако стандартизованные оценки представляют интерес и подлежат интерпретации, если модель по критериям согласия довольно хорошо соответствует исходным данным. В одном из окон между панелью инструментов и Рабочим столом можно видеть краткое резюме результатов оценки, в конце которого приводятся значение хи-квадрат (Chi-square) и число степеней свободы (df). В данном случае они равны, соответственно, 172,3 и 41, то есть хи-квадрат более чем в 4 раза превышает число степеней свободы, что свидетельствует о плохом соответствии модели исходным данным и о необходимости коррекции модели.

Оценка априорной модели КФА

Все результаты оценки модели содержатся в окне вывода Amos Output, которое открывается после щелчка на пункте Text Output (Текст вывода) в меню View (Вид).

Проверка нормальности. Соответствующая информация находится в разделе Assessment of normality (Оценка нормальности (рис. 26.12). Критический коэффициент для многомерного эксцесса $c.r. = 1,871$ значительно меньше пяти, следовательно, предположение о многомерной нормальности переменных не нарушено: выбранный метод оценки корректен. Значения асимметрии и эксцесса для каждой переменной также невелики. Если бы критический коэффициент для многомерного эксцесса превышал 5, следовало бы принять меры, например исключить выбросы, руководствуясь разделом Observations farthest from the centroid (Mahalanobis distance) (Наиболее удаленные от центра наблюдения (Расстояния Махаланобиса)).

Оценка приемлемости объема выборки. В разделе вывода Notes for model (Примечания для модели) содержится общая информация по идентификации модели: количество оцениваемых параметров $T = 35$. Это значение можно получить исходя из количества явных переменных в модели $P = 11$ и числа степеней свободы $df = 41$. Общее число доступных моментов выборки (ковариаций и дисперсий) $P(P + 1)/2 = 66$, следовательно, $T = P(P + 1)/2 - df = 25$. Объем выборки $N = 184$ несколько меньше рекомендуемого соотношения $10T$, но больше предельно малой численности $5T$. Для оценки соответствия модели исходным данным следует применять самые строгие из рекомендуемых величин критериев согласия модели.

Критерии согласия модели, которые содержатся в разделе вывода Model Fit (Подгонка модели) (рис. 26.13), свидетельствуют о слабом соответствии модели исходным данным. Следовательно, модель нуждается в коррекции, а оценки параметров не заслуживают внимания.

Assessment of normality (Group number 1)						
Variable	min	max	skew	c.r.	kurtosis	c.r.
и6	2,000	13,000	-,943	-5,220	,898	2,488
и10	3,000	18,000	-,423	-2,344	,930	2,575
и2	,000	17,000	-,596	-3,303	,743	2,057
и5	4,418	14,861	-,049	-,270	-,652	-1,805
и3	3,000	15,000	-,115	-,636	-,010	-,027
и1	7,000	16,000	,341	1,886	-,394	-1,091
и11	2,323	16,000	-,092	-,509	,306	,848
и9	2,737	18,000	-,857	-4,744	,612	1,695
и4	5,157	17,171	-,107	-,591	,230	,638
и8	3,675	15,000	-,137	-,756	-,112	-,310
и7	3,603	18,516	,045	,247	-,275	-,762
Multivariate					4,664	1,871

Рис. 26.12. Окно вывода: раздел Assessment of normality (Оценка нормальности)

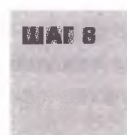
Model Fit Summary					
CMIN					
Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	25	172,268	41	,000	4,202
Saturated model	66	,000	0		
Independence model	11	706,828	55	,000	12,851

RMSEA				
Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,132	,112	,153	,000
Independence model	,254	,238	,271	,000

Рис. 26.13. Окно вывода для априорной модели КФА: фрагменты раздела Model Fit (Подгонка модели)

Коррекция модели КФА

Следует отметить одно неудобство, которое часто возникает при оценке согласия модели и ее коррекции. На диаграмме приведены метки переменных, что, конечно, делает модель наглядной и упрощает ее интерпретацию. Но в выводе фигурируют только имена переменных. И довольно затруднительно соотносить содержание вывода и диаграмму. Для устранения этого неудобства на время коррекции модели целесообразно заменить на диаграмме метки переменных их именами.



1. Выберите в меню View (Вид) команду Interface Properties (Параметры интерфейса).
2. Перейдите на вкладку Misc (Разное) и снимите флажок Display variables labels (Показывать метки переменных).

После выполнения шага 8 метки переменных будут заменены их именами. После коррекции модели можно будет вновь установить упомянутый флажок.

Возможны несколько направлений коррекции модели: 1) фиксирование нулевых значений для статистически недостоверных параметров (удаление «лишних» стрелок); 2) удаление «лишних» переменных, статистически недостоверно связанных с остальными переменными или особенно сильно нарушающих согласие модели; 3) освобождение фиксированных нулевых параметров, повышающих согласие модели (добавление новых стрелок); 4) добавление в модель новых переменных, например вторичных факторов.

Конечно, первое и второе направления предпочтительнее прочих, так как они позволяют получить более лаконичную модель. Информация о статистической достоверности параметров содержится в разделе вывода Estimates (Оценки). Регрессионные коэффициенты и их статистическая значимость приведены на рис. 26.14. Все оцененные регрессионные веса статистически достоверны (три звездочки означают статистическую достоверность $p < 0,001$). Статистическая значимость коэффициентов для параметров, фиксированных как 1, не оценивается, но в данном случае она не вызывает сомнений, так как стандартизованные коэффициенты для них довольно велики. Таким образом, статистическая значимость коэффициентов не дает оснований для удаления каких-либо переменных из модели.

Regression Weights: (Group number 1 - Default model)					
		Estimate	S.E.	C.R.	P
и7 <---	F1	1,000			
и8 <---	F1	,784	,090	8,689	***
и4 <---	F1	,735	,080	9,167	***
и9 <---	F1	,860	,112	7,668	***
и11 <---	F2	1,000			
и1 <---	F2	,875	,160	5,475	***
и3 <---	F2	1,262	,222	5,695	***
и5 <---	F3	1,000			
и2 <---	F3	1,698	,242	7,005	***
и10 <---	F3	1,424	,214	6,649	***
и6 <---	F3	,906	,176	5,151	***

Рис. 26.14. Окно вывода для априорной модели, фрагмент раздела Estimates (Оценки): Regression weights (Регрессионные коэффициенты)

Если вас интересует статистическая значимость параметров, фиксированных как 1, то следует освободить соответствующий параметр в модели, зафиксировав как 1 альтернативный параметр. Например, для определения статистической значимости параметра и4←F1 следует перейти в режим редактирования модели, щелкнув на кнопке . Затем дважды щелкнуть на стрелке, соединяющей F1 и и4 (счет в уме), и в открывшемся окне (см. рис. 26.11) на вкладке Object Properties: Parameters (Свойства объекта: Параметры) в поле Regression weight (Регрессионный коэффициент) убрать 1 и после этого закрыть окно. Затем двойным щелчком на соседней стрелке, например, от F1 к переменной аналогии, снова открыть это окно и зафиксировать значение 1 для этого параметра. После чего снова запустить анализ и открыть окно вывода на вкладке Estimates (Оценки).

В конце раздела Estimates (Оценки) находится таблица Standardized Residual Covariances (Стандартизованные остатки ковариаций) (рис. 26.15). По ней можно судить о переменных, ковариации между которыми плохо воспроизводятся моделью. Экстремальное расхождение эмпирических и модельных ковариаций наблюдается только для одной пары переменных: и10 и и11 (–3,130). Это не дает оснований для уверенного удаления одной из них из модели. Таким образом, нет достаточных оснований для удаления переменных из модели или фиксирования дополнительных нулевых параметров (удаления «лишних» стрелок), и остается 2 пути: освобождение фиксированных параметров (добавление стрелок) и (или) добавление вторичного фактора.

Standardized Residual Covariances (Group number 1 - Default model)											
	и6	и10	и2	и5	и3	и1	и11	и9	и4	и8	и7
и6	,000										
и10	-1,591	,000									
и2	,428	,367	,000								
и5	,443	,262	-,377	,000							
и3	2,280	-,694	,683	,363	,000						
и1	2,057	-2,208	,863	-,680	-,501	,000					
и11	2,477	-3,130	-,665	,812	-,121	,991	,000				
и9	,010	1,296	,041	,067	1,758	-1,659	-1,491	,000			
и4	-1,472	2,548	-1,157	1,182	,228	-1,597	-2,744	-,962	,000		
и8	,405	,373	-,158	-1,152	1,882	,362	,538	-,413	,697	,000	
и7	-,471	-,361	-,530	-,651	1,329	,856	-1,300	1,516	-,044	-,754	,000

Рис. 26.15. Окно вывода для априорной модели. фрагмент раздела Estimates: Standardized Residual Covariances (Оценки: Стандартизованные остатки ковариаций)

Информация для освобождения фиксированных параметров содержится в разделе вывода Modification Indices (Индексы модификации), состоящего из трех таблиц: Covariances (Ковариации), Variances (Дисперсии), Regression Weights (Регрессионные коэффициенты). Таблицы Covariances (Ковариации) (рис. 26.16) и Regression Weights (Регрессионные коэффициенты) содержат значения индексов модификации (M.I.) для ковариаций и регрессий, которые определяют, насколько уменьшится значение хи-квадрат при освобождении данного параметра (добавлении соответствующей связи). Ориентируясь на эти величины, можно последовательно добавлять связи в модель по одной, каждый раз получая уменьшенное значение хи-квадрат и новые значения индексов модификации.

Предположим, мы хотим модифицировать модель таким образом, чтобы избежать пересечения латентных конструкторов, которым соответствуют факторы, то есть добавления связей каждого фактора с переменными, которые являются индикаторами других факторов. Тогда нам следует отказаться от индексов модификации, содержащихся в таблице Regression Weights (Регрессионные коэффициенты), и учитывать только индексы, находящиеся в таблице Covariances (Ковариации), для связывания ошибок (см. рис. 26.16). При этом связи между факторами и ошибками игнорируются как теоретически не обоснованные.

На следующем шаге мы последовательно будем добавлять связи между ошибками до тех пор, пока в таблице Covariances (Ковариации) не останется ни одной пары ошибок.

Covariances (Group number 1 - Default model)			
		M.I.	Par Change
e11 <--> F2		15,195	1,000
e10 <--> F2		16,448	-1,040
e10 <--> F1		8,292	,716
e10 <--> e11		8,261	-,997
e7 <--> F1		6,402	,607
e6 <--> e10		4,675	-,660
e6 <--> e9		4,375	,648
e5 <--> F1		5,183	-,526
e5 <--> e11		6,467	,818
e5 <--> e10		10,874	-1,062
e5 <--> e8		5,456	,584
e4 <--> e7		5,468	,888
e4 <--> e6		5,774	-,830
e1 <--> F2		10,099	-,563
e1 <--> e11		7,594	-,663
e1 <--> e10		22,707	1,148
e1 <--> e9		9,391	-,749
e1 <--> e8		8,903	,558
e1 <--> e5		6,448	-,564
e1 <--> e4		7,420	-,727
e3 <--> F2		6,684	,553
e3 <--> e8		6,432	-,573
e3 <--> e5		5,063	,603
e3 <--> e1		5,581	,455
e2 <--> e10		4,174	-,759
e2 <--> e6		5,278	,744
e2 <--> e4		13,389	1,518
e2 <--> e3		4,774	-,658

Рис. 26.16. Окно вывода для априорной модели, фрагмент раздела Modification Indices: Covariances (Индексы модификации: Ковариации)

ШАГ 8

1. Щелкните на кнопке , чтобы перейти в режим редактирования диаграммы. В таблице Covariances (Ковариации) найдите связь с максимальным M.I. (см. рис. 26.16). Для исходной модели это связь e1–e10.
2. На панели инструментов щелкните на кнопке с двусторонней стрелкой и на диаграмме свяжите ковариацией (двусторонней стрелкой) ошибки e1 и e10. Запустите вычисление оценок (Calculate estimates) из меню Analyse (Анализ) или при помощи соответствующей кнопки на панели инструментов.
3. В меню View (Вид) откройте Text Output (Текст вывода) и перейдите в раздел Modification Indices (Индексы модификации).
4. В таблице Covariances (Ковариации) найдите пару ошибок с максимальным M.I.
5. На панели инструментов щелкните на кнопке с двусторонней стрелкой и на диаграмме свяжите ковариацией (двусторонней стрелкой) соответствующие ошибки. Запустите вычисление оценок (Calculate estimates) из меню Analyse (Анализ) или при помощи соответствующей кнопки на панели инструментов.

6. Повторяйте пункты 3–5 этого шага до тех пор, пока в таблице Covariances (Ковариации) не останется ни одной пары ошибок. В процессе добавления связей между ошибками предшествующие связи могут терять свою статистическую значимость. Тогда их удаляют и процедура продолжается. После добавления каждой связи необходимо контролировать статистическую значимость дисперсий. Если какая-то дисперсия потеряла статистическую значимость при добавлении связи, эту связь удаляют. Если это последняя из пар ошибок в таблице, удаляют и предыдущую из добавленных связей, а вместо них выбирают третью. Процесс продолжается до тех пор, пока в таблице Covariances (Ковариации) не останется ни одной пары ошибок, а все регрессии, ковариации и дисперсии сохраняют статистическую значимость.
7. Отсутствие в таблице Covariances (Ковариации) пар ошибок означает, что достигнуто максимально возможное соответствие модели исходным данным.
8. Перейдите в режим просмотра итоговой диаграммы, щелкнув на кнопке , и отредактируйте диаграмму. Воспользуйтесь «волшебной палочкой» — кнопкой  Touch up a variable (Выравнивание переменной) панели инструментов для выравнивания стрелок: щелкните ею на каждой из ошибок. Щелкните на кнопке  Resize the path diagram to fit on a page (Изменить размер диаграммы под размер страницы).
9. Переместите числовые показатели так, чтобы они были более наглядными. Для этого наведите указатель на стрелку или контур переменной и щелкните правой клавишей мыши. В открывшемся контекстном меню выберите Move Parameter (Перемещать параметр). Рядом с указателем появится надпись move. Теперь, наведя указатель на любой числовой показатель, вы можете его перетаскивать в области соответствующих элементов диаграммы.
10. Сохраните модель под именем КФА2 в папке IQ для дальнейшего анализа ее оценок.

В результате выполнения шага 8 получена апостериорная модель КФА, подобная изображенной на рис. 26.20, оценки которой мы обсудим позже в разделе «Конфирматорный факторный анализ».

Коррекция модели КФА с вторичным фактором

На следующем шаге мы создадим альтернативную априорную модель в предположении наличия вторичного фактора интеллекта (рис. 26.17) и откорректируем ее для наи-

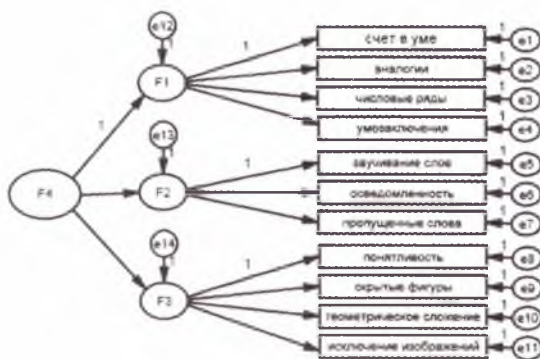


Рис. 26.17. Модель КФА с вторичным фактором

лучшего соответствия исходным данным. Здесь важно следующее: первичный фактор должен включать в себя не менее двух индикаторов, а вторичный — не менее трех индикаторов.



