

Настраиваемые таблицы 25
IBM SPSS

IBM

Примечание

Прежде чем использовать эту информацию и продукт, описанный в ней, прочтите сведения в разделе “Замечания” на стр. 93.

Информация о продукте

Это издание применимо к версии 25, 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.

Содержание

Настраиваемые таблицы 1

Интерфейс строителя таблиц	1
Интерфейс строителя таблиц	1
Построение таблиц	1
Настраиваемые таблицы: вкладка Опции	16
Настраиваемые таблицы: вкладка Заголовки	18
Настраиваемые таблицы: вкладка Статистики критериев	18
Простые таблицы с категориальными переменными	20
Простые таблицы с категориальными переменными	20
Одна категориальная переменная	20
Комбинационная таблица	22
Сортировка и исключение категорий	25
Таблицы с категориальными переменными: стыковывание, вложение и слои	26
Стыковывание категориальных переменных	26
Вложение категориальных переменных	27
Слои	30
Итоги и подытоги для категориальных переменных	32
Простой итог для одной переменной	33
То, что вы видите, зависит от того, что включается в итог	33
Управление расположением итогов	34
Итоги для таблиц с вложениями	34
Итоги для переменных слоя	35
Подытоги	36
Вычисляемые категории для категориальных переменных	38
Простая вычисляемая категория	38
Скрытие категорий в вычисляемой категории	39
Ссылка на подытоги в вычисляемой категории	39
Использование вычисляемых категорий для вывода неисчерпаемых подытогов	41
Таблицы для переменных с одинаковыми категориями	42
Таблица количеств	42
Таблица процентов	43
Итоги и управление категориями	44
Вложенность в таблицах с переменными с одинаковыми категориями	45
Итоговые статистики	45
Переменная - источник итожащих статистик	46
Составные переменные	48

Задаваемые статистики общих итогов для категориальных переменных	49
Подытоживание количественных переменных	51
Подытоживание количественных переменных	51
Сстыковывание количественных переменных	51
Несколько итожащих статистик	52
Частота, количество валидных значений и пропущенные значения	52
Использование различных итожащих статистик для различных переменных	53
Итожащие статистики по категориям	54
Доверительные интервалы	56
Статистические критерии	57
Статистические критерии	57
Критерий независимости (Хи-квадрат)	57
Сравнение средних значений по столбцам	60
Сравнение пропорций по столбцам	63
Замечания по использованию весов и множественных наборов данных	69
Несколько наборов ответов	69
Частоты, ответы, проценты и итоги	70
Использование наборов множественных ответов с другими переменными	72
Категориальные наборы множественных ответов и повторяющиеся ответы	73
Критерии значимости для наборов множественных ответов	75
Пропущенные значения	76
Таблицы без пропущенных значений	77
Включение пропущенных значений в таблицы	77
Форматирование и настройка таблиц	79
Форматирование и настройка таблиц	79
Формат вывода итожащих статистик	79
Вывод меток для итожащих статистик	80
Ширина столбца	81
Вид вывода пустых ячеек	82
Файлы для примеров	83

Замечания 93

Товарные знаки	95
--------------------------	----

Индекс 97

Настраиваемые таблицы

В SPSS Statistics Standard Edition или в модуль Настраиваемые таблицы включены следующие функции настраиваемых таблиц.

Интерфейс построителя таблиц

Интерфейс построителя таблиц

Пользовательские таблицы создаются при помощи удобного интерфейса, который позволяет в процессе создания таблицы видеть, как она будет выглядеть. Процедура также, в отличие от традиционных процедур с "диалоговыми окнами", обеспечивает высокий уровень гибкости, включая возможность изменять размер окна и размер панелей в окне.

Построение таблиц

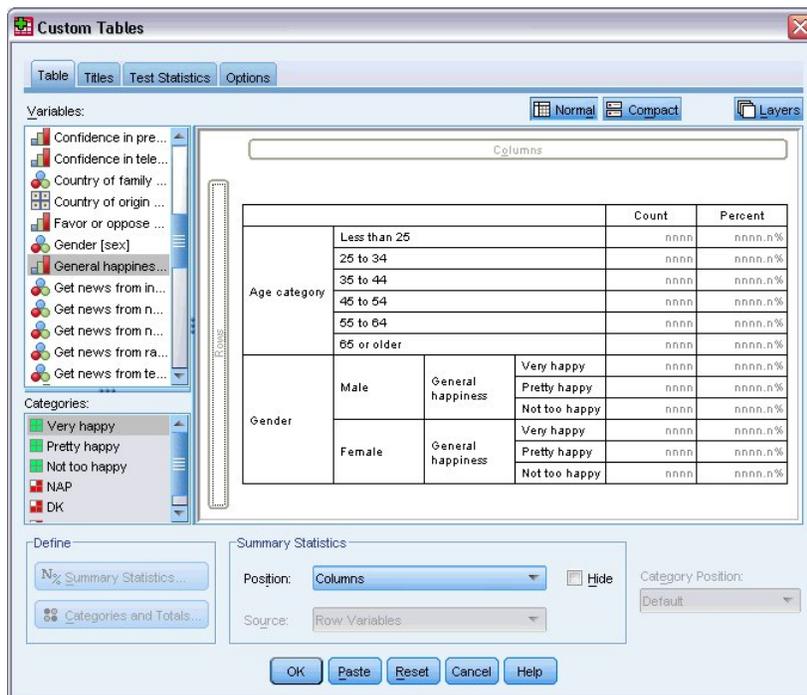


Рисунок 1. Диалоговое окно Настраиваемые таблицы, вкладка Таблица

На вкладке Таблица можно выбрать переменные и итожащие статистики, которые должны появиться в таблицах.

Список переменных. Переменные файла данных отображаются в верхней левой части окна. Процедура Настраиваемые таблицы различает два уровня измерения переменных и различным образом обрабатывает переменные разных уровней измерения:

Категориальное. Это данные с ограниченным числом уникальных значений или категорий (например, пол или религия). Категориальные переменные могут быть текстовыми или числовыми, в которых категории закодированы числовыми кодами (например, 0 = Женский, а 1 = Мужской). Также эти данные называются качественными данными. Категориальные переменные могут быть либо **номинальные**, либо **порядковые**

- *Номинальная.* Переменную можно рассматривать как номинальную, когда ее значения представляют категории без естественного упорядочения, например, подразделение компании, где работает наемный сотрудник. Примеры номинальных переменных включают регион, почтовый индекс или религию.
- *Порядковая.* Переменную можно рассматривать как порядковую, когда ее значения представляют категории с некоторым естественным для них упорядочением, например, уровни удовлетворенности обслуживанием от крайней неудовлетворенности до крайней удовлетворенности. Примеры порядковых переменных включают баллы, представляющие степень удовлетворенности или уверенности, или баллы, оценивающие предпочтение.

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 расстояние между этими зарплатами - 1797 рублей. Такие данные также называют непрерывными.

Категориальные переменные задают в таблице категории (в строках, столбцах или слоях), и статистикой, выводимой по умолчанию, является частота (количество наблюдений в каждой категории). К примеру, таблица для категориальной переменной пола, выводимая с параметрами по умолчанию, будет содержать в себе только количество мужчин и женщин.

Количественные переменные обычно подытоживаются для категорий категориальных переменных, и статистикой, выводимой по умолчанию, является среднее значение. К примеру, таблица с переменной дохода по категориям пола будет содержать в себе средние значения дохода для мужчин и женщин.

Также можно подытоживать количественные переменные сами по себе, без использования категориальных переменных для задания групп. Это очень полезно в случае **состыковывания** итожащих статистик для нескольких количественных переменных. Дополнительную информацию смотрите в разделе “Состыковывание переменных” на стр. 4.

Несколько наборов ответов

Настраиваемые таблицы также позволяют работать со специальными "переменными", которые называются **наборами множественных ответов**. Наборы множественных ответов не являются переменными в обычном смысле. Их нельзя увидеть в Редакторе данных, и они не распознаются другими процедурами. В наборах множественных ответов для ввода ответов на вопросы, на которые можно дать больше одного ответа, используются несколько переменных. Наборы множественных ответов обрабатываются как категориальные переменные, и большинство операций, которые можно выполнять с категориальными переменными, можно делать также и с наборами множественных ответов. Дополнительную информацию смотрите в разделе “Несколько наборов ответов” на стр. 69.

Значок рядом с каждой переменной позволяет идентифицировать тип переменной.

Таблица 1. Значки уровня измерения

	Числовой	Строка	Дата	Время
Количественная (непрерывная)		(не задается)		
Порядковый				
Номинальный				

Таблица 2. Значки наборов множественных ответов

Тип набора множественных ответов	Значок
Набор множественных ответов, множественные категории	
Набор множественных ответов, множественные дихотомии	

Можно изменить уровень измерения переменной в построителе таблиц при помощи щелчка правой кнопкой мыши по переменной в списке переменных и выбора во всплывающем меню вариантов **Категориальная** или **Количественная**. Установить какой-либо уровень измерения для переменной постоянным можно в представлении Переменные окна Редактора данных. Переменные с **номинальным** и **порядковым** уровнями измерения обрабатываются процедурой Настраиваемые таблицы как категориальные.

Категории. Когда вы выделяете в списке переменных категориальную переменную, категории данной переменной отображаются в списке Категории. Когда вы задаете данную переменную в таблице, эти категории выводятся на панели холста. Если переменная не имеет заданных категорий, то в списке категорий выводятся только две категории, задаваемые по умолчанию: *Категория 1* и *Категория 2*.

Категории, показанные в построителе таблиц, основаны на **метках значений**, т.е. описательных метках, присваиваемых различным значениям данных (к примеру, числовые значения 0 и 1 могут иметь, соответственно, метки значений *мужской* и *женский*). Метки значений можно задать на вкладке Переменные окна Редактора данных, или при помощи процедуры Задать свойства переменных в меню Данные окна Редактора данных.

Панель холста. Построение таблицы осуществляется посредством переноса переменных в строки и столбцы панели холста. Панель холста позволяет увидеть, как будет выглядеть таблица. Реальные значения данных не показаны в ячейках таблицы на панели холста, но при помощи панели холста вы всегда можете увидеть холст той таблицы, которую собираетесь построить. Для категориальных переменных построенная таблица может содержать больше категорий, чем таблица на панели макета, в том случае, если в файле данных переменные имеют значения, для которых не заданы метки значений.

- В **Нормальном** виде на панели макета показаны все строки и столбцы, которые будут включены в таблицу, включая строки и/или столбцы для итоговых статистик и категорий категориальных переменных.
- В **Компактном** представлении на панели макета показаны только переменные, используемые в таблице без возможности предварительного просмотра строк и столбцов в таблице.

Основные правила и ограничения процесса построения таблиц

- Для категориальных переменных итоговые статистики основываются на переменной на внутреннем уровне вложения в измерении статистик.
- Определение измерения статистик (строки или столбцы) для категориальных переменных основано на порядке переноса переменных на панель холста. К примеру, если первая переменная переносится в поле строк, то именно измерение строк станет измерением статистик.
- Количественные переменные могут подытоживаться только внутри категорий переменных на внутреннем уровне вложения как в измерении строк, так и в измерении столбцов. (Вы можете располагать количественную переменную на любом уровне в таблице, но подытоживание будет происходить на внутреннем уровне вложения).
- Количественные переменные не могут подытоживаться внутри других количественных переменных. Вы можете состыковывать несколько количественных переменных или подытоживать количественные переменные по категориям категориальных переменных. Но вы не можете вкладывать одну

количественную переменную в другую или располагать одну количественную переменную в измерении строк, а другую количественную переменную - в измерении столбцов.

- Если какая-либо переменная в активном наборе данных содержит более 12 000 определенных меток, пользователь не может использовать построитель таблиц для создания таблиц. Если нет необходимости включать переменные с превышением этого порога в пользовательские таблицы, можно определить и применить наборы, из которых эти переменные будут исключены. Если необходимо использовать какие-либо переменные с более чем 12000 определенными метками значений, воспользуйтесь командой STABLES для создания таких таблиц.

Построение таблицы

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. Перенесите одну или несколько переменных в поле строк и/или поле столбцов на панели холста.
3. Щелкните по **ОК**, чтобы создать таблицу.
4. Выберите (щелкните) переменную на панели холста.
5. Перенесите переменную за границы панели холста, или нажмите клавишу Delete.
Для изменения уровня измерения переменной:
 6. Щелкните правой кнопкой мыши по переменной в списке переменных (изменить тип переменной можно только в списке переменных, но не на панели холста).
 7. Во всплывающем меню выберите **Категориальная** или **Количественная**.

Состыковывание переменных

Состыковывание может рассматриваться как соединение отдельных таблиц в единую таблицу. Например, можно вывести в различных разделах одной таблицы информацию по переменной *Пол* и по переменной *Возрастная категория*.

Для того, чтобы состыковать переменные

1. Выберите в списке переменных все переменные, которые вы хотите состыковать, и перенесите их вместе в строки или столбцы панели холста.
или
2. Переносите переменные по отдельности, размещая каждую переменную ниже или выше переменных, уже перенесенных в измерение строк, или справа или слева переменных, уже перенесенных в измерение столбцов.

Таблица 3. Состыкованные категориальные переменные.

Переменные	Категории	Сводная статистика
Переменная 1	Категория 1	123
	Категория 2	456
Переменная 2	Категория 1	123
	Категория 2	456
	Категория 3	789

Дополнительную информацию смотрите в разделе “Состыковывание категориальных переменных” на стр. 26.

Вложение переменных

Вложение, подобно таблицам сопряженности, предназначено для выявления связи между двумя категориальными переменными. Его особенность состоит в том, что одна переменная вкладывается в другую в одном измерении таблицы. Например, можно вложить переменную *Пол респондента* в переменную *Возрастная категория* в измерении строк, чтобы количество мужчин и женщин было показано для каждой возрастной категории.

Вы также можете вложить количественную переменную в категориальную переменную. Например, вы можете вложить переменную *Доход респондента* в переменную *Пол респондента* для вывода отдельных средних значений (медиан или других итожащих статистик) дохода для мужчин и для женщин.

Для того чтобы вложить переменные,

1. Перенесите категориальную переменную в поле строк или поле столбцов панели холста.
2. Перетащите категориальную или количественную переменную в панель макета и поместите ее левее или правее от категориальной переменной строк или ниже или выше от категориальной переменной столбцов.

Таблица 4. Вложенные категориальные переменные.

Переменная 1	Переменная 2	Сводная статистика
Категория 1	Категория 1	12
	Категория 2	34
	Категория 3	56
Категория 2	Категория 1	12
	Категория 2	34
	Категория 3	56

Дополнительную информацию смотрите в разделе “Вложение категориальных переменных” на стр. 27.

Примечание: Настраиваемые таблицы не поддерживают послойную обработку (сравнение групп) при обработке файла разбиения. Чтобы добиться того же результата, как при послойной обработке файлов разбиения, расположите переменные файла разбиения в таблице в самых ближних внешних слоях вложения.

Слои

Слои позволяют добавлять в таблицы измерение "глубины", создавая, таким образом, трехмерные "кубы". На самом деле, использование слоев во многом похоже на использование вложения. Основное различие состоит в том, что в случае использования слоев в каждый момент времени в таблице показана только одна категория слоя. Например, если использовать *Возрастную категорию* в качестве переменной строк, а *Пол респондента* в качестве переменной слоев, то получится таблица, в которой информация о мужчинах и женщинах будет находиться в разных слоях.

Чтобы задать слой

1. Щелкните по **Слои** на вкладке Таблица построителя таблиц для просмотра списка Слои.
2. Перенесите количественную или категориальную переменные, которые будут задавать слой, в список Слои.

В списке переменных слоя вы не можете указывать одновременно и количественные, и категориальные переменные. Все переменные должны быть переменными одного и того же типа. Наборы множественных ответов в списке Слои обрабатываются как категориальные переменные. Количественные переменные в слоях всегда состыковываются.

Если вы задаете несколько категориальных переменных слоя, слой могут состыковываться или вкладываться друг в друга.

- Вариант **Каждая категория как слой** представляет собой аналог состыковывания. Для каждой категории каждой переменной выводится отдельный слой. Общее количество слоев равно *общему количеству* категорий переменных, указанных в списке слоев. К примеру, если вы задали три переменные слоя, и каждая из них имеет три категории, то количество слоев в таблице будет равно девяти.

- Вариант **Каждая комбинация категорий как слой** представляет собой аналог вложения слоев. Общее количество слоев равно *произведению* количества категорий всех переменных, указанных в списке слоев. К примеру, если вы задали три переменные слоя, и каждая из них имеет три категории, то количество слоев в таблице будет равно 27.

Показ и скрытие имен и/или меток переменных

Возможны следующие варианты вывода имен и меток:

- Выводить только метки переменных. Для переменных с заданными метками переменных будут выводиться метки переменных. Этот параметр выбран по умолчанию.
- Выводить только имена переменных.
- Выводить метки и имена переменных.
- Не выводить метки и имена переменных. Несмотря на то, что строка или столбец с меткой или именем переменной будут показаны в режиме предварительного просмотра, при построении таблицы они выводиться не будут.

Чтобы показать или скрыть метки или имена переменных:

1. Щелкните правой кнопкой мыши по переменной на панели холста.
2. Выберите во всплывающем меню **Показать метку переменной** или **Показать имя переменной**, чтобы включить или отключить вывод меток или имен переменных. Появляющийся переключатель означает, что метка или имя переменной будут выводиться в таблице.

Итоговые статистики

Диалоговое окно Итожащие статистики позволяет:

- Добавлять и удалять итожащие статистики из таблицы.
- Изменять метки статистик.
- Изменять порядок следования статистик.
- Изменять формат статистик, включая количество десятичных знаков.

Какие итожащие статистики (и другие параметры) являются доступными, зависит от уровня измерения переменной, для которой задаются итожащие статистики, как показано в верхней части диалогового окна. Источник итожащих статистик (переменная, для которой будут задаваться итожащие статистики) определяется при помощи следующих правил:

- **Уровень измерения** . Если таблица (или подтаблица в составной таблице) содержит количественную переменную, то итожащие статистики будут основываться на данной количественной переменной.
- **Порядок выбора переменных** . Определение измерения статистик (строки или столбцы) для категориальных переменных основано на порядке переноса переменных на панель холста. К примеру, если первая переменная переносится в поле строк, то именно измерение строк станет измерением статистик.
- **Вложение** . Для категориальных переменных итожащие статистики основываются на переменной на внутреннем уровне вложения в измерении статистик.

Составная таблица может иметь несколько переменных источников статистик (как количественных, так и категориальных), но каждая подтаблица имеет только один источник статистик.

Чтобы выбрать другое измерение в качестве источника статистик

1. Выберите соответствующее измерение (строки, столбцы или слой) в раскрывающемся списке **Расположение** в группе Итожащие статистики на вкладке Таблица.

Для управления итожащими статистиками, выводимыми в таблице

1. Выберите (щелкните) по переменной источника итожащих статистик на панели холста на вкладке Таблица.
2. Нажмите кнопку **Итожащие статистики** в группе Задать на вкладке Таблица.

или

- Щелкните правой кнопкой мыши по переменной источника статистик на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
- Выберите итожащие статистики, которые вы хотите добавить в таблицу. Чтобы переместить статистики из списка Статистики в список Показать, вы можете использовать стрелку или сделать это путем перетаскивания.
- Используйте кнопки со стрелками для изменения расположения выбранных статистик в таблице.
- Выберите формат вывода выбранных итожащих статистик при помощи выпадающего списка Формат.
- Введите в ячейку Десятичные количество десятичных знаков для выделенной итожащей статистики.
- Щелкните **Применить к выделенному** для применения заданных итожащих статистик к выделенным на панели холста переменным.
- Щелкните **Применить ко всем** для применения заданных итожащих статистик ко всем составным на панели холста переменным.

Примечание: **Применить ко всем** отличается от результата выбора варианта **Применить к выделенному** только в случае наличия на панели холста составных переменных одного и того же типа. В обоих случаях заданные итожащие статистики автоматически применяются к новым составным переменным того же типа, которые в дальнейшем добавляются в таблицу.

Итожащие статистики для категориальных переменных: Основными статистиками, доступными для категориальных переменных, являются количества и проценты. Для итогов и подытогов вы можете задать другие итожащие статистики. Задаваемые итожащие статистики включают в себя меры положения центра распределения (такие как среднее и медиана) и дисперсии (такие как стандартное отклонение), которые могут быть применены к некоторым порядковым категориальным переменным. Дополнительную информацию смотрите в разделе “Задаваемые статистики общих итогов для категориальных переменных” на стр. 10.

Частота. Количество наблюдений в каждой ячейке или количество ответов в наборе множественных ответов. Если используется взвешивание, данное значение представляет собой взвешенное количество.

- Если используется взвешивание, это значение представляет собой взвешенное количество.
- Взвешенное количество одинаково и для взвешивания с глобальным набором данных (**Данные > Взвесить наблюдения**), и для взвешивания с переменной эффективных базовых весов, заданной на вкладке Опции в диалоговом окне Пользовательские таблицы.

Невзвешенная частота. Невзвешенное количество наблюдений в каждой ячейке таблицы. Это значение отличается от количества наблюдений только в том случае, если учитывается взвешивание.

Скорректированное количество. Скорректированное количество, используемое при вычислениях эффективных базовых весов. Если вы не используете переменную эффективных базовых весов (вкладка Опции), скорректированное количество совпадает с количеством.

Проценты по столбцу. Проценты по каждому столбцу. Сумма процентов в каждом столбце подтаблицы (простые проценты) равна 100%. Проценты по столбцам обычно полезны в том случае, когда переменная *строк* является категориальной.

Проценты по строке. Проценты в каждой строке. Сумма процентов в каждой строке подтаблицы (простые проценты) равна 100%. Проценты по строкам обычно полезны в том случае, когда переменная *столбцов* является категориальной.

Проценты по строке и столбцу в слое. Сумма процентов по строке или столбцу (простые проценты) в каждой подтаблице таблицы с вложениями равна 100%. Если таблица содержит слои, то сумма процентов по строке или столбцу для всех вложенных подтаблиц в каждом слое равна 100%.

Проценты по слою. Проценты по слою. В случае с простыми процентами, сумма процентов в ячейках каждого слоя составляет 100%. Если у вас нет переменных слоев, это эквивалентно процентам по таблице.

Проценты по таблице. Проценты в каждой ячейке основаны на общем количестве наблюдений в таблице. Все процентные значения основаны на одном и том же общем количестве наблюдений, и их сумма равняется 100% (простые проценты).

Проценты по подтаблице. Проценты в каждой ячейке основаны на общем количестве наблюдений в подтаблице. Все проценты в ячейках в подтаблице основаны на одном и том же общем количестве наблюдений, и их сумма в подтаблице равняется 100%. В таблицах с вложениями, подтаблицы задаются переменной, предшествующей самому внутреннему уровню вложения. Например, в таблице, представляющей переменную *Семейное положение* внутри переменной *Пол респондента* внутри переменной *Возрастная категория*, переменная *Пол респондента* задает подтаблицы.

Доверительные интервалы

- Нижний и верхний доверительные пределы доступны для счетчиков, процентов, средних, медианы, процентилей и суммы.
- Текстовая строка "&[Confidence Level]" в метке включается в метки столбцов для доверительных уровней в таблице.
- Среднеквадратичная ошибка доступна для счетчиков, процентов, средних и суммы.
- Доверительные интервалы и среднеквадратичная ошибка недоступны для наборов множественных ответов.

Уровень

Доверительный уровень для доверительных интервалов, выраженный в процентах. Значение должно быть от 1 до 99.

Несколько наборов ответов

Для наборов множественных ответов можно получить проценты, основанные на количестве наблюдений, количестве ответов или количествах. Дополнительную информацию смотрите в разделе "Итожащие статистики для наборов множественных ответов" на стр. 9.

Составные таблицы

При расчете процентов каждая подтаблица, задаваемая отдельной стыкуемой переменной, рассматривается как отдельная таблица. Сумма процентов по строке в слое, процентов по столбцу в слое и процентов по таблице в каждой составной подтаблице составляет 100% (простые проценты). Базой для расчета процентов является количество наблюдений в каждой составной подтаблице.

База для расчета процентов: Проценты могут быть рассчитаны тремя различными способами, в зависимости от того, каким образом обрабатываются пропущенные значения:

Простые проценты. Проценты основаны на количестве наблюдений, использованных при построении таблицы, и их сумма всегда равна 100%. Если какая-либо категория исключена из таблицы, наблюдения, относящиеся к данной категории, исключаются из общей базы. Наблюдения с системными значениями отсутствия всегда исключаются из базы. Наблюдения с пользовательскими значениями отсутствия исключаются в том случае, если пользовательские категории отсутствия исключены из таблицы (по умолчанию), и включаются в том случае, если пользовательские категории отсутствия включены в таблицу. Любые проценты, которые не имеют в имени % *валидных* или *Итоговый %* являются простыми процентами.

