SPSS Tables™ 11.5

For more information about SPSS® software products, please visit our Web site at *http://www.spss.com*, or contact:

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412
Tel: (312) 651-3000
Fax: (312) 651-3668

SPSS Tables™ 11.5

Printed in the United States of America.

2 3 4 5 6 7 8 9 0 06 05 04 03

ISBN 1-56827-302-9

Preface

SPSS® 11.5 is a comprehensive system for analyzing data. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and complex statistical analyses.

The Tables option is an add-on enhancement that enables you to prepare customized tables suitable for presentation or publication. Through Tables, you can access a wide variety of descriptive statistics and can combine large amounts of information in a single display. The Tables option must be used with the SPSS 11.5 Base and is completely integrated into that system.

Professionals in many different fields will find the Tables procedure beneficial. People in business, for example, can use Tables for periodic status reports and for analyses that support decision making. Market researchers and survey researchers can use Tables to meet the tabular style requirements of academic institutions or professional journals. The flexibility of Tables allows you to follow a prescribed style or, if you choose, design your own.

About This Manual

This manual provides a guide to the Tables option and describes how to obtain the appropriate tables using the dialog box interface. It also gives an item-by-item description of each dialog box. The Tables command syntax, found near the end of the manual, is also included in the *SPSS 11.5 Syntax Reference Guide*, available on the product CD-ROM.

This manual contains two indexes: a subject index and a syntax index. The subject index covers the entire manual. The syntax index applies only to the syntax reference material.

Installation

To install Tables, follow the instructions for adding and removing features in the installation instructions supplied with the SPSS Base. (To start, double-click on the SPSS setup icon.)

Compatibility

The SPSS system is designed to operate on many computer systems. See the installation instructions that came with your system for specific information on minimum and recommended requirements.

Serial Numbers

Your serial number is your identification number with SPSS Inc. You will need this serial number when you contact SPSS Inc. for information regarding support, payment, or an upgraded system. The serial number was provided with your Base system. Before using the system, please copy this number to the registration card.

Registration Card

Don't put it off: *fill out and send us your registration card.* Until we receive your registration card, you have an unregistered system. Even if you have previously sent a card to us, please fill out and return the card enclosed in your Tables package. Registering your system entitles you to:

- Technical support services
- New product announcements and upgrade announcements

Customer Service

If you have any questions concerning your shipment or account, contact your local office, listed on page vi. Please have your serial number ready for identification when calling.

Training Seminars

SPSS Inc. provides both public and onsite training seminars for SPSS. All seminars feature hands-on workshops. SPSS seminars will be offered in major U.S. and European cities on a regular basis. For more information about these seminars, call your local office, listed on page vi.

Technical Support

The services of SPSS Technical Support are available to registered customers of SPSS. Customers may contact Technical Support for assistance in using SPSS products or for installation help for one of the supported hardware environments. To reach Technical Support, see the SPSS Web site at *http://www.spss.com* or call your local office, listed on page vi. Be prepared to identify yourself, your organization, and the serial number of your system.

Additional Publications

Except for academic course adoptions, additional copies of SPSS product manuals may be purchased directly from SPSS Inc. Visit our Web site at *http://www.spss.com/estore*, or contact your local SPSS office, listed on page vi.

SPSS product manuals may also be purchased from Prentice Hall, the exclusive distributor of SPSS publications. To order, fill out and mail the Publications order form included with your system, or call 800-947-7700. If you represent a bookstore or have an account with Prentice Hall, call 800-382-3419. In Canada, call 800-567-3800. Outside of North America, contact your local Prentice Hall office.

Statistical introductions to procedures in the Base, Regression Models, and Advanced Models written by Marija Norušis are planned to be available from Prentice Hall. Check with the publisher or visit the SPSS Web site for announcements regarding availability.

Tell Us Your Thoughts

Your comments are important. Please let us know about your experiences with SPSS products. We especially like to hear about new and interesting applications using the SPSS system. Please send e-mail to *suggest@spss.com*, or write to SPSS Inc., Attn: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Contacting SPSS

If you would like to be on our mailing list, contact one of our offices, listed on page vi, or visit our Web site at *http://www.spss.com*. We will send you a copy of our newsletter and let you know about SPSS Inc. activities in your area.

SPSS Inc.
Chicago, Illinois, U.S.A.
Tel: 1.312.651.3000
or 1.800.543.2185
www.spss.com/corpinfo
Customer Service:
1.800.521.1337
Sales:
1.800.543.2185
sales@spss.com
Training:
1.800.543.6607
Technical Support:
1.312.651.3410
support@spss.com

SPSS Federal Systems
Tel: 1.703.740.2400
or 1.800.860.5762
www.spss.com

SPSS Andino
Tel: +57.1.635.8585
www.spss.com/la

SPSS Argentina SA
Tel: +5411.4371.5031
www.spss.com/la

SPSS Asia Pacific Pte. Ltd.
Tel: +65.245.9110
www.spss.com

SPSS Australasia Pty. Ltd.
Tel: +61.2.9954.5660
www.spss.com

SPSS Belgium
Tel: +32.163.170.70
www.spss.com

SPSS Benelux BV
Tel: +31.183.651777
www.spss.com

SPSS Brasil Ltda.
Tel: +55.11.5505.3644
www.spss.com

SPSS Chile
Tel: +56.2.233.7499
www.spss.com/la

SPSS Czech Republic
Tel: +420.2.24813839
www.spss.cz

SPSS Denmark
Tel: +45.45.46.02.00
www.spss.com

SPSS East Africa
Tel: +254 2 577 262
spss.com

SPSS Finland Oy
Tel: +358.9.4355.920
www.spss.com

SPSS France SARL
Tel: +01.55.35.27.00
www.spss.com

SPSS Germany
Tel: +49.89.4890740
www.spss.com

SPSS BI Greece
Tel: +30.1.6971950
www.spss.com

SPSS Iberica, S.L.U.
Tel: +34.902.123.606
SPSS.com

SPSS Hong Kong Ltd.
Tel: +852.2.811.9662
www.spss.com

SPSS Ireland
Tel: +353.1.415.0234
www.spss.com

SPSS BI Israel
Tel: +972.3.6166616
www.spss.com

SPSS Italia srl
Tel: +800.437300
www.spss.it

SPSS Japan Inc.
Tel: +81.3.5466.5511
www.spss.co.jp

SPSS Korea DataSolution Co.
Tel: +82.2.563.0014
www.spss.co.kr

SPSS Latin America
Tel: +1.312.651.3539
www.spss.com

SPSS Malaysia Sdn Bhd
Tel: +603.6203.2300
www.spss.com

SPSS Miami
Tel: 1.305.627.5700
SPSS.com

SPSS Mexico SA de CV
Tel: +52.5.682.87.68
www.spss.com

SPSS Norway AS
Tel: +47.22.99.25.50
www.spss.com

SPSS Polska
Tel: +48.12.6369680
www.spss.pl

SPSS Russia
Tel: +7.095.125.0069
www.spss.com

SPSS San Bruno
Tel: 1.650.794.2692
www.spss.com

SPSS Schweiz AG
Tel: +41.1.266.90.30
www.spss.com

SPSS BI (Singapore) Pte. Ltd.
Tel: +65.346.2061
www.spss.com

SPSS South Africa
Tel: +27.21.7120929
www.spss.com

SPSS South Asia
Tel: +91.80.2088069
www.spss.com

SPSS Sweden AB
Tel: +46.8.506.105.50
www.spss.com

SPSS Taiwan Corp.
Taipei, Republic of China
Tel: +886.2.25771100
www.sinter.com.tw/spss/main

SPSS (Thailand) Co., Ltd.
Tel: +66.2.260.7070
www.spss.com

SPSS UK Ltd.
Tel: +44.1483.719200
www.spss.com

Contents

Chapter

1

Getting Started with SPSS Tables

Many procedures in SPSS produce results in the form of tables. The SPSS Tables option, however, offers special features designed to support a wide variety of customized reporting capabilities. Many of the custom features are particularly useful for survey analysis and marketing research.

This guide assumes that you already know the basics of using SPSS. If you are unfamiliar with the basic operation of SPSS, see the introductory tutorial provided with the software. From the menu bar in any open SPSS window, choose:

Help
Tutorial

What's New in Tables?

If you have used Tables in the past, you will quickly discover that just about everything is new, including:

- A simple drag-and-drop table builder interface that allows you to preview your table as you select variables and options.
- A single, unified table builder interface instead of multiple menu choices and dialog boxes for different types of tables.
- New, simpler, easy-to-understand CTABLES command syntax in place of TABLES command syntax. (A conversion program is available to convert old TABLES jobs to CTABLES.)
- Subtotals for subsets of categories of a categorical variable.

