

ПРАКТИКУМ ПО ПСИХОДИАГНОСТИКЕ



С.И. КАЛИНИН

# КОМПЬЮТЕРНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ ПСИХОЛОГОВ



  
РЕЧЬ

С. И. Калинин

# КОМПЬЮТЕРНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ ПСИХОЛОГОВ

Системные требования	12
Установка программы	12
Состав пакета SPSS	13
Аутификация продукта SPSS	14
Запуск программы	15
Общий вид программы	16
<b>Глава 2. РЕДАКТИРОВАНИЕ ДАННЫХ В SPSS</b>	<b>22</b>
Создание нового файла данных	22
Импорт данных в SPSS	32
Редактирование таблицы данных	36
<b>Глава 3. АНАЛИЗ ДАННЫХ В SPSS</b>	<b>63</b>
Обзор статистических возможностей программы SPSS	63
Описательные статистики	67
Последовательное выделение статистических характеристик	67
Последовательное построение таблиц	74
Параметры измерения достоверности различий	82
Корреляционный анализ	87
	80



**РЕЧЬ**  
ИЗДАТЕЛЬСТВО

Санкт-Петербург

2002

УДК 159.9

ББК 88.4

К 17

**Научный редактор**  
канд. физ.-мат. наук А. Л. Тулупьев

**Калинин С. И.**

К 17 Компьютерная обработка данных для психологов / Под науч.  
ред. А. Л. Тулупьева. СПб.: «Речь», 2002. 134 с.

ISBN 5-9268-0091-9

Предлагаемое руководство знакомит с общими принципами работы с пакетом SPSS, а также с методами статистического анализа, наиболее распространенными в прикладной и исследовательской работе психолога (описательные статистики, выявление достоверности различий, корреляционный анализ).

Данное руководство будет полезным для психологов — студентов и специалистов — при написании ими научных работ и отчетов об исследованиях, для социологов, маркетологов и других специалистов.

**ББК 88.4**

ISBN 5-9268-0091-9

© С. И. Калинин, 2002  
© «Речь», 2002

# ОГЛАВЛЕНИЕ .

Неформальное предисловие.....	5
Введение.....	10
<b>Глава 1. ОБЩЕЕ ОПИСАНИЕ ПАКЕТА SPSS .....</b>	<b>12</b>
Системные требования.....	12
Установка программы.....	13
Состав пакета SPSS.....	14
Русификация программы SPSS.....	15
Запуск программы.....	16
Общий вид программы.....	18
<b>Глава 2. РЕДАКТИРОВАНИЕ ДАННЫХ В SPSS .....</b>	<b>22</b>
Создание нового файла данных.....	22
Импорт данных в SPSS.....	32
Редактирование таблицы данных .....	36
<b>Глава 3. АНАЛИЗ ДАННЫХ В SPSS .....</b>	<b>63</b>
Обзор статистических возможностей программы SPSS.....	63
Описательные статистики.....	67
Последовательность вычисления частотных характеристик.....	67
Последовательность построения таблиц сопряженности.....	74
Параметрические критерии достоверности различий.....	82
Непараметрические критерии достоверности различий.....	87
Корреляционный анализ.....	108

<b>Глава 4. РЕДАКТИРОВАНИЕ ОТЧЕТА В SPSS.....</b>	<b>113</b>
Редактирование отчета в редакторе	
SPSS Smart Viewer.....	113
Печать и экспорт отчета.....	124
<b>СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ.....</b>	<b>130</b>
<b>ПРИЛОЖЕНИЕ</b>	
Листинг файла поддержки кириллической	
кодировки шрифта в SPSS (CYRIL_UC.BAT).-.....	132

## НЕФОРМАЛЬНОЕ ПРЕДИСЛОВИЕ

Известен расхожий афоризм «Лень — основной двигатель прогресса», согласно которому весь научный и технический прогресс объясняется простым желанием жить легче. Данное пособие не является исключением из этого правила, т. к. оно появилось исключительно ради того, чтобы облегчить жизнь автору и всем небезразличным ему людям, с которыми его сталкивает жизнь.

Зачатки статистической лени, повлекшей за собой написание этого пособия, зародились еще в студенческие годы, когда автор, как и большинство восторженных гуманитариев, мечтал взмывать в высоты и нырять в глубины человеческой души, но никак не мог представить себе, что его ожидают непроходимые дебри учебного курса «Математические методы в психологии». Для истинного гуманитария, чья шаловливая мысль все время норовит сделать финт в сторону и переинтерпретировать уже интерпретированное, математика с ее неумолимой строгостью и последовательностью — нечто Иное, нечто запредельное, вызывающее суеверный страх и отторжение. (Уверен, этот внутренний «математический» холодок знаком почти каждому студенту-психологу младших курсов.)

Кроме инстинктивного ужаса перед математическим Иным, который, впрочем, с помощью толковых и чутких преподавателей почти всегда удается преодолеть, математика раздражала меня своим занудством. Вместо того, что валяться на скрипучей кровати в родном студенческом общежитии с очередным томиком любимого Юнга, приходилось вооружаться калькулятором (т. к. во времена студенческой молодости автора «персоналки» еще только создавались) и фактически вручную коррелировать между собой штук эдак 30 переменных. На подобные ручные расчеты времени уходила уйма, не говоря

уже о моральном ущербе — ведь предполагалось, что математические расчеты были не самоцелью, а всего лишь средством.

Через некоторое время математические страдания довели автора до приобретения компьютера (ах, эта была 286-я!) и статистической программы (ах, это был Statgraphics, версии, кажется, третьей!). Далее началась настоящая «гонка вооружений» — новые, все более мощные статистические программы требовали все более мощного компьютера, а все возрастающие возможности новых компьютеров подталкивали к желанию опробовать в работе новые программы. За последние десять лет автору довелось довольно плотно поработать с такими статистическими программами, как Statgraphics (с 3 по 5 версию включительно), Statistica for Windows (4 и 5 версии), SPSS (с 5 по 9 версию включительно) и ознакомиться с такими программами как SYSTAT, NCSS и др.

В результате столь обширного знакомства с миром статистических программ, автора посетила такая житейская мудрость.

- а) Можно сравнивать технические характеристики программ, но все равно нельзя прийти к достаточно обоснованному выводу, какая из программ «самая лучшая», т. к. каждая из них обычно имеет свою «изюминку», свои уникальные возможности.
- б) Парадоксальным образом все статистические программы одинаковы. Даже если в конкретной программе и имеется возможность расчета какого-либо сверхоригинального и сверхточного раритетного критерия, вопрос о том, насколько реально этот критерий востребован пользователем, остается открытым. А если не учитывать эти редко используемые статистические «изюминки», то основные возможности всех статистических программ одинаковы.
- в) Основной критерий выбора статистической программы — ее субъективная удобность, «совместимость» с пользователем. Автор этого пособия совершенно случайно «совместился» с программой SPSS, хотя он прекрасно понимает, что программа не идеальна (например, графические возможности программы Statistica намного богаче).

Прошли годы. Автору самому довелось стать преподавателем (хорошим или плохим — судить моим студентам) того самого ужасного предмета «Математические методы в психологии». Кроме того, ему

довольно часто приходится консультировать коллег-психологов по вопросам подбора и корректного использования статистических методов для обработки данных, полученных в ходе научных или прикладных исследований. Откровенно замечу, что и студенты, и коллеги-психологи частенько обращаются к автору как с «умными», так и с «глупыми» вопросами.

«Умные» вопросы обычно связаны с поиском адекватных методов статистического анализа для конкретных экспериментальных схем, методик, условий и т. п. Такие вопросы носят характер интеллектуальной головоломки: «А с помощью какого статистического инструментария мы можем описать, проанализировать проблему? А как именно статистический анализ поможет нам в интерпретации результатов экспериментального исследования?» «Умные» вопросы носят концептуальный, методологический характер, включая статистические методы как вспомогательное средство в общую канву исследования. При встрече с «умными» вопросами глаза автора загораются, его охватывает психолого-статистический азарт, а мысль становится невероятно изобретательной в поиске математических отмычек для психологических секретов.

«Глупые» вопросы, на которые автор отвечать не любит (и, надеюсь, при этом не слишком ворчит на своих студентов), обычно касаются технической стороны статистических расчетов. «Глупый» вопрос обычно строится так: «А как сосчитать... (далее следует название статистического метода)». Иногда причиной «глупых» вопросов бывает непонимание, но чаще лень. Тем, кто задает подобные вопросы, просто лень сесть за компьютер, поковыряться часок-другой с занимательной программой и через некоторое время с гордостью возопить: «Я умею это!» Когда автор сталкивается с подобными лентяями, его самого охватывает необузданная лень. Именно благодаря этой лени и появилось на свет данное пособие.

Для того чтобы все окончательно встало на свои места, сделаем несколько решающих оговорок:

1. Это пособие не по статистике. В нем уважаемый читатель не найдет формул и математических выкладок. Даже само назначение описываемых статистических методов и возможности их использования в психологических исследованиях расписаны здесь весьма кратко и скупо. Подобная краткость объясняется тем, что данное пособие



задумывалось прежде всего как справочное руководство для тех, кто *уже знает*, с помощью какого именно статистического метода можно проанализировать полученные эмпирические данные (то есть для тех, кто уже ответил на «умные» методологические вопросы), но при этом плохо умеет работать с компьютерной программой SPSS. В некоторых местах, характеризуя статистические возможности программы, автор умышленно упростил эти описания и теперь готов выслушивать справедливые упреки от математиков. Еще раз подчеркнем, что это пособие не о том, *что* считать, а о том, *как* это делать с помощью конкретного инструмента — программы SPSS. Тем, кто плохо понимает сущность описываемых статистических методов, рекомендуется обратиться к списку литературы, приведенному в конце книги.

2. Это пособие охватывает не все возможности программы SPSS. Статистические возможности программы SPSS огромны! За многие годы работы с программой автору доводилось видеть супер-пользователей SPSS, которые, буквально «на коленке», используя встроенный в программу язык макрокоманд, модифицировали заложенные в программу алгоритмы, создавая собственные изощренные методы анализа данных. Мощностъ программы SPSS возрастает прямо пропорционально вашим математическим знаниям и программистским навыкам, а также опыту работы с ней, и предел этого возрастания — бесконечность. Именно поэтому автор не ставил перед собой задачу «объять необъятное», а решил написать руководство, которое позволило бы «с нуля» освоить базовые возможности программы и сделать ее удобным инструментом в психологических исследованиях для людей, не очень опытных в компьютерных технологиях.
3. Это пособие предельно простое. Известна старая восточная мудрость: «Караван движется со скоростью самого хромого верблюда». Автор не раз убеждался в справедливости этой поговорки за много лет работы со студентами. Хотя компьютер перестал быть элитарной диковинкой и становится все более бытовой вещью, однако даже среди представителей «продвинутого» молодого поколения количество *advanced users*<sup>1</sup> ничтожно мало. Именно поэтому в процессе написания этого пособия автор постоянно представлял себе студента-

<sup>1</sup> *advanced users* (англ.) — продвинутый, опытный пользователь

первокурсника с круглыми глазами, который только что приехал из заповедной деревеньки и второй раз в жизни увидел компьютер.

Эта книга проста как спасательный круг. В ней подробно (и с картинками) расписано, какие кнопки нажимать, чтобы получить желаемый результат. Очень хотелось бы надеяться, что книга сможет реально помочь как «пионерам» от статистики (студентам младших курсов, которым всегда срочно нужно «считать курсовик», потому что завтра *deadline*<sup>2</sup>), так и «пенсионерам» от практики и науки (психологам старшего поколения, которые привыкли обрабатывать данные вручную, т. к. не очень уверенно работают на компьютере).

Напоследок хотел бы предупредить начинающих исследователей. Заполучив в руки такой мощный исследовательский инструмент, как программа SPSS, не обольщайтесь! Мощностъ программы и легкостъ ее использования создают опасную иллюзию всемогущества и подталкивают к некорректному и избыточному использованию статистики. Получить цифры расчетов легко, а вот корректно ли произведены сами вычисления, можно ли их «красиво» проинтерпретировать — это гораздо более важные вопросы, которые иногда забываются в погоне за «показушным» использованием мощных методов анализа. Не забывайте, что психологические открытия скрываются не в компьютеризированном инструменте, а в голове исследователя!

<sup>2</sup> *deadline* (англ.) — последний срок сдачи чего-либо

## **ВВЕДЕНИЕ**

Пакет статистических программ SPSS является мощным и удобным инструментом статистического анализа, который можно использовать для решения как научных, так и прикладных задач. Пакет SPSS начал разрабатываться с 1960-х годов специалистами Чикагского университета и первоначально являлся компьютерной реализацией узко специализированных методов статистического анализа, используемых в социальных науках (психологии, социологии, культуральной антропологии и др.). Данная специализация до сих пор отражается в названии программного продукта — аббревиатура SPSS расшифровывается как Statistical Package for Social Science (Статистический пакет для социальных наук).

Современная версия программы SPSS является многоцелевым статистическим пакетом, включающим ряд различных способов анализа, что позволяет с успехом его использовать в любых отраслях науки. Кроме базового программного ядра, охватывающего наиболее распространенные и универсальные статистические операции, существует около 30 специализированных программных модулей, подключаемых к SPSS. Данные модули в основном содержат специализированные статистические инструменты для решения прикладных экономических задач. Для примера, можно указать на программу Answer Tree, которая позволяет составлять высокоточные маркетинговые прогнозы на основе эффективной сегментации потребителей.

В настоящее время выпущена уже 10 версия статистического пакета SPSS, но, по наблюдениям автора, наиболее распространенным и используемым в России является 8 версия SPSS. До 5 версии включительно пакет SPSS реализован под операционную систему MS DOS, начиная с версии 6 — под операционную систему MS Windows. В дан-

ном руководстве речь пойдет о работе с пакетом SPSS for MS Windows 9-х release 9.0.0, выпущенного в 1999 г.

Пакет SPSS обладает всеми достоинствами прикладных статистических программ. Он позволяет вводить и редактировать данные, содержит в себе более 100 процедур статистического анализа, а также позволяет быстро создавать и редактировать полученные результаты в виде наглядных отчетов (в программе реализованы возможности для построения диаграмм около 50 типов). Кроме того, программа содержит мощную систему помощи и подсказок (в том числе по проведению статистического анализа), а также имеет удобный и интуитивно понятный интерфейс.

Для успешного овладения программой SPSS студентам-психологам рекомендуется предварительное изучение таких вузовских курсов, как «Высшая математика», «Методы математической статистики для психологов», «Психометрия», «Экспериментальная психология». Кроме того, для успешной работы с программой необходимо иметь базовые навыки пользователя ПЭВМ.

Предлагаемое руководство знакомит с общими принципами работы с пакетом SPSS, а также с некоторыми методами статистического анализа, наиболее распространенными в прикладной и исследовательской работе психолога (описательные статистики, выявление достоверности различий, корреляционный анализ). Данное руководство может быть полезно студентам-психологам для обработки экспериментальных данных в ходе написания курсовых или дипломных работ, а также психологам-практикам в процессе написания научно-исследовательских работ и разного рода отчетов о проделанной работе. Кроме этого, данное руководство может быть полезно для социологов, маркетологов и др.

## Глава 1

# ОБЩЕЕ ОПИСАНИЕ ПАКЕТА SPSS



## СИСТЕМНЫЕ ТРЕБОВАНИЯ

Для установки и нормальной работы программы SPSS 9.0 нужно иметь:

1. Процессор не ниже 486 DX (то есть с сопроцессором), что необходимо для обработки больших массивов данных, особенно при использовании сложных алгоритмов — факторного, кластерного и т. п. анализа. Рекомендуется процессор Pentium II и выше;
2. Не менее 12 Мб оперативной памяти (для Windows 95 и не менее 16 Мб для Windows NT). Рекомендуется 32 Мб оперативной памяти и более;
3. Не менее 60 Мб свободного места на жестком диске (только для модуля SPSS Base). При установке всех наиболее важных модулей программы (SPSS Base + SPSS Professional + SPSS Advanced + еще 5 модулей, указанных далее), включая дополнительные драйверы экспорта-импорта данных, требуется около 120 Мб свободного места на жестком диске. Кроме того, необходимо зарезервировать до 10-30 Мб свободного места для файла подкачки (как правило, на диске C:).
4. Операционную систему MS Windows 95—98 или Windows NT 4.0, можно и более поздние версии

## УСТАНОВКА ПРОГРАММЫ

Как правило, установка программы SPSS является простой и удобной процедурой. Достаточно вставить компакт-диск с программой в дисковод, как автоматически запускается программа инсталляции (рис. 1).

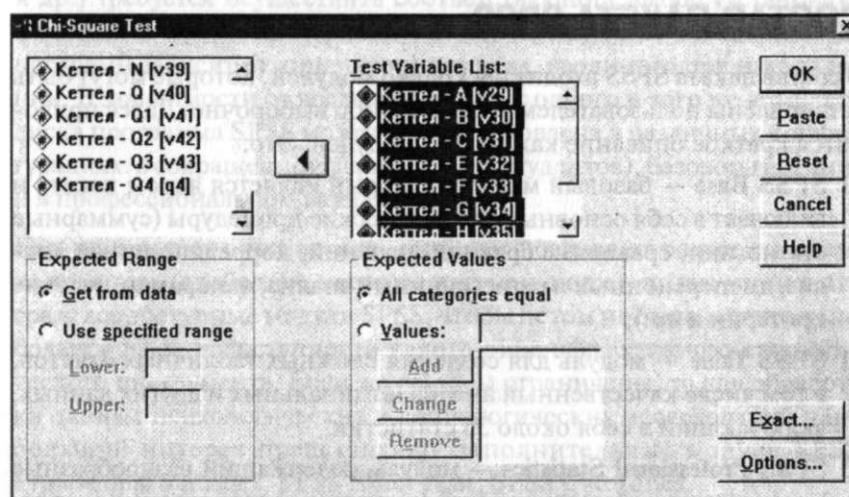


Рис. 1. Окно программы установки SPSS 9.0.

Щелкните левой клавишей мыши по надписи Install SPSS (установить SPSS) и далее четко следуйте инструкциям программы.

После установки самой программы, в ту же директорию что и SPSS, установите редактор статистических отчетов (Install Smart Viewer). Если вы этого не сделаете, то в дальнейшем результаты статистического анализа будут выводиться только в формате чернового отчета (Draft Viewer). Вы лишитесь многих привлекательных возможностей для редактирования отчетов и отображения красиво оформленных таблиц с результатами анализа.

Кроме того, рекомендуется установить к SPSS дополнительные драйверы (для этого нажмите Install SPSS Data Access Pack). Это позволит значительно расширить возможности экспорта-импорта данных различных форматов в программу. Также желательно установить программу Acrobat Reader (нажмите Install Acrobat Reader 4.0), что

позволит вам просматривать файлы в формате \*.PDF, т. к. многие файлы справок и руководства к SPSS созданы именно в этом формате.

Остальные опции окна установки не столь важны, поскольку содержат дополнительную или справочную информацию.

## СОСТАВ ПАКЕТА SPSS

В состав пакета SPSS входит несколько модулей, которые могут быть установлены пользователем либо все, либо выборочно. Далее приводится краткое описание каждого из модулей. Это:

- SPSS Base — базовый модуль, который является ядром пакета и включает в себя основные статистические процедуры (суммарные статистики, сравнение средних значений, корреляционный анализ, дисперсионный и регрессионный анализ, непараметрические критерии и др.);
- SPSS Table — модуль для создания сложных табличных отчетов, в том числе качественный анализ номинальных и других данных, включающий в себя около 35 статистик;
- SPSS Professional Statistics — модуль, содержащий разнообразные методы регрессионного анализа (логистическая регрессия, различные виды нелинейной регрессии, многомерное шкалирование и др.);
- SPSS Advanced Statistics — модуль, включающий в себя различные алгоритмы анализа сложных взаимосвязей (варианты применения обобщенной линейной и логлинейной моделей и др.);
- SPSS Trends — модуль, предназначенный для анализа временных рядов, выявления тенденций и формулирования прогнозов;
- SPSS Categories — модуль для анализа категориальных данных, который содержит в себе ряд методов изучения взаимосвязей в многомерных качественных (номинальных и ранговых) данных и их графического отображения;
- SPSS Exact Tests — модуль для анализа небольших массивов данных, полученных в малых выборках с использованием непараметрических критериев (более 30 критериев);
- SPSS Missing Value Analysis — модуль для анализа пропущенных в данных значений, который позволяет выявлять закономерности в пропусках данных, восстанавливать их и т. д.

Как уже упоминалось ранее, перечисленные модули являются основными. Количество дополнительных модулей может достигать до 10-20.

При установке программы SPSS некоторые из этих модулей устанавливаются по умолчанию (например, SPSS Tables); для установки остальных модулей (например, SPSS Advanced Statistics, SPSS Categories, и др.) требуется осуществить соответствующий выбор (установить флажки) во время инсталляции SPSS. Кроме того, комплектность программ зависит от лицензионного кода, вводимого при инсталляции. В зависимости от введенного кода, с одного и того же компакт-диска программа SPSS может быть установлена в различных конфигурациях: в сокращенном варианте (для студентов), базовом варианте и в профессиональном варианте и др.

Как показывает опыт автора, если вам позволяют ресурсы вашего компьютера (свободное место на жестком диске), лучше установить сразу все доступные модули SPSS, чтобы потом не было «мучительно больно» из-за недоступности какого-либо эффективного статистического инструмента. Если же ресурсы ограничены, то для обработки данных психологических и социологических исследований наибольший интерес представляют дополнительные модули SPSS Professional Statistics, SPSS Exact Tests, SPSS Categories.

**Примечание.** Перечисленные выше 8 модулей отражают внутреннюю структуру программы SPSS. После того как пакет SPSS установлен, выключить модули через меню программы невозможно. Но именно количество установленных модулей определяет количество и разнообразие статистических методов, которые можно выбрать, нажав на опцию Analyze (Анализ),

ч

## РУСИФИКАЦИЯ ПРОГРАММЫ SPSS

В настоящее время полностью русифицированной программы SPSS не существует, хотя и ведутся активные работы в данном направлении. Тем не менее, разработчиками SPSS предусмотрена возможность поддержки программой национальных языковых кодировок.

SPSS поддерживает русские шрифты в редакторе данных (отображает метки на русском языке в процессе выбора алгоритма статистического анализа), а также в окне вывода данных, включая таблицы и диаграммы. Таким образом, достигается «прозрачность» используемой



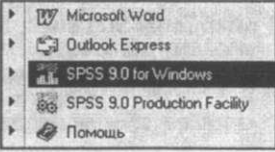


информации на всем протяжении работы с SPSS (ввод данных — статистический анализ — вывод и редактирование отчета).

Для того чтобы программа «понимала» (могла отобразить) кириллическую кодировку и чтобы в ней были прописаны русские шрифты, необходимо в любом текстовом редакторе создать файл CYRILLIC.BAT (листинг файла приводится в Приложении 1). Файл необходимо поместить в корневую директорию программы SPSS и запустить. В процессе запуска, файл автоматически пропишет в SPSS русские шрифты. Данный файл запускается только *однократно*.

**Примечание.** Если программа SPSS Smart Viewer или любой другой модуль SPSS установлены в отдельные директории, то файл CYRILLIC.BAT необходимо скопировать в эти директории и также однократно запустить.

## ЗАПУСК ПРОГРАММЫ

После того как вы установили SPSS, можно начинать работать с программой. Существуют 4 способа начать работу с SPSS.

1. В главном окне Windows нажать на кнопку **Пуск (Start)**, навести курсор на опцию **Программы (Programs)**, выбрать в дополнительном меню пункт **SPSS 9.0 for Windows** и щелкнуть на нем левой кнопкой мыши.
 
2. В корневом каталоге программы SPSS дважды щелкнуть левой кнопкой мыши на иконке программы (файл SPSSWIN.EXE)
 
3. Если вы уже работали с программой и имеете на жестком диске сохраненные в формате SPSS данные (файлы имеют характерные иконки и расширение \*.SAV), вы можете запустить программу, дважды щелкнув левой кнопкой мыши по иконке файла
 
4. Если вы совсем недавно работали с файлами данных SPSS, вы можете быстро загрузить их, нажав в главном окне Windows на кнопку **Пуск (Start)**, после чего необходимо навести курсор на опцию **Документы (Documents)** и выбрать в выпавшем меню необходимый файл. Щелкнув на нем левой кнопкой мыши, вы запустите SPSS.

Для первого или второго вариантов запуска SPSS, программа автоматически откроет следующее меню (рис. 2).

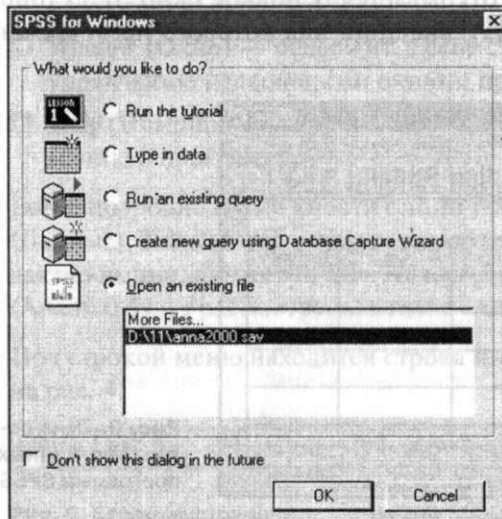


Рис. 2. Стартовое меню программы SPSS

Стартовое меню программы любезно предлагает «Что бы вы хотели сделать?» Вы можете установить флажок напротив соответствующей опции и нажать ОК, если вы хотите:

1. Запустить учебник (Run the tutorial), в котором излагаются основы работы с SPSS на английском языке и сведения из которого учтены при написании данного руководства;
2. Сразу перейти к редактору данных (Type in data) для создания нового файла данных;
3. Запустить ранее сформированный запрос (Run the existing query) на импорт в программу данных другого формата;
4. Создать новый запрос (Create new query using Database Capture Wizard) на импорт в программу данных другого формата с использованием соответствующего мастера;
5. Открыть уже существующий файл (Open an existing file) данных SPSS, с которым вы недавно работали.

## ОБЩИЙ ВИД ПРОГРАММЫ

После запуска программы, открывается основное окно SPSS (оно же — окно редактора данных), внешний вид которого представлен на рис. 3.

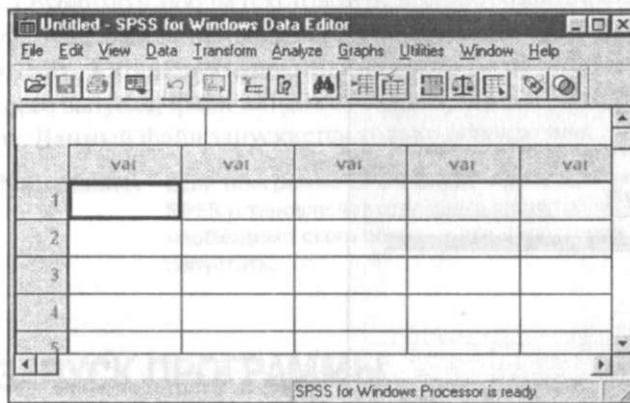


Рис. 3. Окно редактора данных программы SPSS

В строке меню главного окна SPSS находятся 10 опций:

- File** (Файл) — содержит в себе стандартные возможности по созданию новых файлов SPSS, экспорту и импорту данных в различные форматы, выводу на печать и т. п.
- Edit** (Правка) — содержит в себе стандартные возможности по редактированию файлов (вырезать, копировать, отменить действие и т. п.).
- View** (Вид) — определяет внешний вид главного окна SPSS, содержит в себе возможности настройки строки состояния, панели инструментов, шрифтов и т. п.
- Data** (Данные) — содержит основные операции по созданию и редактированию данных SPSS (создание новых переменных, создание шаблона ввода данных, сортировка данных и др.).
- Transform** (Трансформация) — содержит в себе операции по изменению данных SPSS (пересчет данных, перекодирование данных, замена пропущенных и ошибочных данных и др.).
- Analyze** (Анализ) — содержит в себе все статистические операции SPSS.
- Graphs** (Графики) — содержит в себе графические операции SPSS

- **Utilities** (Сервис) — содержит в себе дополнительные сервисные возможности по просмотру свойств переменных, файлов данных, по созданию скриптов и др.
- **Window** (Окно) — содержит в себе стандартные возможности по управлению несколькими окнами программы.
- **Help** (Помощь) — содержит в себе различную справочную информацию о программе.

Более подробно возможности опций **File** (Файл), **Edit** (Правка), **Data** (Данные), **Transform** (Трансформация) будут рассмотрены в главе «Редактирование данных в SPSS». Некоторые возможности опции **Analyze** (Анализ) будут рассмотрены в главе «Анализ данных в SPSS».

Под строкой меню находится строка инструментов, представленная на рис. 4.

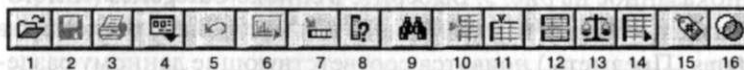


Рис. 4. Строка инструментов программы SPSS

Слева направо представлены пиктограммы следующих операций и инструментов: 1) открыть файл; 2) сохранить файл; 3) печать файла; 4) перейти к окну повтора операций (при нажатии на эту кнопку раскрывается окно, в котором выводится список запомненных программой произведенных ранее действий SPSS; это позволяет быстро находить и запускать часто используемые операции); 5) отмена или повтор произведенного действия; 6) создать рисунок (диаграмму); 7) найти (перейти к...) кейс; 8) показать справочную информацию о переменной; 9) найти (и перейти к...) конкретное значение в данной переменной; 10) добавить (вставить) новый кейс; 11) добавить (вставить) новую переменную; 12) разделить данные; 13) взвесить кейсы; 14) сортировать (отобрать кейсы по определенным правилам) данные; 15) показать/скрыть метки данных; 16) работа с ранее выделенной группой переменных.

Вы также можете настроить строку инструментов «под себя», разместив на ней пиктограммы наиболее необходимых и часто востребованных инструментов (см. рис. 5)

Для изменения строки инструментов вам нужно активизировать меню **View** (Вид), войти в опцию **Toolbars** (Строки инструментов), нажать

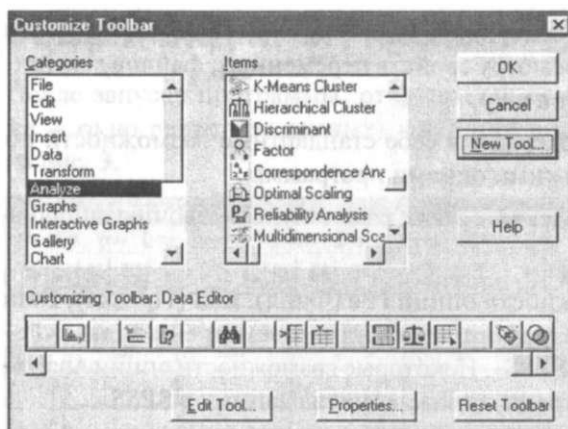


Рис. 5. Окно установки удобных для пользователя настроек строки инструментов

на кнопку *Customize* (Настройки пользователя). После этого появится окно, показанное на рис. 5. Выберите в списке *Categories* (Категории) тот раздел, который вам необходим для настройки, после чего в списке *Items* (Предметы) появятся соответствующие данному разделу пиктограммы инструментов. При наведении курсора на какую-либо пиктограмму из списка *Items* (Предметы) он меняет свою форму (рука вместо стрелки). Для того чтобы добавить выбранную пиктограмму в меню инструментов, необходимо навести на нее курсор, нажать левую кнопку мыши и, удерживая ее нажатой, перетащить пиктограмму вниз, на строку пиктограмм *Customizing Toolbar: Data Editor* (Пользовательское меню: редактор данных).

Когда вы нажмете на клавишу *OK*, строка инструментов программы SPSS примет удобный для вас вид. Если вы захотите вернуть панели инструментов первоначальный стандартный вид, нажмите на кнопку *Reset Toolbar* (Восстановить строку инструментов). Нажав на кнопку *Edit Tool...* (Редактировать инструмент), вы получите доступ к простому графическому редактору, который позволяет изменять внешний вид иконок.

Ниже строки инструментов находится строка ввода данных (см. рис. 6). В левой половине строки ввода данных указываются координаты активной ячейки в редакторе данных (ячейки, на которой в данный момент установлен курсор). Первым указывается номер строки (кейса, испытуемого), потом через двоеточие указывается номер (или название) переменной. Например, на рис. 6 указаны следующие ко-

ординаты 2: var00001, что означает активизацию второй сверху ячейки в первой переменной.

2:var00001		
	var00001	var00002
1	3,00	3,00
2	3,00	3,00
3	2,00	3,00
4	3,00	2,00

Рис. 6. Фрагмент строки ввода данных программы SPSS

В правой половине строки ввода данных указывается значение, размещенное в активной ячейке (здесь — 3).