Итоговые проценты. К базе простых процентов добавляются наблюдения с системными и пользовательскими значениями отсутствия. Сумма процентов может составлять менее 100%.

Допустимые проценты. Наблюдения с пользовательскими значениями отсутствия исключаются из базы простых процентов даже в том случае, если пользовательские категории отсутствия включены в таблицу.

Примечание: В отличие от пользовательских категорий отсутствия наблюдения из вручную исключенных категорий всегда исключаются из базы.

Итожащие статистики для наборов множественных ответов: Для наборов множественных ответов доступными являются следующие дополнительные статистики.

Процент ответов по строке/столбцу/ слою. Процент, основанный на ответах.

% ответов по столбцу/строке/слою (база: наблюдения). В числителе находится количество ответов, в знаменателе - количество респондентов.

% ответов по столбцу/строке/слою (база: ответы). В числителе находится количество, в знаменателе - итоговое количество ответов.

Процент ответов по строке/столбцу в слое. Процент по подтаблицам. Процент, основанный на ответах.

% ответов по столбцу/строке в слое (база: наблюдения). Проценты по подтаблицам. В числителе находится количество ответов, в знаменателе - количество респондентов.

% ответов по столбцу/строке в слое (база: ответы). Проценты по подтаблицам. В числителе находится количество, в знаменателе - итоговое количество ответов.

Ответы. Количество ответов.

Процент ответов по подтаблице/таблице. Процент, основанный на ответах.

% ответов по подтаблице/таблице (база: наблюдения). В числителе находится количество ответов, в знаменателе - количество респондентов.

% по подтаблице/таблице (база: ответы). В числителе находится количество, в знаменателе - итоговое количество ответов.

Итожащие статистики для количественных переменных и итогов категориальных переменных: В дополнение к количествам и процентам, доступным для категориальных переменных, для количественных переменных и задаваемых итогов/подытогов категориальных переменных доступны также следующие итожащие статистики. Эти итожащие статистики не доступны для наборов множественных ответов или текстовых (алфавитно-цифровых) переменных.

Среднее значение. Арифметическое среднее; сумма, деленная на число наблюдений.

Медиана. Значение, выше и ниже которого попадает половина наблюдений; 50-й процентиль.

Режим. Наиболее часто встречающееся значение. Если существует несколько модальных значений, используется наименьшее.

Минимум. Наименьшее значение.

Максимум. Наибольшее значение.

Пропущенные. Количество значений отсутствия (как пользовательских, так и системных).

Процентиль. Можно добавить в таблицу 5-й, 25-й, 75-й, 95-й и/или 99-й процентиль.

Диапазон. Разность между максимумом и минимумом.

Стандартное отклонение. Мера разброса вокруг среднего, выраженная в тех же единицах измерения, что и наблюдения. Равна корню квадратному из дисперсии. При нормальном распределении 68% наблюдений укладываются в одно стандартное отклонение от среднего, и 95% - в два стандартных отклонения. Если, например, средний возраст равен 45 годам со стандартным отклонением 10, то 95% наблюдений должны оказаться между 25 и 65 годами при нормальном распределении.

Сумма. Сумма значений.

Проценты суммы. Проценты, основанные на суммах. Доступны для отдельных строк и столбцов (в подтаблицах), строк и столбцов в целом (по подтаблицам), слоев, подтаблиц и таблиц.

Всего. Количество присутствующих значений, пользовательских значений отсутствия и системных значений отсутствия. Не включает в себя наблюдения из категорий, исключенных вручную и отличающихся от пользовательских категорий отсутствия.

Скорректированное общее число. Скорректированное общее число, используемое при вычислениях эффективных базовых весов. Если вы не используете переменную эффективных базовых весов (вкладка Опции), скорректированное общее число совпадает с общим числом. Этот статистический показатель недоступен для наборов множественных ответов.

Допустимые N. Число присутствующих значений. Не включает в себя наблюдения из категорий, исключенных вручную и отличающихся от пользовательских категорий отсутствия.

Скорректированное число допустимых. Скорректированное число допустимых, используемое при вычислениях эффективных базовых весов. Если вы не используете переменную эффективных базовых весов (вкладка Опции), скорректированное число допустимых совпадает с общим числом допустимых.

Дисперсия. Мера разброса относительно среднего значения. Равна сумме квадратов отклонений от среднего, деленной на число, на единицу меньше числа наблюдений. Дисперсия измеряется в единицах, которые равны квадратам единиц измерения самой переменной (и является квадратом стандартного отклонения).

Доверительные интервалы

- Нижний и верхний доверительные пределы доступны для счетчиков, процентов, средних, медианы, процентилей и суммы.
- Текстовая строка "&[Confidence Level]" в метке включается в метки столбцов для доверительных уровней в таблице.
- Среднеквадратичная ошибка доступна для счетчиков, процентов, средних и суммы.
- Доверительные интервалы и среднеквадратичная ошибка недоступны для наборов множественных ответов.

Уровень

Доверительный уровень для доверительных интервалов, выраженный в процентах. Значение должно быть от 1 до 99.

Составные таблицы

Каждая подтаблица, задаваемая стыкуемой переменной, рассматривается как отдельная таблица, и итоговые статистики подсчитываются соответствующим образом.

Задаваемые статистики общих итогов для категориальных переменных: В таблицах категориальных переменных, содержащих итоги и подытоги, статистики для итогов и подытогов могут отличаться от статистик, показанных для категорий. Например, можно выводить количества и проценты по столбцам для порядковой категориальной переменной строк и медиану в качестве итоговой статистики итога.

Для создания таблицы по категориальной переменной с задаваемой итоговой статистикой итога:

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
Откроется построитель таблиц.
2. Перенесите категориальную переменную в поле строк или столбцов панели холста.
3. Щелкните правой кнопкой мыши по переменной в холсте таблицы и выберите во всплывающем меню пункт **Категории и итоги**.
4. Включите переключатель **Итог**, затем нажмите кнопку **Применить**.
5. Еще раз щелкните правой кнопкой мыши по переменной на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
6. Включите переключатель **Задать статистики для итогов и подытогов**, затем задайте интересующие вас итожащие статистики.

По умолчанию, все итожащие статистики, включая задаваемые итожащие статистики, показываются в измерении, противоположном тому, где находится категориальная переменная. К примеру, если категориальная переменная находится в измерении строк, то итожащие статистики будут задавать столбцы, а именно:

Таблица 5. Категории порядковых переменных в строках, значения и средние итожащих статистик в столбцах.

Переменные	Категории	Количество	Среднее
Переменная 1	1 Согласен	196	2.29
	2 Не уверен	936	
	3 Не согласен	744	
	Всего	1876	

Для того, чтобы итожащие статистики показывались в измерении категориальной переменной:

7. В группе Итожащие статистики вкладки Таблица необходимо указать соответствующее измерение в раскрывающемся списке Расположение.

К примеру, если категориальная переменная выводится в строках, выберите **Строки** в данном раскрывающемся списке.

Формат вывода итожащих статистик: Доступны следующие форматы вывода:

nnnn. Простой числовой.

nnnn%. После значения добавляется знак процента.

Авто. Формат, задаваемый форматом самой переменной, включая количество десятичных знаков.

N=nnnn. Выводит $N=$ перед значением. Данный формат может быть удобен для количеств, *количества* валидных наблюдений и общего *количества* наблюдений в том случае, когда метки итожащих статистик не выводятся.

(nnnn). Все значения заключаются в скобки.

(nnnn)(отр.значение). Все отрицательные значения заключаются в скобки.

(nnnn%). Все значения заключаются в скобки, и в конце каждого значения добавляется знак процента.

n,nnn.n. Формат с запятой. Для разделения разрядов используется запятая, точка используется как разделитель целой и дробной части независимо от региональных установок.

n.nnn,n. Формат с точкой. Для разделения разрядов используется точка, запятая используется как разделитель целой и дробной части независимо от региональных установок.

\$n,nnn.n. Формат со знаком доллара. Перед значением выводится значок доллара; для разделения разрядов используется запятая, точка используется как разделить целой и дробной части независимо от региональных установок.

ССА, ССВ, ССС, ССD, ССЕ. Пользовательские денежные форматы. В списке показывается текущий установленный формат для каждого из 5 вариантов. Эти форматы устанавливаются на вкладке Валюта диалогового окна Параметры (меню Правка, Параметры).

Общие правила и ограничения

- За исключением формата Авто, количество десятичных знаков задается при помощи столбца Десятичн.
- За исключением форматов запятая, доллар и точка, используется разделитель целой и дробной части, установленный на панели управления региональными настройками Windows.
- Хотя в форматах запятая/доллар и точка знаки запятой и точки используются в качестве разделителя разрядов, не существует формата, в котором можно было бы задавать разделитель разрядов, установленный в региональных настройках Windows.

Категории и итоги

Диалоговое окно Категории и итоги позволяет:

- Изменять порядок следования и исключать категории.
- Вставлять итоги и подытоги.
- Вставьте вычисляемые категории.
- Включать или исключать пустые категории.
- Включать или исключать категории, обозначенные как категории пропущенных значений.
- Включать или исключать категории, не имеющие меток значений.
- Данное диалоговое окно доступно только для категориальных переменных и наборов множественных ответов. Оно недоступно для количественных переменных.
- В случае нескольких выделенных переменных с разными категориями нельзя вставлять подытоги, исключать категории или вручную изменять порядок следования категорий. Такая ситуация может возникнуть в том случае, если вы выбрали несколько переменных на панели холста и перешли в данное диалоговое окно. Все эти действия можно выполнять для каждой переменной по отдельности.
- В случае с переменными без меток значений можно только сортировать категории и вставлять итоги.

Для перехода в диалоговое окно Категории и итоги

1. Перенесите категориальную переменную или набор множественных ответов на панель холста.
2. Щелкните правой кнопкой мыши по переменной на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
или
3. Выделите (щелкните) переменную на панели холста, затем нажмите кнопку **Категории и итоги** в группе Задать на вкладке Таблица.
Вы также можете выделить несколько категориальных переменных в одном и том же измерении на панели холста:
4. Щелкните по очереди по всем переменным на панели холста, удерживая клавишу Ctrl.
или
5. Щелкните в любом месте вне таблицы на панели холста, затем щелкните и потяните, чтобы выделить область, которая содержит интересующие вас переменные.
или

- Щелкните правой кнопкой мыши по любой переменной в данном измерении и выберите пункт **Выделить все переменные [измерение]**, чтобы выделить все переменные в измерении.

Для того, чтобы изменить порядок следования категорий

Для того, чтобы вручную изменить порядок следования категорий:

- Выберите (щелкните) категорию в списке.
- Используйте кнопки со стрелками, чтобы перемещать категорию вверх или вниз в списке.
или
- Щелкните по столбцу Значение(я) для данной категории и переместите ее вручную.

Для исключения категорий

- Выберите (щелкните) категорию в списке.
- Используйте кнопку со стрелкой, чтобы переместить данную категорию в список Исключить.
или
- Щелкните по столбцу Значение(я) для данной категории и перетащите ее за пределы списка.

Если исключить хотя бы одну категорию, то категории без меток значений также будут исключены из таблицы.

Для того, чтобы отсортировать категории

Вы можете сортировать категории по значениям данных, меткам значений, количествам в ячейках или итоговым статистикам в порядке возрастания или убывания.

- В группе Сортировка категорий щелкните по раскрывающемуся списку По и выберите нужный способ сортировки: значения, метки, частота или итожащая статистика (такая как среднее, медиана или мода). Доступные итожащие статистики для сортировки зависят от итожащих статистик, выбранных для показа в таблице.
- Щелкните по раскрывающемуся списку Порядок, чтобы указать порядок сортировки (по возрастанию или по убыванию).

Если исключить из таблицы хотя бы одну категорию, то сортировка будет не доступна.

Подытоги

- Выделите в списке (щелкните) категорию, которая должна быть последней в диапазоне категорий, для которых вы хотите вставить подытог.
- Щелкните **Добавить подытог...**
- В диалоговом окне Задать подытог можно изменить текст метки подытога.
- Чтобы показать только подытог и подавить вывод категорий, которые задают подытог, поставьте переключатель **Скрыть категории**.
- Щелкните **Продолжить**, чтобы добавить подытог.

Итоги

- Отметьте переключатель **Итог**. Вы также можете изменять текст метки итога.

Если выбранная переменная вложена в другую переменную, то итоги будут вставляться для каждой подтаблицы.

Управление положением итогов и подытогов

Итоги и подытоги могут выводиться над и под категориями, включенными в итог/подытог.

- Если в группе Итоги выбран пункт **После категорий**, то итоги появляются над каждой подтаблицей, в каждый подытог включаются выбранная категория и все категории выше ее (но ниже любого предшествующего подытога).
- Если в группе Итоги выбран пункт **До категорий**, то итоги появляются под каждой подтаблицей, и в каждый подытог включаются выбранная категория и все категории ниже ее (но выше любого предшествующего подытога).

Внимание: Следует указать положение подытогов до задания каких-либо подытогов. Изменение данного параметра повлияет на расположение всех подытогов (а не только на выбранный подытог), а также *изменит наборы категорий, включенных в подытоги.*

Вычисляемые категории

Можно вывести категории, вычисленные из итожащей статистики, итогов, подытогов и/или констант. Дополнительную информацию смотрите в разделе “Вычисляемые категории”.

Задаваемые итожащие статистики итогов и подытогов

В итогах и подытогах можно задавать другие итожащие статистики, используя диалоговое окно Итожащие статистики. Дополнительную информацию смотрите в разделе “Итожащие статистики для категориальных переменных” на стр. 7.

Примечание: Если вы выберете несколько задаваемых статистик общих итогов, которые также будут находиться в теле таблицы, и скроете метки этих статистик, то вы, возможно, не сможете определить, что каждое из значений этих статистик представляет. Вообще, не является хорошей идеей при выборе нескольких статистик скрыть их метки.

Итоги, подытоги и исключенные категории

Наблюдения из исключенных категорий не используются при вычислении итогов.

Пропущенные значения, пустые категории и значения без меток значений

Пропущенные значения. Данный пункт позволяет управлять выводом **пользовательских пропущенных значений**, или значений, помеченных как пропущенные значения (к примеру, код 99, соответствующий варианту ответа “Вопрос не применим” на вопрос о беременности, задаваемый мужчинам). По умолчанию, пользовательские значения отсутствия исключаются. Выберите этот вариант (установите переключатель) для включения в таблицы пользовательских пропущенных категорий. Хотя переменная может содержать более одной категории пропущенных значений, на панели холста будет показана только одна категория пропущенных значений. Все пользовательские пропущенные категории будут включены в таблицу.

Системные значения отсутствия (пустые ячейки для числовых переменных в Редакторе данных) всегда исключаются из анализа.

Пустые категории. Пустыми категориями являются категории с заданными метками, но без наблюдений в конкретной таблице или подтаблице. По умолчанию, пустые категории включаются в таблицу. Чтобы отменить вывод пустых категорий, выключите переключатель (щелкните) в данном пункте

Другие значения, обнаруженные при просмотре данных. По умолчанию значения категорий в файле данных, не имеющие меток значений, автоматически включаются в таблицы. Чтобы исключить из таблицы значения без меток значений, выключите переключатель в данном пункте. Если вы исключите из таблицы хотя бы одну категорию с заданной меткой значения, все категории без меток значений также будут исключены из таблицы.

Вычисляемые категории

Помимо вывода агрегированных результатов итожащей статистики в таблице могут выводиться одна или более категорий, вычисленных на основе этих агрегированных результатов, значений констант, подытогов,

итогах или их комбинации. Эти результаты известны как вычисляемые категории или пост-категории. Вычисляемая категория действует как категория в одной переменной со следующими сходствами и различиями:

- Вычисляемая категория располагается как и другие категории.
- Вычисляемая категория оперирует с той же статистикой как и другие категории.
- Вычисляемая категория не влияет на подытоги, итоги или критерии значимости.
- По умолчанию значения вычисляемых категорий используют такое же форматирование итожащей статистики как и другие категории. Можно переопределить формат при задании вычисляемой категории.

Так как вычисляемые категории могут использоваться для подведения итогов агрегированных результатов, они могут быть похожи на подытоги. Однако вычисляемые категории имеют преимущества по сравнению с подытогами:

- Вычисляемые категории могут быть рассчитаны по результатам других промежуточных итогов.
- Вычисляемые категории могут накладываться друг на друга в одной и той же категории (или некоторых тех же категориях).
- Вычисляемые категории не должны включать значения для всех других категорий над или под вычисляемой категорией. Таким образом, вычисляемые категории являются неисчерпываемыми.
- Вычисляемые категории могут включать значения из категорий, которые не являются соседними.

В отличие от итогов и подытогов вычисляемые категории вычисляются на основе агрегированных данных, а не исходных данных. Поэтому значения вычисляемых категорий не совпадают с результатами итогов и подытогов. Кроме того, так как пользователь может скрыть категории источника при задании вычисляемой категории, может быть трудно интерпретировать подытоги в результирующей таблице. При использовании вычисляемых категорий рекомендуется указывать настраиваемые метки для подытогов.

Как задать вычисляемую категорию

Вычисляемые категории добавляются из диалогового окна Категории и итоги. Информацию о доступе к этому диалоговому окну смотрите в разделе “Категории и итоги” на стр. 12.

1. В диалоговом окне Категории и итоги щелкните **Добавить категорию...**
2. В **Метке для вычисляемой категории** укажите метку для вычисляемой категории. Можно перетащить категории из списка категорий, чтобы добавить метки для этих категорий.
3. Постройте выражение, выбрав категории и/или итоги и подытоги и используя операторы для определения вычисленных категорий. Также можно ввести константы (например, 500), что включить их в выражение.
4. Чтобы показать только вычисляемую категорию и подавить вывод категорий, которые задают вычисляемую категорию, поставьте переключатель **Скрыть категории**.
5. Чтобы изменить формат вывода и число знаков после запятой для вычисляемой категории, перейдите на вкладку **Формат вывода**. Дополнительную информацию смотрите в разделе “Форматы вывода для вычисленных категорий”.
6. Щелкните **Продолжить**, чтобы добавить вычисляемую категорию.

Форматы вывода для вычисленных категорий: По умолчанию, вычисляемые категории используют те же самые форматы вывода и число знаков после запятой, как и другие категории в переменной. Можно их переопределить на вкладке **Форматы вывода** в диалоговом окне **Задать вычисляемую категорию**. На вкладке **Форматы вывода** кроме форматов вывода и числа знаков после запятой, перечислены текущая суммарная статистика, по которой функционирует вычисляемая категория.

для каждой суммарной статистики можно выполнять следующие действия.

1. Выбирать формат вывода выбранных итоговых статистик в раскрывающемся списке **Формат**. Полный список форматов вывода смотрите в теме “Формат вывода итожащих статистик” на стр. 11.
2. Введите в ячейку **Десятичные количество десятичных знаков** для выделенной итожащей статистики.

Таблицы переменных с общими категориями (таблицы для переменных с одинаковыми вариантами ответа)

В опросах часто бывает много вопросов с одинаковыми наборами вариантов ответов. При помощи стыкования можно поместить все однотипные переменные в одну таблицу таким образом, чтобы одинаковые категории ответов были бы в ее столбцах.

Как построить таблицу для нескольких переменных с общими категориями

Таблица 6. Составные переменные с общими категориями ответов в столбцах

Переменные	Категория 1	Категория 2	Категория 3
Переменная 1	12	34	56
Переменная 2	56	12	34
Переменная 3	34	56	12

Дополнительную информацию смотрите в разделе “Таблицы для переменных с одинаковыми категориями” на стр. 42.

Настройка Построителя таблиц

В отличие от стандартных диалоговых окон, вы можете изменять размеры построителя таблиц точно так же, как вы изменяете размеры любого стандартного окна Windows:

- Щелкните по верхней, нижней, правой, левой границам окна или по любому углу и потяните, чтобы увеличить или уменьшить размер окна.
На вкладке Таблица также можно изменять размер списка переменных, списка категорий и панели холста.
- Щелкните и потяните горизонтальную границу между списком переменных и списком категорий, чтобы сделать списки длиннее или короче. Перемещение границы вниз удлинит список переменных и укоротит список категорий. Перемещение границы вверх позволит выполнить обратную операцию.
- Щелкните и потяните вертикальную границу между списком переменных, списком категорий и панелью холста, чтобы сделать списки шире или уже. Панель холста автоматически изменит свой размер.

Настраиваемые таблицы: вкладка Опции

Значения в ячейках

Позволяет управлять выводом в пустых ячейках и ячейках, для которых не вычисляются статистики.

Пустые ячейки

Для табличных ячеек, не содержащих наблюдений (значения в ячейках равны 0), можно выбрать одну из трех опций вывода: ноль, пустое значение или заданное вами текстовое значение. Размер текстового значения не может превышать 255 символов.

Статистика, которая не подлежит вычислению

Текст, который выводится в том случае, если статистика не может быть подсчитана (например, среднее значение для категории, в которой отсутствуют наблюдения). Размер текстового значения не может превышать 255 символов. Значение по умолчанию равно точке (.).

Ширина столбцов

Позволяет управлять минимальной и максимальной шириной столбцов. Данная установка не оказывает влияния на ширину столбцов с метками строк.

Как в шаблоне таблиц

Используются установки ширины столбца из Текущий шаблон таблиц по умолчанию. Вы можете создавать свои собственные шаблоны таблиц и задавать их в качестве шаблонов, используемых по умолчанию; также при помощи шаблона таблиц можно управлять шириной как столбцов данных, так и столбцов с метками строк.

Пользовательская

Позволяет изменять настройки ширины столбцов текущего шаблона таблиц. Задайте минимальное и максимальное значение для ширины столбцов данных в таблице, а также единицы измерения: пункты, дюймы или сантиметры.

Пропущенные значения для количественных переменных

В случае с построением таблиц, содержащих более одной количественной переменной, можно указать различные варианты обработки пропущенных значений.

Максимальное использование доступных данных (попарное исключение)

Для подсчета итоговых статистик для каждой из количественных переменных будут использованы все наблюдения с валидными значениями для этой переменной.

Использовать только наблюдения без пропущенных значений во всех количественных переменных (исключение целиком)

Наблюдения с пропущенными значениями хотя бы для одной из переменных, включенных в таблицу, не будут использованы при подсчете итоговых статистик для всех переменных.

эффективная база

Если у вас есть переменная, которая представляет корректировочные веса, а не частотные веса, можно использовать эту переменную в качестве переменной эффективных базовых весов. Концепция эффективных базовых весов или эффективных весов размера выборки основана на методах анализа данных из сложных выборок. Эффективный базовый вес позволяет выполнять приблизительную обработку статистического вывода в анализе, включающем произвольные корректировки данных простой случайной выборки при помощи корректировочных весов.

- Эффективный базовый вес влияет на значения взвешенной сводной статистики и критерии средних по столбцам и значимости пропорций столбцов.
- Если для набора данных включено взвешивание, переменная весов набора данных игнорируется, и результаты взвешиваются, как задает переменная эффективного базового веса.
- Переменная эффективного базового веса должна быть числовой.
- Наблюдения с отрицательными, нулевыми и отсутствующими значениями веса исключаются из всех результатов.

Подсчитывать повторяющиеся ответы для наборов множественных ответов

Дублирующий ответ - это один и тот же вариант ответа, встречающийся в двух или более переменных в наборе множественных ответов категориальных переменных. По умолчанию дублирующиеся ответы не учитываются.

Скрывать малые количества

Можно выбрать не показывать частоты, которые меньше заданного целого.

- Скрытые значения выводятся как $\leq N$, где N - заданное целое.
- Заданное целое должно быть больше или равно 2.
- Если для набора данных включено взвешивание или задана переменная эффективного базового веса, используется взвешенное значение.

Веса и округление

- Если вы используете **Данные > Взвесить наблюдения** для взвешивания наблюдений, нецелые веса округляются в ячейках или уровнях категорий для критериев значимости, доверительных интервалов и среднеквадратичных ошибок.
- Если вы выбрали **Использовать переменную эффективных базовых весов**, нецелые значения веса для выбранной переменной весов не округляются.
- Если заданы обе опции, используется переменная эффективных базовых весов, и нецелые веса не округляются.

Настраиваемые таблицы: вкладка Заголовки

Вкладка Заголовки позволяет управлять заголовками, подписями и угловыми метками.

Заголовок. Текст, выводимый над таблицей.

Подпись. Текст, выводимый под таблицей перед сносками.