- Custom control over category display order and ability to selectively show or hide categories.

Figure 1-1
Table builder with table preview

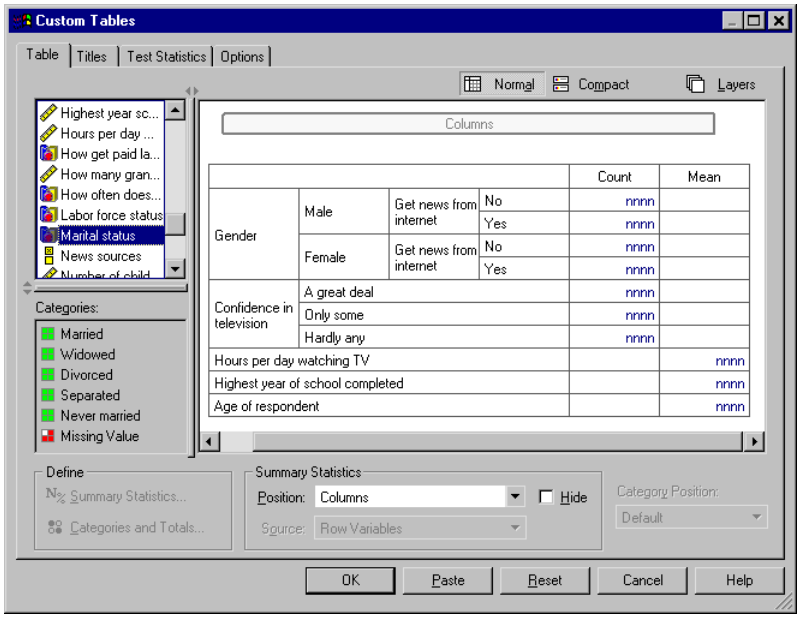


Table Structure and Terminology

SPSS Tables can produce a wide variety of customized tables. While you can discover a great deal of its capabilities simply by experimenting with the table builder interface, it may be helpful to know something about basic table structure in SPSS and the terms we use to describe different structural elements that you can use in a table.

Pivot Tables

Tables produced by SPSS Tables are displayed as **pivot tables** in the Viewer window. Pivot tables provide a great deal of flexibility over the formatting and presentation of tables.

For a general discussion of pivot tables, use the Help system.

- From the menus in any open SPSS window, choose:

  Help
  Topics

- In the Contents pane, click Base System.

- Then click Pivot Tables in the expanded contents list.

Variables and Level of Measurement

To a certain extent, what you can do with a variable in a table is limited by its defined level of measurement. The Tables procedure makes a distinction between two basic types of variables, based on level of measurement:

Categorical. Data with a limited number of distinct values or categories (for example, gender or religion). Also referred to as qualitative data. Categorical variables can be string (alphanumeric) data or numeric variables that use numeric codes to represent categories (for example, 0 = *Female* and 1 = *Male*). Categorical variables can be further divided into:

- **Nominal.** Categorical data where there is no inherent order to the categories. For example, a job category of "sales" isn't higher or lower than a job category of "marketing" or "research."
- **Ordinal.** Categorical data where there is a meaningful order of categories but there isn't a measurable distance between categories. For example, there is an order to the values high, medium, and low, but the "distance" between the values can't be calculated.

Variables defined as nominal or ordinal in the Data Editor are treated as categorical variables in the Tables procedure.

Scale. Data measured on an interval or ratio scale, where the data values indicate both the order of values and the distance between values. For example, a salary of $72,195 is higher than a salary of $52,398, and the distance between the two values is $19,797. Also referred to as quantitative, or continuous, data.

Variables defined as scale in the Data Editor are treated as scale variables in the Tables procedure.

Value Labels

For categorical variables, the preview displayed on the canvas pane in the table builder relies on defined **value labels**. The categories displayed in the table are, in fact, the defined value labels for that variable. If there are no defined value labels for the variable, the preview displays two generic categories. The actual number of categories that will be displayed in the final table is determined by the number of distinct values that occur in the data. The preview simply assumes that there will be at least two categories.

Additionally, some custom table-building features are not available for categorical variables that have no defined value labels.

Rows, Columns, and Cells

Each dimension of a table is defined by a single variable or a combination of variables. Variables that appear down the left side of a table are called **row variables**. They define the rows in a table. Variables that appear across the top of a table are called **column variables**. They define the columns in a table. The body of a table is made up of **cells**, which contain the basic information conveyed by the table—counts, sums, means, percentages, and so on. A cell is formed by the intersection of a row and column of a table.

Stacking

Stacking can be thought of as taking separate tables and pasting them together into the same display. For example, you could display information on *Gender* and *Age category* in separate sections of the same table.

Figure 1-2
Stacked variables

		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

Although the term “stacking” typically denotes a vertical display, you can also stack variables horizontally.

Figure 1-3
Horizontal stacking

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

Crosstabulation

Crosstabulation is a basic technique for examining the relationship between two categorical variables. For example, using *Age category* as a row variable and *Gender* as a column variable, you can create a two-dimensional crosstabulation that shows the number of males and females in each age category.

Figure 1-4
Simple two-dimensional crosstabulation

		Gender	
		Male	Female
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

Nesting

Nesting, like crosstabulation, can show the relationship between two categorical variables, except one variable is nested within the other in the same dimension. For example, you could nest *Gender* within *Age category* in the row dimension, showing the number of males and females in each age category.

In this example, the nested table displays essentially the same information as a crosstabulation of the same two variables.

Figure 1-5
Nested variables

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

Layers

You can use layers to add a dimension of depth to your tables, creating three-dimensional "cubes." Layers are, in fact, quite similar to nesting; the primary difference is that only one layer category is visible at a time. For example, using *Age category* as the row variable and *Gender* as a layer variable produces a table in which information for males and females is displayed in different layers of the table.

Figure 1-6
Layered variables

Gender Female

Age category	Less than 25	134
	25 to 34	351
	35 to 44	370

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

Tables for Variables with Shared Categories

Surveys often contain many questions with a common set of possible responses. For example, our sample survey contains a number of variables concerning confidence in various public and private institutions and services, all with the same set of response categories: 1 = *A great deal*, 2 = *Only some*, and 3 = *Hardly any.* You can use stacking to display these related variables in the same table—and you can display the shared response categories in the columns of the table.

Figure 1-7
Stacked variables with shared response categories in columns

	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

Multiple Response Sets

Multiple response sets use multiple variables to record responses to questions where the respondent can give more than one answer. For example, our sample survey asks the question, “Which of the following sources do you rely on for news?” Respondents can select any combination of five possible choices: Internet, television, radio, newspapers, and news magazines. Each of these choices is stored as a separate variable in the data file, and together they make a multiple response set. With Tables, you can define a multiple response set based on these variables and use that multiple response set in the tables you create.

Figure 1-8
Multiple response set displayed in a table

		Count	Column %
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%

You may notice in this example that the percentages total to more than 100%. This is because the total number of responses can be greater than the total number of respondents, since each respondent may choose more than one answer.

Totals and Subtotals

Tables provides a great deal of control over the display of totals and subtotals, including:

- Overall row and column totals
- Group totals for nested, stacked, and layered tables
- Subgroup totals

Figure 1-9
Subtotals, group totals, and table totals

		Count	Percent
Male	Less than 25	108	8.8%
	25 to 34	276	22.5%
	35 to 44	309	25.2%
	Subtotal < 45	693	56.4%
	45 to 54	221	18.0%
	55 to 64	136	11.1%
	65 or older	178	14.5%
	Subtotal 45+	535	43.6%
	Total	1228	100.0%
Female	Less than 25	134	8.4%
	25 to 34	351	21.9%
	35 to 44	370	23.1%
	Subtotal < 45	855	53.4%
	45 to 54	260	16.3%
	55 to 64	184	11.5%
	65 or older	301	18.8%
	Subtotal 45+	745	46.6%
	Total	1600	100.0%
Total	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	Subtotal < 45	1548	54.7%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Subtotal 45+	1280	45.3%
	Total	2828	100.0%

Custom Summary Statistics for Totals

For tables that contain totals or subtotals, you can have different summary statistics than the summaries displayed for each category. For example, you could display counts for an ordinal categorical row variable and display the mean for the “total” statistic.

Figure 1-10
Categorical variable and summary statistics in the same dimension

Confidence in television	1 A great deal	Count	196
	2 Only some	Count	936
	3 Hardly any	Count	744
	Total	Count	1876
		Mean	2.29

Sample Data File

Most of the examples presented here use the data file *survey_sample.sav*. This data file is a fictitious survey of several thousand people, containing basic demographic information and responses to a variety of questions, ranging from political views to television viewing habits.