1. В меню File (Файл) выберите Open (Открыть) и откройте исходную априорную модель КФА1. Щелкните на кнопке , чтобы перейти в режим редактирования диаграммы. Удалите ковариации (двунаправленные стрелки), связывающие факторы.
2. Щелкните на панели инструментов на кнопке с овалом и нарисуйте латентную переменную, соответствующую вторичному фактору F4 (см. рис. 26.17). Проведите однонаправленные стрелки от F4 к F1, F2, F3.
3. При помощи кнопки  добавьте ошибки к каждому первичному фактору. Присвойте имена латентным переменным: в меню Plugins (Надстройки) щелкните на пункте Name Unobserved Variables (Присвоить имена ненаблюдаемым переменным).
4. Наведите указатель на стрелку, идущую от фактора F4 к фактору F1 (она станет красной), и сделайте двойной щелчок правой клавишей мыши. В открывшемся диалоговом окне Object Properties (Параметры объекта) перейдите на вкладку Parameters (Параметры) (см. рис. 26.11). В поле Regression weight (Регрессионный коэффициент) введите 1. Закрыв окно Object Properties, вы увидите 1 возле стрелки, что означает фиксирование свободного параметра на 1.
5. Для запуска процедуры анализа и вычисления оценок выберите в меню Analyse (Анализ) команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов.
6. В меню View (Вид) откройте Text Output (Текст вывода) и перейдите в раздел Modification Indices (Индексы модификации).
7. В таблице Covariances (Ковариации) найдите пару ошибок с максимальным M.I.
8. На панели инструментов щелкните на кнопке с двусторонней стрелкой и на диаграмме свяжите ковариацией (двусторонней стрелкой) соответствующие ошибки. Запустите вычисление оценок (Calculate estimates) из меню Analyse (Анализ) или при помощи соответствующей кнопки на панели инструментов.
9. Повторяйте пп. 5–7 этого шага до тех пор, пока в таблице Covariances (Ковариации) не останется ни одной пары ошибок. В процессе добавления связей между ошибками предшествующие связи могут терять свою статистическую значимость. Тогда их удаляют и процедура продолжается. После добавления каждой связи необходимо контролировать статистическую значимость дисперсий. Если какая-то дисперсия потеряла статистическую значимость при добавлении связи, эту связь удаляют. Если это последняя из пар ошибок в таблице, удаляют и предыдущую из добавленных связей, а вместо них выбирают третью. Процесс продолжается до тех пор, пока в таблице Covariances (Ковариации) не останется ни одной пары ошибок, а все регрессии, ковариации и дисперсии сохраняют статистическую значимость.

10. Отсутствие в таблице Covariances (Ковариации) пар ошибок означает, что достигнуто максимально возможное соответствие модели исходным данным.
11. Перейдите в режим просмотра итоговой диаграммы и отредактируйте ее (см. шаг 8, пп. 8 и 9). Воспользуйтесь кнопкой Touch up a variable (Выравнивание переменной) панели инструментов для выравнивания стрелок: щелкните ею на каждой из ошибок. Щелкните на кнопке Resize the path diagram to fit on a page (Изменить размер диаграммы под размер страницы). Переместите числовые показатели так, чтобы они были более наглядными, выбрав в контекстном меню Move Parameter (Перемещать параметр).
12. Сохраните модель под именем КФА3 в папке IQ для дальнейшего анализа ее оценок.

В результате выполнения шага 9 получена апостериорная модель КФА с вторичным фактором, подобная изображенной на рис. 26.22, оценки которой мы обсудим позже в разделе «Конфирматорный факторный анализ».

Коррекция модели КФА с пересекающимися факторами

Предположим, исследователь допускает пересечение конструкторов, то есть допускает, что некоторые факторы имеют индикаторы (явные переменные), общие с другими факторами. Обычно если исследователь и допускает пересечения конструкторов, то он стремится к тому, чтобы их минимизировать. Поэтому, добавляя связи факторов с индикаторами других факторов, он ограничивается их минимальным количеством, обеспечивающим приемлемое согласие модели данным. На следующем шаге мы будем руководствоваться этими соображениями, корректируя модель, изображенную на рис. 26.10, с допущением пересечения факторов.

ШАГ 10

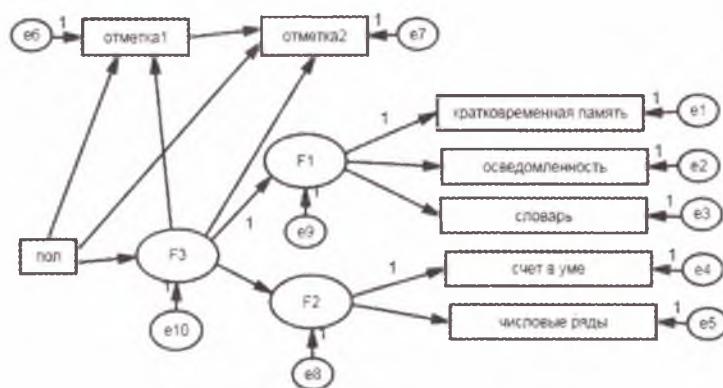
1. В меню File (Файл) выберите Open (Открыть) и откройте апостериорную модель КФА2 (результат выполнения шага 8).
2. Для запуска процедуры анализа и вычисления оценок в меню Analyse (Анализ) выберите команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов.
3. В меню View (Вид) откройте Text Output (Текст вывода), перейдите в раздел Model Fit (Подгонка модели) и убедитесь, что значимость хи-квадрат заметно меньше 0,05, а RMSEA превышает 0,05.
4. В главном окне щелкните на кнопке , чтобы перейти в режим редактирования диаграммы, и добавьте направленные связи (односторонние стрелки) от F1 к и11 (заучивание слов) и от F2 к и6 (исключение изображений).
5. Для запуска процедуры анализа и вычисления оценок в меню Analyse (Анализ) выберите команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов.
6. В меню View (Вид) откройте Text Output (Текст вывода) и перейдите в раздел Estimates (Оценки). Проверьте статистическую значимость параметров. Убедитесь, что ковариации пар ошибок e5–e11 и e10–e11 статистически не достоверны.
7. Перейдите к главному окну и удалите связи ошибок e5– e11 и e10–e11.

8. Запустите процедуру анализа и вычисления оценок. В меню View (Вид) откройте Text Output (Текст вывода) и перейдите в раздел Estimates (Оценки). Убедитесь, что все параметры статистически достоверны. Перейдите в раздел Modification Indices (Индексы модификации): таблицы индексов модификации пусты, возможности улучшения модели исчерпаны. Перейдите в раздел Estimates (Оценки): статистическая значимость хи-квадрат $p = 0,05$, RMSEA = 0,05. Таким образом, достигнуто вполне приемлемое согласие модели.
9. На Рабочем столе отредактируйте модель, как описано в конце шагов 8 и 9. Сохраните модель под именем КФА4 в папке IQ для дальнейшего анализа ее оценок.

В результате выполнения шага 10 у вас должна получиться диаграмма, подобная изображенной на рис. 26.23.

Структурная модель

Для построения структурной модели воспользуемся данными файла ex01.sav, к которому мы неоднократно обращались в первой части книги. Предположим, на основе теоретических представлений и предшествующих результатов исследования была сформирована априорная (теоретическая) модель факторов успешности обучения школьников (рис. 26.18). Модель включает в себя измерительную составляющую — интеллект, представленный двумя первичными и одним вторичным факторами, объясняющими связь пяти его индикаторов (явных переменных). Структурная составляющая модели предполагает влияние общего интеллекта и пола на успешность обучения как в 10-м (отметка1), так и в 11-м (отметка2) классах. Эта модель сохранена в наборе файлов примеров под именем CM.amw.



Хи-квадрат= χ^2 , df=степени свободы, p=вероятность, RMSEA=корень из среднего квадрата остатков

Рис. 26.18. Априорная структурная модель факторов успешности обучения

Как отмечалось ранее, первичный фактор должен включать в себя не менее двух индикаторов, а вторичный — не менее трех индикаторов. В рассматриваемой модели вторичный фактор F3 включает в себя в качестве индикаторов всего 2 фактора интеллекта (F1

и F2). Но его индикаторами являются также переменные *отметка1* и *отметка2*. Поэтому вторичный фактор в модели вполне состоятелен.

Построение и проверка подобных структурных моделей обычно начинается с измерительной ее составляющей. После проверки и коррекции последней к ней добавляется структурная составляющая модели.

Обратите внимание на текст под моделью. Его наличие позволяет получать основные индикаторы соответствия модели без обращения к тексту вывода, что ускоряет процесс коррекции модели.

До применения AMOS нужно выполнить два подготовительных шага. Эти шаги позволяют запустить программу IBM SPSS Amos 20.

На шаге 3а мы построим измерительную модель факторов интеллекта, включающую в себя первичные факторы F1 и F2 (см. рис. 26.18).

Шаг 3а

1. Щелкните на кнопке  Draw a latent variable or add an indicator to a latent variable (Рисовать латентную переменную или добавить индикатор для латентной переменной). Наведите указатель на центр Рабочего стола и щелкните левой клавишей мыши: появится овал, соответствующий латентной переменной (фактору). Каждый последующий щелчок на этом овале будет добавлять один индикатор и его ошибку для данного фактора. Щелкните 3 раза, чтобы добавить 3 индикатора (явные переменные).
2. Щелкните на кнопке  Rotate the indicators of a latent variables (Поворот индикаторов латентных переменных). Наведите указатель на новую латентную переменную и щелкните на ней один раз: ее индикаторы развернутся на 90° по часовой стрелке.
3. Выделите все объекты, щелкнув на кнопке  Выбрать все объекты. Щелкните на кнопке  Дублировать объект, наведите указатель на латентную переменную и перетащите копию вниз для создания контура для фактора F2 и его индикаторов и снимите выделение кнопкой . При помощи кнопки  сотрите лишний индикатор для фактора F2 и его ошибку.
4. Выберите в меню File (Файл) команду Data Files (Файлы данных) или щелкните на соответствующей кнопке  панели инструментов. В открывшемся окне Data Files (Файлы данных) щелкните на кнопке File Name (Имя файла) и в окне Открыть найдите и задайте файл ex01.sav. После чего в окне Data Files (Файлы данных) нажмите OK.
5. Выберите в меню View (Вид) команду Variables in dataset (Переменные в наборе данных) или щелкните на соответствующей кнопке  панели инструментов. Откроется окно со списком имен и меток переменных рабочего файла данных ex01.sav.
6. Ориентируясь на рис. 26.18, перетащите поочередно переменные из окна Variables in dataset (Переменные в наборе данных) в соответствующие прямоугольные контуры явных переменных. Для этого щелчком мыши выделяйте переменную в списке и удерживайте ее при перетаскивании левой клавишей мыши. После задания имен окно Variables in dataset (Переменные в наборе данных) можно закрыть.

7. Выровняйте контуры явных переменных в соответствии с их названиями. Для этого в меню Plugins (Надстройки) щелкните на пункте Resize Observed Variables (Изменить размер наблюдаемых переменных).
8. Задайте имена латентных переменных, включая ошибки. Для этого в меню Plugins (Надстройки) щелкните на пункте Name Unobserved Variables (Присвоить имена ненаблюдаемым переменным). Пустые контуры ненаблюдаемых переменных будут заполнены именами, присваиваемыми программой по умолчанию: F — для латентных переменных, e — для ошибок.
9. Откорректируйте контуры измерительной модели, используя уже знакомые вам кнопки выделения объектов , их перемещения , изменения их формы , снятия выделения , выравнивания переменной .
10. Соедините факторы F1 и F2 ковариацией, воспользовавшись кнопкой .
11. Оформите надпись под диаграммой. Для этого щелкните на кнопке Title – Figure caption (Название – Конфигурация надписи) на панели инструментов, а затем щелкните на пустом месте Рабочего стола под диаграммой. Появится меню Figure caption (рис. 26.19). В поле Caption (Надпись) введите текст
 $\chi^2 = \chi^2$; df = df; p = p; RMSEA = RMSEA
12. Нажмите OK. Под диаграммой появится надпись, как на рис. 26.19.
13. Сохраните построенную модель в предварительно созданной папке CM под именем CM1: File ► Save As ► Примеры_SPSS ► CM ► (Имя файла: CM1) ► Сохранить.

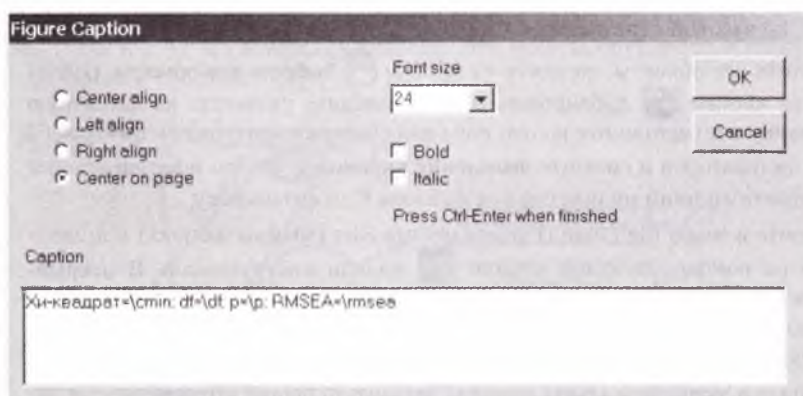


Рис. 26.19. Окно Figure caption (Конфигурация надписи)

В результате выполнения шага 3а создана измерительная составляющая модели интеллекта для проверки ее состоятельности. Надпись под ней позволит быстро получать информацию о соответствии модели данным в процессе анализа без обращения к тексту вывода.

На следующем шаге мы зададим параметры анализа и произведем вычисления, необходимые для проверки состоятельности этой измерительной модели.

ШАГ 4а

1. Выберите в меню View (Вид) команду Analysis Properties (Параметры анализа) или щелкните на соответствующей кнопке  панели инструментов. Откроется одноименное окно на вкладке Estimation (Оценка) (см. рис. 26.3). Поскольку в данных нет пропусков и предполагается приблизительная нормальность переменных, оставляем все параметры на этой вкладке установленными по умолчанию.
2. В окне Analysis Properties (Параметры анализа) перейдите на вкладку Output (Вывод) для настройки вывода (см. рис. 26.4). Для вывода важных результатов анализа установите флажки Standardized estimates (Стандартизованные оценки), Residual moments (Ошибки моментов), Modification indices (Индексы модификации), Tests for normality and outliers (Критерии нормальности и выбросы). Закройте окно Analysis Properties (Параметры анализа).
3. Для запуска процедуры анализа и вычисления оценок выберите в меню Analyse (Анализ) команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов.
4. Для облегчения анализа параметров модели замените метки переменных на диаграмме их именами. Для этого выберите в меню View (Вид) команду Interface Properties (Параметры интерфейса). Перейдите на вкладку Misc (Разное) и снимите флажок Display variables labels (Показывать метки переменных). В дальнейшем можно в любой момент заменить имена переменных их метками, вновь установив этот флажок.

Если модель была задана без ошибок, то программа закончит вычисления, о чем будет свидетельствовать появление красной стрелки на кнопке  View the output path diagram (Обзор результирующей диаграммы путей), которая становится доступной. При щелчке на ней рядом с элементами диаграммы появляются оценки ковариаций и коэффициентов регрессии, а под диаграммой — основные показатели согласия модели. Слишком высокое значение RMSEA свидетельствует о недостаточном соответствии модели исходным данным. Обзор оценок параметров в тексте вывода (View ► Text Output ► Estimates) свидетельствует о том, что причина заключается в слишком «слабом» факторе F2: индикаторов для этого фактора всего два, и один из них связан с фактором статистически не достоверно. Слабая корреляция факторов F1 и F2 является весомым аргументом против состоятельности вторичного фактора интеллекта.