Угол. Текст в левом верхнем углу мобильной таблицы. Угловой текст выводится только в том случае, если таблица содержит переменные строк и для меток строк сводных таблиц в шаблоне установлено свойство **В строках**. Этот параметр *не выбран* по умолчанию.

В заголовки, подписи и угловые метки можно также включать автоматически генерируемый текст:

Дата. Выводится текущий год, месяц и дата в формате, установленном в региональных настройках Windows.

Время. Выводится текущий час, минута и секунда в формате, установленном в региональных настройках Windows.

Выражение. Выводятся переменные, используемые в таблице, и порядок их использования. Если переменная имеет метку переменной, то выводится метка. Для обозначения того, каким образом переменные заданы в таблице, используются следующие символы:

- **+** означает составные переменные.
- **>** означает вложение.
- **ВУ** означает таблицу сопряженности или слою

Настраиваемые таблицы: вкладка Статистики критериев

Вкладка Статистические критерии содержит критерии значимости для пользовательских таблиц.

Эти критерии недоступны для таблиц, в которых метки категорий убраны из измерений таблицы по умолчанию, или для вычисленных категорий.

Критерии средних и пропорций по столбцам

Критерии средних по столбцам доступны для количественных переменных. Критерии пропорций по столбцам доступны для категориальных переменных.

Сравнение средних между столбцами

Парные критерии равенства средних по столбцам. У таблицы должна быть категориальная переменная в столбцах и количественная переменная на внутреннем уровне строк. Таблица должна включать среднее в качестве итожащей статистики.

Для отдельных категориальных переменных можно оценить дисперсию для всех категорий или только для сравниваемых категорий. Для переменных множественных ответов дисперсия для критерия средних всегда основана только на сравниваемых категориях.

Сравнить пропорции столбцов

Парные критерии равенства пропорций столбцов. В таблице должна быть по крайней мере одна категориальная переменная и по столбцам, и по строкам. Таблица должна содержать количества или процентные доли по столбцам.

Вывести значимые различия

Для критериев средних и пропорций по столбцам вы можете вывести значимые результаты в отдельной таблице или в основной таблице.

В отдельной таблице

Результаты применения критериев значимости показаны в отдельной таблице. Если отличия двух значений существенны, ячейка, соответствующая большему значению, содержит ключ, который идентифицирует столбец с меньшим значением.

Вывести значения значимости

Значения значимости выводятся в скобках после каждого значения ключа в ячейке. Эта опция доступна, только когда результаты значимости выводятся в отдельной таблице.

В главной таблице

Результаты критерия значимости выводятся в основной таблице. Каждая категория столбцов в таблице идентифицируется буквенным ключом. Для каждой значимой пары ключ категории с меньшим средним или пропорцией по столбцу выводится в категории с большим средним или пропорцией по столбцу.

- Когда вы ставите указатель на ключ в ячейке метки столбца в опорной таблице, все ячейки в таблице с этим ключом значимости выделяются. Для таблицы с несколькими переменными в измерении столбца выделяются только ячейки в этой подтаблице.
- Чтобы выбрать все ячейки в таблице (или подтаблице) с этим ключом значимости, щелкните правой кнопкой мыши по ячейке метки столбца и выберите **Выбрать > Выбрать все ячейки с этим ключом значимости**.

Использовать нижние индексы в стиле APA

Идентифицировать значимые различия при помощи форматирования в стиле APA, где используются буквенные нижние индексы. Если два значения существенно различны, для этих значений выводятся различные буквенные нижние индексы. Эти нижние индексы не являются сносками. Когда действует этот параметр, заданный стиль сносок в текущем шаблоне таблиц переопределяется, и сноски отображаются в виде чисел в верхнем индексе. Чтобы выделить все ячейки в одной строке с одним и тем же ключом значимости, щелкните правой кнопкой мыши по ячейке с ключом значимости и выберите **Выбрать ячейки с подобной значимостью**.

Уровни значимости

Уровень значимости для критериев средних и пропорций по столбцам.

- Значение должно быть от 0 до 1.
- Если вы задаете для уровня значимости, буквы верхнего регистра используются для идентификации значений значимости, не превосходящих меньшего уровня. Буквы нижнего регистра используются для идентификации значений значимости, не превосходящих большего уровня.
- Если вы выбрали **Использовать нижние индексы в стиле APA**, второе значение игнорируется.

Настроить р-значения на множественность сравнений

Поправка **Бонферрони** применяется для групповой вероятности ошибки (family-wise error rate, FWER). Метод **Беньямини-Хохберга** служит для контроля доли ложных отклонений гипотез (false discovery rate, FDR). Эти методы менее консервативны по сравнению с поправкой Бонферрони.

Проверка независимости (критерий хи-квадрат)

Критерий хи-квадрат независимости для таблиц, в которых по столбцам и по строкам существует хотя бы одна категориальная переменная.

Использовать подытоги вместо подытоженных категорий

Каждый подытог заменяет свои категории для проверки значимости. Если снят, то проверки значимости применяются лишь к тем подытогам, категории которых скрыты.

Включить в проверки переменные множественных ответов

Категории наборов множественных ответов включаются в критерии значимости. Если опция не выбрана, категории наборов множественных ответов не включаются в критерии значимости.

Простые таблицы с категориальными переменными

Простые таблицы с категориальными переменными

Большинство таблиц, которые вы будете создавать, скорее всего будут содержать, как минимум, одну **категориальную переменную**. Категориальная переменная - это переменная с ограниченным числом уникальных значений или категорий (например, пол или религия). Категориальные переменные могут быть либо **номинальные**, либо **порядковые**.

- **Номинальная.** Переменную можно рассматривать как номинальную, когда ее значения представляют категории без естественного упорядочения, например, подразделение компании, где работает наемный сотрудник. Примеры номинальных переменных включают регион, почтовый индекс или религию.
- **Порядковая.** Переменную можно рассматривать как порядковую, когда ее значения представляют категории с некоторым естественным для них упорядочением, например, уровни удовлетворенности обслуживанием от крайней неудовлетворенности до крайней удовлетворенности. Примеры порядковых переменных включают баллы, представляющие степень удовлетворенности или уверенности, или баллы, оценивающие предпочтение.

Значок рядом с каждой переменной позволяет идентифицировать тип переменной.

Таблица 7. Значки уровня измерения

	Числовой	Строка	Дата	Время
Количественная (непрерывная)		(не задается)		
Порядковый				
Номинальный				

Настраиваемые таблицы оптимизированы для использования категориальных переменных, имеющих **метки значений**. Дополнительную информацию смотрите в разделе “Построение таблиц” на стр. 1.

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Одна категориальная переменная

Хотя таблица с одной категориальной переменной и является простейшей таблицей, часто этот вариант таблицы может быть именно тем вариантом, который вам необходим.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.

На панели холста таблицы показана таблица в режиме предварительного просмотра. В режиме предварительного просмотра не показываются действительные значения данных - показываются только заготовки полей таблицы.

- Щелкните по **ОК**, чтобы создать таблицу.

Таблица показана в окне средства просмотра.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Рисунок 2. Одна категориальная переменная в строках

В этой простой таблице заголовок столбца *Количество* в действительности не нужен, и вы легко можете построить таблицу без этого заголовка.

- Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
- В группе Итожащие статистики поставьте переключатель **Скрыть**.
- Щелкните по **ОК**, чтобы создать таблицу.

Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Рисунок 3. Единственная категориальная переменная без метки статистики в столбце

Проценты

Кроме количеств вы можете также выводить в таблицах проценты. Для простой таблицы с одной категориальной переменной, если переменная расположена в строках, Вы, вероятно, захотите посмотреть на проценты по столбцам. Если же переменная выводится в столбцах, Вы, вероятно, захотите посмотреть проценты по строкам.

- Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
- В группе Итожащие статистики выключите переключатель **Скрыть**. Поскольку в таблице будет два столбца, то желательно вывести метки столбцов.
- Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
- В диалоговом окне Итожащие статистики выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
- В ячейке Метка списка Показать удалите метку, задаваемую по умолчанию, и наберите Процент.
- Нажмите кнопку **Применить к выделенному**, а затем щелкните **ОК** в построителе таблиц, чтобы создать таблицу.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%

Рисунок 4. Количества и проценты по столбцу

Итоги

Итоги автоматически не включаются в таблицу, однако их легко можно добавить.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
3. Отметьте (щелкните) элемент **Итог** в диалоговом окне Категории и итоги.
4. Нажмите кнопку **Применить** , а затем щелкните по **ОК** , чтобы создать таблицу .

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Рисунок 5. Количества, проценты по столбцу и итоги

Дополнительную информацию смотрите в разделе “Итоги и подытоги для категориальных переменных” на стр. 32.

Комбинационная таблица

Таблицы сопряженности предназначены для выявления взаимосвязей между двумя категориальными переменными. Например, используя *Возрастная категория* в качестве переменной строк и *Пол респондента* в качестве переменной столбцов, можно создать двумерную таблицу сопряженности, показывающую количество мужчин и женщин в каждой возрастной категории.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
4. Перенесите переменную *Пол респондента* из списка переменных в поле Столбцы панели холста. (Вы можете прокрутить список, чтобы найти данную переменную.)
5. Щелкните по **ОК** , чтобы создать таблицу.

		Gender	
		Male	Female
		Count	Count
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

Рисунок 6. Таблица сопряженности по переменным *Возрастная категория* и *Пол*

Проценты в таблицах сопряженности

В двумерной таблице сопряженности полезными могут быть проценты как по строке, так и по столбцу.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните правой кнопкой мыши по переменной *Пол респондента* на панели холста.
Вы можете заметить, что опция **Итожащие статистики** не активна во всплывающем меню. Причина этого заключается в том, что вы можете выбрать итожащие статистики только для самой внутренней переменной в измерении, являющемся источником статистик. Определение измерения статистик (строки или столбцы) для категориальных переменных основано на порядке переноса переменных на панель холста. В нашем примере первой мы перенесли переменную *Возрастная категория* и, поскольку больше нет никаких переменных в строках, то переменная *Возрастная категория* является источником статистик. Вы можете изменить измерение источника статистик, но в данном случае это не требуется. Дополнительную информацию смотрите в разделе “Итоговые статистики” на стр. 6.
3. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
4. В диалоговом окне **Итожащие статистики** выберите в списке **Статистики % по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список **Показать**.
5. Выберите в списке **Статистики % по строке** и нажмите кнопку со стрелкой, чтобы добавить ее в список **Показать**.
6. Нажмите кнопку **Применить к выделенному**, а затем щелкните **ОК** в построителе таблиц, чтобы создать таблицу.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65 or older	178	14.5%	37.2%	301	18.8%	62.8%

Рисунок 7. Таблица сопряженности с процентами по строке и столбцу

Управление форматом вывода таблицы

Вы можете управлять форматом вывода таблицы, включая число десятичных знаков в итожащих статистиках. Например, по умолчанию проценты выводятся с одним десятичным знаком после запятой и знаком процента. А что если вам нужно, чтобы значения в ячейках были с двумя десятичными знаками и без значка процента?

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.

- Для обеих выбранных статистик (% по столбцу и % по строке) выберите **ппп.п** в раскрывающемся списке и введите 2 в поле Десятичные.
- Щелкните по **ОК** , чтобы создать таблицу.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.79	44.63	134	8.38	55.37
	25 to 34	276	22.48	44.02	351	21.94	55.98
	35 to 44	309	25.16	45.51	370	23.13	54.49
	45 to 54	221	18.00	45.95	260	16.25	54.05
	55 to 64	136	11.07	42.50	184	11.50	57.50
	65 or older	178	14.50	37.16	301	18.81	62.84

Рисунок 8. Отформатированные ячейки с процентами по строке и столбцу

Маргинальные итоги

Достаточно часто в таблицах сопряженности выводят **маргинальные итоги** - итоги для каждой строки и каждого столбца. Поскольку в Настраиваемых таблицах такие итоги не включены по умолчанию, вам нужно явно задать их вывод в таблице.

- Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
- Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
- В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
- Перенесите переменную *Пол респондента* из списка переменных в поле Столбцы панели холста. (Вы можете прокрутить список, чтобы найти данную переменную.)
- Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
- Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить** .
- Щелкните правой кнопкой мыши по переменной *Пол респондента* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
- Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить** .
- В группе Итожащие статистики поставьте переключатель **Скрыть** . (Поскольку в таблице выводятся только количества, то нет необходимости специально идентифицировать, что показывается в ячейках таблицы.)
- Щелкните по **ОК** , чтобы создать таблицу.

		Gender		
		Male	Female	Total
Age category	Less than 25	108	134	242
	25 to 34	276	351	627
	35 to 44	309	370	679
	45 to 54	221	260	481
	55 to 64	136	184	320
	65 or older	178	301	479
	Total	1228	1600	2828

Рисунок 9. Таблица сопряженности с маргинальными итогами

Сортировка и исключение категорий

По умолчанию категории показаны в порядке возрастания значений данных, которые представляют метки значений. Например, хотя категории возраста имеют метки *Менее 25*, *от 25 до 34*, *от 35 до 44* и т.д., значения данных - это значения 1, 2, 3, :, и т.д., и именно эти значения данных определяют порядок вывода категорий.

Вы можете легко изменить порядок следования категорий, а также исключить категории, которые вы не хотите выводить в таблице.

Сортировка категорий

Вы можете вручную переупорядочить категории, а также отсортировать категории в возрастающем или убывающем порядке по:

- значениям данных,
- Метки значений.
- количествам в ячейках,
- Итожащие статистики. Доступные итожащие статистики для сортировки зависят от итожащих статистик, выбранных для показа в таблице.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Если переменная *Возрастная категория* еще не показана в области Строки панели холста, перетащите ее туда.
3. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.

В диалоговом окне показаны значения и метки значений в текущем порядке сортировки, в данном случае в порядке возрастания значений данных.

4. В группе Сортировка категорий в раскрывающемся списке Порядок выберите **Убывание** .
Теперь порядок сортировки изменен на обратный.
5. Выберите в раскрывающемся списке По **Метки** .
Теперь категории отсортированы в порядке убывания меток значений (в обратном алфавитном порядке).
Обратите внимание, что категория *Менее 25* находится вверху списка. При сортировке по алфавиту буквы идут после цифр. Поскольку это единственная метка, начинающаяся с буквы, и порядок сортировки обратный, то она находится вверху списка.
Если вы хотите, чтобы какая-то определенная категория находилась в другом месте списка, то вы можете легко переместить ее.
6. Щелкните по категории *Младше 25* в списке Метка.
7. Нажмите кнопку со стрелкой вниз. Категория перемещается на одну строку ниже в списке.
8. Щелкайте по кнопке со стрелкой вниз, пока категория не окажется в низу списка.

Исключение категорий

Если вы хотите, чтобы некоторые из категорий не появлялись в таблице, вы можете легко исключить их.

1. Щелкните по категории *Младше 25* в списке Метка.
2. Нажмите кнопку со стрелкой, чтобы переместить данную категорию в список Исключить.
3. Щелкните по категории *65 и старше* в списке Метка.
4. Снова нажмите кнопку со стрелкой, чтобы переместить данную категорию в список Исключить.
Две данные категории были перемещены из списка Показать в список Исключить. Если вы измените решение, то сможете снова вернуть эти категории в список Показать.
5. Нажмите кнопку **Применить** , а затем щелкните по **ОК** , чтобы создать таблицу .

		Gender		
		Male	Female	Total
Age category	55 to 64	136	184	320
	45 to 54	221	260	481
	35 to 44	309	370	679
	25 to 34	276	351	627
	Total	942	1165	2107

Рисунок 10. Таблица, отсортированная в порядке убывания меток значений с некоторыми исключенными категориями

Обратите внимание на то, что значения итогов меньше, чем они были ранее до исключения двух категорий. Итоги рассчитываются по категориям, включенным в таблицу. Все исключенные категории исключаются также и из итогов. Дополнительную информацию смотрите в разделе “Итоги и подытоги для категориальных переменных” на стр. 32.

Таблицы с категориальными переменными: состыковывание, вложение и слои

Состыковывание, вложение и слои представляют собой способы просмотра нескольких переменных в одной таблице. В данной главе рассматриваются вопросы применения этих способов в случае категориальных переменных, хотя все сказанное ниже может быть также применено и к количественным переменным.

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Состыковывание категориальных переменных

Состыковывание может рассматриваться как соединение отдельных таблиц в единую таблицу. Например, можно вывести в различных разделах одной таблицы информацию по переменной *Пол* и по переменной *Возрастная категория*.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите переменную *Пол респондента* из списка переменных в поле Строки панели холста.
3. Перенесите переменную *Возрастная категория* из списка переменных в область Строки ниже переменной *Пол респондента*.
Теперь эти две переменные состыкованы в измерении строк.
4. Щелкните по **ОК**, чтобы создать таблицу.

		Count
Gender	Male	1232
	Female	1600
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Рисунок 11. Таблица составных по строкам категориальных переменных

Аналогичным образом можно состыковывать переменные по столбцам.

Состыковывание и таблицы сопряженности

Составная таблица может включать в себя и другие переменные в других измерениях. К примеру, можно построить таблицу сопряженности, которая будет содержать две составные переменные в строках и третью переменную в измерении столбца.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Если переменные *Возрастная категория* и *Пол респондента* не состыкованы по строкам, состыкуйте их, следуя предыдущим инструкциям.
3. Переместите переменную *Получают новости из Интернета* из поля строк в поле Столбцы на панели холста.
4. Щелкните по **ОК** , чтобы создать таблицу.

		Get news from internet	
		No	Yes
		Count	Count
Gender	Male	873	359
	Female	1092	508
Age category	Less than 25	146	96
	25 to 34	368	259
	35 to 44	435	244
	45 to 54	346	135
	55 to 64	252	68
	65 or older	416	63

Рисунок 12. Таблица сопряженности с двумя составными переменными строк и одной переменной столбцов

Примечание: В файле данных присутствуют несколько переменных с меткой, начинающейся со слов *Получают новости...*, поэтому их не всегда легко различить в списке переменных (поскольку метки переменных не полностью показываются в списке переменных). Увидеть метку переменной полностью можно двумя способами:

- При перемещении курсора мыши на переменную появится всплывающая метка переменной.
- Щелкните и потяните вертикальную границу между списком переменных и списком категорий и панелью холста, чтобы сделать списки шире.

Вложение категориальных переменных

Вложение, подобно таблицам сопряженности, предназначено для выявления связи между двумя категориальными переменными. Его особенность состоит в том, что одна переменная вкладывается в другую в одном измерении таблицы. Например, можно вложить переменную *Пол респондента* в переменную *Возрастная категория* в измерении строк, чтобы количество мужчин и женщин было показано для каждой возрастной категории.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.

3. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
4. Перенесите переменную *Пол респондента* из списка переменных на правую границу переменной *Возрастная категория* в поле Строки.

Предварительный просмотр таблицы показывает, что таблица с вложением включает в себя только один столбец, каждая ячейка которого показывает количество мужчин или женщин в конкретной возрастной категории.

Вы можете заметить, что метка переменной *Пол респондента* повторяется для каждой возрастной категории. Минимизировать количество повторений в таблице можно, указав переменную с наименьшим количеством категорий на внешнем уровне вложения.

5. Щелкните по метке переменной *Пол респондента* на панели холста.
6. Переместите переменную на левый край измерения Строки.
Теперь метка переменной *Пол респондента* не повторяется шесть раз; вместо этого всего лишь дважды повторяется метка переменной *Возрастная категория*. Таблица стала выглядеть менее перегруженной и вместе с тем содержит те же самые результаты.
7. Щелкните по **ОК**, чтобы создать таблицу.

				Count
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
			Female	Age category
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

Рисунок 13. Таблица с вложением переменной *Возрастная категория* в переменную *Пол респондента*

Примечание: Настраиваемые таблицы не поддерживают послойную обработку (сравнение групп) при обработке файла разбиения. Чтобы добиться того же результата, как при послойной обработке файлов разбиения, расположите переменные файла разбиения в таблице в самых ближних внешних слоях вложения.

Отключение меток переменных

Другой способ борьбы с часто встречающимися метками переменных в таблицах с вложениями заключается в обычном отключении показа имен переменных или их меток. Поскольку метки значений для переменных *Пол респондента* и *Возрастная категория* являются достаточно информативными сами по себе, мы можем отключить показ меток этих переменных.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выключите переключатель **Показать метку переменной** во всплывающем меню.
3. Сделайте то же самое для переменной *Пол респондента*.
Метки переменных по-прежнему показаны в окне мастера, но они не будут выводиться в таблице.
4. Щелкните по **ОК**, чтобы создать таблицу.

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Рисунок 14. Таблица с вложением без меток переменных

Если вы хотите, чтобы метки переменных были бы включены в таблицу, но без многократного вывода в самой таблице, то вы можете включить их в заголовок таблицы или в угловую метку.

- Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
- Щелкните по вкладке **Заголовки** .
- Щелкните в текстовом окне Заголовок.
- Нажмите кнопку **Выражение** . В текстовом окне Заголовок появится текст $&[Выражение]$. Данная команда будет генерировать заголовок таблицы, включающий в себя метки переменных, задействованных в таблице.
- Щелкните по **ОК** , чтобы создать таблицу.

Gender > Age category

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Рисунок 15. Метки переменных в заголовке таблицы

Знак (>) в заголовке показывает, что переменная *Возрастная категория* вложена в переменную *Пол респондента* .

Таблица сопряженности с вложениями

Таблица с вложениями может включать в себя и другие переменные в других измерениях. К примеру, вы можете вложить переменную *Возрастная категория* в переменную *Пол респондента* в измерении строк и построить таблицу сопряженности с третьей переменной в измерении столбцов.

- Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
- Если *Возрастная категория* еще не вложена в *Пол* в строках, следуйте указаниям в предыдущем примере, чтобы вложить ее.
- Переместите переменную *Получают новости из Интернета* из поля строк в поле Столбцы на панели холста.

Можно заметить, что таблица слишком велика и не видна полностью на панели холста. Вы можете прокрутить панель холста вверх/вниз или вправо/влево, чтобы увидеть всю таблицу, или:

- Щелкнуть **Компактный** в построителе таблиц, чтобы представить таблицу в компактном виде. В этом случае в мастере таблиц будут показаны только метки переменных, без информации о категориях или итожащих статистиках, включенных в таблицу.
- Увеличить размер построителя таблиц, щелкнув и растянув границы окна.

4. Щелкните по **ОК**, чтобы создать таблицу.

				Get news from internet	
				No	Yes
				Count	Count
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
			35 to 44	217	92
			45 to 54	169	52
			55 to 64	112	24
	Female	Age category	65 or older	155	23
			Less than 25	87	47
			25 to 34	209	142
			35 to 44	218	152
			45 to 54	177	83
		55 to 64	140	44	
		65 or older	261	40	

Рисунок 16. Таблица сопряженности с вложениями

Изменение положения строк и столбцов: Что вы бы сделали в том случае, если потратили бы много времени на разработку сложной таблицы, а затем решили бы, что необходимо поменять ориентацию таблицы -- переместить все переменные строк в столбцы и наоборот? К примеру, вы построили таблицу сопряженности с вложением переменной *Возрастная категория* в переменную *Пол респондента* в строках, а теперь хотите, чтобы эти две демографические переменные были бы показаны в столбцах.

- Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
- Щелкните правой кнопкой мыши в любом месте на панели холста и выберите во всплывающем меню меню пункт **Поменять местами переменные строк и столбцов**.
Переменные строк и столбцов поменяются местами.
Перед созданием таблицы необходимо произвести несколько изменений, чтобы таблица получилась более компактной.
- Выберите **Скрыть** для отключения вывода метки итожащих статистик в столбцах.
- Щелкните правой кнопкой мыши на переменной *Пол респондента* на панели холста и отмените выделение пункта **Показать метку переменной**.
- Теперь щелкните **ОК**, чтобы создать таблицу.

		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

Рисунок 17. Таблица сопряженности с вложениями -- демографические переменные в столбцах

Слои

Слои позволяют добавлять в таблицы измерение "глубины", создавая, таким образом, трехмерные "кубы". На самом деле, использование слоев во многом похоже на использование вложения или состыковывания. Основное различие состоит в том, что в случае слоев в каждый момент времени видна только одна

категория слоя. Например, если использовать *Возрастную категорию* в качестве переменной строк, а *Пол респондента* в качестве переменной слоев, то получится таблица, в которой информация о мужчинах и женщинах будет находиться в разных слоях.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
4. Щелкните по **Слой** в верхней части построителя таблиц для просмотра списка Слои.
5. Перенесите переменную *Пол респондента* из списка переменных в список слоев.