Ниже строки ввода данных начинается рабочее пространство редактора данных SPSS, напоминающее обычный листок в клетку и имеющий стандартный для любых электронных таблиц вид.

С левого края расположены номера строк (кейсов), которые программа устанавливает автоматически по умолчанию. Сверху («шапки» столбцов) находятся названия переменных (по умолчанию они маркируются как var; от слова *variable* — переменная) и цифра, означающая порядковый номер переменной. Номера строк пользователь изменить не может; о том, как изменить названия переменных, будет рассказано далее в разделе «Создание нового файла данных».

Слева и снизу в окне редактирования данных расположены традиционные для Windows-программ вертикальная и горизонтальная полосы прокрутки. В самом низу окна находится строка состояния, которая отображает информацию о производимых программой в настоящий момент операциях. Например, на рис. 3 в строке состояния содержится сообщение о готовности программы SPSS к работе (SPSS for Windows Processor is ready).

## Глава 2

# РЕДАКТИРОВАНИЕ ДАНЫХ В SPSS



## СОЗДАНИЕ НОВОГО ФАЙЛА ДАННЫХ

При запуске программы в стартовом меню SPSS (см. рис. 2) выберите опцию **Type in data** (Печатать в редакторе данных) или просто нажмите кнопку **Cancel** (Отмена). После этого вы окажетесь непосредственно в редакторе данных.

Далее необходимо создать переменную. Для этого наведите мышью курсор на «шапку» первой переменной (на серую надпись **var**) и дважды щелкните по ней левой кнопкой мыши или щелкните по ней правой кнопкой мыши и выберите в появившемся меню опцию **Define Variable...** (Назвать переменную). После этого появится окно создания переменной (см. рис. 7):

В поле **Variable Name** (Имя переменной) можно напечатать новое название. Название переменной должно содержать в себе не более 8 букв и/или цифр. Если программа русифицирована (см. ранее), то можно назвать переменную по-русски.

Ниже, в разделе **Variable Description** (Описание переменной) приводится информация о переменной, а именно:

- Type** (Тип данных). На рис. 7 в качестве примера приведен принятый по умолчанию числовой тип данных (**Numeric8.2**). Формат данных **Numeric8.2** означает, что до знака десятичных разрядов (запятой) может быть введено восьмизначное число (состоящее

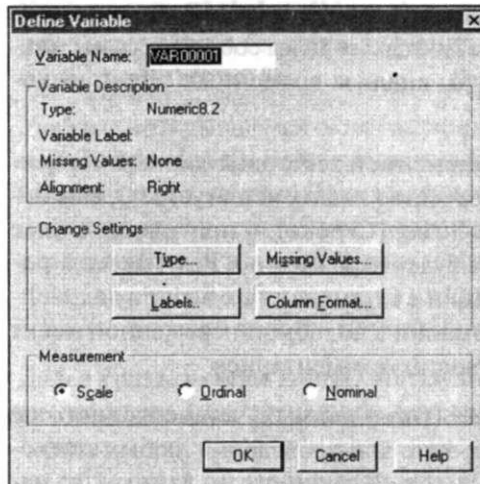


Рис 7. Окно создания новой переменной и редактирования свойств переменной в редакторе данных SPSS

из 8 цифр), а после запятой программа отобразит только две цифры. Разумеется, после запятой может быть гораздо больше цифр (они будут отображаться в строке редактирования справа) и программа их корректно сохранит, но в ячейках редактора данных будут видны только две цифры после запятой;

- Variable Label** (Метка переменной). В данном примере отсутствует;
- Missing Values** (Формат пропущенных значений). В этом случае не определен (**None**);
- Alignment** (Выравнивание данных в ячейке). На рис. 7 указано выравнивание по правому краю (**Right**), принятое в программе по умолчанию.

Для того чтобы изменить любой из параметров описания переменной, необходимо нажать соответствующую кнопку в разделе **Change Settings** (Изменить установки).

Нажав кнопку **Type...** (Тип данных), в открывшемся окне вы сможете выбрать и установить формат (шаблон) ввода данных, соответствующих создаваемой переменной. Наиболее распространен обычный числовой формат (**Numeric**). Можно также изменить количество вводимых цифр до запятой (**Width**) и отображаемых цифр после запятой (**Decimal Places**).

Календарный формат (**Date**) позволяет вводить хронологические данные (день, месяц, год, часы и т. п.) в любой удобной для вас форме



(по шаблону, выбранному из предлагаемого списка). Важно придерживаться календарного формата данных, если вы собираетесь использовать статистические процедуры анализа временных серий, выявления тенденций и т. п.

Если вы хотите разместить в переменной текстовые данные (например, написать фамилию, имя, отчество ваших испытуемых), вам необходимо выбрать тип данных **String** (Строка), в противном случае ввести текст в ячейку редактора будет невозможно. Разумеется, производить статистические операции с типом данных **String** в дальнейшем будет невозможно. Информация в подобной переменной носит в основном характер напоминающих комментариев.

В большинстве случаев формата (типа) данных, установленного по умолчанию (**Numeric8.2**), достаточно для проведения любых статистических процедур. Если вы все-таки изменили тип данных, то нажмите кнопку **Continue** (Продолжить), после чего вы вернетесь в окно создания новой переменной (рис. 7).

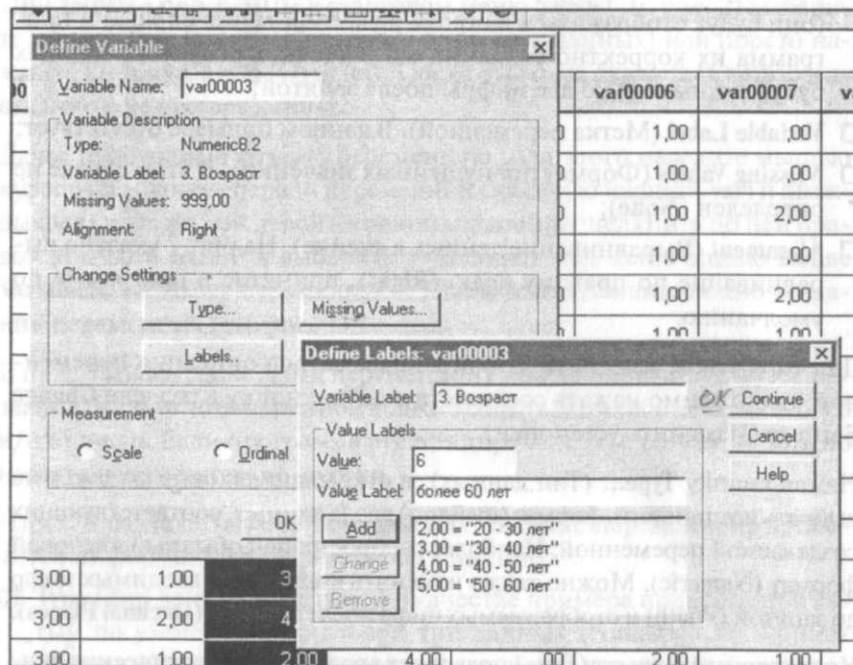


Рис. 8. Окно создания и редактирования меток переменной и значений

Нажав кнопку Labels... (Метки) вы сможете снабдить саму переменную и входящие в ее состав значения произвольными комментариями (метками) (рис. 8).

Укажите в открывшемся окне создания меток комментариев к переменной (Variable Label). Если программа SPSS русифицирована, его можно написать по-русски. Длина метки не должна превышать 256 символов. Например, если производится обработка психологического исследования, в качестве метки переменной может быть приведено полное наименование тестовой шкалы; если производится обработка социологической анкеты, меткой может послужить анкетный вопрос.

Далее в разделе Value Labels (Метки значений) вы можете ввести комментарии ко всем значениям, встречающимся в переменной (длина каждого — не более 60 символов). Разумеется, это целесообразно только тогда, когда в переменной встречается ограниченное число возможных значений. Так часто бывает, если они измерены в шкале наименований (например, варианты ответов на вопрос анкеты). Если в данной переменной возможен сравнительно большой разброс значений (например, сырые баллы по какой-либо из шкал теста 16PF), то метки значений можно совсем не устанавливать.

Для ввода меток значений напишите в графе значение (Value) соответствующее число, в графе Value Label (Метка значения) — требуемый комментарий и нажмите кнопку Add (Добавить). Если вы хотите удалить комментарий к какому-либо значению, выберите его в списке и нажмите кнопку Remove (Убрать). Если вы хотите уже к готовому комментарию подставить новое значение, введите в поле Value (Значение) соответствующее число, выберите в списке нужный комментарий и нажмите кнопку Change (Изменить).

Не ленитесь подробно описывать свои данные, задавая метки переменной и значений! Именно метки обеспечивают «прозрачность» данных на всех этапах работы с ними. И позволяют:

1. Максимально визуализировать данные в редакторе данных. При наведении курсора на название («шапку») переменной SPSS автоматически выводит метку переменной в виде всплывающей подсказки.

В опции View (Вид) можно использовать Value Labels (Метки значений) или нажать на соответствующую кнопку в строке инструментов (номер 15 на рис. 4).

После этого в ячейках редактора данных вместо цифр будут появляться соответствующие им комментарии. Для того чтобы скрыть комментарии, достаточно повторить операцию.

2. Без труда выбирать требуемые переменные для статистического анализа (подробнее см. главу «Анализ данных в SPSS»).
3. Создавать подробные, удобные для работы статистические отчеты и наглядные диаграммы с исчерпывающими комментариями (подробнее см. главу «Редактирование отчета в SPSS»).

После того как вы определили метки переменной и значений, нажмите кнопку Continue (Продолжить) и вернитесь в окно создания новой переменной (рис. 7).

Нажав кнопку Missing Values... (Пропущенные значения), вы сможете определить, какие именно значения в таблице данных считать пропущенными. Это весьма важная характеристика переменной, т. к. пропущенные значения не включаются в обработку при использовании различных статистических процедур.

Начинающие статистики часто совершают ошибку, отождествляя нулевые и пропущенные (отсутствующие) значения. Например, если испытуемый не прошел тест (его данные отсутствуют), то в таблице данных это зачастую ошибочно обозначается как ноль. Но с позиции многих статистических методов ноль — это полноправное число, которое включается в обработку и может значительно повлиять на итоговые результаты. Кроме того, в некоторых методиках испытуемый может на самом деле дать ответ, равный нулю.

Для того, чтобы избежать подобных искажений, программе необходимо напрямую указать, какое именно значение (или значения) счи-

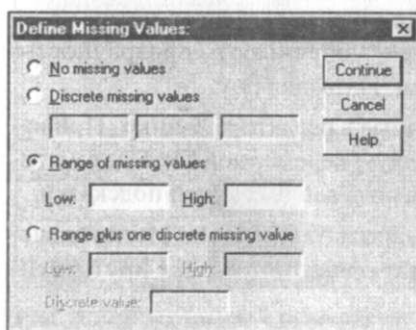


Рис. 9. Окно определения пропущенных значений

тать пропущенным, то есть не включать в обработку, чтобы не происходило искажения результатов.

По умолчанию в программе установлен флажок **No missing values** (Нет пропущенных значений) (рис. 9). В подобном случае программа воспринимает как пропущенные только незаполненные ячейки, в которых вообще нет никакой информации (**system missing**).

Установив флажок на опции **Discrete missing values** (Дискретные пропущенные значения), можно указать от 1 до 3 значений, которые программа будет исключать из обработки данной переменной.

Установив флажок на опции **Range of missing values** (Интервал пропущенных значений), можно указать, ниже и/или выше какого конкретного значения (включая его) все соответствующие этому условию значения будут исключены из обработки.

Опция **Range plus one discrete missing value** (Интервал плюс одно дискретное пропущенное значение) является комбинированной и совмещает в себе возможности двух предыдущих опций.

После того как вы определили, какие значения при обработке считать пропущенными, нажмите кнопку **Continue** (Продолжить) и вернитесь в окно создания новой переменной (рис. 7).

При нажатии на кнопку **Column format** (Формат столбца) появляется окно, в котором можно определить ширину колонки данных в редакторе (по умолчанию — 8 символов), а также выравнивание данных (по умолчанию — вправо). Если вы изменили какие-либо параметры, нажмите кнопку **Continue** (Продолжить) и вернитесь в окно создания новой переменной (рис. 7).

После того как вы побывали во всех четырех опциях окна создания переменной (Тип..., Метки..., Пропущенные значения, Формат столбца) и произвели изменения, они отобразятся в том же окне в разделе **Variable Description** (Описание переменной).

Для завершения создания переменной необходимо в разделе **Measurement** (Измерение) указать, с помощью какой именно измерительной шкалы получены данные. Для этого нужно установить флажок напротив одной из трех опций:

- Scale** (Шкальное) соответствует измерениям, результаты которых представлены в шкале интервалов (отношений);

**Ordinal** (Порядковое) соответствует измерениям, результаты которых представлены в шкале порядка (рангов);

**Nominal** (Номинальное) соответствует измерениям, результаты которых даны в шкале наименований (или в дихотомической).<sup>3</sup>

Правильное указание измерительной шкалы крайне важно, в противном случае методы статистического анализа могут быть использованы некорректно. Например, к измерениям, произведенным с использованием шкалы наименований, неприменимы методы, основанные на сопоставлении средних арифметических величин (для столь любимого психологами t-критерия Стьюдента годятся только интервальные данные, о чем они частенько забывают).

После того как вы указали тип измерительной шкалы, использованной в данной переменной, нажмите на **ОК**, и создание новой переменной будет завершено.

В дальнейшем, в процессе работы с редактором данных, вы можете в любой момент вернуться к окну создания новой переменной и редактировать любое из ее свойств, например, добавить метку к новому значению или переименовать саму переменную, изменить тип данных и т. д. Кроме того, при дальнейшей работе можно всегда получить полную информацию о любой из имеющихся в файле данных переменных через строку меню главного окна, выбрав опцию **Utilities** (Сервис) и затем **Variables...** (Переменные).

В заключение кратко обобщим последовательность обязательных действий по созданию новой переменной в редакторе данных SPSS:

1. Дайте название новой переменной.
2. Задайте тип данных, которые будет содержать переменная.
3. Подпишите метки к самой переменной и содержащимся в ней значениям.
4. Укажите, какие значения в переменной необходимо считать пропущенными.
5. Выберите и укажите, в какой измерительной шкале представлены значения переменной.

<sup>3</sup> Особенности и отличия измерительных шкал подробно описаны в литературе по экспериментальной психологии, психометрии и психодиагностике (см., напр.: Дружинин В. Н. Экспериментальная психология. М., 1997. С.179–189).

После того как вы создали переменную, можно приступить к «забиванию данных» — к непосредственному вводу значений в переменные файла данных SPSS. 2c

В SPSS по умолчанию предусмотрен ввод новых значений столбцом, то есть сначала вы создаете одну переменную и вводите в нее данные по всем испытуемым, затем создаете вторую переменную и также включаете в нее данные. Такой ввод данных наиболее удобен, если вы имеете под рукой уже готовую «бумажную» таблицу, в которую сведены данные по всем испытуемым. Однако такой способ ввода данных не очень удобен с точки зрения практической обработки, например разрозненных бумажных анкет или тестовых протоколов. В этом случае нужно сначала создать все необходимые переменные и только после этого можно вводить данные построчно.

Ввод данных в SPSS является стандартной процедурой, универсальной для всех электронных таблиц — необходимо установить курсор на первую ячейку под названием переменной (ячейка станет активной; она будет обведена жирной рамкой), напечатать требуемое число и нажать «ввод». Значение будет добавлено к переменной, а курсор автоматически опустится на строку ниже. Следующая ячейка активизируется и будет готова к вводу данных.

Обратите внимание на важность тщательного выбора наиболее удобного для вашего исследования кодирования вводимых данных. Некоторые из них не требуют дополнительного кодирования (например, суммарный итоговый балл по тесту КОТ или время выполнения теста) и могут вводиться напрямую. Некоторые другие данные требуют простейшего кодирования (например, при вводе повопросного протокола теста Айзенка (EPI) мы создаем 57 переменных, значения которых могут быть «0» — ответ «нет» или «1» — ответ «да»).

Как правило, сложного кодирования требуют вопросы социологических или психологических анкет, в которых в качестве ответов на один вопрос испытуемые имеют возможность выбирать несколько вариантов ответов из предложенного списка. В данном случае, в зависимости от глубины и целей исследования, данные можно кодировать двумя способами:

1. Каждый из вариантов ответа на вопрос рассматривается как отдельная переменная, которая может принимать одно из двух значений

(«0» — вариант не был выбран, «1» — вариант ответа был выбран). Подобный способ позволяет получить точную и подробную информацию, однако количество анализируемых переменных возрастает в несколько раз.

- Каждый из вариантов ответа не рассматривается как переменная, а все выбранные ответы кодируются одним числом. Например, из шести возможных ответов испытуемый выбрал 1, 4 и 6. Тогда в соответствующую ячейку вводится число 146. Такой способ упрощает анализ данных, но несколько усложняет работу с результатами обработки и отчетом. Кроме того, количество вариантов ответа не должно быть слишком большим.

**Примечание.** При втором способе кодирования нельзя использовать ноль. Также желательно, чтобы число вариантов ответов не превышало девяти. Кроме того, варианты ответов вводятся одним числом без запятых и цифры располагаются в порядке возрастания.

Пример готового файла данных можно увидеть на рис. 10:

	фio	пол	возраст	дата	вопрос1	кот
1	Иванов Иван Иванович	1	21,00	Дата прохождения теста	2	16,00
2	Петров Петр Петрович	1	18,00	21.09.99	4	24,00
3	Сидоров Сидор Сидорович	1	19,50	03.09.98	5	44,00
4	Михайлова Анна Ивановна	2	23,00	13.10.99	125	31,00
5	Николаева Анна Петровна	2	55,00	09.06.99	7	12,00
6	Сидорчук Петр Григорьевич	1	62,00	11.11.99	.	10,00
7	Федорова Анна Федоровна	2	48,00	20.02.99	24	17,00

Рис. 10. Пример готового файла данных SPSS

В этом примере переменная ФИО является текстовой (String), при этом была указана ширина (Characters) поля 30 (по умолчанию было 8, то есть в ячейку поместилось бы всего 8 букв).

Переменная Дата является временной (Data: день, месяц, год). Обратите внимание, что курсор мыши при простом наведении на дата показывает метку переменной Дата прохождения теста.

Переменные Пол, Возраст, Вопрос 1 и КОТ являются числовыми (Numeric), при этом у переменных Пол и Вопрос 1 исключены десятичные разряды (формат Numeric8.0 вместо принятого по умолчанию Numeric8.2).

Переменная Вопрос 1 является иллюстрацией к способу кодирования данных анкетных вопросов с многовариантным выбором. Так, результат, отображенный в активной ячейке (число 125), означает, что испытуемый Михайлова А. И. в качестве ответов на 1 вопрос анкеты выбрала варианты 1, 2 и 5.

Обратите внимание, что ячейка 6: вопрос1 является пустой, пропущенной (system missing) и в дальнейшей обработке программой учитываться не будет. При нажатии на кнопку «показать/скрыть метки данных» (кнопка 15 на рис. 4) можно также будет увидеть, что в переменной «Пол» цифре 1 соответствует комментарий «мужской», а цифре 2 — «женский».

После окончания ввода всех данных не забудьте сохранить полученный файл. Для этого в строке меню выберите опцию File (Файл), далее Save (Сохранить), после чего в стандартном для всех Windows-программ окне сохранения введите имя файла (можно по-русски) и нажмите кнопку ОК.

По умолчанию файл будет сохранен в файловом формате данных SPSS (файлы с расширением \*.SAV). Данный формат является весьма удобным в силу своей компактности, т. к. даже при очень больших объемах информации размер файлов данных SPSS невелик.

При работе с файлами данных SPSS на разных компьютерах необходимо помнить о так называемой вертикальной совместимости файлов данных. Это означает, что программы SPSS более новых версий «понимают» файлы SPSS более старых версий. И наоборот, старые версии SPSS «не понимают» файлы данных, созданные в более новых версиях программы.

**Примечание, форматы файлов данных в SPSS 8, SPSS 9 и SPSS 10 одинаковы.**

Если вы хотите сохранить файл данных в формате более ранней версии SPSS или даже в другом формате, необходимо в строке меню выбрать опцию File (Файл), далее Save as (Сохранить как), после чего можно выбрать новый формат файла, ввести имя файла (лучше



по-английски, т. к. другие программы могут не поддерживать кириллицу), и затем нажать кнопку ОК.

В программе SPSS заложена возможность экспорта файла данных в 13 различных форматах, включая такие распространенные, как Excel (\*.XLS), dBase (\*.DBF) и др. Однако без крайней необходимости лучше не использовать данные форматы, т. к. это обычно связано с потерей какой-либо информации из файла данных SPSS (например, меток значений).

## ИМПОРТ ДАННЫХ В SPSS

Как показывает многолетний опыт работы автора со статистическими программами, эти программы являются гораздо менее распространенными и гораздо более громоздкими, чем стандартные офисные электронные таблицы (обычно это MS Excel). В подавляющем большинстве случаев пользователям гораздо удобнее создать таблицу данных, например, в MS Excel, а потом уже импортировать эти данные в SPSS.

В программах SPSS, начиная с 8 версии и выше, существует возможность как простого, так и сложного импорта таблиц данных, созданных в других форматах (который в этом пособии не рассматривается). Простой импорт данных осуществляется через строку меню, где необходимо выбрать опцию File (Файл), далее — Open (Открыть), после чего указать тип (формат) открываемого файла (все-го SPSS по умолчанию поддерживает импорт около 10 различных форматов), указать имя файла и нажать кнопку ОК. Данные из импортируемого файла окажутся в редакторе данных SPSS, при этом программа автоматически сгенерирует и откроет в окне вывода результатов (SPSS Output) отчет об успешности или неуспешности процедуры переноса данных.

Наиболее удобными для простого импорта данных являются таблицы в формате MS Excel (\*.XLS), а для пользователей очень старых компьютеров — в текстовом формате (\*.DAT, \*.TXT). На рис. 11-13 показан процесс импорта файла данных MS Excel в редактор данных SPSS.

Прежде всего необходимо создать таблицу данных в программе MS Excel. Обратите внимание, что таблица должна иметь простую структуру (не допускаются «двухэтажные» и сложные заголовки пе-

	A	B	C	D	E	F
1	нейротизм	экстраверсия	невротизм	страхи	тревожность 1	тревожно
2		15	12	13	23	43
3		20	9	0	2	45
4		14	12	26	28	13
5		17	10	7	14	7
6		14	14	8	15	35
7		23	15	31	25	30
8		13	11	24	12	55
9		22	13	32	24	60

Рис. 11. Исходный файл данных в программе MS Excel

ременных), все данные должны быть размещены на одном листе рабочей тетради MS Excel.

После создания таблицы надо выбрать в строке меню MS Excel опцию **Файл**, затем **Сохранить как** и выбрать формат сохранения **Файл Microsoft Excel 4.0**.

Далее необходимо запустить программу SPSS, выбрать опцию **File** (Файл), потом **Open** (Открыть), указать, что будет импортироваться файл в формате MS Excel (\*.XLS), указать имя файла и нажать **OK**.

После этого программа откроет окно импорта файла (**Opening File Options**), показанное на рис. 12.

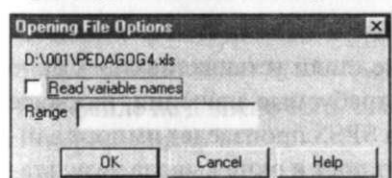
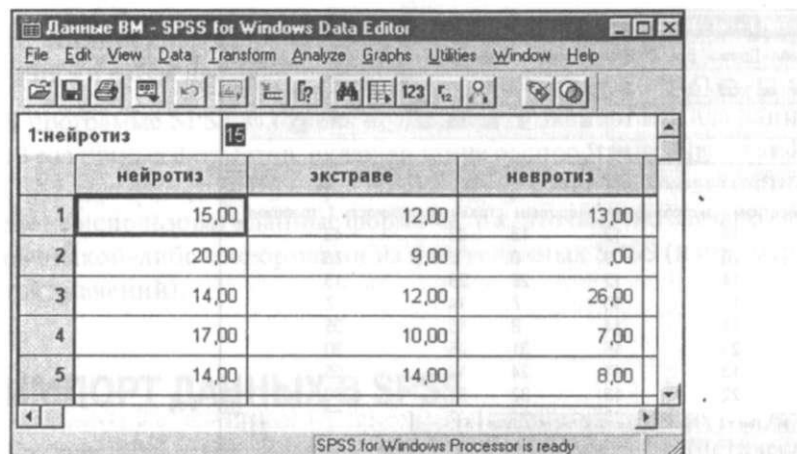


Рис. 12. Окно импорта файла данных в SPSS

Обратите внимание на две опции в окне импорта файла: **Read variable names** (Читать названия переменных), **Range** (Импортировать диапазон). Если вы поставите флажок в опции **Read variable names** (Читать названия переменных), то программа SPSS автоматически перенесет верхнюю текстовую строчку таблицы MS Excel как название переменных в SPSS. Обратите внимание (рис. 13), что при этом текстовые на-



The screenshot shows the SPSS Data Editor window titled 'Данные ВМ - SPSS for Windows Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a table with the following data:

	нейротиз	экстраве	невротиз
1	15,00	12,00	13,00
2	20,00	9,00	,00
3	14,00	12,00	26,00
4	17,00	10,00	7,00
5	14,00	14,00	8,00

The status bar at the bottom indicates 'SPSS for Windows Processor is ready'.

Рис. 13. Результаты импорта файла MS Excel в SPSS

звания из таблицы MS Excel в наименования переменных SPSS будут сокращены до 8 букв. По умолчанию программа SPSS дает импортируемым переменным названия `var00001, var00002` и т. д.

В некоторых случаях не требуется импортировать в SPSS таблицу данных из MS Excel целиком, а только ее фрагмент. Для этого можно воспользоваться опцией Range (Импортировать диапазон). В строке данной опции необходимо указать координаты левой верхней и правой нижней точек импортируемого прямоугольного фрагмента данных MS Excel. Например, указанные в строке опций значения `A3: C8` позволяют импортировать фрагмент общей таблицы размером 3 на 6. Если вы импортируете данные целиком, то в этой строке ничего указывать не нужно.

После того как вы установили (или не стали устанавливать) в окне импорта файла (Opening File Options) требуемые значения, нажмите на кнопку ОК. После этого программа SPSS произведет импорт данных и автоматически сгенерирует и откроет в окне вывода результатов (SPSS Output) отчет (Log) об успешности или неуспешности процедуры переноса данных. Для того, чтобы вернуться из окна вывода результатов в окно редактора данных, воспользуйтесь опцией Window (Окно) в строке меню.

После того как таблица данных из MS Excel была успешно импортирована в редактор данных SPSS (см. рис. 13), необходимо отредакти-

ровать каждую из новых переменных, задав ее тип, метки данных, пропущенные значения и т. д. (см раздел «Создание нового файла данных»). В заключение необходимо сохранить импортированный файл в формате SPSS (с расширением \*.SAV).

**Примечание.** Исходный файл MS Excel обязательно должен быть сохранен в формате Файл Microsoft Excel 4.0. Программа SPSS вплоть до 10 версии не способна импортировать файлы версий Excel 5.0 и выше без дополнительно установленного драйвера. Даже при

- дополнительно установленном драйвере данные Excel 5.0 и выше переносятся в SPSS с помощью процедуры сложного импорта (Database Capture), которая в данном пособии рассматриваться не будет.

Если вы являетесь несчастливым владельцем совсем уж старого и маломощного компьютера, то подготовить файл данных для импорта в SPSS можно в любом текстовом редакторе под MS DOS или MS Windows (см. рис. 14).

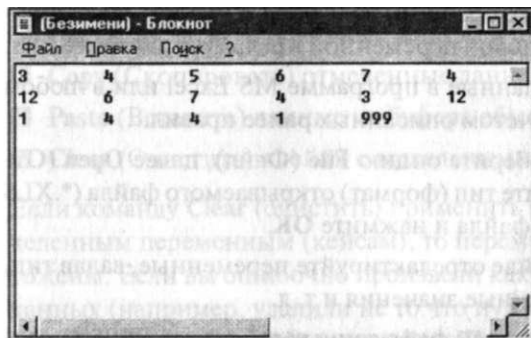


Рис. 14. Исходный файл данных в программе Блокнот (Notebook)

Создание файла данных в текстовом редакторе требует соблюдения нескольких важных правил:

1. В файле не должно быть текстовых надписей, в том числе «шапок» переменных;
2. Таблица должна быть прямоугольной, то есть строки должны быть одинаковой длины и не должно быть пропусков (пустых мест) внутри строк. Обратите внимание на число **999** в приведенном примере (рис. 14). Данное число (или любое другое, значительно отличающееся от основных используемых значений) стоит на месте пропущенного значения. Потом, уже после импортирования таблицы, при редактировании переменных необходимо выбрать опцию **Missing**

Values... (Пропущенные значения), далее указать Discrete missing values (Дискретные пропущенные значения) и напечатать там 999;

3. Ввод данных осуществляется построчно следующим образом: печатается число, затем нажимается клавиша табуляции (Tab): число — табуляция — число — табуляция и т. д. Только после того, как вы напечатали последнее в строке число, нажимаете на Enter (ввод) и переходите к набору следующей строки.

Сохранять подобный исходный файл данных необходимо с расширением \*.DAT или \*.TXT (по умолчанию). Последовательность процедуры импорта аналогична описанной ранее. После того как текстовый файл данных был успешно импортирован в редактор данных SPSS, необходимо отредактировать каждую из новых переменных, задав ее тип, метки данных, пропущенные значения и т. д. (см раздел «Создание нового файла данных»). Теперь нужно сохранить импортированный файл в формате SPSS (с расширением \*.SAV).

В заключение давайте кратко обобщим последовательность обязательных действий по импорту новой переменной в редактор данных SPSS:

1. Подготовьте исходные данные в программе MS Excel или в любом текстовом редакторе с учетом описанных ранее правил.
2. В строке меню SPSS выберите опцию File (Файл), далее Open (Открыть), после чего укажите тип (формат) открываемого файла (\*.XLS или \*.TXT, \*.DAT), имя файла и нажмите ОК.
3. В импортированном файле отредактируйте переменные, задав тип, метки данных, пропущенные значения и т. д.
4. Сохраните импортированный файл данных в формате SPSS (с расширением \*.SAV).

## РЕДАКТИРОВАНИЕ ТАБЛИЦЫ ДАННЫХ

В ряде случаев, после того как вы создали (или импортировали) файл данных SPSS, возникает необходимость внесения изменений в таблицу данных. Основные возможности по редактированию данных представлены в строке главного меню SPSS в опциях Edit (Правка), Data (Данные), Transform (Трансформация).

Опция Edit (Правка) содержит в себе стандартные возможности редактирования, которые можно использовать после выделения про-

извольного массива данных, а также одной или нескольких переменных. Выделение производится следующим образом:

1. Курсор устанавливается на нужную ячейку, далее, нажимая на клавишу **Shift**, необходимо с помощью стрелок-указателей растянуть выделение до требуемого размера.
2. Курсор устанавливается на нужную ячейку, далее, удерживая левую кнопку мыши, растягиваем выделение до требуемого размера.
3. Для выделения переменной щелкаем мышью на названии переменной, для выделения кейса — на номере кейса.

**Примечание.** В редакторе данных можно выделить сразу несколько переменных или кейсов, расположенных рядом, но отсутствует возможность выделения разрозненных фрагментов данных, как, например, это можно сделать в MS Excel.

В опции **Edit** (Правка) к выделенным данным можно применить следующие операции:

- Cut** (Вырезать) отмеченные данные в буфер обмена данными;
- Copy** (Скопировать) отмеченные данные в буфер обмена;
- Paste** (Вставить) данные из буфера обмена;
- Clear** (Очистить) ячейки с выделенными данными.

Если команду **Clear** (очистить) применить не к массиву ячеек, а к выделенным переменным (кейсам), то переменные (кейсы) будут уничтожены. Если вы ошибочно произвели какую-либо операцию правки данных (например, удалили не то что нужно), то вы можете воспользоваться командой **Undo** (отменить). Помните, что команда **Undo** исправляет только последнюю операцию с данными!

В опции **Edit** (Правка) также содержатся команды **Find** (Найти) и **Options** (Опции). С помощью операции **Find** в рамках указанной переменной вы можете найти любое интересующее вас значение. Данная операция имеет смысл, если а) вы работаете с очень большим файлом данных, б) если значения внутри кейсов достаточно разнообразны. В пункте **Options** (Опции) содержатся системные настройки SPSS, рассчитанные на продвинутого пользователя (в данном руководстве они рассматриваться не будут).

