

*IBM SPSS Statistics V27 - Краткое
руководство*



Примечание

Прежде чем использовать эту информацию и продукт, описанный в ней, прочтите сведения в разделе [“Замечания” на стр. 89](#).

Информация о продукте

Это издание применимо к версии 27, выпуск 0, модификация 0 IBM® SPSS Statistics и ко всем последующим версиям и модификациям до тех пор, пока в новых изданиях не будет указано иное.

© Copyright International Business Machines Corporation .

Содержание

Глава 1. Введение.....	1
Файлы для примеров.....	1
Открытие файла данных.....	1
Запуск аналитических процедур	2
Создание диаграмм.....	4
Глава 2. Открытие данных.....	7
Основная структура файлов данных IBM SPSS Statistics.....	7
Чтение файлов Баз Данных IBM SPSS Statistics.....	7
Чтение данных Excel.....	8
Открытие данных из баз данных.....	11
Открытие данных, хранящихся в текстовых файлах.....	14
Глава 3. Работа с Редактором данных.....	19
Ввод числовых данных.....	19
Ввод текстовых данных.....	20
Задание свойств данных.....	21
Задание меток переменных.....	21
Изменение типа и формата переменной.....	22
Добавление меток значений.....	22
Работа с пропущенными данными.....	23
Пропущенные значения для числовых переменных.....	23
Пропущенные значения для текстовых переменных.....	24
Глава 4. Итожащие статистики для отдельных переменных.....	25
Шкала измерения.....	25
Сводные показатели для категориальных данных.....	25
Диаграммы для категориальных данных.....	26
Сводные показатели для количественных переменных.....	27
Гистограммы для количественных переменных.....	28
Глава 5. Создание и редактирование диаграмм.....	31
Основы построения диаграмм.....	31
Использование галереи Построителя диаграмм.....	31
Задание переменных и статистик (итожащих показателей).....	32
Добавление текста.....	34
Построение диаграмм.....	34
Глава 6. Работа с выводом результатов.....	37
Работа со средством просмотра.....	37
Использование Редактора мобильных таблиц.....	38
Получение определений терминов вывода.....	38
Преобразование таблиц.....	39
Создание и показ слоев.....	40
Редактирование таблиц.....	40
Скрытие строк и столбцов.....	41
Изменение формата данных в таблицах.....	42
Шаблоны таблиц.....	43
Использование форматов, заданных заранее.....	43
Изменение шаблонов таблиц.....	44

Изменение заданного по умолчанию формата таблиц.....	46
Изменение параметров вывода.....	46
Вывод меток переменных и значений.....	47
Использование результатов в других приложениях.....	48
Вставка результатов в виде таблиц Word.....	49
Вставка результатов в виде текста.....	49
Изучение результатов в файлах Microsoft Word, PowerPoint и Excel.....	50
Экспорт результатов в PDF.....	55
Экспорт результатов в HTML.....	57
Глава 7. Работа с синтаксисом.....	59
Вставка синтаксиса.....	59
Редактирование синтаксиса.....	60
Открытие и запуск файла синтаксиса.....	61
Использование контрольных точек.....	61
Глава 8. Изменение значений данных.....	63
Создание категориальной переменной из количественной.....	63
Вычисление новых переменных.....	65
Использование функций.....	66
Использование условных выражений (условий).....	67
Работа с датами и временем.....	68
Вычисление продолжительности времени между двумя датами.....	69
Добавление продолжительности времени к дате.....	69
Глава 9. Сортировка и выбор данных.....	71
Сортировка данных.....	71
Анализ с расщеплением файла.....	71
Сортировка наблюдений при анализе с расщеплением файла.....	73
Включение и отключение анализа с расщеплением файла.....	73
Отбор подмножеств наблюдений.....	73
Отбор наблюдений на основе условия.....	74
Отбор случайной подвыборки.....	75
Выбор временного диапазона или диапазона наблюдений.....	76
Действия с неотобранными наблюдениями.....	76
Состояние выбора наблюдения.....	77
Глава 10. Файлы для примеров.....	79
Замечания.....	89
Товарные знаки.....	90
Индекс.....	93

Глава 1. Введение

В этом руководстве показано, как использовать многие базовые возможности. В нем даются практические пошаговые инструкции. Все файлы, показанные в примерах, установлены вместе с программой, так что вы можете выполнять описанный в руководстве анализ, получая те же самые результаты.

Если вам нужны подробные примеры различных методов статистического анализа, попробуйте пошаговый Разбор примеров, который можно вызвать из меню Справка.

Файлы для примеров

В большинстве примеров, представленных здесь, используется файл данных *demo.sav*. Файл данных представляет собой данные вымышленного опроса нескольких тысяч человек, содержащие демографическую информацию и информацию о потреблении.

Если используется студенческая версия, то версия *demo.sav* представляет собой репрезентативную выборку из исходного файла данных, размер которой равен 1500 наблюдений. Результаты, полученные при использовании этого файла данных, будут отличаться от результатов, показанных в Учебнике.

Файлы примеров, установленные вместе с продуктом, находятся во вложенной папке *Образцы* папки, в которой установлена система. В подкаталоге *Samples* есть отдельная папка для каждого из следующих языков: английский, французский, немецкий, итальянский, японский, корейский, польский, русский, упрощенный китайский, испанский и традиционный китайский.

Не все файлы примеров доступны на каждом языке. Если файл примера не доступен на конкретном языке, языковая папка содержит версию этого файла примера на английском языке.

Открытие файла данных

Чтобы открыть файл данных:

1. Выберите в меню:

Файл > Открыть > Данные...

Диалоговое окно открытия файлов.

По умолчанию в окне находятся файлы данных в формате программы IBM SPSS Statistics (с расширением *.sav*).

В этом примере используется файл *demo.sav*.

	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Рисунок 1. Файл demo.sav в Редакторе данных

Теперь в Редакторе данных появились данные из файла данных. В представлении данных при наведении указателя мыши на имя переменной (заголовки столбцов) выводится описательная метка переменной (если для этой переменной определена метка).

По умолчанию, в Редакторе данных показаны значения данных. Чтобы были показаны метки:

2. Из меню выберите:

Вид > Метки значений



Рисунок 2. Кнопка Метки значений

Другой способ - использовать кнопку Метки значений на панели инструментов.

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

Рисунок 3. Метки значений в Редакторе данных

Теперь в Редакторе данных отображаются описательные метки значений, облегчающие интерпретацию ответов.

Запуск аналитических процедур

Если имеются дополнительные модули, то в меню Анализ содержится список процедур создания отчетов и статистических процедур

Начнем с построения простых частотных таблиц. Для этого примера требуется модуль Статистическая база.

1. Из меню выберите:

Анализ > Описательные статистики > Частоты...

На экране появилось диалоговое окно Частоты.

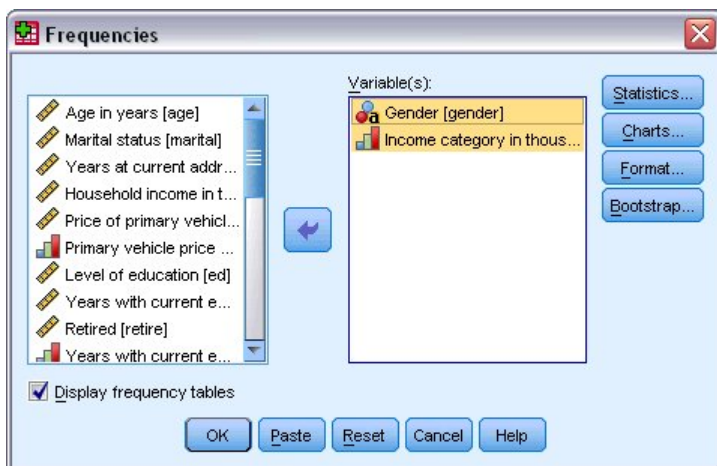


Рисунок 4. Диалоговое окно Частоты

Значки рядом с переменными говорят о типах данных и шкалах измерений.

	Числовой	Строка символов	Дата	Время
Количественная (непрерывная)		(не задается)		
Порядковая				
Номинальная				

Если метка переменной или ее имя отображается в списке в усеченном виде, разместите курсор над переменной для просмотра полного текста метки или имени. Имя переменной *incscat* показано в квадратных скобках после метки с описанием переменной. Категория дохода домохозяйства в тысячах является меткой переменной. Если у переменной нет метки, в списке переменных появляется только ее имя.

Чтобы изменить размеры диалогового окна, щелкните и перетащите мышью границы или углы окна. Например, если растянуть окно по горизонтали, список переменных также станет шире.

В диалоговом окне пользователь выбирает переменные, необходимые для анализа, в исходном списке и перетаскивает их в список переменных в правой части. Кнопка **ОК**, предназначенная для запуска анализа, недоступна, пока в список Переменные не будет перенесена хотя бы одна переменная.

Во многих диалоговых окнах можно получить дополнительную информацию, щелкнув правой кнопкой мыши по имени переменной в списке и выбрав во всплывающем меню **Информация о переменной**.

2. Щелкните *Пол [gender]* в списке исходных переменных и перетащите переменную в список целевых переменных.

- Щелкните переменную *Категория дохода домохозяйства [inccat]* в списке исходных и перетащите ее в список целевых.

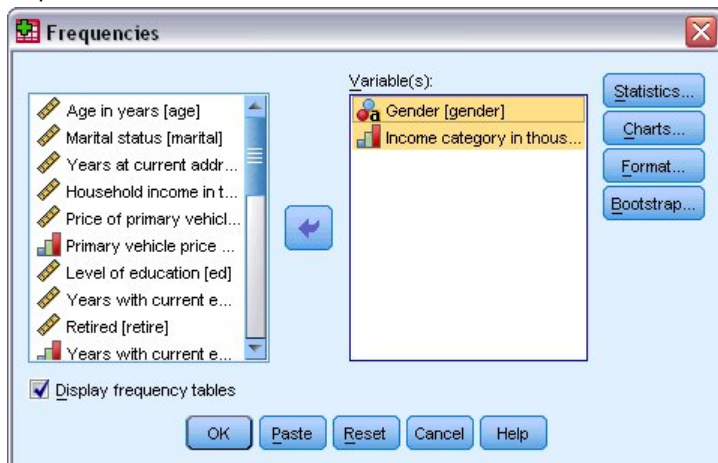


Рисунок 5. Переменные, выбранные для анализа

- Нажмите кнопку **ОК**, чтобы запустить эту процедуру.

Результаты попадают в окно вывода результатов средства просмотра.

	Frequency	Percent	Valid Percent	Cumul: Perce
Valid Female	3179	49.7	49.7	
Male	3221	50.3	50.3	
Total	6400	100.0	100.0	

	Frequency	Percent	Valid Percent	Cum Pe
Valid Under \$25	1174	18.3	18.3	
\$25 - \$49	2388	37.3	37.3	
\$50 - \$74	1120	17.5	17.5	
\$75+	1718	26.8	26.8	
Total	6400	100.0	100.0	

Рисунок 6. Таблица частот для категории дохода

Создание диаграмм

Хотя в некоторых статистических процедурах и создаются диаграммы, для создания диаграмм можно воспользоваться пунктом меню **Графика**.

Например, можно создать диаграмму, в которой будет показана взаимосвязь между использованием мобильной связи и наличием персонального цифрового помощника (PDA).

- Из меню выберите:

Графики > Построитель диаграмм...

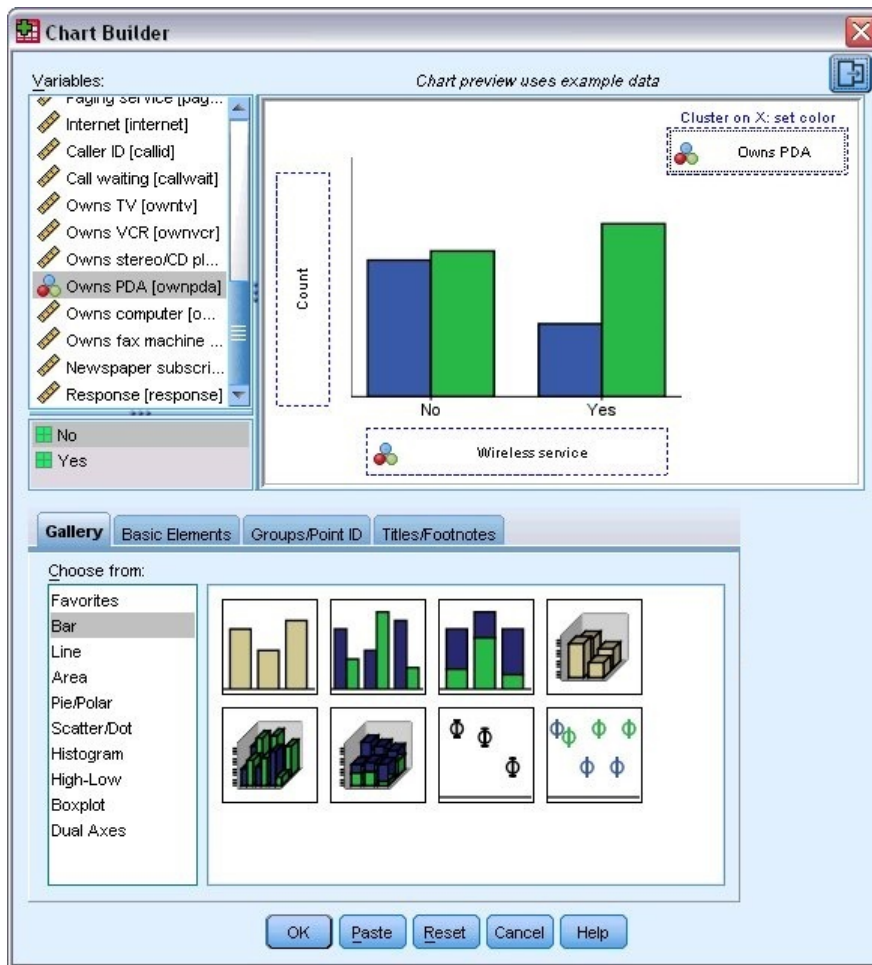


Рисунок 7. Диалоговое окно Построитель диаграмм с заполненными зонами сброса

2. Щелкните по вкладке **Галерея** (если она не выбрана).
3. Выберите **Столбики**, (если этот пункт не выбран).
4. Перенесите значок кластеризованные столбики в панель построения диаграмм (большое пространство над Галереей).
5. Прокрутите вниз список Переменные, щелкните правой кнопкой мыши по переменной *Мобильная связь [wireless]* и выберите **Номинальную** шкалу измерения.
6. Перенесите переменную *Мобильная связь [wireless]* в поле оси X.
7. Щелкните правой кнопкой мыши по *Наличие персонального цифрового помощника (PDA) [ownpda]* и выберите **Номинальную** шкалу измерения.
8. Перенесите переменную *Наличие персонального цифрового помощника (PDA) [ownpda]* в поле Кластеры правом верхнем углу панели построения диаграмм.
9. Нажмите кнопку **ОК**, чтобы создать диаграмму.

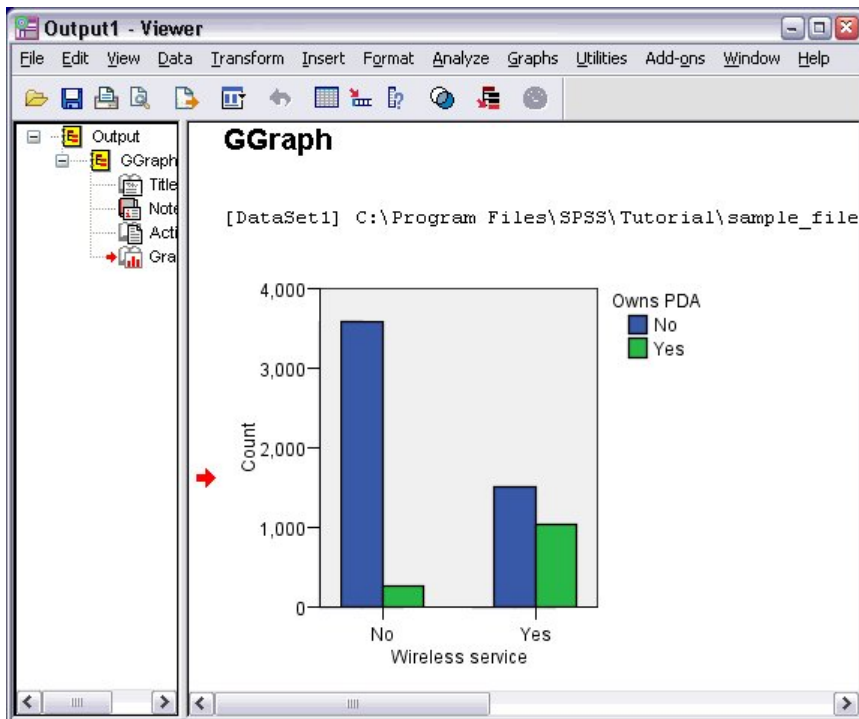


Рисунок 8. Столбчатая диаграмма в окне вывода результатов Средства просмотра.

Столбчатая диаграмма появилась в окне вывода средства просмотра. Диаграмма показывает, что среди людей, пользующихся мобильной связью, гораздо больше владельцев PDA, чем среди людей, не пользующихся мобильной связью.

Диаграммы и таблицы можно редактировать, дважды щелкая по ним на панели содержимого окна просмотра. Можно также копировать и вставлять результаты в другие приложения. Эти возможности будут освещены позже.

The screenshot shows the SPSS Data Editor window titled "demo.sav - Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The main area displays a data table with 11 rows and 6 columns: age, marital, address, income, inccat, and car. The 'age' column is selected, and the value '40' is entered in the top cell. The table data is as follows:

	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Рисунок 10. Открытый файл данных

Чтение данных Excel

Вместо того, чтобы вводить данные непосредственно в Редакторе данных, можно открыть данные, хранящиеся в форматах приложений, таких как Microsoft Excel. Заголовки столбцов можно прочитать как имена переменных.

1. Из меню выберите:

Файл > Импорт данных > Excel

2. Перейдите в папку Samples\English и выберите demo.xlsx.

В диалоговом окне **Чтение файла Excel** выводится предварительный просмотр файла данных. Показано содержимое первого листа в файле. Если файл содержит несколько листов, можно выбрать лист из списка.

Можно увидеть, что некоторые строковые значения для поля *Gender* (пол) содержат начальные пробелы. Некоторые значения для *MaritalStatus* выводятся как точки (.).

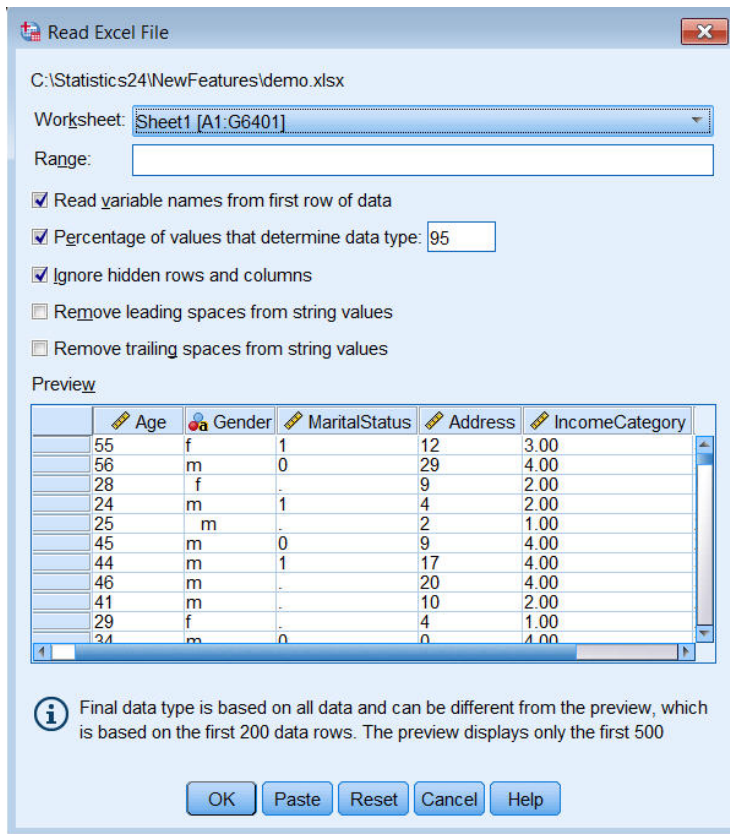
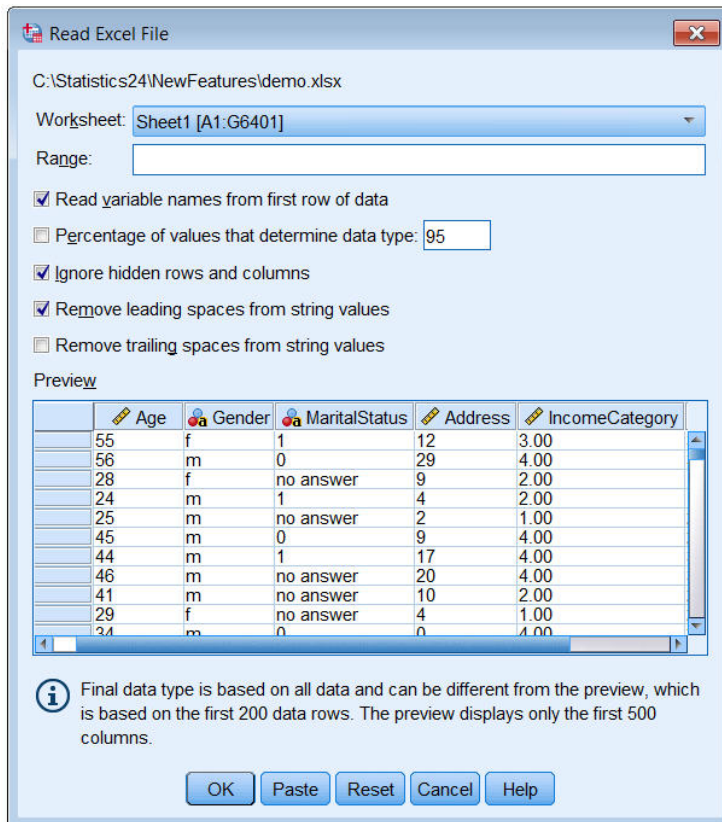


Рисунок 11. Диалоговое окно Чтение файла Excel

3. Убедитесь, что выбран **Читать имена переменных из первой строки данных**. Если заголовки столбцов не соответствуют требованиям к именам переменных, они будут преобразованы в допустимые имена переменных. Исходные заголовки столбцов сохраняются как метки переменных.
4. Выберите **Удалить начальные пробелы из строковых значений**.
5. Отмените выбор **Процент значений, определяющих тип данных**.



Теперь в системных пропущенных ячейках показано строковое значение "нет ответа". Если процент параметра значений не задан, а столбец содержит набор данных, то переменная читается как строковый тип данных. Все значения сохраняются, но числовые значения рассматриваются как строковые значения.

6. Выберите **Процент значений, определяющих тип данных**, чтобы рассматривать *MaritalStatus* как числовую переменную.
7. Нажмите кнопку **OK**, чтобы открыть файл Excel.

Данные появились в Редакторе данных и заголовки столбцов используются в качестве имен переменных. Поскольку имена переменных не могут содержать пробелов, пробелы были удалены из исходных заголовков столбцов. Например, заголовок столбца "Семейное положение" преобразуется в переменную *MaritalStatus*. Исходный заголовок столбца превратится в метку переменной.

The screenshot shows the IBM SPSS Statistics Data Editor window with a data table. The table has 10 rows and 6 columns. The columns are labeled: Age, Gender, MaritalStatus, Address, IncomeCategory, and JobCategory. The data is as follows:

	Age	Gender	MaritalStatus	Address	IncomeCategory	JobCategory
1	55	f	1	12	3.00	3
2	56	m	0	29	4.00	3
3	28	f	.	9	2.00	1
4	24	m	1	4	2.00	1
5	25	m	.	2	1.00	2
6	45	m	0	9	4.00	2
7	44	m	1	17	4.00	3
8	46	m	.	20	4.00	3
9	41	m	.	10	2.00	2
10	29	f	.	4	1.00	2

Рисунок 12. Данные, импортированные из Excel

Информация, связанная с данной

“Файлы для примеров” на стр. 79

Открытие данных из баз данных

Конструктор чтения баз данных позволяет без лишних усилий импортировать данные из баз данных. Любая база данных, использующая драйверы ODBC (Open Database Connectivity) может читаться напрямую после установки драйверов. Драйверы ODBC для множества форматов баз данных поставляются на компакт-диске. Дополнительные драйверы можно получить у поставщиков систем управления базами данных. В этом примере разбирается одна из наиболее распространенных систем управления базами данных - Microsoft Access.

Примечание: Этот пример относится только к Microsoft Windows и для него нужен драйвер ODBC для Access. Драйвер Microsoft Access ODBC работает только с 32-битной версией IBM SPSS Statistics. Описанные шаги подобны шагам для других платформ, но для них может понадобиться драйвер ODBC третьей стороны для Access.

1. Из меню выберите:

Файл > Импорт данных > База данных > Новый запрос...

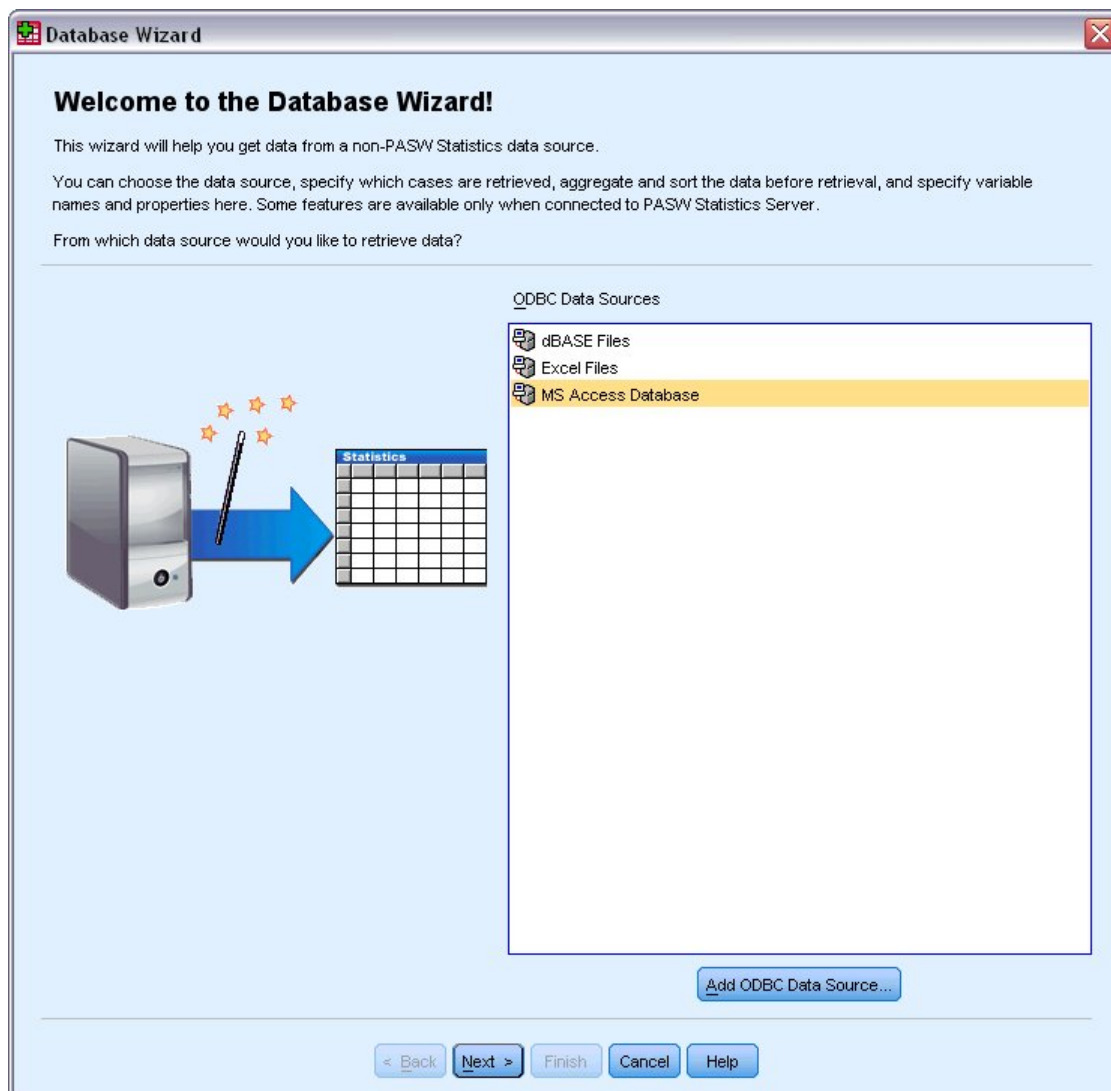


Рисунок 13. Диалоговое окно Добро пожаловать Конструктора чтения баз данных

2. Выберите в списке источников данных **База данных MS Access**, а затем нажмите кнопку **Далее**.

Примечание: В зависимости от установки можно также увидеть список ресурсов данных OLEDB в левой части конструктора (только в ОС Windows), но этот пример использует список ресурсов данных ODBC в правой части.

3. Нажмите кнопку **Обзор**, чтобы найти файл Access, который хотите открыть.
4. Откройте *demo.mdb*. Дополнительную информацию смотрите в разделе [Глава 10, “Файлы для примеров”](#), на стр. 79.
5. Нажмите кнопку **OK** в диалоговом окне Вход в систему.

На следующем шаге необходимо выбрать таблицы данных и переменные для импорта.

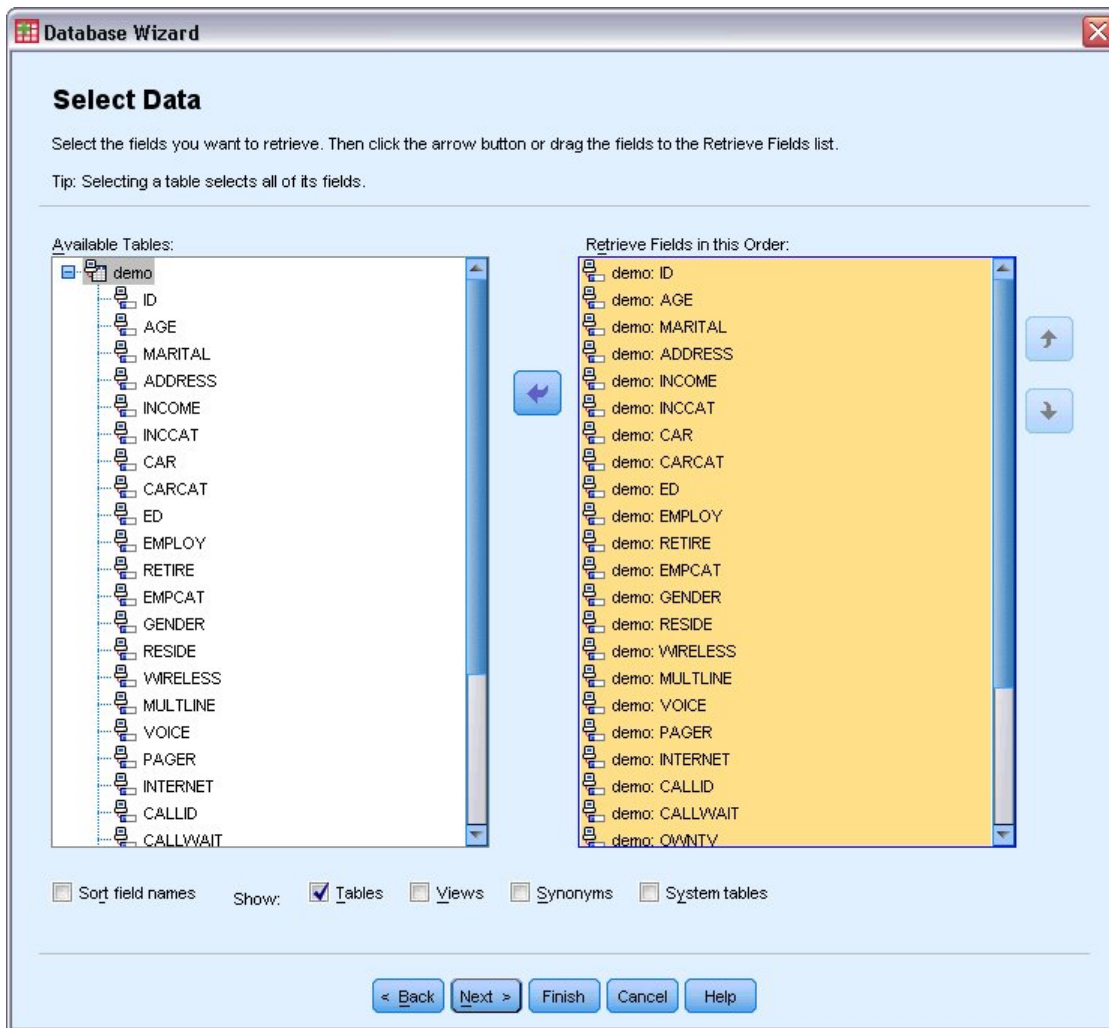


Рисунок 14. Шаг Выбор данных

6. Перенесите таблицу **demo** целиком в список Получить поля в следующем порядке.
7. Нажмите кнопку **Далее**.