Можно заметить, что добавление переменной в слой не оказало никакого влияния на внешний вид таблицы, изображенной на панели холста. Переменные слоя оказывают влияние на изображение на панели холста только в том случае, если она является переменной статистик и пользователь изменяет итожащие статистики.

6. Щелкните по **ОК** , чтобы создать таблицу.

		Count
Age category	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178

Рисунок 18. Простая таблица со слоями

На первый взгляд, эта таблица не сильно отличается от обычной таблицы, построенной по одной категориальной переменной. Единственное отличие заключается в наличии метки *Пол респондента Мужской* в верхней части таблицы.

7. Дважды щелкните по таблице в окне средства просмотра, чтобы активировать ее.
8. Обратите внимание на то, что метка *Пол респондента Мужской* в действительности является раскрывающимся списком.
9. Щелкните по стрелке раскрывающегося списка, чтобы просмотреть весь список слоев.
В данной таблице существует возможность выбора только двух слоев.
10. Выберите *Пол респондента Женский* из раскрывающегося списка.

		Count
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Рисунок 19. Простая таблица со слоями, в которой теперь показан другой слой

Две составные категориальные переменные слоя

Если в качестве переменных слоя указать более одной переменной, то эти переменные могут быть как состыкованы, так и вложены друг в друга. По умолчанию, переменные слоя состыковываются. (*Примечание:* Если среди ваших переменных есть масштабные переменные слоев, переменные слоев можно только состыковать).

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .

2. Если у вас еще нет *Возрастной категории* в строках и *Пола* в слоях, следуйте указаниям в предыдущем примере, чтобы создать многослойную таблицу.
3. Перенесите переменную *Уровень образования* из списка переменных в список слоев под переменную *Пол респондента* .
Можно заметить, что две радио-кнопки, располагающиеся ниже списка слоев, стали активными. По умолчанию отмечен вариант **Каждая категория как слой** . Данный вариант является аналогом состыковывания.
4. Щелкните по **ОК** , чтобы создать таблицу.
5. Дважды щелкните по таблице в окне средства просмотра, чтобы активировать ее.
6. Щелкните по стрелке раскрывающегося списка, чтобы просмотреть весь список слоев.

В этой таблице есть семь слоев: два слоя для категорий *Пол респондента* и пять слоев для пяти категорий *Уровень образования*. Для случая с составными слоями общее количество слоев равняется сумме категорий переменных слоя (включая категории итогов и подытогов, присутствующие в переменных слоя).

Две вложенные категориальные переменные слоя

Использование процедуры вложения при указании переменных слоя создает отдельный слой для каждой комбинации категорий переменных слоя.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Если вы еще не сделали это, следуйте указаниям в предыдущем примере, чтобы создать таблицу наложенных слоев.
3. В группе Вывод слоя выберите вариант **Каждая комбинация категорий как слой** . Данный вариант является аналогом вложения.
4. Щелкните по **ОК** , чтобы создать таблицу.
5. Дважды щелкните по таблице в окне средства просмотра, чтобы активировать ее.
6. Щелкните по стрелке раскрывающегося списка, чтобы просмотреть весь список слоев.

Теперь в таблице присутствуют десять слоев (для того, чтобы увидеть все слои, необходимо прокрутить ниже в раскрывающемся списке), по одному для каждой комбинации переменных *Пол респондента* и *Уровень образования* . Для слоев с вложениями общее количество слоев равно *произведению* количества категорий переменных слоя (в данном примере, $5 \times 2 = 10$).

Печать слоев таблиц: По умолчанию, на печать выводится только один, видимый в данный момент слой. Для печати всех слоев в таблице необходимо сделать следующее:

1. Дважды щелкните по таблице в окне средства просмотра, чтобы активировать ее.
2. В меню окна Viewer выберите:
Формат > Свойства таблицы...
3. Щелкните по вкладке **Печать** .
4. Выберите **Печатать все слои** .

Вы также можете отметить данные установки при изменении свойств шаблонов таблиц, включая и задаваемый по умолчанию шаблон таблиц.

Итоги и подытоги для категориальных переменных

Вы можете включать в настраиваемые таблицы итоги и подытоги. Итоги и подытоги могут быть применены к категориальным переменным на любом уровне вложенности в любом измерении -- строках, столбцах и слоях.

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Простой итог для одной переменной

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В строителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
3. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
4. В диалоговом окне Итожащие статистики выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
5. В ячейке Метка списка Показать удалите метку, задаваемую по умолчанию, и наберите Процент .
6. Нажмите кнопку **Применить к выделенному** .
7. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
8. Отметьте (щелкните) элемент **Итог** в диалоговом окне Категории и итоги.
9. Нажмите кнопку **Применить** , а затем щелкните **ОК** , чтобы создать таблицу .

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
Total		2828	100.0%

Рисунок 20. Простой итог для одной категориальной переменной

То, что вы видите, зависит от того, что включается в итог

Итоги основаны на категориях, показанных в таблицах. Если вы исключите из таблицы некоторые категории, то наблюдения, входящие в эти категории, не будут использованы при подсчете итога.

1. Откройте строитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
3. Щелкните по категории *Младше 25* в списке Метка.
4. Нажмите кнопку со стрелкой, чтобы переместить данную категорию в список Исключить.
5. Щелкните по категории *65 и старше* в списке Метка.
6. Снова нажмите кнопку со стрелкой, чтобы переместить данную категорию в список Исключить.
Две данные категории были перемещены из списка Показать в список Исключить.
7. Нажмите кнопку **Применить** , а затем щелкните по **ОК** , чтобы создать таблицу .

		Count	Percent
Age category	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%
	Total	2107	100.0%

Рисунок 21. Итог в таблице с исключенными категориями

Значение итога в данной таблице равняется 2107, что меньше значения 2828, которое появлялось в случае включения в таблицу всех категорий. В итог включаются только те категории, которые используются в таблице. (Итог в процентах по-прежнему равен 100%, поскольку проценты основываются на общем количестве наблюдений, использованных в таблице, а не на общем количестве наблюдений в файле данных.)

Управление расположением итогов

По умолчанию итоги показаны под подытоживаемыми категориями. Вы можете изменить расположение итогов и задать вывод итогов над подытоживаемыми категориями.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
3. В группе Итоги и подытоги выберите **Над категориями, к которым они относятся**.
4. Нажмите кнопку **Применить** , а затем щелкните по **ОК** , чтобы создать таблицу .

		Count	Percent
Age category	Total	2107	100.0%
	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%

Рисунок 22. Вывод итога над подытоживаемыми категориями

Итоги для таблиц с вложениями

Поскольку итоги могут быть применены к категориальным переменным на любом уровне вложенности, вы можете создавать таблицы, содержащие групповые итоги на различных уровнях вложенности.

Групповые итоги

Итоги для категориальных переменных, вложенных в другие категориальные переменные, называются групповыми итогами.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Перетащите переменную *Пол респондента* к левой границе переменной *Возрастная категория* на панели холста.
3. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
Перед созданием таблицы переместим итоги в нижнюю часть таблицы, под подытоживаемые категории.
4. В группе Итоги и подытоги выберите **Над категориями, к которым они относятся**.
5. Нажмите кнопку **Применить** , чтобы сохранить заданные параметры и вернуться в диалоговое окно Настраиваемые таблицы.
6. Щелкните по **ОК** , чтобы создать таблицу.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%

Рисунок 23. Итоги переменной *Возрастная категория* внутри категорий переменной *Пол респондента*

Теперь таблица содержит два групповых итога: один для мужчин и один для женщин.

Общие итоги

Итоги, примененные к вложенным переменным, всегда являются групповыми итогами, а не общими итогами. Если вы хотите получить итоги для всей таблицы, то вам следует применить итоги к переменной на внешнем уровне вложенности.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните правой кнопкой мыши по переменной *Пол респондента* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
3. Отметьте (щелкните) элемент **Итог** в диалоговом окне Категории и итоги.
4. Нажмите кнопку **Применить**, а затем щелкните по **ОК**, чтобы создать таблицу.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%
	Total	Age category	25 to 34	627	29.8%
			35 to 44	679	32.2%
			45 to 54	481	22.8%
			55 to 64	320	15.2%
			Total	2107	100.0%

Рисунок 24. Общие итоги в таблице с вложениями

Отметим, что общий итог составляет 2107, а не 2828. Причина этого заключается в том, что две возрастные категории по-прежнему исключены из таблицы, поэтому наблюдения в этих категориях не учитываются при подсчете итога.

Итоги для переменных слоя

Итоги для переменных показаны в отдельных слоях таблицы.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните по **Слой** в построителе таблиц для просмотра списка Слои.
3. Переместите переменную *Пол респондента* из поля строк в холсте таблицы в список Слои.

Примечание: Поскольку вы уже указали итоги для переменной *Пол респондента*, повторно делать этого не нужно. Перемещение переменной из одного измерения в другое не изменяет установки, сделанные для данной переменной.

4. Щелкните по **ОК** , чтобы создать таблицу.
5. Дважды щелкните по таблице в средстве просмотра, чтобы активировать ее.
6. Щелкните на стрелке раскрывающегося списка слоя, чтобы просмотреть список всех слоев таблицы.

В таблице - три слоя: *Пол респондента Мужской*, *Пол респондента Женский* и *Пол респондента Итого*.

Управление расположением итогов по слоям

Для итогов переменной в слоях расположение итога (вверху или внизу) определяет расположение слоя итогов. К примеру, если вы укажете для итога переменной слоя **Вверху**, итоговый слой будет показан на первом месте.

Подытоги

Вы можете указывать подытоги для подгрупп категорий переменной. К примеру, вы можете указать подытоги для переменной возраста, представляющие респондентов в возрасте до и после 45 лет.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
4. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
5. Выберите **3,00** в списке Значения.
6. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
7. В поле Текст метки введите Подытог < 45.
8. Затем щелкните по **Продолжить**.
В таблицу будет добавлена строка, содержащая подытог для первых трех возрастных категорий.
9. Выберите **6,00** в списке Значения.
10. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
11. В поле Метка введите Промежуточный итог 45+.
12. Затем щелкните по **Продолжить**.

Важное замечание: Перед заданием подытогов необходимо указать расположение итогов и подытогов (**Над категориями, к которым они относятся** или **Под категориями, к которым они относятся**). Изменение данного параметра повлияет на расположение всех подытогов (а не только на выбранный подытог), а также *изменит наборы категорий, включенных в подытоги*.

13. Нажмите кнопку **Применить** , а затем щелкните по **ОК** , чтобы создать таблицу .

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Subtotal < 45	1548
	45 to 54	481
	55 to 64	320
	65 or older	479
	Subtotal 45+	1280

Рисунок 25. Подытоги для переменной *Возрастная категория*

То, что вы видите, зависит от того, что включается в подытог

Как и в случае с итогами, подытоги основываются на категориях, показанных в таблицах.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.

Примечание: Значение (не метка), выводимое для первого подытога равно **1,00...3,00** , указывая на то, что этот подытог включает все значения в списке, лежащие между 1 и 3.

3. Выберите **1,00** в списке Значения (или щелкните по метке *Менее 25*).
4. Нажмите кнопку со стрелкой, чтобы переместить данную категорию в список Исключить.

Поскольку первая возрастная категория была исключена, для первого подытога отображается значение **2,00...3,00** . Это указывает на тот факт, что исключенные категории не учитываются в подытогах, поскольку подытоги основаны только на категориях, отображаемых в таблицах. Исключение категории автоматически исключает ее из всех подытогов, поэтому вы не можете, к примеру, показать только подытоги без категорий, на которых основываются данные подытоги.

Соккрытие категорий, задающих подытог

Можно подавить вывод категорий, которые задают подытог, и вывести только подытоги, не повлияв на учет категорий, входящих в подытог.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
4. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
5. Выберите **3,00** в списке Значения.
6. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
7. В поле Метка введите Менее чем 45.
8. Поставьте переключатель **Скрыть в таблице подытоженные категории**.
9. Затем щелкните по **Продолжить**.
В таблицу будет добавлена строка, содержащая подытог для первых трех возрастных категорий.
10. Выберите **6,00** в списке Значения.
11. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
12. В поле Метка введите 45 или старше.
13. Выберите **Скрыть в таблице подытоженные** .
14. Затем щелкните по **Продолжить**.
15. Чтобы включить итог с подытогами, поставьте переключатель **Итог** в группе параметров Показывать.
16. Нажмите кнопку **Применить** .

Панель холста отражает тот факт, что подытоги будут выведены, но категории, задающие подытоги, будут исключены.

17. Нажмите **ОК**, чтобы создать эту таблицу.

		Count
Age category	Less than 45	1548
	45 or older	1280
	Total	2828

Рисунок 26. Таблица, показывающая только подытоги и итоги

Подытоги переменных слоя

Как и в случае с итогами, подытоги для переменных слоя показаны в виде отдельных слоев таблицы. По существу, подытоги являются категориями. Каждая категория создает отдельный слой в таблице, и порядок вывода категорий определяется порядком категорий, указанным в диалоговом окне Категории и итоги, включая место расположения категорий промежуточных итогов.

Вычисляемые категории для категориальных переменных

Можно включить вычисляемые категории в пользовательские таблицы. Это новые категории, которые вычисляются из категорий той же переменной на любом уровне вложения во всех измерениях: строке, столбце или уровне. Например, можно включить вычисленную категорию, которая показывает разницу между двумя категориями.

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Простая вычисляемая категория

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
3. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
4. Выберите **3,00** в списке Значения.
5. Щелкните пункт **Добавить категорию**, чтобы открыть диалоговое окно Определение вычисляемой категории.
6. В поле Текст метки для вычисляемой категории введите **Меньше чем 45**.
7. Выберите **Меньше чем 25 (1,00)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории. В выражении показывается [1].
8. Нажмите кнопку оператора плюс (+) в данном диалоговом окне (или нажмите клавишу + на клавиатуре).
9. Выберите **От 25 до 34 (2,00)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
10. Нажмите кнопку оператора плюс (+) в данном диалоговом окне (или нажмите клавишу + на клавиатуре).
11. Выберите **От 35 до 44 (3,00)** в списке "Категории" и нажмите кнопку стрелки, чтобы копировать в текстовое поле "Выражение для вычисляемой категории".
12. Затем щелкните **Продолжить**.
В таблицу будет добавлена строка, содержащая подытог для первых трех возрастных категорий.
13. Выберите **5,00** в списке Значения.
14. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
15. В поле Метка введите **Меньше чем 65**.
16. Затем щелкните **Продолжить**.
В таблицу будет добавлена строка, содержащая подытог для первых пяти категорий.
17. Нажмите кнопку **Применить** , а затем щелкните по **ОК** , чтобы создать таблицу .

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Less than 45	1548
	45 to 54	481
	55 to 64	320
	Less than 65	2349
	65 or older	479

Рисунок 27. Вычисляемая категория с подытогом

В данную таблицу входит вычисляемая категория (*Меньше чем 45*) и подытог (*Меньше чем 65*). Данный подытог включает категории, которые также включены в вычисляемую категорию. Невозможно создать такую же таблицу только с подытогами, поскольку подытоги не могут быть общими для одних категорий.

Скрытие категорий в вычисляемой категории

Как в случае с подытогами, можно подавить вывод категорий, которые используются в выражении вычисляемой категории и показываются только в самой вычисляемой категории. Следующий пример основан на предыдущем.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
3. В списке Значения выберите вычисляемую категорию *Меньше чем 45*.
4. Щелкните пункт **Изменить**, чтобы показать диалоговое окно Определение вычисляемой категории.
5. Выберите пункт **Скрыть в таблице используемые в выражении категории**.
6. Затем щелкните по **Продолжить**.
7. В списке Значения выберите подытог *Меньше чем 65*.
8. Щелкните пункт **Изменить**, чтобы показать диалоговое окно Определение подытога.
9. Установите переключатель **Скрыть в таблице подытоженные категории**.
10. Затем щелкните по **Продолжить**.
11. Нажмите кнопку **Применить**, а затем щелкните по **ОК**, чтобы создать таблицу.

		Count
Age category	Less than 45	1548
	Less than 65	2349
	65 or older	479

Рисунок 28. Вычисляемая категория с подытогом и скрытыми категориями

Так же, как и в предыдущем примере, в таблицу включена вычисляемая категория и подытог. Однако в этом случае категории скрыты. Таким образом, показаны только данные итоги.

Ссылка на подытоги в вычисляемой категории

Можно включить подытоги в выражение вычисляемой категории.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Занятость респондента* из списка переменных в поле Строки панели холста.
4. Перенесите переменную *Семейное положение* из списка переменных в поле Столбцы.

5. Щелкните правой кнопкой мыши по переменной *Занятость респондента* на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
6. Выберите **2** в списке Значения.
7. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
8. В поле Текст метки введите *Работающий*.
9. Установите переключатель **Скрыть в таблице подытоженные категории**.
10. Затем щелкните по **Продолжить**.
В таблицу будет добавлена строка, содержащая подытог для первых двух рабочих категорий.
11. Выберите **8** в списке Значения.
12. Щелкните **Добавить подытог**, чтобы открыть диалоговое окно Задать подытог.
13. В поле Текст метки введите *Неработающий*.
14. Выберите **Скрыть категории, задающие подытоги**.
15. Затем щелкните по **Продолжить**.
В таблицу будет добавлена строка, содержащая подытог для других рабочих категорий.
16. В списке Значения выберите подытог *Неработающий*.
17. Щелкните пункт **Добавить категорию**, чтобы открыть диалоговое окно Определение вычисляемой категории.
18. В поле Текст метки для вычисляемой категории введите *Работающий / Неработающий*.
19. Выберите **Работающий (Работающий #1)** в списке Итоги и подытоги и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
20. Нажмите кнопку оператора деления (/) в данном диалоговом окне (или нажмите клавишу / на клавиатуре).
21. Выберите **Неработающий (Неработающий #2)** в списке Итоги и подытоги и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
По умолчанию в вычисляемой категории используется тот же формат, что и в статистике переменной, которая в данном случае Количество. Поскольку необходимо показать десятичные знаки, которые являются результатом деления в выражении вычисляемой категории, а формат по умолчанию для статистики Количество не включает десятичных знаков, то необходимо изменить формат.
22. Откройте вкладку **Форматы вывода**.
23. Измените настройку **Десятичные знаки** для статистики **Частота** на значение **2**.
24. Затем щелкните по **Продолжить**.
25. Нажмите кнопку **Применить**, а затем щелкните по **ОК**, чтобы создать таблицу.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Count	Count	Count	Count	Count
Labor force status	Working	916	64	330	67	494
	Not Working	429	219	116	26	169
	Working / Not Working	2.14	.29	2.84	2.58	2.92

Рисунок 29. Вычисляемая категория, которая показывает отношение подытогов

Данная таблица включает два подытога и вычисляемую категорию. Вычисляемая категория показывает отношение подытогов. Поэтому можно легко сравнить группы, представленные каждым подытогом. Отношение работающих и вдовых неработающих респондентов намного ниже по сравнению с другими группами. Также, отношение женатых респондентов немного ниже. Возможно, это связано с респондентами, которые покинули работу, чтобы оставаться дома с детьми.

Использование вычисляемых категорий для вывода неисчерпаемых подытогов

Подытоги являются неисчерпаемыми. Это означает, что все подытоги в таблице включают все значения ниже или выше их позиций в таблице. Вычисляемые категории, с другой стороны, не являются исчерпаемыми и позволяют суммировать смешанные категории в таблице.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Самовосприятие в качестве либерала или консерватора* из списка переменных в поле Строки панели холста.
4. Щелкните правой кнопкой мыши по переменной *Самовосприятие в качестве либерала или консерватора* на панели холста таблицы и выберите в контекстном меню пункт **Категории и итоги**.
5. Выберите **3** в списке Значения.
6. Щелкните пункт **Добавить категорию**, чтобы открыть диалоговое окно Определение вычисляемой категории.
7. В поле Текст метки для вычисляемой категории введите Подытог по либералам. Обратите внимание на четыре пробела перед текстом. Эти пробелы используются для отступов в результирующей таблице.
8. Выберите **Крайний либерал (1)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
9. Нажмите кнопку оператора плюс (+) в данном диалоговом окне (или нажмите клавишу + на клавиатуре).
10. Выберите **Либерал (2)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
11. Нажмите кнопку оператора плюс (+) в данном диалоговом окне (или нажмите клавишу + на клавиатуре).
12. Выберите **Мягкий либерал (3)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
13. Нажмите кнопку **Продолжить** .
В таблицу будет добавлена строка, содержащая подытог для либеральных категорий.
14. Выберите **7** в списке Значения.
15. Щелкните пункт **Добавить категорию**, чтобы открыть диалоговое окно Определение вычисляемой категории.
16. В поле Текст метки для вычисляемой категории введите Подытог по консерваторам. Обратите внимание на четыре пробела перед текстом. Эти пробелы используются для отступов в результирующей таблице.
17. Выберите **Мягкий консерватор (5)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
18. Нажмите кнопку оператора плюс (+) в данном диалоговом окне (или нажмите клавишу + на клавиатуре).
19. Выберите **Консерватор (6)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
20. Нажмите кнопку оператора плюс (+) в данном диалоговом окне (или нажмите клавишу + на клавиатуре).
21. Выберите **Крайний консерватор (7)** в списке Категории и нажмите кнопку стрелки, чтобы копировать в текстовое поле Выражение для вычисляемой категории.
22. Нажмите кнопку **Продолжить** .
В таблицу будет добавлена строка, содержащая подытог для консервативных категорий.
23. Нажмите кнопку **Применить** , а затем щелкните **ОК** , чтобы создать таблицу .

		Count
Think of self as liberal or conservative	Extremely liberal	64
	Liberal	357
	Slightly liberal	351
	Liberal Subtotal	772
	Moderate	986
	Slightly conservative	432
	Conservative	415
	Extremely conservative	86
	Conservative Subtotal	933

Рисунок 30. Вычисляемые категории, показывающие неисчерпаемые подытоги

В таблицу входят две вычисляемых категории, которые не включают все категории, показанные в таблице. Категория *Умеренный* не включена ни в одну из двух вычисляемых категорий. Невозможно создать такую же таблицу с подытогами, поскольку подытоги исчерпаемы.

Таблицы для переменных с одинаковыми категориями

В опросах часто бывает много вопросов с одинаковыми наборами вариантов ответов. Например, в рассматриваемом образце опроса содержится несколько переменных, относящихся к конфиденциальности в различных общедоступных и частных предприятиях и на службах, у всех этих вопросов одинаковый набор категорий ответов: 1 = *Полностью доверяю*, 2 = *Скорее доверяю* и 3 = *Не доверяю*. При помощи состыковывания можно поместить все однотипные переменные в одну таблицу таким образом, чтобы одинаковые категории ответов были бы в ее столбцах. Эти функции доступны также, если использовать вычисляемые категории с учетом того, что любая вычисляемая метка категории и любое выражение одинаковы во всех переменных.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Рисунок 31. Таблицы переменных с одинаковыми категориями

Примечание: В предыдущей версии модуля Настраиваемые таблицы эта таблица называлась "таблица частот."

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Таблица количеств

1. Выберите в меню:

Анализ > Таблицы > Настраиваемые таблицы...

- Выберите в списке переменных в построителе таблиц переменную *Доверие банкам...* и затем щелкните левой кнопкой мыши с одновременным удержанием клавиши Shift на переменной *Доверие телевидению*, чтобы выделить все переменные "доверия". (*Примечание:* Предполагается, что метки переменных расположены в алфавитном порядке, а не в том порядке, в каком переменные идут в файле.)
- Перенесите шесть переменных доверия в поле Строки в холсте таблицы.
Переменные состыкованы в измерении строк. По умолчанию, метки категорий для каждой переменной также показаны в строках, что приводит к получению очень длинной, узкой таблицы (6 переменных x 3 категории = 18 строк), но, поскольку все шесть переменных имеют одни и те же метки категорий (метки значений), то вы можете переместить метки категорий в измерение столбцов.
- Выберите в раскрывающемся списке Расположение категорий **Метки строк в столбцах** .
Теперь таблица содержит только шесть строк, по одной для каждой составной переменной, а категории определяют столбцы в таблице.
- Перед созданием таблицы выберите **Скрыть** на панели Статистики, поскольку метка статистики *Частота* не является необходимой в данной таблице.
- Щелкните по **ОК** , чтобы создать таблицу.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Рисунок 32. Таблица с составными переменными строк и общими метками категорий в столбцах

Наряду с таблицей, содержащей переменные в строках, а категории в столбцах, вы можете создавать таблицы, содержащие составные переменные в столбцах, а категории - в строках. Это является целесообразным в том случае, когда количество категорий превышает количество переменных, хотя в нашем случае количество переменных превышает количество категорий.