Таким образом, наиболее целесообразным на следующем шаге было бы исключение из модели фактора F2 и, соответственно, вторичного фактора F3. Однако есть слабая надежда на то, что вторичный фактор интеллекта окажется состоятельным за счет включения в качестве его индикаторов переменных отметка1 и отметка2.

На следующем шаге мы добавим в измерительную модель вторичный фактор F3 и переменные отметка1 и отметка2 в качестве его индикаторов и проверим состоятельность измерительной модели с вторичным фактором. В качестве исходной будем использовать модель, построенную на шаге 4а.

ШАГ 5а

1. Щелкните на кнопке , чтобы перейти в режим редактирования диаграммы. Удалите ковариацию (двунаправленную стрелку), связывающую факторы.
2. Щелкните на панели инструментов на кнопке с овалом и нарисуйте латентную переменную, соответствующую вторичному фактору F3 (см. рис. 26.18). Проведите однонаправленные стрелки от F3 к F1 и F2.

3. Щелкните на панели инструментов на кнопке с прямоугольником и нарисуйте контуры для переменных *отметка1* и *отметка2*. Проведите однонаправленные стрелки от *F3* к этим переменным.
4. При помощи кнопки  добавьте ошибки к каждому первичному фактору, а также к переменным *отметка1* и *отметка2*. Добавляя ошибку, делайте дополнительные щелчки для оптимальной ориентации ошибок.
5. Присвойте имена латентным переменным: в меню *Plugins* (Настройки) щелкните на пункте *Name Unobserved Variables* (Присвоить имена ненаблюдаемым переменным).
6. Наведите указатель на стрелку от фактора *F3* к фактору *F1* (она станет красной) и сделайте двойной щелчок правой клавишей мыши. В открывшемся диалоговом окне *Object Properties* (Параметры объекта) перейдите на вкладку *Parameters* (Параметры) (см. рис. 26.11). В поле *Regression weight* (Регрессионный коэффициент) введите 1. Закрыв окно *Object Properties* (Параметры объекта), вы увидите 1 возле стрелки, что означает фиксирование свободного параметра на 1.
7. Для запуска процедуры анализа и вычисления оценок в меню *Analyse* (Анализ) выберите команду *Calculate Estimates* (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов. Перейдите к обзору результирующей диаграммы.

Показатель RMSEA снова оказывается слишком велик (0,083), а обзор оценок параметров в тексте вывода показывает значительное число статистически недостоверных параметров, связанных с фактором *F2*. Таким образом, надежда на состоятельность этого фактора не оправдалась и его следует удалить из модели вместе с вторичным фактором. Слишком слабая корреляция *F1* и *F2* свидетельствует против предположения о том, что индикаторы фактора *F2* могут служить индикаторами общего фактора интеллекта, включающего в себя все 5 индикаторов. Тем не менее проверим это предположение.

На следующем шаге мы проверим однофакторную измерительную модель в предположении, что фактор интеллекта содержит 5 индикаторов (тест1–тест5) и влияет на успеваемость (*отметка1* и *отметка2*). В качестве исходной модели будем использовать результат выполнения шага 5а.

ШАГ 6а

1. Щелкните на кнопке  , чтобы перейти в режим редактирования диаграммы. Удалите латентные переменные *F2*, *F3*, *e8*, *e9*.
2. Добавьте стрелки, направленные от *F1* к явным переменным *тест1*, *тест2*, *отметка1* и *отметка2*.
3. Для запуска процедуры анализа и вычисления оценок выберите в меню *Analyse* (Анализ) команду *Calculate Estimates* (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов. Перейдите к обзору результирующей диаграммы.

Статистическая значимость хи-квадрат и слишком большое значение RMSEA свидетельствуют о плохом согласии модели. Обзор оценок в тексте вывода показывает статистически недостоверные связи фактора *F1* с переменными *тест1* и *тест2*. Следовательно, эти переменные не могут являться индикаторами фактора *F1*. Регрессионные коэффициенты для остальных 4 переменных статистически достоверны, следовательно их целесообразно оставить в модели.

На следующем шаге мы удалим переменные тест1 и тест2 и добавим переменную пол, связав ее с остальными переменными в соответствии с априорной моделью. В качестве исходной модели будем использовать результат выполнения шага 5а.

ШАГ 7а

1. Щелкните на кнопке , чтобы перейти в режим редактирования диаграммы. Удалите переменные тест1 и тест2.
2. Щелкните на кнопке с прямоугольником и добавьте прямоугольный контур для явной переменной пол. Откройте список переменных в наборе данных при помощи кнопки  и перетащите оттуда переменную пол в ее контур на диаграмме.
3. Добавьте стрелки, направленные от переменной пол к переменным отметка1, отметка2 и F1.
4. В связи с тем, что теперь переменная F1 стала эндогенной (зависимой), добавьте к ней ошибку при помощи кнопки  и ориентируйте ее вниз, щелкнув несколько раз внутри контура F1. Присвойте имя новой ошибке: в меню Plugins (Надстройки) щелкните на пункте Name Unobserved Variables (Присвоить имена ненаблюдаемым переменным).
5. Для запуска процедуры анализа и вычисления оценок выберите в меню Analyse (Анализ) команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке  панели инструментов. Перейдите к обзору результирующей диаграммы.

Модель плохо согласована. Для определения путей ее коррекции следует обратиться, во-первых, к статистической значимости коэффициентов (View ► Text Output ► Estimates), во-вторых, к индексам модификации (View ► Text Output ► Modification Indices). Коэффициенты регрессии пол–отметка1 и пол–отметка2 статистически недостоверны, следовательно, их можно удалить (фиксировать как равные 0). В соответствии с M.I. наиболее существенное улучшение согласия модели будет получено при освобождении параметра регрессии (добавления связи в модель) отметка1–отметка2. На следующем шаге внесем эти изменения в модель.

ШАГ 8а

1. Щелкните на кнопке , чтобы перейти в режим редактирования диаграммы.
2. Удалите стрелки от переменной пол к переменным отметка1 и отметка2. Добавьте стрелку, направленную от переменной отметка1 к переменной отметка2.
3. Для запуска процедуры анализа и вычисления оценок выберите в меню Analyse (Анализ) команду Calculate Estimates (Вычислить оценки) или щелкните на соответствующей кнопке панели инструментов. Перейдите к обзору результирующей диаграммы.

Модель недостаточно хорошо согласована ($RMSEA > 0,05$), поэтому следует обратиться к оцененным параметрам и индексам модификации. Регрессионный коэффициент от F1 к переменной отметка2 статистически недостоверен, следовательно, его можно удалить (фиксировать как равный 0).

ШАГ 9а

1. Перейдите в режим редактирования диаграммы и удалите стрелку от F1 к отметка2.
2. Запустите процедуру анализа и перейдите к обзору результирующей диаграммы.

Показатели согласия модели не улучшились, поэтому снова обратитесь к оцененным параметрам и индексам модификации. Все параметры статистически достоверны, а наибольшее улучшение согласия модели по значениям M.I. следует ожидать при освобождении параметра `тест5--отметка2` (добавлении соответствующей стрелки).

ШАГ 10а

1. Перейдите в режим редактирования диаграммы и добавьте стрелку от `тест5` к `отметка2`.
2. Запустите процедуру анализа и перейдите к обзору результирующей диаграммы.

Теперь все показатели согласия модели находятся в допустимых пределах, а все ее параметры статистически достоверны. Следовательно, можно считать результат коррекции модели окончательным, а все параметры модели доступными для интерпретации.

В заключение следует окончательно отредактировать модель, добавить дополнительные оценки (квадраты множественной корреляции, прямые, не прямые и общие эффекты) и сохранить модель для дальнейшего анализа результатов.

ШАГ 11а

1. Перейдите в режим редактирования диаграммы и отредактируйте ее, используя возможности панели инструментов.
2. Откройте меню параметров анализа (`View ► Analysis Properties`) на вкладке `Output` (Вывод) и установите флажки `Squared multiple correlations` (Квадраты множественной корреляции) и `Indirect, direct & total effects` (Косвенные, прямые и общие эффекты).
3. Запустите процедуру анализа и перейдите к обзору результирующей диаграммы.
4. Сориентируйте лист построения диаграммы оптимальным образом и верните метки переменных. Для этого воспользуйтесь меню `View ► Interface Properties` (Вид ► Параметры интерфейса): на вкладке `Page Layout` установите `Paper size` (Размер бумаги) `Landscape` (Альбомный), а на вкладке `Misc` установите флажок `Display variables labels` (Выводить метки переменных). Щелкните на кнопке `Apply` (Применить) и закройте окно `Interface Properties` (Параметры интерфейса). Щелкните на кнопке , чтобы изменить размер диаграммы в соответствии с размером и ориентацией листа.
5. Отредактируйте диаграмму (см. шаг 8, пп. 8 и 9). Воспользуйтесь кнопкой `Touch up a variable` (Выравнивание переменной) панели инструментов для выравнивания стрелок: щелкните ею на каждой из переменных. Щелкните на кнопке `Resize the path diagram to fit on a page` (Изменить размер диаграммы под размер страницы). Переместите числовые показатели так, чтобы они были более наглядными, выбрав в контекстном меню `Move Parameter` (Перемещать параметр).
6. Сохраните готовую модель: `File ► Save`.

В результате выполнения шага 11а у вас должна получиться диаграмма, аналогичная изображенной на рис. 26.24.

Представление результатов

Конфирматорный факторный анализ

В результате выполнения шага 8 по коррекции исходной априорной (теоретической) модели (см. рис. 26.10) получена апостериорная модель КФА (рис. 26.20). На рис. 26.21 приведены основные критерии согласия этой модели из раздела вывода Model Fit (Подгонка модели).

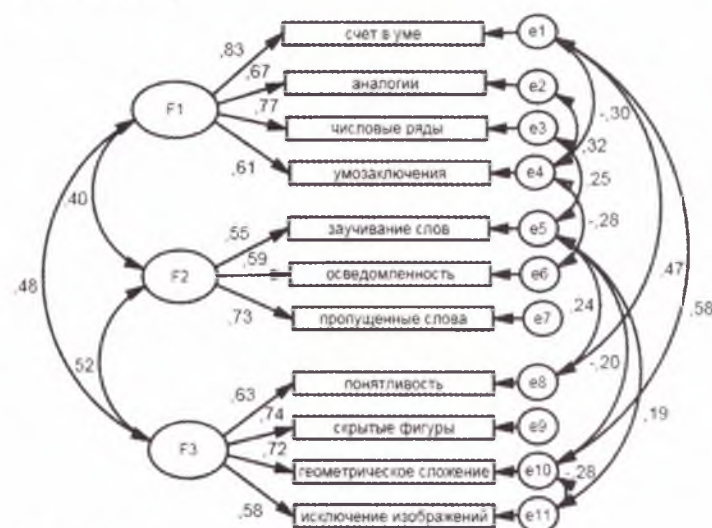


Рис. 26.20. Апостериорная модель КФА

Модель является рекурсивной. Критический коэффициент для многомерного эксцесса с.г. = 1,871 — значительно меньше пяти, следовательно, предположение о многомерной нормальности переменных не нарушено: выбранный для оценки метод максимального правдоподобия корректен.

Количество явных переменных модели $P = 11$, а число степеней свободы $df = 31$. Следовательно, число оцениваемых параметров $T = 66 - 31 = 35$ ($P(P + 1)/2 = 66$, где P — число явных переменных в модели) на 10 больше, чем для априорной модели, в соответствии с 10 добавленными связями (освобожденными для оценки параметрами) между ошибками индикаторов. Соотношение численности выборки ($N = 184$) к числу оцениваемых параметров $N/T = 5,26$ чуть больше предельно малой величины. Следовательно, для принятия решения о согласии модели с данными должны выполняться самые строгие пороги для значений критериев согласия.

Несмотря на то что GFI и CFI превышают 0,95, а соотношение $CMIN/df < 2$, статистическая значимость хи-квадрат ($p < 0,05$) и $RMSEA > 0,05$ не позволяют признать модель эмпирически подтвержденной. Если бы при тех же значениях критериев согласия объем выборки превышал 350 ($N/T > 10$), то модель можно было бы признать соответствующей данным.

На рис. 26.22 приведена апостериорная модель КФА с вторичным фактором после ее коррекции. Основные критерии согласия модели из раздела окна вывода Model Fit (Подгонка модели): хи-квадрат = 55,546; $df = 31$; $p = 0,004$; $CMIN/df = 1,792$; GFI = 0,950;

CFI = 0,962; RMSEA = 0,066. Таким образом, число оцениваемых параметров, соотношение N/T и значения критериев согласия для этой модели идентичны тем же параметрам предшествующей апостериорной модели. Следовательно, и вывод в отношении этой модели тот же: модель эмпирически не подтверждается.

На рис. 26.23 изображена апостериорная модель, построенная в предположении о пересечении конструкторов (факторов) в результате выполнения шага 10.

Число оцениваемых параметров модели $T = 66 - 31 = 35$ ($P(P + 1)/2 = 66$, где P — число явных переменных в модели) такое же, как и для предыдущих моделей, за счет добавления к априорной модели двух свободных связей между факторами и индикаторами и восьми свободных связей для ошибок. Следовательно, для принятия решения о согласии модели с данными также должны выполняться строгие ограничения для значений критериев согласия.