All sample files used in these examples are located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

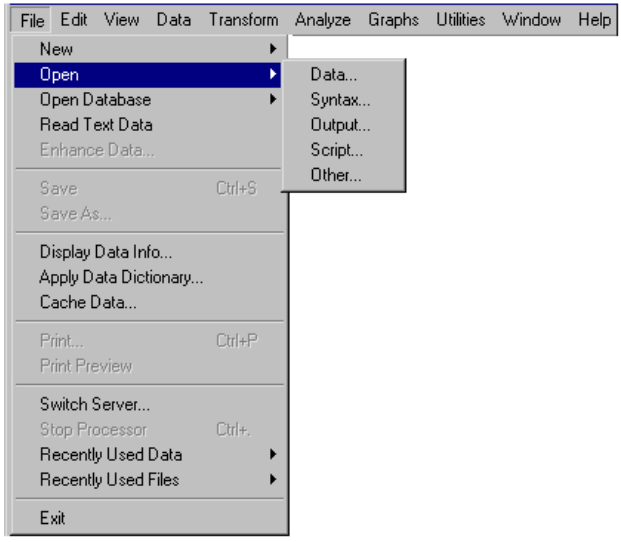
Building a Table

Before you can build a table, you need some data to use in the table.

- From the menus, choose:

File
  Open
    Data...

Figure 1-11
File menu

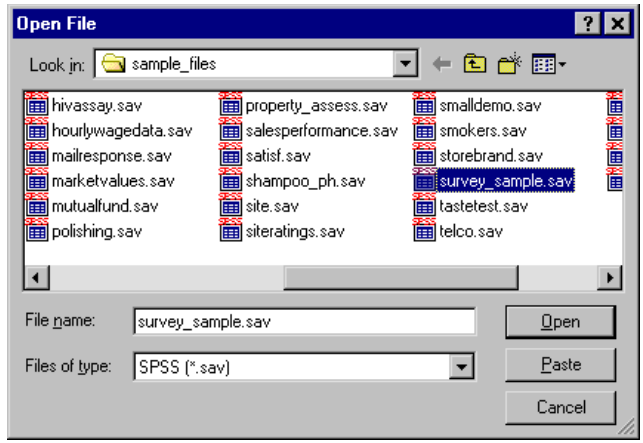


Alternatively, you can use the Open File button on the toolbar.

Figure 1-12
Open File toolbar button

This opens the Open File dialog box.

Figure 1-13
Sample_files folder displayed in Open File dialog box



- To use the data file in this example, use the Open File dialog box to navigate to the *tutorial\sample_files* folder, located in the folder in which SPSS is installed (typically, *c:\program files\spss*).

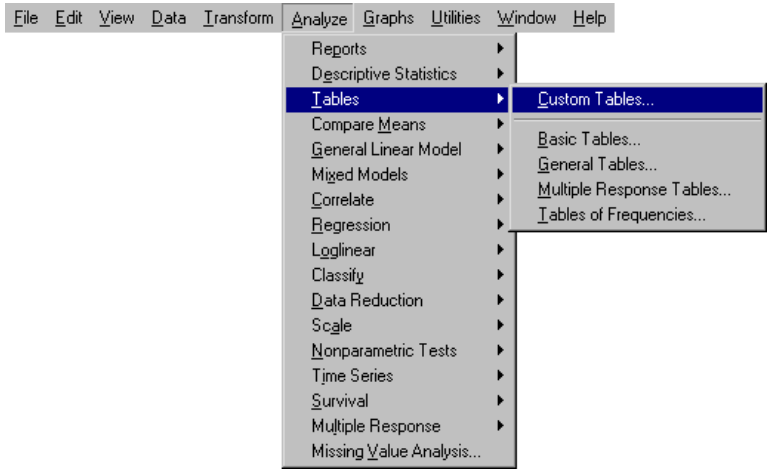
- Select *survey_sample.sav* and then click Open.

Opening the Custom Table Builder

- To open the custom table builder, from the menus, choose:

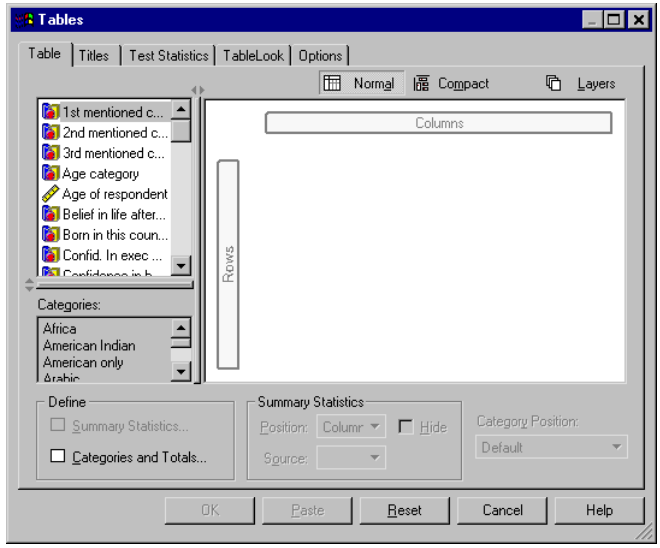
Analyze
  Tables
    Custom Tables...

Figure 1-14
Analyze menu, Tables



This opens the custom table builder.

Figure 1-15
Custom table builder

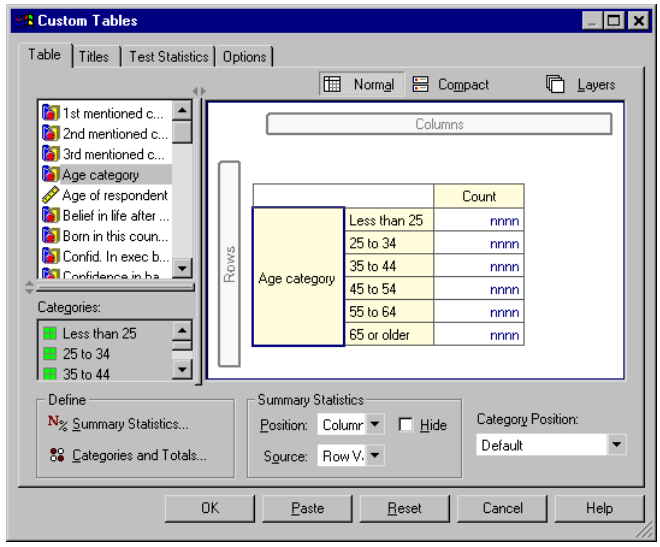


Selecting Row and Column Variables

To create a table, you simply drag and drop variables where you want them to appear in the table.

- Select (click) *Age category* in the variable list and drag and drop it into the Rows area on the canvas pane.

Figure 1-16
Selecting a row variable

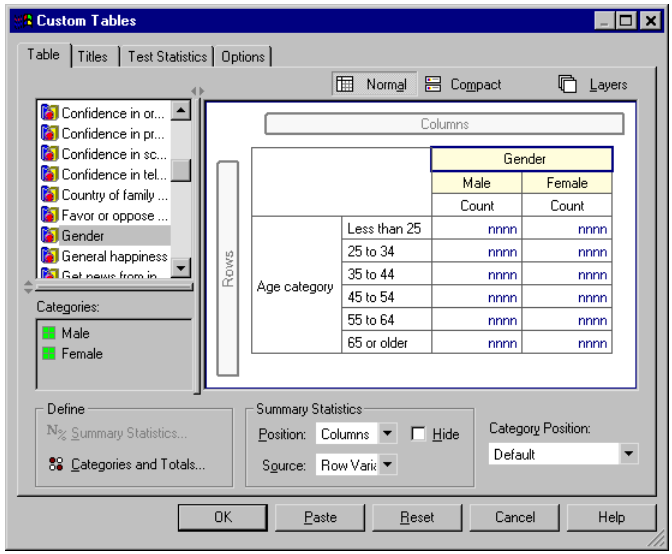


The canvas pane displays the table that would be created using this single row variable.

The preview does not display the actual values that would be displayed in the table; it displays only the basic layout of the table.

- Select *Gender* in the variable list and drag and drop it into the Columns area on the canvas pane (you may have to scroll down the variable list to find this variable).