В опции **Data** (Данные) содержится ряд возможностей по редактированию данных:

- Define Variable...** (Создать новую переменную). Эта операция уже подробно рассматривалась ранее в разделе «Создание нового файла данных»).
- Define Dates...** (Определить формат временной переменной). Этот пункт опции применим только к переменным, которые содержат в себе информацию о времени (датах, часах, минутах и т. п.).
- Templates...** (Задать шаблоны переменных).
- Insert Variable** (Вставить переменную).
- Insert Case** (Вставить кейс).
- Go to Case** (Перейти к кейсу).
- Sort Cases...** (Сортировать кейсы).
- Transpose...** (Транспонировать данные).
- Merge Files** (Присоединить файлы).
- Aggregate** (Агрегировать данные).
- Split File** (Расщепить файл).
- Select Cases...** (Выбрать кейсы).
- Weight Cases...** (Взвесить кейсы).

Рассмотрим некоторые из этих возможностей подробнее.

Если вы постоянно работаете с SPSS на одном и том же компьютере и часто создаете новые файлы данных с однотипными переменными, то вашу работу сможет значительно ускорить использование готовых шаблонов переменных.

Сначала шаблон переменной необходимо создать. Для этого войдите в **Data** (Данные), выберите **Templates...** (Шаблоны переменных) и далее в появившемся окне **Template** (Шаблон) нажмите на кнопку **Define >>** (Определить). После этого раскроется дополнительное окно **Define Template** (Определить шаблон), содержащее в себе стандартные кнопки задания переменной: **Type...** (Тип переменной), **Value labels...** (Метки данных) и др., рассмотренные ранее в разделе «Создание нового файла данных».

После того как вы указали все необходимые параметры в подокне **Define Template** (Определить шаблон), необходимо в окне **Name** (Имя) написать название вашего шаблона (в примере на рис. 15 новое название MYDATA1) и нажать на кнопку **Add** (Добавить). Созданный вами

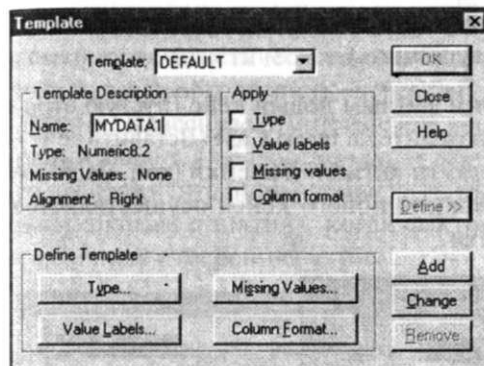


Рис. 15. Окно создания шаблона переменных

новый шаблон появится в списке доступных шаблонов (опция **Template**).

Для того чтобы применить шаблон к переменной, необходимо правой кнопкой мыши кликнуть (щелкнуть) по названию («шапке») переменной и в появившемся меню выбрать пункт **Templates...** (Шаблоны). В открывшемся окне (см. рис. 15) в опции **Template** (Шаблон) нужно выбрать название созданного вами шаблона. После этого в группе элементов управления **Apply** (Применить) установите флажки напротив одного или нескольких параметров шаблона, которые вы хотите присвоить новой переменной (тип переменной, метки данных, формат пропущенных значений и т. д.). Далее нажмите кнопку **OK** для завершения операции.

Шаблон можно использовать и следующим способом: установите курсор на любой ячейке новой переменной, к которой вы хотите применить шаблон, войдите в **Data** (Данные), выберите **Templates...** (Шаблоны переменных), и далее повторите действия, описанные в предыдущем абзаце.

Для добавления новых значений в редактируемом файле данных существуют две типовые операции: **Insert Variable** (Вставить переменную) и **Insert Case** (Вставить кейс).

Если вы хотите добавить к своим данным новую переменную (пустой столбец), войдите в **Data** (Данные) и нажмите на **Insert Variable** (Вставить переменную). Помните, что новая переменная будет вставлена *слева* от активной ячейки (той, на которой установлен курсор). По умолчанию новой переменной присваивается имя var00001.



В дальнейшем новую переменную необходимо отредактировать (см. раздел «Создание нового файла данных»).

Если вы хотите добавить к своим данным новый кейс (пустую строку), войдите в Data (Данные) и нажмите на Insert Case (Вставить кейс). Новый кейс будет вставлен *сверху* от активной ячейки (той, на которой установлен курсор). Нумерация нового кейса осуществляется программой SPSS автоматически.

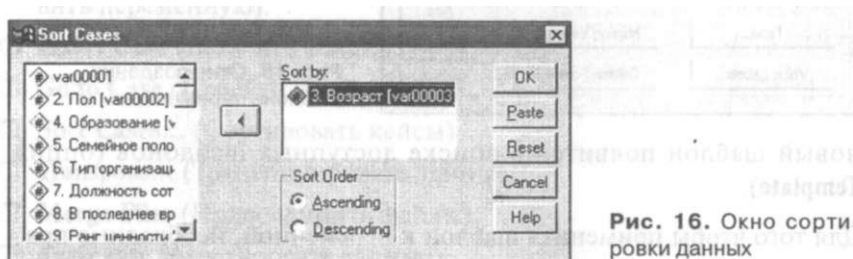


Рис. 16. Окно сортировки данных

Если вам нужно отсортировать данные, вы можете воспользоваться опцией **Sort Cases...** (Сортировать кейсы) (рис. 16).

Войдите в **Data** (Данные), нажмите на **Sort Cases...** (Сортировать кейсы), после чего откроется окно сортировки данных. Выберите в левом поле одну или несколько переменных, по которым вы хотите отсортировать данные и добавьте их в поле **Sort by:** (Сортировать по). После этого в опции **Sort Order** (Порядок сортировки) установите флажок. Данные можно отсортировать как по возрастанию (**Ascending**) — от меньшего значения к большему, так и по убыванию (**Descending**). Для завершения операции нажмите кнопку **OK**.

Если вы добавите в поле **Sort by:** (Сортировать по) несколько переменных, то данные будут *последовательно* отсортированы по всем этим переменным. То есть сначала программа отсортирует данные по переменной, которая была добавлена первой, потом — по второй, по третьей и т. д.

**Примечание.** Процедура сортировки данных является весьма распространенной в прикладной психологической статистике, например при составлении простых или комплексных рейтингов. Однако следует учитывать, что сортировка данных в большом количестве переменных является весьма ресурсоемкой процедурой и на маломощном компьютере может привести к зависанию программы.

В некоторых случаях, например, при использовании процедур многомерного статистического анализа (кластерного, факторного и др.), возникает необходимость «перевернуть» (транспонировать) матрицу данных. Классическим примером востребованности подобной операции является обработка «куба» данных после психосемантического эксперимента (испытуемые) × (шкалы СД) × (оцениваемые по данным шкалам понятия), когда факторизация может проводиться в различных измерениях «куба».

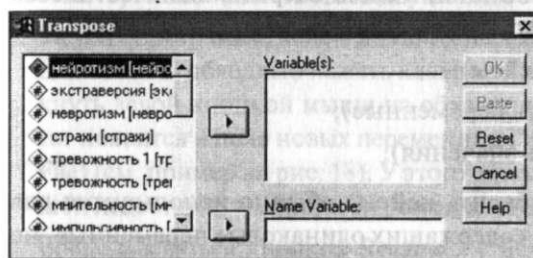


Рис. 17. Окно транспонирования данных

Для транспонирования данных необходимо войти в **Data** (Данные), выбрать **Transpose...** (Транспонировать данные), после чего откроется окно транспонирования данных (см. рис. 17).

Далее выберите в левом поле переменные, которые вы хотите транспонировать и переместите их с помощью кнопки-указателя (с треугольником) в поле **Variable(s):** (Переменные). Если среди ваших переменных есть текстовая (например, ФИО испытуемых), то ее необходимо переместить в поле **Name Variable** (Названия переменных). Если подобная переменная отсутствует, то данное поле можно оставить пустым.

После этого нажмите кнопку **OK**, и программа SPSS перевернет вашу матрицу. При этом в окне вывода результатов (**SPSS Output**) появится отчет (**Log**) об успешности или неуспешности процедуры транспонирования. Для того чтобы вернуться к файлу данных, воспользуйтесь опцией **Window** (Окно) в строке меню.

Перевернутая матрица — это новый файл данных (исходный файл при этом остается без изменений), который необходимо назвать и сохранить. В результате транспонирования первая переменная (столбец) становится первым кейсом (строкой), вторая переменная — вторым кейсом и т. д. В новом файле первая переменная автоматически

получает название `case_1b1` и содержит в себе старые названия переменных (до транспонирования). Новые переменные (бывшие кейсы) автоматически получают названия `var001`, `var002` и т. д. Разумеется, для удобства дальнейшей работы переменные можно переименовать и отредактировать.

Если вам необходимо объединить данные из двух различных файлов данных SPSS, вы можете воспользоваться опцией **Merge Files** (Присоединить файлы). Данная опция, в свою очередь, включает в себя три дополнительные опции:

- Add cases...** (Добавить кейсы);
- Add Variables...** (Добавить переменные);
- Update Values** (Обновить значения).

Операция **Add cases...** (Добавить кейсы) обычно используется для объединения двух файлов, содержащих одинаковые переменные, но различные кейсы. Например, в одном файле у вас хранятся результаты тестирования детей в 6А классе, а в другом файле — результаты тестирования по тем же самым методикам в 6Б классе.

Для начала необходимо нажать на опцию **Add cases...** (Добавить кейсы), затем указать файл, из которого будут добавляться кейсы, после чего откроется нужное окно (см. рис. 18). Слева в окне находится перечень непарных переменных (**Unpaired Variables**) — это все переменные из двух объединяемых файлов, названия которых не совпадают. При этом названия переменных из главного файла (к которому добавляем кейсы) в скобках помечены звездочкой (\*); названия переменных из дополнительного файла (из которого добавляем кейсы) в скобках помечены плюсом (+).

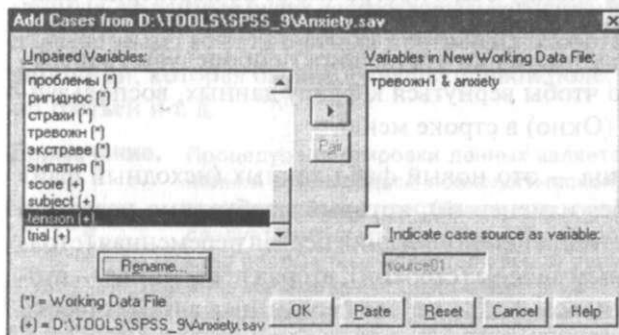


Рис. 18. Окно добавления кейсов

В правой половине окна добавления кейсов необходимо указать, какие именно переменные войдут в новый файл данных (Variables in New Working Data File).

Будьте внимательны, т. к. зачастую названия переменных в файлах данных отличаются (в приведенном на рис. 18 примере есть названия переменных как на русском, так и на английском языках), и в этом случае программа объединит данные некорректно. Чтобы этого избежать, можно действовать двумя способами:

1. Указать программе, кейсы из каких переменных следует объединять. Для этого необходимо нажать клавишу Control и, удерживая ее, щелкнуть левой кнопкой мыши на объединяемых переменных. Результат появится в поле новых переменных (Variables in New Working Data File) (см. пример на рис. 18). У этого способа есть лишь один небольшой недостаток — слишком громоздкое название новой переменной.
2. Воспользоваться кнопкой Rename... (Переименовать), находящейся под полем непарных переменных. Обычно переименовывается дополнительная переменная, ей дается такое же имя, как у основной переменной. После этого повторите операцию объединения (см. выше).

Если вы хотите включить в новый файл данных какие-либо переменные, не объединяя их, то для этого просто выберите необходимые названия в списке непарных переменных (Unpaired Variables:) и с помощью стрелки переместите их в поле новых переменных (Variables in New Working Data File).

Для завершения процедуры нажмите ОК.

**Примечание.** Объединять можно только кейсы из переменных одинакового формата. Например, нельзя объединить числовые переменные с текстовыми названиями. Помните также, что включение в новый файл не объединенных переменных приведет к появлению большого числа «дырок» — пропущенных значений, что может в дальнейшем затруднить обработку данных.

Операция Add Variables... (Добавить переменные) обычно используется для объединения двух файлов, содержащих одинаковые кейсы, но различные переменные. Например, в одном файле у вас хранятся результаты тестирования детей в 6А классе по тесту Кеттела, а в другом файле — результаты тестирования того же самого 6А класса по тесту Айзенка.

Операция добавления переменных аналогична процедуре добавления кейсов, и здесь подробно рассматриваться не будет. Необходимо только учитывать, что для корректного выполнения данной операции оба файла должны содержать одинаковое количество кейсов и все кейсы должны быть отсортированы в одинаковом порядке. Например, при объединении результатов двух тестирований одного и того же класса 1) количество детей в первом и втором тестировании должно быть равным; 2) это должны быть одни и те же дети; 3) они должны быть расположены в том же порядке, например по алфавиту. Только при соблюдении этих правил программа корректно объединит данные.

Опция *Aggregate* (Агрегировать данные) позволяет в значительной мере сжать файл данных за счет группировки и объединения исходных значений по определенным правилам. Фактически, использование этой операции можно считать своеобразным микростатистическим анализом.

Если вы выбираете опцию *Aggregate* (Агрегировать данные), то автоматически открывается окно агрегирования данных (см. рис. 19). Сначала необходимо указать одну или несколько переменных, которые не будут агрегированы (находящиеся в них данные останутся без изменений), но станут основанием для объединения данных в других переменных.

Для этого в левом поле выберите одну или нескольких переменных (в примере на рис. 19 это переменная «возраст») и с помощью кнопки с треугольником переместите их в поле *Break Variables* (Отброшенные переменные). После этого в левом поле выберите одну или несколько переменных, данные в которых необходимо объединить (на рис. 19 это переменная «пед. стаж»). С помощью кнопки с треугольником переместите их в поле *Aggregate Variables* (Агрегированные переменные). В этом поле автоматически появится название новой переменной (в примере на рис. 19 это *пед. ст1*) и функция, с помощью которой произойдет агрегирование данных (в данном примере это *MEAN* — подсчет средней арифметической). Если вас не устраивает автоматически составленное название *пед. ст1*, его можно исправить, воспользовавшись кнопкой *Name & Label* (Название и метка).

По желанию в отдельной переменной (по умолчанию — *N\_BREAK*) можно сохранить количество кейсов, участвовавших в агрегирова-

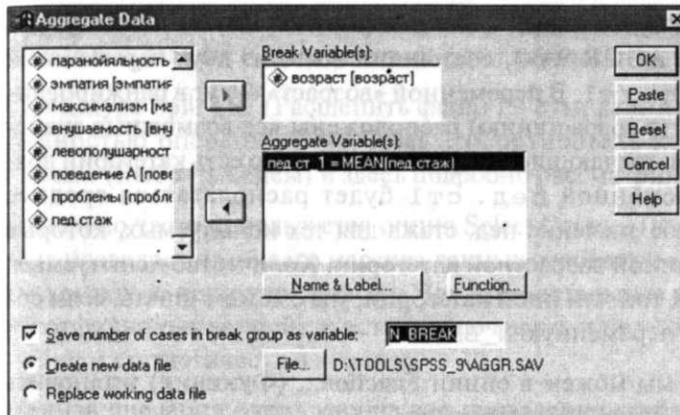


Рис. 19. Окно агрегирования данных

нии данных по каждой из категорий. Для этого необходимо установить флажок в опции **Save number of cases in break group as variable** (Сохранить число кейсов в категориях отброшенной переменной).

Обратите внимание, что объединить данные можно с использованием различных функций (по умолчанию это будет подсчет среднего арифметического значения).

Нажав на кнопку **Function...** (Функция), вы попадете в окно выбора агрегирующей функции (рис. 20).

Для того чтобы правильно воспользоваться функциями агрегирования, необходимо хорошо понимать саму процедуру объединения данных. В рассматриваемом здесь простом примере (рис. 19) результатом

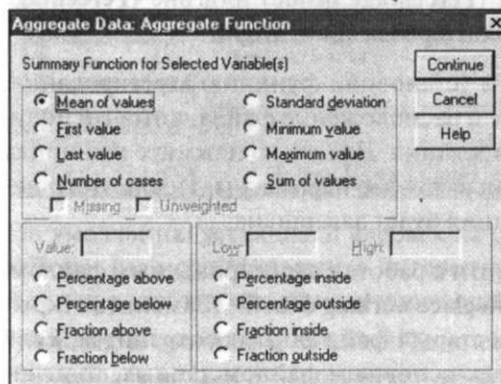


Рис. 20. Окно указания функции объединения данных

агрегирования будет новый файл данных (по умолчанию получающий название AGGR.SAV), состоящий всего из двух переменных: возраст и пед. ст1. В переменной «возраст» будут в ранжированном порядке (по возрастанию) расположены все возможные значения возраста, встречающиеся в файле данных (то есть категории возраста). В переменной пед. ст1 будет располагаться среднее арифметическое значение пед. стажа для тех испытуемых, которые относятся к данной возрастной категории. Количество испытуемых, относящихся к той или иной категории, мы сможем узнать, если создадим новую переменную N\_BREAK.

При желании мы можем в опции **Function...** (Функции) установить следующие функции агрегирования:

- среднее арифметическое (**Mean of values**);
- первое значение (**First value**);
- последнее значение (**Last value**);
- количество пропущенных или невзвешенных кейсов (**Number of cases**);
- стандартное отклонение (**Standard deviation**);
- минимальное значение (**Minimum value**);
- максимальное значение (**Maximum value**);
- сумма значений (**Sum of values**).

Также можно подсчитать процент кейсов, имеющих значение выше (**Percentage above**) или ниже (**Percentage below**) определенного числа (**Value**) находящихся внутри (**Percentage inside**) или вне (**Percentage outside**) заданного числового интервала (**Low, High**).

После того как вы выбрали и установили функцию агрегирования, необходимо указать название и расположение файла, который получится после операции агрегирования. Для этого нажмите на кнопку **File...** (Файл) и введите удобные для вас параметры. После этого нажмите на кнопку **ОК** и операция будет завершена.

Если вы хотите сразу же перейти к работе с агрегированным файлом, установите флажок в опции **Replace working data file** (Заменить открытый файл данных). При этом старый файл данных сохранится. Если вы хотите продолжить работу с основным файлом данных, то игно-

рируйте эту опцию, т. к. по умолчанию в ней установлен параметр Create new data file (Создать новый файл данных).

Процедура Split File (Расщепить файл) по сути дела является разновидностью операций Sort Cases... (Сортировать кейсы) и Select Cases... (Выбрать кейсы) и здесь подробно рассматриваться не будет.

Довольно часто используется опция Select Cases... (Выбрать кейсы). Она позволяет из общего массива данных отобрать по определенному критерию подгруппу кейсов. При этом остальные кейсы временно отфильтровываются, то есть выключаются и не участвуют в дальнейшем статистическом анализе.

Данная операция очень удобна для проведения глубокого и разно-стороннего статистического анализа данных, т. к. с ее помощью возможно выделение практически любого подмножества интересующих исследователя данных.

Щелкнув мышью на надписи Select Cases... (Выбрать кейсы), вы попадаете в окно выбора кейсов (рис. 21).

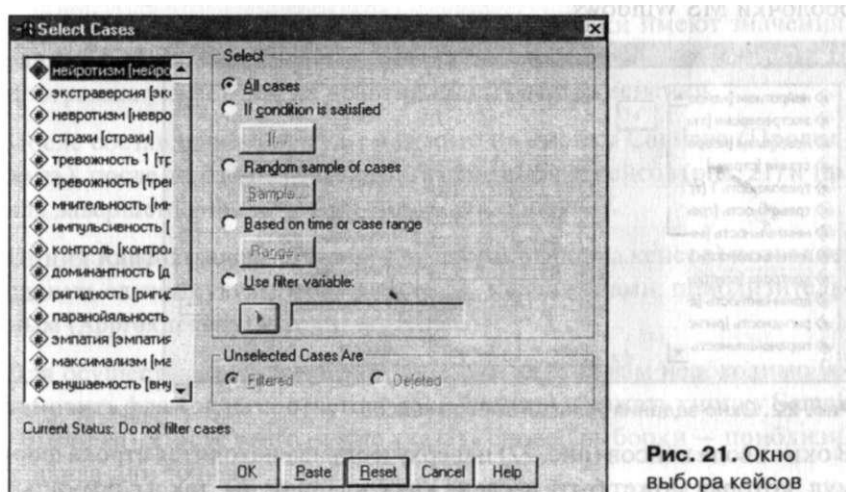


Рис. 21. Окно выбора кейсов

По умолчанию установлен флажок на опции All cases (Выбрать все кейсы). Вы можете отметить интересные вас данные, установив флажок напротив одной из четырех опций:

1. Опция If condition is satisfied (Если соблюдены условия) является наиболее мощной и позволяет делать сложные выборки данных



с использованием арифметических, логических, статистических и других функций.

2. Опция Random sample of cases (Случайная выборка кейсов) позволяет выбрать случайным образом заданное количество кейсов.
3. Опция Based on time or case range (Выборка на основе временного или числового интервала) позволяет выбрать кейсы, попадающие в диапазон между указанными минимальным и максимальным значениями.
4. Опция Use filter variable (Использовать переменную в качестве фильтра) является упрощенным аналогом опции Sort Cases... (Сортировать кейсы) и в данном руководстве рассматриваться не будет.

Если вы хотите сделать сложную выборку данных, то вам необходимо установить флажок на опции If condition is satisfied (Если соблюдены условия) и нажать на кнопку If... (Если).

После этого откроется окно задания условий выбора кейсов (рис. 22), несколько напоминающее стандартную программу «Калькулятор» оболочки MS Windows.

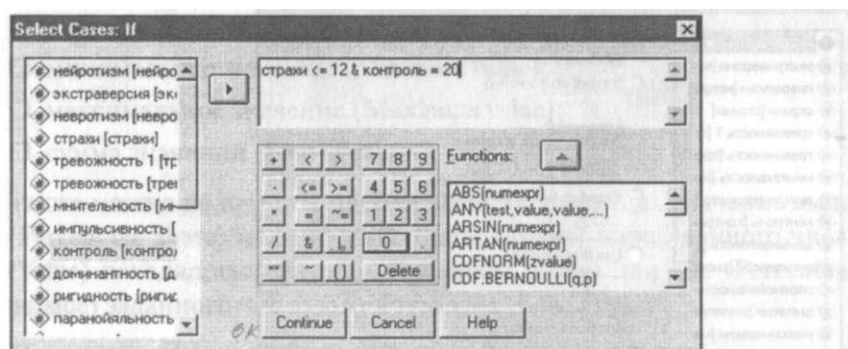


Рис. 22. Окно задания условий выбора кейсов

В окне ввода кейсов (рис. 22) наверху и справа находится строка формул, которая может быть введена как с клавиатуры, так и с помощью мыши с использованием списка переменных (слева), кнопок арифметических операций (под строкой формул) и списка функций (под строкой формул справа).

Возможности ввода формул практически не ограничены (в одном только списке функций их около 50!), однако требуют специальных

знаний по составлению формул и макрокоманд SPSS. Обычно при установке SPSS также устанавливается справочное руководство по макроязыку SPSS (*SPSS Syntax Guide*), в котором содержится описание функций и синтаксических операций, и оно может быть изучено самостоятельно продвинутыми пользователями.

На рис. 22 в строке формул приведена простейшая формула

страхи <=12 & контроль =20,

которая означает, что в дальнейшем будут отфильтрованы (исключены из анализа) все кейсы, которые подходят под следующие два условия:

1. Их значение по переменной *страхи* равно 13 и более.
2. Их значение по переменной *контроль* не равно 20.

Подобные формулы пользователь может составлять самостоятельно в неограниченных вариантах, даже для осуществления «точных» выборок конкретных значений. Например, формула *страхи = 5* or *страхи = 7* or *страхи = 25* (возможна также запись *страхи = 5 I страхи = 7 I страхи = 25*) отфильтрует все кейсы за исключением тех, которые в переменной *страхи* имеют значения, равные 5, 7 или 25. Если же вы некорректно составили формулу, то программа укажет вам на невозможность ее применения.

После составления формулы нажмите на кнопку Continue (Продолжить), после чего вновь окажетесь в окне выбора кейсов (рис. 21) и там для завершения фильтрации нажмите кнопку ОК.

Опция Random sample of cases (Случайная выборка кейсов) позволяет сделать случайную выборку кейсов двумя способами: приблизительным (Approximately) и точным (Exactly).

Для осуществления случайной выборки кейсов вам необходимо установить флажок на соответствующей опции и нажать кнопку Sample (Выборка). После этого нужно указать способ выборки — приблизительная или точная.

Если вы выбрали приблизительную выборку (Approximately), то необходимо указать, сколько процентов от всех кейсов вы хотите выбрать.

Если вы хотите сделать точную выборку (Exactly), то вам необходимо указать, сколько кейсов (обязательно надо написать число), из какого числа (это значение не обязательно) вы хотите отобрать.

Например, использование приблизительной случайной выборки (с параметром 50%) может быть с успехом использовано в психометрии для проверки надежности и валидности теста.

После указания способа случайной выборки нажмите на кнопку **Continue** (Продолжить), после чего вновь окажетесь в окне выбора кейсов (рис. 21) и там для завершения фильтрации нажмите кнопку **OK**.

Если вы хотите осуществить выборку на основе временного или числового интервала (**Based on time or case range**), то после установки соответствующего флажка нажмите на кнопку **Range...** (Диапазон). После этого в появившемся окне укажите номера первого (**First Case**) и последнего (**Last Case**) кейсов в том множестве, которое хотите отсортировать. Все кейсы, которые не войдут в указанный диапазон, будут отфильтрованы.

После указания диапазона выбираемых кейсов нажмите на кнопку **Continue** (Продолжить), после чего вновь окажетесь в окне выбора кейсов (рис. 21) и там для завершения операции нажмите кнопку **OK**.

Процедура **Weight Cases...** (Взвесить кейсы) используется для «улучшения» и подготовки данных для некоторых способов статистического анализа. Например, взвешивание кейсов исключает из обработки нулевые, отрицательные, пропущенные и т. п. данные. (В этом руководстве процедура взвешивания кейсов подробно рассматриваться не будет.)

В опции **Transform** (Трансформация) содержится ряд операций, позволяющих изменить исходные данные:

- Compute...** (Вычислить);
- Random Number Seed...** (Генератор случайных чисел);
- Count...** (Сосчитать);
- Recode** (Перекодировать);
- Categorize Variables...** (Создать категории);
- Rank Cases...** (Ранжировать кейсы);
- Automatic Recode...** (Автоматическое перекодирование);
- Create Time Series...** (Создать временные последовательности);
- Replace Missing Values...** (Заменить пропущенные значения).

Далее будут рассмотрены некоторые операции из опции «Трансформация».

**Compute...** (Вычислить) часто является востребованной, когда необходимо с помощью программы SPSS осуществить математические (арифметические и др.) расчеты с переменными. Например, ваши данные представлены в виде ответов на каждый вопрос теста (каждая переменная — вопрос; в случае ответа «Да» в ячейке стоит цифра 1, в случае ответа «Нет» — цифра 0) и вам необходимо вычислить суммарный балл по тесту для каждого из испытуемых.

В некоторых случаях у исследователя в процессе анализа данных возникает необходимость в создании и вычислении гипотетического коэффициента, например рассчитать среднее арифметическое для нескольких выбранных переменных.

Щелкнув мышью на строке меню **Compute...** (Вычислить), вы вызываете окно (рис. 23) **Compute Variable** (Вычислить переменную).

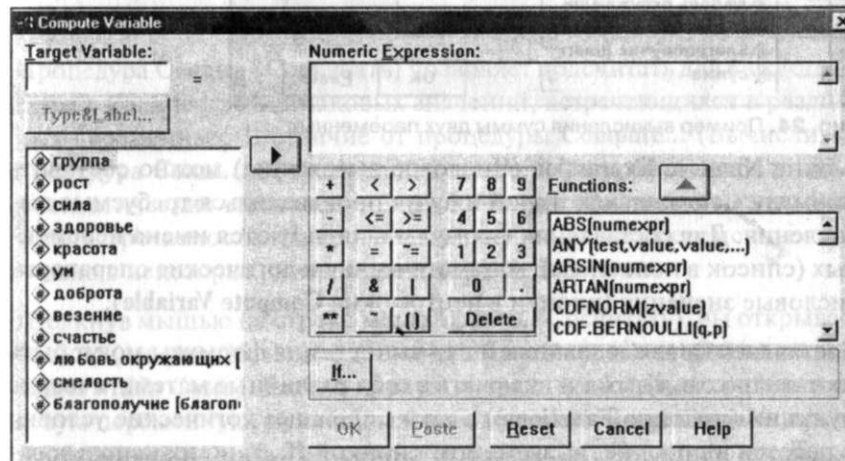


Рис. 23. Окно математических вычислений с переменными

Сначала необходимо задать итоговую переменную (**Target Variable**), в которой будут представлены результаты проведенных расчетов. После ввода имени итоговой переменной, активизируется кнопка **Type & Label** (Тип и метка). Щелкнув по этой кнопке, вы получаете возможность указать свойства новой переменной и прибавить метку к ней.

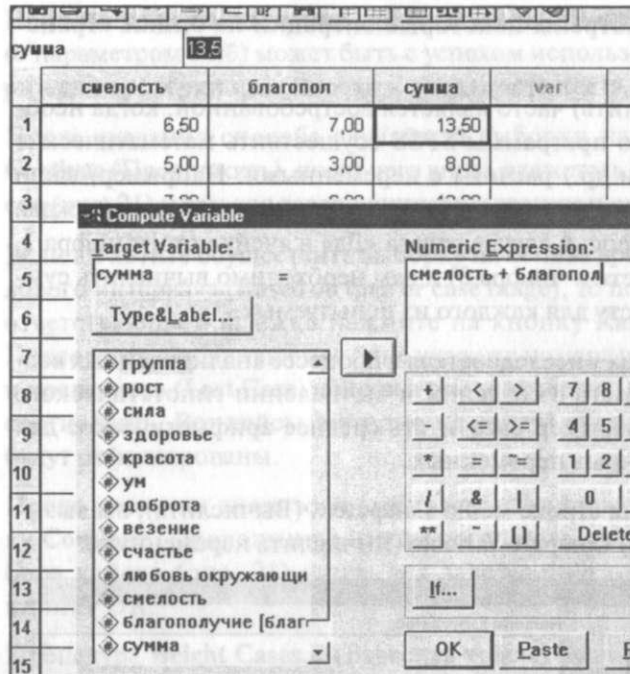


Рис. 24. Пример вычисления суммы двух переменных

В окне **Numeric Expression** (Числовое выражение) можно составить формулу, с помощью которой и будут производиться требуемые вычисления. Для составления формулы используются имена переменных (список в поле слева), математические и логические операторы, числовые значения (кнопки в центре окна **Compute Variable**).

Составляемые исследователем для вычисления формулы могут быть достаточно сложными и включать в себя различные математические функции (подокно **Functions**), а также сложные логические условия (задаются в подокне, вызываемом кнопкой **If...**), но их использование рекомендуется только для опытных пользователей, знакомых с основами программирования. (В данном руководстве эти возможности программы не рассматриваются).

На рис. 24 приводится пример проведения простейшего вычисления суммы между двумя переменными. Итоговая переменная (**Target Variable**) получила название *сумма*, после чего была задана формула для вычисления в такой последовательности:

1. В подокне слева была выбрана переменная смелость и с помощью кнопки-указателя (с треугольником) перемещена в подокно **Numeric Expression** (Числовое выражение).
2. Нажата кнопка «+» в центре окна **Compute Variable**.
3. В левом окне была выбрана переменная благопол и с помощью кнопки-указателя перемещена в подокно «Числовое выражение». После того как в подокне **Numeric Expression** (Числовое выражение) появилась надпись смелость + благопол, была нажата кнопка **ОК**. Результатом расчета являлось создание в таблице данных новой переменной под названием сумма (см. рис. 24), в ячейках которой приведены результаты попарного суммирования значений переменных смелость и благополучие.

Следует добавить, что при составлении формул допустимо использование скобок, в том числе и многократно вложенных. В качестве простейшего примера использования скобок можно привести формулу расчета средней арифметической величины между переменными смелость и благополучие:  $(\text{смелость} + \text{благопол}) / 2$ .

Процедура **Count...** (Сосчитать) позволяет подсчитать для каждого из кейсов количество одинаковых значений, встречающихся в различных переменных. В отличие от процедуры **Compute...** (Вычислить), процедура **Count...** (Сосчитать) не позволяет производить сложные математические вычисления, но удобна для простого подсчета частоты встречаемости конкретного значения во всех переменных или только в некоторых, интересующих исследователя.