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	35	55,546	31	,004	1,792
Saturated model	66	,000	0		
Independence model	11	706,828	55	,000	12,851

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	,435	,950	,894	,446
Saturated model	,000	1,000		
Independence model	1,883	,493	,391	,411

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,921	,861	,964	,933	,962
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,564	,519	,542
Saturated model	,000	,000	,000
Independence model	1,000	,000	,000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,066	,036	,093	,168
Independence model	,254	,238	,271	,000

Рис. 26.21. Окно вывода для апостериорной модели КФА: фрагменты раздела Model Fit (Подгонка модели)

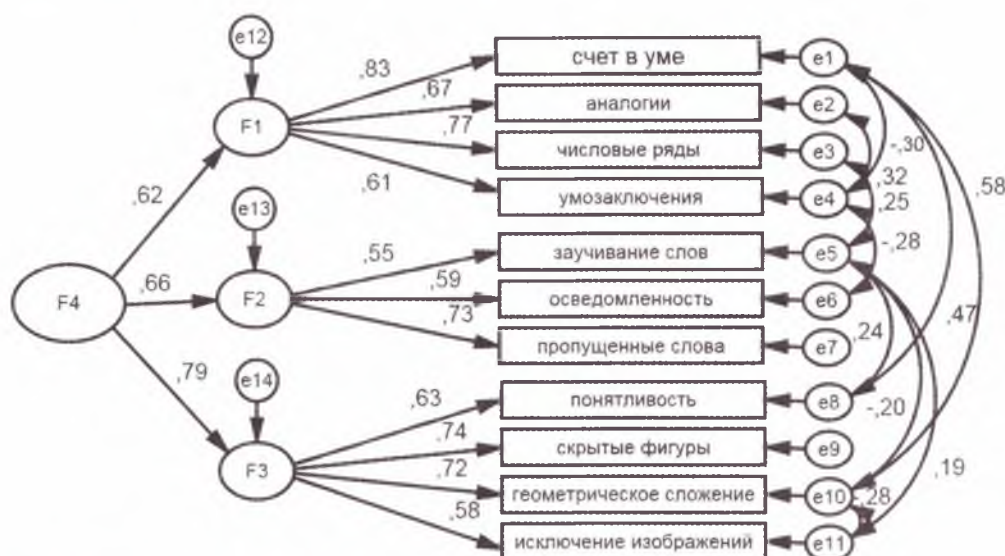


Рис. 26.22. Апостериорная модель КФА с вторичным фактором

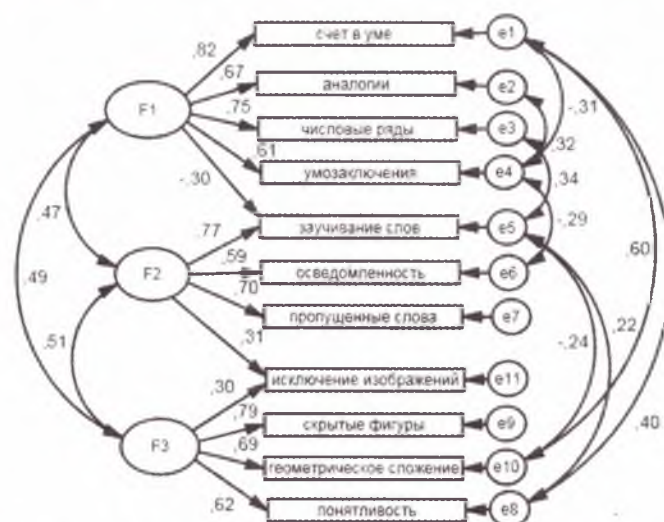


Рис. 26.23. Апостериорная модель КФА с пересечением факторов

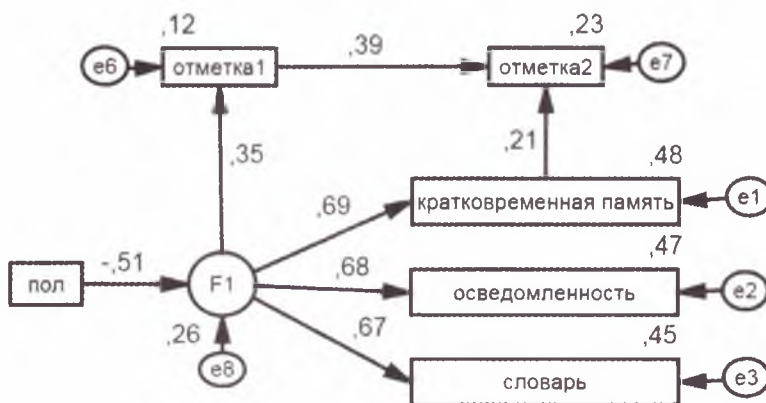
Основные критерии согласия для этой модели из раздела окна вывода Model Fit (Подгонка модели): $\chi^2 = 44,949$; $df = 31$; $p = 0,050$; $CMIN/df = 1,450$; $GFI = 0,960$; $AGFI = 0,916$; $CFI = 0,979$; $RMSEA = 0,050$; $PCLOSE = 0,478$. Таким образом, значения критериев согласия находятся в допустимых пределах даже с учетом малости выборки.

Таким образом, модель с пересекающимися факторами, единственная из трех апостериорных моделей КФА, эмпирически подтверждается. Не менее важным аргументом в пользу состоятельности модели является статистическая значимость всех оцененных дисперсий (Variances), ковариаций (Covariances) и регрессионных коэффициентов (Regression Weights) в разделе вывода Estimates (Оценки).

После констатации состоятельности модели возможна ее содержательная интерпретация, которая, однако, выходит за рамки этой книги. Отметим лишь, что переменная исключение изображений является весьма сомнительным индикатором для факторов F2 и F3: ее общность (сумма квадратов нагрузок) менее 0,20 ($31 \cdot 0,31 + 0,30 \cdot 0,3 = 0,186$). А вклад фактора F2 в дисперсию переменной заучивание слов ($0,77 \cdot 0,77 = 0,593$) более чем в 5 раз превышает вклад фактора F1 ($0,3 \cdot 0,3 = 0,09$). Также следует помнить о том, что отсутствие стрелок на диаграмме означает, что связь между соответствующими переменными строго фиксирована в модели как равная 0.

Структурная модель

На рис. 26.24 изображена апостериорная структурная модель как результат выполнения шагов 3а–11а по формированию, проверке и модификации априорной модели факторов успешности обучения (см. рис. 26.18). Критический коэффициент для многомерного эксцесса с.г. = $-0,210$, предположение о нормальной нормальности не отклоняется: применение для оценки метода максимального правдоподобия корректно.



Хи-квадрат=9,252; df=8; p=,322; RMSEA=,040

Рис. 26.24. Апостериорная структурная модель факторов успешности обучения

Модель является рекурсивной. Количество явных переменных в модели $P = 6$, $df = 8$, число оцениваемых параметров $T = P(P+1)/2 - df = 13$. Соотношение численности выборки ($N = 100$) к числу оцениваемых параметров $N/T = 7,69$ больше пяти (предельно малой величины), но меньше 10 (рекомендуемого соотношения). Следовательно, для принятия решения о согласии модели данным должны выполняться строгие ограничения для значений критериев согласия.

Основные критерии согласия для этой модели из раздела окна вывода Model Fit (Подгонка модели): хи-квадрат = 9,252; $df = 8$; $p = 0,322$; CMIN/ $df = 1,156$; GFI = 0,972; AGFI = 0,925; CFI = 0,989; RMSEA = 0,040; PCLOSE = 0,496. Таким образом, несмотря на малочисленность выборки, значения критериев согласия находятся в допустимых пределах для аргументированного принятия модели как подтвержденной эмпирическими данными.

В разделе Estimates (Оценки) текста вывода все оценки дисперсий (Variances), ковариаций (Covariances) и регрессионных коэффициентов (Regression Weights) статистически достоверны. Это подтверждает состоятельность модели и позволяет содержательно ее интерпретировать.

Измерительная модель интеллекта включает в себя три индикатора, которые являются показателями вербального интеллекта. КМК для каждого индикатора (общность) от 0,45 до 0,48, то есть доля дисперсии каждого из них на 45–48 % объясняется влиянием общего фактора, который может быть назван фактором вербального интеллекта.

Оценки общих эффектов (Total effects), косвенных эффектов (Indirect effects) и прямых эффектов (Direct effects) влияния одних переменных на другие доступны в подразделе Matrices (Матрицы) раздела Estimates (Оценки) текста вывода (рис. 26.25).

Standardized Total Effects (Group number 1 - Default model)

	пол	F1	отметка1	тест5
F1	-,514	,000	,000	,000
отметка1	-,179	,347	,000	,000
тест5	-,356	,691	,000	,000
отметка2	-,143	,278	,387	,207
тест3	-,344	,668	,000	,000
тест4	-,352	,684	,000	,000

Standardized Direct Effects (Group number 1 - Default model)

	пол	F1	отметка1	тест5
F1	-,514	,000	,000	,000
отметка1	,000	,347	,000	,000
тест5	,000	,691	,000	,000
отметка2	,000	,000	,387	,207
тест3	,000	,668	,000	,000
тест4	,000	,684	,000	,000

Standardized Indirect Effects (Group number 1 - Default model)

	пол	F1	отметка1	тест5
F1	,000	,000	,000	,000
отметка1	-,179	,000	,000	,000
тест5	-,356	,000	,000	,000
отметка2	-,143	,278	,000	,000
тест3	-,344	,000	,000	,000
тест4	-,352	,000	,000	,000

Рис. 26.25. Окно вывода для апостериорной структурной модели: фрагменты подраздела Matrices (Матрицы) раздела Estimates (Оценки)

Связи, представленные на диаграмме стрелками с числовыми значениями регрессионных коэффициентов, отражают так называемые *прямые эффекты* (direct effects) влияния одних переменных на другие. Так, фактором успешности обучения в 10-м классе является вербальный интеллект (F1), влияние которого объясняет 12 % дисперсии переменной отметка1. В свою очередь, 26 % дисперсии вербального интеллекта (F2) объясняется половыми различиями учащихся (пол), а отрицательный знак коэффициента регрессии свидетельствует о том, что у девушек (пол = 1) вербальный интеллект в среднем **выше**, чем у юношей (пол = 2). Дисперсия успеваемости в 11-м классе (отметка2) на 23 %

объясняется причинами, учтенными в модели. Из них 15,21 % приходится на влияние успеваемости в 10-м классе (отметка1), а 4,41 % — на влияние кратковременной памяти (остаток — на влияние взаимодействия двух последних переменных).

Помимо прямых эффектов интерес, как правило, представляют и *косвенные эффекты* (indirect effects) переменных. Например, как влияет переменная пол на успеваемость в 10-м классе (отметка1) или в 11-м классе (отметка2). В соответствии с моделью прямое (непосредственное) влияние переменной пол на эти переменные равно 0. Однако можно говорить о том, что влияние переменной пол на успеваемость в 10-м и 11-м классах опосредовано интеллектом (F1). Подобного рода влияние называется косвенным эффектом пола, опосредованным интеллектом. А интеллект в данном случае выступает *медиатором* косвенного влияния пола на успеваемость. Косвенный эффект пола на успеваемость в 10-м классе (отметка1) опосредован одним медиатором (F1), а на успеваемость в 11-м классе (отметка2) — 3 медиаторами: отметка1, F1 и кратковременная память. В косвенном эффекте пола на переменные отметка1 и отметка2 медиатором 1-го порядка является F1, а в отношении эффекта пола на переменную отметка2 переменные отметка1 и кратковременная память являются медиаторами 2-го порядка. Помимо медиаторов в структурной модели выделяют переменные *модераторы*, влияние которых меняет связь между другими переменными. Так, переменная F1 является модератором влияния пола на успеваемость в 10-м и 11-м классах (отметка1 и отметка2). Если исключить переменную F1, то обнаружится статистически достоверное влияние пола на успеваемость. В этом вы можете убедиться, сравнив успеваемость юношей и девушек при помощи *t*-критерия Стьюдента (см. главу 11). «Модерирующее» влияние F1 проявляется в данном случае в том, что эти прямые эффекты обращаются в 0, так как влияние пола на успеваемость опосредовано интеллектом (F1) и F1 выступает как медиатор этих эффектов. Переменные F1 и кратковременная память выступают модераторами связи отметка1 и отметка2, парная корреляция между которыми равна 0,434, а с учетом влияния модераторов снижается до 0,39.

О косвенных эффектах переменных модели позволяет судить таблица Standardized Indirect Effects (Стандартизованные косвенные эффекты) раздела вывода Estimates (Оценки). Так, косвенный эффект влияния пола на переменную отметка1 составляет $(-0,51) \cdot 0,35 = -0,179$: у девушек (пол = 1) успеваемость в 10-м классе выше, чем у юношей (пол = 2). Косвенные эффекты 2-го и более высоких порядков, особенно с несколькими медиаторами, вычислить сложнее. Так, косвенный эффект влияния F1 на переменную отметка2, опосредованный двумя медиаторами, равен 0,278 (положительная связь). А косвенный эффект влияния пола на переменную отметка2 равен $-0,143$ (отрицательная связь).

Терминология, используемая при выводе

- ▶ Asymptotically distribution-free (Приблизительно свободный от распределения) — альтернативный метод оценки параметров модели в AMOS, применяемый для больших выборок в случае отсутствия многомерной нормальности распределения переменных.
- ▶ с.г. (critical ratio — критическое отношение) — в AMOS отношение величины показателя (асимметрии, эксцесса, многомерного эксцесса) к его стандартной ошибке.
- ▶ CFI (Сравнительный индекс согласия) — критерий согласия модели; значение не менее 0,95 — хорошее согласие, не менее 0,90 — приемлемое согласие.

- ▶ CMIN (Критерий хи-квадрат) — проверяет H_0 на предмет того, что разность между эмпирической и воспроизведенной по модели матрицами ковариаций равна 0.
- ▶ Constrained parameter (Ограниченный параметр) — исключенная или отсутствующая связь между переменными модели, своего рода фиксированный параметр, приравненный к 0.
- ▶ Covariance (Ковариация) — произведение корреляции двух переменных и их стандартных отклонений.
- ▶ Default model (Модель по умолчанию) — модель, заданная пользователем для оценки.
- ▶ df (Число степеней свободы) — в SEM определяется как разность числа элементов ковариационной матрицы явных переменных $P(P + 1)/2$ (P — количество явных переменных) и количества свободных параметров модели, подлежащих оценке.
- ▶ Discrepancy (Расхождение) — разность между исходной ковариационной матрицей и репродуцированной (воспроизведенной) в соответствии с моделью.
- ▶ Disturbance (Помеха) — ошибка латентной переменной, обусловленная неучтенными в модели причинами.
- ▶ Endogenous variable (Эндогенная переменная) — переменная, на которую влияют другие переменные модели.
- ▶ Estimates (Оценки) — количественные оценки свободных параметров модели как результат анализа.
- ▶ Exogenous variable (Экзогенная переменная) — переменная, источником изменчивости которой являются причины, находящиеся за пределами модели; всегда только независимая переменная.
- ▶ Fixed parameter (Фиксированный параметр) — параметр модели (коэффициент регрессии, дисперсия, среднее), значение которого заранее задано как 1 или 0.
- ▶ Free parameter (Свободный параметр) — параметр модели, значение которого подлежит оценке (коэффициент регрессии, ковариация, дисперсия); для коэффициентов регрессии и ковариаций свободным параметрам соответствуют стрелки в блок-схеме модели.
- ▶ GFI (Критерий согласия) — критерий согласия модели; значение не менее 0,90 — хорошее согласие.
- ▶ Independence model (Независимая модель) — модель, в которой все корреляции между переменными равны 0.
- ▶ Maximum likelihood (Максимального правдоподобия) — метод оценки параметров модели, используемый в AMOS по умолчанию.
- ▶ Modification Indices, M.I. (Индексы модификации) — величина для пары переменных, равная уменьшению хи-квадрат при добавлении в модель оценки их связи (регрессии или ковариации).
- ▶ Observations farthest from the centroid (Mahalanobis distance) (Наблюдаемые удаления от центроида (Расстояния Махаланобиса), — в AMOS показатель удаленности наблюдения от центроида для диагностики выбросов.
- ▶ Observed variables (Наблюдаемая переменная) — явная переменная, содержащаяся в файле данных.
- ▶ Positive definite (Положительно определенная) — матрица (ковариаций), в которой отсутствуют взаимно однозначные связи между переменными (корреляции, равные единице или близкие к ней).