Figure 1-17
Selecting a column variable

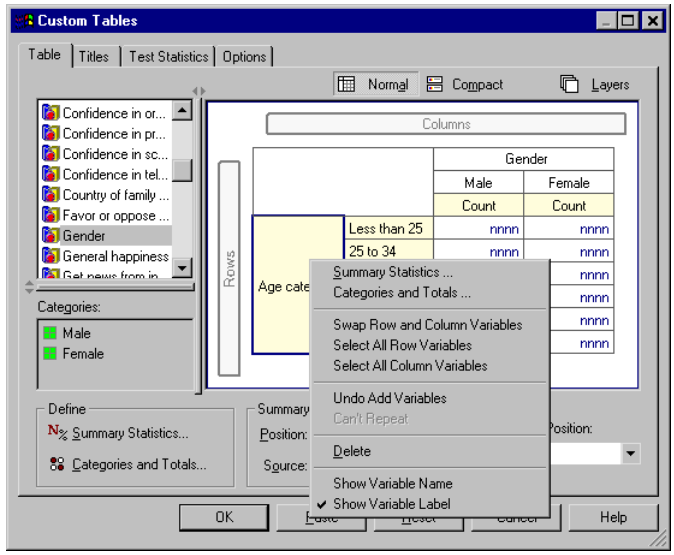


The canvas pane now displays a two-way crosstabulation of *Age category* by *Gender.*

By default, counts are displayed in the cells for categorical variables. You can also display row, column, and/or total percentages.

- Right click on *Age category* on the canvas pane and select Summary Statistics from the pop-up context menu.

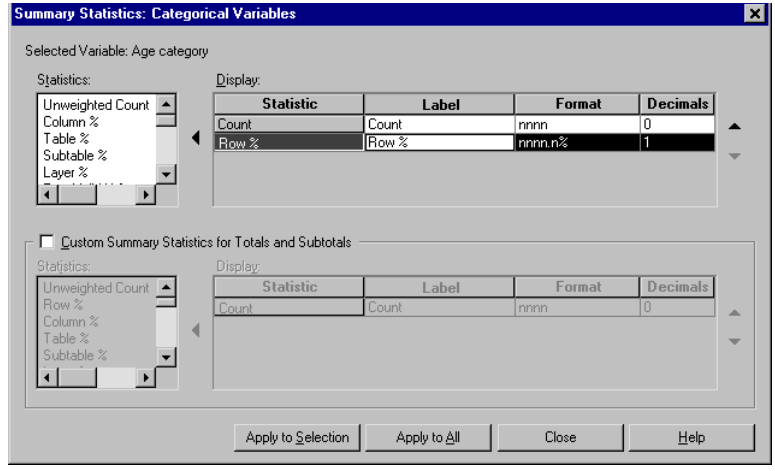
Figure 1-18
Context menu for categorical variables on canvas pane



- In the Summary Statistics dialog box, select Row % in the Statistics list and click the arrow button to add it to the Display list.

Now both the counts and row percentages will be displayed in the table.

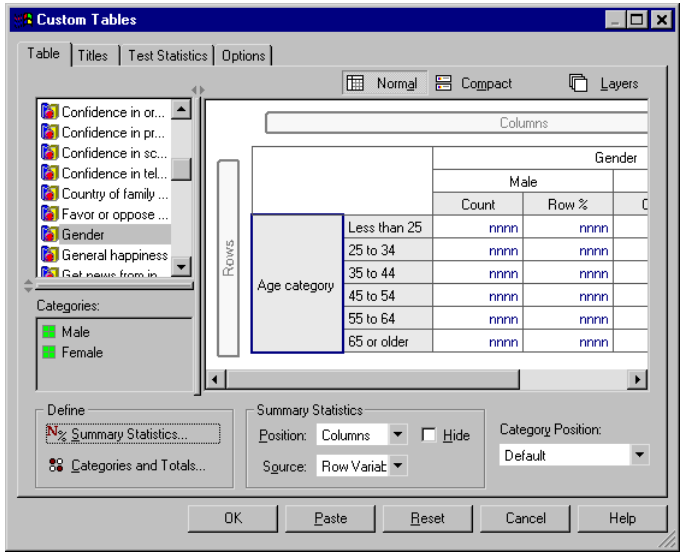
Figure 1-19
Summary Statistics dialog box for categorical variables



- Click Apply to Selection to save these settings and return to the table builder.

The canvas pane reflects the changes you have made, displaying columns for both counts and row percentages.

Figure 1-20
Counts and row percentages displayed on canvas pane



Inserting Totals and Subtotals

Totals are not displayed by default in custom tables, but it's easy to add both totals and subtotals to a table.

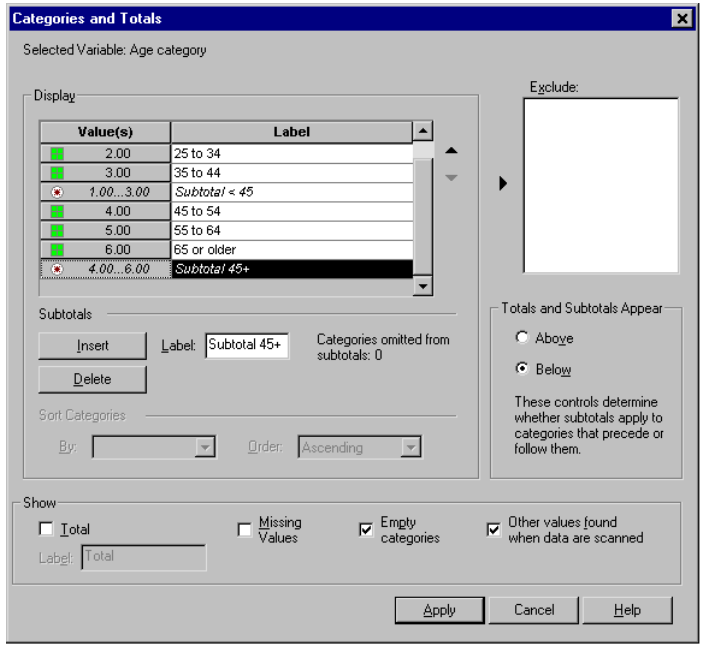
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.
- In the Categories and Totals dialog box, select (click) 3.00 in the Value(s) list.
- In the Label text field next to the Insert button, type Subtotal < 45.
- Then click the Insert button.

This inserts a row containing the subtotal for the first three age categories.

- Select (click) 6.00 in the Value(s) list.
- In the Label text field next to the Insert button, type Subtotal 45+.
- Then click the Insert button.

  This inserts a row containing the subtotal for the last three age categories.
- To include an overall total, select the Total check box.

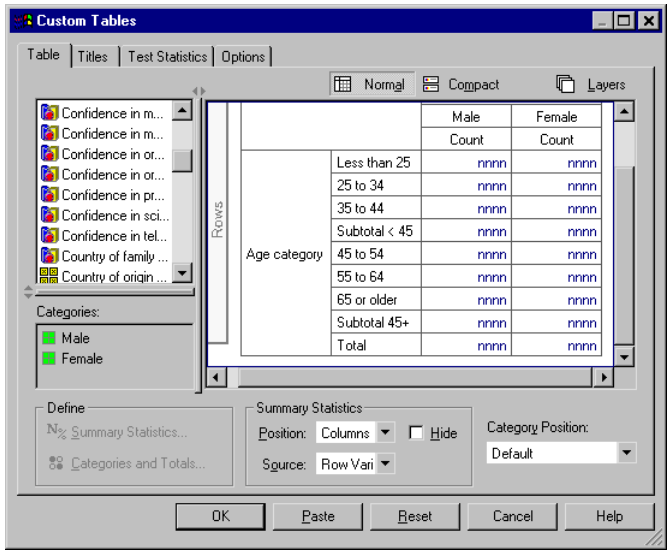
Figure 1-21
Inserting totals and subtotals



- Then click Apply.

The canvas pane preview now includes rows for the two subtotals and the overall total.

Figure 1-22
Total and subtotals on canvas pane



► Click OK to produce this table.

The table is displayed in the Viewer.

Figure 1-23
Crosstabulation with totals and subtotals

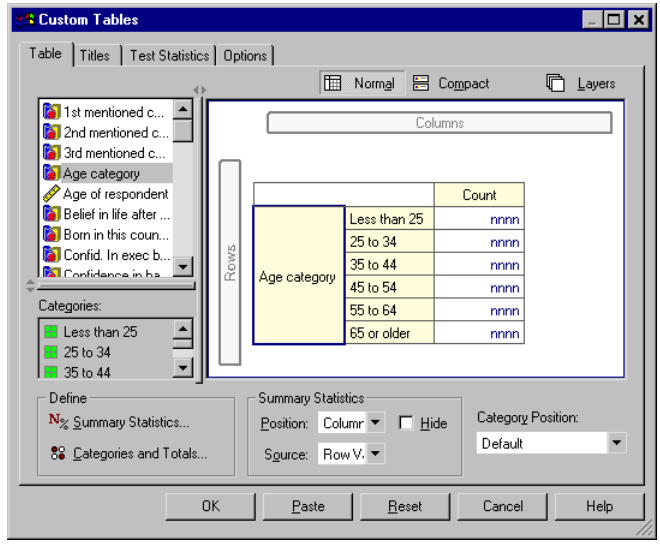
		Gender			
		Male		Female	
		Count	Row %	Count	Row %
Age category	Less than 25	108	44.6%	134	55.4%
	25 to 34	276	44.0%	351	56.0%
	35 to 44	309	45.5%	370	54.5%
	Subtotal < 45	693	44.8%	855	55.2%
	45 to 54	221	45.9%	260	54.1%
	55 to 64	136	42.5%	184	57.5%
	65 or older	178	37.2%	301	62.8%
	Subtotal 45+	535	41.8%	745	58.2%
	Total	1228	43.4%	1600	56.6%

Summarizing Scale Variables

A simple crosstabulation of two categorical variables displays counts or percentages in the cells of the table, but you can also display summaries of scale variables in the cells of the table.