Щелкнув мышью на строке меню **Count...** (Сосчитать), вы открываете окно (см. рис. 25) подсчета значений в кейсах (**Count Occurrences of Values within Cases**). Далее в первую очередь необходимо создать итоговую переменную, присвоив ей имя в подокне **Target Variable** и, при необходимости, задав ей соответствующую метку в подокне **Target Label** (это не обязательно).

Следующим действием является выбор переменных, в которых будет осуществляться подсчет числа интересующих исследователя значений. Для этого в списке переменных (подокно слева) нужно выбрать необходимые переменные и с помощью кнопки-указателя (с треугольником) переместить их в подокно **Numeric Variables** (Числовые переменные).

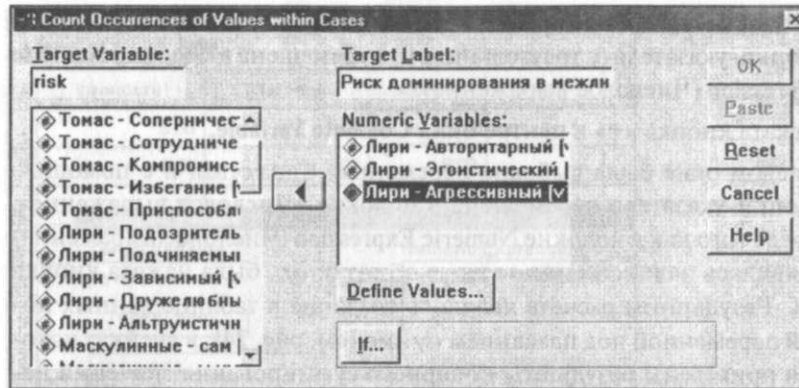


Рис. 25. Окно подсчета количества значений в кейсах

Следующим шагом является определение значений, подсчет которых и будет осуществлен далее в выбранных переменных.

Нажав на кнопку **Define Values...** (Определить значения), вы откроете окно, в котором содержится ряд возможностей для задания искомого значения (рис. 26).

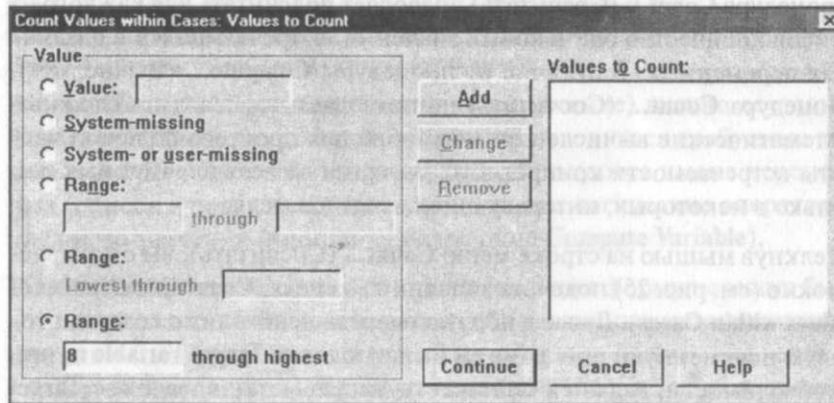


Рис. 26. Окно определения значения для подсчета его количества в кейсах

В этом окне можно указать конкретное число (**Value**); пропущенное значение, заданное программой или пользователем (**System-missing**, а также **System- or user-missing**); числовой интервал **Range-through** (от и до), **Range-Lowest through** (менее чем), и **Range-Through highest** (более чем). После того как значение определено, необходимо нажать на кнопку **Add** (Добавить) — значение будет помещено в подокно **Values to Count**

(Значения для подсчета) и нажать на кнопку **Continue** (Продолжить). Следует отметить, что в данном окне можно задать сразу несколько вариантов значений, частоту встречаемости которых вы хотите подсчитать. В некоторых случаях по недосмотру исследователя задаются логически противоречащие условия подсчета значений, например требования подсчитать значения менее 5, а также значения в интервале от 3 до 7. Разумеется, программа не сможет корректно обработать данный запрос. Также необходимо учитывать, что при задании числового интервала граничные числа включаются в сам интервал.

На рис. 25 и 26 приведен пример использования процедуры **Count...** (Сосчитать). Предположим, у психолога-исследователя есть результаты психодиагностического обследования большой группы людей по ряду методик. Перед психологом стоит практическая задача в короткий срок выявить группу риска — лиц, склонных к проявлению доминирования в межличностных отношениях — и исключить их из этой группы. В качестве диагностического основания для подобного отсева исследователь выбирает результаты испытуемых по шкалам теста ДМО\* (Т. Лири — Л. Н. Собчик) — «авторитарность», «эгоистичность», «агрессивность», выраженные значительно выше среднего уровня. Предполагается, что испытуемый, обладающий одновременно всеми тремя ярко выраженными характеристиками, стремится к деструктивному доминированию в общении и подлежит отсеvu. Последовательность действий исследователя была следующей:

1. В окне (рис. 25) подсчета значений в кейсах (**Count Occurrences of Values within Cases**) была создана итоговая переменная, которой в подокне **Target Variable** было присвоено имя *risk*. После этого в подокне **Target Label** новой переменной была присвоена метка Риск доминирования в межличностных отношениях.
2. Далее из всего массива имеющихся в исследовании диагностических показателей (переменных) были выбраны три ключевые (авторитарность, эгоистичность, агрессивность по тесту ДМО), которые были перемещены в подокно **Numeric Variables** (Числовые переменные);
3. В окне определения значения (рис. 26) для подсчета его количества в трех выделенных переменных выбираем числовой интервал **Range-through highest** (более чем) и вводим число **8** (напомним, что макси-

\* Собчик Л. Н. Диагностика индивидуально-типологических свойств и межличностных отношений. Практическое руководство. СПб.: «Речь», 2002.



мально возможное значение по шкалам ДМО — 12 баллов). Таким образом, мы указываем программе, что нам необходимо сосчитать встречающееся в трех указанных переменных количество ответов, превышающих 8 баллов (то есть выраженность которых значительно выше среднего уровня). После нажатия на кнопку **Add** (Добавить) в подокне **Values to Count** (Значения для подсчета) появляется надпись **8 thru Highest**. С помощью кнопки **Continue** (Продолжить) вновь оказываемся в окне подсчета значений;

4. После того как все необходимые условия заданы, для завершения процедуры подсчета нажмите на кнопку **ОК**. После этого в файле данных будет создана новая переменная *risk*, которая может принимать значения от 0 до 3. При этом «0» означает, что у конкретного испытуемого (в конкретном кейсе) ни один из ответов по трем изучаемым переменным («авторитарность», «эгоистичность», «агрессивность») не превышает 8 баллов. В свою очередь, «3» означает, что у данного испытуемого (в данном кейсе) ответы по каждой из трех изучаемых переменных превышают 8 баллов.

Таким образом, практическая задача успешно решена — достаточно лишь сортировать испытуемых по переменной *risk* (по убыванию; процедура сортировки описана ранее), чтобы получить ранжированный список в верхней части которого окажутся испытуемые, наиболее склонные к проявлению доминирования в межличностных отношениях.

Обратите внимание, что приведенный пример позволяет подсчитать одинаковые значения лишь в сходных переменных. Если необходимо подсчитать количество различных значений в разнородных переменных, то можно несколько раз повторить процедуру подсчета, а потом суммировать полученные частоты; либо воспользоваться опцией, позволяющей задать логические и т. п. сложные условия подсчета. Опция сложных условий подсчета вызывается кнопкой **If...** (Если...) в окне подсчета значений в кейсах (**Count Occurrences of Values within Cases**). (В данном руководстве эта операция не рассматривается).

Довольно часто для удобства представления и дальнейшего анализа данных используется операция **Recode** (Перекодировать). Перекодирование данных можно осуществлять сразу как во всех подобных переменных (**Recode into Same Variables**), так и в отдельных, значительно отличающихся друг от друга переменных (**Recode into Different**

Variables). В данном руководстве будет подробно рассмотрена процедура перекодирования данных в отдельных переменных (Recode into Different Variables) как наиболее распространенная в психологических исследованиях.

Выбрав в подменю опцию перекодирования данных в отдельных переменных (Recode into Different Variables), вы открываете соответствующее окно программы (рис. 27). Последовательность действий будет следующая:.

1. Выберите в списке слева исходную переменную, которую требуется перекодировать и с помощью кнопки-указателя (с треугольником) переместите ее в подокно Numeric Variable — Output (Переменная — результат перекодирования);

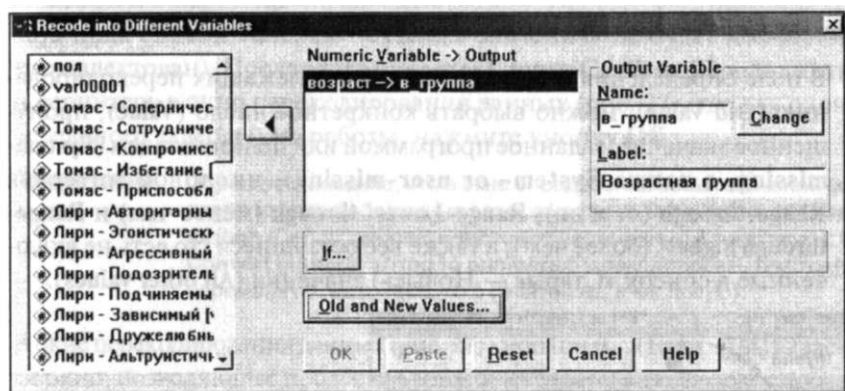


Рис. 27. Окно перекодирования данных в различных переменных

2. В подокне Output Variable (Итоговая переменная) необходимо задать имя (Name:) и метку (Label:) новой переменной.

Далее, для того чтобы новое имя переменной появилось в окне Numeric Variable — Output (Переменная — результат перекодирования), необходимо нажать на кнопку Change (Изменить). По желанию в поле Name: (Имя) можно ввести прежнее имя (в примере на рис. 27 вместо нового в\_группа, повторить еще раз возраст). В этом случае старые данные будут утеряны, а результат перекодирования помещен на месте прежних значений.

3. Щелкнув мышкой по кнопке Old and New Values... (Старые и новые значения), вы открываете окно определения новых значений при перекодировании (рис. 28).

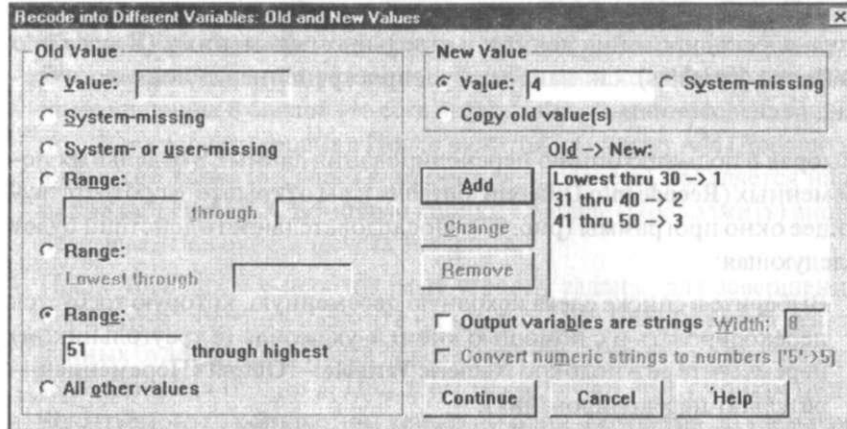


Рис. 28. Окно определения новых значений при перекодировании

В поле определения старых значений, подлежащих перекодированию (**Old Value**), можно выбрать конкретное число (**Value**), пропущенное значение, заданное программой или пользователем (**System-missing**, а также **System- or user-missing**), числовой интервал **Range-through** (от и до), **Range-Lowest through** (менее чем) и **Range-through highest** (более чем), а также все оставшиеся (то есть не включенные в список «Старые — Новые») значения (**All other values**).

в_группа	пол	возраст	в_группа
1	1,00	30,00	1,00
2	1,00	30,00	1,00
3	1,00	47,00	3,00
4	1,00	31,00	2,00
5	1,00	40,00	2,00
6	1,00	35,00	2,00
7	1,00	30,00	1,00
8	1,00	54,00	4,00

Рис. 29. Пример результата процедуры перекодирования данных

4. После того как вы выбрали значение для перекодирования, необходимо в поле **New Value** (Новое значение) указать новое число, которое будет подставлено вместо старого.

В качестве нового значения можно указать число (Value); пометить старое значение как пропущенное (System-missing); сохранить старое значение (Copy old value).

Обратите внимание, что перекодированию должны подвергаться *все* значения, входящие в переменную (именно для этого и существует флажок «сохранить старое значение» — Copy old value), в противном случае программа автоматически пометит все старые не перекодированные значения как пропущенные (System-missing).

С помощью кнопки Add (Добавить) перенесите формулу перекодирования в список Old > New (Старые — Новые значение).

Повторите шаги 1–4 требуемое количество раз, пока всем старым значениям не будут приписаны соответствующие новые значения (то есть пока список «Старые — Новые значения» не будет полностью укомплектован). После этого нажмите кнопку Continue (Продолжить) и вернитесь в окно перекодирования данных (рис. 27), где для окончательного завершения работы, нажмите кнопку ОК.

**Примечание.** Обратите внимание, что в окне определения новых значений содержатся две любопытные возможности: а) **Output variables are strings** (Перекодировать числовую переменную в текстовую); б) **Convert numeric string to numbers** (Перекодировать текстовую переменную, включающую в себя числа, в числовую).

В окне перекодирования данных (рис. 27) кнопкой И... (Если...) вызывается окно, позволяющее провести сложное избирательное перекодирование данных с использованием различных логических условий. (В данном руководстве эта возможность программы не рассматривается.)

На рис. 27 и 28 приведен пример перекодирования переменной «возраст» (данные представлены в интервальной измерительной шкале) в переменную «возрастная группа» (данные представлены в шкале порядка). Результат проведенной процедуры перекодирования можно увидеть на рис. 29. Подобное разделение на возрастные группы может быть полезным в том случае, если, например, в дальнейшем предполагается использование какого-либо критерия для проверки достоверности различий между результатами исследования, проведенного среди испытуемых разного возраста.

Операция Categorize Variables... (Создать категории) является упрощенным вариантом процедуры ранжирования (Rank Cases...) и дает

возможность оперативно разделить испытуемых на требуемое количество групп на основе любой из данных переменных.

Данная операция позволяет автоматически разбить континуальные числовые данные на заданное количество дискретных категорий. В основе процедуры категоризации лежит процентильная шкала. По умолчанию в программе принята категоризация на 4 группы (Number of categories), то есть используется квартильная мера. Напомним, что квартили представляют собой значения, которые делят выборку на четыре равные части. Это означает, что результаты испытуемых по изучаемой переменной, отнесенные к первой категории, во-первых, составляют ровно 25% от всех полученных значений, а во-вторых, являются наименьшими в изучаемой выборке.

Последовательность действий при создании категорий следующая:

1. Щелкнув мышью по строке Categorize Variables... (Создать категорию) вы раскроете окно создания категорий (рис. 30).

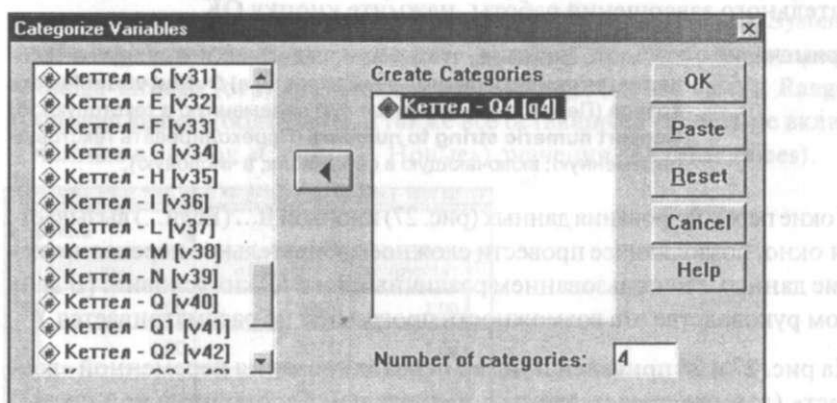


Рис. 30. Окно создания категорий

В окне создания категорий прежде всего нужно выбрать одну или несколько переменных, требующих категоризации. С помощью кнопки-указателя (с треугольником) переместите выбранные переменные в подокно Create Categories (Создать категории).

2. При необходимости можно изменить число выделяемых категорий, напечатав нужную цифру в соответствующем подокне (Number of categories). Для завершения процедуры категоризации нажмите на кнопку **ОК**.

После этого программа автоматически создаст в файле данных новые переменные, содержащие категории, которые далее могут быть названы и отредактированы (см. пример на рис. 31). Для каждой заданной в подокне Create Categories (Создать категории) переменной будет создана новая переменная. Необходимо учитывать, что после проведения процедуры категоризации размер файла данных значительно увеличится.



	q4	quartil	decil
1	6,00	3	7
2	5,00	3	6
3	4,00	2	4
4	4,00	2	4
5	6,00	3	7
6	2,00	1	1
7	4,00	2	4
8	7,00	4	9
9	4,00	2	4
10	2,00	1	1

Рис. 31. Пример результата процедуры создания категорий

На рис. 31 дан пример результата категоризации переменной q4 (одноименная шкала теста 16 PF\*) сначала на 4, а потом на 10 категорий. В основе первой категоризации лежит квартальная мера, в основе второй — децильная.

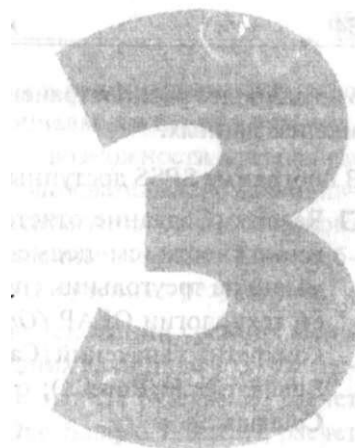
Для практического использования результатов, полученных по шкале Q4 («расслабленность — напряженность») теста Кеттелла 16 PF, удобнее всего квартильная категоризация. В результате подобной категоризации всех испытуемых можно разделить на 4 равные группы, причем в первую войдут имеющие наиболее выраженную расслабленность, во вторую — имеющие расслабленность выше среднего уровня (в пределах изучаемой группы), в третью — имеющие напряженность выше среднего уровня, в четвертую — имеющие наиболее высокую в данной группе напряженность и фрустрированность. Таким образом, получается простая и удобная типология для использования в практической работе психолога с конкретными группами.

\* Капустина А. Н. Многофакторная личностная методика Р. Кеттелла. СПб., «Речь», 2001.

Из своего опыта работы автор сделал вывод, что для составления различного рода рейтингов и экспресс-типологий наиболее удобно деление испытуемых на три категории (с использованием терцильной меры), в первую из которых попадают испытуемые с «низкими» для данной группы результатами, во вторую — со «средними», в третью — с «высокими». При этом необходимо помнить, что данные оценки (низкие, средние и высокие) являются не абсолютными, а относительными, имеющими смысл только внутри изучаемой группы и дающими возможность сравнить испытуемых друг с другом.

## Глава 3

# АНАЛИЗ ДАННЫХ В SPSS



## ОБЗОР СТАТИСТИЧЕСКИХ ВОЗМОЖНОСТЕЙ ПРОГРАММЫ SPSS

Выбрав опцию Analyze (Анализ) в строке главного меню SPSS (рис. 32), вы получаете доступ ко всем статистическим операциям SPSS. Статистические возможности программы поистине огромны (ниже они будут кратко описаны), но, так как данное руководство предназначено в первую очередь для студентов, аспирантов и психологов-практиков, далее будут подробно рассмотрены лишь некото-

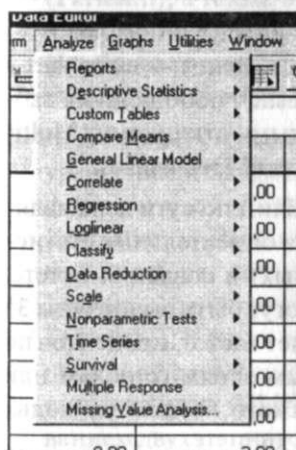


Рис 32. Меню выбора статистических операций SPSS



рые наиболее распространенные и удобные методы статистического анализа данных.

В программе SPSS доступны следующие методы анализа данных:

- Reports** (Создание отчетов) позволяет использовать следующие возможности (см. подменю, которые появляются при наведении мыши на треугольник справа): создание отчетов с использованием технологии OLAP (*Online Analytical Processing*) Cubes; подсчет конкретных значений (**Case Summaries...**); отчет по строкам (**Report Summaries in Rows...**); отчет по столбцам (**Report Summaries in Columns...**).
- Метод создания отчетов позволяет не только получить общее представление о таблице данных, но, по сути дела, является самостоятельным и довольно мощным инструментом описания и анализа данных. Кроме элементарного подсчета частот, в создание отчетов включены богатые возможности по группировке и категоризации данных, что дает возможность получить наиболее наглядную и удобную для исследователя картину. При создании отчетов также можно использовать инструментарий по вычислению различных алгебраических и статистических функций.
- Descriptive Statistics** (Подсчет описательных статистик) — включает в себя частотные характеристики данных (**Frequencies...**), описательные статистики (**Descriptives...**), разведывательные статистики (**Explore...**), кросстабуляции (таблицы сопряженности; **Crosstabs...**). Более подробно процедура расчета описательных статистик будет рассмотрена ниже.
- Custom Tables** (Создание пользовательских таблиц) — включает в себя создание базовых таблиц (**Basic Tables...**), обобщенных таблиц (**General Tables**), таблиц с множественными откликами (**Multi-Response Tables...**), частотных таблиц (**Tables of Frequencies...**).
- Процедура создания пользовательских таблиц по сути дела является комбинацией двух процедур: подсчета описательных статистик (в особенности — кросстабуляционных) и создания отчета. При создании пользовательских таблиц могут быть вычислены 35 статистических показателей, включая расчет частот, процентов по строкам и столбцам, расчет по вложенным слоям (строкам или столбцам), расчет среднего арифметического, медианы, моды, стандартного отклонения, процентилей и др.

- Этот раздел программы SPSS может быть рекомендован в первую очередь социологам, маркетологам, социальным психологам, т. к. здесь содержатся наиболее эффективные возможности для анализа данных анкетных опросов. Создание пользовательских таблиц является мощным самостоятельным инструментом статистического анализа, в особенности анализа качественных данных. Такие таблицы очень удобны для представления полученных результатов.
- **Compare Means** (Сравнение данных на основании средних величин) — включает в себя сравнение средних величин с возможностью однофакторного дисперсионного анализа (**Means...**), расчет Т-критерия для одной переменной (**One-Sample T Test...**), расчет Т-критерия для независимых переменных (**Independent-Samples T Test...**), расчет Т-критерия для парных переменных (**Paired-Samples T Test...**), однофакторный дисперсионный анализ (**One-Way ANOVA...**). Более подробно процедура будет рассмотрена ниже.
- **Общая линейная модель (General Linear Model)** — различные виды дисперсионного анализа. Позволяет производить однофакторный дисперсионный анализ (**Univariate...**), многофакторный дисперсионный анализ (**Multivariate...**), дисперсионный анализ для повторяющихся измерений (**Repeated Measures...**). А также *смешанная дисперсионная модель*, так называемая компонентная дисперсия (**Variance Components...**).
- **Correlate** (Корреляционный анализ) — включает в себя вычисление парных корреляций (**Bivariate...**), частных корреляций (**Partial...**), а также вычисление расстояний между переменными, аналогично кластерному или факторному анализу (**Distances...**). Наиболее распространенная процедура расчета парных корреляций будет подробно рассмотрена далее.
- **Regression** (Регрессионный анализ) — включает в себя построение уравнения линейной регрессии (**Linear...**), а также построение графиков и расчет статистик для регрессионных кривых (**Curve Estimation...**). Кроме того, в программе SPSS также имеется возможность расчета двоичной логистической регрессии (**Binary Logistic...**), множественной логистической регрессии (**Multinomial Logistic...**), а также пробит-анализ (**Probit...**). Доступен расчет нелинейной регрессии (**Nonlinear...**) и оценка полученных регрессионных кривых (**Weight Estimation...**), расчет регрессии с использованием двухэтапного метода наименьших квадратов (**2-Stage Least**

- Squares...**) и метод оптимального шкалирования для ранговых и номинальных данных (**Optimal Scaling...**).
- Loglinear** (Логлинейная модель для многомерных таблиц сопряженности) — включает в себя общие модели (**General...**), логистические (**Logit...**), кроме того, возможность построения пользовательских моделей (**Model Selection...**).
  - Classify** (Автоматическая классификация данных) — включает в себя упрощенный алгоритм кластерного анализа с использованием **K-Means Cluster...** ( $K$  средних), различные алгоритмы иерархического кластерного анализа (**Hierarchical Cluster...**), а также дискриминантный анализ (**Discriminant...**).
  - Data Reduction** (Сокращение размерности данных) — включает в себя различные алгоритмы факторного анализа (**Factor...**), анализ соответствий (процедура аналогична факторному анализу, но применяется для номинальных данных; **Correspondence Analysis...**); метод оптимального шкалирования для ранговых и номинальных данных (**Optimal Scaling...**).
  - Scale** (Шкалирование) — включает в себя анализ надежности (**Reliability Analysis...**) и многомерное шкалирование (**Multi-dimensional Scaling...**).
  - Nonparametric Tests** (Непараметрические критерии) — включает в себя расчет критериев Хи-квадрат (**Chi-Square...**), биномиального критерия для дихотомических данных (**Binomial...**), критерия серий (**Runs...**), критерия Колмогорова—Смирнова для одной выборки (**1-Sample K-S...**), ряда критериев для двух независимых выборок (**2 Independent Samples...**), для  $K$  независимых выборок (**K Independent Samples...**), для двух связанных выборок (**2 Related Samples...**), для  $K$  связанных выборок (**K Related Samples...**). Более подробно процедуры расчета непараметрических критериев будут рассмотрены ниже.
  - Time Series** (Анализ временных рядов) — включает в себя метод экспоненциального сглаживания (**Exponential Smoothing...**), авторегрессионный анализ (**Autoregression...**), модель авторегрессии и проинтегрированного скользящего среднего (**ARIMA...**), модификация модели ARIMA для выявления сезонных тенденций (**X11ARIMA...**), выявление сезонных составляющих в различных временных рядах (**Seasonal Decomposition...**).

- Survival** (Анализ выживаемости) — включает в себя анализ таблиц выживаемости (**Life Tables...**), оценивание временного интервала до наступления события по методу Каплана–Майера (**Kaplan-Meier...**), регрессионный анализ Кокса (**Cox Regression...** и **Cox W/Time-Dep Cov...**).
- Multiple Response** (Анализ множественных откликов).
- Missing Value Analysis** (Анализ пропущенных значений).

## ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ

После активизации в главном меню программы SPSS опции **Analyze** (Анализ), выбираем в выпадающем меню строку **Descriptive Statistics** (Описательные статистики). В свою очередь, опция «Описательные статистики» имеет четыре раздела: 1) подсчет частотных характеристик данных (**Frequencies...**); 2) вычисление собственно описательных статистик (**Descriptives...**); 3) расчет разведывательных статистик (**Explore...**); построение таблиц сопряженности (**Crosstabs...**). Рассмотрим некоторые из этих разделов подробнее.

Опция **Frequencies...** (Вычисление частотных характеристик) позволяет получить базовые статистические показатели четырех типов:

1. Процентные оценки (**Percentile Values**).
2. Меры центральной тенденции (**Central Tendency**).
3. Меры рассеяния (**Dispersion**).
4. Особенности распределения (**Distribution**).

Эти статистические показатели не только позволяют получить наглядное представление о полученных данных и провести их качественный анализ, но и «на глаз» сравнить некоторые переменные друг с другом. Кроме того, данные статистические показатели являются основой для принятия решения об использовании в дальнейшем более сложных методов статистического анализа.

### Последовательность вычисления частотных характеристик

1. Щелкнув мышью на строке выпадающего меню **Frequencies...** (Вычисление частотных характеристик), вы вызываете соответствующее

окно (см. рис. 33). В левом верхнем подокне выберите одну или несколько переменных и с помощью кнопки-указателя (с треугольником) переместите их в подокно Variable(s): (Переменные).

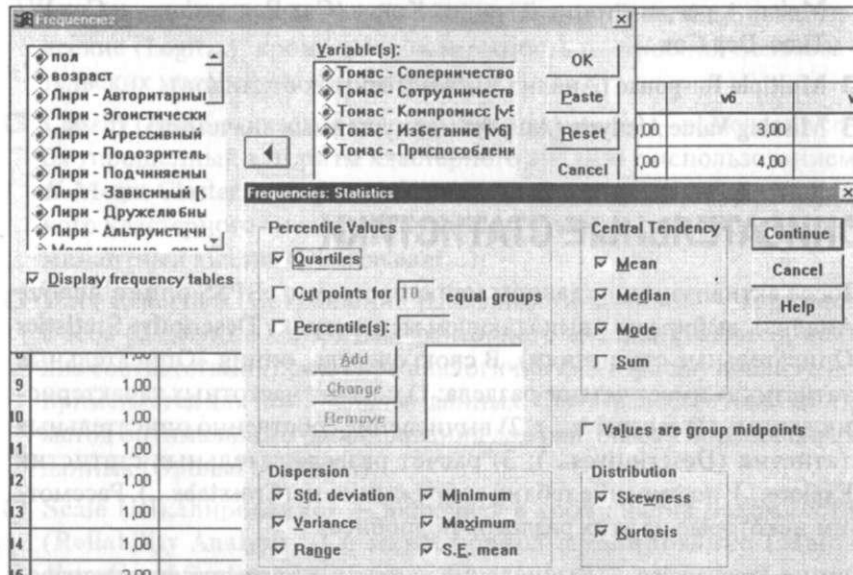


Рис. 33. Вычисление частотных характеристик (отображены базовое окно процедуры (окно **Frequencies** на втором плане) и подокно задания статистик (окно **Frequencies: Statistics** на первом плане)).

2. Нажав в окне **Frequencies** кнопку **Statistics...** (Статистики), вы вызовете подокно задания статистик (**Frequencies: Statistics**; см. рис. 33). Установив соответствующий флажок, вы можете подсчитать статистические показатели всех четырех типов.

### Процентные оценки

Можно выбрать подсчет квартилей (**Quartiles**). По желанию исследователя можно вычислить процентные оценки для  $N$  равных по количеству групп (**Cut points for \_\_\_ equal groups**). Количество интересующих групп задается в соответствующей строке и может находиться в пределах от 2 до 100. Кроме того, можно запросить у программы вычисление сразу нескольких произвольных процентных оценок (**Percentile(s)**). Для этого необходимо внести процентные значения в соответствующую строку, после каждого нового значения нажимая на кнопку **Add** (Добавить).

### Меры центральной тенденции

Операция включает в себя расчет среднего арифметического (Mean), медианы (Median), моды (Mode) и суммы значений по столбцу (Sum).

### Меры рассеяния

Позволяет вычислить стандартное отклонение (Std. deviation), дисперсию (Variance), размах (Range), минимальное значение переменной (Minimum), максимальное значение переменной (Maximum), стандартную ошибку средней (S. E. mean).

### Особенности распределения

Включает в себя вычисление асимметрии (Skewness) и эксцесса (Kurtosis).

После того как вы указали все интересующие вас статистики, нажмите на кнопку Continue (Продолжить) и вернитесь в окно Frequencies.

3. Нажав в окне Frequencies кнопку Charts... (Диаграммы), вы вызовете подокно создания диаграмм (Frequencies: Charts). Установив соответствующий флажок в данном окне, вы можете выбрать создание диаграмм трех типов (Chart Type): а) столбиковые (Bar Charts); б) круговые (Pie Charts); в) гистограммы (Histograms). Если вы выбрали столбиковые или круговые диаграммы, то ниже (Chart Values), установив соответствующий флажок, вы можете установить в качестве отображаемых на диаграмме значений частоты (Frequencies) или проценты (Percentages). Если вы выбрали создание гистограммы, то с помощью соответствующего флажка (with normal curve) вы сможете наложить на изображение кривую нормального распределения. Данная операция очень удобна для визуальной экспресс-оценки соответствия полученных вами данных нормальному распределению.

Обратите внимание, что в подокне создания диаграмм по умолчанию стоит флажок, блокирующий создание диаграмм (None). Это не случайно! Дело в том, что создаваемая программой графика в несколько раз увеличивает размер файла с результатами расчетов (SPSS Output) и требует больших объемов как оперативной, так и дисковой памяти компьютера. Владельцам маломощных компьютеров вообще *не рекомендуется* использовать очень много диаграмм, т. к. это может привести к зависанию программы и потере данных.

После выбора типа диаграммы нажмите на кнопку **Continue** (Продолжить) и вернитесь в основное окно **Frequencies**.