- ▶ Regression Weight (Коэффициент регрессии) — коэффициент для независимой переменной в регрессионном уравнении, пропорциональный соотношению дисперсий зависимой и независимой переменных (зависит от масштаба измерения переменных).
- ▶ RMSEA (Корень квадратный среднеквадратической ошибки аппроксимации) — критерий согласия модели, его значение не более 0,05 — хорошее согласие, не более 0,08 — приемлемое, 0,08 — 0,1 — слабое, более 0,1 — отсутствие согласия.
- ▶ Sample moments (Выборочные моменты) — ковариации и дисперсии явных переменных.
- ▶ Saturated model (Насыщенная модель) — модель, в которой количество оцениваемых параметров максимально возможно и равно количеству элементов ковариационной матрицы явных переменных модели $P(P + 1)/2$, где P — количество явных переменных.
- ▶ Standardized Regression Weight (Стандартизованный коэффициент регрессии) — коэффициент переменной в регрессионном уравнении, пропорциональный вкладу независимой переменной в дисперсию зависимой переменной (не зависит от масштаба измерения переменных).
- ▶ Standardized residuals (Стандартизированные остатки) — в AMOS это стандартизированные (в z -значениях) разности между исходными и воспроизведенными по модели ковариациями.
- ▶ Tests for normality and outliers (Критерии нормальности и выбросы) — раздел окна вывода AMOS, содержащий результаты проверки исходных данных на одномерную и многомерную нормальность явных переменных модели.
- ▶ Threshold for modification indices (Порог индекса модификации) — минимальное значение индекса модификации для выводимых пар переменных в разделе вывода Modification Indices (Индексы модификации).
- ▶ Under-identified model (Недоопределенная модель) — характеризуется тем, что количество свободных параметров превышает число элементов ковариационной матрицы явных переменных и оценка параметров невозможна.
- ▶ Uniqueness (Специфичность) — ошибка явной переменной, обусловленная ее спецификой.
- ▶ Unobserved variables (Ненаблюдаемая переменная) — латентная переменная, предполагаемый исследователем конструкт, проявляющийся через его влияние на индикаторы (явные переменные).

Завершение анализа и выход из программы

Результаты AMOS невозможно экспортировать в другие форматы, а текст вывода не подлежит редактированию в AMOS. Поэтому подготовка к помещению в отчет результатов сводится к копированию и вставке диаграммы и разделов текста вывода. Диаграмма при этом становится недоступной для редактирования. Содержимое текста вывода можно копировать и вставлять в документ Word целыми разделами или отдельными таблицами. При этом вставленные в документ Word фрагменты текста вывода доступны для редактирования.

Для выхода из программы выберите в меню File (Файл) команду Exit (Выход).

Глоссарий

Asymptotically distribution-free. Приблизительно свободный от распределения — альтернативный метод оценки параметров модели в AMOS, применяемый для больших выборок в случае отсутствия многомерной нормальности распределения переменных.

B. Показатели *B* представляют собой набор коэффициентов и константу регрессионного уравнения. Показатель *B* можно рассматривать как весовой коэффициент, характеризующий влияние соответствующей независимой переменной (предиктора) на зависимую переменную (критерий). Положительное значение *B* указывает на то, что с возрастанием предиктора значение критерия возрастает, а отрицательное — на то, что значение критерия убывает.

c.r., critical ratio. Критическое отношение — в AMOS отношение величины показателя (асимметрии, эксцесса, многомерного эксцесса) к его стандартной ошибке.

CFI. Сравнительный индекс согласия — в AMOS критерий согласия модели; значение не менее 0,95 — хорошее согласие, не менее 0,90 — приемлемое согласие.

CMIN. Критерий хи-квадрат — в AMOS проверяет H_0 на предмет того, что разность между эмпирической и воспроизведенной по модели матрицами ковариаций равна 0.

Constrained parameter. Ограниченный параметр — в AMOS исключенная или отсутствующая связь между переменными модели, своего рода фиксированный параметр, приравненный к 0.

Default model. Модель по умолчанию — в AMOS модель, заданная пользователем для оценки.

df. См. *Число степеней свободы*.

Discrepancy. Расхождение — в AMOS разность между исходной ковариационной матрицей и репродуцированной (воспроизведенной) в соответствии с моделью.

Disturbance. Помеха — в AMOS ошибка латентной переменной, обусловленная не учтенными в модели причинами.

Endogenous variable. Эндогенная переменная — в AMOS переменная, на которую влияют другие переменные модели.

Estimates. Оценки — в AMOS количественные оценки свободных параметров модели как результат анализа.

Exogenous variable. Экзогенная переменная — в AMOS переменная, источником изменчивости которой являются причины, находящиеся за пределами модели; всегда только независимая переменная.

Fixed parameter. Фиксированный параметр — в AMOS параметр модели (коэффициент регрессии, дисперсия, среднее), значение которого заранее задано как 1 или 0.

Free parameter. Свободный параметр — в AMOS параметр модели, значение которого подлежит оценке (коэффициент регрессии, ковариация, дисперсия); для коэффициентов регрессии и ковариаций свободным параметрам соответствуют стрелки в блок-схеме модели.

F-критерий. В дисперсионном анализе отношение межгруппового среднего квадрата к внутригрупповому среднему квадрату. Данная величина позволяет сравнить межгрупповую дисперсию с внутригрупповой дисперсией. Если первая окажется значительно выше второй, это будет означать наличие значимого различия между группами. В множественном регрессионном анализе *F*-критерий позволяет определить значимость множественной корреляции.

GFI. Критерий согласия — в AMOS критерий согласия модели; значение не менее 0,90 — хорошее согласие.

Independence model. Независимая модель — в AMOS модель, в которой все корреляции между переменными равны 0.

k. В иерархической логлинейной модели порядок взаимодействия эффектов; $k = 1$ соответствует первому порядку (одна переменная), $k = 2$ — второму порядку (две переменные) и т. д.

Maximum likelihood. Максимального правдоподобия — в AMOS метод оценки параметров модели, используемый по умолчанию.

Modification Indices, M.I. Индексы модификации — в AMOS величина для пары переменных, равная уменьшению хи-квадрата при добавлении в модель оценки их связи (регрессии или ковариации).

Observations farthest from the centroid (Mahalanobis distance). Наблюдаемые удаления от центроида (расстояния Махаланобиса) — в AMOS показатель удаленности наблюдения от центроида для диагностики выбросов.

Observed variables. Наблюдаемая переменная — в AMOS явная переменная, содержащаяся в файле данных.

Positive definite. Положительно определенная — матрица (ковариаций), в которой отсутствуют взаимно однозначные связи между переменными (корреляции, равные единице или близкие к ней).

p-уровень. См. *Значимость*.

R. Множественный коэффициент корреляции между зависимой переменной и двумя или более независимыми переменными. Значение *R* лежит в пределах от 0 до 1 и интерпретируется по аналогии с обычным (двумерным) коэффициентом корреляции.

R². Квадрат коэффициента множественной корреляции (коэффициент детерминации), доля дисперсии зависимой переменной, обусловленная воздействием двух или более независимых переменных.

RMSEA. Корень квадратный средней квадратической ошибки аппроксимации — в AMOS критерий согласия модели; значение не более 0,05 — хорошее согласие, не более 0,08 — приемлемое, 0,08...0,1 — слабое, более 0,1 — согласие отсутствует.

Saturated model. Насыщенная модель — модель в AMOS, в которой количество оцениваемых параметров максимально возможно и равно количеству элементов ковариационной матрицы явных переменных модели $P(P + 1)/2$, где *P* — количество явных переменных.

Standardized residuals. Стандартизированные остатки — в AMOS это стандартизированные (в *z*-значениях) разности между исходными и воспроизведенными по модели ковариациями.

S-стресс. В многомерном шкалировании мера степени соответствия модели исходной матрице различий. Чем меньше это значение, тем лучше соответствие.

Tests for normality and outliers. Критерии нормальности и выбросы — раздел окна вывода AMOS, содержащий результаты проверки исходных данных на одномерную и многомерную нормальность явных переменных модели.

Threshold for modification indices. Порог индекса модификации — в AMOS минимальное значение индекса модификации для выводимых пар переменных в разделе окна вывода Modification Indices (Индексы модификации).

t-критерий в регрессионном анализе. Критерий, определяющий статистическую значимость корреляций, равен отношению коэффициента B к своей стандартной ошибке.

t-критерий для зависимых выборок. Критерий, сравнивающий средние значения двух распределений для одной и той же выборки.

t-критерий для независимых выборок. Критерий, сравнивающий средние значения одной и той же переменной для двух независимых выборок.

t-критерий для одной выборки. Критерий, предназначенный для сравнения среднего значения распределения переменной с некоторой эталонной величиной.

t-критерий. Критерий для определения статистической значимости различия двух средних.

Uniqueness. Специфичность — в AMOS ошибка явной переменной, обусловленная ее спецификой.

Unobserved variables. Ненаблюдаемая переменная — в AMOS латентная переменная, предполагаемый исследователем конструкт, проявляющийся через его влияние на индикаторы (явные переменные).

V Крамера. Мера ассоциации между значениями двух категориальных переменных. Значение V всегда варьируется от 0 до 1 и интерпретируется по аналогии с коэффициентом корреляции (исключая отсутствие отрицательных значений). Нередко используется в контексте χ^2 -анализа; вычисление осуществляется по формуле

$$V = \sqrt{\chi^2 / [N(k-1)]},$$

где k — наименьшее из количеств строк и столбцов.

z-значения. Также называются стандартизованными значениями. После стандартизации (или z-преобразования) значений переменной среднее равно 0, стандартное отклонение равно 1. Стандартизованное значение может характеризовать направление и степень отклонения исходного значения от среднего. Для стандартизованных значений, превышающих по модулю 1,96, уровень значимости оказывается ниже 0,05.

Альфа (α). Мера внутренней согласованности измерительной шкалы, вычисляемая по формуле $\alpha = rk / [1 + (k-1)r]$, где k — число переменных в анализе, r — среднее значение корреляции между пунктами шкалы. Значение α зависит от числа переменных, поэтому его точной интерпретации нет, тем не менее в большинстве случаев действует следующая оценка внутренней согласованности шкалы: $\alpha > 0,9$ — отличная; $\alpha > 0,8$ — хорошая; $\alpha > 0,7$ — приемлемая; $\alpha > 0,6$ — сомнительная; $\alpha > 0,5$ — малопригодная; $\alpha < 0,5$ — недопустимая.

Альфа, если элемент удален. В анализе надежности значение α для шкалы, получающейся путем удаления текущего ее пункта.

Априорная вероятность для каждой группы. Вероятность, характеризующая предполагаемое соотношение численности групп. Для каждой из двух групп она равна 0,5, если предполагается, что их численность одинакова.

Асимметричная матрица. Квадратная матрица, у которой хотя бы в одной паре ячеек, симметрично расположенных относительно главной диагонали, значения различны. Примером асимметричной матрицы является корреляционная матрица.

Асимметрия. Мера отклонения распределения от нормального, характеризующая симметричность графика.

Асимптотические значения. Величины, на которых основано определение оценок параметров. Оценки параметров вычисляются в случаях, когда определение точных значений невозможно, в частности, в регрессионном анализе и некоторых других статистических процедурах.

Бартлетта критерий сферичности. Критерий многомерной нормальности для распределения переменных. Помимо нормальности критерий проверяет, отличаются ли корреляции от 0. Значение p -уровня, меньшее 0,05, указывает на то, что данные вполне приемлемы для проведения факторного анализа

Бета (β). В регрессионном анализе β означает стандартизованный коэффициент регрессии и представляет собой B -коэффициент для нормализованных переменных. Значения β всегда лежат в интервале от -1 до $+1$, значения β для разных переменных могут сравниваться друг с другом.

Бета при включении. Данная величина используется во множественном регрессионном анализе для переменных, не вошедших в уравнение регрессии, и представляет собой коэффициент β , рассчитанный в предположении, что соответствующая переменная была включена в регрессионное уравнение.

Биномиальный критерий. Непараметрический критерий, определяющий степень близости эмпирического распределения бинарной переменной к биномиальному распределению, для которого частоты двух категорий равны.

Бонферрони критерий. Критерий для множественного сравнения средних, если в дисперсионном анализе получен значимый результат.

Вальда критерий. В модели логистической регрессии критерий значимости коэффициента B для соответствующего предиктора. Чем выше его значение (вместе с числом степеней свободы), тем выше значимость.

Вероятность. Ожидаемая относительная частота некоторого события.

Взаимодействие. Эффект совместного влияния на зависимую переменную двух и более независимых переменных, который не сводится к их отдельному влиянию. В случае двух независимых переменных проявляется в том, что одна из них по-разному влияет на разных уровнях другой переменной.

Вилкоксона критерий. Непараметрический критерий, сходный критерием знаков, однако, в отличие от последнего, использующий ранги положительных и отрицательных разностей.

Внутригрупповая сумма квадратов. Сумма квадратов отклонений наблюдаемых значений от среднего для каждой группы.

Внутригрупповой фактор. Фактор, уровням которого соответствуют повторные измерения (зависимые выборки).

Внутригрупповой эффект. Эффект внутригруппового фактора, уровням которого соответствуют повторные измерения (зависимые выборки).

Вращение. Процедура, применяемая в факторном анализе, для того чтобы получить более простую структуру факторов.

Выборка. Подмножество объектов из некоторой генеральной совокупности, выбранное для статистических выводов относительно свойств всей совокупности.

Выборочные моменты. Ковариации и дисперсии явных переменных.

Гистограмма. Столбиковая диаграмма для отображения распределения частот по категориям (диапазонам значений) переменной. Горизонтальная ось графика соответствует значениям переменной, а вертикальная — частотам.

Главный эффект. Воздействие независимой переменной на зависимую переменную. Примеры главных эффектов можно найти в главах 14 и 15.

График собственных значений. Диаграмма, позволяющая выбрать число факторов в факторном анализе на основе критерия «каменистой осыпи» Р. Кеттелла.

Гуттмана критерий половинного расщепления. В анализе надежности половинного расщепления значение надежности, полученное с помощью процедуры нижних пределов.

Дендрограмма. Диаграмма древовидной структуры, иллюстрирующая процесс кластеризации в кластерном анализе. Пример дендрограммы приведен в главе 21.

Детерминант ковариационно-дисперсионной матрицы. Величина, характеризующая степень зависимости между значениями переменных. Чем меньше значение детерминанта, тем сильнее соответствующая зависимость. Эта величина используется при вычислении *M* Бокса. Детерминант общей дисперсионно-ковариационной матрицы учитывает все матрицы, используемые в анализе.

Диаграмма рассеяния. График для анализа связи между двумя переменными, на котором каждый объект представляет собой точку. Положение точки задано парой значений двух переменных для данного объекта. Более подробное описание приведено в главе 9.

Диаграмма регрессии. Диаграмма разброса, включающая в себя сдвиги точек от линии регрессии по вертикальной оси.

Дискриминантный анализ. Процедура создания формулы регрессии, на основе которой производится разбиение объектов на группы, соответствующие категориям зависимой переменной.

Дисперсионный анализ (ANOVA). Статистический анализ, устанавливающий статистическую значимость различий между средними значениями для трех или более выборок.

Дисперсии пунктов. Аналог средних значений элементов (пунктов шкалы) в анализе надежности. Поясняющий пример приведен в разделе «Представление результатов» главы 19.

Дисперсия. Характеристика выборочного распределения переменной, описывающая разброс значений вокруг среднего и вычисляемая как отношение суммы квадратов отклонений к объему выборки, уменьшенному на 1. Кроме того, дисперсия представляет собой квадрат стандартного отклонения.

Дисперсия шкалы, если элемент удален. Дисперсия суммы всех пунктов шкалы, кроме удаленного пункта.

Доверительный интервал. Диапазон, в котором находится большинство значений выборки. Например, термин «доверительный интервал 95 %» означает интервал, в который любое случайное значение из выборки попадает с вероятностью 95 %.

Доверительный интервал 95 %. См. *Доверительный интервал*.

Знаков критерий. Непараметрический критерий, определяющий различие двух измерений для одной выборки на основе знаков разностей пар значений.

Значимость (p -уровень). Мера случайности полученного результата, равная вероятности того, что в генеральной совокупности этот результат (различия, связь) отсутствует. Чем меньше эта вероятность (значение p -уровня), тем выше статистическая значимость результата. Результат считается статистически достоверным (значимым), если p -уровень не превышает 0,05.