На следующем шаге можно выбрать, какие записи (наблюдения) импортировать.

Если вы хотите импортировать не все наблюдения, а только часть (например, мужчин старше 30 лет). Также можно произвести случайный отбор наблюдений. Для сокращения времени обработки данных большого объема может понадобиться ограничить число наблюдений небольшой репрезентативной выборкой.

8. Нажмите кнопку **Далее**, чтобы продолжить работу.

На основе имен полей создаются имена переменных. При необходимости имена преобразуются в допустимые имена переменных. Исходные имена полей сохраняются в виде меток переменных. Имена переменных также можно изменить перед тем, как импортировать данные.

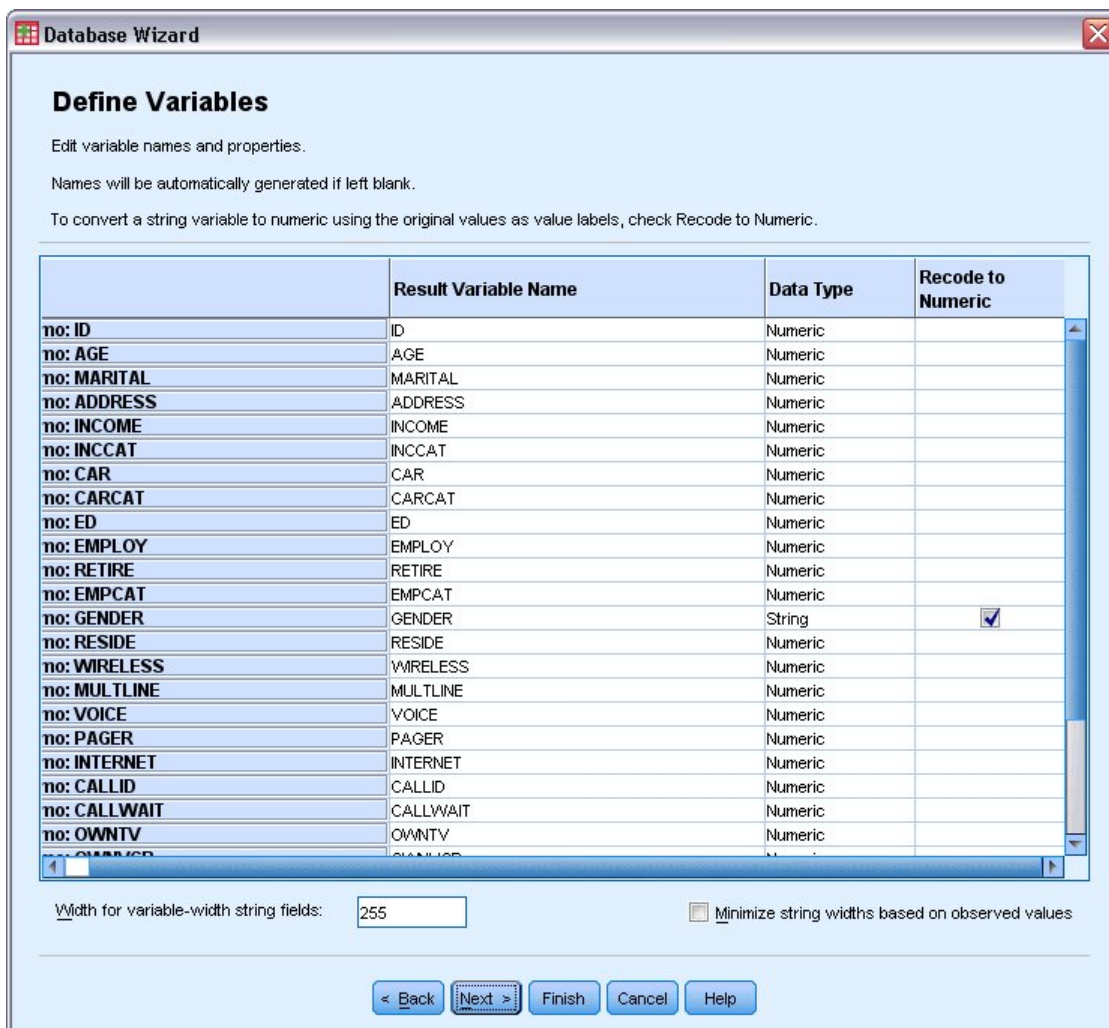


Рисунок 15. Шаг Задать переменные

9. Установите флажок в ячейке **Перекодировать поля Gender** поля Gender. Этот параметр преобразует текстовые переменные в числовые и сохраняет исходные значения в качестве меток значений новых переменных.
10. Нажмите кнопку **Далее**, чтобы продолжить работу.
 На шаге Результаты на основе параметров, выбранных нами на предыдущих шагах Конструктора чтения баз данных, генерируется команда SQL. Эта команда может быть выполнена сейчас или сохранена в файле для использования в дальнейшем.
11. Нажмите кнопку **Готово**, чтобы импортировать данные.
 Все выбранные для импорта данные базы данных Access теперь находятся в Редакторе данных.

Открытие данных, хранящихся в текстовых файлах

Текстовые файлы представляют собой другой часто используемый источник данных. Многие приложения для работы с электронными таблицами и системы управления базами данных могут сохранять данные в одном из нескольких текстовых форматов. Файлы с запятыми или знаками табуляции в качестве разделителей - это файлы, в которых запятые или знаки табуляции разделяют значения одной переменной от значений другой. В этом примере в качестве разделителя данных используются символы табуляции.

1. Из меню выберите:

Файл > Импорт данных > Текстовые данные

2. Перейдите в папку Samples\English и выберите demo.txt.

Мастер импорта текстовых файлов поможет вам указать способ интерпретации указанного текстового файла.

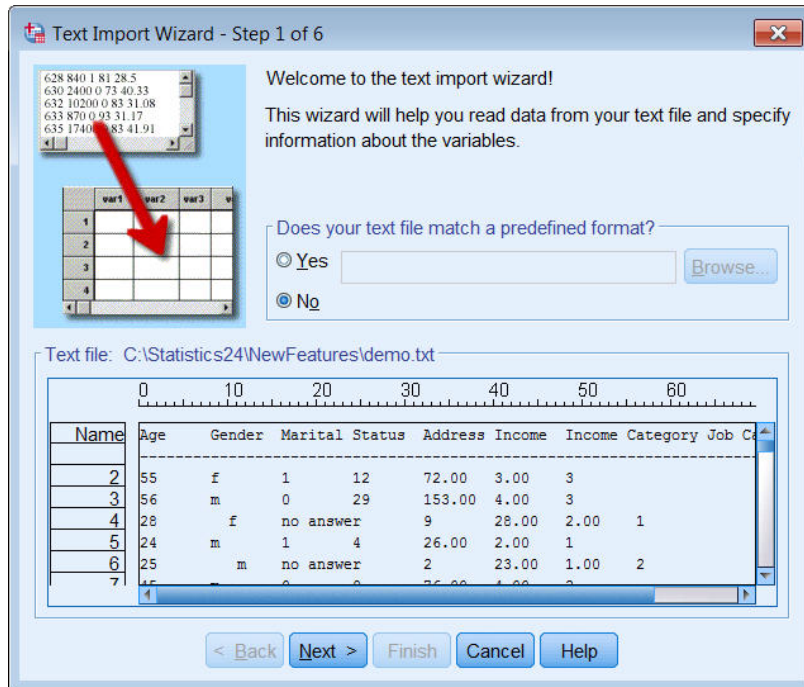


Рисунок 16. Мастер импорта текстовых файлов: шаг 1 из 6

3. На Шаге 1 можно выбрать, соответствуют ли данные созданному заранее специальному файлу формата, или параметры импорта будут задаваться с самого начала. Выберите **Нет**.
4. Нажмите кнопку **Далее**, чтобы продолжить работу.

Как уже говорилось, в нашем файле в качестве разделителей используются знаки табуляции. Кроме того, в первой строке файла указаны имена переменных.

5. На шаге 2 мастера выберите **С разделителями**, чтобы указать, что структура файла - форматирование с разделителями.
6. Выберите **Да**, чтобы указать, что этот файл содержит имена переменных в начале файла.
7. Нажмите кнопку **Далее**, чтобы продолжить работу.
8. На шаге 3 введите 2 как номер строки, где начинается первое наблюдение данных (поскольку в первой строке находятся имена переменных).
9. Оставьте значения по умолчанию для остальной части этого шага и нажмите кнопку **Далее**, чтобы продолжить работу.

Предварительный просмотр данных на шаге 4 - быстрый способ убедиться в правильности чтения файла

10. Выберите **Табуляция** и отмените выбор других вариантов для разделителей. **Пробел** выбран по умолчанию, так как этот файл содержит пробелы. Для этого файла пробелы - это часть значений данных, а не разделители. Надо отменить выбор опции **Пробел**, чтобы правильно прочитать файл.
11. Выберите **Удалить начальные пробелы из строковых значений**. Пробелы в начале строковых значений влияют на то, как строковые значения оцениваются в выражениях. В этом файле некоторые значения для поля *Пол* содержат начальные пробелы, не входящие в значение. Если вы не удалите эти пробелы, значение " f" будет рассматриваться как отличающееся от "f".

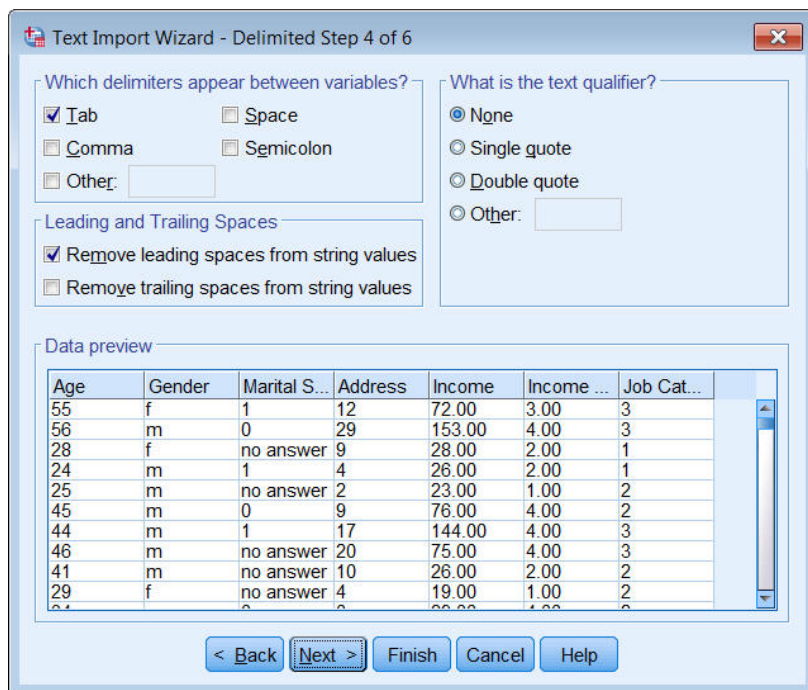


Рисунок 17. Мастер импорта текстовых файлов: шаг 4 из 6

12. Нажмите кнопку **Далее**, чтобы продолжить работу.

Поскольку имена переменных изменены в соответствии с правилами именования, шаг 5 дает возможность редактировать любые нежелательные имена.

Здесь же можно задать типы данных. Например, можно изменить формат переменной *Доход* на денежный формат в долларах.

Чтобы изменить тип данных:

13. На панели **Предварительный просмотр данных** выберите *Доход*.

14. Выберите в выпадающем списке формат переменной **Доллар**.

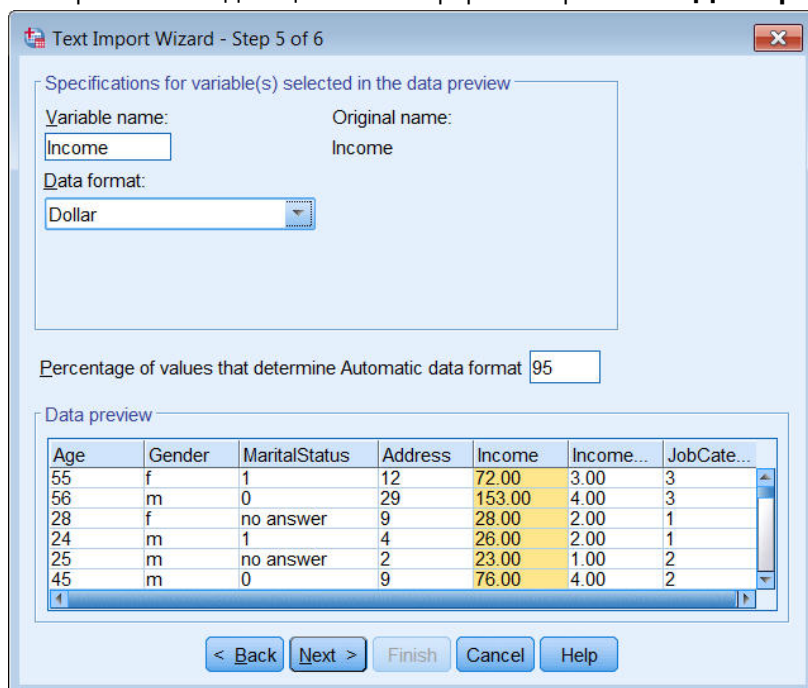


Рисунок 18. Изменение типа данных

Переменная *MaritalStatus* (семейное положение) содержит и строковые, и числовые значения. Строковых значений меньше пяти процентов. При значении по умолчанию 95% для параметра **Процент значений, определяющих автоматический формат данных** эта переменная обрабатывается как числовая, а строковые значения заменяются на системными значениями отсутствия. Если ни у какого формата данных доля не превышает заданного значения в процентах, переменная обрабатывается как строковая. Если вы изменить значение на 100, все значения будут сохранены, но все числовые значения будут обрабатываться как строки.

15. Нажмите кнопку **Далее**, чтобы продолжить работу.
16. Оставьте значения по умолчанию на последнем шаге и нажмите кнопку **Готово**, чтобы импортировать данные.

Глава 3. Работа с Редактором данных

В Редакторе данных отображается содержимое активного набора данных. Информация в Редакторе данных состоит из переменных и наблюдений.

- В закладке Данные столбцы представляют собой переменные, а строки - наблюдения.
- В закладке Переменные, строки представляют переменные, а столбцы - свойства переменных.

Переменные используются для представления анализируемых данных. Для примера можно взять опрос. Любой вопрос, на который можно дать один ответ, представляет собой одну переменную. Переменные бывают различных типов, включая числовые, текстовые, валюту и даты.

Ввод числовых данных

Данные можно вводить в Редакторе данных. Это может иметь смысл, когда объем данных невелик или когда необходимо внести небольшие изменения или добавления в данные большого объема.

1. Щелкните по вкладке **Переменные** в нижней части окна Редактора данных.

Необходимо задать переменные, которые будут использоваться. В данном случае требуются только три переменные: *age* (возраст), *marital status* (семейное положение) и *income* (доход).

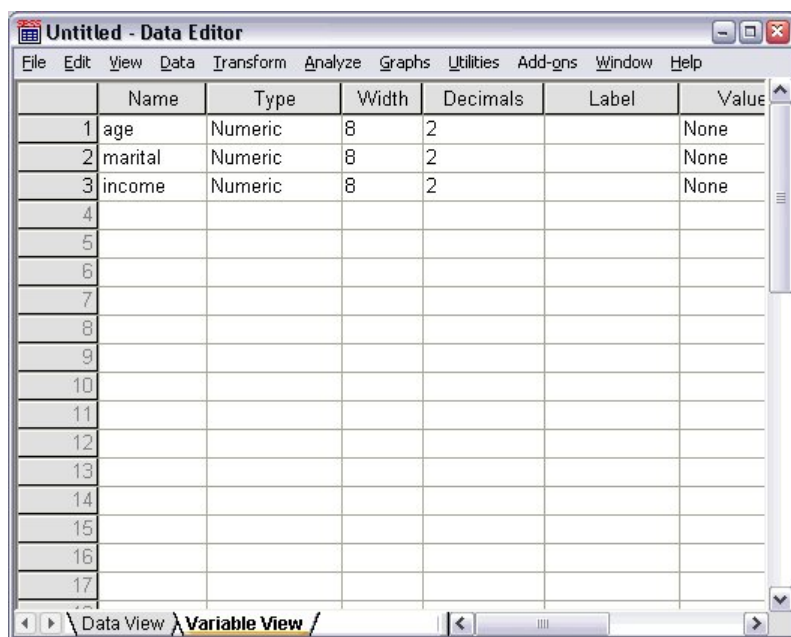


Рисунок 19. Имена переменных в представлении Переменные

2. В первой строке первого столбца введите *age*.
3. Во второй строке введите *marital*.
4. В третьей строке введите *income*.

Новые переменные автоматически становятся числовыми.

Если не вводить имена переменных, уникальные имена будут созданы автоматически. Однако эти имена не несут в себе содержательной информации и использовать их в больших файлах данных не рекомендуется.

5. Щелкните по вкладке **Данные**, чтобы продолжить вводить данные.

Имена, которые Вы ввели в закладке Переменные, появились в заголовках первых трех столбцов в закладке Данные.

Начнем вводить данные, начиная с первой строки первого столбца.

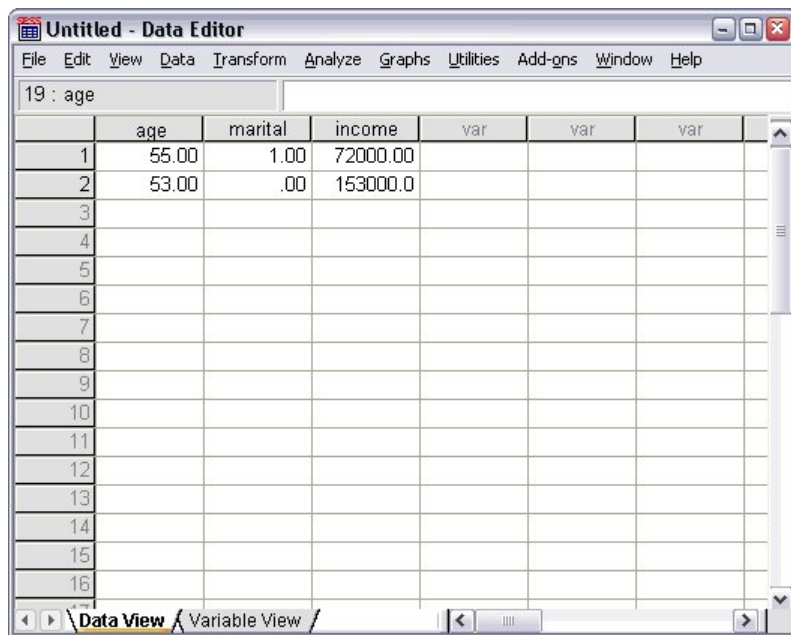


Рисунок 20. Значения, введенные в закладке Данные

6. В столбце *age* введите 55.
7. В столбце *marital* введите 1.
8. В столбце *income* введите 72000.
9. Поместите курсор в первый столбец второй строки, чтобы добавить данные следующего наблюдения.
10. В столбце *age* введите 53.
11. В столбце *marital* введите 0.
12. В столбце *income* введите 153000.

В столбцах *age* и *marital* сейчас видны десятичные знаки, несмотря на то, что их значения должны были быть целыми числами. Чтобы скрыть десятичные знаки в этих переменных:

13. Щелкните по вкладке **Переменные** в нижней части окна Редактора данных.
14. Выберите ячейку на пересечении столбца *Десятичные* и строки *age* и введите 0, чтобы скрыть десятичные знаки.
15. Выберите ячейку на пересечении столбца *Десятичные* и строки *marital* и введите 0, чтобы скрыть десятичные знаки.

Ввод текстовых данных

Нечисловые данные, такие как текстовые строки, также можно вводить в Редакторе данных.

1. Щелкните по вкладке **Переменные** в нижней части окна Редактора данных.
2. В первой ячейке первой пустой строки введите имя переменной - пол.
3. Щелкните по ячейке *Тип* рядом с введенным именем переменной.
4. Нажмите кнопку в правой части ячейки *Тип*, чтобы открыть диалоговое окно Тип переменной.
5. Выберите **Текстовая**, чтобы задать тип переменной.
6. Нажмите кнопку **ОК**, чтобы подтвердить сделанный выбор и вернуться в Редактор данных.

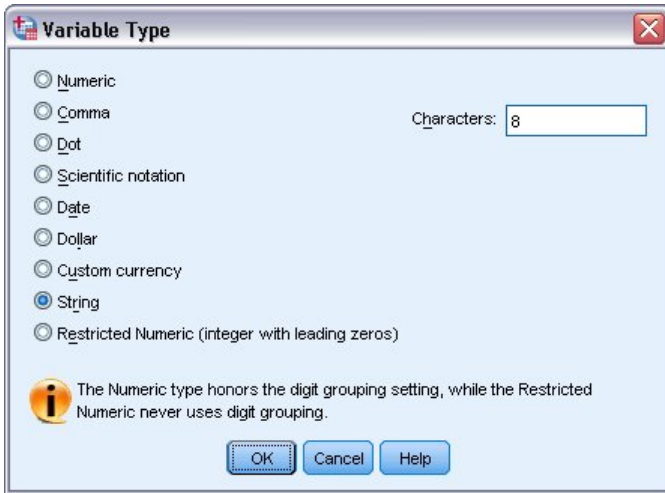


Рисунок 21. Диалоговое окно Тип переменной

Задание свойств данных

Помимо задания типов переменных, можно задавать описательные метки переменных и метки значений данных. Эти описательные метки используются в статистических таблицах и диаграммах.

Задание меток переменных

Метки предназначены для того, чтобы описывать переменные. Эти описания, обычно, представляют собой развернутые версии имен переменных. Длина меток может быть не более 255 символов. Метки используются в выводе для наглядного представления переменных.

1. Щелкните по вкладке **Переменные** в нижней части окна Редактора данных.
2. В столбце *Метка* строки *age* введите Возраст.
3. В столбце *Метка* строки *marital* введите Семейное положение респондента.
4. В столбце *Метка* строки *income* введите Доход домохозяйства.
5. В столбце *Метка* строки *sex* введите Пол респондента.

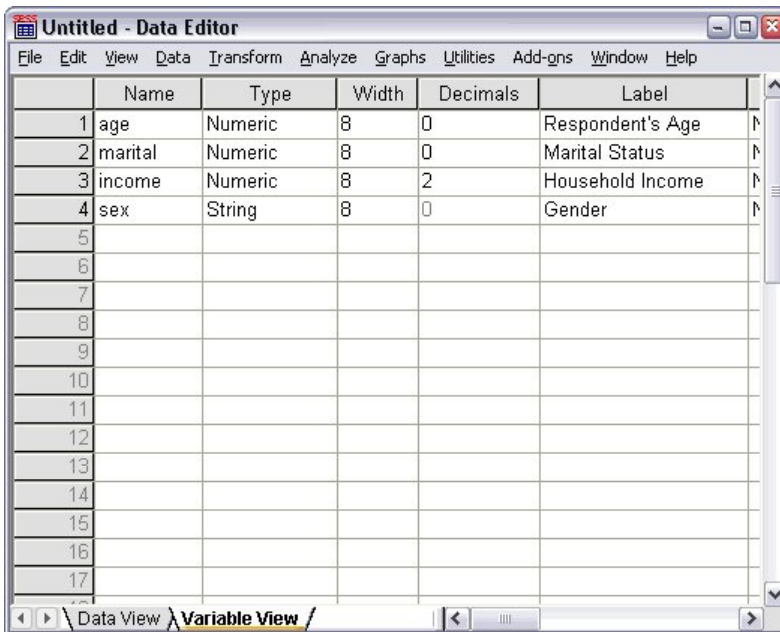


Рисунок 22. Метки переменных, введенные в представлении Переменные

Изменение типа и формата переменной

В столбце *Тип* показаны типы всех переменных. Наиболее распространены числовые и текстовые переменные, но есть еще несколько поддерживаемых типов переменных. В текущем файле данных переменная *доход* задана как числовая.

1. Щелкните по ячейке *Тип* в строке *доход*, а затем нажмите кнопку в правой части ячейки Тип, чтобы открыть диалоговое окно Тип переменной.
2. Выберите **Доллар**.

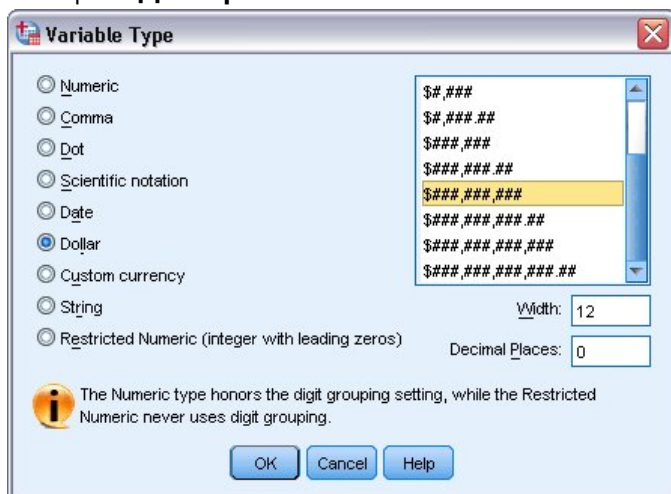


Рисунок 23. Диалоговое окно Тип переменной

В окне показаны параметры форматов, доступные для выбранного типа переменных.

3. Для формата валюты в нашем примере, выберите **\$###,###,###**.
4. Нажмите кнопку **ОК**, чтобы сохранить изменения.

Добавление меток значений

Метки значений обеспечивают сопоставление значений переменных с описывающими эти значения текстовыми метками. В нашем примере у переменной *marital* (семейное положение) есть два допустимых значения. Значение 0 обозначает, что респондент не состоит в браке, а значение 1 обозначает, что респондент состоит в браке.

1. Щелкните по ячейке *Значения* в строке *marital* (семейное положение), а затем нажмите кнопку в правой части ячейки Значения, чтобы открыть диалоговое окно Метки значений.

Значение - это числовое значение переменной.

Метка значения - это текстовая метка, соответствующая указанному числовому значению.

2. Введите 0 в поле Значение.
3. В поле Метка значения ведите Не состоит в браке.
4. Нажмите **Добавить**, чтобы добавить метку в список меток значений переменной.

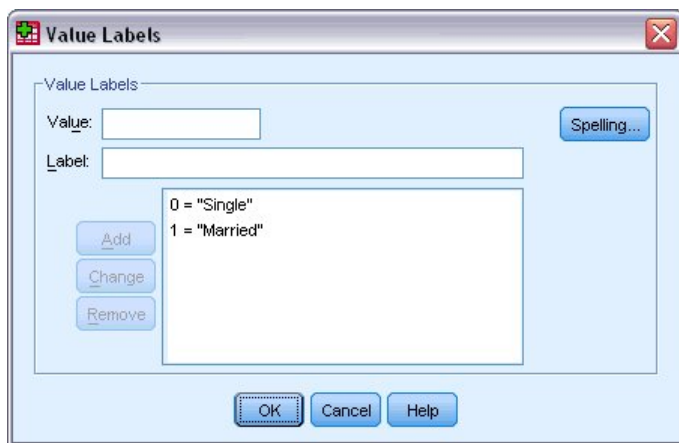


Рисунок 24. Диалоговое окно Метки значений

5. В поле Значение введите 1, в поле Метка значения введите Состоит в браке.
6. Нажмите кнопку **Добавить**, а затем кнопку **ОК**, чтобы сохранить изменения и вернуться в Редактор данных.

Эти метки также можно отображать в закладке Данные, что может упростить восприятие данных.

7. Щелкните по вкладке **Данные** в нижней части окна Редактора данных.
8. Выберите в меню:

Вид > Метки значений

Теперь при вводе данных в Редакторе данных можно не вводить числовые значения, а выбирать метки в выпадающих списках. Это позволяет обеспечить ввод допустимых вариантов ответов в более понятной форме.

Если пункт меню Метки значений уже выбран (рядом с ним стоит галочка), повторный выбор пункта **Метки значений** приведет к *выключению* отображения меток значений.

Работа с пропущенными данными

Обычно, пропущенных или недопустимых данных слишком много, чтобы их игнорировать. Респонденты, принимающие участие в опросах и обследованиях, могут отказываться отвечать на определенные вопросы, могут не знать, как ответить, а также могут давать ответы, не укладывающиеся в рамки предполагаемых вариантов ответов. Если не отфильтровать или не идентифицировать такие данные, результаты анализа данных могут оказаться некорректными.

Для числовых данных пустые поля данных или поля, содержащие недопустимые значения, преобразуются в системные пропущенные значения, показываемые в виде одной точки.

Причины, по которым значения пропущены, возможно, являются важными для анализа. Например, может оказаться полезным сделать различие между респондентами, отказавшимися на вопрос, и респондентами, не ответившими на вопрос, так как вопрос был не применим к ним.

Пропущенные значения для числовых переменных

1. Щелкните по вкладке **Переменные** в нижней части окна Редактора данных.
2. Щелкните по ячейке на пересечении столбца *Пропущенные* и строки *возраст*, а затем нажмите кнопку в правой части этой ячейки, чтобы открыть диалоговое окно Пропущенные значения.

В этом диалоговом окне можно задать три пропущенных значения или же диапазон значений плюс отдельное пропущенное значение.



Рисунок 25. Диалоговое окно Пропущенные значения

3. Выберите **Отдельные пропущенные значения**.
4. Введите 999 в первом поле и оставьте пустыми остальные два.
5. Нажмите кнопку **ОК**, чтобы сохранить изменения и вернуться в Редактор данных.
Теперь, когда информация о пропущенных данных добавлена, можно задать для этого значения метку.
6. Щелкните по ячейке *Значения* в строке *возраст*, а затем нажмите кнопку в правой части ячейки *Значения*, чтобы открыть диалоговое окно *Метки значений*.
7. В поле *Значение* введите 999 в поле *Значение*.
8. В поле *Метка значения* введите *Нет* ответа.
9. Нажмите **Добавить**, чтобы добавить метку в список меток значений переменной.
10. Нажмите кнопку **ОК**, чтобы сохранить изменения и вернуться в Редактор данных.

Пропущенные значения для текстовых переменных

Работа с пропущенными значениями для текстовых переменных похожа на работу с пропущенными значениями для числовых переменных. Однако, в отличие от числовых переменных, пустые текстовые значения не являются системными пропущенными. Такие значения интерпретируются как пустые строки.

1. Щелкните по вкладке **Переменные** в нижней части окна Редактора данных.
2. Щелкните по ячейке на пересечении столбца *Пропущенные* и строки *пол*, а затем нажмите кнопку в правой части этой ячейки, чтобы открыть диалоговое окно *Пропущенные значения*.
3. Выберите **Отдельные пропущенные значения**.
4. Введите *Н0* в первом текстовом поле.
Пропущенные значения для текстовых переменных чувствительны к регистру. Поэтому значение *но* не будет считаться пропущенным.
5. Нажмите кнопку **ОК**, чтобы сохранить изменения и вернуться в Редактор данных.
Теперь можно добавить метки для пропущенных значений.
6. Щелкните по ячейке *Значения* в строке *пол*, а затем нажмите кнопку в правой части ячейки *Значения*, чтобы открыть диалоговое окно *Метки значений*.
7. Введите *Н0* в поле *Значение*.
8. В поле *Метка значения* введите *Нет* ответа.
9. Нажмите **Добавить**, чтобы добавить метку в список меток значений переменной.
10. Нажмите кнопку **ОК**, чтобы сохранить изменения и вернуться в Редактор данных.

Глава 4. Итожащие статистики для отдельных переменных

В этом разделе обсуждаются итожащие меры и то, как уровень измерения переменных влияет на выбор подходящих статистик. Мы будем использовать файл данных *demo.sav*. Дополнительную информацию смотрите в разделе [Глава 10, “Файлы для примеров”](#), на стр. 79.

Шкала измерения

Различные итожащие меры подходят для различных типов данных, в зависимости от уровня измерения данных.