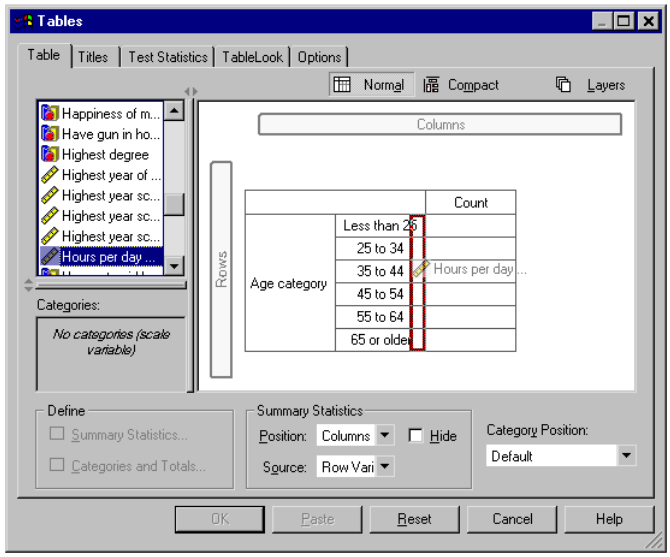
- Open the custom table builder again (Analyze menu, Tables, Custom Tables).
- Click Reset to clear any previous selections.
- Select (click) *Age category* in the variable list and drag and drop it into the Rows area on the canvas pane.

Figure 1-24
Selecting a row variable



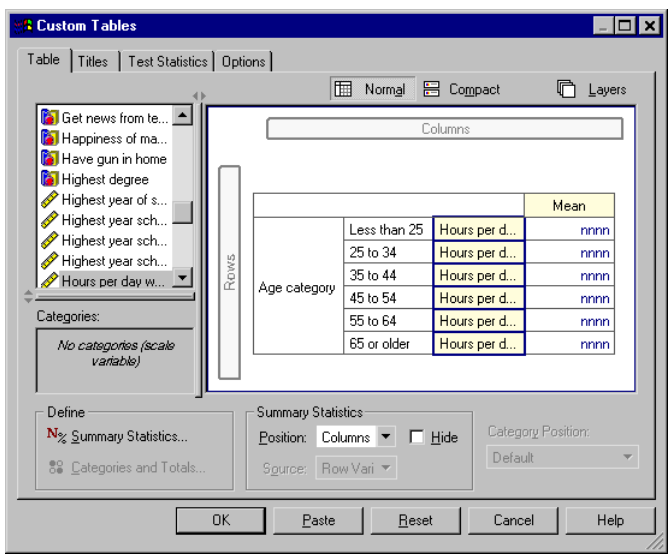
- Select *Hours per day watching TV* in the variable list and drag and drop it to the right of *Age category* in the row dimension of the table.

Figure 1-25
Dragging and dropping a scale variable into the row dimension



Now, instead of category counts, the table will display the mean (average) number of hours of television watched for each age category.

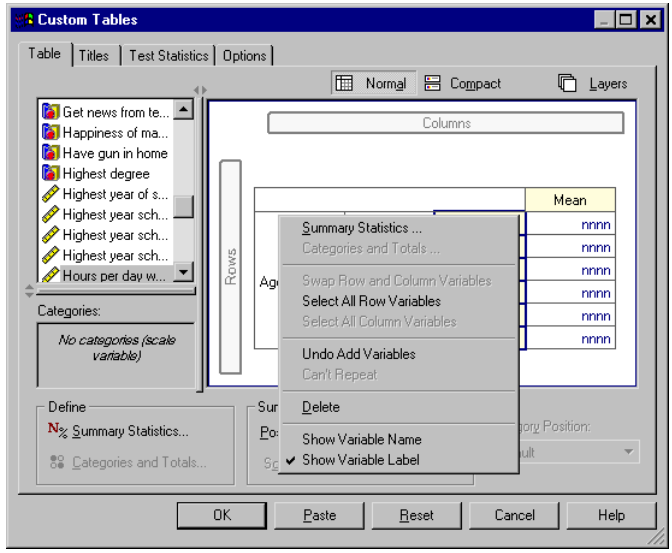
Figure 1-26
Scale variable summarized in table cells



The mean is the default summary statistic for scale variables. You can add or change the summary statistics displayed in the table.

- Right-click the scale variable on the canvas pane, and select Summary Statistics from the pop-up context menu.

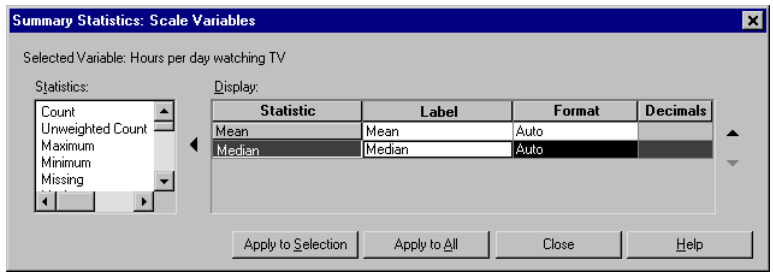
Figure 1-27
Context menu for scale variables in table preview



- In the Summary Statistics dialog box, select Median in the Statistics list and click the arrow button to add it to the Display list.

Now both the mean and the median will be displayed in the table.

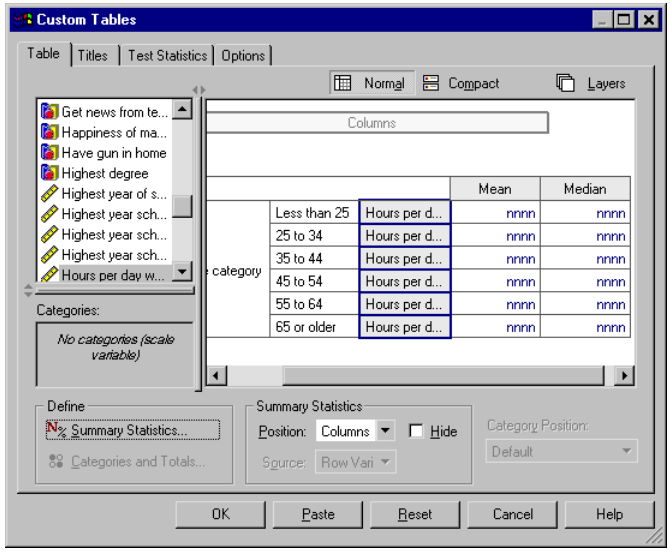
Figure 1-28
Summary Statistics dialog box for scale variables



- Click Apply to Selection to save these settings and return to the table builder.

The canvas pane now shows that both the mean and median will be displayed in the table.

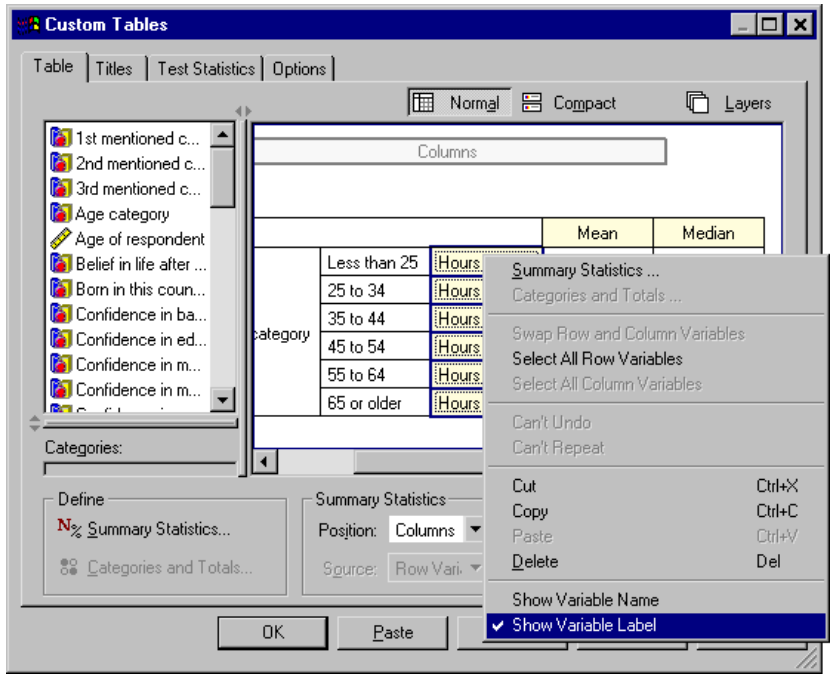
Figure 1-29
Mean and median scale summaries displayed on canvas pane



Before creating this table, let's clean it up a bit.