4. Кнопка форматирования результатов (**Format...**) расчетов позволяет сортировать результаты, задавать формат таблицы выходных данных и т. п. Начинаящему пользователю лучше оставить принятые в программе по умолчанию настройки, поэтому в данном руководстве возможности по форматированию результатов не рассматриваются.

5. Обратите особое внимание на стоящий по умолчанию в окне **Frequencies** флажок, дающий возможность вывода частотных таблиц для каждой из анализируемых переменных (**Display frequency tables**).

Несмотря на то, что вывод частотных таблиц значительно увеличивает размер файла с результатами расчетов, данная операция является необычайно полезной в особенности для наглядного представления и анализа качественных данных (результаты исследований, представленные в шкале наименований), где использование других статистических показателей невозможно.

Если вы анализируете большое количество «неудобных» переменных (результаты представлены в шкале интервалов; существуют очень большой разброс и разнообразие значений), то вывод частотных таблиц в файл результатов можно отключить. Для этого достаточно убрать соответствующий флажок.

6. Для завершения процедуры расчета частотных характеристик нажмите на кнопку **OK**. После этого программа автоматически создаст файл с результатами и откроет окно вывода (**SPSS Output**) полученных статистик (см. пример на рис. 34, 35, 36).

На рис. 34 представлен фрагмент таблицы с результатами расчета описательных статистик для шкал теста Томаса, выявляющего стиль поведения в конфликтных ситуациях. Таблица **Statistics** (Статистики) расположена первой в файле вывода (**SPSS Output**) полученных результатов. В шапке таблицы (см. рис. 34) размещены наименования переменных (шкал теста) для которых производился расчет, в боковой графе таблицы (сверху вниз) располагается следующая информация: количество значений в переменной (**N**), из них валидных, годных для обработки (**Valid**) и пропущенных (**Missing**), средняя арифметическая величина (**Mean**), медиана (**Median**), мода (**Mode**), стандартное отклонение (**Std. Deviation**), дисперсия (**Variance**), асиммет-

рия (**Skewness**) и стандартная ошибка асимметрии (**Std. Error of Skewness**), эксцесс (**Kurtosis**) и его стандартная ошибка (**Std. Error of Kurtosis**), минимальное (**Minimum**) и максимальное (**Maximum**) для данной выборки значения.

Обратите внимание, что в данном примере (рис. 34) также содержатся заданные исследователем (**Percentile Values > Percentile(s):**) расчеты значений, делящих испытуемых на три равные подгруппы (тертили).

		Stati		
		Томас - Соперни- чество	Томас - Сотрудни- чество	То Комп
N	Valid	25	25	
	Missing	0	0	
Mean		4,4000	6,1200	
Median		4,0000	6,0000	
Mode		3,00	6,00	
Std. Deviation		2,6458	1,6155	
Variance		7,0000	2,6100	
Skewness		1,127	,112	
Std. Error of Skewness		,464	,464	
Kurtosis		1,388	-,645	
Std. Error of Kurtosis		,902	,902	
Minimum		1,00	3,00	
Maximum		12,00	9,00	
Percentiles	33	3,0000	5,0000	
	67	5,0000	7,0000	

Рис. 34. Фрагмент таблицы с результатами расчета описательных статистик

Например, для показателя Томас – Соперничество треть испытуемых имеет результаты от 1 до 3 баллов (**Minimum** – 33%), еще треть от 3 до 5 баллов (33–67%), и оставшаяся треть от 5 до 12 баллов (67% – **Maximum**). Подобный поверхностный взгляд уже позволяет сделать вывод, что большая часть изучаемых испытуемых избегает соперничества в конфликтных ситуациях.

		Томас - Сотрудничество			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3,00	1	4,0	4,0	4,0
	4,00	3	12,0	12,0	16,0
	5,00	5	20,0	20,0	36,0
	6,00	7	28,0	28,0	64,0
	7,00	3	12,0	12,0	76,0
	8,00	4	16,0	16,0	92,0
	9,00	2	8,0	8,0	100,0
Total		25	100,0	100,0	

Рис. 35. Пример частотной таблицы



После таблицы со статистиками в файле вывода результатов (обычно ниже) располагаются частотные таблицы (см. рис. 35)

На рис. 35 приводится пример частотной таблицы для переменной Томас - Сотрудничество. В боковой графе таблицы перечислены все встречающиеся у наших испытуемых значения переменной (**Valid**), а ниже автоматически подсчитывается их общее количество (**Total**). Если в переменной имеются пропущенные значения, то внизу также появляется соответствующая строка (**Missing**).

В шапке частотной таблицы содержится следующая информация: частота встречаемости каждого значения (**Frequency**), процентная доля данного значения от всей выборки (**Percent**), валидный процент (строгий процент, вычисленный с учетом пропущенных значений; **Valid Percent**), накопленные проценты (**Cumulative Percent**).

На рис. 36 приведен пример гистограммы, построенной программой для переменной Томас - Сотрудничество. Обычно гистограммы располагаются в файле отчета после частотных таблиц. Обратите внимание, что данная гистограмма построена с наложением кривой нормального распределения (**Frequencies: Charts > Histograms > With normal curve**). По оси абсцисс расположены значения, по оси ординат — частоты появления данных значений. Справа от рисунка в качестве дополнительной информации располагаются показатели стандартного



Рис. 36. Пример гистограммы

отклонения (**Std. Dev**), средней арифметической (**Mean**), количества испытуемых (**N**).

Специально для начинающих статистиков напомним, что корректное использование описательных статистик зависит от типа измерительной шкалы, в формате которой представлены ваши данные.

Для результатов, представленных в шкале наименований (номинальной) допускается построение частотных таблиц (**Display frequency tables**) и вычисление моды (**Mode**). Возможно также построение столбиковой или круговой диаграмм (**Bar Charts** или **Pie Charts**).

Для результатов, представленных в шкале рангов (порядковой, ordinalной), допускается построение частотных таблиц (**Display frequency tables**), вычисление моды (**Mode**), медианы (**Median**), вычисление любых процентных оценок (**Percentile Values**). Возможно также построение столбиковой или круговой диаграмм (**Bar Charts** или **Pie Charts**). В некоторых случаях (например, при сравнении двух ранговых переменных) допускается вычисление среднего ранга, минимального и максимального ранга, размаха и др.

Для результатов, представленных в шкале интервалов (или отношений), допускается использование любых описательных статистик, относящихся к процентным оценкам, мерам центральной тенденции, мерам рассеяния, особенностям распределения. В этом случае графическим операциям добавляется построение гистограмм с наложением кривой плотности нормального распределения (гауссианы).

Опция описательных статистик (**Descriptives...**) содержит ряд таких же возможностей, что и опция **Frequencies...** (Вычисление частотных характеристик). Нажав на кнопку **Options...** (Опции) в окне **Descriptives** можно произвести подсчет среднего арифметического (**Mean**), суммы (**Sum**), стандартного отклонения (**Std. deviation**), дисперсии (**Variance**), размаха (**Range**), минимума (**Minimum**), максимума (**Maximum**), стандартной ошибки среднего (**S. E. mean**), асимметрии (**Kurtosis**) и эксцесса (**Skewness**). Процедура подсчета всех этих показателей аналогична подсчету частотных характеристик в опции **Frequencies...**, подробно описанной выше, и поэтому рассматриваться не будет.

Однако стоит упомянуть, что в главном окне опции расчета описательных статистик (**Descriptives...**) содержится одна важная операция —

расчет z-оценок. Для этого достаточно поставить флажок **Save standardized values as variables** (Сохранить стандартизированные значения как новые переменные) в окне подсчета описательных статистик (**Descriptives...**). По окончании расчета программа автоматически добавит в файл данных z-оценки для каждой из стандартизируемых переменных; каждая новая переменная сохранит своё название, но с добавлением буквы «z» в начале имени.

Напомним, что расчет z-оценок — важный этап в разработке теста. Такой расчет (или z-трансформация) является линейным преобразованием, который позволяет перейти от «сырых» тестовых оценок к стандартным показателям. Для практика расчет z-оценок удобен в первую очередь тем, что дает возможность привести «сырые» результаты, полученные по различным методикам, к одному общему «масштабу» и, следовательно, получить возможность прямого сопоставления разнородных данных.

Подсчет z-оценок возможен *только* для данных, измеренных в шкале интервалов (или отношений). Положительное значение z-оценки свидетельствует о том, что результаты испытуемого выше среднего по выборке, отрицательное — что результаты ниже среднего.

Построение и анализ таблиц сопряженности (**Analyze > Descriptive Statistics > Crosstabs...**) является одной из наиболее востребованных статистических процедур в социально-психологических, социологических, маркетинговых и т. п. исследованиях, результаты которых представлены в основном в шкале наименований (Explore...). (Расчет разведывательных статистик в данном руководстве не рассматривается.)

Напомним, что таблицы сопряженности позволяют получить наглядное отображение (в виде частотной таблицы) совместного распределения двух переменных, а также проверить гипотезу о наличии или отсутствии связи между ними.

## Последовательность построения таблиц сопряженности

### 1 шаг

Щелкнув мышью на строке выпадающего меню **Crosstabs...** (Таблицы сопряженности), вы вызываете соответствующее окно (см.

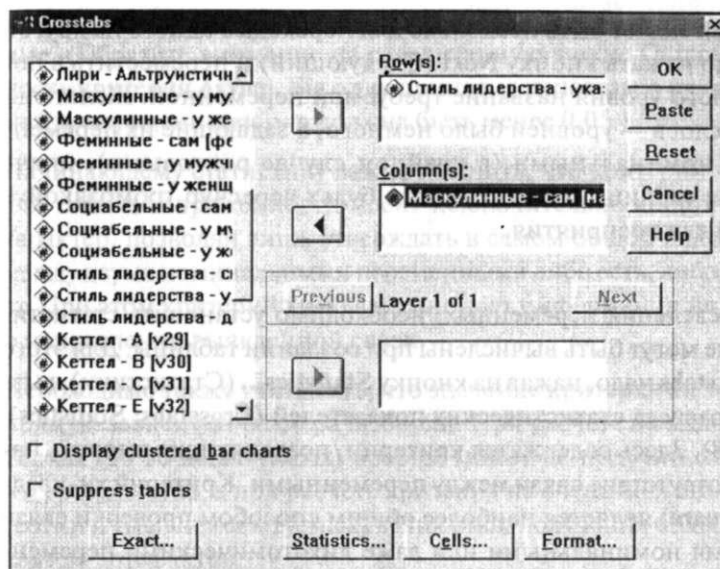


Рис. 37. Окно построения таблиц сопряженности

рис. 37). В левом подокне выберите одну или несколько переменных и с помощью кнопки-указателя (с треугольником) переместите их в подокно **Row(s)** (Строки). Далее переместите одну или несколько переменных в подокно **Column(s)** (Столбцы). После построения таблицы сопряженности значения переменных, помещенных в **Row(s)** (Строки), появятся в боковой графе таблицы; значения переменных, помещенных в **Column(s)** (Столбцы), окажутся в шапке таблицы.

Простейший вариант построения таблицы сопряженности — когда в подокнах **Row(s)** (Строки) и **Column(s)**: (Столбцы) содержится только по одной переменной. Если хотя бы в одном из подокон находится несколько переменных, то программа построит несколько таблиц сопряженности, исчерпывающих все возможные парные комбинации между строками и столбцами.

Обратите внимание, что в программе SPSS предусмотрена возможность построения многоуровневых таблиц сопряженности, позволяющая «вкладывать» изучаемые переменные друг в друга. Значения переменной, перемещенные в подокно **Layer** (Слой), в дальнейшем будут выступать в качестве критерия разделения значений переменных в поле **Row(s)** на соответствующие им подгруп-

пы. Уровней может быть несколько для перехода с одного на другой необходимо нажать кнопку Next (Следующий) и переместить в подокно второго уровня название требуемой переменной. Желательно, чтобы слоев — уровней было немного, а задающие их переменные были номинальными (в крайнем случае ранговыми), иначе полученная таблица сопряженности будет чересчур громоздкой и неудобной для восприятия.

## II шаг

После определения переменных, необходимо установить статистики, которые могут быть вычислены при создании таблицы. Для этого в окне Crosstabs надо, нажав на кнопку Statistics... (Статистики), вызвать окно расчета статистических показателей (Crosstabs: Statistics), (см. рис. 38). Здесь содержатся критерии, позволяющие выявить наличие или отсутствие связи между переменными. Критерий хи-квадрат (Chi-square) является наиболее общим способом проверки связи между двумя номинальными или даже дихотомическими переменными. Напомним, что критерий хи-квадрат Пирсона достоверно указывает на существование связи между переменными на основании нормированной суммы квадратов различий между эмпирическими частотами встречаемости значения и теоретическими частотами, вычисленными на основе предположения о независимости этих переменных. Таким образом, чем меньше абсолютное значение критерия

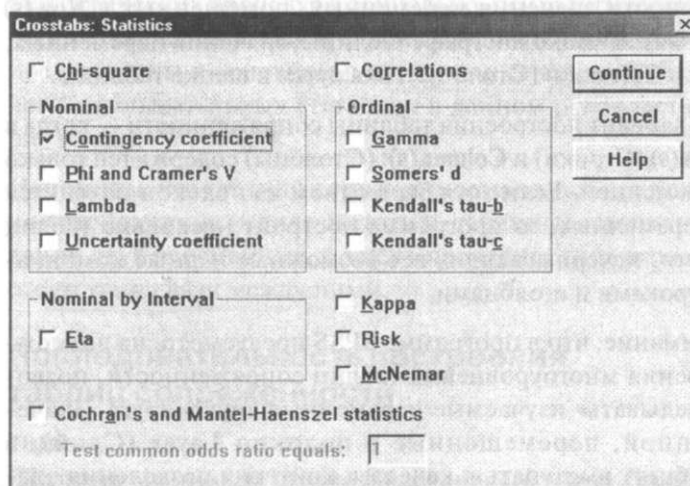


Рис. 38. Окно расчета статистик для таблиц сопряженности

хи-квадрат, тем более зависимы, взаимосвязаны изучаемые переменные. (Обратите внимание на рассчитанную в окне Output достоверность критерия *Asymp. Sig.*; для принятия гипотезы о наличии связи между переменными она должна быть менее 0,05.)

Начинающему статистику важно помнить, что критерий хи-квадрат для таблиц сопряженности носит исключительно констатирующий характер, позволяя лишь утверждать в самом общем виде, что связь между двумя номинальными переменными либо есть, либо отсутствует. При этом критерий хи-квадрат не дает информации ни о силе, ни о направлении выявленной связи.

Необходимо также учитывать, что значение критерия хи-квадрат напрямую зависит от размера выборки. При расчете на маленьких выборках (до 30 испытуемых) вообще может не получиться достоверных результатов, а при расчете критерия на очень больших выборках (сотни и тысячи испытуемых) велик риск получения большого числа связей-артефактов.

Использование операции расчета корреляций (*Correlations*) в окне *Crosstabs: Statistics* допустимо только для переменных, содержащих ранговые или интервальные данные. Связь между ранговыми переменными будет рассчитана с помощью коэффициента ранговой корреляции Спирмена; связь между интервальными данными — с помощью коэффициента линейной корреляции Пирсона. Обратите внимание, что для корректного вычисления корреляций в таблицах сопряженности ранговые переменные должны сопоставляться с ранговыми, интервальные — с интервальными.

В окне *Crosstabs: Statistics* содержится также ряд критериев, позволяющих проверить связь между номинальными данными (*Nominal*): коэффициент сопряженности (*Contingency coefficient*), Фи-критерий и V-критерий Крамера (*Phi and Cramer's V*), критерий лямбда Гудмена-Крускала (*Lambda*), а также коэффициент неопределенности (*Uncertainty coefficient*).

Данные коэффициенты являются модификациями критерия хи-квадрат, и их основное назначение — конкретизировать и более наглядно представить связь между двумя номинальными переменными. Например, в отличие от «безразмерного» хи-квадрата, коэффициент сопряженности (*Contingency coefficient*) может принимать значения

от 0 до 1, где ноль указывает на отсутствие связи между переменными, а единица — на наличие тесной связи между ними. Критерий лямбда (Lambda) показывает относительную оценку эффективности использования независимой переменной для предсказания значения зависимой переменной. Значения критерия также меняются от 0 до 1, при этом ноль определяет невозможность подобного предсказания, а единица свидетельствует о том, что по значению независимой переменной всегда можно точно предсказать значение зависимой переменной. Аналогичный смысл имеет коэффициент неопределенности (Uncertainty coefficient). Например, коэффициент неопределенности 0,82 означает, что зная значение одной переменной, мы на 82% уменьшаем ошибку в предсказании значения другой переменной. Если коэффициент неопределенности равен единице, то это указывает на существование четкого соответствия (связи) между значениями изучаемых переменных.

Если данные в изучаемых переменных представлены в ранговых (порядковых) шкалах, то также можно воспользоваться специализированными критериями связи (Ordinal): критерием гамма Гудмена и Краскела (Gamma), критерием D Соммерса (Sommer's d), а также критериями тау-б и тау-це Кендалла (Kendall's tau-b и Kendall's tau-c). Для начинающих статистиков напомним, что данные критерии можно интерпретировать как показатели корреляционной связи, т. к. они не только указывают на наличие или отсутствие связи между переменными (как критерии для номинальных шкал), но также дают информацию о силе и направлении этой связи. Значение всех четырех критериев может изменяться в пределах от  $-1$  до  $+1$ , при этом близкие к нулю значения свидетельствуют об отсутствии (или слабой) связи между признаками, знак « $-$ » указывает на обратную, а знак « $+$ » — на прямую зависимость между признаками. Использование критерия тау-б Кендалла (Kendall's tau-b) наиболее эффективно для квадратных таблиц сопряженности, не содержащих нулевых значений; критерии гамма (Gamma) и тау-це (Kendall's tau-c) работают несколько хуже при сопоставлении переменных, имеющих много парных совпадающих значений. Этого недостатка лишен критерий Соммерса (Sommer's d).

Для выявления связи между номинальной и интервальной переменными используется коэффициент эта (Eta), который может принимать значения от нуля (при отсутствии связи) до единицы (при нали-

ции сильной связи). Для корректного подсчета коэффициента в качестве независимой переменной должна выступать номинальная, а в качестве зависимой — интервальная. Таким образом, коэффициент выявляет, влияет ли принадлежность испытуемого к дискретной группе (например, к половой) на проявление каких-либо непрерывных качеств (например, показателей по тестам интеллекта). По умолчанию программа SPSS рассчитывает два значения критерия: первое определяется исходя из предположения, что интервальная переменная была задана в таблице сопряженности по строке (Row(s)); второе значение вычисляется на основе предположения, что интервальная переменная была задана по столбцу (Column(s)). Разумеется, единственное корректное значение коэффициента эта должен выбрать сам исследователь.

В окне Crosstabs: Statistics содержится также ряд узкоспециализированных критериев связи. Первый из них — каппа Козна (Kappa) используется для выявления согласия оценок двух разных экспертов, оценивающих один и тот же объект с помощью одной и той же системы оценок. Критерий каппа может принимать значение от нуля (отсутствие согласия, случайное совпадение отдельных оценок экспертов) до единицы (полное согласие, закономерное совпадение большинства оценок). Для корректного использования критерия оцениваемые переменные должны быть одинаковой длины и не должны содержать пропущенных значений.

Вычисление относительного риска (Risk) используется для прогнозирования вклада значений одной переменной в значения другой («итоговой»). Вычисляется только для таблиц 2x2, не имеющих пустых ячеек.

Критерий Мак Немары (McNemar) специализирован для выявления степени связи между двумя зависимыми дихотомическими переменными. Часто используется для экспресс-оценки результатов экспериментального воздействия, когда проверяется сила связи между результатами, полученными до и после экспериментального воздействия на одну и ту же группу испытуемых. Экспериментальное воздействие является эффективным, если значение полученного критерия Мак Немары имеет достоверность (Exact Sig.) 0,05 и меньше.

Статистики Кохрана и Мантеля-Хензела (Cochran's and Mantel-Haenszel statistics) позволяют выявить линейную зависимость между



двумя дихотомическими переменными на основе параметров одной или нескольких переменных-уровней.

После того как вы установили флажок напротив одного или нескольких подходящих статистических критериев связи для таблиц сопряженности, нажмите на кнопку Continue (Продолжить), чтобы вернуться в окно построения таблиц сопряженности (Crosstabs).

### III шаг

Нажав на кнопку Cells... (Ячейки), вы получаете возможность определить статистическую информацию, которая будет размещена в ячейках таблицы сопряженности после ее генерации. По умолчанию в программе SPSS уже установлен флажок подсчета актуально наблюдаемых частот (Observed), но также можно задать вычисление ожидаемых частот (Expected), которые должны находиться в ячейке, если вам необходимо подтверждение гипотезы о статистической независимости сопоставляемых переменных. Заметим, что вычисление ожидаемых частот является не обязательной процедурой, т. к. при расчете различных критериев связи в таблицах сопряженности программа автоматически проводит требуемые вычисления ожидаемых частот.

В разделе Percentages (Проценты), установив соответствующие флажки произвести у программы расчет процентов по строкам таблицы сопряженности (Row), по столбцам (Column), либо общий, по всей таблице в целом (Total). Для начинающих исследователей обычно наиболее востребованной является процедура расчета общих процентов по таблице (Total), тогда как расчеты процентов по строкам и столбцам могут лишь перегружать таблицу, делая ее неудобной для восприятия.

В разделе Residuals (Остатки) можно сделать вычисление разностей между изучаемыми переменными. Разности (остатки) помогают более наглядно представить зависимость или независимость изучаемых переменных и являются промежуточными расчетами при вычислении ряда критериев связи. В программе SPSS существует возможность расчета простых остатков (Unstandardized), нормированных (Standardized) и приведенных нормированных (Adj. standardized).

Окончательно определив свойства ячеек таблицы сопряженности, после установки соответствующих флажков, нажмите на кнопку Continue (Продолжить), чтобы вернуться в окно построения таблиц сопряженности (Crosstabs).

#### IV шаг

В базовом окне построения таблиц сопряженности (Crosstabs) остались еще две кнопки: Exact... (Точность) и Format... (Формат). При нажатии кнопки Exact... (Точность) появляется окно, в котором определяется способ вычисления доверительного интервала, что особенно важно при проверке уровня достоверности произведенных расчетов. С помощью кнопки Format... (Формат) пользователь получает возможность по своему усмотрению сортировать данные в построенной таблице сопряженности. Начинающему исследователю лучше не использовать эти опции, оставив настройки, принятые в программе по умолчанию.

Кроме того, в базовом окне построения таблиц сопряженности содержатся возможности создания сгруппированных столбиковых диаграмм (Display clustered bar charts). Подобная диаграмма отображает частоты совместного распределения значений в строках и столбцах таблицы и носит не столько информативный характер, сколько наглядный. Необходимо учитывать, что при построении большого количества таблиц сопряженности диаграммы будут построены для каждой из них, что может потребовать значительных затрат памяти компьютера для отображения графики. В этом же окне находится опция отключения построения таблицы (Suppress tables), которая нужна в тех случаях, когда необходимо сразу получить результаты статистических расчетов по таблице (Statistics...), а сама таблица не нужна, либо является чрезмерно громоздкой и неудобной для отображения на экране. (По мнению автора, начинающему исследователю данные возможности лучше не использовать.)

#### V шаг

Для завершения процедуры построения таблиц сопряженности нажмите кнопку **ОК**, после чего программа автоматически сгенерирует и откроет файл вывода результатов SPSS Output. В файле вывода результатов обычно сначала располагается сама таблица сопряженности, ниже следуют рассчитанные статистические критерии, еще ниже — столбиковые диаграммы (если было задано их построение).

## ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ ДОСТОВЕРНОСТИ РАЗЛИЧИЙ

После активизации в главном меню программы SPSS опции **Analyze** (Анализ) выбираем в выпадающем меню строку **Compare Means** (Сравнение средних). Опция «Сравнение средних» содержит 5 разделов: средние (**Means...**), Т-критерий Стьюдента для одной выборки (**One-Sample T Test...**), Т-критерий Стьюдента для независимых выборок (**Independent Samples T Test...**), Т-критерий Стьюдента для парных выборок (**Paired-Samples T Test...**), однофакторный дисперсионный анализ (**One-Way ANOVA...**). В данном руководстве не рассматриваются сравнение средних и однофакторный дисперсионный анализ.

Напомним, что в основе параметрических статистических критериев лежит расчет таких параметров распределения, как среднее арифметическое и дисперсия. Для корректного использования параметрических критериев, экспериментальные данные должны соответствовать следующим требованиям:

- результаты измерений должны быть представлены в шкале интервалов;
- распределение результатов измерений не должно значимо отличаться от нормального распределения (о проверке соответствия эмпирического распределения нормальному смотри в разделе «Непараметрические критерии достоверности различий», критерий Колмогорова—Смирнова для одной выборки).

Т-критерий Стьюдента для одной выборки (**Analyze > Compare Means > One-Sample T Test...**) используется в тех случаях, когда необходимо проверить достоверность отличия результатов группы испытуемых от единственного заданного значения.

Вызвав соответствующее базовое окно (**One-Sample T Test**), необходимо в левом подокне выбрать одну или несколько переменных и с помощью кнопки-указателя (с треугольником) переместить их в подокно **Variable(s)** (Переменные). Далее, в строке **Test Value** (Проверяемое значение) необходимо напечатать интересующее вас число, отличие которого от результатов группы вы хотите проверить. Кнопку **Options...** (Опции) можно не трогать, оставив принятые по умолчанию в программе настройки доверительного интервала и пропущен-

ных значений. Для завершения расчетов нажмите кнопку ОК, и вы получите файл вывода результатов SPSS Output.

В файле вывода результатов будут размещены две таблицы. В первой (One-Sample Statistics) содержатся описательные статистики, на основе которых рассчитывается критерий (среднее арифметическое, стандартное отклонение, стандартная ошибка средней). Во второй таблице (One-Sample Test) приводятся значение Т-критерия ( $t$ ), число степеней свободы ( $df$ ), уровень достоверности критерия (Sig), отличие средней арифметической величины групповых результатов от тестируемого значения (Mean Difference), а также верхняя и нижняя границы 95%-ного доверительного интервала обнаруженного отличия (95% Confidence Interval of the Difference). О достоверности различий между результатами группы и заданным значением можно говорить только в том случае, если уровень достоверности критерия (Sig) меньше чем 0,05.

В психологических исследованиях данный критерий удобно использовать, например, для проверки однородности группы. Предположим, что исследователя интересует, насколько коллектив научных сотрудников однороден по уровню интеллекта. Протестировав всех членов коллектива с помощью какого-либо известного теста для определения уровня развития интеллекта (например, WAIS), исследователь задает балл в качестве контрольного значения (Test Value), например 120, предполагая, что у всех членов научного коллектива интеллектуальный уровень должен быть выше среднего уровня (то есть выше 100 баллов по тесту Векслера). Если после вычисления Т-критерия уровень его достоверности (Sig) больше, чем 0,05, то можно сделать вывод об отсутствии значимых отличий результатов всех членов коллектива от проверяемого значения, то есть об однородности группы по уровню интеллекта.

Т-критерий Стьюдента для независимых выборок (Analyze > Compare Means > Independent Samples T Test...) используется для выявления достоверности различий между двумя независимыми группами на основе сравнения выборочных средних.

После вызова соответствующего базового окна (Independent Samples T Test, см. рис. 39) необходимо в левом подокне выбрать одну или несколько переменных и с помощью кнопки-указателя (с треуголь-

ником) переместить их в подокно **Test Variable(s)** (Проверяемые переменные).

Далее, с помощью кнопки-указателя необходимо выбрать и переместить в соответствующее подокно группирующую переменную (**Grouping Variable**), в которой на основании значений программа выделит две сравниваемые группы.

С помощью кнопки **Define Groups...** (Задать группы) исследователь получает возможность определить сравниваемые группы двумя способами: 1) с использованием конкретных значений (**Use specified values**) и 2) с использованием точки разбиения (**Cut point**).

В первом случае необходимо задать два конкретных значения группирующей (**Group 1** и **Group 2**), в соответствии с которыми все испытуемые будут поделены на две группы. Данный способ определения сравниваемых групп удобен, если группирующая переменная является номинальной или ранговой. Классическим случаем является приведенный на рис. 39 пример, в котором в качестве группирующей переменной выступает пол испытуемых и значение «1» (**Group 1**) соответствует мужчинам, а значение «2» (**Group 2**) — женщинам.

Если группирующая переменная является интервальной (например, возраст испытуемых), то задавать сравниваемые группы удобнее с использованием точки разбиения (**Cut point**). В данном случае, напечатав в соответствующей строке (**Cut point**) конкретное число (напри-

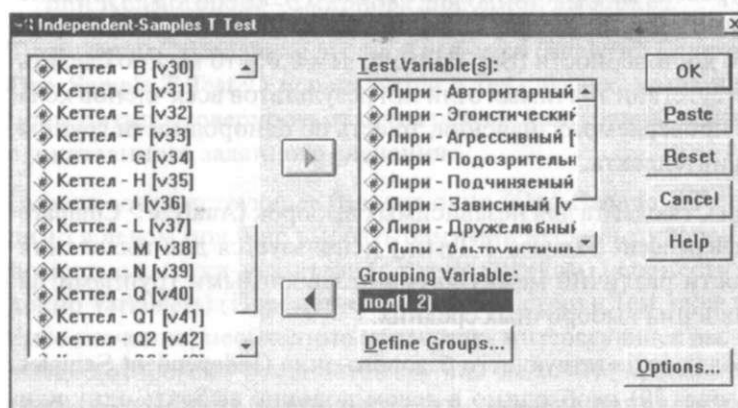


Рис. 39. Окно расчета Т-критерий Стьюдента для независимых выборок

мер, 33 года), мы получим разбиение испытуемых по возрасту на две группы. При этом в первую группу войдут испытуемые, имеющие возраст 33 года и больше, во вторую — 32 года и меньше.

Кнопку **Options...** (Опции) можно не использовать, оставив принятые по умолчанию в программе настройки доверительного интервала и пропущенных значений. Для завершения расчетов нажмите кнопку **ОК**.

В файле вывода результатов содержатся две таблицы. В первой таблице (**Group Statistics**) находятся описательные статистики для каждой из сравниваемых групп, на основе которых рассчитывается критерий (среднее арифметическое, стандартное отклонение, стандартная ошибка средней). Во второй таблице (**Independent Samples Test**) приводятся: критерий Ливина (**Levene's Test for Equality of Variances**), его значение (**F**) и достоверность (**Sig.**). Критерий Ливина (в отечественной литературе также называемый критерием Фишера) выявляет достоверность различия между группами на основе сопоставления дисперсий двух изучаемых групп. Далее в таблице приводится Т-критерий Стьюдента (**t-test for Equality of Means**), включая его значение (**t**), число степеней свободы (**df**), уровень достоверности критерия (**Sig.**), разница между средними арифметическими сравниваемых групп (**Mean Difference**), стандартная ошибка приведенной разницы (**Std. Error Difference**), а также верхняя и нижняя границы 95%-ного доверительного интервала для этой разницы (**95% Confidence Interval of the Difference**). Обратите внимание, что Т-критерий подсчитывается в двух вариантах (строки таблицы **Independent Samples Test**) с учетом общей дисперсии для двух групп (**Equal variances assumed**) и с учетом дисперсий по группам (**Equal variances not assumed**); второй способ расчета является более точным.

Т-критерий Стьюдента — один из любимых методов выявления достоверности различий между группами испытуемых в психологических исследованиях. При этом начинающим исследователям необходимо помнить как о математической корректности применения этого критерия (см. выше), так и о психологической корректности его использования. Психологическая корректность заключается в тщательно обоснованной независимости сравниваемых между собой групп, что обеспечивается за счет правильной организации исследования (например, засчитывать результаты одного испытуемого можно только в одной из сравниваемых групп).

Кроме того, сам признак, на основании которого выделяются независимые группы испытуемых, не должен быть напрямую связан с теми показателями, различия в которых мы пытаемся выявить. Например, пытаюсь обнаружить достоверные различия в стиле межличностных отношений по тесту ДМО у мужчин и женщин (см. пример на рис. 39), мы предполагаем, что не существует прямой и явной зависимости между полом и стилем межличностных отношений. Если бы подобная связь существовала в явном виде, то применение критерия не имело бы смысла.

Т-критерий Стьюдента для парных выборок (**Analyze > Compare Means > Paired-Samples T Test...**) используется для выявления достоверности различий между результатами измерений, проведенных с помощью одной и той же методики на одной и той же группе испытуемых. Классическим является использование Т-критерия Стьюдента для обнаружения степени изменений, проявившихся после экспериментального воздействия. При этом сравниваются результаты тестов, полученные до и после эксперимента на одной и той же группе испытуемых.