Изменение R^2 . Изменение величины R^2 в результате введения новой переменной в уравнение регрессии.

Исправленная величина R^2 . Во множественном регрессионном анализе величина R^2 является точной для выборок, однако в генеральной совокупности ее значение лишь приблизительно. Исправленная величина R^2 представляет собой более точную оценку R^2 для генеральной совокупности и используется при сравнениях моделей, содержащих различное число независимых переменных.

Итерация. Стадия процесса формирования регрессионного (дискриминантного) уравнения, на которой происходит включение или исключение очередной переменной. Процесс продолжается до тех пор, пока не перестанет удовлетворяться заданный в процедуре критерий.

Кайзера — Мейера — Олкина критерий адекватности выборки. Величина, характеризующая степень применимости факторного анализа к данной выборке. Правило интерпретации этого критерия следующее: более 0,9 — безусловная адекватность; более 0,8 — высокая адекватность; более 0,7 — приемлемая адекватность; более 0,6 — удовлетворительная адекватность; более 0,5 — низкая адекватность; менее 0,5 — факторный анализ неприменим к выборке.

Канонические дискриминантные функции. Одно или более линейное дискриминантное уравнение, построенное таким образом, что классификация объектов по уровням зависимой переменной происходит наиболее точно. Более подробное описание вы можете найти в главе 22.

Канонические коэффициенты. В дискриминантном анализе канонический коэффициент представляет собой корреляцию между оценками дискриминантной функции и уровнями зависимой переменной. Более подробное описание вы можете найти в главе 22.

Категориальная (номинальная) переменная. Переменная, каждое значение которой указывает на принадлежность объекта к определенной группе (категории). Категориальная переменная не является количественной, она разделяет все объекты на непересекающиеся группы по определенному признаку (пол, хобби, класс и пр.), но не позволяет сравнивать объекты по уровню выраженности этого признака.

Квадрат евклидова расстояния. Мера, используемая по умолчанию в кластерном анализе для определения расстояния между объектами и кластерами и вычисляемая как сумма квадратов разностей между значениями переменных двух объектов.

Квадрат эта (η^2). Доля дисперсии зависимой переменной, обусловленная воздействием со стороны независимой переменной. Так, $\eta^2 = 0,044$ означает, что 4,4 % дисперсии зависимой переменной обусловлено данной независимой переменной.

Квадратная матрица. Матрица, строки и столбцы которой соответствуют одной и той же последовательности элементов (переменных или объектов).

Квартили. 25, 50 и 75-й процентиля.

Кластерный анализ. Процедура, на основе заданного правила объединяющая объекты или переменные в группы, называемые кластерами.

Ковариата. Количественная переменная, имеющая значительную корреляцию с зависимой переменной и включаемая в анализ для более точной проверки воздействий факторов на зависимую переменную.

Количественная переменная. Значения количественной переменной (в отличие от категориальной) отражают уровень выраженности у объектов соответствующего признака в метрической или порядковой шкале.

Колмогорова — Смирнова критерий для одной выборки. Непараметрический критерий, определяющий, отличается ли данное эмпирическое распределение от теоретического распределения (нормального, равномерного, Пуассона или экспоненциального).

Контрасты. Метод контрастов — это метод множественного сравнения средних в дисперсионном анализе, который позволяет сравнивать выборки по градациям независимой переменной. Например, метод контрастов позволяет сравнивать одну градацию с другой, одну градацию со всеми остальными или разбить все градации на 2 группы и затем сравнить их между собой.

Корреляция. Мера степени и направления связи между значениями двух переменных. Корреляции посвящена глава 9.

Корреляция между формами. В анализе надежности половинного расщепления приближенное значение надежности измерения в предположении, что обе половины содержат одинаковое число пунктов.

Корреляция между элементами. В анализе надежности это описательная информация о корреляциях каждого пункта с суммой всех остальных пунктов.

Коэффициент корреляции. Мера связи двух переменных, обозначаемая символом r и принимающая значения от -1 до $+1$.

Коэффициент регрессии. Коэффициент для независимой переменной в регрессионном уравнении, пропорциональный соотношению дисперсий зависимой и независимой переменных (зависит от масштаба измерения переменных).

Критерий согласия. В логлинейном анализе критерий χ^2 для определения степени адекватности модели исходным данным. Чем выше его значения и чем ниже соответствующие уровни значимости, тем хуже модель соответствует данным.

Ливиния критерий. Критерий, предназначенный для проверки гипотезы о том, что все распределения зависимой переменной для сравниваемых выборок имеют одинаковые дисперсии.

Линия регрессии. Линия на графике двумерного рассеяния, отражающая наиболее точные прогнозируемые значения (линия наилучшего соответствия).

Логарифмический детерминант. В дискриминантном анализе натуральный логарифм определителя каждой ковариационной матрицы. Логарифмический детерминант используется для вычисления M Бокса.

Логит. Натуральный логарифм шанса. Описание понятия «логит» приводится в главе 24.

Лямбда Уилкса. Отношение внутригрупповой суммы квадратов к общей сумме квадратов, характеризующее дисперсию оценок дискриминантной функции, не обусловленную различиями между двумя группами. Единичное значение λ принимает, если наблюдаемые средние значения групп равны; значения, близкие к нулю, означают, что внутригрупповая дисперсия мала по сравнению с общей дисперсией.

М Бокса (М). Критерий равенства дисперсионно-ковариационных матриц, основанный на близости значений определителей матриц ковариаций двух или более групп. Аналог критерия Ливиня для многомерного случая.

Максимум. Наибольшее наблюдаемое значение распределения переменной.

Манна — Уитни и Вилкоксона критерий ранговых сумм. Непараметрический аналог t -критерия, определяющий различие между двумя выборками на основе рангов.

Матрица различий. Матрица, каждое значение которой соответствует различию между двумя объектами.

Матрица трансформации факторов. Результатом перемножения этой матрицы и матрицы факторных нагрузок до вращения является матрица нагрузок факторов после вращения.

Медиана. Значение переменной, делящее упорядоченное множество всех значений выборки ровно пополам: у половины объектов выборки значения переменной больше, а у другой половины меньше медианы.

Медианный критерий к выборок. Непараметрический критерий для сравнения двух и более выборок по уровню выраженности признака; основан на подсчете числа элементов, лежащих выше и ниже главной медианы.

Межгрупповая сумма квадратов. Сумма квадратов разностей между главным средним значением и средними значениями групп, умноженных на весовые коэффициенты, равные числу объектов в соответствующих группах.

Метод главных компонент. Метод, применяемый SPSS по умолчанию в факторном анализе для извлечения факторов.

Метод иерархического слияния. Метод, используемый в кластерном анализе, при выполнении которого объекты объединяются в кластеры по одному на каждом шаге до тех пор, пока не будет образован единственный кластер, охватывающий все объекты. Кластерный анализ подробно описан в главе 21.

Метрическая переменная. Количественная переменная, соответствующая измерению признака в шкале интервалов или отношений. В отличие от ранговой (порядковой) переменной, при сравнении объектов позволяет судить не только о том, больше или меньше выражен признак, но и о том, насколько больше (меньше) он выражен.

Минимум. Наименьшее наблюдаемое значение распределения переменной.

Многомерные критерии значимости. В многомерном дисперсионном анализе набор критериев, позволяющих определить влияние факторов и их взаимодействий на совокупность зависимых переменных. Наиболее мощным считается критерий Пилая.

Многомерный дисперсионный анализ (ОЛМ-многомерная, MANOVA). Отличие многомерного дисперсионного анализа от одномерного (ANOVA) заключается в том, что число зависимых переменных в нем может быть теоретически любым.

Многомерный ковариационный анализ (MANCOVA). Многомерный дисперсионный анализ с включением в анализ ковариат.

Многомерный дисперсионный анализ с повторными измерениями (ОЛМ-повторные измерения). Вид дисперсионного анализа, в котором одна и та же группа объектов подвергается действию каждого уровня независимой переменной. С точки зрения вычисления этот анализ можно назвать внутригрупповым.

Многомерный критерий однородности матриц ковариаций. Критерий М Бокса определяет, являются ли ковариационные матрицы одинаковыми. Для каждого из значений вычисляется p -уровень, а также величина F или χ^2 .

Многомерное шкалирование. Метод, позволяющий на основе матрицы различий между объектами построить одно-, двух- или трехмерное изображение, иллюстрирующее удаленность этих объектов друг от друга.

Множественный регрессионный анализ. Метод, позволяющий спрогнозировать значения зависимой переменной на основе известных значений независимых переменных.

Мода. Наиболее часто повторяющееся значение распределения переменной.

Моучли критерий сферичности. Критерий многомерной нормальности. SPSS вычисляет приблизительное значение χ^2 и соответствующий уровень значимости. Если уровень значимости оказывается менее 0,05, то, вероятно, данные не являются нормально распределенными.

Наблюдаемое значение или частота. В χ^2 -анализе фактическая частота категории.

Надежность половинного расщепления. Мера надежности, для вычисления которой все пункты шкалы делятся на две эквивалентные группы, а затем на основе корреляции между двумя половинами шкалы устанавливается ее внутренняя согласованность.

Наименьшая ожидаемая частота. Наименьшая частота в ячейке таблицы сопряженности, определяемая в процессе применения критерия χ^2 .

Наименьшей значимой разности критерий (НЗР). Критерий множественного сравнения средних, представляющий собой серию t -критериев; применяется, если в дисперсионном анализе получен значимый результат.

Накопленная частота. Суммарное число объектов, имеющих значение переменной не большее, чем указано.

Накопленный процент. Процент объектов от общего числа, имеющих значение переменной не большее, чем указано.

Насыщенная модель. Логлинейная модель, включающая все взаимодействия и главные эффекты факторов.

Нелинейная регрессия. Процедура вычисления параметров нелинейного регрессионного уравнения.

Неортогональное вращение. Процедура, используемая в факторном анализе, допускающая результат, в котором угол между факторами отклоняется от прямого. Это иногда желательно для получения более простой структуры.

Непараметрические критерии. Серия критериев, каждый из которых применяется без предварительных допущений относительно нормальности распределения. Непараметрические критерии основаны на ранжировании, попарных сравнениях и других средствах, не требующих нормальности распределения переменных.

Нестандартизованные коэффициенты канонической дискриминантной функции. Список коэффициентов и константа дискриминантного уравнения.

Номинальная шкала. См. *Категориальная переменная*.

Нормальное распределение. Распределение частот (вероятностей), графически представляемое в виде симметричной кривой, имеющей пик в центре и асимптотически приближающейся к горизонтальной оси по краям. Идеальное нормальное распределение характеризуется нулевыми значениями асимметрии и эксцесса.

Общая внутригрупповая ковариационная матрица. Матрица, состоящая из средних значений ковариационных матриц, вычисленных для каждого уровня зависимой переменной.

М Бокса (М). Критерий равенства дисперсионно-ковариационных матриц, основанный на близости значений определителей матриц ковариаций двух или более групп. Аналог критерия Ливиня для многомерного случая.

Максимум. Наибольшее наблюдаемое значение распределения переменной.

Манна — Уитни и Вилкоксона критерий ранговых сумм. Непараметрический аналог t -критерия, определяющий различие между двумя выборками на основе рангов.

Матрица различий. Матрица, каждое значение которой соответствует различию между двумя объектами.

Матрица трансформации факторов. Результатом перемножения этой матрицы и матрицы факторных нагрузок до вращения является матрица нагрузок факторов после вращения.

Медиана. Значение переменной, делящее упорядоченное множество всех значений выборки ровно пополам: у половины объектов выборки значения переменной больше, а у другой половины меньше медианы.

Медианный критерий к выборок. Непараметрический критерий для сравнения двух и более выборок по уровню выраженности признака; основан на подсчете числа элементов, лежащих выше и ниже главной медианы.

Межгрупповая сумма квадратов. Сумма квадратов разностей между главным средним значением и средними значениями групп, умноженных на весовые коэффициенты, равные числу объектов в соответствующих группах.

Метод главных компонент. Метод, применяемый SPSS по умолчанию в факторном анализе для извлечения факторов.

Метод иерархического слияния. Метод, используемый в кластерном анализе, при выполнении которого объекты объединяются в кластеры по одному на каждом шаге до тех пор, пока не будет образован единственный кластер, охватывающий все объекты. Кластерный анализ подробно описан в главе 21.

Метрическая переменная. Количественная переменная, соответствующая измерению признака в шкале интервалов или отношений. В отличие от ранговой (порядковой) переменной, при сравнении объектов позволяет судить не только о том, больше или меньше выражен признак, но и о том, насколько больше (меньше) он выражен.

Минимум. Наименьшее наблюдаемое значение распределения переменной.

Многомерные критерии значимости. В многомерном дисперсионном анализе набор критериев, позволяющих определить влияние факторов и их взаимодействий на совокупность зависимых переменных. Наиболее мощным считается критерий Пилая.

Многомерный дисперсионный анализ (ОЛМ-многомерная, MANOVA). Отличие многомерного дисперсионного анализа от одномерного (ANOVA) заключается в том, что число зависимых переменных в нем может быть теоретически любым.

Многомерный ковариационный анализ (MANCOVA). Многомерный дисперсионный анализ с включением в анализ ковариат.

Многомерный дисперсионный анализ с повторными измерениями (ОЛМ-повторные измерения). Вид дисперсионного анализа, в котором одна и та же группа объектов подвергается действию каждого уровня независимой переменной. С точки зрения вычислений этот анализ можно назвать внутригрупповым.

Многомерный критерий однородности матриц ковариаций. Критерий M Бокса определяет, являются ли ковариационные матрицы одинаковыми. Для каждого из значений вычисляется p -уровень, а также величина F или χ^2 .

Многомерное шкалирование. Метод, позволяющий на основе матрицы различий между объектами построить одно-, двух- или трехмерное изображение, иллюстрирующее удаленность этих объектов друг от друга.

Множественный регрессионный анализ. Метод, позволяющий спрогнозировать значения зависимой переменной на основе известных значений независимых переменных.

Мода. Наиболее часто повторяющееся значение распределения переменной.

Моучли критерий сферичности. Критерий многомерной нормальности. SPSS вычисляет приблизительное значение χ^2 и соответствующий уровень значимости. Если уровень значимости оказывается менее 0,05, то, вероятно, данные не являются нормально распределенными.

Наблюдаемое значение или частота. В χ^2 -анализе фактическая частота категории.

Надежность половинного расщепления. Мера надежности, для вычисления которой все пункты шкалы делятся на две эквивалентные группы, а затем на основе корреляции между двумя половинами шкалы устанавливается ее внутренняя согласованность.

Наименьшая ожидаемая частота. Наименьшая частота в ячейке таблицы сопряженности, определяемая в процессе применения критерия χ^2 .

Наименьшей значимой разности критерий (НЗР). Критерий множественного сравнения средних, представляющий собой серию t -критериев; применяется, если в дисперсионном анализе получен значимый результат.

Накопленная частота. Суммарное число объектов, имеющих значение переменной не большее, чем указано.

Накопленный процент. Процент объектов от общего числа, имеющих значение переменной не большее, чем указано.

Насыщенная модель. Логлинейная модель, включающая все взаимодействия и главные эффекты факторов.

Нелинейная регрессия. Процедура вычисления параметров нелинейного регрессионного уравнения.

Неортогональное вращение. Процедура, используемая в факторном анализе, допускающая результат, в котором угол между факторами отклоняется от прямого. Это иногда желательно для получения более простой структуры.

Непараметрические критерии. Серия критериев, каждый из которых применяется без предварительных допущений относительно нормальности распределения. Непараметрические критерии основаны на ранжировании, попарных сравнениях и других средствах, не требующих нормальности распределения переменных.