- Right-click on *Hours per day...* on the canvas pane and deselect (uncheck) Show Variable Label on the pop-up context menu.

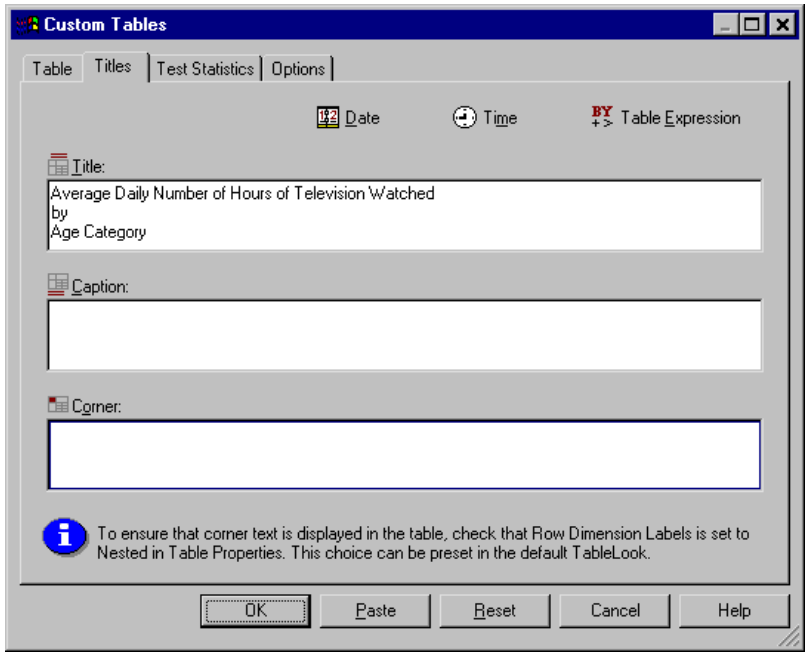
Figure 1-30
Suppressing the display of variable labels



The column is still displayed in the table preview (with the variable label text grayed out), but this column will not be displayed in the final table.

- Click the Titles tab in the table builder.
- Enter a descriptive title for the table, such as Average Daily Number of Hours of Television Watched by Age Category.

Figure 1-31
Custom Tables, Titles tab



▶ Click OK to create the table.

The table is displayed in the Viewer window.

Figure 1-32
Mean and median number of TV hours by age category

Average Daily Number of Hours of Television Watched
by
Age Category

		Mean	Median
Age category	Less than 25	2.85	2.00
	25 to 34	2.78	2.00
	35 to 44	2.56	2.00
	45 to 54	2.58	2.00
	55 to 64	3.02	2.50
	65 or older	3.58	3.00

Chapter

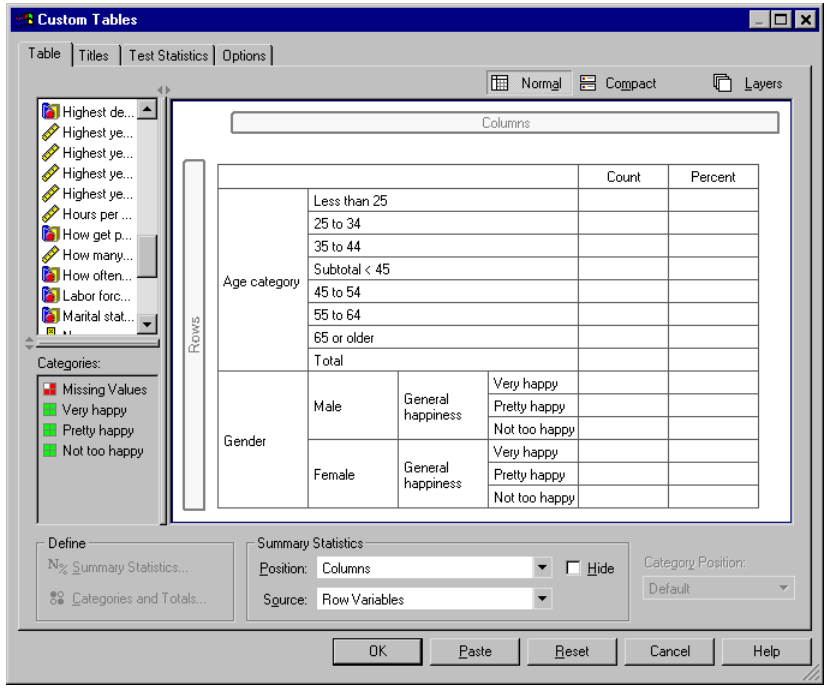
2

Table Builder Interface

Custom Tables uses a simple drag-and-drop table builder interface that allows you to preview your table as you select variables and options. It also provides a level of flexibility not found in a typical “dialog box,” including the ability to change the size of the window and the size of the panes within the window.

Building Tables

Figure 2-1
Custom Tables: Table tab



You select the variables and summary measures that will appear in your tables on the Table tab in the table builder.

Variable list. The variables in the data file are displayed in the top left pane of the window. Custom Tables distinguishes between two different measurement levels for variables and handles them differently depending on the measurement level:

- **Categorical.** Data with a limited number of distinct values or categories (for example, gender or religion). Categorical variables can be string (alphanumeric) or numeric variables that use numeric codes to represent categories (for example, 0 = male and 1 = female). Also referred to as qualitative data.
- **Scale.** Data measured on an interval or ratio scale, where the data values indicate both the order of values and the distance between values. For example, a salary of $72,195 is higher than a salary of $52,398, and the distance between the two values is $19,797. Also referred to as quantitative or continuous data.

Categorical variables define categories (row, columns, and layers) in the table, and the default summary statistic is the count (number of cases in each category). For example, a default table of a categorical gender variable would simply display the number of males and the number of females.

Scale variables are typically summarized within categories of categorical variables, and the default summary statistic is the mean. For example, a default table of income within gender categories would display the mean income for males and the mean income for females.

You can also summarize scale variables by themselves, without using a categorical variable to define groups. This is primarily useful for **stacking** summaries of multiple scale variables. For more information, see "Stacking Variables" below.

Multiple Response Sets

Custom Tables also supports a special kind of "variable" called a **multiple response set**. Multiple response sets aren't really "variables" in the normal sense. You can't see them in the Data Editor, and other procedures don't recognize them. Multiple response sets use multiple variables to record responses to questions where the respondent can give more than one answer. Multiple response sets are treated like categorical variables, and most of the things you can do with categorical variables, you can also do with multiple response sets. For more information, see the chapter *Multiple Response Sets*.

An icon next to each variable in the variable list identifies the variable type.


Scale


Categorical


Multiple response set, multiple categories


Multiple response set, multiple dichotomies

You can change the measurement level of a variable in the table builder by right-clicking the variable in the variable list and selecting Categorical or Scale from the pop-up context menu. You can permanently change a variable's measurement level in

the Variable View of the Data Editor. Variables defined as **nominal** or **ordinal** are treated as categorical by Custom Tables.

Categories. When you select a categorical variable in the variable list, the defined categories for the variable are displayed in the Categories list. These categories will also be displayed on the canvas pane when you use the variable in a table. If the variable has no defined categories, the Categories list and the canvas pane will display two placeholder categories: *Category 1* and *Category 2*.

The defined categories displayed in the table builder are based on **value labels**, descriptive labels assigned to different data values (for example, numeric values of 0 and 1, with value labels of *male* and *female*). You can define value labels in Variable View of the Data Editor or with Define Variable Properties on the Data menu in the Data Editor window.

Canvas pane. You build a table by dragging and dropping variables onto the rows and columns of the canvas pane. The canvas pane displays a preview of the table that will be created. The canvas pane doesn't show actual data values in the cells, but it should provide a fairly accurate view of the layout of the final table. For categorical variables, the actual table may contain more categories than the preview if the data file contains unique values for which no value labels have been defined.