Таблица процентов

В таблице, содержащей составные переменные по строкам и категории в столбцах, лучше всего в качестве процентных значений использовать проценты по строкам. (Для таблицы, содержащей составные переменные в столбцах и категории в строках, вы можете указать проценты по столбцам.)

- Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
- Щелкните правой кнопкой мыши на любой переменной доверия на панели холста и выберите во всплывающем меню **Итожащие статистики**.
- Выберите в списке Статистики **% по строке** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
- Щелкните по любой ячейке в строке *Количество* в списке Показать и нажмите кнопку со стрелкой, чтобы убрать данную статистику из списка Показать.
- Щелкните **Применить ко всем** , чтобы применить данную статистику ко всем составным переменным в таблице.

Примечание: Если ваша таблица не выглядит как таблица на данном рисунке, то возможно, вы щелкнули **Применить к выделенному** вместо **Применить ко всем**, тем самым применяя новую итожащую статистику только к выделенной переменной. В данном примере это может сказаться на двух столбцах для каждой категории: в одном для всех других переменных выводятся значения по умолчанию, а в другом для

выбранной переменной выводится значение по умолчанию для процентной доли по строкам. В результате получилась бы именно такая таблица, но наша задача заключается в построении *другой* таблицы.

6. Щелкните по **ОК** , чтобы создать таблицу.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	26.3%	57.3%	16.4%
Confidence in education	27.2%	56.1%	16.7%
Confidence in major companies	27.5%	59.2%	13.3%
Confidence in medicine	45.0%	46.1%	8.9%
Confidence in press	9.5%	47.2%	43.4%
Confidence in television	10.4%	49.9%	39.7%

Рисунок 33. Таблица процентов по строкам для переменных, составных в строках, категории показаны в столбцах

Примечание: В таблицу с одинаковыми для разных переменных категориями вы можете включать любое количество итоговых статистик. В наших примерах мы будем использовать при каждом построении таблицы только одну статистику.

Итоги и управление категориями

Вы можете создавать таблицы с выводом категорий в измерении, противоположном измерению переменных, только в том случае, если все переменные в таблице имеют одни и те же категории, упорядоченные одним и тем же образом. Это касается итогов, промежуточных итогов и любых других изменений в категориях. Это означает, что любые изменения, которые вы делаете в диалоговом окне Категории и итоги, должны быть сделаны для всех переменных, которые имеют те же самые категории.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните правой кнопкой мыши по первой переменной доверия на панели холста и выберите во всплывающем меню **Итожащие статистики**.
3. Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить**.

Первое, что сразу бросается в глаза - метки категорий были перемещены из столбцов обратно в строки. Вы также можете отметить, что управление расположением категорий стало невозможным. Причина этого заключается в том, что теперь переменные не обладают одним и тем же набором категорий. Одна из переменных теперь имеет категорию Итого.

4. Щелкните правой кнопкой мыши по любой переменной доверия в холсте таблиц и выберите во всплывающем меню вариант **Выбрать все переменные строк** или щелкните правой кнопкой мыши, удерживая клавишу Ctrl, по каждой составной переменной на панели холста, пока все переменные не будут выбраны (для этого может потребоваться прокрутка панели вниз или расширение окна построителя).
5. Щелкните **Категории и Итоги** на панели Определить.
6. Если вариант **Итог** все еще не выбран (не отмечен) в диалоговом окне Категории и итоги, то выберите его и щелкните **Применить**.
7. Раскрывающийся список Расположение категорий снова стал доступен, поскольку теперь все переменные имеют дополнительную итоговую категорию, поэтому выберите **Метки строк в столбцах**.
8. Щелкните по **ОК** , чтобы создать таблицу.

	A great deal	Only some	Hardly any	Total
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%
Confidence in television	10.4%	49.9%	39.7%	100.0%

Рисунок 34. Таблица процентов по строкам для переменных, составных в строках, категории и итоговые значения показаны в столбцах

Вложенность в таблицах с переменными с одинаковыми категориями

В таблицах с вложениями составные переменные с одинаковыми категориями должны находиться на внутреннем уровне вложенности их измерения, если вы хотите показать метки категорий в противоположном измерении.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Перенесите переменную *Пол респондента* из списка переменных на левую границу поля Строки. Составные переменные с одинаковыми категориями сейчас вложены в категории переменной пола в мастере таблиц.
3. Теперь перенесите переменную *Пол респондента* на правую границу одной из составных переменных доверия в окне мастера таблиц.

Метки категорий снова были перенесены в измерение строк, а управление расположением категорий стало невозможным. Теперь вы имеете одну составную переменную, в которую вложена переменная *Пол респондента* , в то время как другие составные переменные не имеют вложенных переменных. Конечно, вы можете добавить переменную *Пол респондента* в качестве вложенной переменной и к другим составным переменным, но последующая попытка перемещения меток строк в столбцы приведет к появлению в столбцах меток категорий переменной *Пол респондента*, а не меток категорий составных переменных с одинаковыми категориями. Причина этого состоит в том, что переменная *Пол респондента* теперь является вложенной на внутреннем уровне переменной и изменение положения категорий будет применено именно к переменной на внутреннем уровне вложения.

Итоговые статистики

Итожащие статистики включают в себя самые различные статистики, начиная от простых количеств для категориальных переменных и заканчивая мерами дисперсии, такими как стандартная ошибка среднего значения для количественных переменных. Они *не* включают в себя статистические критерии, доступные на вкладке Статистические критерии в диалоговом окне Настраиваемые таблицы. Дополнительную информацию смотрите в разделе “Статистические критерии” на стр. 57.

Итожащие статистики для категориальных переменных и наборов множественных ответов включают в себя количества и различные процентные значения, а именно:

- Проценты по строке
- Проценты по столбцу
- Проценты по подтаблице
- Проценты по таблице
- Валидные проценты

В дополнение к итожащим статистикам для категориальных переменных, итожащие статистики для количественных переменных и итоги для категориальных переменных включают в себя:

- Среднее
- Медиана
- Процентили
- Сумма
- Стандартное отклонение
- Диапазон
- Минимальное и максимальное значение

Для наборов множественных ответов доступными являются и дополнительные статистики. Доступен также полный список итожащих статистик. Дополнительную информацию смотрите в разделе “Итоговые статистики” на стр. 6.

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Переменная - источник итожащих статистик

В дополнение к вышесказанному, возможность выбора итожащих статистик зависит от уровня измерения переменной источника итожащих статистик. Источник итожащих статистик (переменная, для которой будут задаваться итожащие статистики) определяется при помощи следующих правил:

- **Уровень измерения** . Если таблица (или подтаблица в составной таблице) содержит количественную переменную, то итожащие статистики будут основываться на данной количественной переменной.
- **Порядок выбора переменных** . Определение измерения статистик (строки или столбцы) для категориальных переменных основано на порядке переноса переменных на панель холста. К примеру, если первая переменная переносится в поле строк, то именно измерение строк станет измерением статистик.
- **Вложение** . Для категориальных переменных итожащие статистики основываются на переменной на внутреннем уровне вложения в измерении статистик.

Составная таблица может иметь несколько переменных источников статистик (как количественных, так и категориальных), но каждая подтаблица имеет только один источник статистик.

Источник итожащих статистик для категориальных переменных

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
3. Щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**. (Поскольку таблица содержит только одну переменную, эта переменная является переменной - источником статистик.)
4. В диалоговом окне Итожащие статистики выберите в списке Статистики *% по столбцу* и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
5. Нажмите кнопку **Применить к выделенному** .
6. В построителе таблиц перенесите переменную *Получают новости из Интернета* вправо от переменной *Возрастная категория* на панели холста.

7. Щелкните правой кнопкой мыши по переменной *Возрастная категория* еще раз. Пункт **Итожащие статистики** во всплывающем меню теперь недоступен, поскольку переменная *Возрастная категория* не находится на самом внутреннем уровне вложения в измерении итожащих статистик.
8. Щелкните правой кнопкой мыши по *Получают новости из Интернета* на панели холста. Пункт **Итожащие статистики** является доступным, поскольку данная переменная является переменной - источником итожащих статистик, находясь на самом внутреннем уровне вложения в измерении источника итожащих статистик. (Так как у этой таблицы есть только строки с одним измерением, это измерение источника статистик).
9. Переместите переменную *Получают новости из Интернета* из поля строк в поле столбцов на панели холста.
10. Снова щелкните правой кнопкой мыши по *Получают новости из Интернета* на панели холста. Пункт **Итожащие статистики** во всплывающем меню теперь недоступен, поскольку данная переменная больше не находится в измерении источника статистик.

Переменная *Возрастная категория* теперь опять является переменной источника статистик, поскольку измерением источника статистик для категориальных переменных, задаваемым по умолчанию, является первое по порядку измерение, в которое вы помещаете переменные при построении таблицы. В данном примере, мы вначале поместили переменные в измерение строк. Таким образом, измерение строк является измерением источника статистик, задаваемым по умолчанию, а поскольку переменная *Возрастная категория* является единственной переменной в данном измерении, эта переменная является источником статистик.

Источник итожащих статистик для количественных переменных

1. Переместите количественную переменную *Число часов в день, проводимых у телевизора* левее переменной *Возрастная категория* в поле строк на панели холста.
Первое, на что нужно обратить внимание, - итожащие значения *Количество и % по столбцу* были заменены на *Среднее*, и если щелкнуть правой кнопкой мыши по *Число часов в день, проводимых у телевизора* на панели холста, вы увидите, что теперь это переменная источника итожащей статистики. Для таблиц с количественной переменной, именно количественная переменная всегда является источником итожащих статистик, вне зависимости от уровня вложения или измерения, и итожащей статистикой, задаваемой по умолчанию для количественных переменных, является среднее значение.
2. Перенесите переменную *Число часов в день, проводимых у телевизора* из поля строк на панели макета в поле столбцов выше переменной *Получают новости из Интернета*.
3. Щелкните правой кнопкой мыши по *Число часов в день, проводимых у телевизора* и во всплывающем меню выберите **Итожащая статистика**. (Данная переменная по-прежнему является источником статистик, несмотря на то, что вы перенесли ее в другое измерение.)
4. В диалоговом окне **Итожащие статистики** щелкните на поле **Формат** для среднего значения в группе **Показать** и выберите **pppp** из раскрывающегося списка. (Вы можете прокрутить список, чтобы найти данный вариант.)
5. В поле **Десятичн** наберите 2.
6. Нажмите кнопку **Применить к выделенному**.
Предварительный просмотр таблицы на панели холста позволяет сделать вывод, что средние значения теперь выводятся с двумя десятичными знаками.
7. Щелкните по **ОК**, чтобы создать таблицу.

		Hours per day watching TV	
		Get news from internet	
		No	Yes
		Mean	Mean
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

Рисунок 35. Количественная переменная, подытоженная в категориях таблицы сопряженности категориальных переменных

Составные переменные

Поскольку составная таблица может содержать несколько переменных - источников статистик, и вы для каждой из этих переменных можете указать различные итожащие статистики, существует несколько специальных рекомендаций по поводу указания итожащих статистик в составных таблицах.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. Щелкните по переменной *Получают новости из газет* в списке переменных, а затем, удерживая клавишу Shift, щелкните по переменной *Получают новости из телевидения* в списке переменных, чтобы выделить все переменные, связанные с получением новостей. (*Примечание:* Предполагается, что метки переменных расположены в алфавитном порядке, а не в том порядке, в каком переменные идут в файле.)
4. Перенесите пять "новостных" переменных в поле Строки панели холста.
Теперь пять "новостных" переменных состыкованы в измерении строк.
5. Щелкните по переменной *Получают новости из Интернета* на панели холста таким образом, чтобы выделена была только эта переменная.
6. Теперь щелкните правой кнопкой мыши по переменной *Получают новости из Интернета* и выберите во всплывающем меню **Итожащие статистики**.
7. В диалоговом окне Итожащие статистики выберите в списке Статистики *% по столбцу* и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать. (Для переноса выбранных статистик из списка Статистики в список Показать воспользуйтесь кнопкой со стрелкой или просто перетащите выбранные статистики из списка Статистики в список Показать).
8. Затем нажмите кнопку **Применить к выделенному** .
Для процента по столбцам в таблице был добавлен еще один столбец, но предварительный просмотр таблицы на панели холста позволяет сделать вывод, что проценты по столбцам показаны только для одной переменной. Причина заключается в том, что данная таблица является составной и может иметь несколько переменных - источников статистик, и каждая из этих переменных может иметь разные итожащие статистики. В данном примере, однако, мы хотим вывести одни и те же итожащие статистики для всех переменных.
9. Щелкните правой кнопкой мыши по переменной *Получают новости из газет* на панели холста и выберите во всплывающем меню **Итожащие статистики**.
10. В диалоговом окне Итожащие статистики выберите в списке Статистики *% по столбцу* и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
11. Затем нажмите кнопку **Применить ко всем** .

Теперь предварительный просмотр таблицы показывает, что процент по столбцам выводится для всех составных переменных.

Задаваемые статистики общих итогов для категориальных переменных

Для категориальных переменных источников статистик, вы можете сами задать статистики общих итогов, которые будут отличаться от итожащих статистик, подсчитываемых для категорий переменной. К примеру, для порядковой переменной вы можете вывести процент для каждой категории и среднее или медиану как задаваемую статистику общих итогов.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Нажмите кнопку **Сброс** , чтобы очистить заданные ранее в диалоговом окне параметры.
3. Щелкните по переменной *Доверие к прессе* в списке переменных, а затем, удерживая клавишу Ctrl, щелкните по переменной *Доверие к телевидению* , чтобы выбрать две переменные.
4. Перенесите две выбранные переменные в поле строки на панели холста. Они будут состыкованы в измерении строк.
5. Щелкните правой кнопкой мыши по переменной на панели холста и выберите во всплывающем меню пункт **Выделить все переменные строк**. (Эти переменные уже могут быть одновременно выделены, но мы хотим быть в этом уверенными.)
6. Снова щелкните правой кнопкой мыши по переменной на панели макета таблицы и выберите во всплывающем меню пункт **Категории и итоги**.
7. Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить** .
Предварительный просмотр таблицы на панели холста позволяет сделать вывод, что для каждой переменной в таблице добавилась строка Итог. Для того чтобы вывести задаваемые статистики общих итогов, итоги и/или подытоги должны быть заданы в данной таблицы.
8. Щелкните правой кнопкой мыши по любой переменной на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
9. В диалоговом окне Итожащие статистики щелкните по статистике *Количество* в списке Показать, а затем щелкните по стрелке, чтобы удалить данную статистику из списка Показать.
10. Выберите в списке Статистики *% по столбцу* и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
11. Щелкните (отметьте) **Задать статистики для итогов и подытогов**.
12. В диалоговом окне Итожащие статистики щелкните по статистике *Количество* в списке Показать, а затем щелкните по стрелке, чтобы удалить данную статистику из списка Показать.
13. Щелкните *Среднее* в списке задаваемых статистик и нажмите кнопку со стрелкой, чтобы добавить данную статистику в список Показать.
14. Щелкните на поле **Формат** для среднего значения в группе Показать и выберите **nnnn** из раскрывающегося списка. (Вы можете прокрутить список, чтобы найти данный вариант.)
15. В поле Десятичн наберите 2.
16. Щелкните **Применить ко всем** , чтобы применить данные установки к обеим переменным в таблице.
Для задаваемых статистик общих итогов был добавлен новый столбец, и это делает таблицу не очень удобной для анализа, поскольку она содержит в себе большое количество пустых ячеек.
17. В построителе таблиц в группе Итожащие статистики выберите **Строки** в раскрывающемся списке Расположение.
Данная операция перемещает все итожащие статистики в измерение строк, и названия итожащих статистик появляются в таблице в виде отдельного столбца.
18. Щелкните по **ОК** , чтобы создать таблицу.

Confidence in press	A great deal	Column N %	9.5%
	Only some	Column N %	47.2%
	Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	A great deal	Column N %	10.4%
	Only some	Column N %	49.9%
	Hardly any	Column N %	39.7%
	Total	Mean	2.29

Рисунок 36. Категориальные переменные с задаваемыми статистиками общих итогов статистик

Вывод значений категорий

С только что полученной таблицей связана одна небольшая проблема -- у пользователя могут возникнуть трудности с интерпретацией среднего значения, поскольку ему непонятно, на основе каких значений было получено данное среднее значение. Является ли среднее значение, равное 2.34, чем-то средним между *Полностью доверяю* и *Скорее доверяю*, или чем-то средним между *Скорее доверяю* и *Не доверяю*?

Хотя мы не сможем решить данную проблему непосредственно в настраиваемых таблицах, она может быть решена более общим способом.

1. Выберите в меню:
Правка > Опции...
2. В диалоговом окне Параметры щелкните по вкладке **Метки в выводе**.
3. В группе Метки в сводных таблицах выберите **Метки и значения** в раскрывающемся списке **Значения переменных в метках**.
4. Нажмите кнопку **ОК**, чтобы сохранить изменения.
5. Откройте построитель таблиц (Анализ, Таблицы, Настраиваемые таблицы) и снова щелкните **ОК** для создания таблицы.

Confidence in press	1 A great deal	Column N %	9.5%
	2 Only some	Column N %	47.2%
	3 Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	1 A great deal	Column N %	10.4%
	2 Only some	Column N %	49.9%
	3 Hardly any	Column N %	39.7%
	Total	Mean	2.29

Рисунок 37. Значения и метки категорий переменных

Значения, соответствующие категориям, позволяют сделать вывод, что среднее значение 2,34 соответствует промежутку между *Скорее доверяю* и *Не доверяю*. Вывод в таблице значений категорий существенно упрощает интерпретацию значений таких задаваемых статистик общих итогов, как среднее значение.

Данное изменение в установках является глобальным изменением, которое будет постоянно влиять на сводные таблицы, появляющиеся в результате выполнения различных процедур, до тех пор, пока мы снова не изменим его. Для того чтобы в таблицах снова показывались только метки значений, необходимо сделать следующее:

6. Выберите в меню:
Правка > Опции...
7. В диалоговом окне Параметры щелкните по вкладке **Метки в выводе**.
8. В группе Метки в сводных таблицах выберите **Метки** в раскрывающемся списке **Значения переменных в метках**.
9. Нажмите кнопку **ОК**, чтобы сохранить изменения.

Подытоживание количественных переменных

Подытоживание количественных переменных

Для количественных переменных существует большое количество итожащих статистик. В дополнение к количествам и процентам, которые могут подсчитываться для категориальных переменных, итожащие статистики для количественных переменных включают в себя:

- Среднее
- Медиана
- Процентили
- Сумма
- Стандартное отклонение
- Диапазон
- Минимальное и максимальное значение

Дополнительную информацию смотрите в разделе “Итожащие статистики для количественных переменных и итогов категориальных переменных” на стр. 9.

Файл данных примера

Для примеров в данной главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Состыковывание количественных переменных

При помощи состыковывания вы можете подытожить несколько количественных переменных в одной таблице.

1. Выберите в меню:

Анализ > Таблицы > Настраиваемые таблицы...

2. В построителе таблиц щелкните в списке переменных *Возраст респондента*, затем, удерживая клавишу **Ctrl**, щелкните *Количество лет, потраченных на образование* и *Число часов в день, проводимых у телевизора*, чтобы выделить эти три переменные.

3. Перенесите три выбранные переменные в поле строки на панели холста.

Теперь три выделенные переменные состыкованы в измерении строк. Поскольку данные переменные являются количественными, категории в таблице не выводятся, и итожащей статистикой, выводимой по умолчанию, является среднее значение.

4. Щелкните по **ОК**, чтобы создать таблицу.

	Mean
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

Рисунок 38. Таблица средних значений по составным количественным переменным

Несколько итожащих статистик

По умолчанию для количественных переменных выводится только среднее значение; однако пользователь может выбрать и другие итожащие статистики для количественных переменных, и одновременно могут выводиться несколько итожащих статистик.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните правой кнопкой мыши по любой из трех количественных переменных на панели холста в мастере таблиц и выберите во всплывающем меню **Итожащие статистики**.
3. В диалоговом окне Итожащие статистики выберите в списке статистик *Медиана* и щелкните по стрелке, чтобы перенести данную статистику в список Показать. (Для перемещения статистик из списка Статистики в список Показать вы можете использовать стрелку или перетаскивать статистики в список Показать.)
4. Щелкните на поле **Формат** для медианы в группе Показать и выберите **nnnn** из раскрывающегося списка форматов.
5. В поле Десятичн наберите 1 .
6. Аналогичные изменения произведите и для среднего значения в списке Показать.
7. Щелкните **Применить ко всем** , чтобы применить сделанные изменения ко всем трем количественным переменным.
8. Щелкните **ОК** в построителе таблиц для создания таблицы.

	Mean	Median
Age of respondent	45.6	42.0
Highest year of school completed	13.3	13.0
Hours per day watching TV	2.9	2.0

Рисунок 39. Среднее и медиана, выведенные в таблице с составными количественными переменными

Частота, количество валидных значений и пропущенные значения

Часто бывает полезным показать количество наблюдений, которые использовались для подсчета итожащих статистик, таких как среднее значение, и вы можете небезосновательно полагать, что эту информацию можно получить при помощи итожащей статистики *Количество*. Однако эта статистика не даст вам правдивую информацию в том случае, если в данных присутствуют пропущенные наблюдения. Для получения реального количества наблюдений, использованных в расчетах, необходимо задать статистику *Валидные* .

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните правой кнопкой мыши по любой из трех количественных переменных на панели холста в мастере таблиц и выберите во всплывающем меню **Итожащие статистики**.
3. В диалоговом окне Итожащие статистики выберите в списке Статистики **Количество** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
4. Выберите в списке Статистики **Валидные** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
5. Щелкните **Применить ко всем** , чтобы применить сделанные изменения ко всем трем количественным переменным.
6. Щелкните **ОК** в построителе таблиц для создания таблицы.

	Mean	Median	Count	Valid N
Age of respondent	45.6	42.0	2832	2828
Highest year of school completed	13.3	13.0	2832	2820
Hours per day watching TV	2.9	2.0	2832	2337

Рисунок 40. Сравнение значения частоты с количеством валидных значений

Для всех трех переменных значение статистики *Частота* одинаково: 2832. И это не является случайностью, поскольку таково общее количество наблюдений в файле данных. Поскольку количественные переменные не являются вложенными в категориальные переменные, *Частота* представляет собой общее количество наблюдений в файле данных.

Значения статистики *Валидные*, с другой стороны, отличаются для различных переменных; особенно сильно значение данной статистики отличается от значения статистики *Количество* для переменной *Число часов в день, проводимых у телевизора*. Причина заключается в том, что данная переменная имеет большое количество **пропущенных значений** -- наблюдений, в которых данная переменная не имеет значений или имеет значения, помеченные пользователем как пропущенные (такие как код 99, описывающий вариант *Не применим* для мужчин при вопросе о беременности).

7. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
8. Щелкните правой кнопкой мыши по любой из трех количественных переменных на панели холста в мастере таблиц и выберите во всплывающем меню **Итожащие статистики**.
9. В диалоговом окне Итожащие статистики щелкните по статистике **Валидные** в списке Показать, а затем нажмите кнопку со стрелкой, чтобы удалить данную статистику из списка Показать.
10. В диалоговом окне Итожащие статистики щелкните по статистике **Количество** в списке Показать, а затем нажмите кнопку со стрелкой, чтобы удалить данную статистику из списка Показать.
11. Выберите в списке Статистики **Пропущенные** и нажмите кнопку со стрелкой, чтобы добавить данную статистику в список Показать.
12. Щелкните **Применить ко всем**, чтобы применить сделанные изменения ко всем трем количественным переменным.
13. Щелкните **ОК** в построителе таблиц для создания таблицы.

	Mean	Median	Missing
Age of respondent	45.6	42.0	4
Highest year of school completed	13.3	13.0	12
Hours per day watching TV	2.9	2.0	495

Рисунок 41. Число пропущенных значений, выведенное в таблице вместе с итожащими статистиками для количественных переменных

Теперь в таблице показано количество пропущенных значений для каждой количественной переменной. Теперь мы видим, что переменная *Число часов в день, проводимых у телевизора* имеет большое количество пропущенных значений, в то время как две другие переменные имеют мало пропущенных значений. Данный фактор нужно учитывать до того, как принимать какие-либо решения на основе полученных итожащих статистик.

Использование различных итожащих статистик для различных переменных

В дополнение к выводу нескольких итожащих статистик, вы можете также задавать вывод различных итожащих статистик для различных количественных переменных, составных в таблице. К примеру, предыдущая таблица позволила сделать вывод, что одна из трех переменных имеет большое количество пропущенных значений; и вы можете задать вывод количества пропущенных значений только для одной переменной.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).