Категориальное. Это данные с ограниченным числом уникальных значений или категорий (например, пол или семейное положение). Также такие данные называют **количественными данными**. Категориальные переменные могут быть текстовыми или числовыми, в которых категории закодированы числовыми кодами (например, 0 = *Не состоит в браке*, а 1 = *Состоит в браке*). Существует два основных типа категориальных данных.

- **Номинальные.** Это категориальные данные, для которых не существует естественного порядка категорий. Например, для переменной Отдел категория *продажи* не может быть больше или меньше категории *маркетинг* или категории *бухгалтерия*.
- **Порядковые.** Порядковые данные - это категориальные данные, для которых существует естественный порядок категорий, но расстояние между категориями невозможно измерить. Например, для значений *высокий*, *средний* и *низкий* существует естественный порядок, но рассчитать "расстояния" между значениями - невозможно.

Масштаб. Это данные, измеренные на интервальной шкале или на шкале отношений, для которых существует и порядок значений, и расстояния между значениями. Например, зарплата 7195 рублей больше зарплаты 5398 рублей, а разница между этими зарплатами - 1797 рублей. Также такие данные называют **количественными** или **непрерывными**.

Сводные показатели для категориальных данных

Наиболее распространенная итожащая мера для категориальных данных - это количество и процент наблюдений в каждой категории. **Мода** - это категория с наибольшим количеством наблюдений. Для порядковых данных, особенно с большим числом категорий, полезной итожащей мерой является **медиана** (значение, выше и ниже которого находится половина наблюдений).

Процедура Частоты создает частотные таблицы с количествами и процентами наблюдений для всех наблюдаемых значений переменной.

1. Выберите в меню:

Анализ > Описательные статистики > Частоты...

Примечание: Для этой функциональной возможности требуется модуль Statistics Base.

2. Выберите переменные *Наличие персонального цифрового помощника (PDA) [ownprda]* и *Наличие телевизора [owntv]* и перенесите их в список Переменные.

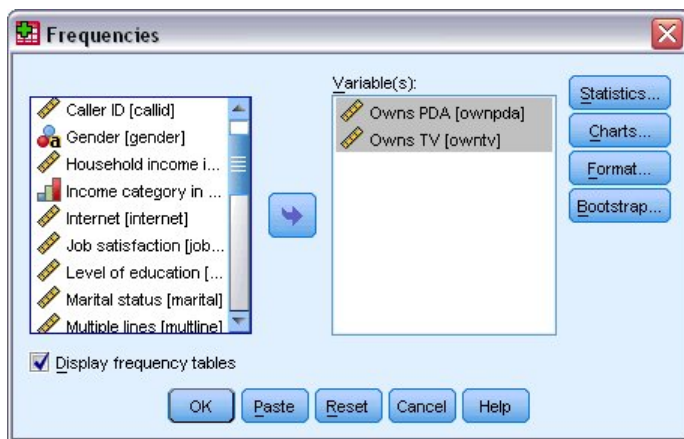


Рисунок 26. Выбранные для анализа категориальные переменные

3. Нажмите кнопку **OK**, чтобы запустить эту процедуру.

Frequency Table

Owns PDA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	5093	79.6	79.6	79.6
	Yes	1307	20.4	20.4	100.0
Total		6400	100.0	100.0	

Owns TV

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	63	1.0	1.0	1.0
	Yes	6337	99.0	99.0	100.0
Total		6400	100.0	100.0	

Рисунок 27. Таблицы Частота

Таблицы частот выводятся в окне просмотра. Таблицы частот показывают, что только 20,4% респондентов владеет персональными цифровыми помощниками, но почти все - (99%) - владеют телевизорами. Это может и не быть интересным открытием, хотя нас может заинтересовать небольшая группа людей, у которых нет телевизоров.

Диаграммы для категориальных данных

При помощи столбчатой или круговой диаграммы информацию в частотной таблице можно представить в графическом виде.

1. Снова откройте диалоговое окно Частоты. (Обе переменные должны, по-прежнему, быть выбраны.)

Можно воспользоваться кнопкой Повторный вызов диалога на панели инструментов для быстрого перехода к выполненным недавно процедурам.



Рисунок 28. Кнопка Повторный вызов диалога

2. Нажмите кнопку **Диаграммы**.
3. Выберите **Столбчатые**, а затем нажмите кнопку **Продолжить**.
4. Нажмите кнопку **ОК** в главном диалоговом окне, чтобы запустить процедуру.

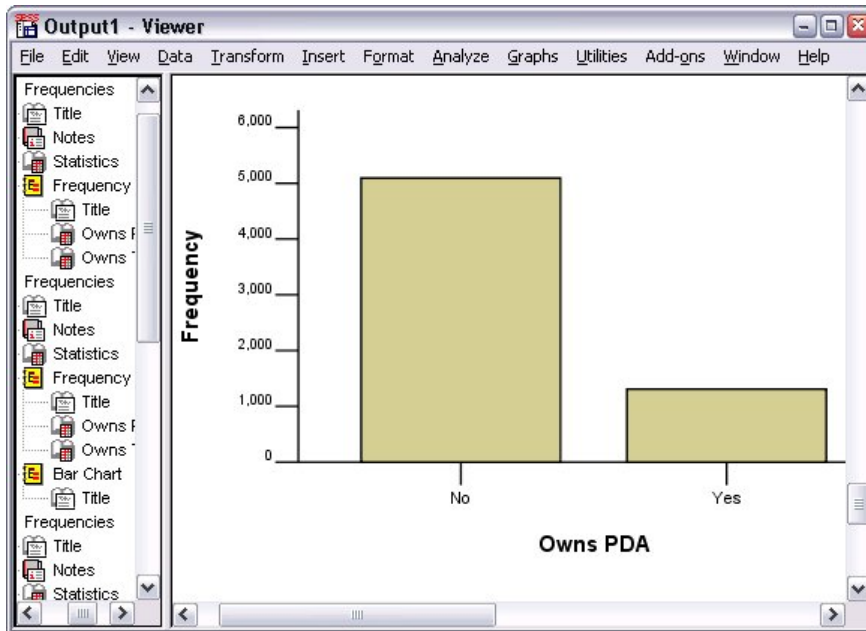


Рисунок 29. Столбчатая диаграмма

Теперь, та же самая информация, что и в частотных таблицах, показана в форме столбчатой диаграммы, что позволяет еще лучше увидеть, что большинство респондентов не владеют персональными цифровыми помощниками и практически все владеют телевизорами.

Сводные показатели для количественных переменных

Для количественных переменных существует целый ряд сводных показателей, среди которых:

- **Показатели центральной тенденции.** Наиболее распространенные меры центральной тенденции - это **среднее значение** (среднее арифметическое) и **медиана** (значение, выше и ниже которого находится ровно половина наблюдений).
 - **Показатели дисперсии.** Статистики, измеряющие вариацию или разброс в данных, включают стандартное отклонение, минимум и максимум.
1. Снова откройте диалоговое окно Частоты.
 2. Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
 3. Выберите *Доход домохозяйства в тысячах [income]* и перенесите в список Переменные.
 4. Нажмите кнопку **Статистики**.
 5. Выберите **Среднее, Медиана, Стандартное отклонение, Минимум и Максимум**.
 6. Нажмите кнопку **Продолжить**.
 7. Удалите флажок **Вывести частотные таблицы** в основном диалоговом окне. (Частотные таблицы, обычно, не очень полезны для количественных переменных, поскольку в них может быть почти столько же уникальных значений, сколько наблюдений в данных.)
 8. Нажмите кнопку **ОК**, чтобы запустить эту процедуру.

Статистики процедуры Частоты появились в окне вывода результатов средства просмотра.

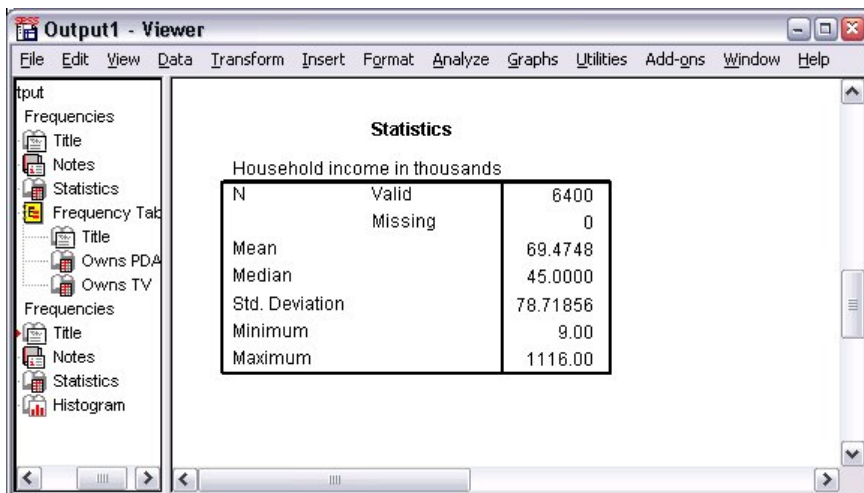


Рисунок 30. Таблица Статистики процедуры Частоты

В этом примере между средним значением и медианой наблюдается большая разница. Среднее значение почти на 25000 больше, чем медиана. Это означает, что значения не распределены нормально. Визуально представление о распределении можно получить при помощи гистограммы.

Гистограммы для количественных переменных

1. Снова откройте диалоговое окно Частоты.
2. Нажмите кнопку **Диаграммы**.
3. Выберите **Гистограммы** и установите переключатель **С нормальной кривой**.
4. Нажмите **Продолжить**, а затем **ОК** в главном диалоговом окне, чтобы запустить процедуру.

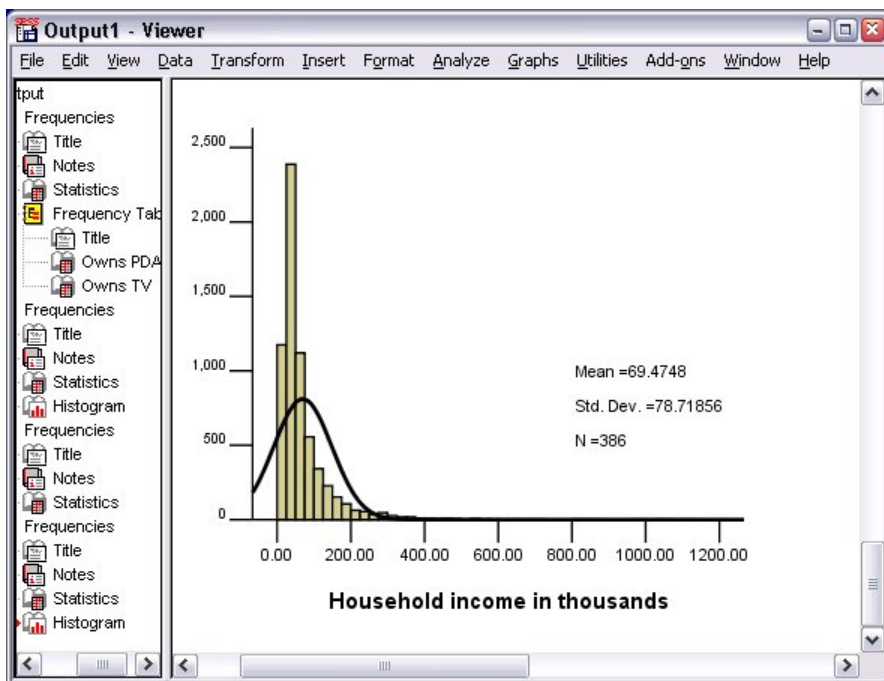


Рисунок 31. Гистограмма

Большая часть наблюдений сконцентрирована в начале диапазона значений (на отрезке от 0 до 100 000). Тем не менее, есть достаточное количество наблюдений и за пределами этого отрезка (это можно увидеть, даже не преобразуя диаграмму). Эти большие значения существенно влияют

на среднее, но не оказывают влияния на медиану, что делает последнюю лучшим индикатором центральной тенденции для нашего случая.

Глава 5. Создание и редактирование диаграмм

Пользователь может создавать и редактировать различные типы диаграмм. В этой главе рассказывается о том, как создать и изменить столбчатую диаграмму. Пользователь применяет те же принципы при работе с другими типами диаграмм.

Основы построения диаграмм

Для демонстрации основ создания диаграмм, мы создадим столбчатую диаграмму средних доходов для различных уровней удовлетворенности работой. В этом примере использован файл данных *demo.sav*. Дополнительную информацию смотрите в разделе Глава 10, “Файлы для примеров”, на стр. 79.

1. Выберите в меню:

Графики > Построитель диаграмм...

Диалоговое окно Построителя диаграмм - это интерактивное окно, позволяющее видеть, как будет выглядеть диаграмма, когда вы ее построите.

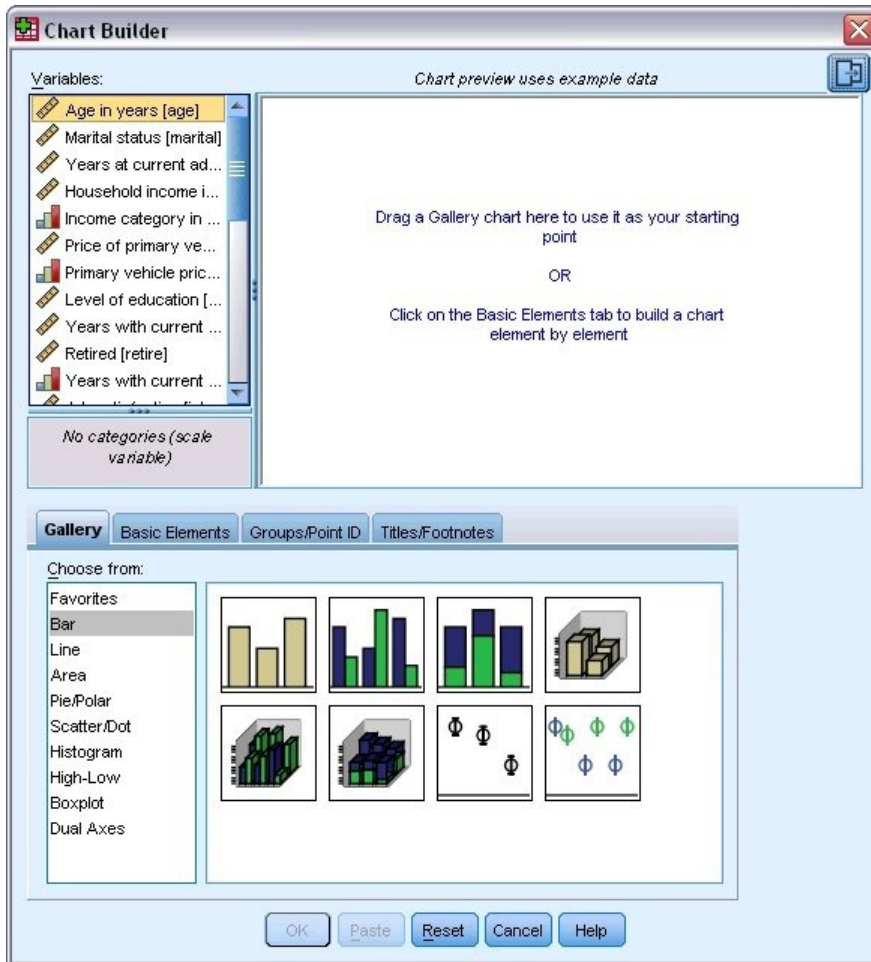


Рисунок 32. Диалоговое окно Построитель диаграмм

Использование галереи Построителя диаграмм

1. Щелкните по вкладке **Галерея** (если она не выбрана).

В Галерее содержатся различные заданные заранее диаграммы, которые организованы по типам. Вкладка Базовые элементы содержит базовые элементы (такие как оси и графические элементы) для создания диаграмм с чистого листа, но использовать Галерею проще.

2. Выберите **Столбики**, (если этот пункт не выбран).

Значками представлены доступные типы столбчатых диаграмм в галерее, расположенной в данном диалоговом окне. Значки дают достаточно информации, чтобы понять каким типам диаграмм они соответствуют. Если, все же, информации недостаточно, можно подвести курсор мыши к значку, чтобы на экране появилась всплывающая подсказка.

3. Перенесите (при помощи мыши) значок простой столбчатой диаграммы в панель макета диаграммы - большую область над Галереей. В панели макета диаграммы теперь можно увидеть, как (приблизительно) будет выглядеть диаграмма. Обратите внимание на то, что изображение диаграммы не основывается на данных, открытых в Редакторе данных. Это данные из примера.

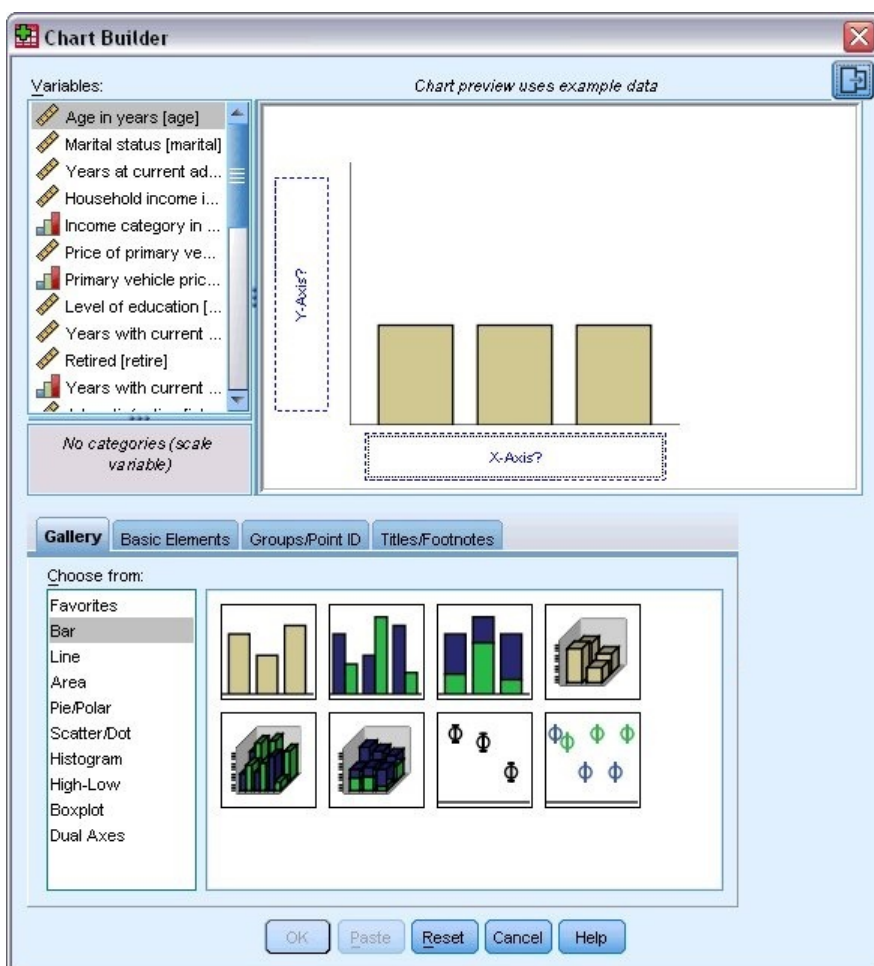


Рисунок 33. Столбчатая диаграмма на панели макеты диаграммы

Задание переменных и статистик (итожащих показателей)

Хотя на панели макета диаграммы и находится диаграмма, процесс ее построения не закончен, так как переменная или статистика, задающая высоту столбиков, а также переменная, категориям которой соответствуют столбики, - не заданы. Невозможно построить диаграмму без переменных и без статистик. Переменные можно добавлять, перетаскивая их из списка Переменные, расположенного слева от панели макета диаграммы.

Шкала измерения переменной играет большую роль в Построителе диаграмм. В поле оси x мы перенесем переменную *Удовлетворенность работой*. Однако значок (в виде линейки) рядом с переменной показывает, что для этой переменной задана количественная шкала измерения.

Чтобы создать требуемую диаграмму, необходимо использовать категориальную шкалу измерения. Вместо того, чтобы возвращаться в Редактор данных и там изменять шкалу измерения, можно сделать это (на время) в Конструкторе диаграмм.

1. Щелкните правой кнопкой мыши по переменной *Удовлетворенность работой* в списке Переменные и выберите **Порядковая**. Порядковая шкала подходит для измерения *Удовлетворенности работой*, поскольку категории этой переменной имеют естественный порядок. Обратите внимание, что значок изменился после изменения шкалы измерения.
2. Теперь перенесем *Удовлетворенность работой* из списка Переменные в поле оси x.

По умолчанию в поле оси x находится статистика (итожащий показатель) *Частота*. Если необходимо использовать другую статистику (например, процент или среднее значение), поле оси x можно изменить. В нашем примере мы не будем использовать ни одну из этих статистик, однако будет полезно познакомиться с процессом изменения статистик, чтобы иметь возможность делать это в дальнейшем.

3. Щелкните по вкладке **Свойства элемента** в боковой панели Конструктора диаграмм. (Если боковая панель не показана, то нажмите кнопку в правом верхнем углу Конструктора диаграмм, и боковая панель появится.)



Рисунок 34. Свойства элемента

Свойства элементов позволяют изменять свойства различных элементов диаграммы. Среди этих элементов - элементы графики (например, столбики столбчатой диаграммы) и оси диаграммы. Выберите один из элементов в списке Редактировать свойства, чтобы изменить свойства этого элемента. Обратите внимание на кнопку красным значком X, расположенную справа от списка. При помощи этой кнопки можно убрать элемент графики из панели макета диаграммы. Поскольку в списке выбран элемент **Столбики1**, показанные свойства относятся к элементу графики, а точнее, к элементу графики столбики.

В выпадающем списке Статистика показаны доступные статистики. Как правило, для всех типов диаграмм доступен один и тот же набор статистик. Необходимо учитывать, что для некоторых статистик требуется, чтобы в поле оси x находилась переменная.

4. Перенесите переменную *Доход семьи в тысячах* из списка Переменные в поле оси y. Поскольку переменная, выбранная для оси y - количественная, а переменная, выбранная для оси x - категориальная (а точнее порядковая), для оси y по умолчанию выбрана статистика *Среднее*.

Таким образом, мы выбрали переменные и статистики, которые хотели, поэтому необходимость в изменении свойств элементов отсутствует.

Добавление текста

{\f3 Пользователь может добавлять на диаграмму заголовки и сноски}{\f5 .}

1. Откройте вкладку **Заголовки и сноски** .
2. Выберите **Заголовок 1**.

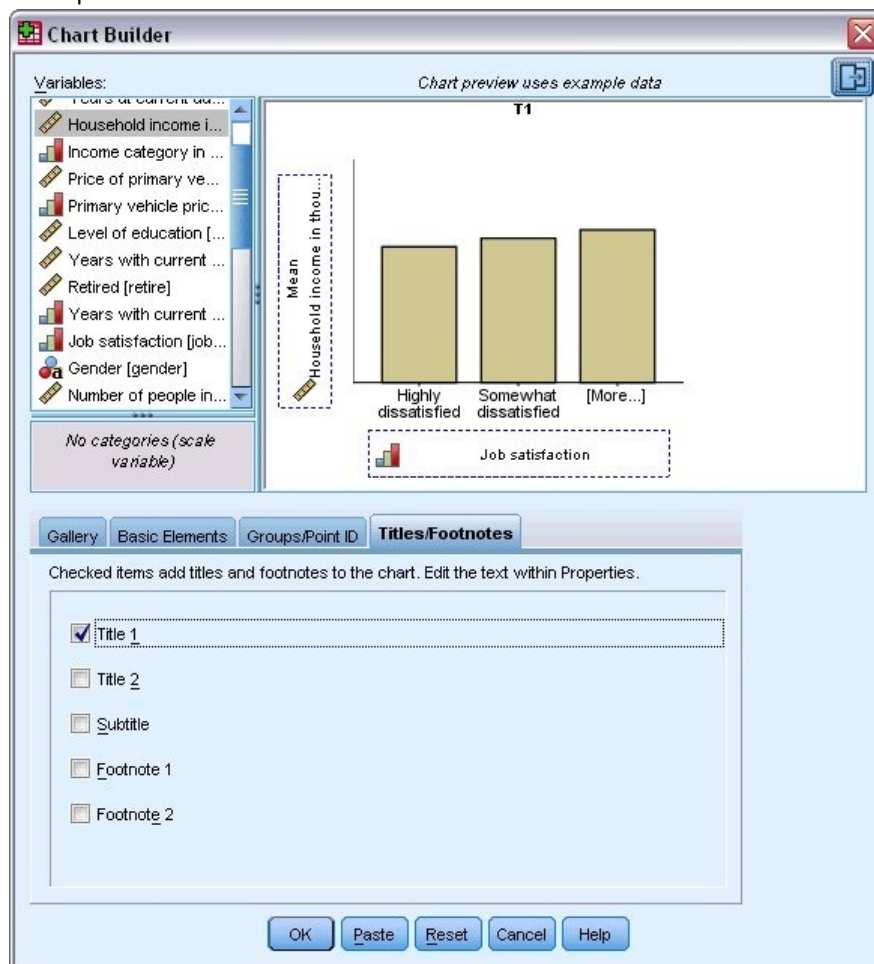


Рисунок 35. Заголовок 1 на панели макета диаграммы

Заголовок появился на панели макета диаграммы с меткой **31**.

3. На вкладке **Свойства элемента** выберите **Заголовок 1** в списке Редактировать свойства.
4. В поле Содержимое введите Доход и удовлетворенность работой. Это текст, который появится в заголовке.

Построение диаграмм

1. Нажмите кнопку **ОК**, чтобы построить диаграмму.

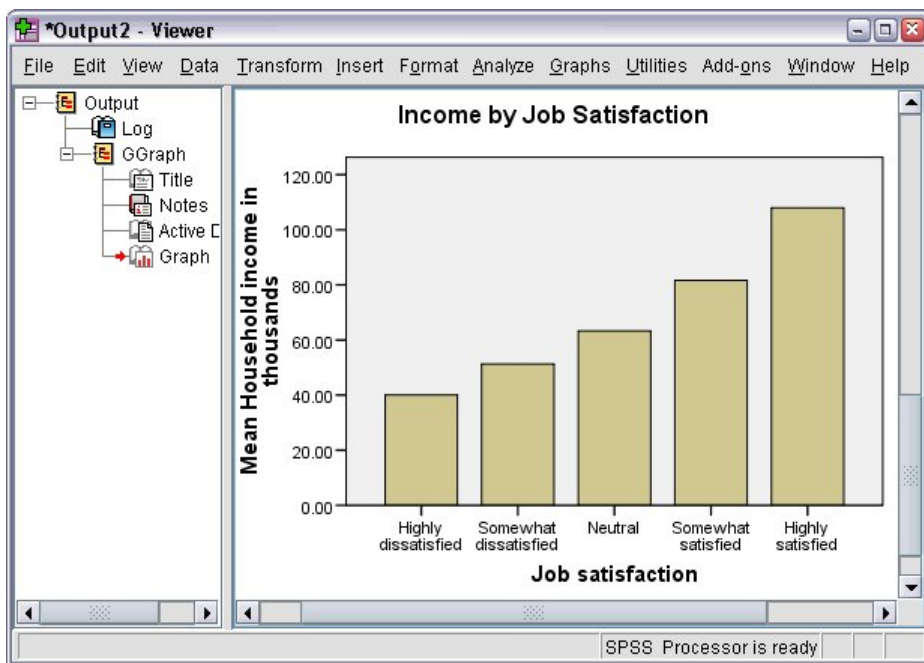


Рисунок 36. Столчатая диаграмма

Столчатая диаграмма показывает, что у тех, кто больше удовлетворен работой, выше доход домохозяйства.

Глава 6. Работа с выводом результатов

Результаты выполнения статистических процедур выводятся в средстве просмотра. Результаты могут выводиться в виде статистических таблиц, диаграмм и графиков, а также в виде текста, в зависимости от параметров, выбираемых при запуске процедур. В этом разделе используются файлы *viewertut.spv* и *demo.sav*. Дополнительную информацию смотрите в разделе Глава 10, “Файлы для примеров”, на стр. 79.

Работа со средством просмотра

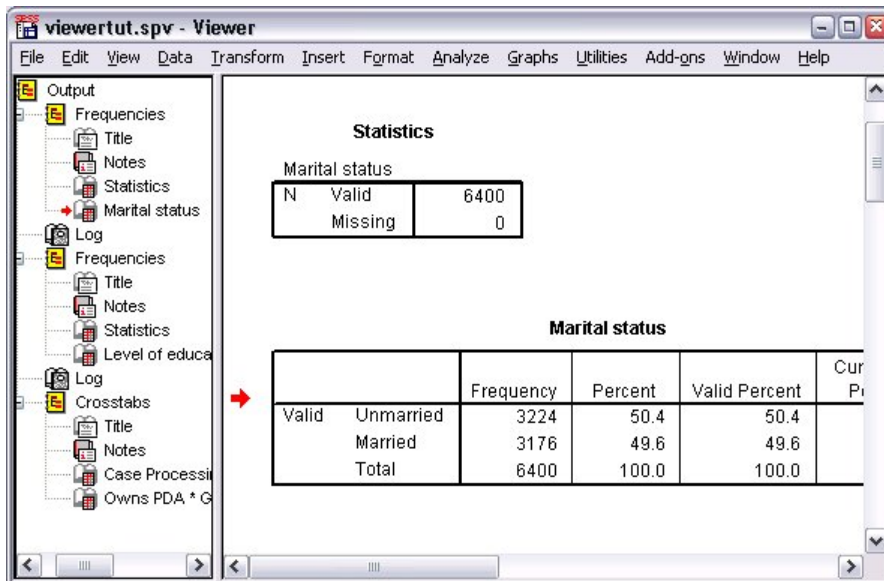


Рисунок 37. Средство просмотра

Окно средства просмотра делится на две панели. В **панели схемы вывода** находится схема всей информации, содержащейся в средстве просмотра. В панели **содержимого** находятся сами результаты - статистические таблицы, диаграммы и текстовый вывод.

Перемещайтесь по содержимому окон с помощью вертикальной и горизонтальной полос прокрутки. Для ускорения перемещения, можно, щелкая по элементам на панели схемы перемещаться к соответствующим элементам на панели содержимого.

1. Перетащите правую границу панели схемы, чтобы изменить ее ширину.

Значок на панели схемы в виде открытой книги означает, что соответствующий элемент показан в средстве просмотра, хотя сам элемент в настоящий момент может быть и не виден на экране.

2. Чтобы скрыть таблицу или диаграмму, дважды щелкните по значку в виде книги на панели схемы.

Значок в виде открытой книги изменяется на значок в виде закрытой книги, указывая на то, что связанная с ним информация теперь скрыта.

3. Чтобы снова показать скрытый вывод, дважды щелкните по значку в виде закрытой книги.

Можно скрыть, как вывод конкретной статистической процедуры, так и весь вывод в средстве просмотра.

4. Щелкните по полю со знаком минус (-) слева от процедуры, результаты которой вы хотите скрыть, или щелкните по переключателю рядом с верхним элементом на панели схемы, чтобы скрыть весь вывод.

Схема вывода сворачивается, что означает, что результаты оказываются скрытыми.

Также можно изменить порядок элементов вывода.

5. На панели схемы щелкните по элементам, которые вы хотите переместить.

6. Перетащите выбранные элементы в новое место в схеме.

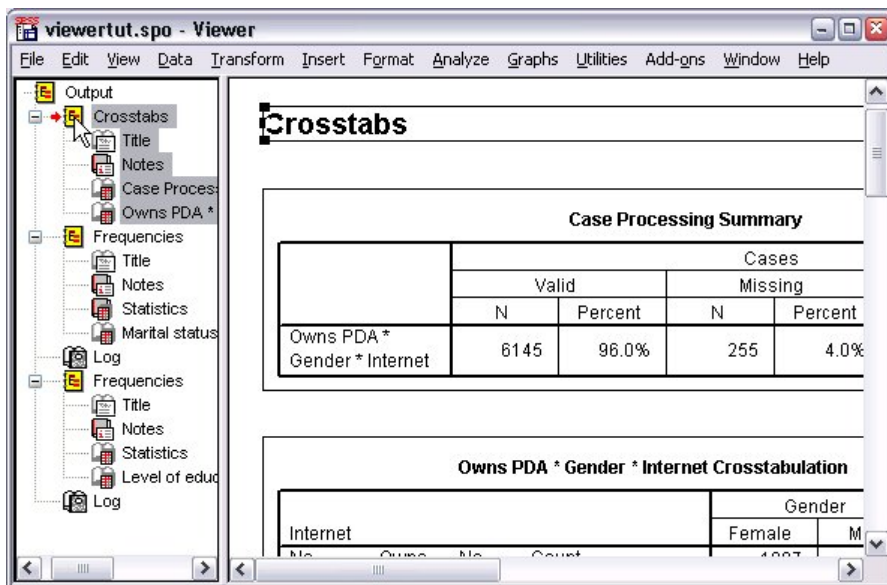


Рисунок 38. Изменение порядка вывода в средстве просмотра

Вы также можете перемещать элементы, перетаскивая их на панели содержимого.