- Normal view displays all of the rows and columns that will be included in the table, including rows and/or columns for summary statistics and categories of categorical variables.
- Compact view shows only the variables that will be in the table, without a preview of the rows and columns that the table will contain.

Basic Rules and Limitations for Building a Table

- For categorical variables, summary statistics are based on the innermost variable in the statistics source dimension.
- The default statistics source dimension (row or column) for categorical variables is based on the order in which you drag and drop variables into the canvas pane. For example, if you drag a variable to the rows tray first, the row dimension is the default statistics source dimension.
- Scale variables can be summarized only within categories of the innermost variable in either the row or column dimension. (You can position the scale variable at any level of the table, but it is summarized at the innermost level.)

- Scale variables cannot be summarized within other scale variables. You can stack summaries of multiple scale variables or summarize scale variables within categories of categorical variables. You cannot nest one scale variable within another or put one scale variable in the row dimension and another scale variable in the column dimension.

To Build a Table

▶ From the menus, choose:

Analyze
  Tables
    Custom Tables...

▶ Drag and drop one or more variables to the row and/or column areas of the canvas pane.

▶ Click OK to create the table.

To delete a variable from the canvas pane in the table builder:

▶ Select (click) the variable on the canvas pane.

▶ Drag the variable anywhere outside the canvas pane, or press the Delete key.

To change the measurement level of a variable:

▶ Right-click the variable in the variable list (you can do this only in the variable list, not on the canvas).

▶ Select Categorical or Scale from the pop-up context menu.

Stacking Variables

Stacking can be thought of as taking separate tables and pasting them together into the same display. For example, you could display information on gender and age category in separate sections of the same table.

To Stack Variables

- In the variable list, select all of the variables you want to stack, and drag and drop them together into the rows or columns of the canvas pane.

  or

- Drag and drop variables separately, dropping each variable either above or below existing variables in the rows or to the right or left of existing variables in the columns.

Figure 2-2
Stacked variables

		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

For more information, see “Stacking Categorical Variables” in the chapter *Stacking, Nesting, and Layers with Categorical Variables*.

Nesting Variables

Nesting, like crosstabulation, can show the relationship between two categorical variables, except that one variable is nested within the other in the same dimension. For example, you could nest gender within age category in the row dimension, showing the number of males and females in each age category.

You can also nest a scale variable within a categorical variable. For example, you could nest income within gender, showing separate mean (or median or other summary measure) income values for males and females.

To Nest Variables

- Drag and drop a categorical variable into the row or column area of the canvas pane.
- Drag and drop a categorical or scale variable to the left or right of the categorical row variable or above or below the categorical column variable.

Figure 2-3
Nested categorical variables

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

Figure 2-4
Scale variable nested within a categorical variable

Average Daily Number of Hours of Television Watched
by
Age Category

		Mean	Median
Age category	Less than 25	2.85	2.00
	25 to 34	2.78	2.00
	35 to 44	2.56	2.00
	45 to 54	2.58	2.00
	55 to 64	3.02	2.50
	65 or older	3.58	3.00

Note: Technically, the preceding table is an example of a categorical variable nested within a scale variable, but the resulting information conveyed in the table is essentially the same as nesting the scale variable within the categorical variable, without redundant labels for the scale variable. (Try it the other way around and you'll understand.)

For more information, see "Nesting Categorical Variables" in the chapter *Stacking, Nesting, and Layers with Categorical Variables*.

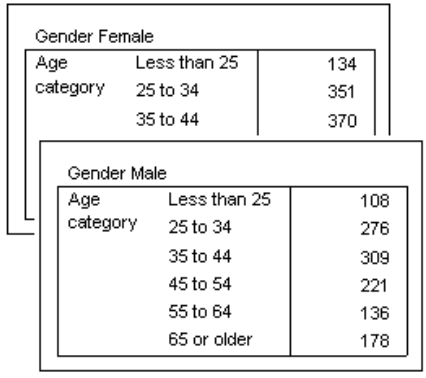
Layers

You can use layers to add a dimension of depth to your tables, creating three-dimensional "cubes." Layers are similar to nesting or stacking; the primary difference is that only one layer category is visible at a time. For example, using age category as the row variable and gender as a layer variable produces a table in which information for males and females is displayed in different layers of the table.

To Create Layers

- Click Layers on the Table tab in the table builder to display the Layers list.
- Drag and drop the scale or categorical variable(s) that will define the layers into the Layers list. You can also drag and drop variables onto the Layers button without displaying the Layers list.

Figure 2-5
Layered variables



You cannot mix scale and categorical variables in the Layers list. All variables must be of the same type. Multiple response sets are treated as categorical for the Layers list. Scale variables in the layers are always stacked.

If you have multiple categorical layer variables, layers can be stacked or nested.

- Show each category as a layer is equivalent to stacking. A separate layer will be displayed for each category of each layer variable. The total number of layers is simply the *sum* of the number of categories for each layer variable. For

example, if you have three layer variables, each with three categories, the table will have nine layers.

- Show each combination of categories as a layer is equivalent to nesting or crosstabulating layers. The total number of layers is the *product* of the number of categories for each layer variable. For example, if you have three variables, each with three categories, the table will have 27 layers.

Showing and Hiding Variable Names and/or Labels

The following options are available for the display of variable names and labels:

- Show only variable labels. For any variables without defined variable labels, the variable name is displayed. This is the default setting.
- Show only variable names.
- Show both variable labels and variable names.
- Don't show variable names or variable labels. Although the column/row that contains the variable label or name will still be displayed in the table preview on the canvas pane, this column/row will not be displayed in the actual table.

To show or hide variable labels or variable names:

▶ Right-click the variable in the table preview on the canvas pane.

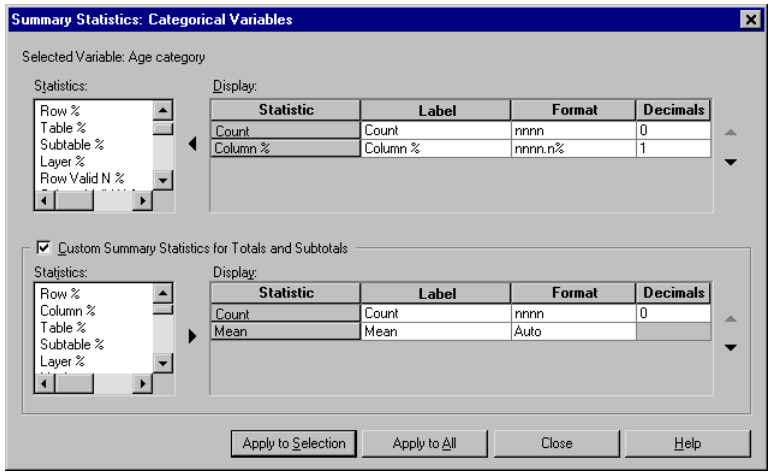
▶ Select Show Variable Label or Show Variable Name from the pop-up context menu to toggle the display of labels or names on or off. A check mark next to the selection indicates that it will be displayed.

Summary Statistics

The Summary Statistics dialog box allows you to:

- Add and remove summary statistics from a table.
- Change the labels for the statistics.
- Change the order of the statistics.
- Change the format of the statistics, including the number of decimal positions.

Figure 2-6
Summary Statistics dialog box



The summary statistics (and other options) available here depend on the measurement level of the summary statistics source variable. The source of summary statistics (the variable on which the summary statistics are based) is determined by:

- **Measurement level**. If a table (or a table section in a stacked table) contains a scale variable, summary statistics are based on the scale variable.
- **Variable selection order**. The default statistics source dimension (row or column) for categorical variables is based on the order in which you drag and drop variables onto the canvas pane. For example, if you drag a variable to the rows area first, the row dimension is the default statistics source dimension.
- **Nesting**. For categorical variables, summary statistics are based on the innermost variable in the statistics source dimension.

A stacked table may have multiple summary statistics source variables (both scale and categorical), but each table section has only one summary statistics source.

To Change the Summary Statistics Source Dimension

▶ Select the dimension (rows, columns, or layers) from the Source drop-down list in the Summary Statistics group of the Table tab.

To Control the Summary Statistics Displayed in a Table

- Select (click) the summary statistics source variable on the canvas pane of the Table tab.
- In the Define group of the Table tab, click Summary Statistics.

  or

- Right-click the summary statistics source variable on the canvas pane and select Summary Statistics from the pop-up context menu.
- Select the summary statistics you want to include in the table. You can use the arrow to move selected statistics from the Statistics list to the the Display list, or you can drag and drop selected statistics from the Statistics list into the Display list.
- Click the up or down arrows to change the display position of the currently selected summary statistic.
- Select a display format from the Format drop-down list for the selected summary statistic.
- Enter the number of decimals to display in the Decimals cell for the selected summary statistic.
- Click Apply to Selection to include the selected summary statistics for the currently selected variables on the canvas pane.
- Click Apply to All to include the selected summary statistics for all stacked variables of the same type on the canvas pane.