В психологических исследованиях использование Т-критерия Стьюдента для парных выборок также удобно для выявления достоверности различий между оценками различных объектов, данных одной и той же группой испытуемых с помощью одной и той же оценочной (самооценочной) шкалы. В примере на рис. 40 демонстрируется выявление различий по параметру маскулинности (по оценочной шкале андрогинии BSRI Сандры Бем), которые по мнению испытуемых существуют у мужчин и женщин руководителей.

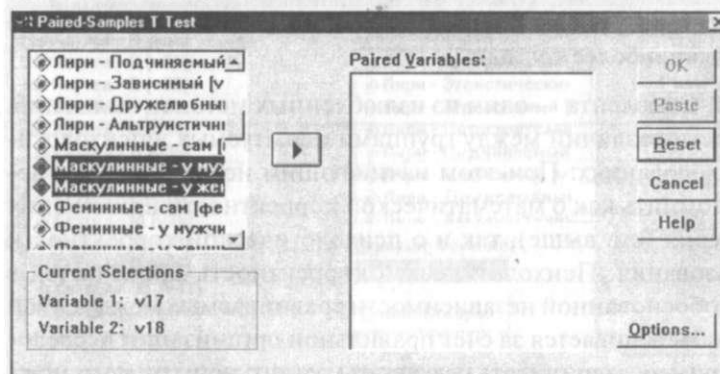


Рис. 40. Окно расчета Т-критерия Стьюдента для парных выборок

После вызова базового окна (**Paired-Samples T Test;**) (см. рис. 40) необходимо в подокне слева выбрать две сравниваемые между собой переменные. Для этого достаточно просто щелкнуть мышью сначала на одной, потом на другой переменной — их обозначение появится в разделе **Current Selections — Variable 1** (Текущий выбор). После того как вы выбрали обе сравниваемые переменные, с помощью кнопки-указателя переместите сформированную пару в подокно **Paired Variables:** (Парные переменные). По желанию в этом подокне можно задать сравнение одной или нескольких пар, при этом в различных парах имена переменных могут повторяться.

Кнопку **Options...** (Опции) можно не использовать, оставив принятые по умолчанию в программе настройки доверительного интервала и учета пропущенных значений. Для завершения расчетов нажмите кнопку **OK**, и программа автоматически сгенерирует и откроет файл вывода результатов **SPSS Output**.

Файл вывода результатов содержит три таблицы: 1) описательные статистики для парных выборок (**Paired Samples Statistics**), в которой приводятся среднее арифметическое, стандартное отклонение, стандартная ошибка средней; 2) коэффициент параметрической корреляции Пирсона (**Paired Samples Correlations**), включая значение коэффициента корреляции (**Correlations**) и его достоверность (**Sig.**); 3) Т-критерий Стьюдента для парных выборок (**Paired Samples Test**), включая расчет парных различий (**Paired Differences**) — средней арифметической (**Mean**), стандартного отклонения (**Std. Deviation**), стандартной ошибки средней (**Std. Error Mean**), верхняя и нижняя границы 95%-ного доверительного интервала для этих различий (**95% Confidence Interval of the Difference**), а также само значение критерия (**t**), число степеней свободы (**df**), уровень достоверности критерия (**Sig.**).

## НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ ДОСТОВЕРНОСТИ РАЗЛИЧИЙ

После активизации в главном меню SPSS опции **Analyze** (Анализ), выбираем в дополнительном меню строку **Nonparametric Tests** (Непараметрические критерии). Опция «Непараметрические критерии» включает в себя 8 разделов: расчет критерия Хи-квадрат (**Chi-**



Square...), биномиального критерия для дихотомических данных (Binomial...), критерия серий (Runs...), критерия Колмогорова—Смирнова для одной выборки (1-Sample K-S...), ряда критериев для двух независимых выборок (2 Independent Samples...), для ^независимых выборок (K Independent Samples...), для двух связанных выборок (2 Related Samples...), для .^связанных выборок (ATRelated Samples...).

В основе непараметрических статистических критериев лежит оперирование частотами или рангами эмпирических данных, при этом, в отличие от параметрических критериев, тип распределения данных не обязательно должен соответствовать нормальному. Для расчета непараметрических критериев результаты измерений должны быть представлены в шкале наименований, рангов или в шкале интервалов (если распределение интервальных данных значимо отличается от нормального, что бывает довольно часто при малом размере выборки). Тип распределения данных при использовании непараметрических критериев может быть любым, но при этом необходимо учитывать ограничения, специфичные для некоторых критериев.

В разделе непараметрических критериев метод Хи-квадрат (Analyze > Nonparametric Tests > Chi-Square...) используется в качестве критерия согласия, то есть позволяет проверить соответствие (пропорциональность) наблюдаемых частот значений изучаемого признака ожидаемым частотам. При этом ожидаемые частоты могут задаваться по умолчанию, исходя из теоретического предположения о равномерном распределении результатов при их случайном появлении (то есть при отсутствии какой-либо закономерности в полученных данных), или назначаться по желанию исследователя, исходя из принятого им предположения о том, какое именно распределение результатов должно получиться после эксперимента.

В самом простом варианте Хи-квадрат рассчитывается следующим образом: после вызова базового окна (Chi-Square Tests) (см. рис. 41) необходимо в подокне слева выбрать одну или несколько переменных и с помощью кнопки-указателя переместить их в подокно Test Variable List (Список проверяемых переменных). Для завершения вычисления критерия Хи-квадрат необходимо нажать кнопку ОК.

В файл вывода результатов входят частотные таблицы (Frequencies) для каждой из проверяемых переменных, при этом в боковой графе таб-

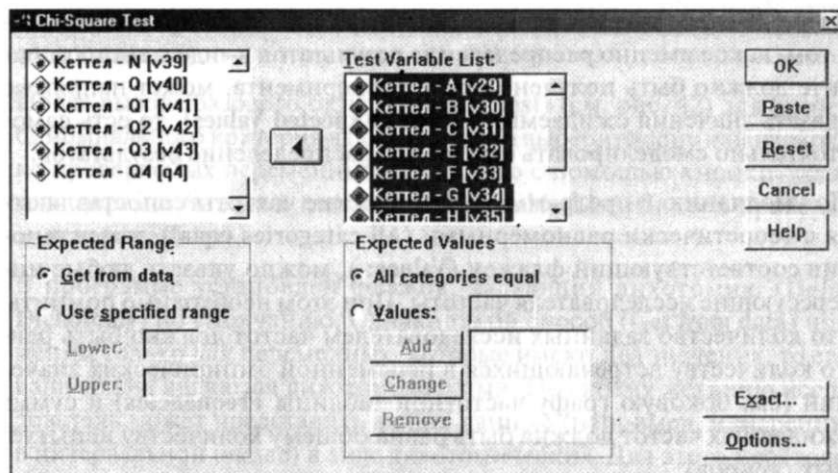


Рис. 41. Окно расчета непараметрического критерия Хи-квадрат

лицы содержатся значения, которые может принимать переменная, а в шапке таблицы — реальные частоты (Observed N), ожидаемые частоты (Expected N) и разница между ними (Residual). После частотных таблиц в файле вывода данных приводится общая таблица (Test Statistics), где по строкам указаны значение критерия Хи-квадрат (Chi-Square), степени свободы (df) и статистическая достоверность (Asymp. Sig.). В данном случае, чем четче выражена статистическая достоверность полученного критерия (Asymp. Sig. должна быть равна 0,05 и меньше), тем больше распределение значений переменной отличается от равномерного, то есть является не случайным.

По желанию исследователя расчет критерия может быть модифицирован. Например, он может изменить диапазон эмпирических значений, на основании частоты встречаемости которых будут рассчитываться ожидаемые частоты (Expected Range; по умолчанию учитываются все значения — Get from data). Для этого необходимо установить флажок Use specified range (Использовать конкретный диапазон) и указать нижнюю (Lower) и верхнюю (Upper) границы диапазона. Все эмпирические значения, оказавшиеся за пределами указанного диапазона, не будут учитываться при расчете критерия. Подобное ограничение диапазона позволяет повысить точность расчетов, например, за счет отбраковки крайних, редко встречающихся минимальных и максимальных значений.

Кроме того, исследователь, исходя из принятого им предположения о том, какое именно распределение результатов в «идеальном» варианте должно быть получено после эксперимента, может напрямую указать значения ожидаемых частот (Expected Values), то есть самостоятельно смоделировать ожидаемое распределение результатов.

По умолчанию в программе эмпирические частоты сопоставляются с теоретически равномерными (All categories equal), но, установив соответствующий флажок (Values: ), можно указать любые интересующие исследователя частоты. При этом необходимо помнить, что количество заданных исследователем частот должно быть равно количеству встречающихся в переменной эмпирических значений (см. боковую графу частотной таблицы Frequencies)^ сумма ожидаемых частот должна быть равна общему количеству испытуемых (кейсов).

Напомним, что корректно рассчитать критерий Хи-квадрат можно только на большой выборке (более 30 человек) или на данных с малой вариативностью значений, которые может принимать изучаемая переменная. Критерий также не может быть корректно рассчитан, если ожидаемая частота для ячеек таблицы — менее 5. В данном случае программа SPSS под таблицей Test Statistics автоматически сообщает о том, что определенное число ячеек содержит менее 5 ожидаемых частот (например, «10 cells (100%) have expected frequencies less than 5»). В подобном случае необходимо либо собрать большее количество измерений, либо уменьшить вариативность значений, объединив их в группы (категории).

Кнопки Exact... (Точность) и Options... (Опции) пользователь может не трогать, оставив принятые в программе по умолчанию способы расчета доверительного интервала и учета пропущенных значений.

Биномиальный критерий для дихотомических данных (Analyze > Nonparametric Tests > Binomial...) позволяет сравнить наблюдаемые частоты двух значений дихотомической переменной (или любых других переменных, представленных в виде дихотомических) с теоретически ожидаемыми частотами. В качестве условия сравнения выступает соответствие эмпирического соотношения частот теоретическому биномиальному распределению с заданным параметром вероятности. Основное назначение биномиального критерия — выявить, насколько значимо эмпирические частоты (одного из значе-

ний дихотомической переменной) превышают заданную вероятность их проявления.

После вызова базового окна (Binomial Test) (см. рис. 42) для расчета биномиального критерия в подокне слева выберите одну или несколько проверяемых переменных, после чего с помощью кнопки-указателя переместите их в подокно Test Variable List (Список проверяемых переменных).

В программе установлен флажок, задающий дихотомию (Define Dichotomy) по умолчанию. Однако такой способ (Get from data) применим только для переменных, которые имеют два значения, то есть изначально являются дихотомическими. По своему желанию исследователь может представить любые данные (например, измеренные в интервальной шкале) в виде дихотомических. Для этого необходимо установить флажок Cut point: (Точка разбиения) и указать в активизированном поле конкретное значение. Все значения переменной равные ему, и меньшие значения будут отнесены к первой категории, все значения больше указанного — ко второй категории.

Обратите внимание на пример использования точки разбиения, приведенный на рис. 42. Исследование, посвященное лидерским качествам сотрудников, проводилось на выборке взрослых трудоспособных испытуемых, и перед исследователем возник вопрос о сбалансированности выборки по возрастному составу. В качестве точки разбиения был выбран возраст 33 года, в результате чего в первую

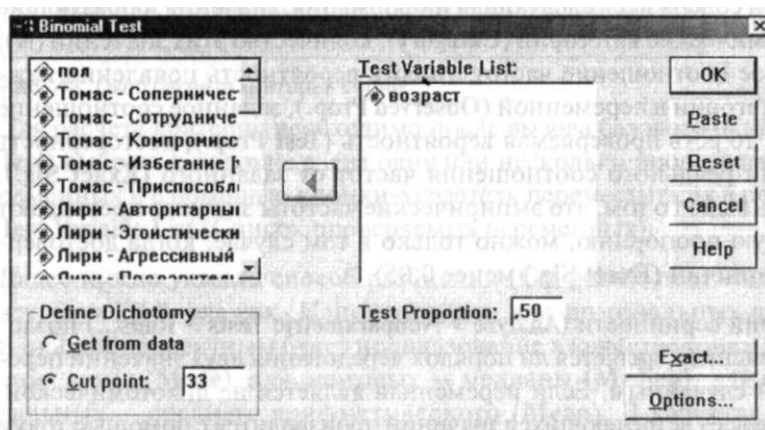


Рис. 42. Окно расчета биномиального критерия

категорию вошли испытуемые на стадии ранней взрослости (33 года и моложе), во вторую категорию — на стадии средней взрослости (34 года и старше).

Далее необходимо установить проверяемый параметр вероятности (Test Proportion:), который в программе по умолчанию равен 0,5. Это означает, что оба значения проверяемой переменной должны появляться равное количество раз (классический пример — бросание монетки, где при достаточном количестве бросков и орел, и решка должны появляться с вероятностью 0,5). В примере на рис. 42 проверяется предположение о том, что соотношение испытуемых в возрасте ранней и средней взрослости является равным (то есть 0,5), сбалансированным.

Проверяемая вероятность (Test Proportion:) может меняться от 0,001 до 0,999 для первого значения переменной. В свою очередь, вероятность для второго значения переменной будет, соответственно, меняться от 1-0,001 до 1-0,999.

Пользователь может не трогать кнопки Exact... (Точность) и Options... (Опции), оставив принятые в программе по умолчанию способы расчета доверительного интервала и учета пропущенных значений. Для завершения процедуры расчета биномиального критерия необходимо нажать на кнопку ОК.

В файле вывода результатов SPSS Output будет содержаться единственная таблица с результатами расчетов (Binomial Test), в столбцах которой содержится следующая информация: значения, образующие дихотомические категории (Category), количество этих значений (N), реальное соотношение частот, то есть вероятность появления каждой категории в переменной (Observed Prop.), заданное соотношение частот, то есть проверяемая вероятность (Test Prop.), достоверность отличий реального соотношения частот от заданного (Exact Sig.). (Делать вывод о том, что эмпирические частоты значительно превышают заданную пропорцию, можно только в том случае, когда достоверность отличий (Exact Sig.) менее 0,05).

Критерий серийности (Analyze > Nonparametric Tests > Runs...) позволяет выяснить, является ли порядок чередования двух значений переменной случайным. Если переменная является не дихотомической, деление всех встречающихся значений производится с помощью точки разбиения (Cut point). Результаты исследований, в которых встречается

ся достаточно много (или мало) серий — повторяющихся последовательностей значений — являются не случайными, а тенденциозными.

В психологии данный критерий может быть использован для выявления социальной значимости в ответах испытуемых на вопросы тестов. На рис. 43 приведен пример использования критерия серий для выявления социальной желательности в ответах руководителей предприятий на вопросы об организационных ценностях, принятых на этих предприятиях. Характерно то, что результаты ответов на вопрос о значимости таких ценностей, как «Власть», «Престиж», «Богатство» ит. п. оказались достаточно разнообразными (случайными, поскольку тенденция отсутствует), зато ценность «Общественное благо» в ответах прослеживается как явная тенденциозность (достоверность критерия *Asymp. Sig* = 0,008).

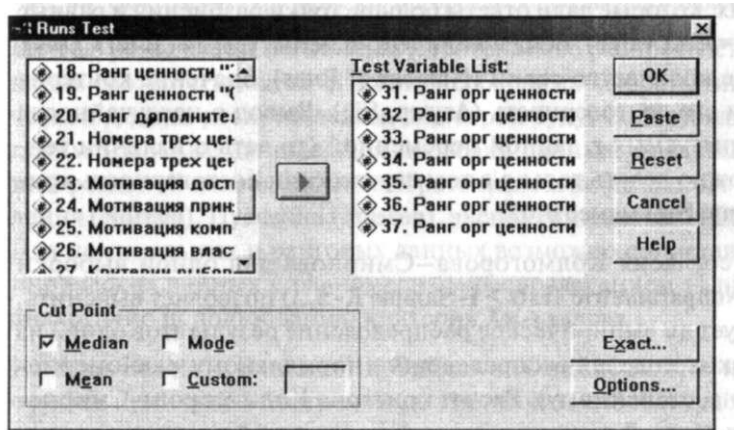


Рис. 43. Окно расчета критерия серий<sup>4</sup>

Для расчета критерия необходимо после вызова базового окна (Runs Test) выбрать в подокне слева одну или несколько проверяемых переменных и с помощью кнопки-указателя переместить их в подокно Test Variable List (Список проверяемых переменных).

Далее нужно указать способ разбиения (Cut point), поставив соответствующий флажок. Напомним, что для номинальных данных наиболее корректным будет использование в качестве точки разбиения моды (Mode), для ранговых — медианы (Median), для интервальных — среднего арифметического (Mean). В качестве точки разбиения можно использовать и любое значение по выбору ис-

следователя (Custom:), что в некоторых случаях удобно для работы с интервальными данными.

Пользователь может не трогать кнопки Exact... (Точность) и Options... (Опции), оставив принятые в программе по умолчанию способы расчета доверительного интервала и учета пропущенных значений. Для завершения процедуры расчета биномиального критерия необходимо нажать на кнопку ОК.

В файле вывода результатов SPSS Output появится только одна таблица с результатами расчетов (Runs Test), в ячейках которой находится следующая информация: значение выбранной точки разбиения для данной переменной (Test Value), количество испытуемых, которые дали ответы меньше точки разбиения ( $Cases < Test Value$ ), количество испытуемых, которые дали ответы больше точки разбиения и равные ей ( $Cases \geq Test Value$ ), общее количество испытуемых (Total Cases), выявленное количество серий (Number of Runs), значение критерия серий (Z) и его достоверность (Asymp. Sig). Вывод о неслучайности ответов испытуемых в данной переменной (то есть о наличии тенденции) можно делать только в том случае, если достоверность критерия (Asymp. Sig) менее 0,05.

Критерий согласия Колмогорова-Смирнова для одной выборки (Analyze > Nonparametric Tests > 1-Sample K-S...) позволяет выяснить, соответствует ли эмпирическое распределение результатов одной из теоретических моделей распределения (нормальному, равномерному, Пуассона, степенному). Расчет критерия Колмогорова-Смирнова является важной предварительной процедурой в анализе экспериментальных данных, полученных с помощью интервальных измерительных шкал, но на малых выборках. Именно этот критерий позволяет оценить, соответствует ли распределение подобных результатов нормальному, и в случае обнаружения такого соответствия дает «зеленый свет» для применения к таким данным параметрических методов статистического анализа.

Для расчета критерия после вызова базового окна (One-Sample Kolmogorov-Smirnov Test) (см. рис. 44) необходимо в подокне слева выбрать одну или несколько проверяемых переменных и с помощью кнопки-указателя переместите их в подокно Test Variable List (Список проверяемых переменных).

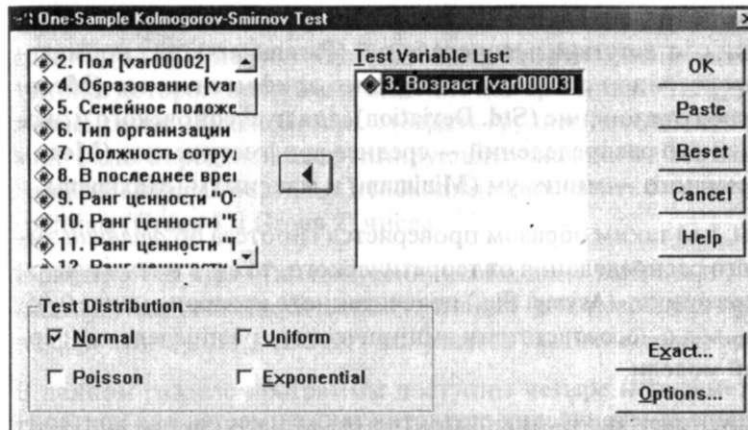


Рис. 44. Окно расчета критерия Колмогорова–Смирнова для одной выборки

Далее в разделе **Test Distribution** (Проверяемое распределение), установив соответствующий флажок, необходимо выбрать тип теоретической модели распределения, с которой будут сопоставляться экспериментальные данные. Для проверки данных, представленных в интервальной шкале, можно выбрать модели нормального распределения (**Normal**), Пуассона (**Poisson**), экспоненциального (**Exponential**). Для номинальных и ранговых данных возможно сопоставление эмпирических данных с равномерным распределением (**Uniform**), что аналогично использованию критерия Хи-квадрат.

Кнопки **Exact...** (Точность) и **Options...** (Опции) можно не трогать, оставив принятые в программе по умолчанию способы расчета доверительного интервала и учета пропущенных значений. Для завершения процедуры расчета критерия Колмогорова–Смирнова для одной выборки необходимо нажать на кнопку **ОК**.

В файле вывода результатов SPSS Output появится несколько таблиц (отдельно для каждого проверяемого типа распределения) с результатами расчетов (**One-Sample Kolmogorov-Smirnov Test**). В строках каждой таблицы содержится следующая информация: количество испытаний (**N**), параметры, лежащие в основе проверки соответствия распределений (**Parameters**), наиболее сильные отклонения от данных параметров (**Most Extreme Differences**; абсолютное — **Absolute**; положительное — **Positive**; отрицательное — **Negative**), само значение критерия Колмогорова–Смирнова (**Kolmogorov-Smirnov Z**) и его



достоверность (Asymp. Sig.). В качестве параметров, лежащих в основе проверки соответствия распределений (Parameters), для нормального распределения используются среднее арифметическое (Mean) и стандартное отклонение (Std. Deviation), для пуассоновского и экспоненциального распределений — среднее арифметическое (Mean), для равномерного — минимум (Minimum) и максимум (Maximum).

Напомним, что таким образом проверяется гипотеза об *отличии* эмпирического распределения от теоретического, то есть в тех случаях, когда достоверность (Asymp. Sig.) подсчитанного критерия менее 0,05, делается вывод о несоответствии эмпирического распределения теоретической модели.

В разделе непараметрических статистик также имеется ряд критериев (Analyze > Nonparametric Tests > 2 Independent Samples...), позволяющих обнаружить достоверные различия между результатами исследований двух независимых выборок.

После вызова базового окна (Two-Independent-Samples Tests) (рис. 45) необходимо в подокне слева выбрать одну или несколько переменных и с помощью кнопки-указателя переместить их в подокно Test Variable List (Список проверяемых переменных).

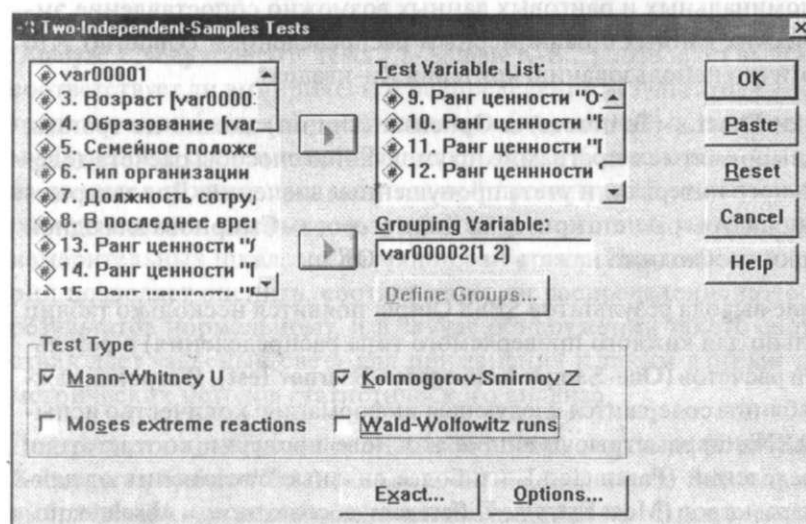


Рис. 45. Окно расчета критериев достоверности различий для двух независимых выборок

Далее необходимо выбрать и переместить в соответствующее подокно группирующую переменную (Grouping Variable:), на основании значений в которой программа выделит две сравниваемые группы. Нажав на кнопку Define Groups... (Задать группы), исследователь вызывает окно определения группирующих значений (Two Independent Samples: Define Groups), где необходимо напечатать соответствующие группам (Group 1 и Group 2) числа.

В разделе Test Type (Тип критерия) необходимо выбрать и с помощью флажка указать статистический критерий (или несколько критериев), на основании которого будет выявляться различие между группами.

В данном разделе программы доступны четыре непараметрических критерия проверки достоверности различий: U-критерий Манна-Уитни (Mann-Whitney U; включает в себя также расчет критерия Вилкоксона), Z-критерий Колмогорова—Смирнова (Kolmogorov-Smirnov Z), критерия крайних значений Мозеса (Moses extreme reactions), критерий серий Вальда—Вольфовица (Wald-Wolfowitz runs). Напомним, что данные критерии корректно использовать только для выявления достоверности различий между ранговыми и интервальными данными.

Наиболее удобными и распространенными в психологических исследованиях являются U-критерий Манна—Уитни (Mann-Whitney U) и Z-критерий Колмогорова—Смирнова (Kolmogorov-Smirnov Z).

Критерий Манна—Уитни проверяет сходство расположения эмпирических распределений результатов двух сравниваемых групп. Результаты обеих групп объединяются в общий ранжированный ряд, и если количество чередующихся значений («перекрещиваний») обеих групп достаточно велико, то можно сделать вывод о схожести данных. Если же количество «перекрещиваний» невелико, то речь пойдет о значительном несовпадении в расположении результатов групп, то есть о достоверности различий между ними.

Критерий Колмогорова—Смирнова в данном разделе проверяет достоверность различий между двумя эмпирическими распределениями посредством выявления максимальной точки расхождения между накопленными частотами значений. Если максимальная точка расхождения между накопленными частотами превышает критическое значение, то различия между двумя эмпирическими распределениями можно считать достоверными.

Кнопки Exact... (Точность) и Options... (Опции) можно не использовать, оставив принятые в программе по умолчанию способы расчета доверительного интервала и учета пропущенных значений. Для завершения процедуры расчета критериев достоверности различий между двумя выборками необходимо нажать на кнопку ОК.

В файле вывода результатов SPSS Output содержится несколько таблиц — по две для каждого указанного критерия. В первой таблице даны промежуточные статистические показатели, на основании которых рассчитывается критерий. Например, для U-критерия Манна—Уитни в первой таблице Ranks (Ранги) указаны количество испытуемых в каждой из сравниваемых групп (N), средний ранг результатов испытуемых каждой группы в общем ранжированном ряду (Mean Rank), сумма рангов для каждой группы (Sum of Ranks). Во второй (основной) таблице приведены результаты расчета критерия.

Для U-критерия Манна—Уитни в основной таблице (Test Statistics) по строкам приводится следующая информация: значение U-критерия Манна-Уитни (Mann-Whitney U), значение W-критерия Вилкоксона, который является эквивалентом критерия Манна—Уитни, но более точен при сравнении равных по количеству групп испытуемых (Wilcoxon W), Z-оценка, которая вычисляется в результате нормальной аппроксимации и позволяет наглядно сопоставлять между собой результаты проверки достоверности различий между группами по различным переменным (Z), достоверность выявленного различия по критерию Манна—Уитни (Asymp. Sig.). Напомним, что делать вывод о существовании достоверных различий между двумя сравниваемыми группами можно только в том случае, если статистическая достоверность (Asymp. Sig.) различий менее 0,05.

Для критерия Колмогорова-Смирнова в основной таблице (Test Statistics) по строкам приводится следующая информация: наиболее сильные различия между двумя сравниваемыми группами (Most Extreme Differences; абсолютное — Absolute; положительное — Positive; отрицательное — Negative), само значение критерия Колмогорова-Смирнова (Kolmogorov-Smirnov Z) и его достоверность (Asymp. Sig.). Вывод о существовании достоверных различий между двумя сравниваемыми группами также делается только в том случае, если статистическая достоверность (Asymp. Sig.) различий менее 0,05.

Для нескольких независимых выборок (Analyze > Nonparametric Tests > K Independent Samples...) также существует ряд критериев, позволяющих выявить достоверность различий между ними по одной или нескольким переменным.

После вызова базового окна (Test for Several Independent Samples) (рис. 46) в подокне слева необходимо выбрать одну или несколько переменных и с помощью кнопки-указателя переместить их в подокно Test Variable List (Список проверяемых переменных).

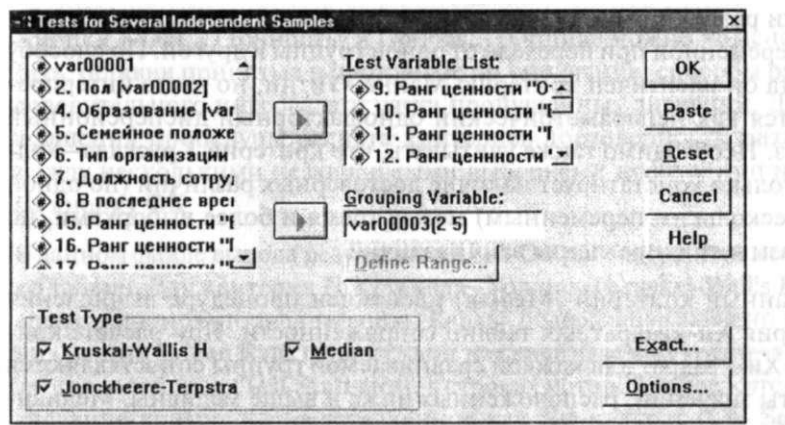


Рис. 46. Окно расчета критериев достоверности различий для нескольких независимых выборок

Далее, необходимо выбрать и переместить в соответствующее подокно группирующую переменную (Grouping Variable:), на основании значений в которой программа сформирует сравниваемые группы. Нажав на кнопку Define Range... (Задать диапазон), исследователь вызывает окно определения группирующих значений (Several Independent Samples: Define Range), где нужно напечатать числа, задающие нижнюю (Minimum) и верхнюю (Maximum) границы значений, определяющих сравниваемые группы.

Заданный диапазон должен включать три и более значений, однако не может быть слишком большим, в противном случае использование критериев будет мало информативным. Значения, оказавшиеся за границами заданного диапазона, в расчете критерия учитываться не будут. При расчете критериев Крускала—Уоллиса и медианного в качестве группирующей переменной можно использовать номинальные

и ранговые данные; для расчета критерия Джонкира–Терпстры — только ранговые.

После определения группирующей переменной необходимо выбрать тип критерия (**Test Type**), с помощью которого будут выявляться различия между группами. В данном разделе программы доступны критерии Н Крускала–Уоллиса (**Kruskal-Wallis H**), медианный (**Median**) и Джонкира–Терпстры (**Jonckheere-Terpstra**).

Критерий Н Крускала–Уоллиса (Kruskal-Wallis H) используется для оценки различий между несколькими группами по уровню изменений переменной при переходе от одной группы к другой. По способу расчета он идентичен критерию Манна–Уитни, но часто интерпретируется как непараметрический однофакторный дисперсионный анализ. Необходимо также учитывать, что критерий Крускала–Уоллиса только констатирует наличие достоверных различий (по одной или нескольким переменным) между тремя и более выборками, но не указывает направление этих различий.

Медианный критерий (Median) идентичен процедуре вычисления критерия Хи-квадрат для таблиц сопряженности. При расчете критерия Хи-квадрат для каждой сравниваемой группы сопоставляются частоты значений, расположенных ниже и выше медианы. Медианный критерий, по сравнению с критериями Крускала–Уоллиса и Джонкира–Терпстры, является наиболее общим и слабым. Для этого критерия также справедливы все те ограничения, которые необходимо учитывать при расчете критерия Хи-квадрат (например, частота в ячейках таблицы сопряженности должна быть не менее 5).

Критерий Джонкира–Терпстры (Jonckheere-Terpstra) используется не только для оценки различий между несколькими группами по уровню изменений переменной при переходе от одной группы к другой, но и выявляет тенденцию (направление) этих различий. Напомним, что в качестве группирующей переменной (**Grouping Variable:**) для вычисления критерия может быть использована только ранговая переменная, то есть предполагается выделение групп, упорядоченных по возрастанию какого-либо качества. В примере на рис. 46 в качестве группирующей выступает переменная «Возрастные группы» (на рисунке — var00003), имеющая диапазон от 2 до 5. Двойка соответствует возрастной группе 21–30 лет, тройка — 31–40 лет и т. д., то есть ранги строго упорядочены по возрастанию.

Критерий Джонкира–Терпстры в данном разделе является наиболее сильным, т. к. не только определяет, существуют ли достоверные различия между группами, но и указывает на то, что эти различия также упорядочены по возрастанию. По сути дела, критерий Джонкира–Терпстры близок к корреляционным показателям связи между двумя переменными и позволяет делать вывод, что выявленные различия между группами по проверяемым переменным связаны именно с тем качеством, которое легло в основу категоризации (то есть с группирующей переменной).

Кнопки **Exact...** (Точность) и **Options...** (Опции) можно не использовать, оставив принятые в программе по умолчанию способы расчета доверительного интервала и учета пропущенных значений. Для завершения процедуры расчета критериев достоверности различий между несколькими независимыми выборками необходимо нажать на кнопку **ОК**.