Использование Редактора мобильных таблиц

Результаты выполнения большинства статистических процедур попадают в **мобильные таблицы**.

Получение определений терминов вывода

В выводе появляется много статистических терминов. Определения этих терминов можно получить непосредственно в средстве просмотра.

1. Дважды щелкните по таблице *Таблица сопряженности Наличие персонального цифрового помощника (PDA) * Пол * Интернет*.
2. Щелкните правой кнопкой мыши по тексту в таблице *Ожидаемая частота* и выберите во всплывающем меню **Что это такое?**.

Определение выводится во всплывающем окне.

Owns PDA * Gender * Internet Crosstabulation

Internet			Gender		Total	
			Female	Male		
No	Owns PDA	Count	1897	1962	3859	
		Expected Count	1903.4	1955.6	3859.0	
	% within Owns PDA	50.3%	49.7%	100.0%		
	Count	650	650.0	100.0%		
Yes	Owns PDA	Count	513	512	1025	
		Expected Count	512.5	512.5	1025.0	
	% within Owns PDA	50.0%	50.0%	100.0%		
	Count	611	611.0	100.0%		
Total			Count	2224	2285	4509
			Expected Count	2224.0	2285.0	4509.0
			% within Owns PDA	49.3%	50.7%	100.0%

The number of cases that would be expected in the cell if the row and column variables are statistically independent or unrelated to one another.

Рисунок 39. Всплывающее определение

Преобразование таблиц

В таблицах, создаваемых по умолчанию, информация может быть показана недостаточно ясно. При помощи сводных таблиц можно поменять местами строки и столбцы (транспонировать таблицу), настроить порядок данных в таблице и изменить таблицу многими другими способами. Например, можно превратить короткую, широкую таблицу в длинную и узкую, транспонировав строки и столбцы. Изменение макета таблицы не влияет на результаты. Однако оно позволяет представить информацию в другом, более наглядном виде.

1. Если данная таблица не активирована, щелкните дважды *Наличие персонального цифрового помощника (PDA) * Пол * Интернет*
2. Если окна Поля вращения нет на экране, выберите в меню:

Вращение > Поля вращения

Поля вращения позволяют перемещать данные между столбцами, строками и слоями.

Internet			Gender		Total	
			Female	Male		
No	Owns PDA	Count	1897	1962	3859	
		Expected Count	1903.4	1955.6	3859.0	
	% within Owns PDA	49.2%	50.8%	100.0%		
	Count	327	323	650		
Yes	Owns PDA	Count	513	512	1025	
		Expected Count	512.5	512.5	1025.0	
	% within Owns PDA	50.0%	50.0%	100.0%		
	Count	611	611.0	100.0%		
Total			Count	2224	2285	4509
			Expected Count	2224.0	2285.0	4509.0
			% within Owns PDA	49.3%	50.7%	100.0%

Pivoting Trays

Layer: [Empty]

Column: Gender

Row: Internet, Owns PDA, Statistics

Рисунок 40. Поля вращения

3. Перетащите элемент *Статистика* из измерения Строки в измерение Столбцы, ниже переменной *Пол*. Таблица сразу же изменяется в соответствии с внесенными изменениями.

Порядок элементов в поле вращения отражает порядок представления соответствующих элементов в таблице.

4. Перетащите элемент *Наличие персонального цифрового помощника (PDA)* в позицию перед элементом *Интернет* в измерении строк, чтобы поменять их местами.

			Gender					Total			
			Female		Male						
Internet			Count	Expected Count	% within Owns PDA	Count	Expected Count	% within Owns PDA	Count	Expected Count	% within Owns PDA
Owns PDA	No	No	1897	1903.4	49.2%	1897	1903.4	49.2%	3859.0	3859.0	100.0%
	Yes	No	650.0	650.0	100.0%	611.0	611.0	100.0%	1261.0	1261.0	100.0%
Total	No	Yes	2547.0	2547.0	100.0%	2508.0	2508.0	100.0%	5055.0	5055.0	100.0%
	Yes	Yes	1636.0	1636.0	100.0%	1636.0	1636.0	100.0%	3272.0	3272.0	100.0%

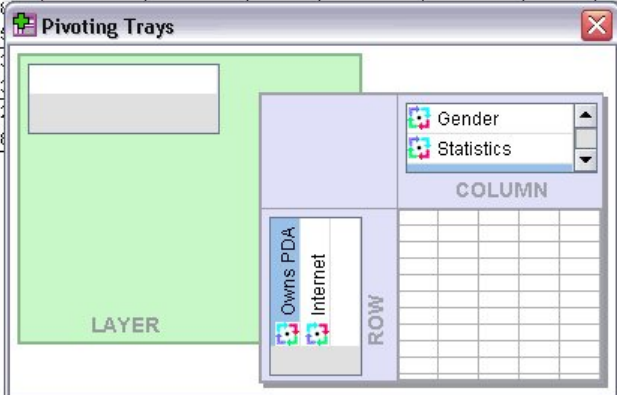


Рисунок 41. Перестановка строк

Создание и показ слоев

Слои могут быть полезны в больших таблицах со вложенными категориями. Создавая слои, можно упростить внешний вид таблицы и сделать ее более удобочитаемой.

1. Перетащите элемент *Пол* из измерения Столбцы в измерение Слои.

Gender			Female		
Internet			Count	Expected Count	% within Owns PDA
Owns PDA	No	No	1897	1903.4	49.2%
	Yes	No	650.0	650.0	100.0%
Total	No	Yes	2547.0	2547.0	100.0%
	Yes	Yes	1636.0	1636.0	100.0%

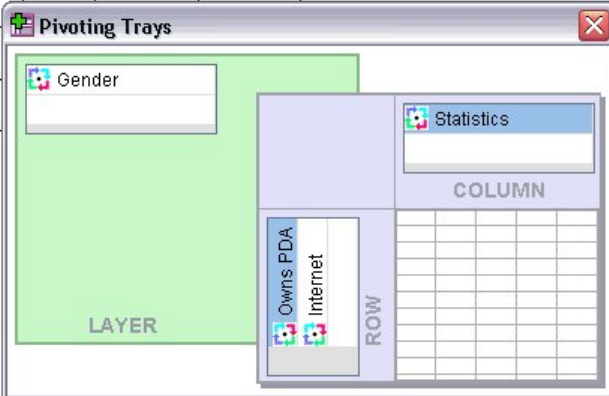


Рисунок 42. Значок Пол в измерении Слои

Для просмотра другого слоя выберите категорию из списка в таблице.

Редактирование таблиц

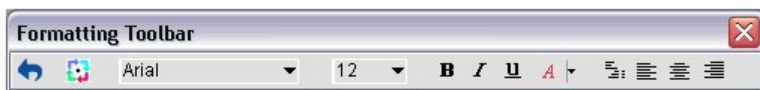
Если пользовательский шаблон таблиц не создан, создаваемые мобильные таблицы имеют стандартное форматирование. Формат отображения любого текста в таблице можно изменить.

Можно изменить такие параметры, как шрифт, размер шрифта, начертание шрифта (полужирный или курсив), а также цвет шрифта.

1. Дважды щелкните по таблице *Уровень образования*.
2. Если панель форматирования не видна, выберите в меню:

Вид > Панель инструментов

3. Щелкните по тексту заголовка *Уровень образования*.
4. В выпадающем списке размеров шрифтов на панели инструментом выберите **12**.
5. Чтобы изменить цвет текста заголовка, щелкните по инструменту Цвет текста и выберите цвет.



Level of education					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7	21.7
	High school degree	1936	30.3	30.3	52.0
	Some college	1360	21.3	21.3	73.2
	College degree	1355	21.2	21.2	94.4
	Post-undergraduate degree	359	5.6	5.6	100.0
Total		6400	100.0	100.0	

Рисунок 43. Измененный текст заголовка мобильной таблицы

Также можно редактировать содержимое таблиц и метки. Например, можно изменить заголовок этой таблицы.

6. Дважды щелкните по таблице.
7. Введите Образование в качестве нового заголовка.

Примечание: После изменения значений в таблице итоги и другие статистики не пересчитываются.

Скрытие строк и столбцов

Некоторые данные в таблице могут быть бесполезными или без необходимости усложнять таблицу. К счастью, в таблицах можно скрывать строки и столбцы без потери данных.

1. Если данная таблица не активирована, щелкните дважды *Наличие персонального цифрового помощника (PDA) * Пол * Интернет*
2. Чтобы выделить столбец, щелкните по метке столбца *Валидный процент*.
3. Выберите в меню Правка или во всплывающем меню:

Выбрать > Ячейки данных и метки

4. В меню Вид выберите **Скрыть** или во всплывающем меню выберите **Скрыть категорию**.

Теперь столбец скрыт, но не удален.

Education Level					
		Frequency	Percent	Cumulative Percent	
Valid	Did not complete high school	1390	21.7	21.7	
	High school degree	1936	30.3	52.0	
	Some college	1360	21.3	73.2	
	College degree	1355	21.2	94.4	
	Post-undergraduate degree	359	5.6	100.0	
Total		6400	100.0		

Рисунок 44. Столбец Валидный процент скрыт

Чтобы снова показать скрытый столбец:

5. Выберите в меню:

Вид > Показать все

По аналогии со столбцами можно скрывать и показывать строки.

Изменение формата данных в таблицах

Формат представления данных в мобильных таблицах можно изменить.

1. Если данная таблица не активирована, щелкните дважды *Наличие персонального цифрового помощника (PDA) * Пол * Интернет*
2. Чтобы выделить столбец, щелкните по метке столбца *Процент*.
3. Выберите в меню Правка или во всплывающем меню:

Выбрать > Ячейки данных

4. В меню Формат или во всплывающем меню, вызываемом щелчком правой кнопкой мыши, выберите **Свойства ячейки**.
5. Перейдите на вкладку **Формат значения**.
6. Введите 0 в поле Десятичных знаков, чтобы скрыть все десятичные знаки в этом столбце.

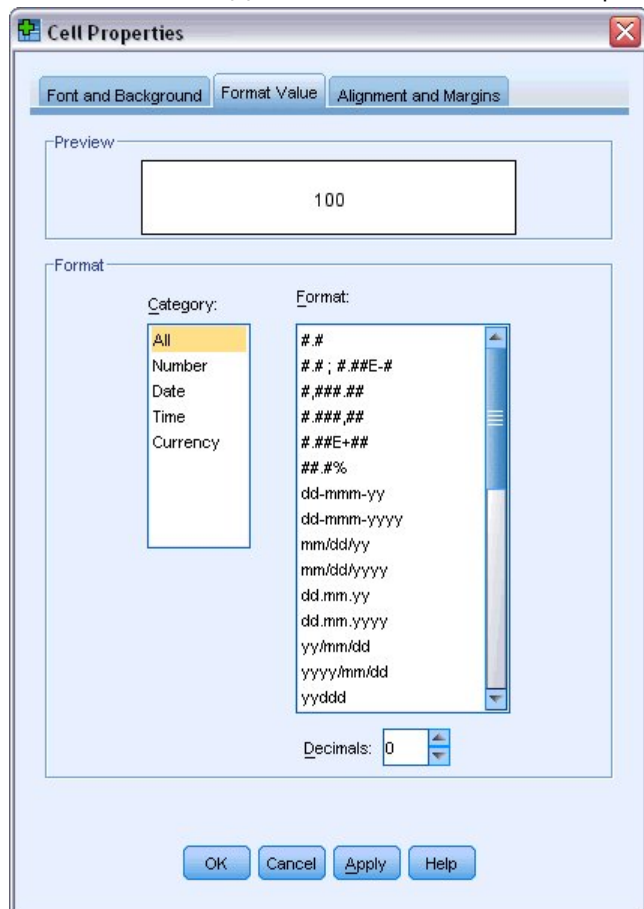


Рисунок 45. Свойства ячеек, вкладка Формат значения.

Также, в этом диалоговом окне можно изменить тип и формат данных.

7. Выберите тип данных в списке Категория, а затем выберите формат для выбранного типа данных в списке Формат.
8. Нажмите кнопку **ОК** или **Применить** для сохранения изменений.

Education Level

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	22	21.7
	High school degree	1936	30	52.0
	Some college	1360	21	73.2
	College degree	1355	21	94.4
	Post-undergraduate degree	359	6	100.0
	Total	6400	100	

Рисунок 46. Скрытые десятичные знаки в столбце Процент

Десятичные знаки в столбце Процент теперь оказались скрытыми.

Шаблоны таблиц

Наглядность и четкость представления результатов анализа во многом определяется внешним видом таблиц. Если таблица сложна для восприятия, информация, содержащаяся в ней, может оказаться непонятной.

Использование форматов, заданных заранее

1. Дважды щелкните по таблице Семейное положение.
2. Из меню выберите:

Формат > Шаблоны таблиц...

В диалоговом окне Шаблоны таблиц перечислено несколько заданных заранее шаблонов. Выбрав шаблон в списке, можно в правой части окна увидеть, как будет выглядеть таблица, к которой будет применен выбранный шаблон.

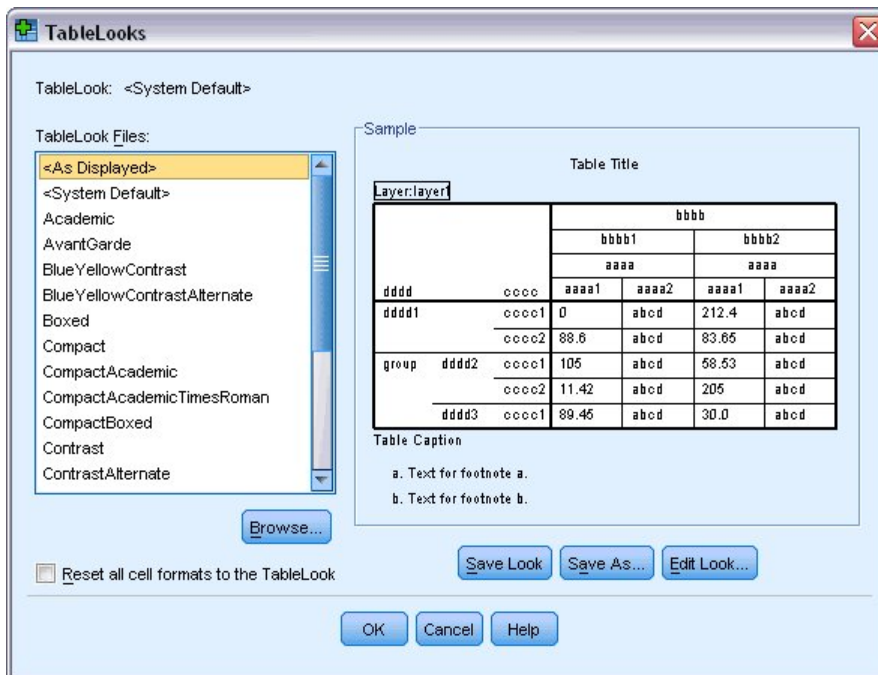


Рисунок 47. Диалоговое окно Шаблоны таблиц

Можно применить шаблон, не внося в него изменений. А можно изменить существующий шаблон, чтобы он лучше соответствовал Вашим потребностям.

3. Чтобы применить шаблон, выберите его и нажмите кнопку **OK**.

Изменение шаблонов таблиц

Можно настроить формат в соответствии с вашими потребностями. Практически любой аспект таблицы, начиная с фона и заканчивая стилями границ, можно изменить.

1. Дважды щелкните по таблице *Семейное положение*.
2. Из меню выберите:

Формат > Шаблоны таблиц...

3. Выберите стиль, ближайший к нужному вам формату, и нажмите кнопку **Изменить вид**.
4. Щелкните по вкладке **Формат ячеек**, чтобы перейти к заданию параметров форматирования.

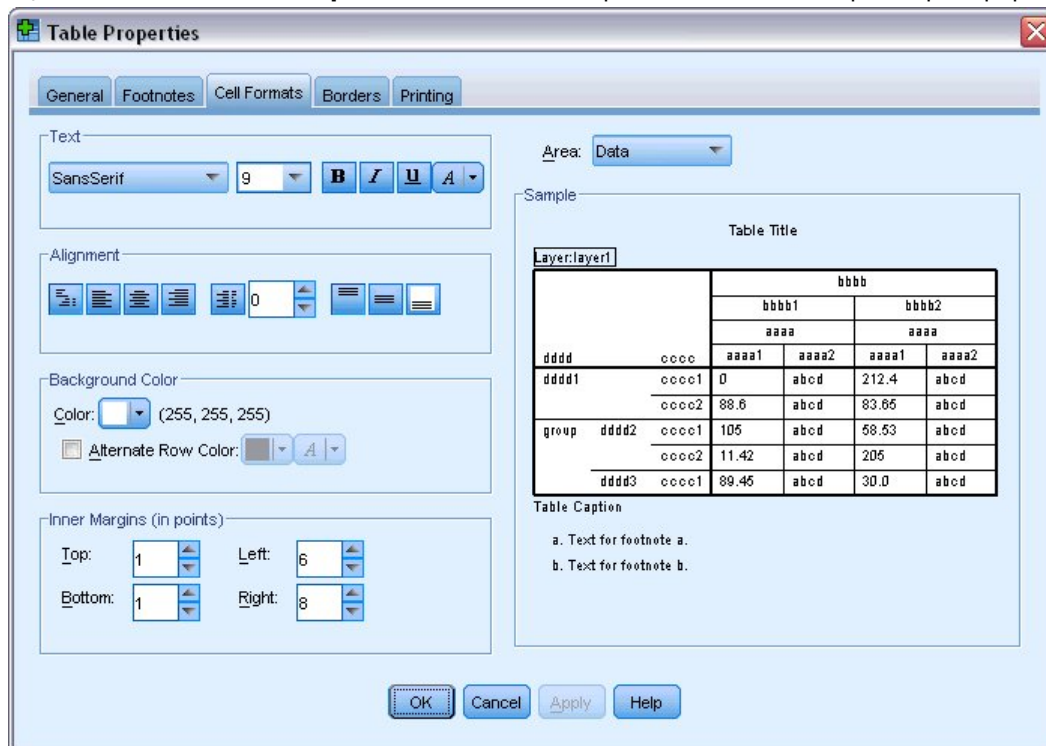


Рисунок 48. Диалоговое окно Свойства таблицы

Параметры форматирования включают шрифт, а также его размер, начертание и цвет. Кроме того, можно задать такие параметры, как выравнивание, цвет шрифта, цвет фона и размер полей.

В окне Образец справа выводится предварительный просмотр того, как изменения форматирования повлияют на таблицу. В каждой области таблицы могут быть свои стили форматирования. Например, вы, возможно не захотите, чтобы заголовок таблицы выглядел так же, как и данные. Выбрать область таблицы для редактирования можно по имени в выпадающем списке Область или же щелчком по этой области в окне Образец.

5. В выпадающем списке Область выберите **Данные**.
6. В выпадающем списке Фон выберите новый цвет.
7. Затем выберите новый цвет шрифта.

Вносимые изменения отражаются на панели Образец.

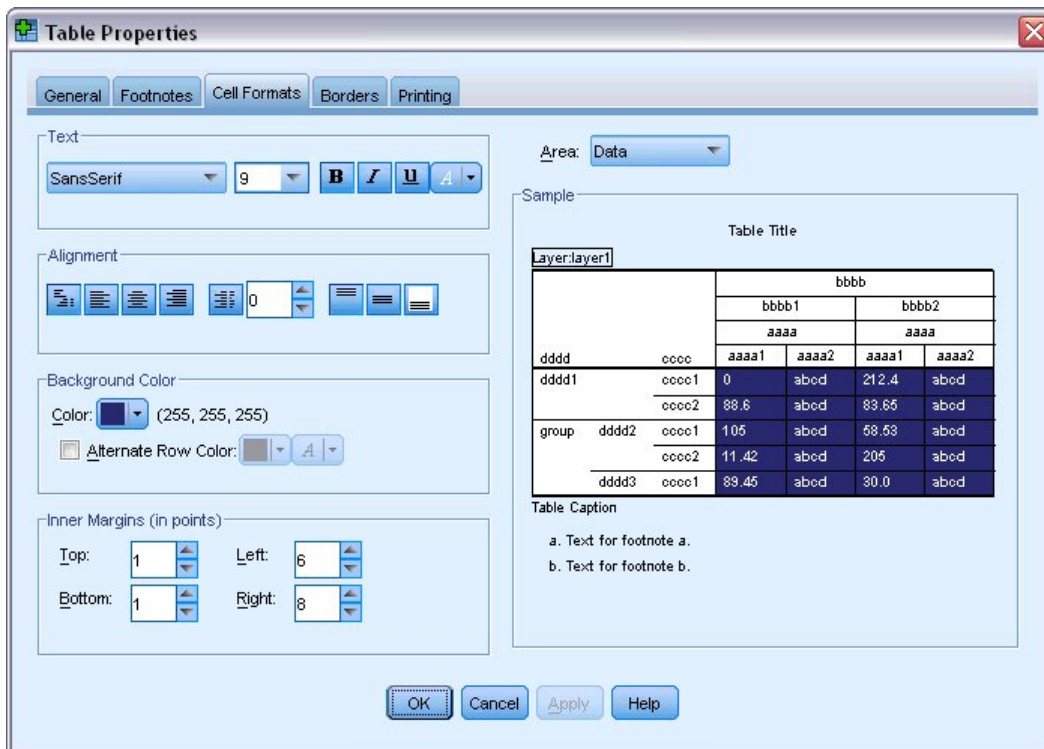


Рисунок 49. Изменение форматов ячеек таблицы

8. Нажмите кнопку **ОК**, чтобы вернуться в диалоговое окно Шаблоны таблиц.

Внесенные изменения можно сохранить в виде нового шаблона и применять этот шаблон к другим таблицам.

9. Нажмите кнопку **Сохранить как**.

10. Перейдите в каталог назначения и введите имя нового стиля в текстовом поле Имя файла.

11. Нажмите кнопку **Сохранить**.

12. Нажмите кнопку **ОК**, чтобы применить внесенные изменения и вернуться в средство просмотра.

Формат таблицы изменился в соответствии со сделанными изменениями.

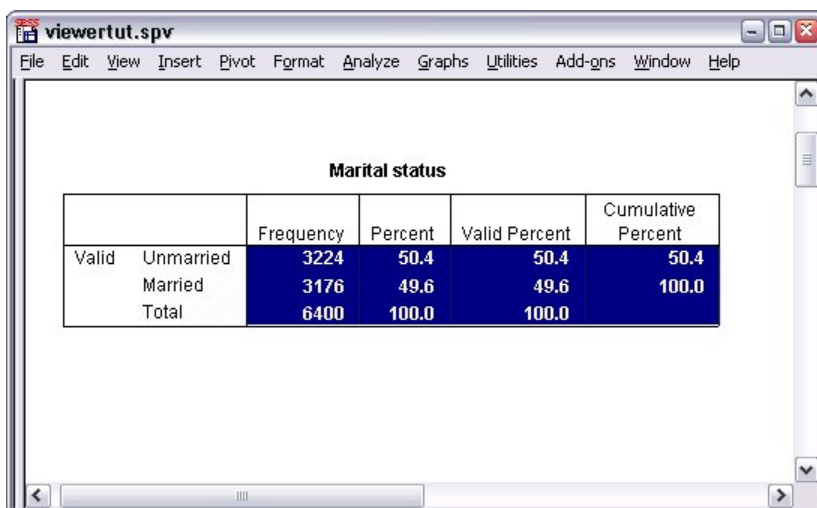


Рисунок 50. Измененный шаблон таблиц

Изменение заданного по умолчанию формата таблиц

Несмотря на то, что формат любой таблицы можно изменить после ее создания, может оказаться эффективнее изменить шаблон таблиц, заданный по умолчанию, чтобы не менять формат каждый раз при создании новой таблицы.

Чтобы изменить шаблон таблиц, заданный по умолчанию, выберите в меню:

Правка > Опции...

1. Щелкните по вкладке **Мобильные таблицы** в диалоговом окне Параметры.

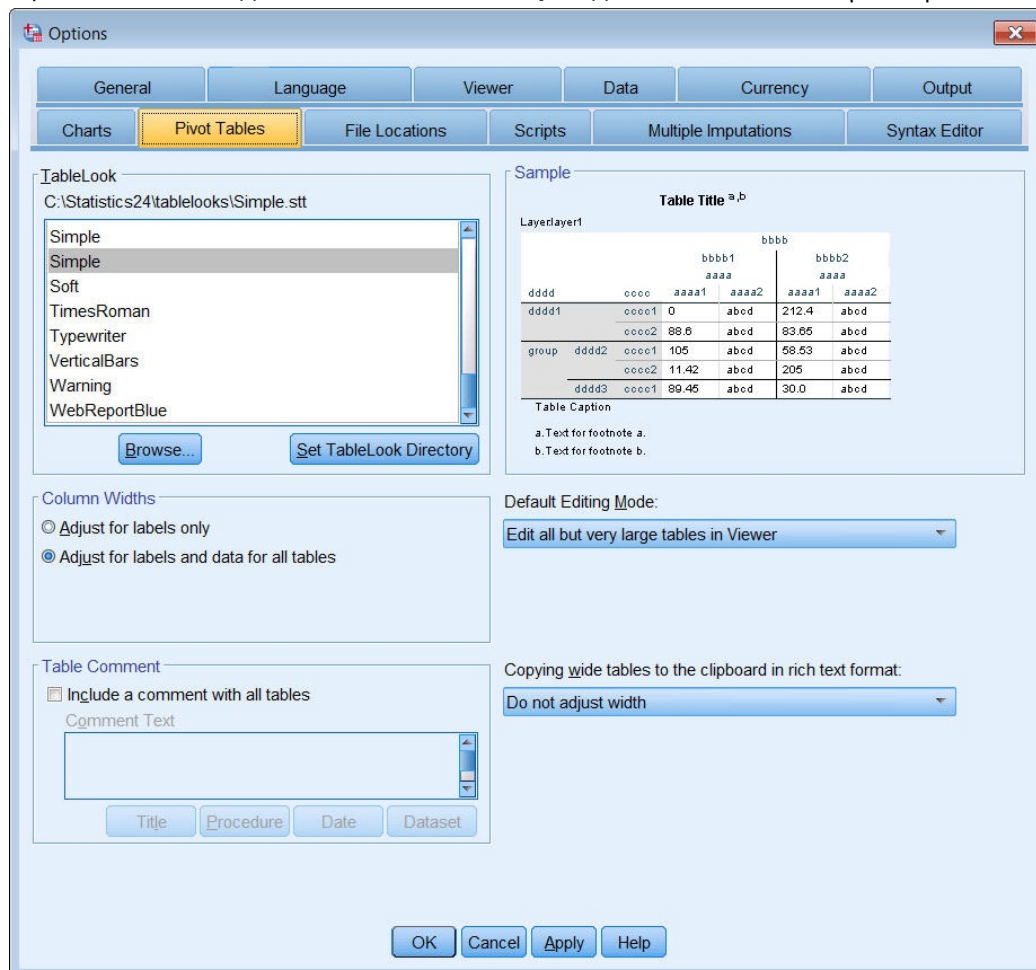


Рисунок 51. Диалоговое окно Параметры

2. Выберите шаблон, который Вы хотите использовать для всех новых таблиц.

В правой части диалогового окна показан образец шаблона таблиц, выделенного в списке Шаблон таблиц.

3. Нажмите кнопку **ОК**, чтобы сохранить выбранные параметры и закрыть диалоговое окно.

Все таблицы, которые будут создаваться после задания нового шаблона таблиц по умолчанию, будут соответствовать новому заданному шаблону.

Изменение параметров вывода

Среди параметров вывода результатов: выравнивание объектов в средстве просмотра, вывод или скрытие объектов по умолчанию, а также ширина окна просмотра. Чтобы изменить эти параметры:

1. Из меню выберите:

Правка > Опции...

2. Щелкните по вкладке **Средство просмотра**.

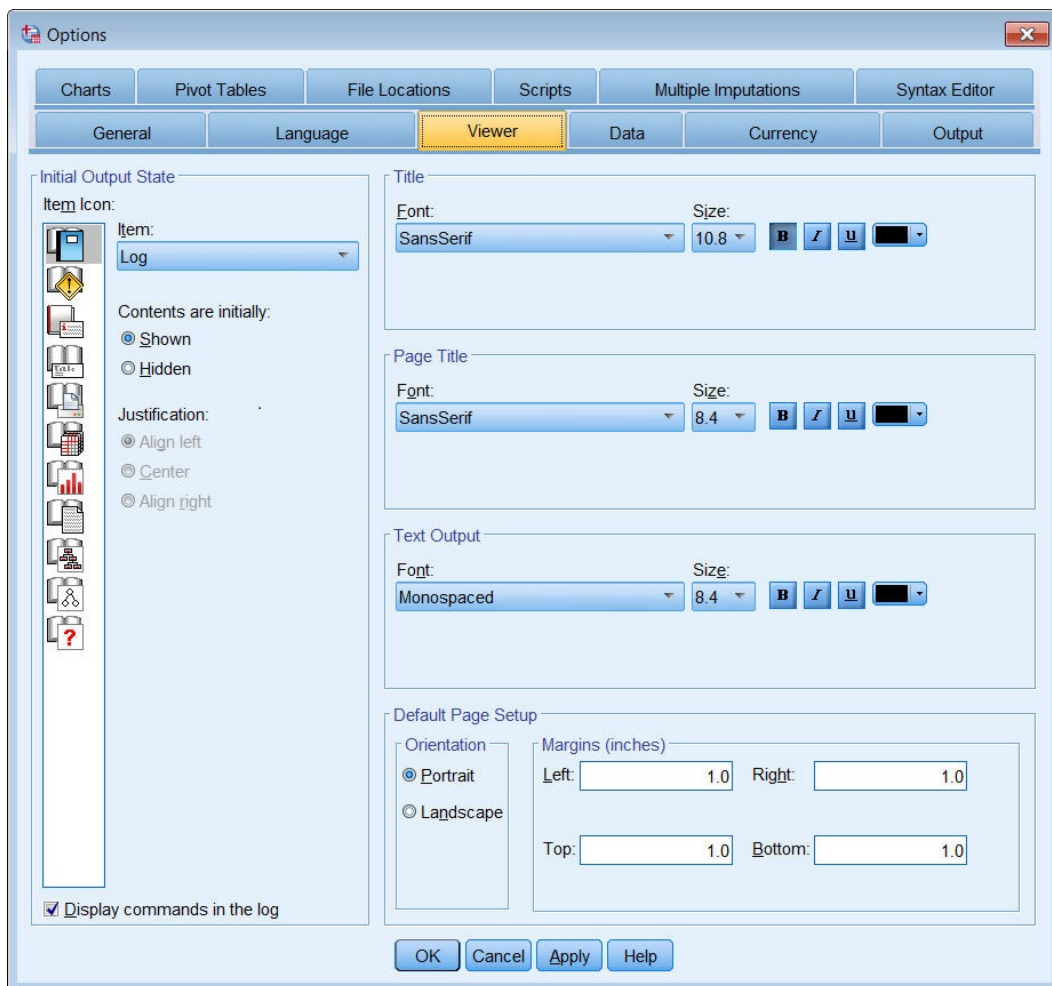


Рисунок 52. Опции средства просмотра

Параметры применяются на уровне объектов. Например, можно изменить параметры вывода диаграмм, оставив неизменными параметры вывода таблиц. Просто выберите объект, который вы хотите настроить, и внесите изменения.

3. Щелкните по значку **Заголовок**, чтобы задать параметры заголовков.

4. Щелкните **По центру**, чтобы все заголовки выводились в средстве просмотра по центру.

Можно также скрыть элементы, такие как сообщения журнала и предупреждения, которые перегружают вывод. Двойной щелчок по значку автоматически изменяет свойство вывода объекта.

5. Дважды щелкните по значку **Предупреждения**, чтобы скрыть предупреждения, которые будут появляться в выводе.

6. Нажмите кнопку **ОК**, чтобы сохранить изменения и закрыть диалоговое окно.

Вывод меток переменных и значений

В большинстве случаев вывод меток для переменных и значений эффективнее, чем вывод на экран имени переменной или фактического значения данных. Однако возможны случаи, когда вы захотите показывать и имена, и метки.

1. Из меню выберите:

Правка > Опции...

2. Щелкните по вкладке **Метки в выводе**.