Нестандартизованные коэффициенты канонической дискриминантной функции. Список коэффициентов и константа дискриминантного уравнения.

Номинальная шкала. См. *Категориальная переменная*.

Нормальное распределение. Распределение частот (вероятностей), графически представляемое в виде симметричной кривой, имеющей пик в центре и асимптотически приближающейся к горизонтальной оси по краям. Идеальное нормальное распределение характеризуется нулевыми значениями асимметрии и эксцесса.

Общая внутригрупповая ковариационная матрица. Матрица, состоящая из средних значений ковариационных матриц, вычисленных для каждого уровня зависимой переменной.

Общая сумма квадратов. Сумма квадратов отклонений всех значений от среднего значения всего распределения.

Общность. В факторном анализе мера, характеризующая долю дисперсии переменной, обусловленную воздействием всех факторов.

Одномерные F-критерии. В многомерном дисперсионном анализе критерии, которые характеризуют влияние независимых переменных и их взаимодействий на каждую зависимую переменную в отдельности.

Ожидаемое значение. В перекрестной таблице при использовании критерия χ^2 значение, вычисляемое в предположении, что все переменные являются полностью независимыми друг от друга. В регрессионном анализе термин «ожидаемое значение» эквивалентен термину «прогнозируемое значение» и означает величину, получаемую для каждого объекта в результате подстановки значений переменных для него в уравнение регрессии.

Остатки и стандартизованные остатки. В логлинейных моделях остатки представляют собой разности между ожидаемыми и наблюдаемыми частотами. Чем выше значения остатков, тем менее адекватной является модель. SPSS подсчитывает исправленные величины остатков с использованием оценок стандартного отклонения. Исправленные остатки представлены в единицах нормального распределения; как правило, значения остатков, по модулю превышающие 1,96, являются значимыми.

Остаток. Как правило, разность между наблюдаемым и ожидаемым значениями. Эта величина относится к части дисперсии, которая не объясняется воздействием независимых переменных.

Отклонение. Расстояние и направление (отрицательное или положительное) между средним и данным значениями.

Оценка дискриминантной функции. Значение, получаемое для каждого объекта путем подстановки значений его переменных в уравнение дискриминантной функции.

Параметр. Некоторая числовая характеристика генеральной совокупности.

Параметрические критерии. Критерии, применяемые в предположении о нормальном распределении переменных в генеральной совокупности.

Переменные в уравнении. При выводе результатов пошагового регрессионного анализа SPSS включает для каждого шага статистики тех переменных, которые вошли в уравнение регрессии.

Пирсона коэффициент корреляции. Мера корреляции, идеально подходящая для двух непрерывных (метрических) переменных.

Порядковая (ранговая) шкала. См. *Ранговая переменная*.

Прямоугольная матрица. Матрица, строкам и столбцам которой соответствуют разные последовательности элементов (объектов или переменных).

Размах. Характеристика распределения, равная разности между минимумом и максимумом распределения.

Ранговая (порядковая) переменная. Количественная переменная, отражающая измеренное качество на уровне порядка: в большей или меньшей степени оно выражено. В отличие от метрической шкалы, не позволяет судить о том, насколько больше или меньше выражено качество, поэтому не допускает применения арифметических операций.

Распределение. Статистическое понятие, обозначающее соотношение значений признака и частот (вероятностей) их встречаемости. Распределение (вероятностей, частот) может быть представлено в виде формулы для функции распределения вероятностей,

графика распределения частот (гистограммы, столбиковой диаграммы), таблицы распределения частот.

Регрессионный анализ. Инструмент статистики, позволяющий прогнозировать значения зависимой переменной с помощью известных значений независимых переменных. Регрессионный анализ подробно рассматривается в главах 17 и 18.

Регрессия. В множественном регрессионном анализе этим термином обозначается статистика, отражающая влияние предикторов на зависимую переменную.

Серий критерий. Непараметрический критерий, определяющий, является последовательность бинарных величин (событий) случайной или упорядоченной.

Симметричная матрица. Квадратная матрица, у которой в каждой паре ячеек, расположенных симметрично относительно главной диагонали, содержатся одинаковые значения. Типичным примером симметричной матрицы является корреляционная матрица.

Скорректированная корреляция пункта и суммы. В анализе надежности корреляция между пунктом шкалы и суммой всех остальных пунктов.

Собственное значение. В факторном анализе эта величина пропорциональна доле дисперсии, обусловленной влиянием данного фактора; в дискриминантном анализе отношение межгрупповой суммы квадратов к внутригрупповой сумме квадратов. Чем больше собственное значение, тем выше точность дискриминантной функции.

Спирмена — Брауна критерий неэквивалентных форм. В анализе надежности половинного расщепления надежность, вычисленная для случая, когда «половины» имеют неравный размер.

Спирмена — Брауна критерий эквивалентных форм. Используется в анализе надежности, когда число элементов в «половинах» одинаково (вычисляется коэффициент корреляции).

Среднее шкалы при удалении пункта. В анализе надежности для каждого пункта шкалы вычисляется сумма остальных пунктов по всем объектам выборки; отношение указанной суммы к числу объектов является средним шкалы, если данный элемент удален.

Средние значения пунктов. В анализе надежности (с применением α Кронбаха) описательная информация, касающаяся средних значений пунктов шкалы по всем наблюдениям. Поясняющий пример приведен в разделе «Представление результатов» главы 19.

Средний квадрат. Отношение суммы квадратов к числу степеней свободы. В однофакторном дисперсионном анализе, как правило, средний квадрат вычисляется для внутригрупповой и межгрупповой сумм квадратов, а в регрессионном анализе — для регрессионной и остаточной сумм квадратов. Во всех перечисленных случаях средний квадрат используется для вычисления F -критерия.

Стандартизованный коэффициент α . В анализе надежности значение α , полученное в случае, если перед проведением анализа стандартизовать распределения всех элементов шкалы.

Стандартизованный коэффициент регрессии. Коэффициент переменной в регрессионном уравнении, пропорциональный вкладу независимой переменной в дисперсию зависимой переменной (не зависит от масштаба измерения переменных).

Стандартная ошибка. Стандартное отклонение величины, получаемое в результате ее многократного вычисления для случайных выборок. Как правило, стандартная ошибка вычисляется для среднего значения распределения.

Стандартное отклонение. Мера разброса значений распределения вокруг среднего. Стандартное отклонение определяется как квадратный корень дисперсии (суммы квадратов отклонений от среднего, деленной на $N - 1$, где N — объем выборки).

Статистики шкалы. В анализе надежности статистические характеристики суммы всех переменных по объектам.

Столбиковая диаграмма. График распределения частот по категориям (значениям) переменной. Каждый столбец на графике соответствует одному значению признака, а его высота пропорциональна частоте встречаемости этого значения. Аналогичное средство для количественных переменных, имеющих большое число возможных значений, обычно называется гистограммой.

Стресс. В многомерном шкалировании мера соответствия модели исходной матрице различий. Чем меньше значение стресса, тем лучше соответствие модели.

Сумма квадратов. Стандартная мера разброса, представляющая собой сумму квадратов отклонений всех значений величины от среднего.

Таблица распределения (частот). Таблица, устанавливающая соотношение между категориями (значениями) признака и частотами их встречаемости.

Таблица сопряженности (кросстабуляции). Обычно таблица совместного распределения частот для двух категориальных или дискретных переменных; строки соответствуют категориям (значениям) одной переменной, а столбцы — другой. Подробно таблицы сопряженности описаны в главе 8.

Толерантность. Мера линейной зависимости между одной переменной и набором других переменных. Если при дискриминантном анализе уровень толерантности составляет менее 0,001, это означает, что линейная зависимость для данной переменной настолько высока, что ее включение в дискриминантное уравнение недопустимо.

Тьюки критерий подлинной значимости. Критерий множественного сравнения, позволяющий попарно сравнивать средние значения; применяется, если дисперсионный анализ показал значимый результат.

Фактор. В факторном анализе объединение нескольких переменных, чья взаимная корреляция исчерпывает определенную долю общей дисперсии. После процедуры вращения каждый фактор интерпретируется как некоторая общая причина взаимосвязи группы переменных.

Факторный анализ. Метод, позволяющий свести большое количество исходных переменных к значительно меньшему числу факторов, каждый из которых объединяет исходные переменные, имеющие сходный смысл.

Фи (φ). Мера связи (корреляции) двух категориальных переменных, обычно применяемая наряду с критерием χ^2 при анализе таблиц сопряженности и вычисляемая по формуле

$$\phi = \sqrt{\chi^2 / N}.$$

Фридмана дисперсионный анализ. Непараметрическая процедура, определяющая, различаются ли между собой три или более измерения для одной и той же выборки, на основе среднего ранга каждого измерения.

Хи-квадрат (χ^2) для модели. В анализе логистической регрессии величина, позволяющая определить, оказывают ли переменные, входящие в состав регрессионного уравне-

ния, значимое влияние на зависимую переменную. Чем выше полученное значение, тем значительнее воздействие.

Хи-квадрат (χ^2) критерий для одной выборки. Непараметрический критерий, определяющий отличие наблюдаемого распределения переменной от ожидаемого (теоретического) распределения.

Хи-квадрат критерий (критерий χ^2). Непараметрический критерий для сравнения ожидаемых и наблюдаемых частот (как правило, для таблиц сопряженности). Критерий χ^2 может использоваться для оценки адекватности структурных и логлинейных моделей. В любом случае, χ^2 -анализ всегда отвечает на вопрос: отличаются ли ожидаемые частоты модели от наблюдаемых. Коэффициент χ^2 Пирсона вычисляется по формуле

$$\chi^2 = \sum [(f_o - f_e)^2 / f_e].$$

Центроиды групп. В дискриминантном анализе средние значения дискриминантных функций для каждой из двух или более групп. Если число групп равно 2, центроиды будут иметь одинаковые абсолютные величины и разные знаки. Чем ближе объект к центроиду группы, тем больше вероятность того, что он принадлежит к этой группе.

Частичный критерий χ^2 . Значение критерия χ^2 , характеризующее долю воздействия очередной независимой переменной на зависимую переменную.

Частота (абсолютная). Количество объектов в выборке, имеющих данное значение признака.

Частота относительная. Доля объектов в выборке, имеющих данное значение признака; равна отношению абсолютной частоты к объему выборки.

Число степеней свободы (df). Количество возможных направлений изменчивости статистического показателя, наряду с эмпирическим значением критерия служит для определения p -уровня значимости.

Шаговый отбор переменных. Процедура, включающая и исключающая переменные из дискриминантного или регрессионного уравнения в соответствии с выбранными критериями.

Шанс. Отношение вероятности того, что событие произойдет, к вероятности того, что не произойдет.

Шеффе критерий. Процедура, позволяющая осуществлять попарные множественные сравнения средних значений после получения статистически достоверного результата дисперсионного анализа.

Экспонента В. В логистическом регрессионном анализе величина e^B используется в одной из форм регрессионного уравнения и позволяет создать одну из интерпретаций коэффициентов регрессии.

Эксцесс. Мера сглаженности (островершинности или плосковершинности) распределения. Если значение эксцесса близко к 0, это означает, что форма распределения близка к нормальному виду. Детальное описание эксцесса приведено в главе 7.

Эта (η). Мера корреляции между двумя переменными в случае, если одна из них является категориальной.

Англо-русский словарь терминов

Далее перечислены основные термины, с которыми может встретиться пользователь программы SPSS (в частности, в окне вывода). Более подробную информацию можно найти в разделах «Представление результатов» каждой из глав.

Термин	Перевод
Advanced models	Модуль дополнительных моделей
Agglomeration schedule	Шаги агломерации
Alignment	Выравнивание
Alpha	Альфа
Analysis Of Variances (ANOVA)	Дисперсионный анализ
Approximate significance	Приблизительная значимость
Asymptotic significance	Асимптотическая значимость
Asymptotically distribution-free	Приблизительно свободный от распределения, метод (AMOS)
Bar graph	Столбиковая диаграмма
Bartlett's test of sphericity	Критерий сферичности Бартлетта
Base (Statistics)	Основной системный модуль SPSS
Between-groups linkage	Межгрупповое связывание
Binomial test	Биномиальный критерий
Box plot	Коробчатая диаграмма
Box's M	M Бокса
c.r., critical ratio	Критическое отношение (показателя к его стандартной ошибке)
Canonical analysis	Канонический анализ
Category (x) axes	Ось категорий (x)
Category axes label	Метка оси категорий

Термин	Перевод
Cell	Ячейка (таблицы)
Central tendency	Центральная тенденция (мера)
Centroid	Центроид
CFI, Comparative Fit Index	Сравнительный индекс согласия (AMOS)
Chi-square	Хи-квадрат
Chronbach's alpha	Альфа Кронбаха
Cluster analysis	Кластерный анализ
Cluster method	Метод кластеризации
CMIN	Критерий хи-квадрат (AMOS)
Coefficient	Коэффициент
Commonality	Общность
Comparison	Сравнение
Complete linkage	Полная связь
Confidence interval	Доверительный интервал
Constrained parameter	Ограниченный параметр (AMOS)
Continuity correction	Поправка на непрерывность
Contrast	Контраст
Convergence	Сходимость
Correlation	Корреляция
Correlation matrix	Корреляционная матрица
Count measure	Количественный показатель (мера)
Covariance	Ковариация
Covariate	Ковариата
Cox & Snell R Square	R-квадрат Кокса и Снелла
Cramer's V	V Крамера
Criteria (criterion)	Условие

Термин	Перевод
Crosstabulation (crosstab)	Таблица сопряженности (кросс-табуляции)
Cumulative frequencies	Кумулятивные (накопленные) частоты
Curve	Кривая (линия)
Curvilinear analysis	Криволинейный анализ
Cut point	Точка деления
Data reduction	Сокращение данных
Default model	Модель по умолчанию (AMOS)
Degrees of Freedom (<i>df</i>)	Число степеней свободы
Deletion	Удаление (исключение)
Dendrogram	Дендрограмма
Density function	Функция плотности вероятности
Dependent sample	Зависимая выборка
Descriptive statistic	Описательная статистика
Determinant	Детерминант (определитель) матрицы
Deviation	Отклонение
Difference	Разность, различие
Dimension	Шкала
Directory	Каталог
Discrepancy	Расхождение, разность
Discriminant analysis	Дискриминантный анализ
Dispersion	Изменчивость
Dissimilarity	Различие
Distance	Расстояние
Distribution	Распределение
Distribution function	Функция распределения вероятности
Disturbance	Помеха, ошибка латентной переменной (AMOS)

Термин	Перевод
Eigenvalue	Собственное значение
Endogenous variable	Эндогенная переменная (AMOS)
Enter	Включение
Epsilon corrected	С эpsilon-коррекцией
Equality of variances	Равенство дисперсий
Equal-length Spearman — Brown	Коэффициент эквивалентных форм Спирмена — Брауна
Equation	Уравнение
Error	Ошибка
Error bar chart	Диаграмма столбцов ошибок
Eta Squared	Квадрат эта
Euclidian distance	Евклидово расстояние
Exogenous variable	Экзогенная переменная (AMOS)
Expected frequency	Ожидаемая (теоретическая) частота
Extraction method	Метод извлечения (факторов)
Factor	Фактор
Factor analysis	Факторный анализ
Factor score	Факторная оценка (значение фактора)
Factor score coefficient	Коэффициент факторной оценки (значений факторов)
Fisher's exact test	Точный критерий Фишера
Fixed factor	Фиксированный фактор
Fixed parameter	Фиксированный параметр (AMOS)
Footnote	Сноска
Free parameter	Свободный параметр (AMOS)
Frequency	Частота
Frequency table	Таблица частот
Friedman test	Критерий Фридмана