В данном файле вывода результатов SPSS Output содержится несколько таблиц. Для критерия Н Крускала–Уоллиса (**Kruskal-Wallis H**) выводится промежуточная таблица (**Ranks**), содержащая значения средних рангов (**Mean Rank**) для каждой из сравниваемых групп, а также итоговая таблица (**Test Statistics**), в строках которой содержится само значение критерия, представленное как Хи-квадрат (**Chi-Square**), число степеней свободы (**df**) и статистическая достоверность различий (**Asymp. Sig.**). Вывод о существовании достоверных различий между всеми сравниваемыми группами делается только в том случае, если статистическая достоверность (**Asymp. Sig.**) различий менее 0,05.

Для медианного критерия (**Median**) выводится промежуточная таблица (**Frequencies**) с частотами значений, которые больше медианы или меньше, или равны ей, для каждой из сравниваемых групп. Ниже располагается итоговая таблица (**Test Statistics**), в строках которой содержится информация о количестве испытуемых (**N**), медиане (**Median**), значении критерия Хи-квадрат (**Chi-Square**), число степеней свободы (**df**) и статистическая достоверность различий (**Asymp. Sig.**). Вывод о существовании достоверных различий между всеми сравниваемыми группами делается только в том случае, если статистическая достоверность (**Asymp. Sig.**) различий менее 0,05.

Для критерия Джонкира–Терпстры выводится только одна общая таблица с результатами вычислений (**Jonckheere-Terpstra Test**), в которой

содержится следующая информация: количество категорий в группирующей переменной (**Number of Levels in**), общее количество испытуемых (**N**), эмпирическое значение критерия (**Observed J-T Statistic**), среднее значение критерия (**Mean J-T Statistic**), стандартное отклонение критерия (**Std. Deviation of J-T Statistic**), нормализованное значение критерия (**Std. J-T Statistic**), достоверность (**Asymp. Sig.**) критерия. Для принятия гипотезы о существовании достоверных различий между группами необходимо, чтобы выявленная достоверность (**Asymp. Sig.**) различий была менее 0,05.

Для определения направления связи (аналогично корреляции) удобно использовать нормализованное значение критерия (**Std. J-T Statistic**), которое является положительным в случае прямой связи и отрицательным — в случае обратной.

В разделе непараметрических статистик существует ряд критериев (**Analyze > Nonparametric Tests > 2 Related Samples...**), позволяющих выявить различия между результатами измерений, проведенных с помощью одной и той же методики на одной и той же группе испытуемых. Подобного рода критерии часто используются для выявления значимых изменений, возникших после экспериментального воздействия, при этом сравниваются результаты замеров, сделанные до и после эксперимента на одной и той же группе испытуемых (см.

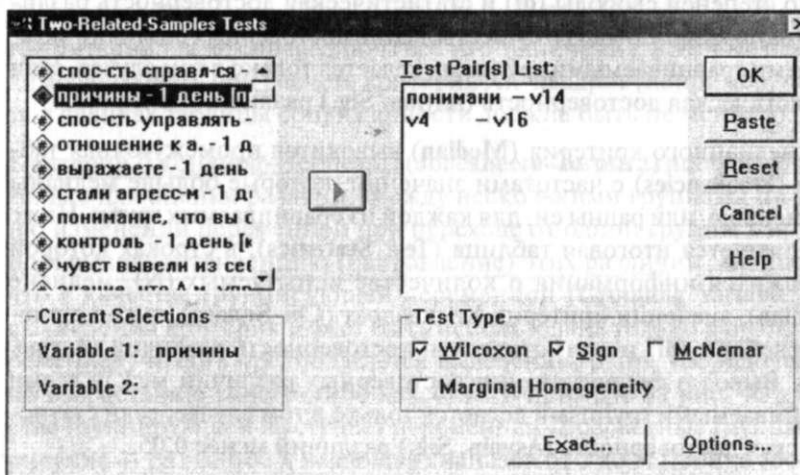


Рис. 47. Окно расчета критериев достоверности различий для двух связанных выборок

также аналогичное использование параметрического Т-критерия Стьюдента для парных выборок.)

После вызова базового окна (**Two-Related-Samples Test**) (см. рис. 47) в подокне слева необходимо выбрать две сравниваемые между собой переменные. Для этого нужно щелкнуть мышью сначала на первой переменной, и ее обозначение **Variable 1** (Текущий выбор) появится в разделе **Current Selections**; затем — на второй переменной, и ее обозначение **Variable 2** (Текущий выбор) также появится в разделе **Current Selections**. После того как вы выбрали обе сравниваемые переменные, с помощью кнопки-указателя переместите сформированную пару в подокно **Test Pair(s) List:** (Список проверяемых пар). По желанию в подокне **Test Pair(s) List:** можно задать сравнение одной или нескольких пар, при этом в различных парах имена переменных могут повторяться.

Далее, необходимо выбрать критерий (**Test Type**), с помощью которого будет выявляться достоверность различий между переменными. В этом разделе программы предлагается использование четырех критериев: критерия Вилкоксона (**Wilcoxon**), критерия знаков (**Sign**), критерия Мак Немары (**McNemar**) и критерия маргинальной однородности (**Marginal Homogeneity**).

- ✓ Критерий Вилкоксона (**Wilcoxon**) является наиболее мощным из всех предлагаемых критериев и выявляет не только степень (достоверность) различий между переменными, но и направление этих различий. Для корректного вычисления он требует интервальных или ранговых данных и наиболее информативен в тех случаях, когда исходные данные варьируют в достаточно широком диапазоне.
- ✓ Критерий знаков (**Sign**) является наиболее слабым эквивалентом критерия Вилкоксона, однако он более эффективен в тех случаях, когда сравниваемые данные варьируют в небольшом диапазоне и не имеют схожего распределения. Этот критерий также требует, чтобы исходные данные были интервальными или ранговыми.
- ✓ Критерии Мак Немары (**McNemar**) и маргинальной однородности (**Marginal Homogeneity**) являются более специализированными. Оба разработаны на базе математического аппарата критерия Хи-квадрат. Критерий Мак Немары выявляет достоверность различий между двумя парными дихотомическими переменными, а критерий маргинальной однородности — между двумя ранговыми переменными.



После выбора соответствующего критерия для завершения процедуры вычисления достоверности различий между двумя связанными выборками необходимо нажать на кнопку **ОК**.

В файле вывода результатов SPSS Output содержится по две таблицы для каждого указанного критерия. В первой таблице указаны промежуточные статистические показатели, на основании которых рассчитывается критерий. Например, для критерия Вилкоксона в первой таблице **Ranks** (Ранги) указаны количество негативных (**Negative Ranks**), позитивных (**Positive Ranks**) и равных рангов (**Ties**), а также общее количество ранжируемых значений (**Total**). Для позитивных и негативных рангов вычислен средний ранг (**Mean Rank**) и сумма рангов (**Sum of Ranks**). Под первой таблицей содержатся важные комментарии, отображающие способ расчета рангов.

Для критерия знаков приводится аналогичная таблица, но содержащая не ранги, а частоты (**Frequencies**) позитивных (**Positive Differences**), негативных (**Negative Differences**) и отсутствующих (**Ties**) сдвигов в значениях сравниваемых переменных.

Во второй (основной) таблице приведены результаты расчета критерия (**Test Statistics**). Для критерия Вилкоксона приводится значение критерия (**Z**) и его статистическая достоверность (**Asymp. Sig.**). Для критерия знаков сразу указывается достоверность различий (**Exact Sig.**) между сравниваемыми переменными. Напомним, что принимать гипотезу о существовании значимых различий между сравниваемыми парными измерениями (переменными) можно только в том случае, если статистическая достоверность (**Asymp. Sig.** или **Exact Sig.**) менее 0,05.

Направление различий, выявляемое с помощью критерия Вилкоксона, можно определить по знаку самого критерия (**Z**), но гораздо проще и удобнее сделать это с помощью первой таблицы **Ranks** (Ранги). Если, например, при соответствующей достоверности самого критерия (**Asymp. Sig. < 0,05**) количество отрицательных рангов (**Negative Ranks**) преобладает над количеством положительных (**Positive Ranks**), то мы можем достоверно утверждать, что результаты испытуемых по изучаемой переменной снизились.

Ряд непараметрических критериев (**Analyze > Nonparametric Tests > K Related Samples...**) позволяет оценить различия между результата-

ми нескольких повторных измерений, проведенных с помощью одной методики на одной и той же группе испытуемых.

После вызова базового окна (**Test for Several Related Samples**) (рис. 48) в подокне слева необходимо выбрать несколько сравниваемых между собой переменных. Учтите, что количество сравниваемых переменных должно быть более трех несмотря на то, что расчет доступных нам критериев возможен и для двух переменных (в случае сравнения двух связанных переменных приведенные выше критерии Вилкоксона и др. являются более мощными). Выбранные переменные с помощью кнопки-указателя перемещаются в подокно **Test Variables:** (Проверяемые переменные).

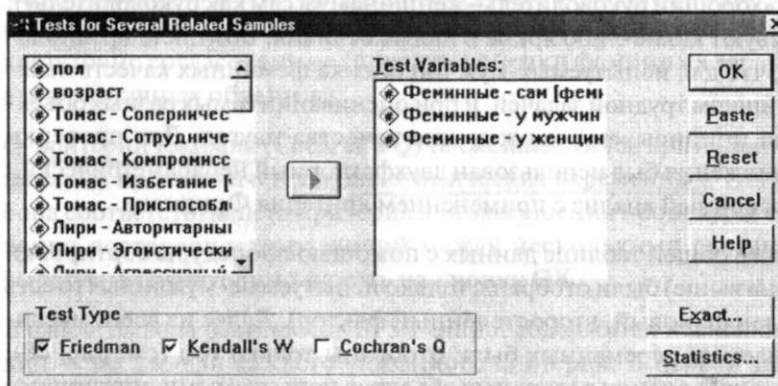


Рис. 48. Окно расчета критериев достоверности различий для нескольких связанных выборок

Далее необходимо выбрать критерий (**Test Type**), с помощью которого будет выявляться достоверность различий между несколькими связанными переменными. В данном случае имеется возможность расчета трех критериев: критерия Фридмана (**Friedman**), W-критерия Кендалла (**Kendall's W**) и Q-критерия Кохрана (**Cochran's Q**).

Наиболее популярным для выявления достоверности различий трех и более измерений, произведенных на одной и той же выборке испытуемых, является критерий Фридмана (**Friedman**). В некоторых руководствах критерий Фридмана рассматривается как метод двухфакторного непараметрического дисперсионного анализа, позволяющего оценить эффект влияния главного исследуемого фактора на фоне другого (второстепенного) фактора. В данном случае столбцы (переменные)

соответствуют значениям главного фактора, а строки (испытуемые) — значениям второстепенного фактора. Для расчета критерия можно использовать только интервальные и ранговые данные.

Приведем пример. Цель проводимого гипотетического исследования — выявление различий в представлениях мужчин- и женщин-руководителей о выраженности маскулинных и феминных качеств в личности идеального «хорошего руководителя» (включая аутостереотип). Исследование проводилось с помощью оценочной шкалы андрогинии BSRI Сандры Бем. После подсчета описательных статистик (см. выше) выяснилось, что у мужчин по одной из шкал BSRI, «феминность» в оценках всех трех сравниваемых объектов («хороший руководитель—мужчина», «хороший руководитель—женщина», «я сам как руководитель»), отсутствуют какие-либо яркие и явные отличия. Возникло предположение, что для испытуемых-мужчин оценка феминных качеств является слишком трудной задачей, и при оценивании самых разных объектов они приписывают им женские качества наугад. Для проверки предположения был использован двухфакторный непараметрический дисперсионный анализ с применением критерия Фридмана.

Сначала в общей таблице данных с помощью процедуры сортировки (описана выше) были отобраны только испытуемые-мужчины (то есть был задан фоновый, второстепенный фактор). Далее из всего массива изучаемых переменных были отобраны только три (см. рис. 48), отражающие оценки различных объектов испытуемыми-мужчинами по параметру феминности (то есть был задан главный фактор). В качестве конкретных методов анализа достоверности различий были выбраны критерии Фридмана и Кендалла (см. рис. 48). По критерию Фридмана была выявлена высокая достоверность в подсчитанном различии (Asymp. Sig. = 0,003) между оценками трех объектов по параметру феминности. Таким образом, предположение о том, что оценки мужчин (пол — фоновый фактор), данные различным объектам по феминным качествам (главный фактор), являются случайными (не отличающимися друг от друга), не подтвердилось. Мужчины являются «компетентными» экспертами в оценке феминных качеств различных объектов, несмотря на отсутствие наглядных отличий в описательных статистиках.

W-критерий Кендалла (Kendall's W) аналогичен критерию Фридмана, но кроме расчета непосредственно критерия Фридмана он

включает в себя также расчет нормализованного W-критерия, который также часто называют критерием конкордантности (согласованности оценок) Кендалла. Этот критерий интерпретируется как согласие (единообразие, систематичность, тенденциозность) испытуемых в оценивании ими всех сравниваемых объектов (переменных). Значение W-критерия может изменяться от нуля, в случае полного отсутствия согласия между испытуемыми, до единицы, в случае полного единообразия в оценках. В рассмотренном выше примере несмотря на то, что были выявлены достоверные различия в оценивании трех различных объектов по параметру феминности, было выявлено значение W-критерия, равное 0,7. Это позволяет сделать предположение о том, что испытуемые-мужчины имеют некую имплицитную «теорию феминности», которая позволяет им давать достаточно согласованные (похожие) оценки феминных характеристик различных объектов.

Q-критерий Кохрана (Cochran's Q) также аналогичен критерию Фридмана, но предназначен для дихотомических переменных. После выбора соответствующего критерия для завершения процедуры вычисления достоверности различий между несколькими связанными выборками необходимо нажать на кнопку ОК.

В файле вывода результатов SPSS Output содержится несколько таблиц — по две для каждого указанного критерия. В первой таблице указаны промежуточные статистические показатели, на основании которых рассчитывается критерий. Например, для критерия Фридмана в первой таблице Ranks (Ранги) указаны средние ранги (Mean Rank) для каждой из сравниваемых переменных.

В основной таблице с результатами расчетов (Test Statistics) для критерия Фридмана приводится следующая информация: количество испытуемых (N), значение критерия, представленное как Хи-квадрат (Chi-Square), число степеней свободы (df) и статистическая достоверность различий (Asymp. Sig.). Вывод о существовании достоверных различий между всеми сравниваемыми группами делается только в том случае, если статистическая достоверность (Asymp. Sig.) различий менее 0,05.

Для W-критерия Кендалла (Kendall's W) в итоговой таблице (Test Statistics), кроме всех вычисленных для критерия Фридмана показателей (см. выше), рассчитывается само значение W-критерия (Kendall's W).

## КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

В основном меню программы SPSS в опции **Analyze > Correlate** (Корреляционный анализ) представлены следующие статистические методы: вычисление парных корреляций (**Bivariate...**), частных корреляций (**Partial...**) а также вычисление расстояний между переменными, аналогично кластерному или факторному анализу (**Distances...**). В данном руководстве рассматривается только процедура расчета парных корреляций (**Analyze > Correlate > Bivariate...**), которая наиболее распространена и востребована в психологических исследованиях.

Основное назначение корреляционного анализа — выявление корреляционной связи между двумя или более изучаемыми переменными. Корреляционная связь — это совместное согласованное изменение двух изучаемых характеристик. Данная совместная изменчивость обладает тремя основными характеристиками: формой, направлением и силой.

По форме такая связь может быть линейной или нелинейной. Наиболее удобной для выявления и последующей интерпретации корреляционных связей является линейная форма. (Следует заметить, что рассматриваемый в данном руководстве расчет парных корреляций адекватно описывает именно линейные корреляции.) Отсутствие линейной корреляции (близость коэффициента корреляции к нулю) не говорит о том, что переменные не зависимы, между ними может существовать более сложная нелинейная связь. В программе SPSS для выявления криволинейной корреляционной связи могут быть использованы методы вычисления частных корреляций (**Partial...**), а также вычисление корреляционных расстояний (**Distances...**).

В случае линейной корреляционной связи можно выделить два основных направления: положительное («прямая связь») и отрицательное («обратная связь»). В случае положительной корреляционной связи между переменными можно говорить об их совместной однонаправленной изменчивости, то есть с ростом значений одной переменной увеличиваются значения и другой и, наоборот, с уменьшением значений первой переменной уменьшаются значения и другой. В случае отрицательной корреляционной связи между переменными можно говорить об их совместной разнонаправленной изменчивости, то есть с ростом значений одной переменной значения другой

уменьшаются и соответственно с уменьшением значений первой переменной значения другой увеличиваются. После расчетов направление корреляционной связи отображается как знак, стоящий перед вычисленным значением коэффициента корреляции: для положительной связи это знак «+», для отрицательной — знак «-».

Напомним также, что корреляционная связь является двунаправленной, то есть не позволяет сделать однозначный вывод о том, какая именно из переменных является причиной обнаруженной совместной изменчивости, так как их вклад в расчеты равноправен. Исследователю следует помнить, что причиной совместной изменчивости может быть влияние либо одной из двух анализируемых переменных, либо их совместное взаимодействие, либо воздействие какого-либо внешнего фактора (или комплекса факторов).

К интерпретации выявленных корреляционных связей следует подходить крайне осторожно, избегая категоричные фразы типа «переменная X является причиной увеличения показателя Y». Даже если с содержательной точки зрения для исследователя очевидно, какой именно показатель «должен быть» причиной, а какой следствием, подобные утверждения следует формулировать как предположения, которые кроме интуиции исследователя должны быть строго обоснованы теоретически.

Важнейшей характеристикой корреляционной связи кроме формы и направления является ее сила (степень, теснота). Сила связи напрямую указывает, насколько ярко проявляется совместная изменчивость изучаемых переменных и может принимать значение от нуля (связь не обнаружена) до единицы (предельно сильная связь; совместная изменчивость, прослеживаемая во всех парах значений сравниваемых переменных). Как правило, после проведения расчетов исследователь из всего массива полученных корреляций отбирает только наиболее сильные (независимо от знака, то есть направления корреляции), которые в дальнейшем интерпретируются. Критерием для отбора «достаточно сильных» корреляций может быть как абсолютное значение самого коэффициента корреляции (обычно «сильной» считается корреляция от 0,7 до 1), так и относительная величина этого коэффициента, определяемая по уровню статистической значимости (**Significance level** от 0,01 до 0,1), зависящему от размера выборки. Делая вывод из опыта работы, автору хотелось бы заметить, что в ма-

лых выборках для дальнейшей интерпретации корректнее отбирать сильные корреляции на основании уровня статистической значимости. Для исследований, проведенных на больших выборках, лучше использовать сильные корреляции по абсолютному значению (от 0,7 и выше), в противном случае исследователь рискует утонуть в огромном массиве статистически значимых корреляций.

Проведение корреляционного анализа с помощью программы SPSS — простая и удобная операция.

После вызова базового окна (**Analyze > Correlate > Bivariate...**) (рис. 49) нужно в подокне слева выбрать две или более переменных, которые необходимо прокоррелировать. Затем с помощью кнопки-указателя можно переместить выбранные переменные в подокно справа (**Variables:**).

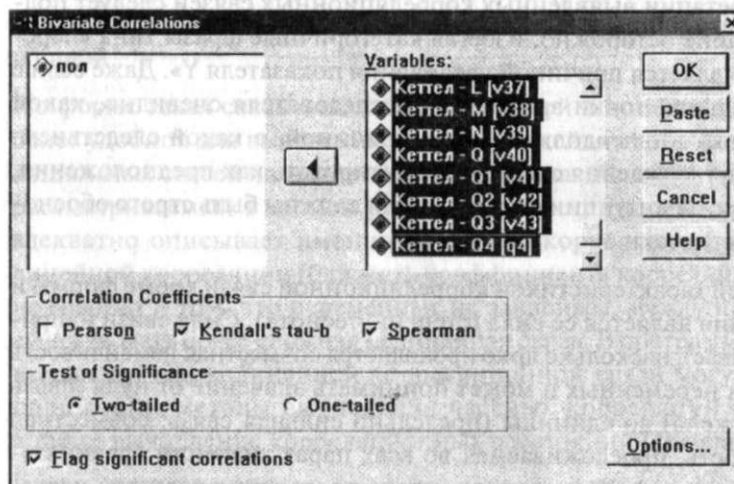


Рис. 49. Окно расчета парных корреляций

Далее необходимо выбрать метод (**Correlation Coefficients**), с помощью которого будет выявляться корреляционная связь между переменными. Можно взять следующие критерии: коэффициент линейной корреляции Пирсона (**Pearson**) и коэффициенты ранговой корреляции тау-б Кендалла (**Kendall's tau-b**) и Спирмена (**Spearman**). Напомним, что коэффициент линейной корреляции Пирсона (**Pearson**) является параметрическим методом и его корректное применение возможно только в том случае, если результаты измерений представлены в шкале

интервалов, а само распределение значений в анализируемых переменных отличается от нормального в незначительной степени.

Коэффициент ранговой корреляции Спирмена (**Spearman**) является непараметрическим аналогом классического коэффициента корреляции Пирсона, но при его расчете учитываются не связанные с распределением показатели сравниваемых переменных (среднее арифметическое и дисперсия), а ранги. Коэффициент ранговой корреляции Кендалла (**Kendall's tau-b**), иногда называемый коэффициентом конкордации, является самостоятельным оригинальным методом, опирающимся на вычисление соотношения пар значений двух выборок, имеющих одинаковые или отличающиеся тенденции (возрастание или убывание значений). Для корректного вычисления обоих коэффициентов (Спирмена и Кендалла) результаты измерений должны быть представлены в шкале рангов или интервалов. Следует также заметить, что принципиальных отличий между двумя ранговыми корреляционными критериями не существует, но принято считать, что коэффициент Кендалла является более «содержательным», т. к. он более полно и детально анализирует связи между переменными, перебирая все возможные соответствия между парами значений, а коэффициент Спирмена более точно учитывает именно количественную степень связи между переменными.

После выбора требуемого коэффициента в окне **Correlation Coefficients** необходимо указать способ расчета статистической значимости (**Test of Significance**) вычисляемых коэффициентов. По умолчанию в программе принят двусторонний (**Two-tailed**) способ вычисления достоверности, который является более строгим, но исследователь по своему желанию сможет изменить его на односторонний (**One-tailed**). Подобное изменение приведет к появлению большего количества достоверных корреляций, требующих последующей интерпретации. Для большей наглядности и удобства во время работы с полученными результатами следует оставить установленный по умолчанию флажок **Flag significant correlations** (Отмечать значимые корреляции).

Кнопку **Options...** (Опции) можно не использовать, оставив принятые в программе по умолчанию способы расчета промежуточных статистик для коэффициента корреляции Пирсона и учета пропущенных значений. Для завершения процедуры корреляционного анализа необходимо нажать на кнопку **OK**.

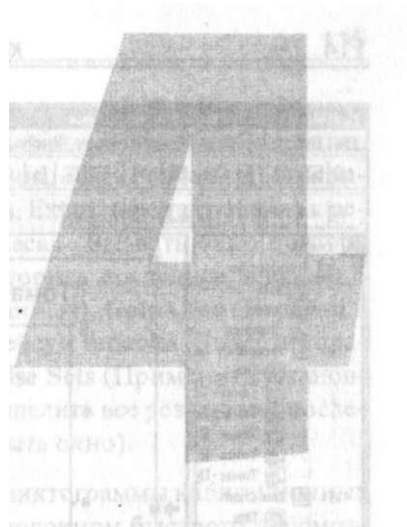


В файле вывода результатов корреляционного анализа SPSS Output для каждого рассчитываемого критерия содержится квадратная таблица (Correlations). В каждой ячейке таблицы приведены: само значение коэффициента корреляции (Correlation Coefficient), статистическая значимость рассчитанного коэффициента (Sig.), количество испытуемых (N). В шапке и боковой графе полученной корреляционной таблицы содержатся названия переменных.

Напомним, что диагональ (левый верхний — правый нижний угол) таблицы состоит из единиц, т. к. корреляция любой переменной с самой собой является максимальной, то есть равна единице. Таблица симметрична относительно этой диагонали (т. к. корреляция переменной А с переменной В тождественна корреляции В с А), поэтому нет смысла искать значимые «сильные» корреляции во всей таблице — достаточно лишь просмотреть треугольный фрагмент полученной матрицы над или под диагональю. Если вы оставили принятый в программе по умолчанию флажок «Отмечать значимые корреляции», то в итоговой корреляционной таблице будут отмечены статистически значимые коэффициенты: на уровне 0,05 и меньше — одной звездочкой (\*), а на уровне 0,01 — двумя звездочками (\*\*).

## **Глава 4**

# **РЕДАКТИРОВАНИЕ ОТЧЕТА В SPSS**



## **РЕДАКТИРОВАНИЕ ОТЧЕТА В РЕДАКТОРЕ SPSS SMART VIEWER**

В программе SPSS содержится встроенный редактор статистических отчетов SPSS Smart Viewer (рис. 50), который автоматически запускается в новом окне, как только программа завершила процедуру статистического анализа и сгенерировала файл вывода полученных результатов SPSS Output. Редактор статистических отчетов содержит все возможности для написания, редактирования и последующего вывода на печать предельно полного статистического отчета, включающего в себя таблицы, графики, текстовые комментарии.

В основном меню базового окна редактора отчетов SPSS находятся 10 опций. Опции File (Файл), Edit (Правка), View (Вид), Analyze (Анализ), Graphs (Графики), Utilities (Сервис), Window (Окно), Help (Помощь) являются идентичными одноименным опциям окна редактора данных программы SPSS (см. раздел «Общий вид программы»).

Опция Insert (Вставка) содержит в себе возможности: по внесению в файл отчета SPSS Output разрыва страницы (Page Break), и, наоборот, удалению разрыва страницы (Clear Page Break); по вставке нового верхнего колонтитула (New Heading), нового заголовка (New Title), нового примечания, нижнего колонтитула (New Page Title), текстового блока (New Text); интерактивной двухмерной диаграммы (Interactive 2-D Graph), интерактивной трехмерной диаграммы

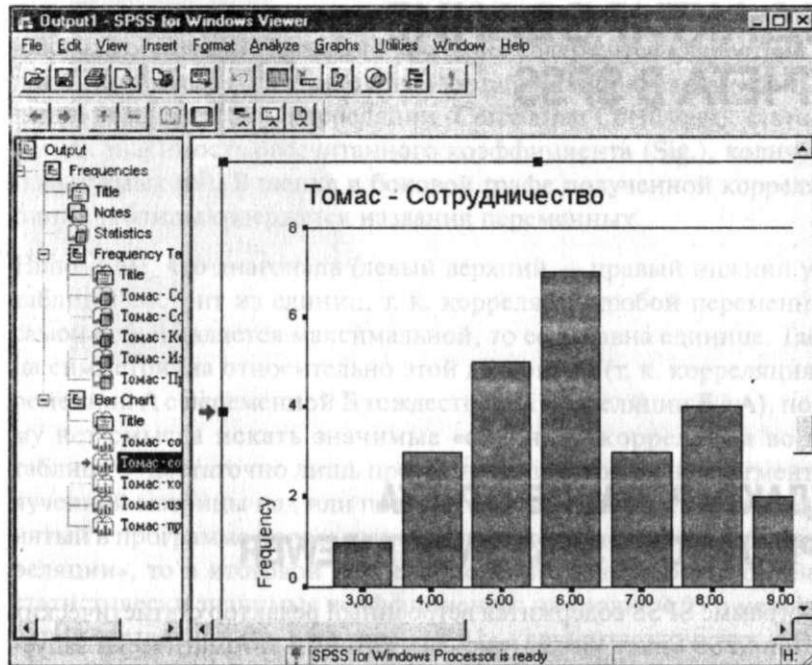


Рис. 50. Окно редактора статистического отчета программы SPSS

(**Interactive 3-D Graph**), диаграммы, созданной в старых версиях SPSS с использованием **Chart Carousel (Old Graph...)**, текстовых файлов SPSS Output, созданных в старых версиях SPSS или в режиме **Draft Viewer (Text File)**; по импорту любого объекта, созданного в любых Windows-приложениях, имеющих соответствующие драйвера импорта (**Object**).

Опция **Format (Формат)** предназначена для операций по выравниванию любых объектов SPSS Output (таблиц, текста, диаграмм) на странице. Она включает в себя возможность выравнивания по левому краю страницы (**Align Left**), этот способ принят в программе по умолчанию; по центру (**Center**) и по правому краю (**Align Right**). Подобные возможности форматирования имеются в любом текстовом редакторе и позволяют красиво оформить документ для печати.

Под строкой меню расположены две панели инструментов: первая содержит общие инструменты редактора отчетов, вторая — навигационные кнопки и некоторые возможности редактирования.

На первой панели инструментов находятся следующие пиктограммы (слева направо): Open File (Открыть файл), Save File (Сохранить файл), Print (Печать файла с результатами), Print Preview (Предварительный просмотр файла перед печатью), Export (Экспорт файла с результатами в другой формат), Dialog Recall (Вызвать окно повтора операций), Undo/Redo (Отменить/повторить последнее действие), Goto Data (Перейти в окно редактора данных), Goto Case (Перейти в окно редактора данных к конкретному кейсу), Variables (Вывести справочную информацию о переменных), Use Sets (Применить установки пользователя), Select Last Output (Выделить все результаты последних расчетов), Designate Window (Закреть окно).

Вторая панель инструментов содержит пиктограммы навигационных инструментов, облегчающих работу с подокном быстрого просмотра: кнопки прокрутки названий объектов (Promote и Demote), кнопки развертки (Expand) и сжатия (Collapse) уровней навигационного дерева, кнопки показа (Show) и скрытия (Hide) отдельных объектов, кнопки вставки верхнего колонтитула (Insert Heading), заголовка (Insert Title) и текстового поля (Insert Text).

Под строками инструментов расположено базовое окно (справа), отображающее результаты произведенных вычислений, и подокно быстрого просмотра (слева), которое позволяет увидеть древовидную структуру объектов (таблиц, текста, диаграмм), входящих в файл с результатами. Следует отметить, что один итоговый файл SPSS Output может содержать в себе результаты применения нескольких различных методов статистического анализа или результаты нескольких повторных (с модификация<sup>м</sup> применений одного и того же метода. Следовательно, файл SPSS Output может содержать множество объектов и иметь очень сложную структуру. Подокно быстрого просмотра обеспечивает максимальное удобство в просмотре именно больших файлов с результатами и максимальную быстроту при переходе к нужному объекту. Достаточно один раз щелкнуть мышью на названии объекта в подокне быстрого просмотра, и в базовом окне тут же будет отображен этот объект.

С помощью подокна быстрого просмотра также можно легко и быстро изменить расположение объектов (таблиц, текста, диаграмм) в файле SPSS Output. Для этого нужно щелкнуть мышью на объекте и, удерживая левую клавишу мыши нажатой, перетащить его на новое

место в подокне быстрого просмотра. Например, на рис. 50 в окне редактора отчетов содержатся результаты вычисления описательных статистик, при этом сначала расположена общая таблица со статистиками (Statistics), далее идут частотные таблицы для каждой из изучаемых переменных (Frequency Tables), далее — столбиковые диаграммы для каждой переменной (Bar Chart). По желанию исследователя каждая из диаграмм может быть перенесена и расположена под соответствующей частотной таблицей.

В подокне быстрого просмотра можно выделить сразу несколько объектов файла, расположенных рядом или в различных местах (для этого необходимо удерживать нажатой клавишу <Ctrl> и щелкнуть мышью на выделяемых объектах). После выделения объекты SPSS Output можно удалить, скопировать, экспортировать или использовать любые другие стандартные действия, содержащиеся в опциях File (Файл) или Edit (Правка).

Подокно быстрого просмотра помогает также ускорить процесс написания статистического отчета, поскольку здесь можно осуществить быструю вставку любого текстового элемента — заголовка, колонтитула, текстового поля. Для этого достаточно выбрать нужный объект в подокне быстрой навигации, а далее для вставки требуемого элемента использовать возможности опции Insert (Вставка) или аналогичные кнопки на второй панели инструментов. Вставка любого текстового элемента (включая заголовки и верхние колонтитулы) всегда осуществляется *ниже* активного (выбранного щелчком мыши) объекта.

Ранее уже упоминалось, что в файле результатов SPSS Output содержатся объекты трех типов:

1. Таблицы.
2. Диаграммы.
3. Текстовые элементы (к ним относятся заголовки, колонтитулы и поля текстовых комментариев).

Далее кратко рассмотрим основные возможности по редактированию каждого из трех типов объектов.