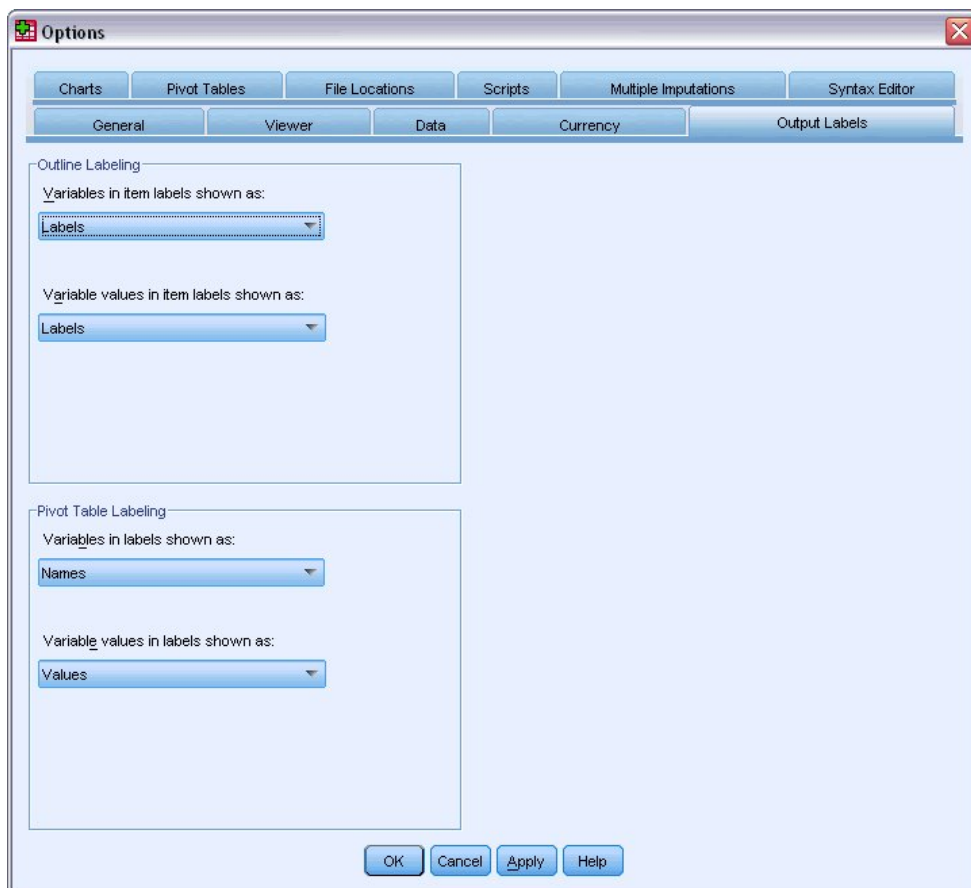


Рисунок 53. Группа параметров Метки в мобильных таблицах

Можно задавать разные параметры для панели схемы и панели содержания. Например, задать для панели схемы вывод меток, а для панели содержания вывод имен переменных и значений данных.

3. Выберите **Имена** в списке Переменные в метках, в группе параметром Метки в мобильных таблицах, чтобы в таблицах вместо меток выводились имена переменных.
4. Затем, выберите **Значения** в списке Значения переменных в метках, чтобы вместо меток значений в таблицах выводились сами значения.

Последующие таблицы, созданные в течение данного сеанса, отображают эти изменения.

marital

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3224	50.4	50.4	50.4
	1	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

Рисунок 54. Вывод имен и значений переменных

Использование результатов в других приложениях

Ваши результаты можно использовать во многих приложениях. Например, может понадобиться вставить таблицу или диаграмму в презентацию или отчет.

Хотя в примерах, приведенных ниже, используется Microsoft Word, вставка результатов производится аналогично и в других текстовых редакторах.

Вставка результатов в виде таблиц Word

Мобильные таблицы можно вставлять в Word в качестве собственных таблиц Word. Все атрибуты таблиц, такие как размеры шрифтов и цвета, сохраняются. Так как таблица вставляется в виде таблицы Word, ее можно редактировать в Word точно также, как и любую другую таблицу Word.

1. Щелкните по таблице в Средстве просмотра, чтобы выбрать его.
2. Из меню выберите:

Правка > Копировать

3. Откройте текстовый редактор.
4. Выберите в меню текстового редактора:

Правка > Специальная вставка...

5. Выберите в диалоговом окне Специальная вставка **Текст в формате RTF**.
6. Нажмите кнопку **ОК**, чтобы вставить результаты в текущий документ.

Таблица появилась в документе. Ее размеры и форматирование можно изменять, данные в ней можно редактировать.

Вставка результатов в виде текста

Мобильные таблицы можно вставлять в другие приложения в виде текста. При этом форматирование не сохраняется, но после вставки данные в таблице можно редактировать.

1. Щелкните по таблице в Средстве просмотра, чтобы выбрать его.
2. Из меню выберите:

Правка > Копировать

3. Откройте текстовый редактор.
4. Выберите в меню текстового редактора:

Правка > Специальная вставка...

5. Выберите в диалоговом окне Специальная вставка **Неформатированный текст**.
6. Нажмите кнопку **ОК**, чтобы вставить результаты в текущий документ.

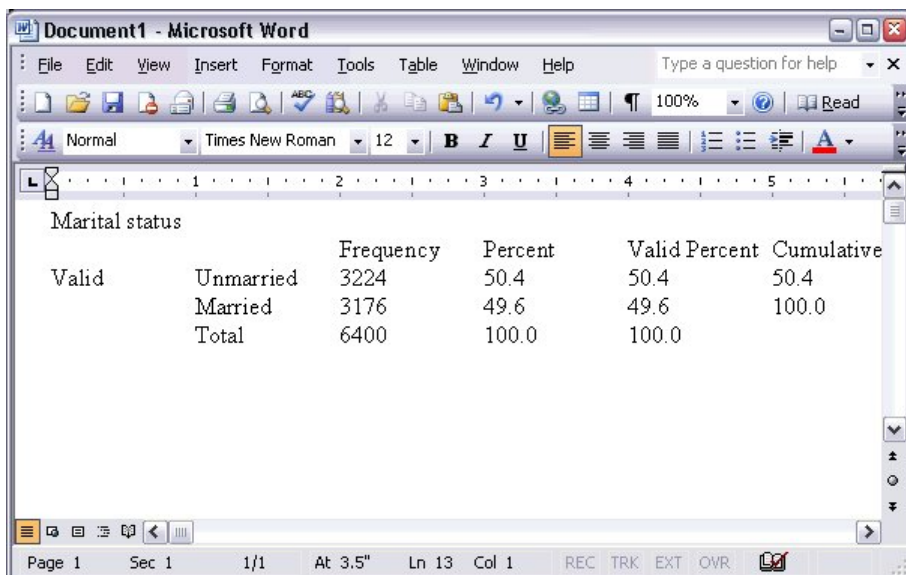


Рисунок 55. Мобильная таблица в Word

Столбцы таблицы разделены знаками табуляции. Ширину столбцов можно изменить, устанавливая позиции табуляции в текстовом редакторе.

Изучение результатов в файлах Microsoft Word, PowerPoint и Excel

Пользователь может экспортировать результаты в файлы Microsoft Word, PowerPoint или Excel. Можно экспортировать выделенные элементы средства просмотра или все элементы. В этом разделе используются файлы *msouttut.spv* и *demo.sav*. Дополнительную информацию смотрите в разделе [Глава 10, “Файлы для примеров”](#), на стр. 79.

Примечание: Экспорт в PowerPoint доступен только в операционных системах Windows и недоступен в студенческой версии.

На панели навигации (схемы) в средстве просмотра можно выбирать конкретные элементы для экспорта, а также экспортировать все элементы или все видимые элементы.

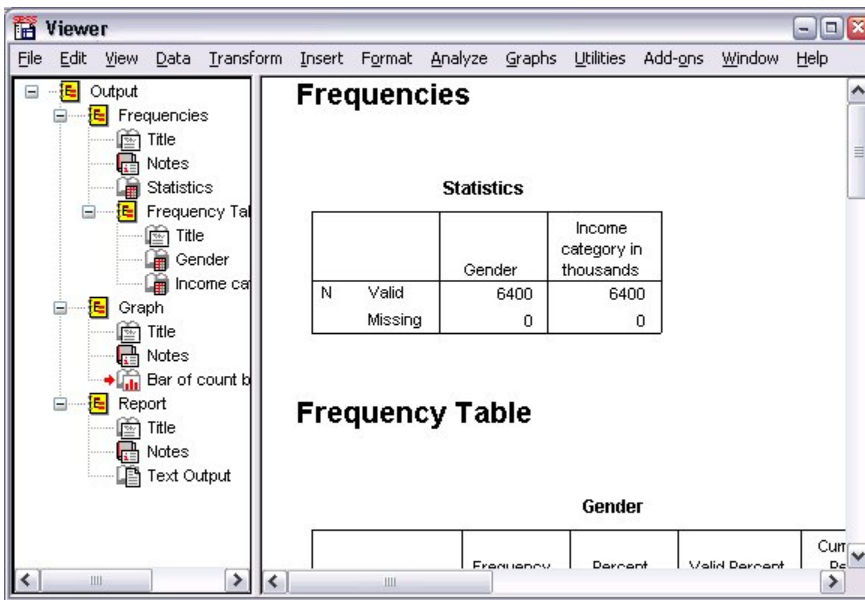


Рисунок 56. Средство просмотра

1. В меню средства просмотра:

Файл > Экспорт...

Вместо экспорта всех объектов (элементов) в окне просмотра можно экспортировать только видимые объекты (открытые книги на панели схемы) или объекты, выделенные на панели схемы. Если на панели схемы не выделен ни один объект (элемент), вариант Выделенные объекты недоступен.

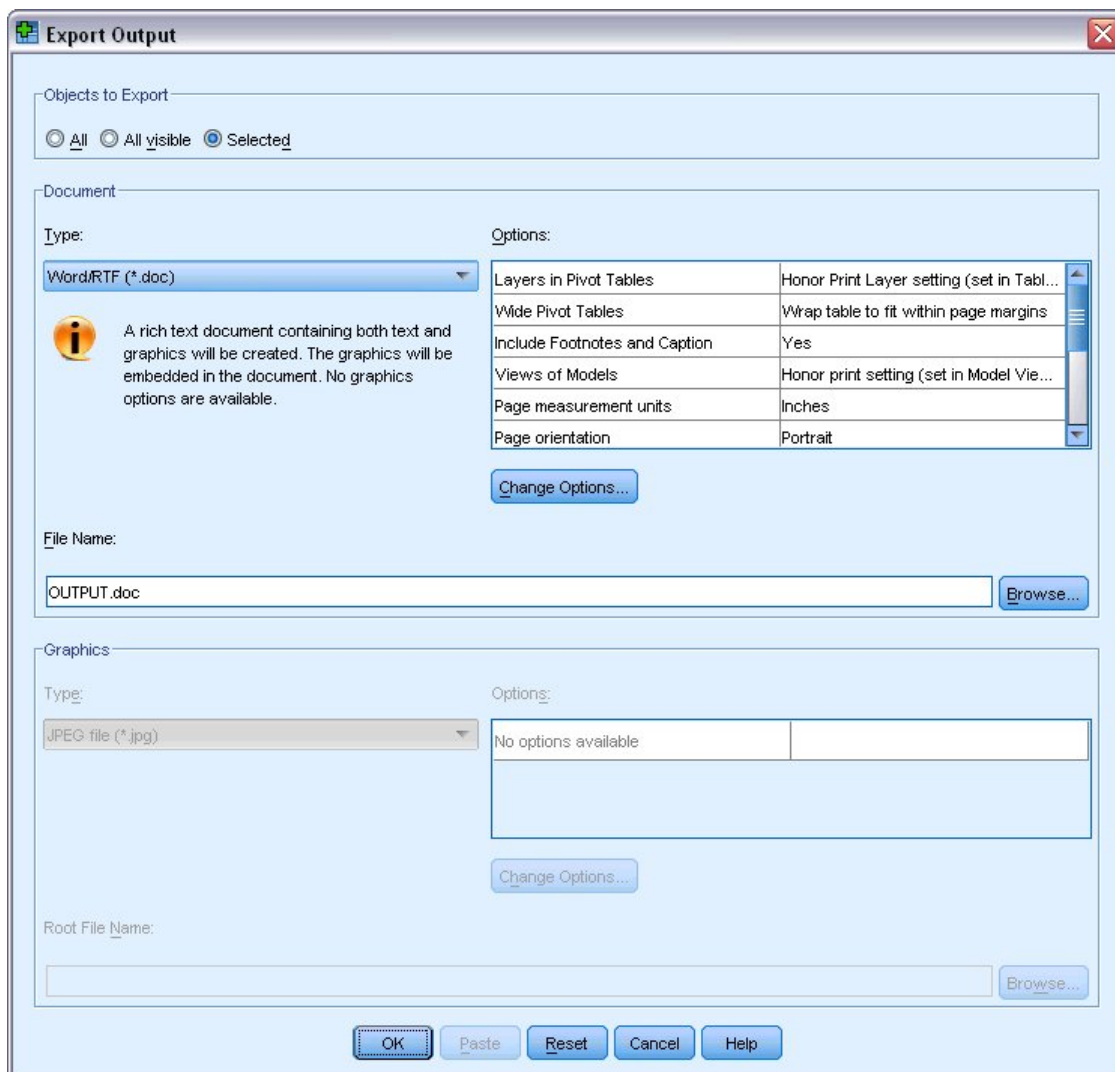


Рисунок 57. Диалоговое окно Экспортировать вывод

2. Выберите **Все** в группе Экспортируемые объекты.
3. Выберите в выпадающем списке Тип файла **Word/RTF (*.doc)**.
4. Нажмите кнопку **OK**, чтобы сгенерировать файл Word.

Открыв экспортированный файл в Word, можно увидеть, как выглядят экспортированные результаты. Замечания, не являвшиеся видимыми объектами, появляются в Word, так как мы экспортировали все объекты.

Мобильные таблицы превратились в таблицы Word, а все форматирование исходных мобильных таблиц, включая шрифты, цвета, границы и так далее, сохраняется.

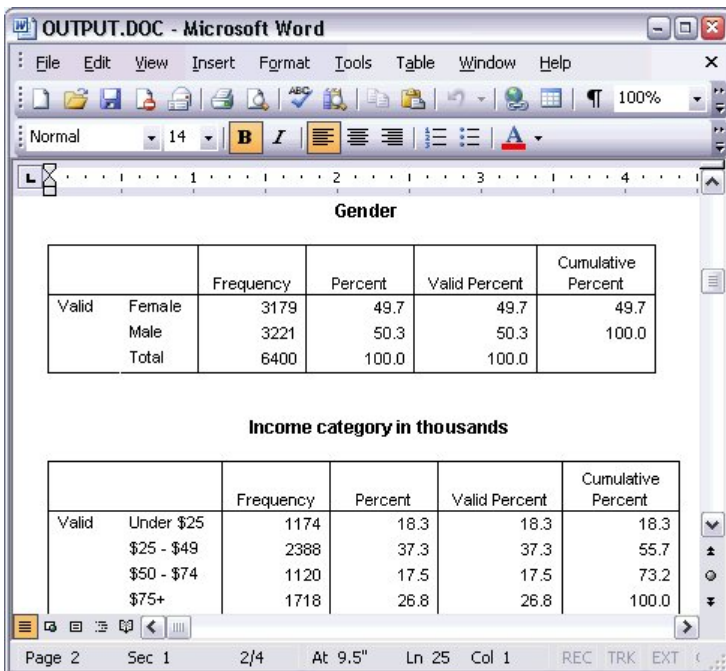


Рисунок 58. Мобильные таблицы в Word

Диаграммы вставлены в документ Word как графические объекты.

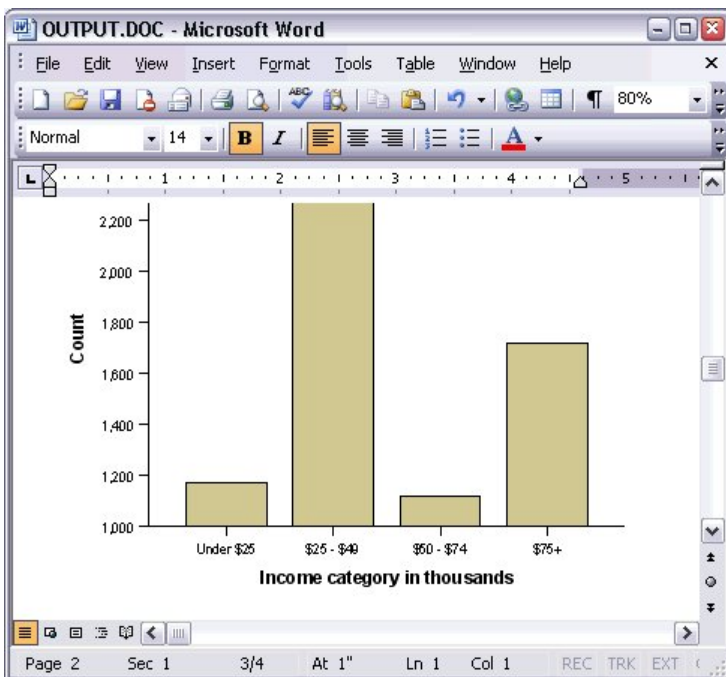


Рисунок 59. Диаграммы в Word

Текстовый вывод показан таким же шрифтом, как и текстовый объект в окне средства просмотра. Для верного формата вывода текста необходимо использовать моноширинные шрифты.

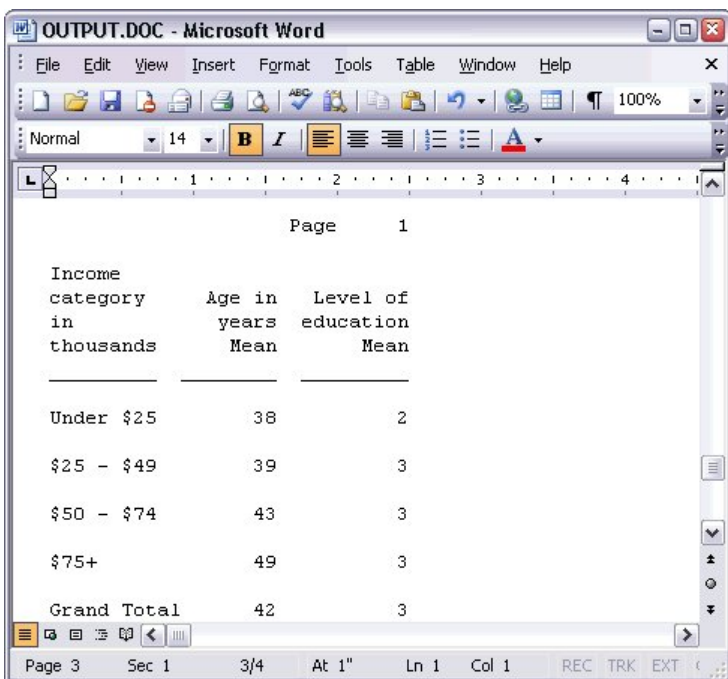


Рисунок 60. Текстовый вывод в Word

При экспорте в файл PowerPoint каждый экспортируемый объект помещается на отдельный слайд. Мобильные таблицы, экспортированные в PowerPoint превратились в таблицы Word, а все форматирование исходных мобильных таблиц, включая шрифты, цвета, границы и так далее, сохраняется.

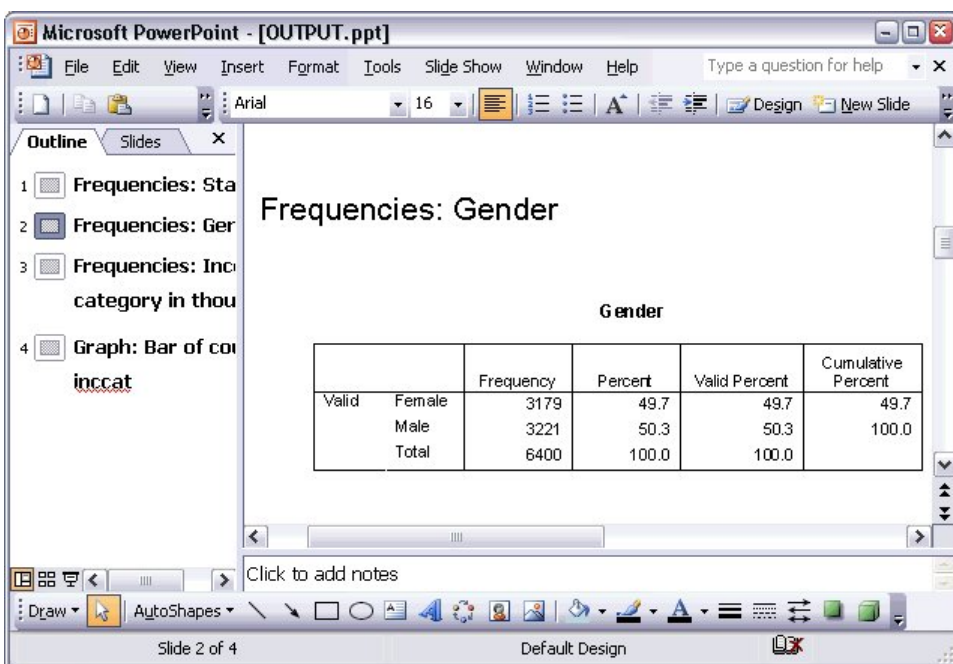


Рисунок 61. Мобильные таблицы в PowerPoint

Диаграммы, экспортируемые в PowerPoint, встраиваются в файл PowerPoint.

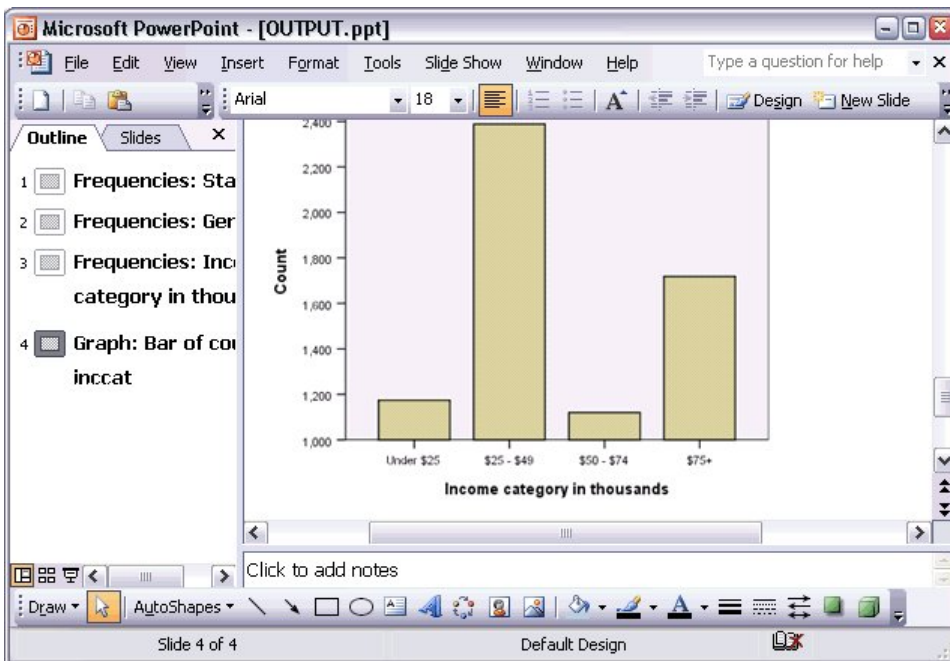


Рисунок 62. Диаграммы в PowerPoint

Примечание: Экспорт в PowerPoint доступен только в операционных системах Windows и недоступен в студенческой версии.

При экспорте в Excel результаты экспортируются по-другому.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3,179	49.7	49.7	49.7
	Male	3,221	50.3	50.3	100.0
	Total	6,400	100.0	100.0	

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under \$25	1,174	18.3	18.3	18.3
	\$25 - \$49	2,388	37.3	37.3	55.7
	\$50 - \$74	1,120	17.5	17.5	73.2
	\$75+	1,718	26.8	26.8	100.0
	Total	6,400	100.0	100.0	

Рисунок 63. Output.xls в Excel

Строки, столбцы и ячейки мобильных таблиц становятся строками, столбцами и ячейками Excel.

Каждая строка текстового вывода является строкой файла Excel, причем весь текст строки помещается в одну ячейку Excel.

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - OUTPUT.xls". The spreadsheet contains a table with the following data:

	A	B	C	D	E
68	Income				
69	category	Age in	Level of		
70	in	years	education		
71	thousands	Mean	Mean		
72					
73					
74	Under \$25	38	2		
75					
76	\$25 - \$49	39	3		
77					
78	\$50 - \$74	43	3		
79					
80	\$75+	49	3		
81					
82	Grand Total	42	3		

Рисунок 64. Текстовый вывод в Excel

Экспорт результатов в PDF

Пользователь может экспортировать все или выбранные объекты из средства просмотра в PDF-файл.

1. В меню в окне просмотра, которое содержит результат для экспорта данных в PDF, выберите:

Файл > Экспорт...

2. В диалоговом окне Выходные данные экспорта, в списке Формат файла экспорта выберите **PDF (.pdf)**.

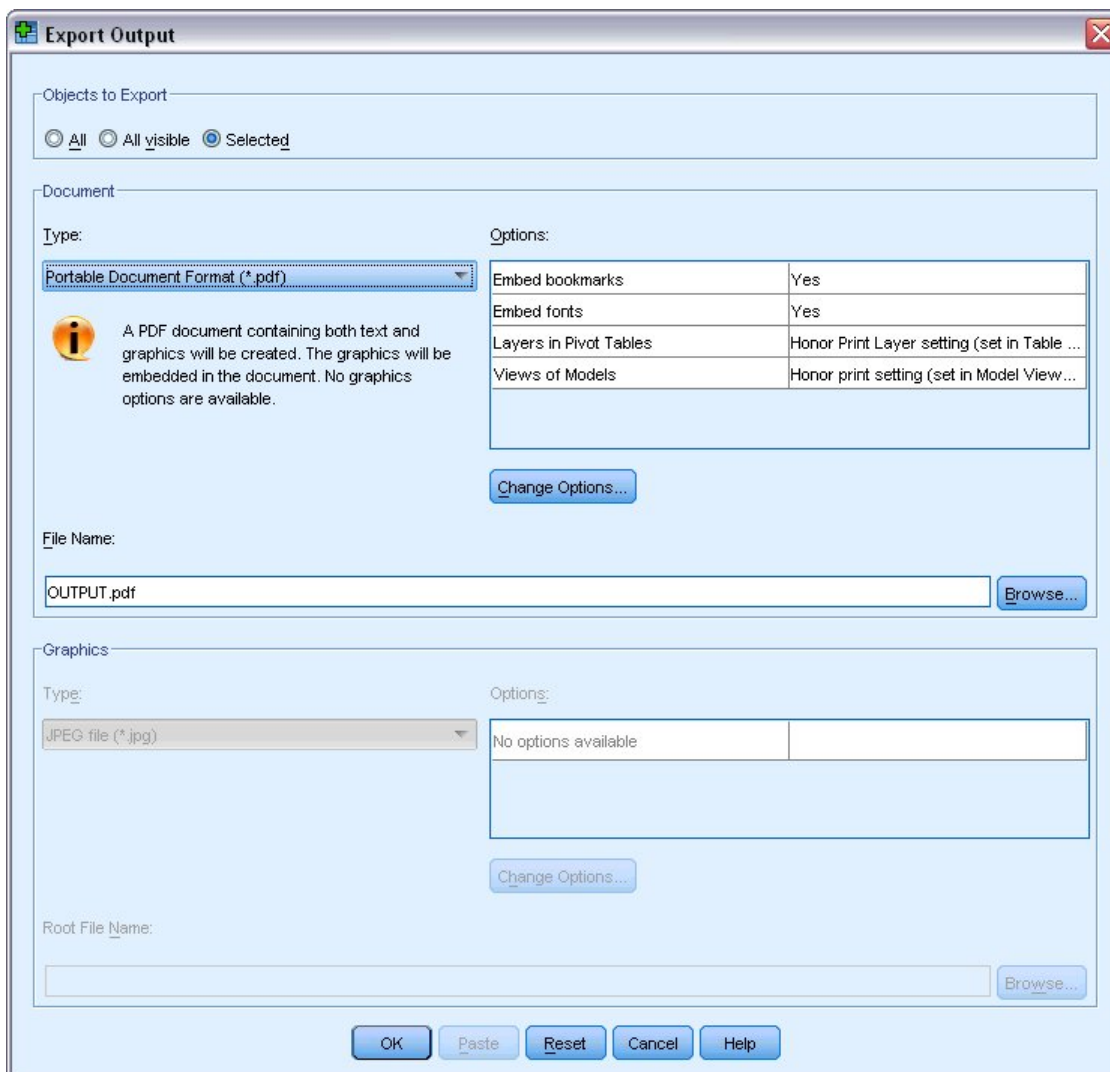


Рисунок 65. Диалоговое окно Экспортировать вывод

- Панель структуры документа в средстве просмотра конвертируется в закладки PDF-файла для упрощения навигации.
- Размер бумаги, расположение листа, размеры полей, содержимое и отображение колонтитулов, а также размер печатаемых диаграмм в документах PDF управляется параметрами страницы (меню Файл, Макет страницы окна просмотра).
- Разрешение (DPI) документа PDF принимается равным разрешению выбранного принтера или принтера по умолчанию (который можно указать в Макете страницы). Максимальное разрешение - 1200 DPI. Если в настройках принтера указано более высокое значение, разрешение документа PDF будет соответствовать 1200 DPI. *Примечание:* При печати документов высокого разрешения на принтерах с низким разрешением могут быть получены неудовлетворительные результаты.

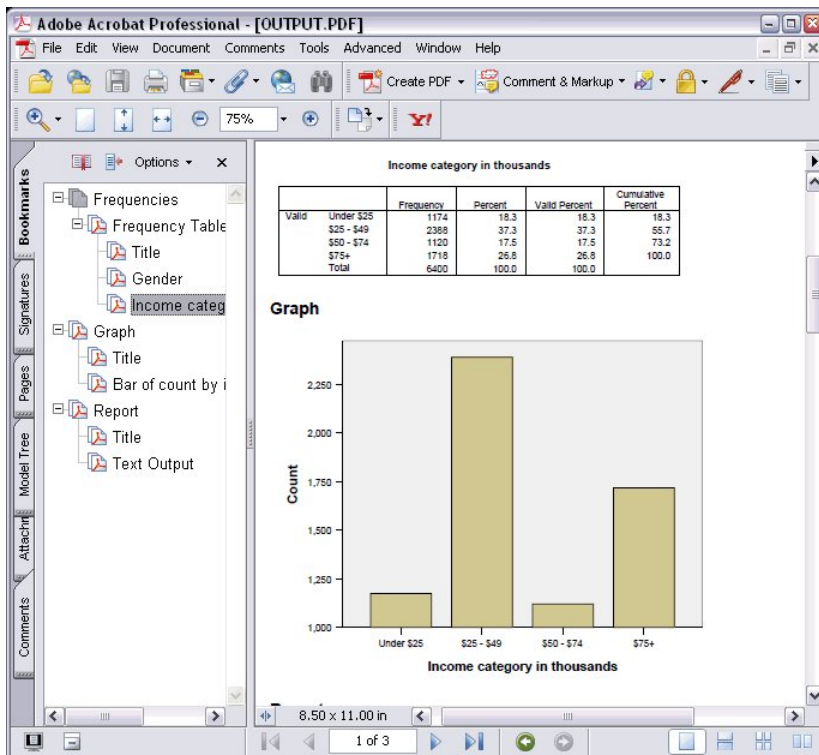


Рисунок 66. PDF-файл с закладками

Экспорт результатов в HTML

Результаты также могут быть экспортированы в формат HTML. При сохранении в формате HTML весь неграфический вывод экспортируется в единый файл HTML.

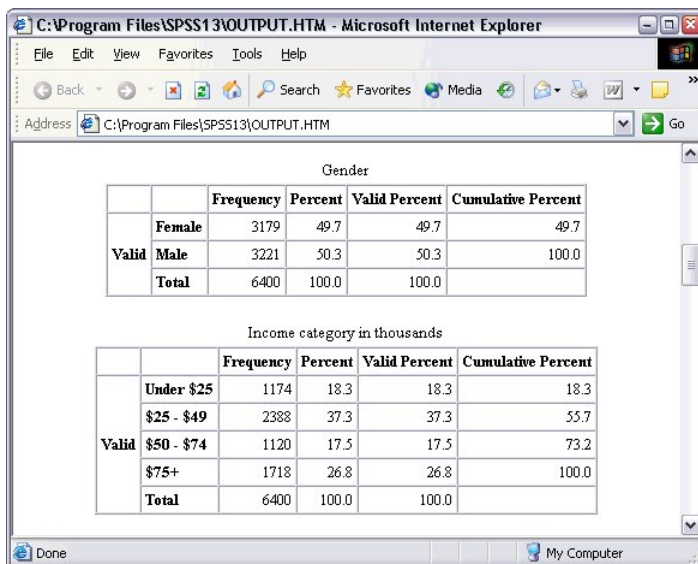


Рисунок 67. Output.htm в Обозревателе Интернет

При экспорте в формат HTML диаграмм также можно экспортировать, но не в один файл.