Термин	Перевод
F-value	F-критерий
General linear model	Общая линейная модель
Goodness-of-Fit Index (GFI)	Критерий согласия (AMOS)
Graph	График
Grid line	Линия сетки
Group centroid	Центроид группы
Group membership	Принадлежность к группе
Grouping variable	Группирующая переменная
Guttman Split-Half	Критерий надежности половинного расщепления Гуттмана
Hierarchical cluster analysis	Иерархический кластерный анализ
Histogram	Гистограмма
Homogeneity of variances	Гомогенность (однородность) дисперсий
Honestly Significant Difference (HSD)	Критерий подлинной значимости
Hotelling T-square	T-квадрат Хотеллинга
Icicle	Диаграмма накопления (слияния), сосульчатая диаграмма
Image factoring	Факторный анализ образов
Improvement	Улучшение
Independence model	Независимая модель (AMOS)
Independent sample	Независимая выборка
Individual differences model	Модель индивидуальных различий
Initial Eigenvalue	Исходное собственное значение
Inner frame	Внутренняя рамка
Interaction	Взаимодействие
Inverse distribution function	Обратная функция распределения
Item	Пункт (элемент шкалы)
Item variance	Дисперсия пункта (элемента шкалы)

Термин	Перевод
Iteration	Итерация
Kaiser — Meyer — Olkin Measure of Sampling Adequacy	Критерий адекватности выборки Кайзера — Мейера — Олкина
Kendall's tau	Тау Кендалла
Kendall's tau-b	Тау- <i>b</i> Кендалла
KMO and Bartlett test of sphericity	Критерии КМО и сферичности Бартлетта
Kolmogorov — Smirnov test	Критерий Колмогорова — Смирнова
Kruskal — Wallis one-way analysis of variance	Однофакторный дисперсионный анализ Краскела — Уоллеса
Kurtosis	Экссесс
Legend	Легенда (условные обозначения)
Legend label	Метка легенды
Level of measurement	Уровень (шкала) измерения
Levene's test	Критерий Ливиня
Likelihood	Правдоподобие
Line (style: dotted)	Пунктирная линия
Line (style: solid)	Сплошная линия
Line graph	Линейный график
Linear regression analysis	Линейный регрессионный анализ
Linear-by-linear association	Линейная связь между ~~перекрестными~~
Linkage	Соединение, связь
Listwise	Построчно
Loading	Нагрузка (факторная, компонентная)
Logistic regression	Логистическая регрессия
Logit	Логит
Loglinear analysis	Логлинейный анализ
Mahalanobis distance	Расстояние Махаланобиса
Main effect	Главный эффект

Термин	Перевод
Mann — Whitney <i>U</i>	Критерий <i>U</i> Манна — Уитни
Marker	Маркер
Matrix	Матрица
Mauchly's test of sphericity	Критерий сферичности Моучли
Maximum	Максимум
Maximum likelihood (ML)	Максимального правдоподобия, метод (AMOS)
McNemar test	Критерий Мак-Нимара
Mean	Среднее значение
Mean difference	Разность средних
Mean of Squares (MS)	Средний квадрат
Mean Rank	Средний ранг
Mean Square	Средний квадрат
Median	Медиана
Minimum	Минимум
Minimum expected count	Минимальная ожидаемая частота
Mode	Мода
Model selection	Подбор модели
Modification Indices, M.I.	Индексы модификации (AMOS)
Multidimensional scaling	Многомерное шкалирование
Multiple comparisons	Множественные сравнения (средних)
Multivariate Analysis Of Covariance (MANCOVA)	Многомерный ковариационный анализ
Multivariate Analysis Of Variances (MANOVA)	Многомерный дисперсионный анализ
Nagelkerke <i>R</i> Square	<i>R</i> -квадрат Найджелкерка
No correlation	Отсутствие корреляции
Nonparametric test	Непараметрический критерий
Normal distribution	Нормальное распределение

Термин	Перевод
Oblique rotation	Облическое (неортогональное) вращение
Observations farthest from the centroid	Наблюдаемые удаления от центроида (AMOS)
Observed frequency	Наблюдаемая (эмпирическая) частота
Observed variable	Наблюдаемая переменная (AMOS)
One sample	Одна выборка
One-sample Kolmogorov — Smirnov test	Критерий Колмогорова — Смирнова для одной выборки
One-tailed significance	Односторонняя значимость
Outer frame	Внешняя рамка
Paired samples	Парные (зависимые) выборки
Pairwise	Попарно
Parametric test	Параметрический критерий
Pareto chart	Парето-диаграмма
Partial correlation	Частная корреляция
Partial Eta Squared	Частичный квадрат эта
Pattern	Образец
Pearson Chi-Square	Критерий хи-квадрат Пирсона
Pearson correlation	Корреляция Пирсона
Percentile	Процентиль
Perfect negative correlation	Строгая отрицательная корреляция
Perfect positive correlation	Строгая положительная корреляция
Phi	Фи
Pie chart	Круговая диаграмма
Pillai's trace	След Пиллая
Plot	Диаграмма
Principal axis factoring	Факторный анализ методом главных осей
Principle component analysis	Анализ главных компонентов

Термин	Перевод
Prior probability	Априорная вероятность
Probability	Вероятность
Probability of group membership	Вероятность принадлежности к группе
Proximity	Близость
Quadratic relationship	Квадратичная связь (зависимость)
Quantile	Квантиль
Quartile	Квартиль
R Square	Коэффициент детерминации (квадрат коэффициента корреляции)
R Square change	Изменение коэффициента детерминации (квадрата коэффициента корреляции)
Random factor	Случайный фактор
Range	Размах
Rank	Ранг
Rectangular matrix	Прямоугольная матрица
Reference line	Вспомогательная линия
Regression coefficient	Коэффициент регрессии
Regression line	Линия регрессии
Regression model unit	Модуль регрессионных моделей
Related sample	Зависимая (связанная) выборка
Reliability analysis	Анализ надежности
Remove method	Метод исключения
Repeated-measures ANOVA	Анализ ANOVA с повторными измерениями
Residual	Остаток
Residual analysis	Анализ остатков
Root mean square error of approximation (RMSEA)	Квадратный корень средней квадратической ошибки аппроксимации (AMOS)

Термин	Перевод
Rotation method	Метод вращения
Sample	Выборка
Sample moments	Выборочные моменты (распределения)
Saturated model	Насыщенная модель
Saturated model	Насыщенная модель (AMOS)
Scale (y) axes	Ось значений (y)
Scale axes title	Заголовок оси значений
Scaling	Шкалирование
Scaling model	Модель шкалирования
Scatter plot	Диаграмма рассеяния
Scheffe test	Критерий Шеффе
Scree plot	График собственных значений («каменистой осыпи»)
Selection	Подбор
Sign	Знак
Sign test	Критерий знаков
Significance	Значимость
Significance level	Уровень значимости
Similarity	Сходство
Skewness	Асимметрия
Solution	Решение
Source	Источник
Spearman's rho	Коэффициент корреляции Спирмена
Split-half reliability	Надежность половинного расщепления
Square asymmetric matrix	Квадратная асимметричная матрица
Square symmetric matrix	Квадратная симметричная матрица
Squared Euclidean distance	Квадрат евклидова расстояния

Термин	Перевод
Squared multiple correlation	Квадрат множественной корреляции
s-stress	s-стресс
s-stress convergence	Величина сходимости s-стресса
Standard (std.) error	Стандартная ошибка
Standard deviation (std. dev.)	Стандартное отклонение
Standard error kurtosis	Стандартная ошибка эксцесса
Standard error mean	Стандартная ошибка среднего
Standard error skewness	Стандартная ошибка асимметрии
Standardized residuals	Стандартизированные остатки
Statistical Package for the Social Science (SPSS)	Статистический пакет для социальных наук
Statistical test	Статистический критерий
Stepwise method	Пошаговый метод
Stimulus	Стимул
Stress	Стресс
Subset	Подмножество
Subtitle	Подзаголовок
Sum	Сумма
Sum of Squares (SS)	Сумма квадратов
System missing value	Системное пропущенное значение
Territorial map	Карта территории
Test	Статистический критерий
Test value	Заданное (тестовое) значение
Tests for normality and outliers	Критерии нормальности и выбросы (AMOS)
Tick	Риск
Tied Ranks	Связанные (повторяющиеся) ранги
Title	Заголовок

Термин	Перевод
Tolerance	Толерантность
Total	Итог
Tree diagram	Древовидная диаграмма
True value	Истинное значение
Tukey's test of additivity	Критерий аддитивности Тьюки
t-value	t-критерий
Two-tailed significance	Двусторонняя значимость
Unequal-length Spearman — Brown	Критерий неэквивалентных форм Спирмена — Брауна
Uniqueness	Специфичность, ошибка явной переменной (AMOS)
Univariate procedure (approach)	Одномерный метод (подход)
Unobserved variable	Ненаблюдаемая, латентная переменная (AMOS)
Unstandardized coefficient	Нестандартизированный коэффициент
User missing value	Пользовательское пропущенное значение
Value	Значение
Value label	Метка значений
Variable (var)	Переменная
Variance	Дисперсия
Wald test	Критерий Вальда
Weight	Вес
Weirdness	Предсказуемость
Wilcoxon test	Критерий Вилкоксона
Wilks' Lambda	Лямбда Уилкса
Within-groups linkage	Внутригрупповое связывание
Working directory	Рабочий каталог
z-score	z-значение (стандартизованное значение)

Рекомендуемая литература

1. *Agresti, Alan* (1996). *An Introduction to Categorical Data Analysis*. New York: Wiley.
2. *Arbuckle, J. L.* Amos 7.0 user's guide. Chicago, IL: SPSS, 2006. 583 p.
3. *Bentler, P. M.* EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc., 2006. 418 p.
4. *Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W.* (1975). *Discrete multivariate analysis*. Cambridge, MA: MIT Press.
5. *Byrne, B. M.* *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. 2nd ed. Multivariate applications series. New York: Taylor & Francis Group, 2010. 396 p.
6. *Harrington, D.* *Confirmatory factor analysis*. New York: Oxford University Press, Inc., 2009. 122 p.
7. *Kline, R. B.* *Principles and practice of structural equation modeling*. 3rd ed. New York: The Guilford Press, 2011. 432 p.
8. *Айвазян С. А. и др.* Прикладная статистика. Классификация и снижение размерности. М., 1989.
9. *Боровиков В.* Statistica. Искусство анализа данных на компьютере: Для профессионалов. СПб., 2003.
10. *Гмурман В. Е.* Теория вероятностей и математическая статистика. М., 1997.
11. *Гублер Е. В., Генкин А. А.* Применение непараметрических критериев статистики в медико-биологических исследованиях. Л., 1973.
12. *Гусев А. Н.* Дисперсионный анализ в экспериментальной психологии: Учебное пособие для студентов факультетов психологии. М., 2000.
13. *Дэйвисон М.* Многомерное шкалирование: методы наглядного представления данных. М., 1988.
14. *Закс Л.* Статистическое оценивание. М., 1976.
15. *Иберла К.* Факторный анализ. М., 1980.
16. *Кендалл М., Стьюарт А.* Статистические выводы и связи. М., 1973.
17. *Крамер Д.* Математическая обработка данных в социальных науках: современные методы: Учеб. пособие. М.: Академия, 2007. 288 с.
18. *Митина О. В.* Моделирование латентных изменений с помощью структурных уравнений // Экспериментальная психология. 2008. № 1. С. 131–148.
19. *Митина О. В.* Основные идеи и принципы структурного моделирования // Ученые записки кафедры общей психологии МГУ. 2006. Вып. 2. С. 272–296.
20. *Митина О. В., Михайловская И. Б.* Факторный анализ для психологов. М., 2001.

21. *Наследов А. Д.* Математические методы психологического исследования. Анализ и интерпретация данных. СПб., 2008.
22. *Пацюрковский В. В., Пацюрковская В. В.* SPSS для социологов. Учеб. пособие. М.: ИСЭПН РАН, 2005. 433 с.
23. *Рунион Р.* Справочник по непараметрической статистике. М., 1982.
24. *Сидоренко Е. В.* Методы математической обработки в психологии. СПб., 1996.
25. Справочник по прикладной статистике. В 2-х т. / Под ред. Э. Ллойда, У. Ледермана, Ю. Торина. М., 1989.
26. *Суходольский Г. В.* Основы математической статистики для психологов. СПб., 1998.
27. *Суходольский Г. В.* Математические методы психологии. СПб., 2003.
28. *Терехина А. Ю.* Анализ данных методами многомерного шкалирования. М., 1986.
29. Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. М., 1989.
30. *Харман Г.* Современный факторный анализ. М., 1972.

Андрей Наследов
**IBM SPSS Statistics 20 и AMOS: профессиональный
статистический анализ данных**

Заведующий редакцией
Руководитель проекта
Ведущий редактор
Технический редактор
Литературный редактор
Художественный редактор
Корректор
Верстка

*А. Кривцов
А. Юрченко
Ю. Сергиенко
Е. Милокова
Н. Рощина
Л. Абуевская
И. Тимофеева
Е. Волошина*

ООО «Питер Пресс», 192102, Санкт-Петербург, ул. Андреевская (д. Волкова), д. 3, литер А, пом. 7Н.
Налоговая льгота — общероссийский классификатор продукции ОК 005-93, том 2: 95 3005 — литература учебная.
Подписано в печать 19.11.12. Формат 70х100/16. Усл. п. л. 33,540. Тираж 1500. Заказ № 1006.
Отпечатали с готовых диапозитивов в ГППО «Исковская областная типография».
180004, Псков, ул. Ротная, 34.



САЛД

САНКТ-ПЕТЕРБУРГСКАЯ
АНТИВИРУСНАЯ
ЛАБОРАТОРИЯ
ДАНИЛОВА

www.SALD.ru
8 (812) 336-3739

АНТИВИРУСНЫЕ
ПРОГРАММНЫЕ ПРОДУКТЫ

Книга представляет собой практическое руководство по анализу данных с помощью самой мощной и популярной программы статистической обработки информации — IBM SPSS Statistics 20. В издании подробно описываются основы работы с пакетом SPSS, рассматривается большинство методов обработки и анализа данных, а также способов табличного и графического представления полученных результатов. Кроме того, книга содержит подробную информацию о программе IBM SPSS AMOS, позволяющей использовать в исследованиях популярную и эффективную методологию моделирования структурными уравнениями (SEM — structural equation modeling).

Материал книги организован таким образом, чтобы удовлетворить запросы как новичка, впервые приступающего к анализу данных на компьютере, так и опытного исследователя, желающего воспользоваться самыми современными методами. Основное содержание глав составляют пошаговые инструкции по реализации различных видов математико-статистического анализа в SPSS. Особое внимание уделяется полученным результатам и их интерпретации. В конце книги приведен глоссарий, содержащий определения большинства статистических терминов.

Тема: **Статистика на компьютере**

Уровень пользователя: **начинающий/опытный**

ISBN: 978-5-496-00107-6



9 785496 001076

ПИТЕР

Заказ книг:

197198, Санкт-Петербург, а/я 127

тел.: (812) 703-73-74, postbook@piter.com

61093, Харьков-93, а/я 9130

тел.: (057) 758-41-45, 751-10-02, piter@kharkov.piter.com

www.piter.com — вся информация о книгах и веб-магазин