- Щелкните по переменной *Возраст респондента* на панели холста в мастере таблиц, а затем, удерживая клавишу Ctrl, щелкните по переменной *Количество лет, потраченных на образование*, чтобы выделить эти две переменные.
- Щелкните правой кнопкой мыши на любой из двух выделенных переменных и выберите во всплывающем меню **Итожащие статистики**.
- В диалоговом окне **Итожащие статистики** щелкните по статистике **Пропущенные** в списке Показать, а затем нажмите кнопку со стрелкой, чтобы удалить данную статистику из списка Показать.
- Щелкните **Применить к выделенному**, чтобы применить изменения только к двум выделенным переменным.
Пустые ячейки в таблице показывают, что количество пропущенных значений будет показано только для переменной *Число часов в день, проводимых у телевизора*.
- Щелкните по **ОК**, чтобы создать таблицу.

	Mean	Median	Missing
Age of respondent	45.6	42.0	
Highest year of school completed	13.3	13.0	
Hours per day watching TV	2.9	2.0	495

Рисунок 42. Таблица с разными итожащими статистиками для разных переменных

Хотя полученная таблица содержит всю необходимую информацию, формат этой таблицы затрудняет ее интерпретацию. Ваш коллега, работающий с данной таблицей, может подумать, что пустые ячейки в столбце *Пропущенные* говорят о том, что переменные не имеют пропущенных значений.

- Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
- В построителе таблиц в группе **Итожащие статистики** выберите **Строки** в раскрывающемся списке **Расположение**.
- Щелкните по **ОК**, чтобы создать таблицу.

Age of respondent	Mean	45.6
	Median	42.0
Highest year of school completed	Mean	13.3
	Median	13.0
Hours per day watching TV	Mean	2.9
	Median	2.0
	Missing	495

Рисунок 43. Итожащие статистики и переменные, выведенные в измерении строк

Теперь становится понятно, что таблица содержит статистику количества пропущенных наблюдений только для одной переменной.

Итожащие статистики по категориям

Вы можете использовать категориальные переменные в качестве группирующих переменных, с подсчетом итожащих статистик по количественным переменным внутри групп, образуемых вариантами ответов категориальных переменных.

- Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
- Перенесите переменную *Пол респондента* из списка переменных в поле **Столбцы** панели холста.
Если щелкнуть правой кнопкой мыши по переменной *Пол респондента* на панели холста, то можно увидеть, что опция **Итожащие статистики** во всплывающем меню не является активной. Причина заключается в том, что в таблице с количественными переменными статистики могут задаваться только для количественных переменных.
- Щелкните по **ОК**, чтобы создать таблицу.

		Gender	
		Male	Female
Age of respondent	Mean	44.6	46.3
	Median	42.0	43.0
Highest year of school completed	Mean	13.4	13.2
	Median	13.0	13.0
Hours per day watching TV	Mean	2.8	2.9
	Median	2.0	2.0
	Missing	213	282

Рисунок 44. Итоговые статистики для количественных переменных по категориям переменной столбцов

Таблица позволяет сравнивать средние значения (среднее и медиану) для мужчин и женщин, и это сравнение показывает, что особых различий в данном случае нет, что не является предметом особого интереса, но может оказаться полезной информацией.

Несколько группирующих переменных

Вы можете разделить полученные группы на подгруппы при помощи использования процедуры вложения и/или одновременного использования категориальных переменных в строках и столбцах.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Перенесите переменную *Получают новости из Интернета* из поля строк в левую часть поля Строки на панели холста. Убедитесь, что в новой таблице все три количественные переменные (а не одна) вложены в переменную Получение новостей из Интернета.

Хотя возможны ситуации, когда необходимо получить таблицу, сходную с таблицей из второго примера, в данном случае мы остановимся на другом примере.

3. Щелкните по **ОК** , чтобы создать таблицу.

				Gender	
				Male	Female
Get news from internet	No	Age of respondent	Mean	47.0	48.8
			Median	45.0	46.0
		Highest year of school completed	Mean	13.4	13.1
			Median	13.0	12.0
		Hours per day watching TV	Mean	3.2	3.4
			Median	2.0	3.0
	Yes	Age of respondent	Mean	38.7	41.1
			Median	35.0	38.0
		Highest year of school completed	Mean	13.2	13.3
			Median	13.0	13.0
		Hours per day watching TV	Mean	2.1	2.1
			Median	2.0	2.0
		Missing	0	0	

Рисунок 45. Итоговые статистики для количественных переменных по группам, образованным категориальными переменными строк и столбцов

Вложение категориальной переменной в количественные переменные

Хотя изображенная выше таблица содержит всю необходимую информацию, это не лучший формат для интерпретации результатов. К примеру, вы можете сравнить средний возраст мужчин, получающих новости из Интернета, и мужчин, не получающих информацию из Интернета. Это было бы гораздо проще сделать в том случае, если бы значения располагались рядом, а не в разных частях таблицы. Изменение положения переменных строк с вложением категориальной группирующей переменной в три количественные переменные, могло бы улучшить бы таблицу. В случае с количественными переменными уровень вложения не влияет на то, какие переменные будут определять статистики в таблице. Вне зависимости от уровня вложения, количественные переменные всегда будут определять статистики в таблице.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .

2. В мастере таблиц щелкните на панели холста по переменной *Возраст респондента*, затем, удерживая клавишу Ctrl, щелкните *Количество лет, потраченных на образование* и *Число часов в день, проводимых у телевизора*, чтобы выделить эти три переменные.
3. Перенесите все три количественные переменные на левый край области Строки таким образом, чтобы переменная *Получают новости из Интернета* была вложена в каждую из трех количественных переменных.
4. Щелкните по **ОК**, чтобы создать таблицу.

				Gender	
				Male	Female
Age of respondent	Get news from internet	No	Mean	47.0	48.8
			Median	45.0	46.0
	Yes	Mean	38.7	41.1	
		Median	35.0	38.0	
Highest year of school completed	Get news from internet	No	Mean	13.4	13.1
			Median	13.0	12.0
	Yes	Mean	13.2	13.3	
		Median	13.0	13.0	
Hours per day watching TV	Get news from internet	No	Mean	3.2	3.4
			Median	2.0	3.0
			Missing	213	282
	Yes	Mean	2.1	2.1	
		Median	2.0	2.0	
		Missing	0	0	

Рисунок 46. Категориальная переменная строк, вложенная в составные количественные переменные

Выбор порядка вложения зависит от того, какие связи или сравнения вы хотите подчеркнуть в таблице. Изменение порядка вложения количественных переменных не влияет на значения итожащих статистик, а только изменяет их относительное расположение в таблице.

Доверительные интервалы

Доверительные интервалы и среднеквадратичные ошибки доступны для многих табличных статистик.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите *Высшая степень* в область строк на панели холста.
3. Щелкните по **Итожащие статистики...**
4. В списке **Статистики** диалогового окна Итожащие статистики щелкните по значку рядом со **Счетчиком**, чтобы раскрыть список связанных со счетчиком статистических показателей.
5. Перенесите следующие статистические показатели в область **Вывод**: Процент по столбцу, Нижняя граница ДИ для процента по столбцу, верхняя граница ДИ для процента по столбцу и Среднеквадратичная ошибка для процента по столбцу.
6. В группе **Доверительные интервалы** в поле **Уровень (%)** введите 99.
7. Нажмите кнопку **Применить к выбору**, а затем **Заккрыть**.
8. Щелкните по **ОК**, чтобы создать таблицу.

		Count	Column N %	99.0% Lower CL for Column N %	99.0% Upper CL for Column N %	Standard Error of Column N %
Highest degree	LT High school	430	15.2%	13.6%	17.0%	0.7%
	High school	1500	53.2%	50.7%	55.6%	0.9%
	Junior college	209	7.4%	6.2%	8.7%	0.5%
	Bachelor	478	16.9%	15.2%	18.8%	0.7%
	Graduate	205	7.3%	6.1%	8.6%	0.5%

Рисунок 47. Таблица количеств, процентов по столбцам и доверительных интервалов

9. Откройте диалоговое окно Пользовательские таблицы и щелкните по **Итожащие статистики...**
10. Для нижнего доверительного предела измените метку на "Нижний доверительный предел (&[Confidence Level])". Строка "&[Confidence Level]" вставляет значение заданного доверительного уровня в положение метки.
11. Нажмите кнопку **Применить к выбору**, а затем **Заккрыть**.
12. Щелкните по **ОК**, чтобы создать таблицу.

		Count	Column N %	Lower Confidence Limit (99.0%)	99.0% Upper CL for Column N %	Standard Error of Column N %
Highest degree	LT High school	430	15.2%	13.6%	17.0%	0.7%
	High school	1500	53.2%	50.7%	55.6%	0.9%
	Junior college	209	7.4%	6.2%	8.7%	0.5%
	Bachelor	478	16.9%	15.2%	18.8%	0.7%
	Graduate	205	7.3%	6.1%	8.6%	0.5%

Рисунок 48. Таблица с измененной меткой доверительного интервала

Статистические критерии

Статистические критерии

Для оценки степени связи между переменными строк и столбцов в процедуре построения таблиц имеются три различных статистических критерия. В этом разделе рассматривается вывод результатов, получаемых при вычислении каждого критерия, причем особое внимание уделяется вопросам влияния вложения и состыковывания. Дополнительную информацию смотрите в разделе "Таблицы с категориальными переменными: состыковывание, вложение и слов" на стр. 26.

Файл данных примера

Для примеров в этой главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Критерий независимости (Хи-квадрат)

Критерий независимости хи-квадрат используется для выяснения того, есть ли связь между двумя категориальными переменными. К примеру, вы хотите выяснить, связана ли переменная *Занятость респондента* с переменной *Семейное положение*.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите переменную *Занятость респондента* из списка переменных в поле Строки панели холста.

3. Перенесите переменную *Семейное положение* из списка переменных в поле Столбцы.
4. Укажите, что итожащие статистики будут задаваться в измерении **Строки**.
5. Выделите переменную *Занятость респондента* и щелкните **Итожащие статистики** в группе Определить.
6. Выделите **% по столбцу** в списке Статистики и добавьте данную статистику в список Показать
7. Нажмите кнопку **Применить к выделенному**.
8. В диалоговом окне Настраиваемые таблицы щелкните по вкладке **Статистические критерии**.
9. Выберите **Критерий независимости (Хи-квадрат)**.
10. Щелкните **ОК** для создания таблицы и вычисления критерия хи-квадрат.

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
Labor force status	Working full time	Count	778	44	295	58	392
		Column %	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	Count	138	20	35	9	102
		Column %	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	Count	23	2	9	1	11
		Column %	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	Count	13	3	10	0	32
		Column %	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	Count	168	150	53	6	17
		Column %	12.5%	53.0%	11.9%	6.5%	2.6%
	School	Count	9	1	7	2	60
		Column %	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	Count	200	55	25	13	35
		Column %	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	Count	16	8	12	4	14
		Column %	1.2%	2.8%	2.7%	4.3%	2.1%

Рисунок 49. Зависимость занятости от семейного положения

Эта таблица представляет собой таблицу сопряженности переменных *Занятость респондента* и *Семейное положение*, содержащую в себе в качестве итожащих статистик значения количеств и процентов по столбцу. Проценты по столбцу подсчитываются таким образом, что сумма процентов по столбцу составляет 100%. Если связь между двумя переменными отсутствует, то значения процентов по различным столбцам в каждой строке не должны сильно различаться. В нашем случае определенные различия в процентах присутствуют, но для уверенности в том, что эти различия являются значимым, вы должны проверить результаты вычисления критерия хи-квадрат.

		Marital status
Labor force status	Chi-square	729.242
	df	28
	Sig.	.000*

*. The Chi-square statistic is significant at the 0.05 level.

Рисунок 50. Критерий хи-квадрат Пирсона

При вычислении критерия независимости проверяется гипотеза о том, что между переменными *Занятость респондента* и *Семейное положение* связи нет. Это означает, что процентные распределения по различным столбцам не отличаются друг от друга, а наблюдаемые отличия являются чисто случайными. Статистика хи-квадрат измеряет общее различие между наблюдаемыми значениями в ячейках таблицы и теми значениями, которые были бы в этих ячейках, если бы процентные распределения в столбцах были бы одинаковыми. Большее значение статистики хи-квадрат указывает на большее различие между наблюдаемыми и ожидаемыми количествами в ячейках и в большей степени свидетельствует в пользу того, что процентные распределения по столбцам не являются равными, гипотеза о независимости переменных неверна, и, следовательно, переменные *Занятость респондента* и *Семейное положение* связаны между собой.

Полученное значение статистики хи-квадрат равняется 729,242. Для определения того, является ли это значение достаточным для отклонения гипотезы о независимости, рассчитывается уровень значимости данной статистики. Показатель значимости представляет собой вероятность случайного извлечения из совокупности, распределенной по закону хи-квадрат с 28 степенями свободы, значения, большего, чем 729,242. Поскольку в нашем случае это значение меньше, чем значение альфа, указанное на вкладке Статистические критерии, вы можете отвергнуть гипотезу о независимости на уровне 0,05. Таким образом, переменные *Занятость респондента* и *Семейное положение* связаны между собой.

Влияние вложения и состыковывания на вычисление критерия независимости

Правила для критериев независимости следующие: вычисление отдельного критерия проводится для каждой внутренней подтаблицы. Для понимания того, как вложение влияет на вычисление критерия, повторите предыдущий пример с вложением переменной *Семейное положение* в переменную *Пол респондента*.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Перенесите переменную *Пол респондента* из списка переменных в поле Столбцы, поместив ее над переменной *Семейное положение*.
3. Щелкните по **ОК**, чтобы создать таблицу.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{*,1,2}	.000 ^{*,1,2}

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Рисунок 51. Критерий хи-квадрат Пирсона

С вложением переменной *Семейное положение* в переменную *Пол респондента* критерий вычисляется два раза -- по одному разу для каждого уровня переменной *Пол респондента*. Показатель значимости для каждого вычисления критерия позволяет сделать вывод, что можно отвергнуть гипотезу о независимости переменных *Семейное положение* и *Занятость респондента* как для мужчин, так и для женщин. Однако примечания в таблице заставляют обратить внимание на тот факт, что более 20% ячеек в каждой таблице имеют ожидаемое количество менее 5 и минимальное ожидаемое значение количества в ячейке (или ячейках) меньше единицы. Эти примечания говорят о том, что для данных таблиц не соблюдаются основные предположения критерия хи-квадрат; таким образом, результаты вычисления критерия могут быть поставлены под сомнение.

Примечание: Примечания могут обрезаться границами ячеек. Чтобы сделать их видимыми, измените параметр выравнивания в ячейках в диалоговом окне Свойства ячеек.

Теперь посмотрим на то, как состыковывание влияет на вычисление критерия:

4. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
5. Перенесите переменную *Уровень образования* из списка переменных в поле Строки под переменную *Занятость респондента*.
6. Щелкните по **ОК**, чтобы создать таблицу.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^a .1,2	.000 ^a .1,2
Highest degree	Chi-square	43.844	105.506
	df	16	16
	Sig.	.000*	.000*

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Рисунок 52. Критерий хи-квадрат Пирсона

При состыковывании переменных *Уровень образования* и *Занятость респондента* выполняются четыре теста: внутри каждой категории переменной *Пол респондента* проверяется независимость переменных *Семейное положение* и *Занятость респондента*, а также переменных *Семейное положение* и *Уровень образования*. Результаты вычисления критерия для переменных *Семейное положение* и *Занятость респондента* аналогичны предыдущим результатам. Результаты вычисления критерия для переменных *Семейное положение* и *Уровень образования* показывают, что эти переменные не являются независимыми.

Сравнение средних значений по столбцам

Критерий сравнения средних значений по столбцам проводится с целью определения того, есть ли связь между категориальной переменной в поле *Столбцы* и количественной переменной в поле *Строки*. Более того, вы можете использовать результаты вычисления критерия для того, чтобы определить относительный порядок категорий категориальной переменной в терминах средних значений количественной переменной. К примеру, можно определить, связана ли переменная *Число часов в день, проводимых у телевизора* с переменной *Получают новости из газет*.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. Щелкните **Сброс** для восстановления первоначальных установок во всех вкладках.
3. В построителе таблиц перенесите переменную *Число часов в день, проводимых у телевизора* из списка переменных в поле *Строки* панели холста.
4. Перенесите переменную *Получают новости из газет* из списка переменных в поле *Столбцы*.
5. Выделите переменную *Число часов в день, проводимых у телевизора* и щелкните **Итожащие статистики** в группе *Определить*.
6. Укажите формат **nnnn**.
7. Укажите количество показанных десятичных знаков - **2**. Обратите внимание, что представление формата изменилось на **nnnn.nn**.
8. Нажмите кнопку **Применить к выделенному**.
9. В диалоговом окне *Настраиваемые таблицы* щелкните по вкладке **Статистические критерии**.
10. Выберите **Сравнение средних по столбцам (t-критерий)**.
11. Щелкните **ОК** для создания таблицы и вычисления критерия сравнения средних значений.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92	2.74

Рисунок 53. Средние значения переменной Число часов в день, проводимых у телевизора в зависимости от переменной Получают новости из газет

В таблице отображаются средние значения переменной *Число часов в день, проводимых у телевизора* для респондентов, которые получают и не получают новости из газет. Средние значения позволяют сделать вывод о том, что респонденты, не получающие новостей из газет, в среднем в день тратят на просмотр телевизора на 0,18 часа больше, чем респонденты, получающие новости из газет. Для того, чтобы понять, может ли такое различие считаться случайным, обратимся к вычисленному критерию сравнения средних по столбцам.

	Get news from newspapers	
	No	Yes
	(A)	(B)
Hours per day watching TV		

Рисунок 54. Сравнение средних по столбцам

Критерий сравнения средних по столбцам присваивает каждой категории переменной столбцов некоторую букву. Для переменной *Получают новости из газет* категории *Нет* присваивается буква А, а категории *Да* - буква В. Каждая пара средних по столбцам сравнивается при помощи *t*-критерия. Поскольку в нашем случае имеется только одна пара, критерий вычисляется только один раз. Если в паре имеются значимые различия, то буква для категории с меньшим средним значением располагается в столбце, соответствующем категории с большим средним значением. Поскольку в ячейках таблицы буквы отсутствуют, мы можем сделать вывод, что средние по столбцам значимо не различаются.

Результаты значимости в записи в стиле АРА

если не нужно, чтобы результаты значимости были в отдельной таблице, можно выбрать их вывод на экран в основной таблице. Результаты значимости определяются с помощью записи в стиле АРА с использованием нижнего индекса. Завершите действия предыдущих шагов для сравнения средних значений столбцов, однако внесите следующее изменение на вкладку Статистические критерии.

1. В области Обнаружить значимые различия выберите пункт **В основной таблице, используя подстрочный индекс АРА**.
2. Нажмите кнопку **ОК** для создания таблицы и вычисления критериев средних значений по столбцам с помощью записи в стиле АРА.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92 _a	2.74 _a

Рисунок 55. Сравнение средних значений по столбцам с помощью записи в стиле АРА

Критерий сравнения средних по столбцам присваивает категориям переменной столбцов нижний индекс. Для каждой пары столбцов, средние значения столбца сравниваются с помощью *t*-критерия. Если значения в паре значимо различаются, то эти значения имеют *различные* приписанные им нижние индексы. Поскольку в

нашем случае имеется только одна пара, критерий вычисляется только один раз. Поскольку средние значения столбцов в данном примере имеют общие нижние индексы, то средние значения для данных столбцов статистически не различаются.

Влияние эффектов вложения и состыковывания на вычисление критериев сравнения средних по столбцам

Правило для вычисления критериев сравнения средних по столбцам следующее: отдельный набор попарных вычислений критерия выполняется для каждой внутренней подтаблицы. Для понимания того, как вложение влияет на вычисление критериев, повторите предыдущий пример с вложением переменной *Число часов в день, проводимых у телевизора* в переменную *Занятость респондента*.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Перенесите переменную *Занятость респондента* из списка переменных в поле Строки панели холста.
3. Щелкните по **ОК**, чтобы создать таблицу.

			Get news from newspapers	
			No	Yes
			(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B	
	Working part-time	Hours per day watching TV		
	Temporarily not working	Hours per day watching TV		
	Unemployed, laid off	Hours per day watching TV		
	Retired	Hours per day watching TV		
	School	Hours per day watching TV		
	Keeping house	Hours per day watching TV		
	Other	Hours per day watching TV		

Рисунок 56. Сравнение средних по столбцам

Если переменная *Число часов в день, проводимых у телевизора* вложена в уровни переменной *Занятость респондента*, вычисляется семь наборов средних по столбцам: по одному для каждого уровня переменной *Занятость респондента*. Категориям переменной *Получают новости из газет* были присвоены те же самые буквы. Для группы респондентов с *полной занятостью* в столбце А появилась буква В. Это означает, что в группе респондентов, работающих полный рабочий день, среднее значение переменной *Число часов в день, проводимых у телевизора* меньше для тех, кто получает новости из газет. Поскольку других букв в таблице нет, можно сделать вывод, что других значимых различий между средними по столбцам нет.

Поправки Бонферрони. При неоднократном вычислении критерия к результатам применяется поправка Бонферрони, обеспечивающая постоянство значения уровня альфа, указанного на вкладке Статистические критерии, в применении к каждому *вычислению* критерия. В данной таблице поправка Бонферрони не применяется по причине того, что (хотя и было вычислено семь наборов критериев) в каждом наборе критерий вычислялся только по одной паре значений.

Теперь посмотрим на то, как состыковывание влияет на вычисление критерия:

4. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
5. Перенесите переменную *Получают новости из Интернета* из списка переменных в поле Столбцы слева от переменной *Получают новости из газет*.
6. Щелкните по **ОК**, чтобы создать таблицу.

			Get news from internet		Get news from newspapers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B		B	
	Working part-time	Hours per day watching TV	B			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	B			
	Retired	Hours per day watching TV	B			
	School	Hours per day watching TV	B			
	Keeping house	Hours per day watching TV	B			
	Other	Hours per day watching TV	B			

Рисунок 57. Сравнение средних по столбцам

В результате состыковывания переменной *Получают новости из Интернета* с переменной *Получают новости из газет* было вычислено 14 наборов критериев сравнения средних - по одному набору для каждой категории переменной *Занятость респондента* для переменных *Получают новости из Интернета* и *Получают новости из газет*. Поправка Бонферрони опять-таки не применялась, поскольку в каждом наборе критерий вычислялся только один раз. Результаты вычисления критерия для переменной *Получают новости из газет* выглядят так же, как и ранее. Для переменной *Получают новости из Интернета* категории *Нет* присваивается буква А, а категории *Да* присваивается буква В. Буква В показана в столбце А для каждого критерия на сравнение средних, исключая подгруппу *временно не работающих* респондентов. Это означает, что среднее значение переменной *Число часов в день, проводимых у телевизора* в целом ниже для тех респондентов, которые получают новости из Интернета, чем для тех, которые не получают новости из Интернета. Для категории *временно не работающих* не выводятся буквы в ячейках таблицы; это означает, что средние по столбцам для данных респондентов статистически не различаются.

Сравнение пропорций по столбцам

Критерий сравнения пропорций по столбцам используется для сравнения категорий, задаваемых переменной столбцов, посредством изучения распределений переменной строк. К примеру, если в результате использования критерия хи-квадрат вы пришли к выводу, что переменные *Занятость респондента* и *Семейное положение* не являются независимыми, то вы можете выяснить, какие строки и столбцы обуславливают наличие связи.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. Щелкните **Сброс** для восстановления первоначальных установок во всех вкладках.
3. В построителе таблиц перенесите переменную *Занятость респондента* из списка переменных в поле Строки панели холста.
4. Перенесите переменную *Семейное положение* из списка переменных в поле Столбцы.
5. Выделите переменную *Занятость респондента* и щелкните **Итожащие статистики** в группе Определить.
6. Выделите **% по столбцу** в списке Статистики и добавьте данную статистику в список Показать.
7. Удалите статистику **Встречаемость** из списка Показать.
8. Нажмите кнопку **Применить к выделенному**.
9. В диалоговом окне **Настраиваемые таблицы** щелкните по вкладке **Статистические критерии**.
10. Выберите **Сравнение пропорций по столбцам (z-критерий)**.