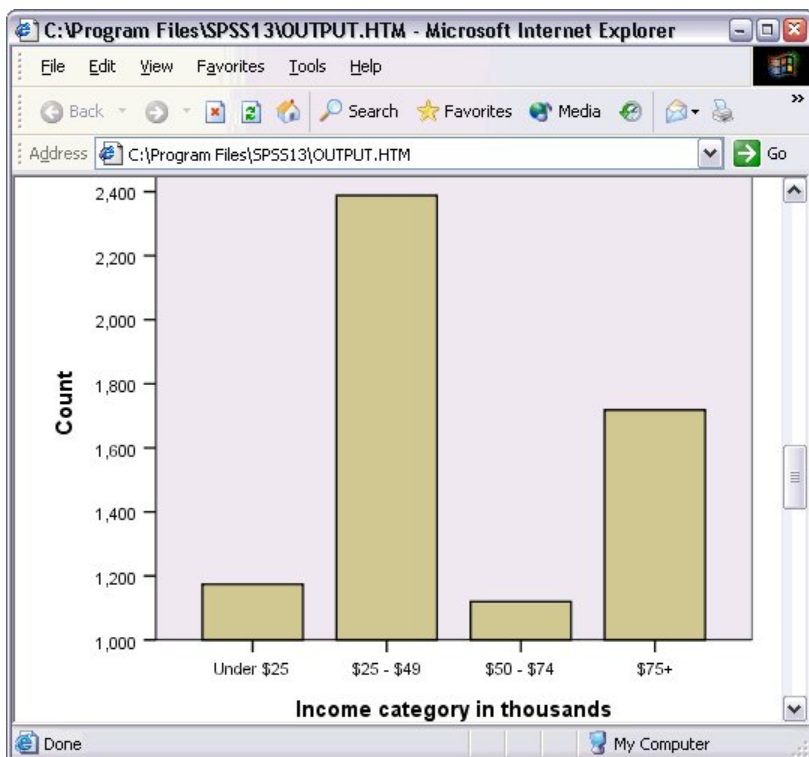


Рисунок 68. Диаграмма в HTML

Все диаграммы сохраняются в отдельных файлах в формате, который задан пользователем, а ссылки на эти файлы размещаются в сохраненные документы в формате HTML. Кроме того, можно экспортировать в отдельные файлы все диаграммы или только выделенные диаграммы.

Глава 7. Работа с синтаксисом

С помощью мощного алгоритмического языка пользователь может сохранять и автоматизировать широкий спектр часто встречающихся задач. Командный язык также позволяет пользоваться некоторыми функциональными возможностями, недоступными через меню и диалоговые окна. Большинство команд можно запустить через меню и диалоговые окна. Однако некоторые процедуры и их параметры доступны только при использовании командного языка - синтаксиса. Кроме того, язык команд позволяет пользователю сохранять задания в файле синтаксиса, что позволяет повторять анализ позднее.

Файл синтаксиса команды - это простой текстовый файл, в котором содержатся команды синтаксиса IBM SPSS Statistics. Можно открыть окно синтаксиса и ввести команды вручную, но обычно бывает проще сгенерировать команды при помощи диалоговых окон.

Для примеров в этой главе используется файл данных *demo.sav*. Дополнительную информацию смотрите в разделе Глава 10, “Файлы для примеров”, на стр. 79.

Примечание: Командный синтаксис недоступен в версии Student Version.

Вставка синтаксиса

Самый простой способ создания синтаксиса - это использование кнопки Вставка, которая есть в большинстве диалоговых окон.

1. Откройте файл данных *demo.sav*. Дополнительную информацию смотрите в разделе Глава 10, “Файлы для примеров”, на стр. 79.
2. Из меню выберите:
Анализ > Описательные статистики > Частоты...
3. Выберите *Семейное положение [marital]* и перенесите в список Переменные.
4. Нажмите кнопку **Диаграммы**.
5. В диалоговом окне Диаграммы выберите **Столбчатые**.
6. В группе параметров Значения на диаграмме выберите **Проценты**.
7. Щелкните по **Продолжить**. Нажмите кнопку **Вставка**, чтобы скопировать синтаксис, созданный в результате выбора параметров в диалоговом окне, в Редактор синтаксиса.

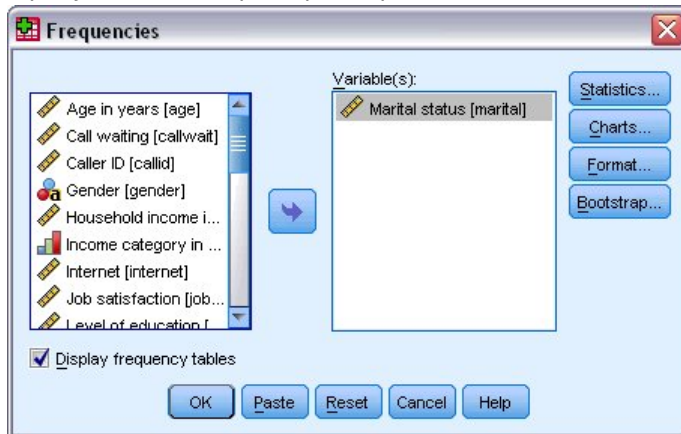


Рисунок 69. Диалоговое окно Частоты

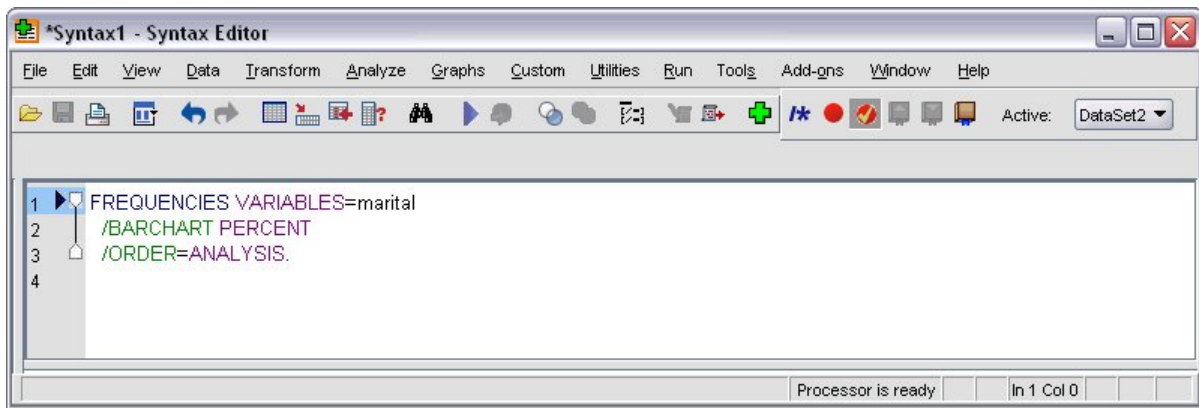


Рисунок 70. Синтаксис процедуры частоты

8. Чтобы запустить отображаемый в текущий момент синтаксис выберите в меню:

Выполнить > Выбор

Редактирование синтаксиса

В окне синтаксиса можно редактировать синтаксис. Например, можно изменить подкоманду / BARCHART, чтобы на диаграмме выводились частоты, а не проценты. (Подкоманда обозначается косой чертой (/).) Если известно ключевое слово для вывода частот, можно ввести его напрямую. Если ключевое слово неизвестно, можно получить список доступных ключевых слов для подкоманды, установив курсор после имени подкоманды и нажав клавиши Ctrl+пробел. При этом показан элемент управления автозаполнения для подкоманды.

Удалите ключевое слово PERCENT из подкоманды BARCHART.

Нажмите клавиши Ctrl-пробел.

1. щелкните по пункту с меткой **FREQ** для использования частот. При щелчке по пункту в управляющем элементе автозаполнения он будет вставлен в текущее положение курсора.
По умолчанию элемент управления автозаполнения показывает список доступных терминов по мере ввода команды. Например, необходимо включить круговую диаграмму в столбчатую диаграмму. Круговая диаграмма обозначается отдельной подкомандой.
2. Нажмите клавишу Enter после ключевого слова **FREQ** и введите косую черту (/) для обозначения начала подкоманды.

Редактор синтаксиса отображает список подкоманд для текущей команды.

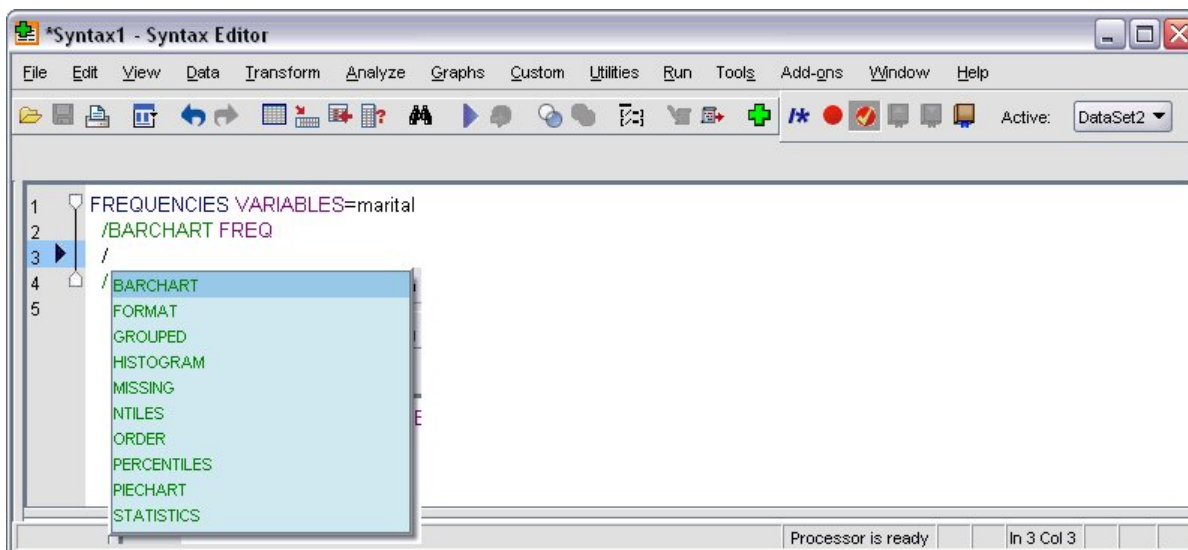


Рисунок 71. Элемент управления Автозаполнение с подкомандами

Для получения более подробной справки по текущей команде нажмите клавишу F1. При этом сразу отобразится справка по синтаксису для текущей команды.

Возможно, вы заметили, что текст в окне синтаксиса выделяется цветом. Цветовое кодирование позволяет быстро определять нераспознанные термины, поскольку цветом выделяются только распознанные термины. Например, подкоманда FORMAT по ошибке написана как FRMAT. По умолчанию подкоманды выделяются зеленым цветом, но текст FRMAT не выделен цветом, поскольку не распознан.

Открытие и запуск файла синтаксиса

1. Чтобы открыть сохраненный файл синтаксиса, выберите в меню:

Файл > Открыть > Синтаксис...

Отображается стандартное диалоговое окно открытия файлов.

2. Выберите файл синтаксиса. Если в диалоговом окне нет ни одного файла синтаксиса, убедитесь, что в качестве типа файла, который необходимо открыть, выбран **Синтаксис (*.sps)**.

3. Нажмите кнопку **Открыть**.

4. Для запуска команд используйте меню **Запуск** в окне синтаксиса.

Если команды применяются к определенному файлу данных, перед запуском команд необходимо открыть файл данных или вставить в синтаксис команду, которая открывает файл данных. Команду такого типа можно вставить из диалоговых окон, предназначенных для открытия файлов данных.

Использование контрольных точек

С помощью контрольных точек можно останавливать выполнение синтаксиса команды на указанных точках в окне синтаксиса и продолжить выполнение при готовности. Это позволяет просматривать вывод или данные в промежуточной точке в задании синтаксиса или выполнять синтаксис команды, который отображает информацию о текущем состоянии данных, например FREQUENCIES. Контрольные точки можно установить только на уровне команды, а не на конкретных строках внутри команды.

Чтобы вставить контрольную точку в команду:

1. Щелкните в любом месте области слева от текста, связанного с командой.

Вне зависимости от того, в каком месте области был выполнен щелчок, контрольная точка представлена красным кружком в области слева от текста команды и в той же строке, где находится имя команды.

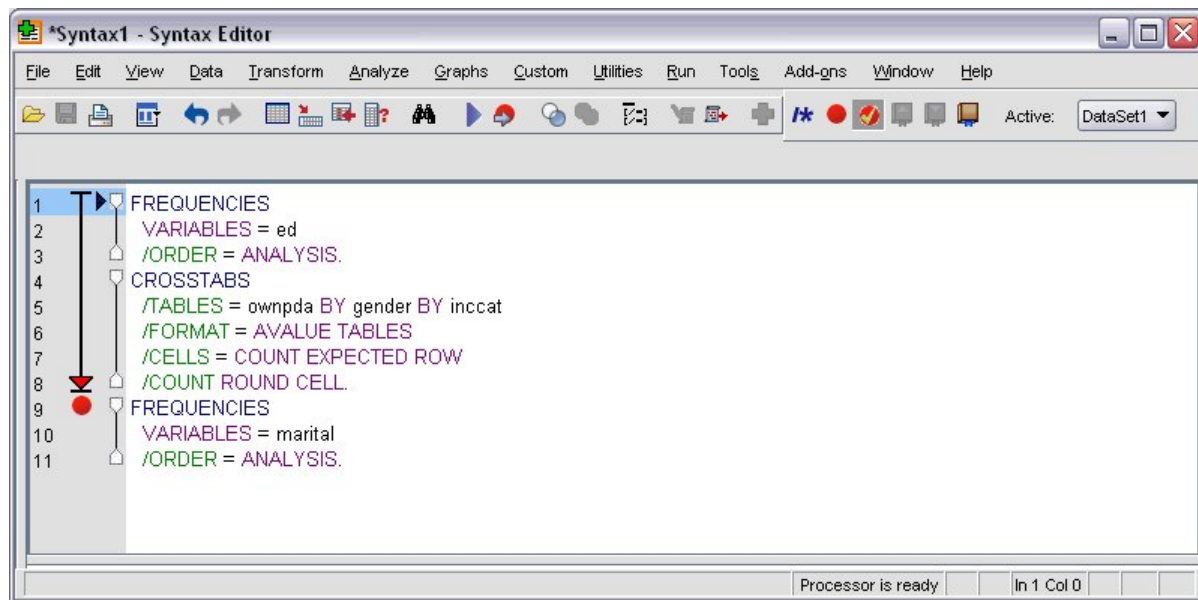


Рисунок 72. Выполнение прервано на контрольной точке

При прогоне синтаксиса команды, содержащего контрольные точки, выполнение прекращается перед каждой командой с контрольной точкой.

Стрелка, указывающая вниз и расположенная слева от текста команды, обозначает ход выполнения синтаксиса. Она охватывает область от прогона первой команды до прогона последней команды и особенно полезна при выполнении синтаксиса команды с контрольными точками.

Чтобы возобновить выполнение после контрольной точки:

2. Выберите в меню Редактора синтаксиса:

Выполнить > Продолжить

Глава 8. Изменение значений данных

Данные, с которыми вам придется работать, не всегда удобно организованы для проведения анализа и создания отчетов. Например, вам может понадобиться:

- Создать категориальную переменную из количественной.
- Свести несколько категорий ответов в одну категорию.
- Создать новую переменную, в которую будут записаны значения разностей между двумя существующими переменными.
- Вычислить количество времени, прошедшее между двумя датами

В этой главе используется файл данных *demo.sav*. Дополнительную информацию смотрите в разделе [Глава 10, “Файлы для примеров”](#), на стр. 79.

Создание категориальной переменной из количественной

Несколько категориальных переменных в файле данных *demo.sav*, на самом деле, были вычислены на основе нескольких количественных переменных этого файла. Например, переменная *Категория дохода домохозяйства [inccat]*, на самом деле, является сгруппированной в четыре категории переменной *Доход домохозяйства в тысячах [income]*. В категориальной переменной целые числа 1-4 используются для представления следующих категорий дохода (в тысячах долларов): менее 25 тысяч долларов, 25 – 49 тысяч долларов, 50 – 74 тысяч долларов и 75 или более тысяч долларов.

Для создания категориальной переменной *Категория дохода домохозяйства [inccat]*:

1. Выберите в меню в окне Редактора данных:

Преобразовать > Визуальная категоризация...

В первом окне процедуры Визуальная категоризация выберите количественные и/или порядковые переменные, для которых будут создаваться новые категориальные переменные. Термин **категоризация** означает группировку двух или более последовательных значений и одно значение (категию).

Процедура Визуальная категоризация отталкивается от реальных значений в файле данных, поэтому сначала ей необходимо прочесть данные. Поскольку чтение данных может занять много времени, если в файле данных большое число наблюдений, первое окно процедуры позволяет ограничить количество считываемых (сканируемых) наблюдений. В нашем случае устанавливать такое ограничение не имеет смысла, так как объем данных - невелик. Даже несмотря на то, что в нашем файле данных более 6000 наблюдений, сканирование файла такого размера не займет много времени.

2. Перенесите переменную *Доход домохозяйства в тысячах [income]* из списка Переменные в список Переменные для категоризации, а затем нажмите кнопку **Продолжить**.

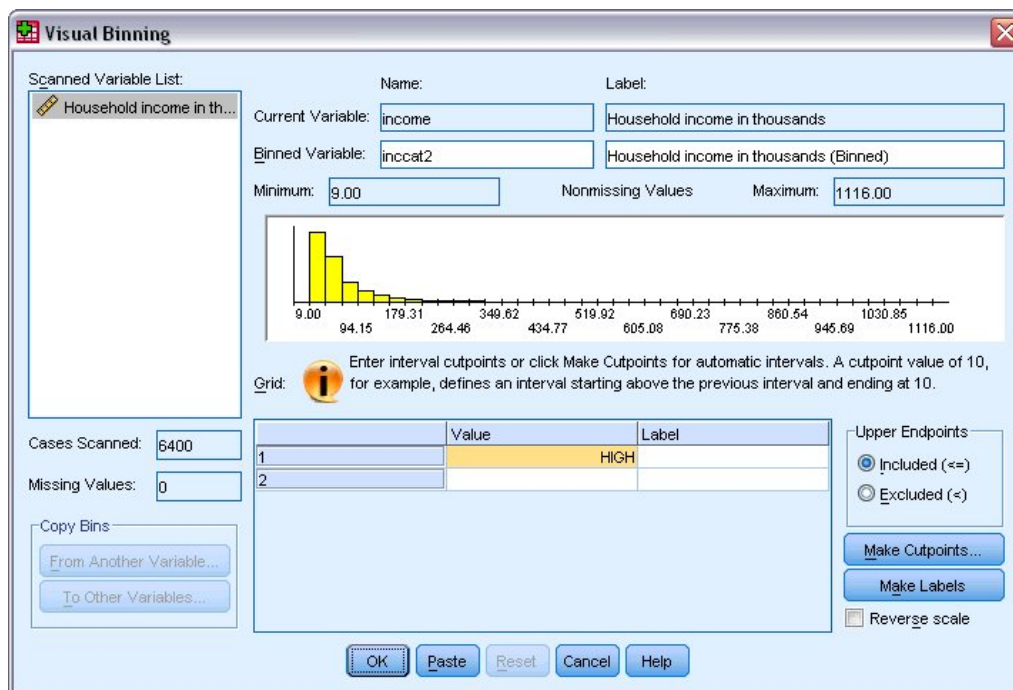


Рисунок 73. Главное окно визуальной категоризации

3. В появившемся диалоговом окне Визуальная категоризация выделите *Доход домохозяйства в тысячах [income]* в списке Просмотренные переменные.

Распределение выбранной переменной выводится в виде гистограммы (и в нашем случае оно несимметричное).

4. Введите `incscat2` в качестве имени новой (категоризованной) переменной: и Категория дохода домохозяйства в качестве метки переменной.
5. Нажмите кнопку **Границы интервалов**.
6. Выберите **Равные интервалы**.
7. Введите 25 в поле Местоположение первой границы, 3 в поле Количество границ и 25 в поле Ширина.

Количество категорий в новой переменной равно количеству границ плюс один. В нашем примере, у новой категоризованной переменной будет четыре категории. Диапазоны первых трех категорий будут равны 25 (тысячам), а последняя категория будет включать все значения выше наибольшей границы (75).

8. Нажмите кнопку **Применить**.

Значения на панели Сетка - это заданные границы, точнее верхние границы категорий. Вертикальные линии на гистограмме также показывают местоположение границ.

По умолчанию, сами граничные значения включаются в соответствующие категории. Например, первый интервал (Значение = 25) будет включать все значения меньше либо равные 25. Но в нашем примере мы хотим получить следующие категории: меньше 25, 25-49, 50-74 и 75 или более.

9. Выберите в группе параметров Верхние границы **Исключены (<)**.
10. Затем нажмите кнопку **Создать метки**.

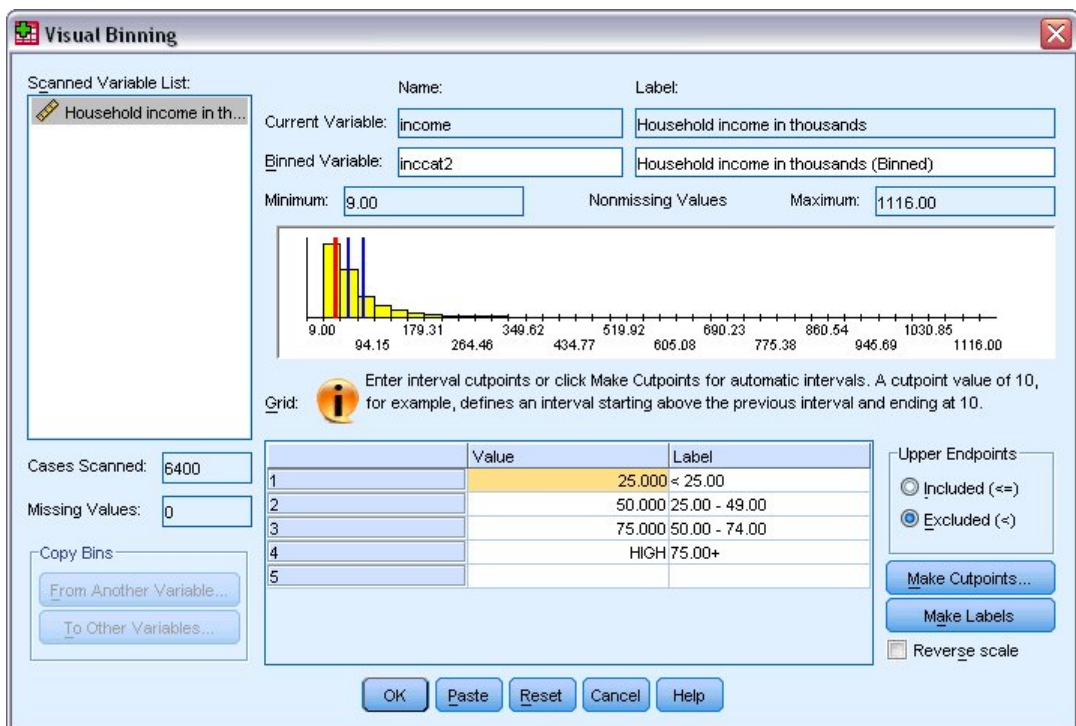


Рисунок 74. Автоматически сгенерированные метки значений

Описательные метки значений создаются автоматически для каждой категории. Поскольку значения, присвоенные новой категоризованной переменной, являются последовательными целыми числами, начинающимися с 1, метки значений могут оказаться очень полезными.

Вы также можете задавать и изменять границы и метки вручную в сетке, изменять расположение границ, перетаскивая линии границ в гистограмме, а также стирать границы, перетаскивая линии границ за пределы гистограммы.

11. Нажмите **ОК**, чтобы создать новую категоризованную переменную.

В Редакторе данных появилась новая переменная. Поскольку новые переменные добавляются в конец файла, она находится в крайнем правом столбце в закладке Данные и в последней строке в закладке Переменные.

Вычисление новых переменных

Вычислять новые переменные можно, используя широкий спектр математических функций (даже очень сложные формулы). В нашем примере, мы произведем простое вычисление новой переменной, вычтя значения одной переменной из значений другой.

В файле данных *demo.sav* есть переменная Возраст [age] и переменная Количество лет на текущем месте работы [employ]. А вот переменной, содержащей возраст во время поступления на последнее место работы, в нашем файле нет. Мы можем создать новую переменную, в которой будет вычислена разность между возрастом в настоящее время и количеством лет на текущем месте работы, то есть примерный возраст в момент поступления на текущее место работы.

1. Выберите в меню в окне Редактора данных:

Преобразовать > Вычислить переменные...

2. Введите в поле Вычисляемая переменная `jobstart`.
3. Выберите в списке исходных переменных *Возраст [age]* и нажмите кнопку со стрелкой вправо, чтобы скопировать переменную в поле Числовое выражение.
4. Нажмите кнопку (-) минус на клавиатуре калькулятора в диалоговом окне (или клавишу минус на клавиатуре компьютера).

5. Выберите *Количество лет на текущем месте работы [employ]* и нажмите кнопку в виде стрелки вправо, чтобы скопировать переменную в поле числового выражения.

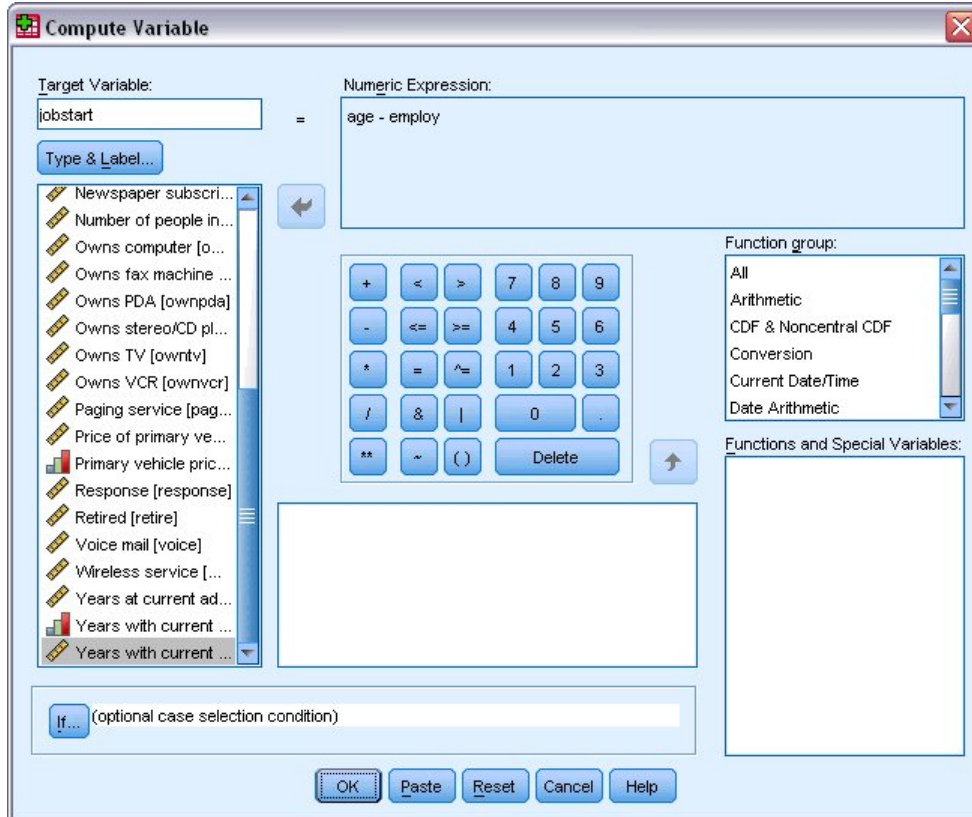


Рисунок 75. Диалоговое окно *Вычислить переменную*

Примечание: Убедитесь, что вы выбрали правильную переменную. Помимо переменной *Количество лет на текущем месте работы [employ]* в данных есть еще и (*не требующая нам в данном случае*) перекодированная категориальная переменная *Категория количества лет на текущем месте работы [empcat]*. Числовое выражение должно быть *age-employ*, а не *age-empcat*.

6. Нажмите кнопку **OK**, чтобы вычислить новую переменную.

В Редакторе данных появилась новая переменная. Поскольку новые переменные добавляются в конец файла, она находится в крайнем правом столбце в закладке *Данные* и в последней строке в закладке *Переменные*.

Использование функций

Для числовых выражений можно также использовать встроенные функции. Доступно более 70 встроенных функций, включая:

- Арифметические функции
- Статистические функции
- Функции распределения
- Логические функции
- Функции агрегации и извлечения данных и времени
- Функции для работы с пропущенными значениями
- Функции для работы с несколькими наблюдениями
- Функции для работы с текстовыми значениями

Функции разбиты на группы по своему назначению. К примеру, есть группа арифметических функций или группа статистических функций. Для удобства в группы функций также включено несколько часто используемых системных переменных, таких как *\$TIME* (текущая дата и время).

Вставка функции в выражение

Для того чтобы вставить функцию в выражение:

1. Установите курсор в поле Числовое выражение, в то место, куда нужно вставить функцию.
2. Выберите в списке Группы функций подходящую группу. В группе **Все** представлены все доступные функции и системные переменные.
3. В списке Функции и специальные значения дважды щелкните по функции (или выберите функцию и щелкните по стрелке рядом со списком Группы функций).

Функция вставляется в выражение. Если вы выделите часть выражения, а затем вставите функцию, выделенная часть выражения станет первым аргументом функции.

Редактирование функции в поле Выражение

Функция не является полной, пока не будут введены аргументы, представленные во вставленной функции знаками вопросов. Количество знаков вопроса указывает на минимальное количество аргументов, которые требуются, чтобы сделать функцию полной.

1. Выделите знак(и) вопроса во вставленной функции.
2. Введите аргументы. Если аргументами являются имена переменных, их можно вставить из списка переменных.

Использование условных выражений (условий).

Для применения преобразований к подмножеству наблюдений используются условные выражения (также называемые логическими выражениями). Условное выражение возвращает для каждого наблюдения значение истина, ложь или пропущенное. Если результат условного выражения - истина, преобразование применяется к этому наблюдению. Если результат - ложь или пропущенное значение, преобразование не применяется к этому наблюдению.

Чтобы задать условное выражение:

1. Нажмите кнопку **Если** в диалоговом окне Вычислить переменную. Открывается диалоговое окно Отбор наблюдений.
2. Выберите **Включить наблюдения, удовлетворяющие условию**.
3. Введите выражение условия.

В большинстве условных выражений содержится хотя бы один оператор сравнения, как, например, в:

```
age>=21
```

или

```
income*3<100
```

В первом примере выбраны только наблюдения со значением 21 или выше для *Возраст [age]*. Во втором примере *Доход домохозяйства в тысячах долларов [income]*, умноженный на 3, должен быть меньше 100, чтобы наблюдение отбиралось.

Также можно связать несколько условных выражений, используя логические операторы, как, например, в:

```
age>=21 | ed>=4
```

или

```
income*3<100 & ed= 5
```

В первом примере отбираются наблюдения, удовлетворяющие либо условию *Возраст [age]*, либо условию *Уровень образования [ed]*. Во втором примере для выбора наблюдения должны

выполняться оба условия - *Доход домохозяйства в тысячах долларов[income]*, и *Уровень образования [ed]*.

Работа с датами и временем

При помощи Конструктора дат и времени можно выполнять целый ряд задач, часто встречающихся при работе с датами и временем. При помощи Конструктора можно:

- Создавать переменные дат/времени из текстовых переменных, содержащих даты и время.
- Конструировать переменные дат/времени, объединяя переменные, в которых содержатся различные составляющие дат и времени.
- Складывать и вычитать значения переменных дат/времени, включая сложение и вычитание двух переменных дат/времени.
- Извлекать части переменных дат/времени; например, день месяца из переменной дат/времени, имеющей формат мм/дд/гггг.

В качестве файла данных для примеров в этом разделе используется *upgrade.sav*. Дополнительную информацию смотрите в разделе Глава 10, "Файлы для примеров", на стр. 79.