Сначала в подокне быстрой навигации нужно выбрать объект, который необходимо отредактировать. Для активации объекта достаточно просто щелкнуть на его названии мышью, после этого объект бу-

дет отображен в главном окне редактора отчетов. Признаками того, что объект активирован, являются красная стрелка слева от него (см. на рис. 51), а также тонкая черная рамка вокруг.

Statistics		Названия шкал теста			
		Томас - Соперн ичество	Томас - Сотрудн ичество	Томас - Компромисс	Томас Избеган
N	Valid Missing	25 0	25 0	25 0	25 0
Среднее арифметическое		4,4000	6,1200	7,4400	6,7000
Медиана		4,0000	6,0000	8,0000	6,0000
Мода		3,00	6,00	10,00	6,00
<b>Std. Deviation</b>		2,6458	1,6155	2,0429	2,1155
Skewness		1,127	,112	-,432	,112
Std. Error of Skewness		,464	,464	,464	,464
Kurtosis		1,388	-,645	-,690	-,645
Std. Error of Kurtosis		,902	,902	,902	,902
Minimum		1,00	3,00	3,00	3,00
Maximum		12,00	9,00	10,00	11,00

Рис. 51. Пример редактирования таблицы в редакторе отчетов SPSS

Для дальнейшего редактирования необходимо:

1. Дважды щелкнуть левой кнопкой мыши на активированном объекте.
2. Один раз щелкнуть правой кнопкой мыши и выбрать в самом низу открывшегося меню **SPSS Pivot Table Object > Edit** (Объект мобильная таблица SPSS > Правка).
3. Выбрать в главном меню редактора отчетов **Edit > SPSS Pivot Table Object > Edit** (Правка > Объект мобильная таблица SPSS > Правка). Готовый к редактированию объект будет выделен штриховой рамкой (см. рис. 51).

**Примечание.** Выше описана процедура перехода к редактированию объекта-таблицы. При переходе к редактированию текста вместо **Объект мобильная таблица SPSS** необходимо выбрать в аналогичном меню **Объект SPSS Rtf Document**, а при переходе к редактированию диаграммы — **SPSS Chart Object**.

При редактировании таблиц в отчете **SPSS Output** наиболее востребованной является процедура их «русификации», поскольку названия

всех рассчитанных статистических показателей обычно приведены по-английски. Кроме того, если при создании файла данных пользователь забыл указать метки переменных, то вместо них в файле результатов везде будут использованы технические обозначения переменных — var00001, var00002 и т. д. «Русификация» осуществляется весьма просто — нужно дважды щелкнуть левой кнопкой мыши на табличной ячейке, надпись в которой необходимо исправить. После этого ячейка будет активирована, а написанное в ней будет выделено синим цветом (на рис. 51 это надпись Std. Deviation). Далее можно просто «поверх» выделения напечатать требуемое русское название. Пример подобной работы с таблицей отражен на рис. 51. Обратите внимание, что кроме «русификации» боковой графы таблицы, содержащей названия вычисленных статистик, исправлены также вспомогательные надписи. Например, вместо заголовка Frequencies сделана подпись Описательные статистики для стиля поведения в конфликте, вместо надписи Variables внесена запись Названия шкал теста и т. п.

Обратите внимание, что при редактировании таблицы меняется вид основного меню базового окна редактора отчетов SPSS. Кроме новой опции Pivot (Вращение), добавляется ряд новых возможностей, связанных с редактированием и форматированием таблицы.

Одной из наиболее любопытных возможностей редактирования таблицы является построение диаграммы, отображающей информацию, которая содержится во всей таблице в целом, либо в отдельных ее строках (столбцах). Для построения диаграммы на основе таблицы необходимо сначала выделить одну или несколько строк (столбцов) с помощью левой кнопки мыши (при этом нужно удерживать нажатой клавишу <Ctrl>). После того как требуемые строки (столбцы) выделены, необходимо зайти в опцию Edit (Правка), в самом низу выбрать команду Create Graph (Создать диаграмму) и один из семи типов доступных диаграмм. В созданную диаграмму автоматически будут включены подписи данных, расположенные в боковой графе и шапке таблицы. Диаграмма будет расположена под таблицей-прототипом и в окне быстрой навигации появится соответствующий комментарий (Chart from Pivot Table). Полученная подобным образом диаграмма далее может быть отредактирована как любой графический объект SPSS Output.

Описательные статистики для стилей поведения в конфл

Статистики	Назван		
	Томас - Соперни- чество	Томас - Сотрудни- чество	Ков
Среднее арифметическое	4,4000	4,4000	
Медиана			
Мода			
Стандартное отклонение			
Асимметрия			
Стандартная ошибка асим			
Эксцесс			
Стандартная ошибка эксцесса	,902	,902	

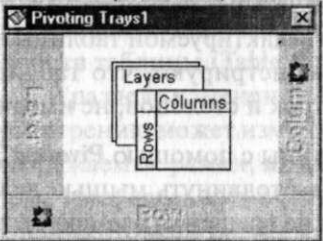


Рис. 52. Окно вращения элементов таблицы

Весьма мощные возможности редактирования таблиц содержатся в опции **Pivot** (Вращение). Данная операция позволяет добавить наиболее интересные для исследователя части таблицы, а также быстро поменять местами строки и столбцы (**Transpose Rows and Columns**).

Описательные статистики для стилей поведения в конфликте (по тесту Томаса)

	Томас - Соперничество	4,4000
	Томас - Сотрудничество	4,4000
Среднее арифметическое	Томас - Компромисс	6,1200
	Томас - Избегание	7,4400
	Томас - Сотрудничество	6,7600
	Томас - Соперничество	4,0000
Медиана	Томас - Избегание	6,0000
	Томас - Компромисс	8,0000
	Томас - Сотрудничество	6,0000
	Томас - Соперничество	5,0000
	Томас - Избегание	3,00
Мода	Томас - Сотрудничество	6,00
	Томас - Компромисс	10,00
	Томас - Избегание	6,00



Рис. 53. Пример использования возможностей окна вращения элементов таблицы

Если таблица многоуровневая, то порядок уровней также можно изменить. Вид таблицы можно поменять по своему усмотрению через перекомпоновку — слой в строки (**Move Layers to Rows**) или столбцы (**Move Layers to Columns**). Если пользователь не удовлетворен получившимся результатом, то можно вернуть таблице первоначальный вид (**Reset Pivots to Defaults**).



Строка Pivoting Trays (Панель вращения) открывает окно вращения элементов таблицы (рис. 52), позволяющее пользователю легко изменить внешний вид элементов таблицы. Окно вращения элементов имеет поля, соответствующие формальным элементам таблицы — слоям (Layer), строкам (Row), столбцам (Column), а также несколько значков (в виде кольца из четырех разноцветных стрелок), отражающих реальную структуру редактируемой таблицы. В примере на рис. 52 значки-указатели демонстрируют, что таблица имеет простую структуру и состоит из строк и столбцов, не имея вложенных слоев.

Для изменения вида таблицы с помощью Pivoting Trays (Панели вращения) достаточно лишь отодвинуть мышью любой из значков на соответствующее поле окна вращения элементов таблицы. Для этого нужно навести курсор на значок и, удерживая левую кнопку мыши, перетащить на требуемое поле. Пример приведен на рис. 53: значения столбцов были перенесены в строки. Новая таблица получилась вытянутой в высоту (а не в ширину, как было), что удобно для вывода на печать. Кроме того, скомпонованные подобным образом одинаковые статистические показатели по разным шкалам теста удобнее сравнивать между собой.

На рис. 54 демонстрируется пример превращения простой таблицы в динамическую, в которой столбцы представлены в виде слоев. Одна-

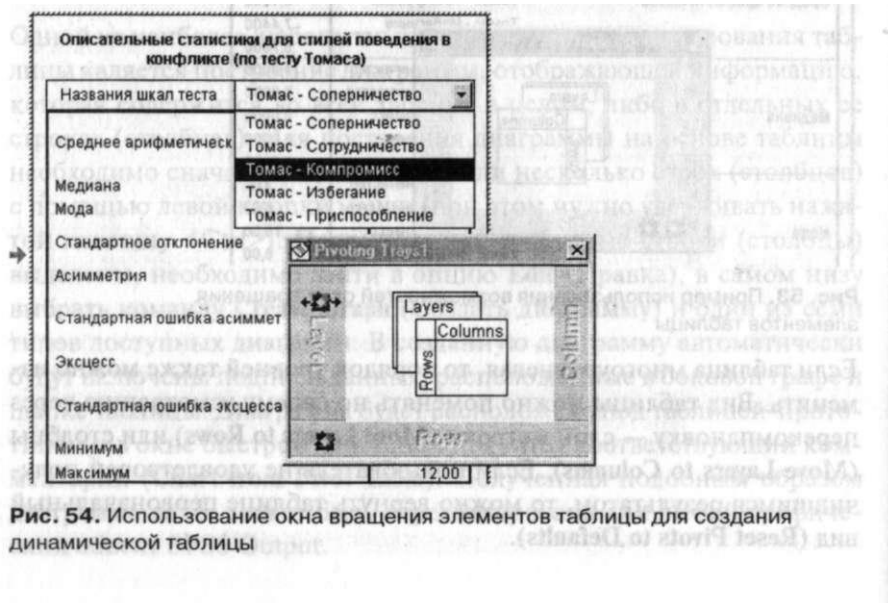


Рис. 54. Использование окна вращения элементов таблицы для создания динамической таблицы

ко следует помнить, что с подобного рода динамической таблицей удобно работать только в редакторе отчетов SPSS Output (компактность, интерактивность), но такую таблицу невозможно распечатать или экспортировать.

В опции Format (Форматирование) также имеются дополнительные возможности редактирования таблицы. Особый интерес представляет изменение свойств таблицы (Table Properties...) и ее внешнего вида (Table Looks...). В разделе изменения свойств таблицы пользователь по своему усмотрению может изменить такие параметры, как размеры ячеек, отображаемые рамки, их цвет, выравнивание текста в ячейках и др. Кроме того, здесь есть очень важный раздел, связанный с предпечатной подготовкой таблицы (Printing), позволяющий, например, подогнать большую таблицу по ширине или высоте страницы или определить характер переноса таблицы с одной страницы на другую. В разделе изменения внешнего вида таблицы (Table Looks...) содержится несколько десятков готовых стилей оформления, которые можно применить к редактируемой таблице (в редакторе MS Word аналогичная возможность содержится в опции Таблица > Автоформат).

Из личного опыта работы с программой автор осмелится сделать вывод, что в подавляющем большинстве случаев все доступные в опции Format (Форматирование) дизайнерские изыски являются избыточными. Например, для оформления научных отчетов вполне достаточно использовать принятый в программе SPSS по умолчанию стиль оформления таблиц, ничего в нем не меняя. При использовании готовых стилей оформления (Table Looks...) также возможно возникновение проблем с корректным отображением таблицы, т. к. в шаблонах стилей содержатся английские шрифты, которые могут не соответствовать установленным в системе русским шрифтам.

Статистические таблицы с результатами расчетов являются основными объектами, содержащимися в файле результатов статистического анализа SPSS Output. Кроме них, для более наглядного отображения произведенных статистических расчетов часто используется построение разного рода диаграмм. Для построения диаграмм можно напрямую воспользоваться соответствующей опцией базового меню — Graphs (Диаграммы) (в данном руководстве возможности этой опции не рассматриваются). Кроме того, возможности построения допустимых

диаграмм обычно включены в соответствующие разделы статистического анализа (например, в разделе описательных статистик имеется возможность построения столбиковых и круговых диаграмм и гистограмм). Различные диаграммы могут быть построены уже на основании имеющейся в файле SPSS Output таблицы с результатами (см. выше).

Для редактирования графического объекта в файле SPSS Output необходимо дважды щелкнуть на нем левой кнопкой мыши (или использовать любой из способов перехода к редактированию, описанных выше), после чего объект появится в новом окне графического редактора SPSS (SPSS for Windows Chart Editor).

В основном меню базового окна графического редактора SPSS (см. рис. 55) имеется ряд традиционных для всей программы SPSS опций (File, Edit, View, Analyze, Graphs, Help), а также несколько специальных опций, связанных непосредственно с редактированием графического объекта (диаграммы). К данным опциям относятся: Gallery (Галерея), в ней содержится набор доступных типов диаграмм, в которые может быть трансформирована исходная диаграмма; Chart (Диаграмма), предназначенная для редактирования общих свойств изображения — заголовков, легенды, вида осей, рамок и др.; Series (Серии), специализированная на создание серийных диаграмм; Format (Форматирование), здесь содержатся возможности по изменению внешнего вида отдельных элементов диаграммы (цвета, заливки, маркеров и др.).

Под основным меню базового окна графического редактора SPSS находятся две панели инструментов: в первой содержатся четыре пиктограммы, во второй — пятнадцать (рис. 55).

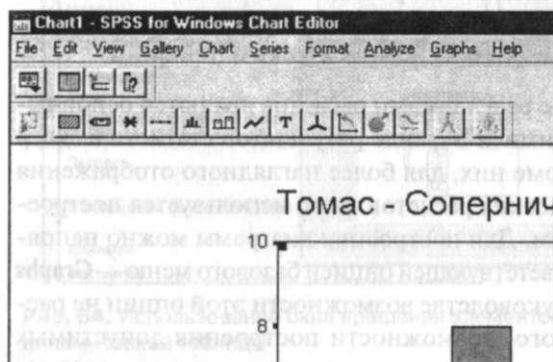


Рис. 55. Фрагмент окна графического редактора SPSS

В первой строке инструментов содержатся следующие пиктограммы (слева направо): Вызвать окно повтора операций (Dialog Recall), Перейти в окно редактора данных (Goto Data), Перейти в окно редактора данных к конкретному кейсу (Goto Case), Вывести справочную информацию о переменных (Variables).

Вторая панель содержит пиктограммы, соответствующие инструментам, предназначенным для изменения внешнего вида отдельных элементов диаграммы (идентичны содержащимся в меню Format (Форматирование): Point ID (Установить идентификатор точек), Fill Pattern (Изменить характер заливки изображения), Color (Изменить цвет графического элемента), Marker (Изменить стиль маркеров), Line Style (Изменить стиль линий), Bar Style (Изменить стиль столбиков диаграммы), Bar Label Style (Использовать подписи столбиков диаграммы), Interpolation (Построить интерполяционную кривую), Text (Добавить надпись), 3-D Rotation (Вращение диаграммы в трехмерном пространстве), Swap Axes (Поменять оси диаграммы местами), Explode Slice (Разделить изображение на части), Break Lines at Missing (Исключить линии как пропущенные), Chart options (Графические опции), Set/Exit spin mode (Установить/отключить режим вращения).

Для того чтобы изменить свойства всей диаграммы, можно воспользоваться опциями Gallery (Галерея) и Chart (Диаграмма). Для изменения свойств отдельных элементов необходимо: дважды щелкнуть левой кнопкой мыши по элементу, после чего автоматически откроется окно с доступными возможностями редактирования, или один раз щелкнуть левой кнопкой мыши по элементу диаграммы, а после этого воспользоваться опцией Format (Форматирование) или второй пиктографической панелью инструментов.

В заключение этого краткого обзора возможностей графического редактора SPSS хотелось бы заметить, что автоматически сгенерированные программой SPSS диаграммы чаще всего нуждаются в минимальном редактировании, например в замене английских подписей на русские.

Кроме таблиц и диаграмм в файле отчета SPSS Output также содержатся различные текстовые вставки — заголовки, текстовые блоки, примечания, колонтитулы. Ряд текстовых объектов ставится программой SPSS автоматически (например, заголовки каждого нового подуровня в древовидной структуре файла SPSS Output), но для придания

статистическому отчету законченного вида большую часть текстов приходится добавлять самому пользователю (например, текстовые аналитические комментарии к таблицам или диаграммам).

Для редактирования уже существующих текстовых элементов (например, заголовков), достаточно дважды щелкнуть на элементе левой кнопкой мыши, после чего текст будет выделен штриховой рамкой (активизирован), а наверху появится новая пиктографическая строка инструментов.

Возможности по редактированию текста в программе SPSS минимальны: пользователь по своему желанию может изменить используемый шрифт, его размер, характер начертания (жирный, курсив, подчеркнутый), цвет, а также характер выравнивания текста (по умолчанию — по левому краю, можно задать выравнивание по центру или по правому краю).

При создании нового текстового элемента (Insert Text) создается и отображается только одна строка текста. Это может смутить начинающего пользователя, т. к. при вводе большого количества текста часть его «пропадает». Для исправления этой ошибки необходимо только *один раз* щелкнуть левой кнопкой мыши по текстовому фрагменту, после чего он будет выделен тонкой рамкой. Далее нужно привести курсор на нижний край этой рамки и, удерживая левую кнопку мыши, потянуть край рамки вниз, пока весь текстовый блок не будет отображен полностью. Возможности по редактированию вновь созданного текстового блока полностью идентичны описанным выше.

## ПЕЧАТЬ И ЭКСПОРТ ОТЧЕТА

После того как редактирование статистического отчета завершено, он сохраняется в специальном формате файлов отчета SPSS — SPSS Output с расширением файла \*.SPO. Данный формат позволяет наиболее полно, без потерь информации, сохранить любые объекты статистического отчета SPSS — таблицы, диаграммы, текст, с учетом их форматирования.

Файлы SPSS Output (\*.SPO) могут быть просмотрены, отредактированы и выведены на печать на любом компьютере, если на нем установлена программа SPSS или программа SPSS Smart Viewer. Следует

заметить, что возможности программы SPSS Smart Viewer по редактированию файлов SPSS Output минимальны, но сама программа может помочь при печати отчетов на слабом компьютере, т. к. занимает на жестком диске всего около 20 Мб и может работать без установки всей программы SPSS.

Вывод на печать файла SPSS Output осуществляется с помощью стандартных для многих компьютерных программ опций: File (Файл) > Print... (Печать). Можно для вывода на печать также использовать стандартную иконку с изображением принтера на панели инструментов. Перед выводом файла на печать можно использовать Page Setup... (Параметры страницы) и Print Preview (Предварительный просмотр), которые тоже находятся в опции File (Файл).

Вызвав окно Page Setup... (Параметры страницы), важно убедиться, что вы правильно установили размер и ориентацию листа бумаги для печати. В этом же окне можно изменить размер полей страницы. Нажав на кнопку Options... (Опции), пользователь получает возможность задать и отформатировать верхний и/или нижний колонтитулы (Header/Footer), а также определить размер выводимых на печать диаграмм (Printed Chart Size). По умолчанию диаграммы выводятся на печать в их естественном размере (As is), но можно самостоятельно изменить их размер на высоту страницы (Full page height), а также на половину (Half page height) или четверть (Quarter page height) страницы. Здесь же (Options) имеются полезные функции по изменению расстояния между выводимыми на печать объектами SPSS Output (Space between items) и указанию номера, с которого следует начинать нумерацию страниц, выводимых на печать (Number pages starting with).

Опция Print Preview (Предварительный просмотр) позволяет использовать стандартную для всех текстовых редакторов возможность — увидеть, как будет выглядеть файл отчета при выводе на печать. В окне предварительного просмотра имеется кнопка Page Setup... (Параметры страницы), которая позволяет, не выходя из режима просмотра, изменять предпечатные настройки.

**Примечание.** Довольно неудобными для вывода на печать являются таблицы. Не забывайте, что кроме стандартной подготовки к печати в опциях «Параметры страницы» и т. п. существует отдельная возможность предпечатной подготовки таблиц во время их редактирования (Format > Table Properties... > Printing). Также не забывайте, что при печати положение объектов SPSS Output на

**странице будет зависеть от выравнивания, заданного в опции Format основного меню редактора отчетов.**

Для завершения процедуры вывода на печать файла отчета необходимо использовать соответствующую опцию — File (Файл) > Print... (Печать) и в открывшемся окне указать требуемое количество копий. Помните, что на печать можно вывести не только весь файл отчета SPSS Output, но также отдельные объекты (или группу объектов). Для вывода на печать отдельных объектов необходимо сначала выделить их в подокне быстрой навигации редактора отчетов, а потом сделать описанную выше процедуру вывода на печать.

В некоторых случаях у пользователя возникает необходимость экспорта файла SPSS Output в другие файловые форматы. Часто эта необходимость возникает из-за того, что пользователю нужно продолжить работу над отчетом, но у него на компьютере не установлена программа SPSS. К сожалению, в типовой установке программы SPSS существует возможность экспорта файла отчета только в два файловых формата — текстовый (\*.TXT) и текстовый для Интернет (\*.HTML).

Экспорт в текстовый формат крайне неудобен, т. к. в этом случае невозможно экспортировать графические объекты, перенос сложных таблиц может осуществляться некорректно, кроме того, теряется форматирование объектов, произведенное пользователем в файле SPSS Output. К тому же полученный текстовый файл неудобен для вывода на печать (например, отображаемые таблицы не ограничены по ширине).

Экспорт в формат \*.HTML более удобен, т. к. позволяет экспортировать диаграммы и почти полностью сохраняет структуру и оформление сложных таблиц. Файл в формате \*.HTML довольно легко можно вывести на печать с помощью любого Интернет-браузера либо импортировать его в распространенные и удобные для пользователя редакторы MS Word или MS Excel. Однако, как свидетельствует опыт автора, такая многократная передача файла отчета из программы в программу, из формата в формат возможна только для небольших по размеру файлов (в противном случае компьютер зависает или производит операцию экспорта некорректно). Подобный двухэтапный экспорт также в значительной мере искажает исходное форматирование файла отчета, и фактически его приходится редактировать заново.

После вызова окна экспорта (File > Export) (рис. 56) необходимо определить тип экспорта (Export). Существуют три возможности: 1) экспортировать файл отчета целиком (Output Document), 2) без графических объектов (Output Document - No Charts) и 3) экспортировать только графические объекты (Charts Only). Нажав кнопку Options... (Опции) можно определить формат, в который будут экспортированы графические объекты (по умолчанию в формат \*.JPG), а также включение в экспорт дополнительной информации — примечаний, слоев и т. п.

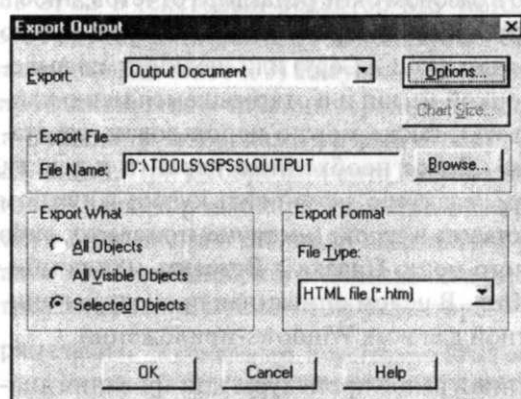


Рис. 56. Окно экспорта отчета SPSS Output в другие файловые форматы

Далее необходимо указать место положения (с помощью кнопки Browse...) и имя файла (File Name), в который будет произведен экспорт отчета. Потом нужно поставить флажок, указывающий, что именно нужно экспортировать (Export What) из файла отчета: все объекты (All Objects), все видимые объекты, кроме скрытых (All Visible Objects), только выделенные объекты (Selected Objects).

Для завершения процедуры экспорта необходимо определить формат файла экспорта (Export Format), выбрав тип файла (File Type), и нажать кнопку ОК.

В психологических исследованиях создание статистического отчета чаще всего не самоцель. Результаты статистических изысканий (таблицы и диаграммы) обычно являются крайне незначительной добавкой к основному тексту научной статьи, к тексту экспериментальной части курсовой или дипломной работы, диссертационного исследования. Обычно научные тексты пишутся в простом текстовом редакторе, например MS Word. Именно поэтому для психолога-исследователя



более востребованным является не умение создавать «красивый» статистический отчет в редакторе SPSS, а умение переносить необходимую информацию (объекты файла SPSS Output — таблицы или диаграммы) из редактора отчетов SPSS в текстовый редактор.

Перенос объектов редактора данных SPSS Output в любой текстовый редактор, работающий в оболочке Windows (здесь рассматривается на примере MS Word), является простой процедурой. Для этого необходимо выделить объект (таблицу или диаграмму), щелкнув на нем левой кнопкой мыши, либо в базовом окне редактора отчетов, либо в подокне быстрой навигации. Далее в строке главного меню нужно выбрать — **Правка > Копировать (Edit > Copy)** или щелкнуть на выделенном объекте правой кнопкой мыши и в открывшемся меню найти команду **Copy** (Копировать), также можно использовать сочетание клавиш <Ctrl> + <Ins>. Далее необходимо перейти к тексту, который открыт в текстовом редакторе, установить курсор в нужном месте и нажать на иконку **Вставить** в строке инструментов сверху, либо использовать опцию главного меню **Правка > Вставить** или комбинацию клавиш <Shift> + <Ins>. В целом описанная процедура копирования является стандартной для всех Windows-приложений.

При использовании описанной выше процедуры копирования диаграммы переносятся в текстовый редактор как готовые изображения, которые уже невозможно отредактировать (то есть вся «чистовая» правка диаграмм должна быть произведена в редакторе отчетов SPSS), а статистические таблицы переносятся в виде обычных таблиц, которые можно продолжать редактировать уже в текстовом редакторе.

При копировании таблиц из редактора отчетов SPSS в текстовый редактор особенности их оформления и заданные в редакторе отчетов SPSS размеры утрачиваются. Таблицы переносятся в текстовый редактор в простом виде (сохраняется лишь структура) и нуждаются в форматировании «с нуля» (определение размеров столбцов и строк, шрифта, линий, цвета, заливки и др.). На основании своего опыта работы с программой автор рекомендует переносить в текстовый редактор описанным способом лишь небольшие таблицы не с очень сложной структурой, которые легко умещаются на странице. Если таблицы слишком громоздки (например, результаты корреляционного анализа большого числа переменных) или имеют сложную струк-

туру, то гораздо удобнее описанным выше способом скопировать их в специализированный табличный редактор (обычно в MS Excel) и в нем уже править.

В некоторых случаях при передаче данных в MS Word удобнее воспользоваться специальными возможностями копирования (**Правка > Специальная вставка**). При копировании таблицы, активировав строку «Специальная вставка», вы вызываете окно, в котором содержатся три варианта копирования таблицы из редактора отчетов SPSS в текстовый редактор MS Word: 1) **Текст в формате RTF** (этот способ копирования используется по умолчанию); 2) **Неформатированный текст** и 3) **Рисунок**. При выборе варианта «Неформатированный текст» таблица переносится как простой текст, где строки разделены символами разрыв строки, а столбцы — символами табуляции. При выборе варианта **Рисунок** таблица переносится в текстовый редактор как картинка, при этом сохраняются все особенности оформления таблицы, заданные в редакторе отчетов SPSS. Разумеется, таблица, перенесенная как рисунок, уже не может быть отредактирована в текстовом редакторе.

**Примечание.** К сожалению, в программе SPSS отсутствует возможность копирования нескольких объектов сразу. Описанным выше способом копировать объекты из редактора отчетов SPSS Output в текстовый редактор можно только по одному.

## СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. *Антон Г.* Анализ таблиц сопряженности. М., 1982.
2. *Гласс Дж., Стенли Дж.* Статистические методы в педагогике и психологии. М., 1976.
3. *Глотов Н. В.* и др. Биометрия. Л., 1982.
4. *Гмурман В. Е.* Теория вероятностей и математическая статистика. М., 1997.
5. *Гублер Е. В., Генкин А. А.* Применение непараметрических критериев статистики в медико-биологических исследованиях. Л., 1973.
6. *Захаров В. П.* Применение математических методов в социально-психологических исследованиях. Л., 1985.
7. *Ивантер Э. В., Коросова В.* Основы биометрии. Петрозаводск, 1992.
8. *Кендэлл М.* Ранговые корреляции. М., 1975.
9. *Куликов Л. В.* Психологическое исследование. СПб., 1995.
10. *Лакин Г. Ф.* Биометрия. М., 1990.
11. *Плохинский Н. А.* Биометрия. М., 1970.
12. *Плохинский Н. А.* Алгоритмы биометрии. М., 1980.
13. *Рунион Р.* Справочник по непараметрической статистике. М., 1982.
14. *Сидоренко Е. В.* Методы математической обработки в психологии. СПб., 2000 (СПб., 1996).
15. *Суходольский Г. В.* Основы математической статистики для психологов. СПб., 1998 (Л., 1972).
16. *Тарасов С. Г.* Основы применения математических методов в психологии. СПб., 1998.

17. *Тюрин Ю. Н.* Непараметрические методы статистики. М., 1978.
18. *Урбах В. Ю.* Биометрические методы. М., 1965.
19. *Урбах В. Ю.* Статистический анализ в биологических и медицинских исследованиях. М., 1975.
20. *Холлендер М., Вулф Д. А.* Непараметрические методы статистики. М., 1983.

## Приложение

### **ЛИСТИНГ ФАЙЛА ПОДДЕРЖКИ КИРИЛЛИЧЕСКОЙ КОДИРОВКИ ШРИФТА В SPSS (CYRILLIC.BAT)**

```
@echo off
dregedit -C -k "Software\SPSS\%R%\Spsswin" -n "TypeFace"
-v "Arial Cyr"
dregedit -C -k "Software\SPSS\%R%\Spsswin\SyntaxFont"
-n "Name" -v "Courier New Cyr"
dregedit -C -k "Software\SPSS\%R%\Syntax\Font" -n "Name"
-v "Courier Cyr"
dregedit -C -k "Software\SPSS\%R%\Syntax\Font" -n
"CharSet" -v "204"
dregedit -C -k "Software\SPSS\%R%\Spsswin" -n
"DialogFont" -v "9,MS Sans Serif Cyr"
dregedit -C -k "Software\SPSS\%R%\OutputViewer\Font"
-n "Name" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\%R%\OutputViewer\Font"
-n "CharSet" -v "204"
dregedit -C -k "Software\SPSS\%R%\OutputViewer\Page
Title Font" -n "Name" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\%R%\OutputViewer\Page
Title Font" -n "CharSet" -v "204"
dregedit -C -k "Software\SPSS\%R%\OutputViewer\Outline
Font" -n "Name" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\%R%\OutputViewer\Outline
Font" -n "CharSet" -v "204"
dregedit -C -k "Software\SPSS\%R%\DraftViewer\Font"
-n "Name" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\%R%\DraftViewer\Font"
-n "CharSet" -v "204"
```

```
dregedit -C -k "Software\SPSS\$$\SPSS Text Block\Font"
-n "Name" -v "Courier New Cyr"
dregedit -C -k "Software\SPSS\$$\SPSS Text Block" -n
"CharSet" -v "204"
dregedit -C -k "Software\SPSS\$$\DataEditor\Font" -n
"Name" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\$$\DataEditor\Font" -n
"CharSet" -v "204"
dregedit -C -k "Software\SPSS\$$\SPSS Pivot Table"
-n "Name" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\$$\SPSS Pivot Table"
-n "CharSet" -v "204"
dregedit -C -k "Software\SPSS\$$\ScriptEditor" -n
"Name" -v "Courier New Cyr"
dregedit -C -k "Software\SPSS\$$\ScriptEditor" -n
"CharSet" -v "204"
dregedit -C -k "Software\SPSS\$$\Graphics\Options"
-n "FontName" -v "Arial Cyr"
dregedit -C -k "Software\SPSS\$$\Graphics\Options"
-n "CharSet" -v "204"
```

Сергей Иванович Калинин  
**Компьютерная обработка данных  
для психологов**

Главный редактор *И. Ю. Авидон*  
Заведующая редакцией *Т. В. Тулупьева*  
Художественный редактор *П. В. Борозенец*  
Научный редактор *А. Л. Тулупьев*  
Литературный редактор *Е. В. Перелетова*  
Технический редактор *О. В. Колесниченко*  
Генеральный директор *Л. В. Янковский*

Лицензия ЛП № 000364 от 29.12.99 г. Подписано в печать 25.01.2002 г.

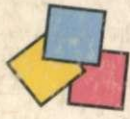
Формат 60x88'/j6. Бумага офсетная. Печать офсетная.

Тираж 2000 экз. Заказ № 28.

ООО Издательство «Речь».

199004, Санкт-Петербург, ВО, 3-я линия, 6 (лит. "А"),  
тел.: (812) 323-76-70, 323-90-63; e-mail: [rech@mail.lanck.net](mailto:rech@mail.lanck.net); [www.rech.spb.ru](http://www.rech.spb.ru).

Отпечатано с готовых диапозитивов в ООО «ИПК Бионт».  
199026, Санкт-Петербург, Средний пр. В. О., д. 86, тел. (812) 322-68-43



## ПРАКТИКУМ ПО ПСИХОДИАГНОСТИКЕ



- Общее описание статистического пакета для социальных наук (SPSS)
- Обзор возможностей программы SPSS
- Первичная обработка данных
- Параметрические и непараметрические критерии достоверности различий
- Корреляционный анализ

ISBN 5-9268-0091-9



9 785926 800910