11. Щелкните **ОК** для создания таблицы и вычисления критерия сравнения пропорций по столбцам.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column %	Column %	Column %	Column %	Column %
Labor force status	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	12.5%	53.0%	11.9%	6.5%	2.6%
	School	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Рисунок 58. Зависимость занятости от семейного положения

Эта таблица представляет собой таблицу сопряженности переменных *Занятость респондента* и *Семейное положение*, содержащую в себе в качестве итожащей статистики проценты по столбцам.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Labor force status	Working full time	B		A B	B	B
	Working part-time					A B C
	Temporarily not working					A B
	Unemployed, laid off					A B
	Retired	E	A C D E	E		
	School					A B C
	Keeping house	C E	C E		C E	
	Other					

Рисунок 59. Сравнение пропорций по столбцам

Критерий сравнения пропорций по столбцам присваивает каждой категории переменной столбцов некоторую букву. Для переменной *Семейное положение* категории *Состоит в браке* присваивается букв А, *Вдовец/вдова* - В, и т.д., вплоть до категории *Никогда не состоял в браке*, которой присваивается буква Е. Для каждой пары столбцов производится сравнение пропорций с вычислением z-критерия. Таким образом, вычисляется семь наборов критериев сравнения пропорций по столбцам, по одному набору для каждой категории переменной *Занятость респондента*. Поскольку переменная *Семейное положение* имеет 5 вариантов ответа, для каждого варианта переменной *Занятость* вычисляется $5 \times 4 / 2 = 10$ критериев, и для корректировки значимости в данном случае применяется поправка Бонферрони. Если в паре обнаруживаются значимые различия, то буква для категории с меньшей долей располагается в столбце, ответственном за категорию с большей долей.

В наборе тестов для категории *Полная занятость* буква В появляется во всех других столбцах. Помимо этого, буква А появляется в столбце С. Другие буквы для данной категории в столбцах не появляются. Вы можете сделать вывод, что доля разведенных респондентов, работающих полный рабочий день, больше доли состоящих в браке респондентов, работающих полный рабочий день, которая, в свою очередь, больше доли овдовевших респондентов, работающих полный рабочий день. Доли тех респондентов, которые проживают раздельно или никогда не состояли в браке и работают полный рабочий день, не отличаются от долей разведенных или состоящих в браке респондентов, работающих полный день; но эти доли значимо больше долей вдовых респондентов, работающих полный рабочий день.

В тестах для категорий *Частичная занятость* или *Учащийся*, в столбце Е появляются буквы А, В и С. Другие буквы для данной категории в столбцах не появляются. Таким образом, доли респондентов, которые никогда не состояли в браке и являются учащимися или работающими неполный рабочий день, значимо больше долей состоящих в браке, вдовых или разведенных респондентов, которые являются учащимися или работающими неполный рабочий день.

В тестах для категорий трудового статуса *Временно не работает* или *Другое*, в столбцах нет никаких букв. Поэтому нет значимых различий между долями, относящимися к разным категориям семейного положения, для респондентов, которые относятся к данным категориям трудового статуса.

Тесты для категории *Пенсионер* показывают, что доля овдовевших респондентов-пенсионеров больше, чем доли пенсионеров во всех других категориях респондентов по семейному положению.. Более того, доли состоящих в браке или разведенных респондентов-пенсионеров также больше доли никогда не состоявших в браке респондентов-пенсионеров.

Доля респондентов, отметивших вариант *Домохозяйка*, значимо выше среди состоящих в браке, овдовевших или проживающих отдельно респондентов, чем среди разведенных респондентов или респондентов, которые никогда не состояли в браке.

Доля *безработных* среди респондентов, которые никогда не состояли в браке, значимо выше, чем аналогичная доля среди состоящих в браке или овдовевших респондентов. Также необходимо обратить внимание на то, что в столбце *Проживают отдельно* для категории *Безработный* стоит символ ".", обозначающий, что наблюдаемая доля безработных среди проживающих отдельно равна 0 или 1, и данная доля не может сравниваться с остальными долями для рассматриваемой категории.

Объединение результатов значимости в основную таблицу

если не нужно, чтобы результаты значимости были в отдельной таблице, можно выбрать их вывод на экран в основной таблице. Завершите действия предыдущих шагов для сравнения пропорций по столбцам, однако внесите следующее изменение на вкладку *Статистические критерии*.

1. В области *Обнаружить значимые различия* выберите пункт **В основной таблице**.
2. Щелкните по **ОК**, чтобы создать таблицу.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Рисунок 60. Результаты значимости, объединенные в основную таблицу

3. Дважды щелкните по таблице в окне Средство просмотра, чтобы активировать ее.
4. Поставьте указатель мыши на одну из ячеек меток столбцов, содержащую буквенные ключи. Например, поставьте указатель на ячейку с меткой "(a)".

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Рисунок 61. Выделение всех ячеек, содержащих выбранный буквенный ключ

Все ячейки, содержащие данный ключ, выделяются в таблице.

5. Щелкните правой кнопкой мыши по ячейке с меткой столбца, содержащей буквенный ключ. Выберите из контекстного меню **Выбрать > Выбрать все ячейки с этим ключом значимости**.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Рисунок 62. Выбор всех ячеек, содержащих выбранный буквенный ключ

Все ячейки, содержащие данный ключ, выбираются в таблице.

Влияние эффектов вложения и состыковывания на вычисление критерия сравнения пропорций по столбцам

Правило для вычисления критериев пропорций по столбцам следующее: отдельный набор попарных вычислений критерия выполняется для каждой внутренней подтаблицы. Для понимания того, как вложение влияет на вычисление критерия, повторите предыдущий пример с вложением переменной *Занятость респондента* в переменную *Пол респондента*.

1. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Перенесите переменную *Пол респондента* из списка переменных в поле Строки панели холста.
3. Щелкните по **ОК**, чтобы создать таблицу.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Other				A			
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
Keeping house			C E	C E		C		
Other								

Рисунок 63. Сравнение пропорций по столбцам

В результате вложения переменной *Занятость респондента* в переменную *Пол респондента* вычисляется 14 наборов критериев на сравнение пропорций по столбцам - по одному набору для каждой категории переменной *Занятость респондента* внутри категорий переменной *Пол респондента*. Категориям переменной *Семейное положение* были присвоены те же самые буквы.

По поводу результатов в таблице необходимо сделать несколько замечаний:

- С увеличением общего количества вычисляемых критериев в таблице появляется больше ячеек с нулевыми долями. Эти ячейки встречаются в группах овдовевших и проживающих отдельно респондентов.
- Различия между столбцами, обнаруженные в предыдущем примере для респондентов, отметивших категорию *Домохозяйка*, вызваны различиями в группе женщин.

Теперь посмотрим на то, как состыковывание влияет на вычисление критерия:

4. Снова откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
5. Перенесите переменную *Уровень образования* из списка переменных в поле Строки под переменную *Пол респондента*.
6. Щелкните по **ОК**, чтобы создать таблицу.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
			Keeping house	C E	C E		C	
Highest degree		Other						
			LT High school		A C E			
			High school					
			Junior college	B		B		B
			Bachelor	B				B
Graduate	B							

Рисунок 64. Сравнение пропорций по столбцам

В результате состыковывания переменной *Уровень образования* с переменной *Пол респондента* вычисляется 19 наборов критериев сравнения пропорций по столбцам - 14 наборов, которые уже обсуждались выше, и по одному набору для каждой категории переменной *Уровень образования*. Категориям переменной *Семейное положение* были присвоены те же самые буквы.

По поводу результатов в таблице необходимо сделать несколько замечаний:

- Результаты 14 вычисленных ранее наборов критериев остались теми же самыми.
- Доля респондентов с образованием на уровне неоконченной средней школы среди овдовевших больше, чем среди состоящих в браке, разведенных или никогда не состоявших в браке респондентов.
- Доля респондентов с образованием на уровне колледжа больше среди состоящих в браке, разведенных или никогда не состоявших в браке респондентов, чем среди овдовевших респондентов.

Замечания по использованию весов и множественных наборов данных

Весы наблюдений всегда основываются на количествах, а не на ответах, даже если одна из переменных является переменной с множественными ответами.

Несколько наборов ответов

Настраиваемые таблицы и Построитель диаграмм позволяют работать со специальными "переменными", которые называются **наборами множественных ответов**. Наборы множественных ответов не являются "переменными" в обычном смысле этого слова. Их нельзя увидеть в Редакторе данных, и они не распознаются другими процедурами. В наборах множественных ответов для ввода ответов на вопросы, на

которые можно дать больше одного ответа, используются несколько переменных. Наборы множественных ответов обрабатываются как категориальные переменные, и большинство операций, которые можно выполнять с категориальными переменными, можно делать также и с наборами множественных ответов.

Наборы множественных ответов создаются из нескольких переменных в файле данных. Набор множественных ответов является специальной конструкцией в файле данных. Вы можете задавать и сохранять наборы множественных ответов в IBM® SPSS Statistics файле данных, но вы не можете импортировать или экспортировать наборы множественных ответов из файлов или в файлы других форматов. Наборы множественных ответов из другого файла данных IBM SPSS Statistics могут быть скопированы при помощи процедуры Копировать свойства данных в меню Данные в окне Редактора данных.

Файл данных примера

Для примеров в этой главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Частоты, ответы, проценты и итоги

Все итожащие статистики, доступные для категориальных переменных, могут быть использованы и для наборов множественных ответов. Для наборов множественных ответов доступными являются и дополнительные статистики.

1. Выберите в меню:

Анализ > Таблицы > Пользовательские таблицы...

2. Переместите набор *Источники новостей* (это метка набора множественных ответов *\$m1tnews*) из списка переменных в поле строк панели холста.

{\f3 Значок рядом с данной } {\f4 " } {\f3 переменной} {\f4 " } {\f3 в списке переменных указывает на то, что данный набор является дихотомическим набором множественных ответов.}



Рисунок 65. Значок дихотомического множественного набора

В случае с дихотомическим набором множественных ответов каждая "категория" является в реальности отдельной переменной, а метки категорий являются метками переменных (или именами переменных, если переменные не имеют меток переменных). В данном примере значения частот показывают количество наблюдений, имеющих вариант ответа *Да* по каждой переменной в наборе.

- Щелкните правой кнопкой по набору *Источники новостей* на панели холста и выберите во всплывающем меню **Категории и итоги**.
- Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить**.
- Теперь щелкните правой кнопкой мыши по переменной *Получают новости из Интернета* и выберите во всплывающем меню **Итожащие статистики**.
- В диалоговом окне Итожащие статистики выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
- Нажмите кнопку **Применить к выделенному**, а затем нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Column N %
News sources	Get news from internet	867	41.7%
	Get news from radio	551	26.5%
	Get news from television	1077	51.8%
	Get news from news magazines	294	14.1%
	Get news from newspapers	805	38.7%
Total		2081	100.0%

Рисунок 66. Частоты и проценты по столбцу для дихотомического набора множественных ответов

Итоги, которые не подытоживают

Если посмотреть на значения в таблице, то можно увидеть, что имеет место несоответствие между "итогами" и значениями, которые должны были подытоживаться, т.е. итоги значительно меньше, чем они должны были бы быть. Причина такого несоответствия заключается в том, что встречаемость для каждой "категории" в таблице показывает количество наблюдений со значением 1 (ответ *Да*) для каждой переменной, и общее количество ответов *Да* для всех пяти переменных в дихотомическом наборе может превышать общее количество наблюдений в файле данных.

Итоговая "частота", однако, показывает общее количество наблюдений, для которых ответ *Да* встретился по крайней мере в одной переменной в наборе; и она никогда не будет превышать общего количества наблюдений в файле данных. В данном примере итоговая частота составляет 2081 наблюдение, что почти на 800 наблюдений меньше, чем общее количество наблюдений в файле данных. Если ни одна из этих переменных не имеет пропущенных значений, то это означает, что почти 800 опрошенных не получают новости ни из каких источников. Общая встречаемость представляет собой некоторую базу для подсчета процентов по столбцам; поэтому сумма процентов по столбцу в данном примере будет принимать значение большее, чем значение 100%, показанное в качестве итогового значения для процентов по столбцу.

Итоги, которые подытоживают

Несмотря на то, что "частота" является достаточно ясным термином, приведенный выше пример позволяет понять, насколько легко можно допустить ошибку при анализе итогов для наборов множественных ответов. Поэтому достаточно часто в качестве итоговой статистики для множественных наборов используется количество ответов (ответы).

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
2. Щелкните правой кнопкой по набору *Источники новостей* на панели холста и выберите во всплывающем меню **Категории и итоги**.
3. В диалоговом окне Итожащие статистики выберите в списке Статистики **Ответы** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
4. Выберите в списке Статистики **% ответов по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
5. Нажмите кнопку **Применить к выделенному**, а затем нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Column N %	Responses	Column Responses %
News sources	Get news from internet	867	41.7%	867	24.1%
	Get news from radio	551	26.5%	551	15.3%
	Get news from television	1077	51.8%	1077	30.0%
	Get news from news magazines	294	14.1%	294	8.2%
	Get news from newspapers	805	38.7%	805	22.4%
Total		2081	100.0%	3594	100.0%

Рисунок 67. Ответы и проценты по столбцу для дихотомического набора множественных ответов

Для каждой "категории" в дихотомическом наборе множественных ответов *Ответы* и *Частоты* идентичны, и такая ситуация всегда будет иметь место в случае с дихотомическими наборами множественных ответов.

Однако итоги выглядят по-разному. Общее количество ответов составляет 3594, что более чем на 1500 превышает общую встречаемость и более чем на 700 превышает общее количество наблюдений в файле данных.

Для процентных значений итоги в столбцах *% по столбцу* и *% ответов по столбцу* равны 100%, но проценты для каждой категории в дихотомических наборах множественных ответов ниже для откликов по столбцу. Причина заключается в том, что в качестве базы для расчета процента ответов берется общее количество ответов, в нашем случае равное 3594, что и приводит к меньшему значению процентов, по сравнению с обычными процентами, база для которых равняется 2081.

Итоговые проценты, большие 100%

Хотя и для процентов по столбцу, основанных на частотах, и для процентов ответов в таблице указывается итоговое значение, равное 100%, очевидно, что сумма для столбца *% по столбцу* дает больше 100%. Что же следует делать в том случае, если необходимо отобразить в таблице проценты, основанные на частотах, таким образом, чтобы итоговое процентное значение полностью отражало сумму отдельных процентов?

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
2. Щелкните правой кнопкой по набору *Источники новостей* на панели холста и выберите во всплывающем меню **Категории и итоги**.
3. В диалоговом окне Сводная статистика выберите **% ответов по столбцу (база: наблюдения)** в списке Статистика и щелкните по стрелке, чтобы добавить значение в список Показать.
4. Нажмите кнопку **Применить к выделенному**, а затем нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news from newspapers	805	38.7%	805	22.4%	38.7%
	Total	2081	100.0%	3594	100.0%	172.7%

Рисунок 68. Проценты ответов по столбцу со значением частоты в качестве базы

Использование наборов множественных ответов с другими переменными

Вообще говоря, с наборами множественных ответов можно работать как с категориальными переменными. К примеру, вы можете построить таблицу по набору множественных ответов и категориальной переменной или вложить набор множественных ответов в категориальную переменную.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
2. Перенесите переменную *Пол респондента* из списка переменных на левую сторону области Строки панели предварительного просмотра, вкладывая набор множественных ответов *Источники новостей* в категории переменной пола.
3. Щелкните правой кнопкой мыши по переменной *Пол респондента* в Панели холста таблицы и очистите переключатель **Показать метку переменной** во всплывающем меню.
4. Сделайте то же самое для набора *Источники новостей*.
Эти операции позволят удалить из таблицы столбцы с метками переменных (поскольку в данном случае их присутствие не является необходимым).
5. Нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	375	41.9%	375	24.4%	41.9%
	Total	896	100.0%	1539	100.0%	171.8%
Female	Get news from internet	508	42.9%	508	24.7%	42.9%
	Get news from radio	318	26.8%	318	15.5%	26.8%
	Get news from television	626	52.8%	626	30.5%	52.8%
	Get news from news magazines	173	14.6%	173	8.4%	14.6%
	Get news from newspapers	430	36.3%	430	20.9%	36.3%
	Total	1185	100.0%	2055	100.0%	173.4%

Рисунок 69. Набор множественных ответов, вложенный в категориальную переменную

Переменная источника статистик и доступные итожащие статистики

При отсутствии в таблице переменной масштаба, категориальные переменные и наборы множественных ответов рассматриваются одинаково в отношении переменной источника статистики: переменная с наибольшим уровнем вложенности в измерениях источников статистики - это переменная источника статистики. Поскольку некоторые итожащие статистики могут применяться только к наборам множественных ответов, то в том случае, если вы хотите использовать некоторые специальные статистики для набора множественных ответов, этот набор должен находиться на самом внутреннем уровне вложения в измерении статистик.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
2. Перенесите набор *Источники новостей* налево от переменной *Пол респондента*, изменяя порядок вложения.

Как можно увидеть, все специальные итожащие статистики для наборов множественных ответов - количество ответов, процент количества ответов - были удалены из данной таблицы, поскольку теперь на самом внутреннем уровне вложения находится категориальная переменная *Пол респондента* и именно данная переменная теперь является переменной источника статистик.

Однако Построитель таблиц "запоминает" все предшествующие установки. Если вы опять вложите набор *Источники новостей* в переменную *Пол респондента*, все итожащие статистики для наборов множественных ответов опять появятся в предварительном просмотре таблиц.

Категориальные наборы множественных ответов и повторяющиеся ответы

Категориальные наборы множественных ответов предоставляют одну возможность, недоступную для дихотомических наборов множественных ответов, а именно, учет повторяющихся ответов. В большинстве случаев повторяющиеся варианты ответов в категориальных наборах множественных ответов - это ошибки кодировки. К примеру, если на вопрос "Какие три страны, по вашему мнению, производят лучшие автомобили?"> дается ответ *Швеция, Германия и Швеция*, такой ответ не является правильным.

Однако в других случаях повторяющиеся ответы могут быть абсолютно правильными. К примеру, если вопрос формулируется таким образом: "В какой стране были произведены ваши последние три автомобиля?", ответ *Швеция, Германия и Швеция* в данном случае имеет смысл.

Пользовательские таблицы позволяют работать с повторяющимися ответами в категориальных наборах множественных ответов. По умолчанию повторяющиеся ответы не подсчитываются, но вы можете задать подсчет количества повторяющихся ответов.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).

2. Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
3. Перенесите набор *Производитель автомобилей* из списка переменных в поле Строки панели холста.

{\f3 Значок рядом с данной } {\f4 " } {\f3 переменной} {\f4 " } {\f3 в списке переменных показывает, что данный набор является множественным категориальным набором.}



Рисунок 70. Значок множественного категориального набора

Для множественных категориальных наборов отображаемые категории представляют собой совокупность всех меток значений всех переменных, входящих в набор (тогда как для множественных дихотомических наборов - это метки значений каждой переменной, входящей в набор).

4. Щелкните правой кнопкой по набору *Производитель автомобилей* на панели холста и выберите во всплывающем меню **Категории и итоги**.
5. Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить**.
6. Теперь щелкните правой кнопкой мыши по набору *Производитель автомобилей* и выберите во всплывающем меню **Итожащие статистики**.
7. В диалоговом окне Итожащие статистики выберите в списке Статистики **Ответы** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
8. Нажмите кнопку **Применить к выделенному**, а затем нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Responses
Car maker, most recent cars	American	1938	1938
	Japanese	1327	1327
	Korean	695	695
	German	693	693
	Swedish	360	360
	Other	343	343
	Total	2832	5356

Рисунок 71. Категориальный набор множественных ответов без дубликатов

Как уже было сказано выше, по умолчанию повторяющиеся ответы не подсчитываются; поэтому в данной таблице значения для всех категорий в столбцах *Частота* и *Ответы* являются идентичными. Различаются только итоги.

9. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
10. Щелкните по вкладке **Параметры**.
11. Отметьте элемент **Подсчитывать количество повторяющихся ответов для наборов множественных ответов**.
12. Нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Responses
Car maker, most recent cars	American	1938	2797
	Japanese	1327	1717
	Korean	695	760
	German	693	754
	Swedish	360	383
	Other	343	359
	Total	2832	6770

Рисунок 72. Категориальный набор множественных ответов с повторяющимися ответами

В полученной таблице можно заметить существенные различия между значениями в столбцах *Частота* и *Ответы*, особенно для американских машин, что говорит о том, что многие респонденты несколько раз владели американскими машинами.

Критерии значимости для наборов множественных ответов

Вы можете использовать наборы множественных ответов в критериях значимости по существу так же, как вы использовали категориальные переменные.

- Критерий независимости (хи-квадрат) и сравнение пропорций столбцов (z-критерий) основаны на частотах, и для них Частота должна быть одной из итожащих статистик, выводимых в таблице.
- Для множественных категориальных наборов критерии, сравнивающих пропорции столбцов или средние для столбцов (t-критерий), не применяются, если на вкладке Параметры включен переключатель **Подсчитывать количество дублирующих ответов для наборов множественных ответов**. Дополнительную информацию смотрите в разделе “Настраиваемые таблицы: вкладка Опции” на стр. 16.

Критерии независимости для наборов множественных ответов

В данном примере формируется таблица сопряженности категориальной переменной и набора множественных ответов, и к ней применяется критерий независимости хи-квадрат.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
2. Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
3. Переместите набор *Источники новостей* (это метка набора множественных ответов \$m1tnews) из списка переменных в поле строк панели холста.
4. Перенесите переменную *Пол респондента* из списка переменных в поле Строки панели холста.
5. Щелкните по вкладке **Статистические критерии**.
6. Выберите **Критерий независимости (хи-квадрат)**.
7. Если не включен переключатель **Включить в проверки переменные множественных ответов**, включите его.
8. Нажмите **ОК**, чтобы запустить эту процедуру.

		News sources
Gender	Chi-square	10.266
	df	5
	Sig.	.068

Рисунок 73. Результаты применения критерия хи-квадрат

Уровень значимости критерия хи-квадрат, равный 0,068, означает, что мужчины и женщины значимо различаются в выборе новых источников (в предположении, что пороговым значением, при значении уровня значимости больше которого констатируется отсутствие значимости, выбрано значение, равное 0,05).

Сравнение средних по столбцам для наборов множественных ответов

В данном примере вычисляются средние значения количественной переменной по категориям, заданным набором множественных ответов, и среднее значение для каждой категории сравнивается со средними для всех остальных категорий, чтобы выявить значимые различия.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Пользовательские таблицы).
2. Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
3. Переместите набор *Источники новостей* (это метка набора множественных ответов \$m1tnews) из списка переменных в поле строк панели холста.
4. Перенесите переменную *Возраст респондента* из списка переменных в поле Строки панели холста.
5. Щелкните по вкладке **Статистические критерии**.
6. Включите переключатель **Сравнение средних по столбцам (t-критерий)**.
7. Если не включен переключатель **Включить в проверки переменные множественных ответов**, включите его.
8. Нажмите **ОК**, чтобы запустить эту процедуру.

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	Mean	Mean	Mean	Mean	Mean
Age of respondent	52	40	48	40	40

Comparisons of Column Means

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	(A)	(B)	(C)	(D)	(E)
Age of respondent	B C D E		B D E		

Results are based on two-sided tests assuming equal variances with significance level 0.05. For each significant pair, the key of the smaller category appears under the category with larger mean.

Рисунок 74. Результаты применения критерия значимости

- Каждая категория набора множественных ответов задается буквой (A, B, C, D, E), и для каждой категории выводится буква представляющая категорию, для которой среднее одновременно меньше и значимо отличается от среднего данной категории.
- *Получают новости из газет (A)* - наибольший средний возраст, а все другие средние по категориям существенно отличаются от этого среднего.
- *Получают новости по телевидению (C)* - следующий по величине максимальный средний возраст, все остальные средние по категориям (B, D и E) существенно отличаются. (C также значимо отличается от A, как показано выше.)
- Средние возрасты для категорий *Получают новости из журналов (B)*, *Получают новости, используя радио (D)* и *Получают новости из Интернета (E)* значимо не различаются между собой.

Пропущенные значения

Многие файлы данных содержат некоторое количество пропущенных значений. Возникновение пропущенных значений может объясняться различными причинами. К примеру, отсутствие значений может объясняться отсутствием ответов респондентов на некоторые вопросы, неприменимостью определенных переменных для отдельных наблюдений и ошибками кодировки.

В программе IBM SPSS Statistics существуют два вида значений отсутствия:

- **Пользовательские значения отсутствия.** Значения, соответствующие пропущенным данным. Для обозначения того, почему данные значения являются пропущенными, им могут присваиваться метки значений (такие как метка значения *Вопрос неприменим* для кода 99 в вопросе о беременности, задаваемом мужчинам).
- **Системные значения отсутствия.** Если для какого-либо наблюдения числовая переменная не содержит значения, такому наблюдению присваивается системное значение отсутствия. В представлении Данные Редактора данных это значение обозначается точкой или запятой, в зависимости от региональных установок на компьютере.