Чтобы открыть Конструктор дат и времени

1. Выберите в меню:

Преобразовать > Мастер дат и времени...



Рисунок 76. Начальный экран мастера дат и времени

Первое диалоговое окно Конструктора дат и времени позволяет выбрать задачу, которую Вы собираетесь решать. Задачи, решение которых невозможно для текущего набора данных, неактивны. К примеру, в файле данных *upgrade.sav* нет текстовых переменных, поэтому задача "Создать переменную дат/времени из текстовой переменной, содержащей даты или время" - неактивна.

Если вы недостаточно знакомы с датами и временем в IBM SPSS Statistics, можно выбрать **Узнать, как представляются значения даты и времени** и щелкнуть по **Далее**. На экране появится краткий обзор переменных дат/времени и ссылка (в виде кнопки Справка) на более подробную информацию.

Вычисление продолжительности времени между двумя датами

Одна из наиболее часто встречающихся при работе с датами задач состоит в вычислении продолжительности времени между двумя датами. В качестве примера рассмотрим компанию разработчика программного обеспечения, которая хочет проанализировать покупки обновлений лицензий, для чего необходимо вычислить, сколько лет назад каждый из клиентов покупал обновление. В файле данных *upgrade.sav* для каждого клиента есть переменная даты последней покупки обновления лицензии, но нет переменной количества лет, прошедших со времени этой покупки. Такую переменную мы и создадим.

Чтобы вычислить продолжительность времени между двумя датами:

1. В первом диалоговом окне Мастера дат/времени выберите **Провести вычисления с датами и временем** и нажмите кнопку **Далее**.
2. Выберите **Вычислить продолжительность времени, прошедшего между двумя датами** и нажмите кнопку **Далее**.

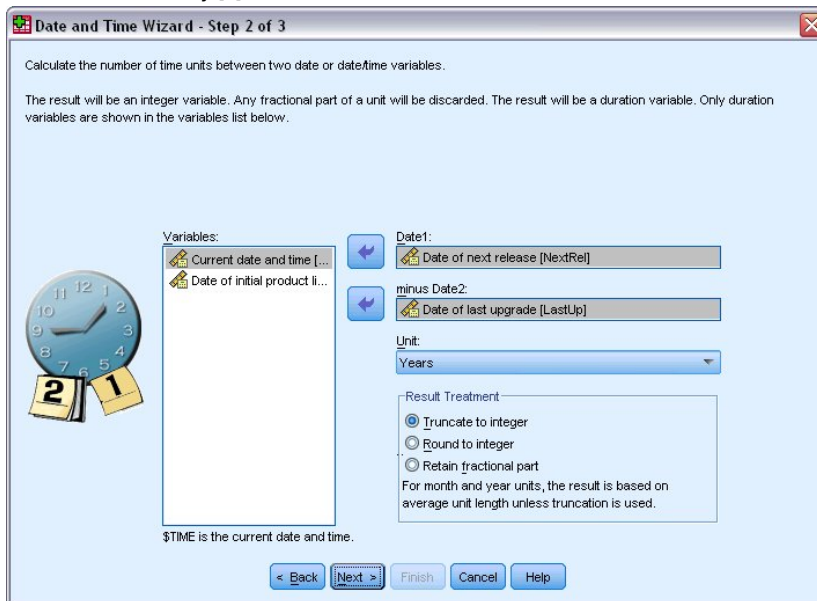


Рисунок 77. Вычисление продолжительности времени, прошедшего между двумя датами: Шаг 2

3. На шаге 2 выберите *Дата следующего выпуска* для *Даты1*.
4. Перенесите переменную *Дата покупки последнего обновления [LastUp]* в поле *Минус Дата2*.
5. Выберите **Годы** в качестве единиц измерения и **Усечение до целого** в качестве значения поля *Обработка результата*. (Это значения по умолчанию.)
6. Нажмите кнопку **Далее**.
7. На шаге 3 введите *YearsLastUp* в качестве имени переменной результата. Имена выходных переменных должны отличаться от имен существующих переменных.
8. Задайте метку выходной переменной - *Количество лет, прошедшее со времени покупки последнего обновления*. Метки для выходных переменных можно не задавать.
9. Оставьте выбранный по умолчанию вариант **Создать переменную сейчас** и нажмите кнопку **Готово**, чтобы создать новую переменную.

В редакторе данных появилась новая переменная *ЛетПослеОбновления*, представляющая целочисленное количество лет между двумя датами. Дробные части лет были урезаны.

Добавление продолжительности времени к дате

Продолжительности времени (например 10 дней или 7 месяцев) можно складывать и вычитать. Продолжим рассмотрение примера компании разработчика программного обеспечения и поставим задачу вычислить дату окончания первоначального контракта на техническую

поддержку. В файле данных *upgrade.sav* есть переменная Количество лет технической поддержки [Support] и переменная Дата первоначальной покупки лицензии [PurDate]. С их помощью можно вычислить дату окончания первоначального контракта на техническую поддержку путем сложения количества лет технической поддержки и дате первоначальной покупки лицензии.

Чтобы добавить продолжительность к дате:

1. В первом диалоговом окне Мастера дат/времени выберите **Провести вычисления с датами и временем** и нажмите кнопку **Далее**.
2. На шаге 1 выберите **Добавить или вычесть продолжительность из даты** и нажмите кнопку **Далее**.

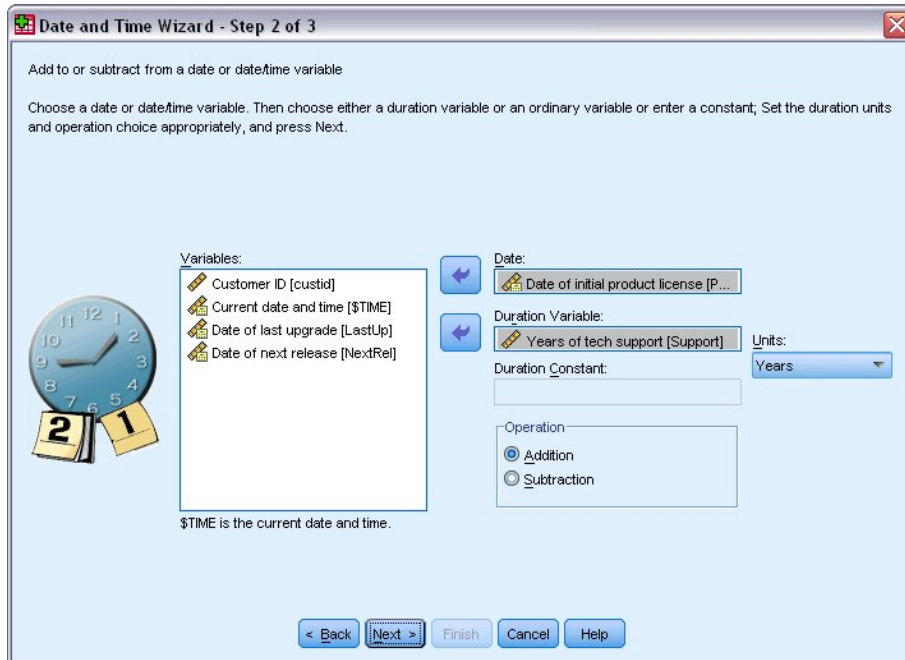


Рисунок 78. Добавление продолжительности к дате: Шаг 2

3. Перенесите переменную *Дата первоначальной покупки лицензии [PurDate]* в поле *Дата*.
 4. На шаге 2 выберите *Срок технической поддержки в годах* для переменной продолжительности. Поскольку *Срок технической поддержки в годах* - это простая числовая переменная, необходимо указать единицу измерения, которая будет использоваться в операции сложения этой переменной.
 5. В выпадающем списке *Единицы измерения* выберите **Годы**.
 6. Нажмите кнопку **Далее**.
 7. На шаге 3 введите *SupEndDate* в качестве имени переменной результата. Имена выходных переменных должны отличаться от имен существующих переменных.
 8. Задайте метку выходной переменной - *Дата окончания контракта на техническую поддержку*. Метки для выходных переменных можно не задавать.
 9. Нажмите кнопку **Готово**, чтобы создать новую переменную.
- В Редакторе данных появилась новая переменная.

Глава 9. Сортировка и выбор данных

Файлы данных не всегда не всегда имеют структуру, наилучшим образом подходящую для решения задач, стоящих перед вами. При подготовке данных для анализа пользователь может использовать разнообразные преобразования, в том числе:

- **Сортировка данных.** Вы можете отсортировать данные на основе значений одной или более переменных.
- **Отбор подмножеств наблюдений.** Вы можете ограничить анализ подмножеством наблюдений или выполнить анализ для разных подмножеств.

Для примеров в этой главе используется файл данных *demo.sav*. Дополнительную информацию смотрите в разделе [Глава 10, “Файлы для примеров”](#), на стр. 79.

Сортировка данных

Сортировка наблюдений (сортировка строк в файле данных) зачастую бывает полезна, а иногда просто необходима для определенных типов анализа.

Для того чтобы задать одну или несколько переменных, по значениям которых упорядочиваются наблюдения в файле данных:

1. Из меню выберите:

Данные > Сортировать наблюдения...

Открывается диалоговое окно Сортировка наблюдений.

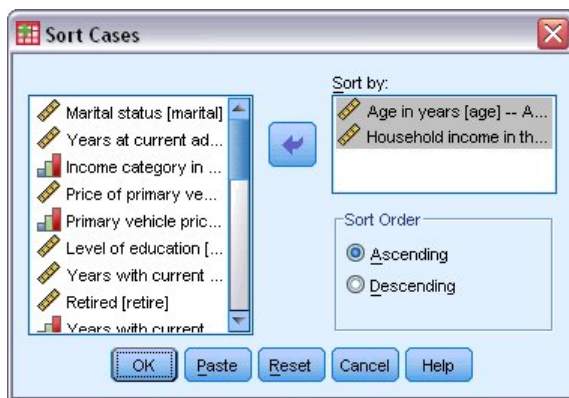


Рисунок 79. Диалоговое окно Сортировка наблюдений

2. Добавьте переменные *Возраст [age]* и *Доход домохозяйства в тысячах долларов [income]* в список Сортировать по.

Если добавлено несколько переменных для сортировки, порядок их расположения в списке Сортировать по определяет порядок сортировки наблюдений. Например, наблюдения сортируются по значениям *Доход домохозяйства в тысячах [income]* в категориях *Возраст [age]* на основе расположения этих пунктов в списке Сортировать по. Текстовые значения сортируются так, что заглавные буквы располагаются раньше строчных (например, текст *Да* располагается перед текстом *да* при сортировке).

Анализ с расщеплением файла

Для расщепления файла на отдельные группы выполните следующие действия:

1. Из меню выберите:

Данные > Расщепить файл...

Открывается диалоговое окно Расщепить файл.

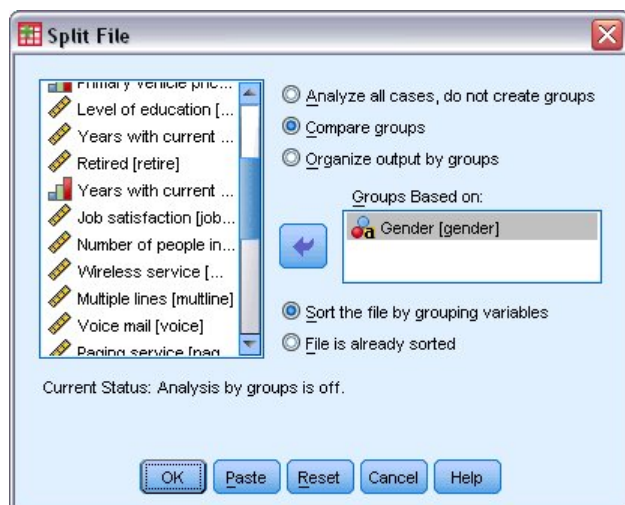


Рисунок 80. Диалоговое окно Расщепить файл

2. Выберите **Сравнить группы** или **Организовать вывод по группам**. (В примерах ниже иллюстрируется разница между двумя этими пунктами.)
3. Выберите *Пол [gender]* для расщепления файла на отдельные группы по значению этих переменных.

В качестве группирующих переменных можно использовать числовые, короткие текстовые и длинные текстовые переменные. Для каждой из подгрупп, определенных значениями группирующих переменных, выполняется отдельный анализ. Если добавлено несколько переменных для сортировки, порядок их расположения в списке Группы образуются по определяет порядок группировки наблюдений.

Если выбран пункт **Сравнить группы**, результаты каждой группы расщепления включаются в подобную таблицу или таблицы, как показано в созданной процедурой Частоты таблице итоговой статистики ниже.

Statistics

Household income in thousands

Female	N	Valid	3179
		Missing	0
	Mean		68.7798
	Median		44.0000
	Std. Deviation		75.73510
Male	N	Valid	3221
		Missing	0
	Mean		70.1608
	Median		45.0000
	Std. Deviation		81.56216

Рисунок 81. Вывод расщепленного файла с единой мобильной таблицей

Если выбрать опцию **Организовать вывод по группам** и запустить процедуру Частоты, будут созданы две сводные таблицы: одна таблица для женщин, а другая - для мужчин.

Statistics^a

Household income in thousands

N	Valid	3179
	Missing	0
Mean		68.7798
Median		44.0000
Std. Deviation		75.73510

a. Gender = Female

Рисунок 82. Вывод расщепленного файла с мобильной таблицей для женского пола

Statistics^a

Household income in thousands

N	Valid	3221
	Missing	0
Mean		70.1608
Median		45.0000
Std. Deviation		81.56216

a. Gender = Male

Рисунок 83. Вывод расщепленного файла с мобильной таблицей для мужского пола

Сортировка наблюдений при анализе с расщеплением файла

Процедура расщепления файла создает новую подгруппу для каждого изменения значения одной из группирующих переменных. Таким образом, важно выполнять сортировку по значениям группирующих переменных перед анализом с расщеплением файла.

По умолчанию процедура расщепления файла автоматически сортирует данные на основе значений группирующих переменных. Если файл уже отсортирован надлежащим образом, можно сократить время обработки, если выбрать пункт **Файл уже отсортирован**.

Включение и отключение анализа с расщеплением файла

После активации анализа с расщеплением файла эта функция остается активной до конца сеанса, пока не будет отключена пользователем.

- **Анализ всех наблюдений.** Этот пункт выключает анализ с расщеплением файла.
- **Сравнить группы и Организовать вывод по группам.** Этот пункт включает анализ с расщеплением файла.

Если анализ с расщеплением файла включен, в строке состояния в нижней части окна приложения возникает сообщение **Расщепление файла**.

Отбор подмножеств наблюдений

Пользователь может ограничить область анализа указанной подгруппой на основе заданных критериев отбора по значениям переменных и сложных выражений. Можно также отобразить случайную подвыборку наблюдений. Критерии, используемые для отбора наблюдений:

- Значения и диапазоны значений переменных
- Диапазоны дат и времени
- Номера наблюдений (строк)
- Арифметические выражения
- Логические выражения
- Функции

Для отбора подмножества наблюдений для анализа:

1. Выберите в меню:

Данные > Отобразить наблюдения...

Открывается диалоговое окно Отбор наблюдений.

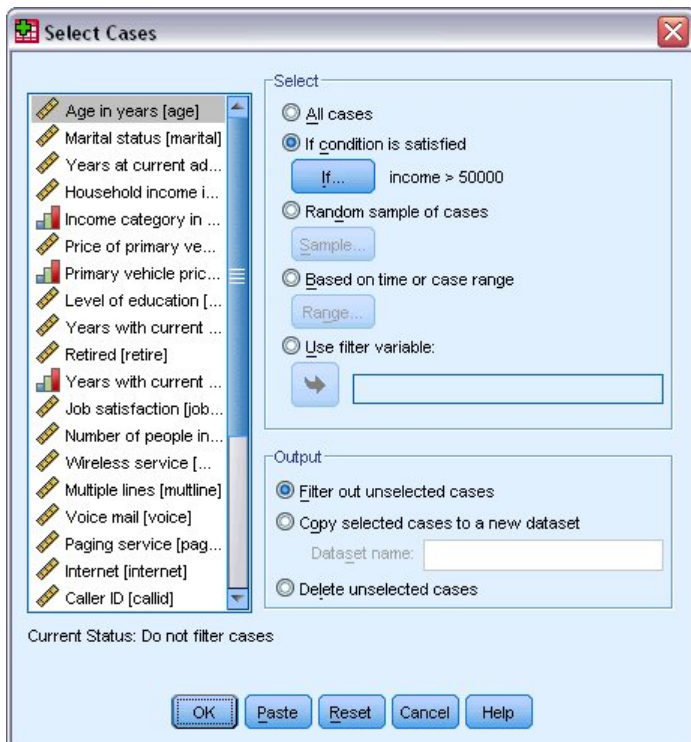


Рисунок 84. Диалоговое окно Отбор наблюдений

Отбор наблюдений на основе условия

Для отбора наблюдений по значению логического выражения:

1. Выберите пункт **Если выполнено условие** и щелкните **Если** в диалоговом окне Отбор наблюдений.

Открывается диалоговое окно Отобразить наблюдения: Условие.

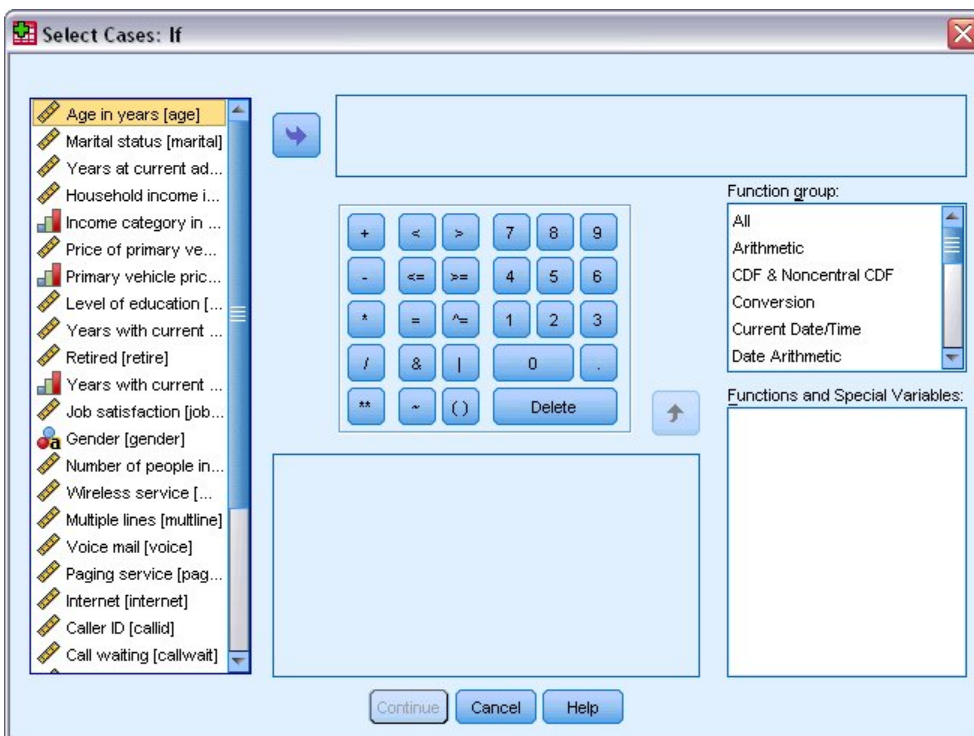


Рисунок 85. Диалоговое окно Отобразить наблюдения: Условие

В логическом выражении могут использоваться имена существующих переменных, константы, арифметические, реляционные и логические операторы, а также функции. Выражение можно вводить и редактировать в текстовом поле как обычный текст в окне вывода. Кроме того, при вводе элементов выражения можно использовать панель калькулятора, список переменных и список функций. Дополнительную информацию смотрите в разделе [“Использование условных выражений \(условий\).”](#) на стр. 67.

Отбор случайной подвыборки

Для получения случайной подвыборки:

1. Выберите пункт **Случайная подвыборка** в диалоговом окне Отобразить наблюдения.
2. Нажмите кнопку **Подвыборка**.

Открывается диалоговое окно Отобразить наблюдения: Случайная подвыборка.

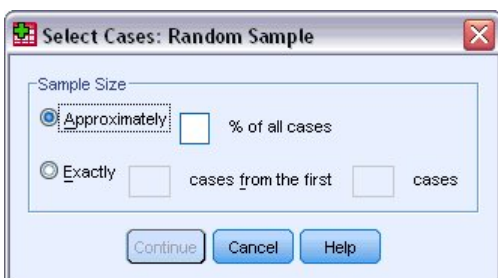


Рисунок 86. Диалоговое окно Отобразить наблюдения: Случайная подвыборка

Пользователь может выбрать одно из следующих значений размера подвыборки:

- **Приблизительно.** Пользователь указывает долю в процентах. Генерируется случайная выборка, объем которой составляет примерно заданный процент от общего числа наблюдений.
- **Точно.** При выборе этой опции необходимо задать точное число наблюдений в подвыборке. Необходимо также задать число наблюдений, из которых генерируется выборка. Это второе

число не должно быть больше числа наблюдений в файле данных. Если же это число превосходит общее число наблюдений в файле данных, то сгенерированная выборка будет содержать наблюдений пропорционально меньше запрошенного числа.

Выбор временного диапазона или диапазона наблюдений

Для выбора диапазона наблюдений на основе даты, времени или номера наблюдения:

1. Выберите пункт **Временной диапазон или диапазон наблюдений** и щелкните **Диапазон** в диалоговом окне Выбрать наблюдения.

При этом открывается диалоговое окно Отобразить наблюдения: Диапазон, в котором пользователь может выбрать диапазон номеров наблюдений (строки).

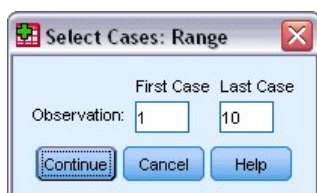


Рисунок 87. Диалоговое окно Отобразить наблюдения: Диапазон

- **Первое наблюдение.** Введите начальную дату и/или время диапазона. Если значение даты не определено, введите номер начального наблюдения (номер строки в редакторе данных, если не включена функция расщепления файла). Если значение параметра Последнее наблюдение не указано, отбираются все наблюдения от начальной даты и времени до конца временного ряда.
- **Последнее наблюдение.** Введите конечную дату и/или время диапазона. Если значение даты не определено, введите номер конечного наблюдения (номер строки в редакторе данных, если не включена функция расщепления файла). Если значение параметра Первое наблюдение не указано, отбираются все наблюдения с начала временного ряда до конечной даты и времени.

Для временного ряда данных с определенными переменными дат можно выбрать ряд диапазон дат/времени на основе определенных переменных дат. Каждая запись представляет собой наблюдение, произведенное в разное время, а файл сортируется в хронологическом порядке.



Рисунок 88. Диалоговое окно Отобразить наблюдения: Диапазон (временные ряды)

Для создания переменных даты для данных временных рядов:

2. Выберите в меню:

Данные > Задать даты...

Действия с неотобранными наблюдениями

Вы можете выбрать один из следующих вариантов:

- **Отфильтровать неотобранные наблюдения.** Неотобранные наблюдения не включаются в анализ, но остаются в наборе данных. Вы сможете использовать неотобранные наблюдения позднее в данном сеансе, если отключите фильтрацию. Если вы отбираете случайную выборку

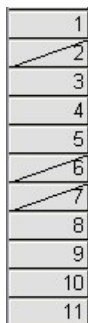
или отбираете наблюдения на основе условия, то генерируется переменная с именем *filter_\$*, имеющая значения 1 для отобранных наблюдений и значение 0 для неотобранных наблюдений.

- **Копировать отобранные наблюдения в новый набор данных.** Отобранные наблюдения копируются в новый набор данных, при этом исходный набор данных остается неизменным. Неотобранные наблюдения не включаются в новый набор данных и остаются в прежнем виде в исходном наборе данных.
- **Неотобранные наблюдения удаляются.** Неотобранные наблюдения удаляются из набора данных. Удаленные наблюдения могут быть восстановлены только путем выхода без сохранения каких-либо изменений и затем повторным открытием файла. Удаленные наблюдения не восстанавливаются, если вы сохраняете изменения в файле данных.

Примечание: Если вы удаляете неотобранные наблюдения и сохраняете файл, наблюдения восстановить невозможно.

Состояние выбора наблюдения

Если вы отобрали подгруппу наблюдений, но не удалили невыбранные наблюдения, не отобранные наблюдения будут перечеркнуты в Редакторе данных диагональной линией через номера строк.



1
2
3
4
5
6
7
8
9
10
11

Рисунок 89. Состояние выбора наблюдения

Глава 10. Файлы для примеров

Файлы примеров, установленные вместе с продуктом, находятся во вложенной папке *Образцы* папки, в которой установлена система. В подкаталоге Samples есть отдельная папка для каждого из следующих языков: английский, французский, немецкий, итальянский, японский, корейский, польский, русский, упрощенный китайский, испанский и традиционный китайский.

Не все файлы примеров доступны на каждом языке. Если файл примера не доступен на конкретном языке, языковая папка содержит версию этого файла примера на английском языке.

Описание

Ниже дано краткое описание файлов, используемых в различных примерах в данной документации.

- **accidents.sav.** Это файл с гипотетическими данными, относящимися к страховой компании, изучающей факторы риска возраста и пола для дорожно-транспортных происшествий в заданном районе. Каждое наблюдение соответствует перекрестной классификации возрастной категории и пола.
- **adl.sav.** Это файл гипотетических данных, относящихся к усилиям по определению преимуществ предлагаемого вида лечения больных с инсультом. Врачи разбили случайным образом пациентов-женщин с инсультом на две группы. Первая группа получала стандартную физиотерапию, а вторая получала дополнительную эмоциональную терапию. Через три месяца после лечения была произведена оценка в виде порядковых переменных способности каждой пациентки выполнять действия повседневной жизни.
- **advert.sav.** Это файл гипотетических данных, относящихся к усилиям розничных торговцев установить зависимость между деньгами, затрачиваемыми на рекламу, и результатами продаж. Для этого они собрали данные о прошлых продажах и связанные с ними затраты на рекламу
- **aflatoxin.sav.** Это файл гипотетических данных, относящихся к тестированию урожая кукурузы на афлатоксин, яд, концентрация которого колеблется в больших пределах между партиями урожая и в пределах одной партии урожая. Обработчик зерен получил 16 образцов от каждой из 8 партий урожая и измерил уровни афлатоксина в частях на миллиард (PPB).
- **anorectic.sav.** Работая над стандартизацией симптоматиологии аноректичного/булимичного поведения, исследователи ¹ исследовали 55 подростков с известными нарушениями питания. Каждый пациент обследовался четыре раза за четыре года, что дало в сумме 220 обследований. При каждом обследовании пациентов оценивали по каждому из 16 симптомов. Оценки симптомов пропущены для пациента 71 на 2 обследовании, пациента 76 на втором обследовании и пациента 47 на третьем обследовании, что дает 217 валидных наблюдений.
- **bankloan.sav.** Это файл гипотетических данных, относящихся к усилиям банка снизить частоту неуплат задолженностей по кредитам. Файл содержит финансовую и демографическую информацию о 850 бывших и потенциальных клиентах. Первые 700 наблюдений - это клиенты, которые ранее получали кредиты. Последние 150 наблюдений представляют собой потенциальных клиентов, которых банку нужно классифицировать как хорошие или плохие риски кредитования.
- **bankloan_binning.sav.** Это файл гипотетических данных, содержащий финансовую и демографическую информацию о 5000 бывших клиентах.
- **behavior.sav.** В классическом примере ² 52 студентов попросили оценить комбинации 15 ситуаций и 15 поведений по 10-балльной шкале от 0="очень подходит" to 9="абсолютно не подходит." С усреднением по индивидуумам значения принимались как различия.

¹ Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

- **behavior_ini.sav.** Этот файл данных содержит исходную конфигурацию двумерного решения для *behavior.sav*.
- **brakes.sav.** Это файл гипотетических данных, относящихся к управлению качеством на заводе, выпускающем дисковые тормоза для высококлассных автомобилей. Файл данных содержит измерения диаметра 16 дисков от каждой из 16 производственных машин. Целевой диаметр для дисков составляет 322 миллиметра.
- **breakfast.sav.** В классическом исследовании ³, 21 магистров бизнеса - выпускников школы Уортон и их супруг попросили оценить 15 объектов завтрака в порядке предпочтения, от 1="наиболее предпочтительного" до 15="наименее предпочтительного." Их предпочтения регистрировались по шести различным сценариям, от "Общего предпочтения" до "Легкая закуска, только с напитками."
- **breakfast-overall.sav.** Этот файл данных содержит предпочтения объектов завтрака для первого сценария, только "Общее предпочтение".
- **broadband_1.sav.** Это файл гипотетических данных, содержащий количество абонентов национальных широкополосных услуг. Этот файл данных содержит номера ежемесячных абонентов по 85 регионам за четырехлетний период.
- **broadband_1.sav.** Этот файл данных идентичен файлу *broadband_1.sav*, но содержит данные для трех дополнительных месяцев.
- **car_insurance_claims.sav.** Набор данных, представленный и проанализированный в другой работе ⁴, касается исков за повреждение автомобилей. Среднюю сумму иска можно смоделировать как имеющую гамма-распределение, использующее функцию инверсной связи для соотношения среднего зависимой переменной с линейной комбинацией возраста обладателя страховки, типа автомобиля и возраста автомобиля. Количество зарегистрированных исков можно использовать в качестве масштабирующего веса.
- **car_sales.sav.** Этот файл данных содержит гипотетические оценки продаж, прайс-листы и технические характеристики различных марок и моделей автомобилей. Прайс-листы и технические характеристики были получены поочередно от *edmunds.com* и сайтов производителей.
- **car_sales_uprepared.sav.** Это измененная версия файла *car_sales.sav*, в которую не входят все трансформированные версии полей.
- **carpet.sav.** В известном примере ⁵ компания, заинтересованная в маркетинге нового средства чистки ковров, хочет изучить влияние пяти факторов на потребительские предпочтения - дизайн упаковки, фирменное название, цена, знак *Идеальный дом* и гарантия возврата средств. Есть три уровня факторов для дизайна упаковки, каждый из которых отличается размещением щетки аппликатора; три фирменных названия (*K2R*, *Glory* и *Bissell*); три ценовых уровня и два уровня (либо да, либо нет) для каждого из последних двух факторов. Десять потребителей ранжировали 22 профиля, определяемых этими факторами. Переменная *Предпочтение* содержит ранг среднего ранжирования для каждого профиля. Низкое ранжирование соответствует высокому предпочтению. Эта переменная отражает общую меру предпочтения для каждого профиля.
- **carpet_prefs.sav.** Этот файл данных основан на том же примере, что и описанный для *carpet.sav*, но он содержит фактическое ранжирование, полученное от каждого из 10 потребителей. Потребителей попросили ранжировать 22 профилей изделия от наиболее предпочтительного до наименее предпочтительного. Переменные с *PREF1* по *PREF22* содержат идентификаторы ассоциированных профилей, как они определены в *carpet_plan.sav*.
- **catalog.sav.** Этот файл данных содержит цифры гипотетических ежемесячных продаж трех товаров, продаваемых компанией, торгующей по каталогу. Сюда также включены данные для пяти возможных переменных предикторов.

² Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

³ Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

⁴ McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

⁵ Green, P. E., and Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **catalog_seasfac.sav.** Этот файл данных такой же, что и *catalog.sav*, за исключением того, что в него добавлен набор сезонных факторов, рассчитанных с помощью процедуры Сезонная декомпозиция, вместе с сопровождающими переменными дат.
- **cellular.sav.** Это файл гипотетических данных, относящихся к компании сотовой связи, старающейся уменьшить отток абонентов. Баллы предрасположенности к оттоку абонентов применяются к учетным записям в ранге от 0 до 100. Учетные записи, набирающие 50 и выше баллов, возможно, собираются сменить провайдера.
- **ceramics.sav.** Это файл гипотетических данных, относящихся к усилиям производителя определить, имеет ли новый высококачественный сплав более высокую жаростойкость, чем стандартный сплав. Каждое наблюдение представляет собой отдельный тест одного из сплавов. Регистрируется температура, при которой отказывает подшипник.
- **cereal.sav.** Это файл гипотетических данных, относящихся к опросу 880 людей об их предпочтениях за завтраком, с указанием их возраста, пола, семейного положения и ведут ли они активный образ жизни (исходя из того, делают ли они физические упражнения по крайней мере два раза в неделю). Каждое наблюдение представляет собой отдельного респондента.
- **clothing_defects.sav.** Это файл гипотетических данных, относящихся к процессу управления качеством на швейной фабрике. Из каждой партии, изготовленной на фабрике, инспекторы берут образец одежды и считают количество забракованной одежды.
- **coffee.sav.** Этот файл данных относится к воспринимаемым образам шести марок кофе со льдом⁶. Для каждого из 23 атрибутов образа кофе со льдом люди подбирали все марки, которые описывал данный атрибут. Шесть марок были обозначены как AA, BB, CC, DD, EE и FF, чтобы сохранить конфиденциальность.
- **contacts.sav.** Это файл гипотетических данных, относящихся к спискам контактов для группы корпоративных торговых представителей по продаже компьютеров. Каждое контактное лицо классифицировалось по отделению компании, в котором этот человек работает, и по должности, которую он занимает. Также регистрировался объем последней продажи, время с момента последней продажи и величина компании контактного лица.
- **creditpromo.sav.** Это файл гипотетических данных, относящихся к усилиям банка снизить частоту невозврата кредитов. Для этого были случайным образом отобраны 500 владельцев кредитных карточек. Половина из них получила рекламное объявление с предложением пониженной процентной ставки на покупки, которые будут сделаны в течение следующих трех месяцев. Половина получила стандартное сезонное рекламное объявление.
- **customer_dbase.sav.** Это файл гипотетических данных, относящихся к усилиям компании использовать информацию в своем хранилище данных, чтобы сделать особые предложения клиентам, которые вероятнее всего откликнутся. Подмножество базы данных клиентов было выбрано случайным образом, этим клиентам было сделано особое предложение и зарегистрирована их реакция.
- **customer_information.sav.** Это файл гипотетических данных, содержащий почтовые сведения о пользователе, например имя и адрес.
- **customer_subset.sav.** Подмножество из 80 наблюдений из файла *customer_dbase.sav*.
- **debate.sav.** Это файл гипотетических данных, относящихся к парным ответам на опрос участников политических дебатов до и после этих дебатов. Каждое наблюдение представляет собой отдельного респондента.
- **debate_aggregate.sav.** Этот файл гипотетических данных, агрегирующий ответы в *debate.sav*. Каждое наблюдение представляет собой перекрестную классификацию предпочтения до и после дебатов.
- **demo.sav.** Это файл гипотетических данных, относящихся к базе данных клиентов, сделавших покупки, с целью рассылки ежемесячных предложений. Регистрируется, отреагировал ли клиент на предложение, а также различная демографическая информация.

⁶ Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

- **demo_cs_1.sav.** Это файл гипотетических данных, относящихся к первому этапу усилий компании составить базу данных с информацией опросов. Каждое наблюдение соответствует отдельному городу, также регистрируется идентификация региона, провинции, района и города.
- **demo_cs_2.sav.** Это файл гипотетических данных, относящихся ко второму этапу усилий компании составить базу данных с информацией опросов. Каждое наблюдение представляет собой отдельную семью из городов, отобранных на первом этапе. Также регистрируется идентификация региона, провинции, района, города, городского района. Также включена информация о выборках на первых двух этапах проекта.
- **demo_cs.sav.** Это файл гипотетических данных, содержащий информацию опроса, собранную с помощью проекта комплексной выборки. Каждое наблюдение соответствует отдельной семье. Регистрируется также различная демографическая информация и информация о выборках.
- **diabetes_costs.sav.** Это файл гипотетических данных, который содержит информацию, поддерживаемую страховой компанией о держателях полисов, страдающих диабетом. Каждое наблюдение соответствует одному держателю полиса.
- **dietstudy.sav.** Этот файл гипотетических данных содержит результаты исследования "диеты Стилмана"⁷. Каждое наблюдение соответствует отдельному субъекту и регистрирует его или ее вес до и после в фунтах и уровень триглицерида в мг/100 мл.
- **dmdata.sav.** Это гипотетические данные, которые содержат сведения о демографии и покупках для компании, занимающейся прямым маркетингом. Файл *dmdata2.sav* содержит информацию для подмножества адресов, по которым получена пробная рассылка, а *dmdata3.sav* содержит информацию об остальных адресах, по которым пробная рассылка не получена.
- **dvdplayer.sav.** Это файл гипотетических данных, относящихся к разработке нового DVD-плеера. Маркетинговая команда собрала с помощью прототипа данные о целевой группе. Каждое наблюдение соответствует отдельному отслеживаемому пользователю и фиксирует некоторую демографическую информацию о нем и его ответах на вопросы о прототипе.
- **german_credit.sav.** Этот файл данных взят из набора данных "German credit" в репозитории баз данных машинного обучения⁸ Калифорнийского университета, город Ирвин.
- **grocery_1month.sav.** Этот файл гипотетических данных - файл данных *grocery_coupons.sav* со "свернутыми" еженедельными покупками, так что каждое наблюдение соответствует отдельному клиенту. Некоторые еженедельно изменявшиеся переменные в результате исчезают, и теперь регистрируемая потраченная сумма представляет собой сумму затрат на покупки, сделанные в течение четырех недель исследования.
- **grocery_coupons.sav.** Это гипотетический файл данных, содержащий данные опроса, проведенного сетью магазинов бакалейных товаров, заинтересованной в покупательских привычках своих клиентов. Каждый покупатель отслеживался в течение четырех недель, и каждое наблюдение соответствует отдельной паре клиент-неделя и регистрирует информацию о том, где и как покупает клиент, включая сумму, потраченную на бакалейные товары в течение этой недели.
- **guttman.sav.** Белл⁹ представил таблицу для иллюстрации возможных социальных групп. Гуттман¹⁰ использовал часть этой таблицы, в которой пять переменных, описывающих такие вещи, как социальное взаимодействие, чувство принадлежности к группе, физическая близость членов группы и формальность отношений были пересечены с семью теоретическими социальными группами, включая толпы (например, публика на футбольном матче), аудитории (например, люди в театре или на лекции), публику (например, аудитория газет или телевидения), сборища (подобные толке, но с гораздо более сильным взаимодействием), первичные группы (близкие),

⁷ Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

⁸ Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

⁹ Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

¹⁰ Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

вторичные группы (добровольцы) и современное общество (слабая конфедерация, возникающая в результате физической близости и потребности в специализированных услугах).

- **health_funding.sav.** Это файл гипотетических данных, содержащий данные о финансировании здравоохранения (сумма на 100 человек населения), заболеваемости (заболеваемость на 10000 человек населения) и посещениях провайдеров медицинских услуг (частота на 100000 человек). Все примеры представляют разные города.
- **hivassay.sav.** Это файл гипотетических данных, относящихся к усилиям фармацевтической лаборатории разработать быстрый тест для обнаружения ВИЧ-инфекции. Результаты теста - восемь все более насыщенных оттенков красного, причем более темный цвет указывает на большую вероятность инфекции. Лабораторные исследования проводились на 2000 образцах крови, половина из которых была ВИЧ-инфицирована, а другая половина - чистая.
- **hourlywagedata.sav.** Этот файл гипотетических данных, относящихся к почасовой оплате медицинских сестер, работающих в офисе и госпитале и имеющих различный опыт.
- **insurance_claims.sav.** Этот файл гипотетических данных, относящихся к страховой компании, которой необходимо построить модель для того, чтобы отмечать подозрительные, потенциально обманные требования. Каждое наблюдение представляет собой отдельное требование.
- **insure.sav.** Это гипотетический файл данных, относящихся к страховой компании, которая изучает факторы риска того, что клиент подаст заявку на выплату по 10-летнему контракту страхования жизни. Каждое наблюдение в файле представляет собой пару контрактов, в одном из которых зарегистрирована заявка, а в другом - нет, подобранных по полу и возрасту.
- **judges.sav.** Это файл гипотетических данных, относящихся к баллам, присужденным опытными судьями (плюс один энтузиаст) 300 гимнастическим упражнениям. Каждая строка представляет собой отдельное упражнение. Судьи видели одни и те же упражнения.
- **kinship_dat.sav.** Розенберг и Ким ¹¹ проанализировали 15 терминов родства (тетя, брат, двоюродный брат, дочь, отец, внучка, дедушка, бабушка, внук, племянник, племянница, сестра, сын, дядя). Они попросили четыре группы студентов колледжа (две женских и две мужских) рассортировать эти термины на основе подобия. Две группы (одну женскую и одну мужскую) попросили отсортировать два раза, причем вторая сортировка основывалась на критерии, отличающемся от критерия первой сортировки. Таким образом, всего получено шесть "источников". Каждый источник соответствует матрице близости размером 15 x 15, значения в ячейках которой равно разности между количеством людей в источнике и числом случаев, когда объекты оказывались в одном разделе в этом источнике.
- **kinship_ini.sav.** Этот файл данных содержит начальную конфигурацию двумерного решения для *behavior.sav*.
- **kinship_var.sav.** Этот файл данных содержит независимые переменные *gender*, *gener* (поколение) и *degree* (степень разделения), которые можно использовать для интерпретации измерений в решении для *kinship_dat.sav*. В частности, их можно использовать для ограничения пространства решений линейной комбинацией этих переменных.
- **marketvalues.sav.** Этот файл данных относится к продаже жилья в новостройках Алгонкин, Иллинойс за период после 1999–2000. Сведения об этих продажах общедоступны.
- **nhis2000_subset.sav.** Проект National Health Interview Survey (NHIS) - это большое, основанное на населении исследование гражданского населения США. Опросы проводились лицом к лицу в национально репрезентативной выборке домохозяйств. Для членов каждой семьи получали демографическую информацию и наблюдения за поведением в отношении здоровья и состоянием здоровья. Этот файл данных содержит подмножество информации из исследования 2000 года. Национальный центр статистики здравоохранения. National Health Interview Survey, 2000. Файл данных и документация для общественного пользования. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Доступ получен в 2003 году.
- **ozone.sav.** Данные включают 330 наблюдений шести метеорологических переменных для прогнозирования концентрации озона по оставшимся переменным. Среди прочих, предыдущие

¹¹ Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

исследователи ¹², ¹³, обнаружили среди этих переменных нелинейности, затрудняющие применение стандартных методов регрессионного анализа.

- **pain_medication.sav.** Этот файл гипотетических данных содержит результаты клинических испытаний противовоспалительного препарата для лечения хронической боли в суставах. Особый интерес представляет время, необходимое для того, чтобы лекарство начало действовать, и как это сравнимо с существующим препаратом.
- **patient_los.sav.** Этот файл гипотетических данных содержит истории болезни пациентов, поступивших в госпиталь с подозрением на инфаркт миокарда (ИМ или "сердечный приступ"). Каждое наблюдение соответствует отдельному пациенту и фиксирует много переменных, связанных с пребыванием пациента в госпитале.
- **patlos_sample.sav.** Этот файл гипотетических данных содержит истории болезни выборки пациентов, получавших тромболитики во время лечения инфаркта миокарда (ИМ или "сердечного приступа"). Каждое наблюдение соответствует отдельному пациенту и фиксирует много переменных, связанных с пребыванием пациента в госпитале.
- **poll_cs.sav.** Это файл гипотетических данных, относящихся к усилиям специалистов по опросам определить уровень общественной поддержки законопроекта перед тем, как он пройдет утверждение. Наблюдения соответствуют зарегистрированным избирателям. Каждое наблюдение соответствует округу, городу и окружению, в котором живет избиратель.
- **poll_cs_sample.sav.** Этот файл гипотетических данных содержит выборку избирателей, перечисленных в *poll_cs.sav*. Выборка осуществлялась в соответствии с проектом, заданным в файле плана *poll.csplan*, и данный файл данных содержит вероятности включения и веса выборки. Заметьте, однако, что из-за того, что в плане выборки используется метод вероятности, пропорциональной размеру (PPS), есть также файл, содержащий вероятности совместной выборки (*poll_jointprob.sav*). Дополнительные переменные, соответствующие демографическим данным избирателя и его мнению о предлагаемом законопроекте, собирались и добавлялись в файл данных после того, как была сделана выборка.
- **property_assess.sav.** Это файл гипотетических данных, относящийся к усилиям асессора округа поддерживать обновление оценок недвижимости при ограниченных ресурсах. Наблюдения соответствуют недвижимости, проданной в округе за последний год. Каждое наблюдение в файле данных регистрирует участок, на котором расположена недвижимость, имя асессора, который посещал недвижимость последним, время, прошедшее с момента последней оценки, оценку, сделанную в то время и продажную стоимость недвижимости.
- **property_assess_cs.sav.** Это файл гипотетических данных, относящийся к усилиям асессора штата поддерживать обновление оценок недвижимости при ограниченных ресурсах. Наблюдения соответствуют недвижимости в штате. Каждое наблюдение в файле данных регистрирует округ, участок и окружение, в котором расположена недвижимость, время, прошедшее с момента последней оценки и сделанную тогда оценку.
- **property_assess_cs_sample.sav.** Этот файл гипотетических данных содержит выборку недвижимости, перечисленной в *property_assess_cs.sav*. Выборка была сделана в соответствии с проектом, заданным в файле плана *property_assess.csplan*, и данный файл данных содержит вероятности включения и веса выборки. Дополнительную переменную *Текущее значение* собирали и добавляли в файл данных после того, как была сделана выборка.
- **recidivism.sav.** Это файл гипотетических данных, относящихся к усилиям государственного судебного исполнительного органа понять уровень рецидивизма в своей области юрисдикции. Каждое наблюдение соответствует лицу, ранее совершившему правонарушение, и регистрирует демографическую информацию о нем, некоторые подробности первого правонарушения и время, прошедшее до повторного ареста, если он произошел в течение двух лет после первого ареста.
- **recidivism_cs_sample.sav.** Это файл гипотетических данных, относящихся к усилиям государственного судебного исполнительного органа понять уровень рецидивизма в своей

¹² Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

¹³ Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

области юрисдикции. Каждое наблюдение соответствует лицу, ранее совершившему правонарушение, освобожденному после первого ареста в течение июня 2003 года, и регистрирует демографическую информацию о нем, некоторые подробности первого правонарушения и дату второго ареста, если он произошел до конца июня 2006 года. Правонарушители выбирались из выборочных отделений в соответствии с планом выборки, заданным в *recidivism_cs.csplan*; из-за того, что в плане выборки используется метод вероятности, пропорциональной размеру (PPS), есть также файл, содержащий вероятности совместной выборки (*recidivism_cs_jointprob.sav*).

- **rfm_transactions.sav.** Это файл гипотетических данных, содержащий данные о покупках, включая дату покупки, перечень приобретенных товаров, а также сумму покупки для каждой такой транзакции.
- **salesperformance.sav.** Это файл гипотетических данных, относящихся к оценке двух новых курсов по обучению продажам. Шестьдесят сотрудников, разбитых на три группы, получали стандартное обучение. Кроме того, в группе 2 проводилось техническое обучение, а в группе 3 - практические занятия. По окончании курсов обучения каждый сотрудник был протестирован и его оценка зафиксирована. Каждое наблюдение в файле данных представляет собой отдельного обучаемого и регистрирует группу, которой он был назначен, и оценку, которую он получил на экзамене.
- **satisf.sav.** Это файл гипотетических данных, относящихся к опросу об удовлетворении, проведенному компанией розничной торговли в 4 местах расположения магазинов. Всего было опрошено 582 клиента, и каждое наблюдение представляет собой ответы одного клиента.
- **screws.sav.** Этот файл данных содержит информацию о технических характеристиках винтов, болтов, гаек и гвоздей¹⁴.
- **shampoo_ph.sav.** Это файл гипотетических данных, относящихся к управлению качеством на фабрике, производящей средства для ухода за волосами. Через равные промежутки времени измерялись шесть отдельных выходных партий и регистрировался их pH. Диапазон назначения равен 4.5–5.5.
- **ships.sav.** Набор данных, представленный и проанализированный в другой работе¹⁵, относящийся к повреждениям, причиненным грузовым судам волнами. Количество инцидентов можно смоделировать в виде распределения Пуассона при заданных типе судна, времени постройки и продолжительности эксплуатации. Агрегированные месяцы эксплуатации для каждой ячейки таблицы, образованной перекрестной классификацией факторов, дают значения подверженности риску.
- **site.sav.** Это файл гипотетических данных, относящихся к усилиям компании выбрать новые площадки для своего расширяющегося бизнеса. Она наняла двух консультантов, чтобы они по отдельности оценили площадки. Эти консультанты, в дополнение к подробному отчету, оценили каждую площадку как "хорошую," "подходящую" или "плохую".
- **smokers.sav.** Этот файл данных представляет собой выборку из 1998 National Household Survey of Drug Abuse и является вероятностной выборкой американских домохозяйств. (<http://dx.doi.org/10.3886/ICPSR02934>) Таким образом, первый шаг в анализе этого файла данных должен состоять во взвешивании, отражающем тенденции совокупности.
- **stocks.sav** Этот файл гипотетических данных содержит биржевые цены и объемы продаж за один год.
- **stroke_clean.sav.** Этот файл гипотетических данных содержит состояние медицинской базы данных после того, как ее почистили с помощью процедур опции Подготовка данных.
- **stroke_invalid.sav.** Этот файл гипотетических данных содержит первоначальное состояние медицинской базы данных, а также несколько ошибок в записях данных.
- **stroke_survival.** Это файл гипотетических данных, относящихся к срокам дожития для пациентов, выходящих из реабилитационной программы Post-Ischemic Stroke Face a Number of Challenges. Регистрировались пост-инсультное состояние, возникновение инфаркта миокарда, ишемический

¹⁴ Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

¹⁵ McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

инсульт или геморрагический инсульт и время события. Выборка усечена слева, поскольку она включает только пациентов, выживших к концу программы пост-инсультной реабилитации.

- **stroke_valid.sav.** Этот файл гипотетических данных содержит состояние медицинской базы данных после проверки значений с помощью процедуры Проверить данные. Она по-прежнему содержит потенциально ненормальные наблюдения.
- **survey_sample.sav.** Этот файл данных содержит данные опроса, включая демографические данные и различные показатели, характеризующие отношение. Он основывается на подмножестве переменных из 1998 NORC General Social Survey, но для демонстрационных целей были изменены некоторые значения данных и добавлены дополнительные фиктивные переменные.
- **tcm_kpi.sav.** Это файл гипотетических данных, который содержит еженедельные значения ключевых показателей эффективности для бизнеса. Он содержит также еженедельные данные для многих контролируемых показателей за тот же период времени.
- **tcm_kpi_upd.sav.** Этот файл данных идентичен файлу *tcm_kpi.sav*, но содержит данные для 4 дополнительных недель.
- **telco.sav.** Это файл гипотетических данных, относящихся к усилиям телекоммуникационной компании уменьшить отток абонентов в своей базе данных клиентов. Каждое наблюдение соответствует отдельному клиенту и регистрирует различную демографическую информацию и информацию о пользовании услугами.
- **telco_extra.sav.** Этот файл данных аналогичен файлу данных *telco.sav*, но переменные "срок пребывания" и log-преобразованные затраты клиента были удалены и заменены стандартизованными переменными log-преобразованные затраты клиента.
- **telco_missing.sav.** Этот файл данных является подмножеством файла данных *telco.sav*, но некоторые значения демографических данных были заменены значениями отсутствия.
- **testmarket.sav.** Это файл гипотетических данных, относящихся к планам сети быстрого питания добавить новое блюдо в свое меню. Есть три возможных компании по продвижению нового продукта, и новое блюдо в торговых точках на нескольких выбранных случайным образом рынках. В каждой точке использовался различный метод продвижения, а еженедельные продажи нового блюда регистрировались для первых четырех недель. Каждое наблюдение соответствует отдельной точке-неделе.
- **testmarket_1month.sav.** Этот файл гипотетических данных - файл данных *testmarket.sav* со "свернутыми" еженедельными продажами, причем каждое наблюдение соответствует отдельному расположению. Некоторые еженедельно изменявшиеся переменные в результате исчезают, и теперь регистрируемые продажи представляют собой сумму продаж в течение четырех недель исследования.
- **tree_car.sav.** Это файл гипотетических данных, содержащий демографические данные и данные продаж цен автомобилей.
- **tree_credit.sav.** Это файл гипотетических данных, содержащий демографические данные и данные истории банковских займов.
- **tree_missing_data.sav** Это файл гипотетических данных, содержащий демографические данные и данные хронологии банковских займов с большим количеством пропущенных значений.
- **tree_score_car.sav.** Это файл гипотетических данных, содержащий демографические данные и данные продаж цен автомобилей.
- **tree_textdata.sav.** Простой файл данных только с двумя переменными, предназначенный прежде всего для того, чтобы показать состояние по умолчанию переменных перед назначением уровня измерения и меток переменных.
- **tv-survey.sav.** Это файл гипотетических данных, относящихся опросу, проведенному телестудией о том, нужно ли расширять успешную программу. 906 респондентов спросили, будут ли они смотреть программу при различных условиях. Каждая строка представляет собой отдельного респондента, а каждый столбец - отдельное условие.
- **ulcer_recurrence.sav.** Этот файл содержит частичную информацию из исследования, предназначенного для сравнения эффективности двух методов лечения для предотвращения

повторного возникновения язвы желудка. Он представляет собой хороший пример интервал-цензурированных данных и был представлен и проанализирован в другой работе ¹⁶.

- **ulcer_recurrence_recoded.sav.** Этот файл реорганизует информацию в *ulcer_recurrence.sav*, чтобы можно было моделировать вероятность события для каждого интервала исследования, а не просто вероятность события в конце исследования. Он был представлен и проанализирован в другой работе ¹⁷.
- **verd1985.sav.** Этот файл данных относится к опросу ¹⁸. Регистрировались ответы 15 субъектов на 8 переменных. Эти переменные были разбиты на три набора. В набор 1 входят *age* и *marital*, в набор 2 входят *pet* и *news* и в набор 3 входят *music* и *live*. Переменная *pet* является множественной полиномиальной, *age* является порядковой. Все остальные переменные являются одиночными номинальными.
- **virus.sav.** Это файл гипотетических данных, относящихся к усилиям провайдера Интернет-услуг (ISP) определить влияние вируса на свои сети. Они отслежили процент (приблизительный) зараженного e-mail трафика по своим сетям по времени от момента обнаружения до устранения угрозы.
- **wheeze_steubenville.sav.** Это подмножество длительного исследования влияния загрязнения воздуха на здоровье детей ¹⁹. Данные содержат повторяющиеся двоичные измерения состояния свистящего дыхания у детей из Штойбенвилля, штат Огайо, в возрасте 7, 8, 9 и 10 лет, вместе с фиксированной регистрацией того, курила ли мать в течение первого года исследования.
- **workprog.sav.** Это файл гипотетических данных, относящихся к государственной программе работ, которая старается предоставить неимущим людям лучшую работу. Отслеживалась выборка потенциальных участников программы, причем некоторые из них были выбраны для участия в программе, а другие - нет. Каждое наблюдение представляет собой отдельного участника программы.
- **worldsales.sav** Этот гипотетический файл данных содержит товарооборот по континентам и товарам..

¹⁶ Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

¹⁷ Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

¹⁸ Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

¹⁹ Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

Замечания

Эта публикация разрабатывалась для продуктов и услуг, предлагаемых в США. Этот материал может быть доступен от IBM на других языках. Однако для его получения может понадобиться приобрести продукт или версию продукта на нужном языке.

IBM может не предоставлять в других странах продукты, услуги и аппаратные средства, описанные в данном документе. За информацией о продуктах и услугах, предоставляемых в вашей стране, обращайтесь к местному представителю IBM. Ссылки на продукты, программы или услуги IBM не означают и не предполагают, что можно использовать только указанные продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любого продукта, программы или сервиса, не произведенного корпорацией IBM, лежит на пользователе.

IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Предъявление данного документа не предоставляет какую-либо лицензию на эти патенты. Вы можете послать письменный запрос о лицензии по адресу:

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране или направьте запрос в письменной форме по адресу:

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan*

КОРПОРАЦИЯ INTERNATIONAL BUSINESS MACHINES ПРЕДОСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ ГАРАНТИЙ, ВКЛЮЧАЯ, НО НЕ ОГРАНИЧИВАЯСЬ ТАКОВЫМИ, ПОДРАЗУМЕВАЕМЫЕ ГАРАНТИИ ОТСУТСТВИЯ НАРУШЕНИЙ, КОММЕРЧЕСКОЙ ПРИГОДНОСТИ ИЛИ СООТВЕТСТВИЯ КАКОЙ-ЛИБО КОНКРЕТНОЙ ЦЕЛИ. В некоторых странах для ряда сделок не допускается отказ от явных или предполагаемых гарантий; в таком случае данное положение к вам не относится.

Эта информация может содержать технические неточности и типографские ошибки. В представленную здесь информацию периодически вносятся изменения; эти изменения будут включаться в новые издания данной публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые ссылки в данной информации на сайты, не принадлежащие IBM, приводятся только для удобства и никоим образом не означают поддержки этих сайтов. Материалы на этих сайтах не входят в число материалов по данному продукту IBM, и весь риск пользования этими сайтами несете вы сами.

IBM может использовать или распространять предоставленную вами информацию любым способом, как фирма сочтет нужным, без каких-либо обязательств перед вами.

Если обладателю лицензии на данную программу понадобится информация о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

Такая информация может быть доступна при соответствующих условиях и соглашениях, включая в некоторых случаях взимание платы.

Описанную в данном документе лицензионную программу и все прилагаемые к ней лицензированные материалы IBM предоставляет на основе положений Соглашения между IBM и Заказчиком, Международного Соглашения о Лицензиях на Программы IBM или любого эквивалентного соглашения между IBM и заказчиком.

Упомянутые данные о производительности и примеры клиентов представлены только для иллюстративных целей. Фактические результаты производительности могут быть иными в зависимости от определенных конфигураций и конкретных условий.

Информация, касающаяся продуктов других компаний (не IBM) была получена от поставщиков этих продуктов, из опубликованных ими заявлений или из прочих общедоступных источников. IBM не проводила тестирования этой продукции и не может подтвердить или опровергнуть информацию о точности ее работы и совместимости, а также другие заявления относительно продуктов других производителей (не IBM). Вопросы относительно возможностей продуктов других компаний (не IBM) следует адресовать поставщикам этих продуктов.

Утверждения, касающиеся намерений и планов IBM, могут быть изменены без предварительного предупреждения; они приведены здесь только для обозначения целей и задач IBM.

Эти сведения содержат примеры данных и отчетов, используемых в повседневных деловых операциях. Чтобы проиллюстрировать их настолько полно, насколько это возможно, данные примеры включают имена индивидуумов, названия компаний, брендов и продуктов. Все эти имена и названия вымышлены и любое их сходство с реальными именами и названиями компаний полностью случайно.

ЛИЦЕНЗИЯ НА КОПИРОВАНИЕ:

Эта информация содержит примеры исходных текстов прикладных программ, которые иллюстрируют приемы программирования на различных платформах. Разрешается копировать, изменять и распространять эти примеры программ в любой форме без оплаты фирме IBM для целей разработки, использования, сбыта или распространения прикладных программ, соответствующих интерфейсу прикладного программирования операционных платформ, для которых эти примеры программ написаны. Эти примеры не были всесторонне проверены во всех возможных условиях. Поэтому IBM не может гарантировать их надежность, пригодность и функционирование. Примеры программ предоставляются "КАК ЕСТЬ", без каких-либо гарантий. IBM не несет никакой ответственности за какой либо ущерб, причиненный в результате использования этих программ.

Каждая копия или каждая часть этих примеров программ или работы, основанной на них, должна содержать следующее замечание об авторских правах:

© Copyright IBM Corp. 2020. Части этого кода получены из примеров программ IBM Corp.

© Copyright IBM Corp. 1989 - 2020. Все права защищены.

Товарные знаки

IBM, логотип IBM, и ibm.com являются товарными знаками или зарегистрированными товарными знаками компании International Business Machines Corp., зарегистрированными во многих странах мира. Прочие наименования продуктов и услуг могут быть товарными знаками, принадлежащими IBM или другим компаниям. Текущий список товарных знаков IBM можно найти в Интернете в разделе "Copyright and trademark information" ("Информация об авторских правах и товарных знаках") по адресу www.ibm.com/legal/copytrade.shtml.

Adobe, логотип Adobe, PostScript и логотип PostScript являются либо зарегистрированными товарными знаками, либо товарными знаками корпорации Adobe Systems в Соединенных Штатах и/или других странах.

Intel, логотип Intel, Intel Inside, логотип Intel Inside, Intel Centrino, логотип Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium и Pentium являются товарными знаками или зарегистрированными товарными знаками компании Intel или ее дочерних компаний в Соединенных Штатах и других странах.

Linux является зарегистрированным товарным знаком Linus Torvalds в Соединенных Штатах и других странах.

Microsoft, Windows, Windows NT и логотип Windows являются товарными знаками корпорации Microsoft в Соединенных Штатах и других странах.

UNIX является зарегистрированным товарным знаком The Open Group в Соединенных Штатах и других странах.

Java и все основанные на Java товарные знаки и логотипы являются товарными знаками Oracle и/или его филиалов.

Индекс

A

Access (Microsoft) [11](#)

E

Excel (Microsoft)
экспорт результатов в [50](#)

H

HTML
экспорт результатов [57](#)

P

PowerPoint (Microsoft)
экспорт результатов в [50](#)

W

Word (Microsoft)
экспорт результатов в [50](#)

A

анализ с файлом разбиения [71](#)

B

ввод данных
нечисловые [20](#)
числовые [19](#)
Вставка синтаксиса
из диалогового окна [59](#)
Вычисление новых переменных [65](#)

G

гистограммы [28](#)

D

диаграммы
гистограммы [28](#)
построение графиков [31](#)
Создание диаграмм [31](#)
столбец [26](#), [31](#)

З

задание меток переменных [21](#)

И

Инструмент Справочник по синтаксису команд [60](#)
Интервальные данные [25](#)

K

категориальные данные
сводные показатели [25](#)
качественные данные [25](#)
количественные данные [25](#)
количественные переменные
сводные показатели [27](#)
Конструктор импорта текстовых файлов [14](#)

M

Мастер баз данных [11](#)
Мастер дат и времени [68](#)
метки значений
контроль за выводом в средстве просмотра [22](#)
назначение [22](#)
Числовые переменные [22](#)
метки переменных
создание [21](#)
мобильные таблицы
получение определений [38](#)
Поля вращения [39](#)
редактирование [40](#)
скрытие десятичных запятых [42](#)
Скрытие строк и столбцов [41](#)
слои [40](#)
типы данных в ячейках [42](#)
транспонирование строк и столбцов [39](#)
формат ячеек [42](#)
форматирование [40](#)

H

наблюдений
выделение [73](#)
сортировка [71](#), [73](#)
непрерывные данные [25](#)
номинальные данные [25](#)

O

окна синтаксиса
автозавершение [60](#)
вставка команд [59](#)
запуск команд [59](#), [61](#)
Команды редактирования [60](#)
контрольные точки [61](#)
цветовое кодирование [60](#)
отбор наблюдений [73](#)
отношения [25](#)

П

перекодирование значений [63](#)
переменные
метки [21](#)
типы данных [22](#)
переменные дат и времени [68](#)
перемещение
элементы в мобильных таблицах [39](#)
элементы в средстве просмотра [37](#)
подмножества наблюдений
выделение [73](#)
Если выполнено условие [74](#)
на основе даты и времени [76](#)
случайная выборка [75](#)
удаление неотобранных наблюдений [76](#)
условные выражения [74](#)
фильтрация неотобранных наблюдений [76](#)
порядковые данные [25](#)
пропущенные значения
для нечисловых переменных [24](#)
для числовых переменных [23](#)
Системное пропущенное [23](#)

Р

редактирование мобильных таблиц [40](#)
Редактор данных
ввод нечисловых данных [20](#)
ввод числовых данных [19](#)

С

сводные показатели
категориальные данные [25](#)
количественные переменные [27](#)
синтаксис [59](#)
системные значения отсутствия [23](#)
скрытие строк и столбцов в мобильных таблицах [41](#)
слои
создание в мобильных таблицах [40](#)
сортировка наблюдений [71](#)
средство просмотра
перемещение вывода [37](#)
скрытие и показ вывода [37](#)
столбчатые диаграммы [26](#)

Т

текстовые данные
ввод данных [20](#)
текстовые файлы данных
чтение [14](#)
типы данных
для переменных [22](#)
транспонирование строк и столбцов в мобильных таблицах [39](#)

У

условные выражения [67](#)

Ф

файлы баз данных
чтение [11](#)
файлы для примеров
местоположение [79](#)
файлы синтаксиса
открытие [61](#)
файлы электронных таблиц
чтение [8](#)
чтение имен переменных [8](#)
Файлы Excel
чтение [8](#)
функции в выражениях [66](#)

Ч

частотные таблицы [25](#)
частоты
таблицы частот [25](#)
числовые данные [19](#)

Ш

шкала измерений [25](#)
шкала измерения [25](#)

Э

экспорт результатов
в Excel [50](#)
в PowerPoint [50](#)
в Word [50](#)
HTML [57](#)