Существует ряд возможностей, которые могут помочь компенсировать влияние пропущенных значений, а также анализировать структуры в пропущенных данных. Однако у этой главы цель проще - описать, как настраиваемые таблицы обрабатывают отсутствующие данные и как отсутствующие данные влияют на вычисление итоговых статистик.

Файл данных примера

Для примеров в данной главе используется файл данных *missing_values.sav*. Дополнительную информацию смотрите в разделе файлы данных. Это очень простой полностью искусственный файл данных только с одной переменной и десятью наблюдениями, предназначенный для иллюстрации основных понятий, касающихся отсутствующих значений.

Таблицы без пропущенных значений

По умолчанию, пользовательские пропущенные значений не выводятся в настраиваемых таблицах (а системные значения отсутствия не выводятся никогда).

1. Выберите в меню:

Анализ > Таблицы > Настраиваемые таблицы...

2. В построителе таблиц перенесите переменную *Переменная с пропущенными значениями* (единственную переменную в файле) из списка переменных в поле Строки панели холста.
3. Щелкните правой кнопкой мыши по переменной на панели холста и выберите во всплывающем меню пункт **Категории и итоги**.
4. Щелкните (отметьте) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить**.
5. Щелкните правой кнопкой мыши по переменной *Переменная с пропущенными значениями* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
6. В диалоговом окне Итожащие статистики выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
7. Нажмите кнопку **Применить к выделенному**.

Можно заметить некоторое несоответствие между категориями, показанными на панели холста, и категориями, показанными в списке Категории (под списком переменных в левой части построителя таблиц). Список Категории содержит категорию с меткой *Пропущенные значения*, которая не включена в таблицу, поскольку категории, обозначающие пропущенные значения, по умолчанию не включаются в таблицу. Поскольку в метке содержится слово "категории", можно сделать вывод, что переменная имеет две или более пользовательские пропущенные категории.

8. Щелкните по **ОК**, чтобы создать таблицу.

		Count	Column N %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	Total	7	100.0%

Рисунок 75. Таблица без пропущенных значений

Таблица выглядит правильным образом. Сумма значений соответствует итогу, и проценты в таблице показывают, что в качестве базы для расчета было использовано итоговое количество наблюдений (к примеру, $3/7=0.429$, или 42.9%). Однако итоговое количество наблюдений не равно общему количеству наблюдений в файле данных; оно представляет собой общее количество наблюдений с **не пропущенными** значениями, или наблюдений, не имеющих пользовательских пропущенных значений или системных пропущенных значений в данной переменной.

Включение пропущенных значений в таблицы

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните правой кнопкой мыши по переменной *Переменная с пропущенными значениями* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
3. Щелкните (отметьте) элемент **Пропущенные значения** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить**.

Теперь таблица включает в себя категорию *Пропущенные значения*. Хотя в режиме предварительного просмотра показана только одна категория значений отсутствия, в таблице будут показаны все категории пользовательских значений отсутствия.

4. Щелкните правой кнопкой мыши по переменной *Переменная с пропущенными значениями* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
5. В диалоговом окне Итожащие статистики выберите (поставьте) **Задать статистики для итогов и подытогов**.
6. Выберите в списке Статистики **Валидные** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
7. Сделайте то же самое для статистики **Всего**.
8. Нажмите кнопку **Применить к выделенному**, а затем щелкните **ОК**, чтобы создать таблицу.

		Count	Column N %	Valid N	Total N
Variable with missing values	Low	2	22.2%		
	Medium	3	33.3%		
	High	2	22.2%		
	Don't know	1	11.1%		
	Not applicable	1	11.1%		
	Total	9	100.0%	7	10

Рисунок 76. Таблица с пропущенными значениями

Две пользовательские категории отсутствия значений, *Неизвестно* и *Неприменимо*, теперь выводятся в таблице, и общее количество составляет 9 вместо 7 из-за добавления двух наблюдений с пользовательскими значениями отсутствия (по одному в каждой пользовательской категории отсутствия). Значения процентов по столбцу также изменились, поскольку теперь в качестве базы используется суммарное количество присутствующих значений и пользовательских значений отсутствия. В подсчете процентов не учитываются только системные значения отсутствия.

Допустимые N показывает общее количество непропущенных наблюдений (7), а *Общее N* - полное количество наблюдений, в том числе и с пользовательскими и системными значениями отсутствия. Общее количество наблюдений равно 10, то есть на одно больше, чем количество присутствующих значений и пользовательских значений отсутствия, показанное в итоге для столбца *Количество*. Причина заключается в наличии одного наблюдения с системным значением отсутствия.

9. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
10. Щелкните правой кнопкой мыши по переменной *Переменная с пропущенными значениями* на панели холста и выберите во всплывающем меню пункт **Итожащие статистики**.
11. Выберите в верхнем списке Статистики (не в списке задаваемых статистик для итогов и подытогов) **% валидных по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
12. Сделайте то же самое для **Итоговый % по столбцу**.
13. Вы также можете добавить обе эти статистики в список задаваемых итожащих статистик для итогов и подытогов.
14. Нажмите кнопку **Применить к выделенному**, а затем нажмите кнопку **ОК**, чтобы создать таблицу.

		Count	Column N %	Column Valid N %	Column Total N %	Valid N	Total N
Variable with missing values	Low	2	22.2%	28.6%	20.0%		
	Medium	3	33.3%	42.9%	30.0%		
	High	2	22.2%	28.6%	20.0%		
	Don't know	1	11.1%	.0%	10.0%		
	Not applicable	1	11.1%	.0%	10.0%		
	Total	9	100.0%	100.0%	100.0%	7	10

Рисунок 77. Таблица с пропущенными значениями и процентами относительно валидных и всех наблюдений

- *% по столбцу* представляет собой процент в каждой категории, основанный на количестве присутствующих значений и пользовательских значений отсутствия (поскольку пользовательские значения отсутствия были включены в таблицу).
- *% валидных по столбцу* представляет собой процент, основанный только на количестве валидных, присутствующих значений. Эти значения равны процентам по столбцам в исходной таблице, в которую не были включены пользовательские значения отсутствия.
- *Итоговый % по столбцу* представляет собой процент, основанный на общем количестве наблюдений, включая наблюдения с пользовательскими и системными значениями отсутствия. Если сложить проценты в данном столбце, то сумма их будет составлять 90%, поскольку одно наблюдение из 10 (10%) содержит системное значение отсутствия. И хотя это наблюдение используется при расчете процентов, в таблице не выводится отдельная категория для системных значений отсутствия.

Форматирование и настройка таблиц

Форматирование и настройка таблиц

Настраиваемые таблицы обеспечивают возможность управления параметрами форматирования таблицы в процессе ее построения, включая следующие параметры:

- Формат и метки итожащих статистик
- Минимальная и максимальная ширина столбцов
- Текст или значение, выводимое в пустой ячейке

Эти установки задаются и остаются неизменными в окне построителя таблиц (до тех пор, пока вы не измените их, не сбросите все установки процедуры или не откроете новый файл данных), позволяя вам создавать несколько таблиц с одними и теми же параметрами форматирования, без ручного изменения этих таблиц. Параметры форматирования можно сохранить вместе с другими параметрами построения таблицы при помощи кнопки Вставка в диалоговом окне Настраиваемые таблицы, вставляя командный синтаксис в окно синтаксиса, который затем можно сохранить в виде файла.

Вы также можете изменять параметры форматирования таблиц после их создания, используя возможности форматирования сводных таблиц, доступные в окне средства просмотра. В этом разделе основное внимание, однако, уделяется свойствам форматирования, задаваемым перед созданием таблицы. Для получения дополнительной информации воспользуйтесь вкладкой Индекс в системе помощи и наберите в качестве ключевого слова сводные таблицы .

Файл данных примера

Для примеров в этой главе используется файл данных *survey_sample.sav*. Дополнительную информацию смотрите в разделе файлы данных.

Во всех примерах в диалоговых окнах выводятся метки переменных, отсортированные в алфавитном порядке. Вид, в котором выводится список переменных, устанавливается на вкладке Общие свойства в диалоговом окне Параметры (меню Правка, Параметры).

Формат вывода итожащих статистик

Настраиваемые таблицы по умолчанию предлагают достаточно удобные форматы вывода различных итожащих статистик, но возможно возникновение ситуаций, когда вы захотите использовать другие форматы.

1. Выберите в меню:
Анализ > Таблицы > Настраиваемые таблицы...
2. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.

3. Перенесите переменную *Доверие к телевидению* в поле Строки под переменную *Возрастная категория*, состыковывая эти две переменные в измерении строк.
4. Щелкните правой кнопкой по набору *Возрастная категория* на панели холста и выберите во всплывающем меню **Выделить все переменные строк**.
5. Снова щелкните правой кнопкой мыши по переменной *Возрастная категория* на панели макета таблицы и выберите во всплывающем меню пункт **Категории и итоги**.
6. Щелкните (установите переключатель) элемент **Итог** в диалоговом окне Категории и итоги, а затем нажмите кнопку **Применить**.
7. Снова щелкните правой кнопкой по любой переменной на панели холста и выберите во всплывающем меню **Итожащие статистики**.
8. Выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
9. Установите переключатель **Задать статистики для итогов и подытогов**.
10. В диалоговом окне Итожащие статистики выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
11. Сделайте то же самое для статистики **Среднее**.
12. Затем нажмите кнопку **Применить ко всем**.

Значения итожащих статистик, показанные в режиме предварительного просмотра таблицы, выводятся в формате, задаваемом по умолчанию.

- Для количеств формат по умолчанию имеет вид **nnnn** --целые числа без десятичных знаков.
- Для процентов формат по умолчанию имеет вид **nnnn.n%** --числа с одним десятичным знаком и знаком процента после значения.
- Формат, задаваемый по умолчанию для среднего значения, *отличается* для двух переменных, использованных в таблице.

Для итожащих статистик, которые не являются количествами (включая статистики **Все** и **Валидные**) или процентами, форматом, задаваемым по умолчанию, является формат вывода переменных в Редакторе данных. Если вы посмотрите на эти переменные в представлении **Переменные** окна Редактора Данных, то увидите, что переменная *Возрастная категория* (переменная *agecat*) имеет два десятичных знака, в то время как переменная *Доверие к телевидению* (переменная *contv*) не имеет десятичных знаков.

Это одна из тех ситуаций, когда формат по умолчанию не является наиболее подходящим форматом, поскольку было бы лучше, чтобы оба значения средних выводились с одним и тем же числом десятичных знаков.

13. Снова щелкните правой кнопкой по любой переменной на панели холста и выберите во всплывающем меню **Итожащие статистики**.
Для среднего значения в ячейке **Формат списка Показать** указан формат *Авто*, означающий, что в качестве формата используется формат переменной в Редакторе данных, и ячейка **Десятич** является не активной. Для того чтобы указать количество десятичных знаков, необходимо выбрать другой формат.
14. В списке задаваемых итожащих статистик **Показать** щелкните по ячейке **Формат для среднего значения**, и выберите **nnnn** из раскрывающегося списка форматов.
15. В ячейке **Десятич** введите значение **1**.
16. Затем щелкните **Применить ко всем**, чтобы применить данную установку к обоим переменным.

Теперь в режиме предварительного просмотра оба средних значения выводятся с одним десятичным знаком. (Вы можете создать таблицу, но вы должны понимать, что "среднее" значение для переменной *Возрастная категория* интерпретировать достаточно сложно, поскольку значения этой переменной изменяются от 1 до 6.)

Вывод меток для итожащих статистик

В дополнение к формату вывода итожащих статистик, вы можете также управлять метками для каждой итожащей статистики.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Нажмите кнопку **Сброс**, чтобы очистить заданные ранее в диалоговом окне параметры.
3. В построителе таблиц перенесите переменную *Возрастная категория* из списка переменных в поле Строки панели холста.
4. Перенесите переменную *Получение зарплаты за последнюю неделю* из списка переменных в поле Столбцы панели холста.
5. Щелкните правой кнопкой по переменной *Возрастная категория* на панели холста и выберите в контекстном меню **Итожащие статистики**.
6. Выберите в списке Статистики **% по столбцу** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
7. Дважды щелкните на слове *Столбец* в ячейке Метки в списке Показать, чтобы отредактировать содержимое данной ячейки. Удалите слово *Столбец* из метки, изменив метку на **%**.
8. Отредактируйте содержимое ячейки *Количество* таким же образом, изменив метку на **N**.
Изменим также формат статистики % по столбцу, убрав ненужный знак процента (поскольку метка столбца уже показывает, что столбец содержит проценты).
9. Щелкните по ячейке Формат для *% по столбцу* и выберите **nnnn.n** из раскрывающегося списка форматов.
10. Затем нажмите кнопку **Применить к выделенному**.
Теперь в режиме предварительного просмотра показывается измененный формат и метки.
11. Щелкните по **ОК**, чтобы создать таблицу.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Рисунок 78. Таблица с измененными метками итожащих статистик

Ширина столбца

Вы можете заметить, что таблица в предыдущем примере очень широкая. Один из способов решения данной проблемы заключается в том, чтобы поменять местами переменные строк и столбцов. Другое решение заключается в том, чтобы сделать столбцы более узкими, поскольку они кажутся более широкими, чем это необходимо. (Причина, по которой мы укоротили метки итожащих статистик, заключается в том, что мы хотели бы сделать столбцы более узкими.)

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы).
2. Щелкните по вкладке **Параметры**.
3. В группе Ширина столбцов выберите **Задаваемая**.
4. В поле Максимум наберите 36. (Убедитесь, что единицами измерения являются **Пункты**.)
5. Щелкните по **ОК**, чтобы создать таблицу.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Рисунок 79. Таблица с уменьшенной шириной столбцов

Теперь таблица выглядит более компактной.

Вид вывода пустых ячеек

По умолчанию, в пустых ячейках (ячейках, не содержащих наблюдений) выводится значение 0. Вы также можете указать вывод в ячейке пробела или указать текст, выводимый в пустых ячейках.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Щелкните по вкладке **Параметры**.
3. В группе Значения в ячейках в пункте Вид пустых ячеек выберите **Текст** и наберите Нет .
4. Щелкните по **ОК** , чтобы создать таблицу.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	None	None	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	None	None	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Рисунок 80. Таблица с текстом "Нет", выводимым в пустых ячейках

Теперь в четырех пустых ячейках в таблице вместо значения 0 выводится текст *Нет* .

Значения, показанные для тех статистик, которые невозможно рассчитать

Если статистика не может быть подсчитана, по умолчанию в ячейке выводится значение (.), причем этот символ используется и для отображения пропущенных значений в Редакторе данных. Эта ситуация отличается от ситуации с "пустой" ячейкой, поэтому символы, выводимые для пропущенных статистик, отличаются от символов, выводимых для ячеек, не содержащих наблюдений.

1. Откройте построитель таблиц (меню Анализ, Таблицы, Настраиваемые таблицы) .
2. Перенесите переменную *Число часов в день, проводимых у телевизора* из списка переменных в верхнюю часть поля Столбцы панели холста, над переменной *Получение зарплаты за последнюю неделю* .
Поскольку переменная *Число часов в день, проводимых у телевизора* является количественной переменной, она автоматически становится переменной-источником статистик и итожащая статистика меняется на среднее значение.
3. Щелкните правой кнопкой мыши по переменной *Число часов в день, проводимых у телевизора* на панели холста таблицы и выберите в контекстном меню пункт **Итожащие статистики** .
4. Выберите в списке Статистики **Валидные** и нажмите кнопку со стрелкой, чтобы добавить ее в список Показать.
5. Нажмите кнопку **Применить к выделенному** .
6. Щелкните по вкладке **Параметры**.
7. В текстовом поле Статистики, которые не могут быть подсчитаны введите Н0 .
8. Щелкните по **ОК** , чтобы создать таблицу.

		Hours per day watching TV											
		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N
Age category	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8
	25 to 34	3	134	5	2	2	30	2	29	2	52	2	22
	35 to 44	3	136	2	5	3	30	2	34	2	47	3	46
	45 to 54	2	90	2	4	2	22	2	36	2	45	2	34
	55 to 64	3	40	NA	None	3	7	2	15	2	23	3	15
	65 or older	3	18	2	2	1	1	NA	0	1	2	3	11

Рисунок 81. Таблица с текстом "НО", выводимым вместо статистик, которые не могут быть подсчитаны

Текст *НО* выводится вместо среднего значения в трех ячейках данной таблицы. В каждом случае соответствующее значение *Допустимые N* объясняет причину: Нет наблюдений, с помощью которых можно вычислить среднее.

Однако вы можете обратить внимание на некоторое видимое противоречие - одно из этих трех значений *Допустимые N* выводится как 0, а не с помощью метки *Нет*, которую можно предположить для ячейки без наблюдений. Причина заключается в том, что даже если в категории нет наблюдений для подсчета среднего значения, категория может не являться пустой. Если вернуться в исходную таблицу с двумя категориальными переменными, то можно увидеть, что в реальности в данной категории есть три наблюдения. Однако эти наблюдения не являются валидными, поскольку все они имеют пропущенные значения для количественной переменной *Число часов в день, проводимых у телевизора*.

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 ²52 студентов попросили оценить комбинации 15 ситуаций и 15 поведений по 10-балльной шкале от 0="очень подходит" до 9="абсолютно не подходит." С усреднением по индивидуумам значения принимались как различия.
- **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.

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.

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.

- **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 профиля 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.
- **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 льдом ⁶. 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.

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

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.

- **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.
- **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.

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/~mlearn/MLRepository.html>.

- **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), 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 Ким¹¹ 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*.

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.

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.

- **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 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

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.

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 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 продажи 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".

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.

- **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 в Statistics Base Edition.
- **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 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 плану 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 продажных 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 продажных 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.
- **whoeeze_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:

© (название вашей компании) (год). Части этого кода получены из примеров программ IBM Corp.

© Copyright IBM Corp. _введите год или годы_. 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

А

анализ с файлом разбиения
настраиваемые таблицы 4

В

Валидные 52, 77
вложение переменных
количественные переменные 55
Настраиваемые таблицы 27, 29
время
включение в настраиваемые таблицы
текущего времени 18
Всего 77
вывод значений категорий 50
вычисляемые категории
из подытогов 39
Настраиваемые таблицы 14, 38
скрытие категорий в выражении 39
форматы вывода 15

Г

групповые итоги 34
количественные переменные 54

Д

дата
включение в настраиваемые таблицы
текущей даты 18
десятичные знаки
управление количеством выводимых
десятичных знаков в настраиваемых
таблицах 6, 23, 79
диапазон
Настраиваемые таблицы 9
дисперсия
Настраиваемые таблицы 9
доверительные интервалы 56
допустимые N
Настраиваемые таблицы 9

З

заголовки
Настраиваемые таблицы 18
задаваемые статистики общих итогов 49
значения
вывод значений и меток категорий 50
значения и метки значений 50
значения отсутствия 52, 76

И

изменение порядка категорий
Настраиваемые таблицы 12

исключение категорий
Настраиваемые таблицы 12, 25
итоги
групповые итоги 34
исключенные категории 33
маргинальные итоги в настраиваемых
таблицах 24
Настраиваемые таблицы 12, 22, 32
слои 35
таблицы с вложениями 34
управление расположением 34
итоги подгрупп 34
итожащие статистики 45
задаваемые статистики общих
итогов 49
изменение текста метки 80
измерение источника 46
переменная источника 46
разные итожащие статистики для
разных переменных в составных
таблицах 53
составные таблицы 48
формат вывода 79

К

количественные переменные
вложение 55
групповые итоги 54
итоги по категориальным переменным
в строках и столбцах 55
итожащие статистики 51
несколько итожащих статистик 52
состыковывание 51
критерии значимости
наборы множественных ответов 75
Настраиваемые таблицы 18

М

максимум
Настраиваемые таблицы 9
медiana 52
Настраиваемые таблицы 9
метки
изменение текста меток итожащих
статистик 80
метки переменных
скрытие в настраиваемых таблицах 6
минимум
Настраиваемые таблицы 9

Н

наборы множественных ответов 69
дубликаты ответов в множественных
категориальных наборах 16
критерии значимости 69, 75
проценты 9

настраиваемые таблицы
анализ с файлом разбиения 4
заголовки 18
задаваемые итоги 10
изменение измеренных итожащих
статистик 10
изменение уровня измерения 1
исключение пропущенных значений для
количественных итогов 16
итожащие статистики 7, 8, 9
категориальные переменные 1
количественные переменные 1
метки значений для категориальных
переменных 1
множественные категориальные
наборы 16
наборы множественных ответов 1
подписи 18
проценты 7, 8
проценты для наборов множественных
ответов 9
пустые ячейки 16
статистические критерии 18
таблицы средних и частот 10
угловые метки 18
Ширина столбца 16
Настраиваемые таблицы
вложение переменных 27, 29
вложение переменных слоя 32
вычисляемые категории 12, 14, 38
изменение меток итожащих
статистик 21
изменение порядка категорий 12
измерение источника статистик 23
исключение категорий 12, 25
итоги 12, 22, 32
итоги в таблице с исключенными
категориями 25
как построить таблицу 4
компактный вид 29
критерии значимости и множественные
ответы 69
маргинальные итоги 24
наборы множественных ответов 69
объединение категорий 37
переменные слоя 30, 31, 32
печать слоев таблиц 32
подытоги 12, 32
показ и скрытие имен и меток
переменных 6
Поменять местами переменные строк и
столбцов 30
пост-вычисляемые категории 14, 38
простые таблицы с категориальными
переменными 20
проценты 21, 23
проценты по строкам и столбцам 21
скрытие меток статистик 20
сокрытие категорий, задающих
подытог 37
сортировка категорий 25

Настраиваемые таблицы *(продолжение)*
состыковывание переменных 26, 27
статистические критерии 57
таблица сопряженности 22
таблица частот 16, 42
таблицы для переменных с
одинаковыми вариантами
ответа 16, 42
таблицы переменных с одинаковыми
вариантами ответов 16, 42
управление количеством выводимых
десятичных знаков 6
формат вывода итожащих
статистик 11
форматы вывода 6

О

объединение категорий
Настраиваемые таблицы 37
отключение категорий
Настраиваемые таблицы 25

П

переменная источника итожащих
статистик
количественные переменные 55
переменные слоя
вложение переменных слоя 32
Настраиваемые таблицы 30, 31, 32
печать слоев таблиц 32
состыковывание переменных слоя 31
печать таблиц со слоями 32
подписи
Настраиваемые таблицы 18
подытоги 36
Настраиваемые таблицы 12, 32
сокрытие категорий, задающих
подытог 37
пользовательские пропущенные
значения 76
пост-вычисляемые категории
Настраиваемые таблицы 14, 38
пропорции по столбцам
настраиваемые таблицы 63
пропущенные значения
включение в настраиваемые
таблицы 77
влияние на расчет процентов 77
проценты
в настраиваемых таблицах 7, 8, 21, 23
значения отсутствия 77
наборы множественных ответов 9
пустые ячейки
выводимое значение в настраиваемых
таблицах 16, 82

Р

различные итожащие статистики для
различных переменных
составные таблицы 53
режим
Настраиваемые таблицы 9

С

системные пропущенные значения 76
сокрытие меток статистик в
настраиваемых таблицах 20
сортировка категорий
Настраиваемые таблицы 25
состыковывание переменных
количественные переменные 51
Настраиваемые таблицы 26, 27
несколько переменных источника
итожащих статистик 48
различные итожащие статистики для
различных переменных 53
состыковывание переменных слоя 31
среднее 51
Настраиваемые таблицы 9
среднеквадратичное отклонение
Настраиваемые таблицы 9
средние значения по столбцам
настраиваемые таблицы 60
статистики
задаваемые статистики общих
итогов 49
итожащие статистики 45
составные таблицы 48
статистики критериев
Настраиваемые таблицы 18
статистические критерии
Настраиваемые таблицы 57
сумма
Настраиваемые таблицы 9

Т

таблица сопряженности
Настраиваемые таблицы 22
таблица частот
Настраиваемые таблицы 16, 42
таблицы
Настраиваемые таблицы 1
таблицы для переменных с одинаковыми
вариантами ответа 16, 42
таблицы средних и частот 10, 49

У

угловые метки
Настраиваемые таблицы 18
удаление категорий
Настраиваемые таблицы 12, 25
управление количеством выводимых
десятичных знаков 23
уровень измерения
изменение в сводных таблицах 1

Ф

файлы для примеров
местоположение 83
форматы вывода 23
итожащие статистики в настраиваемых
таблицах 11, 79

Х

хи-квадрат
Настраиваемые таблицы 57

Ч

частота
валидные 52

Ш

ширина столбца
управление в настраиваемых
таблицах 16, 81



Напечатано в Дании