  Note: Apply to All differs from Apply to Selection only for stacked variables of the same type already on the canvas pane. In both cases, the selected summary statistics are automatically included for any additional stacked variables of the same type that you add to the table.

Summary Statistics for Categorical Variables

The basic statistics available for categorical variables are counts and percentages. You can also specify custom summary statistics for totals and subtotals. These custom summary statistics include measures of central tendency (such as mean and median) and dispersion (such as standard deviation) that may be suitable for some ordinal categorical variables. For more information, see "Custom Total Summary Statistics for Categorical Variables" below.

Count. Number of cases in each cell of the table or number of responses for multiple response sets.

Unweighted Count. Unweighted number of cases in each cell of the table.

Column percentages. Percentages within each column. The percentages in each column of a subtable (for simple percentages) sum to 100%. Column percentages are typically useful only if you have a categorical *row* variable.

Row percentages. Percentages within each row. The percentages in each row of a subtable (for simple percentages) sum to 100%. Row percentages are typically useful only if you have a categorical *column* variable.

Layer Row and Layer Column percentages. Row or column percentages (for simple percentages) sum to 100% across all subtables in a nested table. If the table contains layers, row or column percentages sum to 100% across all nested subtables in each layer.

Layer percentages. Percentages within each layer. For simple percentages, cell percentages within the currently visible layer sum to 100%. If you don't have any layer variables, this is equivalent to table percentages.

Table percentages. Percentages for each cell are based on the entire table. All cell percentages are based on the same total number of cases and sum to 100% (for simple percentages) over the entire table.

Subtable percentages. Percentages in each cell are based on the subtable. All cell percentages in the subtable are based the same total number of cases and sum to 100% within the subtable. In nested tables, the variable that precedes the innermost nesting level defines subtables. For example, in a table of marital status within gender within age category, gender defines subtables.

Multiple response sets can have percentages based on cases, responses, or counts. For more information, see "Summary Statistics for Multiple Response Sets" below.

Stacked Tables

For percentage calculations, each table section defined by a stacking variable is treated as a separate table. Layer Row, Layer Column, and Table percentages sum to 100% (for simple percentages) within each stacked table section. The percentage base for different percentage calculations is based on the cases in each stacked table section.

Percentage Base

Percentages can be calculated in three different ways, determined by the treatment of missing values in the base (denominator):

Simple percentage. Percentatges are based on the number of cases used in the table and always sum to 100%. If a category is excluded from the table, cases in that category are excluded from the base. Cases with system-missing values are always excluded from the base. Cases with user-missing values are excluded if user-missing categories are excluded from the table (the default) or included if user-missing categories are included in the table. Any percentage that doesn't have "Valid N" or "Total N" in its name is a simple percentage.

Total N percentage. Cases with system-missing and user-missing values are added to the Simple percentage base. Percentages may sum to less than 100%.

Valid N percentage. Cases with user-missing values are removed from the Simple percentage base even if user-missing categories are included in the table. Percentages may sum to more than 100%.

Note: Cases in manually excluded categories other than user-missing categories are always excluded from the base.

Summary Statistics for Multiple Response Sets

The following additional summary statistics are available for multiple response sets.

Col/Row/Layer Responses %. Percentage based on responses.

Col/Row/Layer Responses % (Base: Count). Responses are the numerator and total count is the denominator.

Col/Row/Layer Count % (Base: Responses). Count is the numerator and total responses are the denominator.

Layer Col/Row Responses %. Percentage across subtables. Percentage based on responses.

Layer Col/Row Responses % (Base: Count). Percentages across subtables. Responses are the numerator and total count is the denominator.

Layer Col/RowResponses % (Base: Responses). Percentages across subtables. Count is the numerator and total responses is the denominator.

Responses. Count of responses.

Subtable/Table Responses %. Percentage based on responses.

Subtable/Table Responses % (Base: Count). Responses are the numerator and total count is the denominator.

Subtable/Table Count % (Base: Responses). Count is the numerator and total responses are the denominator.

Summary Statistics for Scale Variables and Categorical Custom Totals

In addition to the counts and percentages available for categorical variables, the following summary statistics are available for scale variables and as custom total and subtotal summaries for categorical variables. These summary statistics are not available for multiple response sets or string (alphanumeric) variables.

Mean. Arithmetic average; the sum divided by the number of cases.

Median. Value above and below which half of the cases fall; the 50th percentile.

Mode. Most frequent value. If there is a tie, the smallest value is shown.

Minimum. Smallest (lowest) value.

Maximum. Largest (highest) value.

Missing. Count of missing values (both user- and system-missing).

Percentile. You can include the 5th, 25th, 75th, 95th, and/or 99th percentiles.

Range. Difference between maximum and minimum values.

Standard error of the mean. A measure of how much the value of the mean may vary from sample to sample taken from the same distribution. It can be used to roughly compare the observed mean to a hypothesized value (that is, you can conclude that the

two values are different if the ratio of the difference to the standard error is less than –2 or greater than +2).

Standard deviation. A measure of dispersion around the mean. In a normal distribution, 68% of the cases fall within one standard deviation of the mean and 95% of the cases fall within two standard deviations. For example, if the mean age is 45, with a standard deviation of 10, 95% of the cases would be between 25 and 65 in a normal distribution (the square root of the variance).

Sum. Sum of the values.

Sum percentage. Percentages based on sums. Available for rows and columns (within subtables), entire rows and columns (across subtables), layers, subtables, and entire tables.

Total N. Count of non-missing, user-missing, and system-missing values. Does not include cases in manually excluded categories other than user-missing categories.

Valid N. Count of non-missing values. Does not include cases in manually excluded categories other than user-missing categories.

Variance. A measure of dispersion around the mean, equal to the sum of squared deviations from the mean divided by one less than the number of cases. The variance is measured in units that are the square of those of the variable itself (the square of the standard deviation).

Stacked Tables

Each table section defined by a stacking variable is treated as a separate table, and summary statistics are calculated accordingly.

Custom Total Summary Statistics for Categorical Variables

For tables of categorical variables that contain totals or subtotals, you can have different summary statistics than the summaries displayed for each category. For example, you could display counts and column percentages for an ordinal categorical row variable and display the median for the “total” statistic.

To create a table for a categorical variable with a custom total summary statistic:

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Drag and drop a categorical variable into the Rows or Columns area of the canvas.
- Right-click on the variable on the canvas and select Categories and Totals from the pop-up context menu.
- Click (check) the Total check box, and then click Apply.
- Right-click the variable again on the canvas and select Summary Statistics from the pop-up context menu.
- Click (check) Custom Summary Statistics for Totals and Subtotals, and then select the custom summary statistics you want.

By default, all summary statistics, including custom summaries, are displayed in the opposite dimension from the dimension containing the categorical variable. For example, if you have a categorical row variable, summary statistics define columns in the table, as in:

Figure 2-7
Default position of summary statistics

		Count	Mean
Confidence in television	1 A great deal	196	
	2 Only some	936	
	3 Hardly any	744	
	Total	1876	2.29

To display summary statistics in the same dimension as the categorical variable:

- On the Table tab in the table builder, in the Summary Statistics group, select the dimension from the Position drop-down list.

  For example, if the categorical variable is displayed in the rows, select Rows from the drop-down list.

Figure 2-8
Categorical variable and summary statistics in the same dimension

Confidence in television	1 A great deal	Count	196
	2 Only some	Count	936
	3 Hardly any	Count	744
	Total	Count	1876
		Mean	2.29

Summary Statistics Display Formats

The following display format options are available:

nnnn. Simple numeric.

nnnn%. Percent sign appended to end of value.

Auto. Defined variable display format, including number of decimals.

N=nnnn. Displays "N=" before the value. This can be useful for counts, valid N, and total N in tables where the summary statistics labels are not displayed.

(nnnn). All values enclosed in parentheses.

(nnnn)(neg. value). Only negative values enclosed in parentheses.

(nnnn%). All values enclosed in parentheses and percent sign appended to end of values.

n,nnn.n. Comma format. Comma used as grouping separator and period used as decimal indicator regardless of locale settings.

n.nnn,n. Dot format. Period used as grouping separator and comma used as decimal indicator regardless of locale settings.

$n,nnn.n. Dollar format. Dollar sign displayed in front of value; comma used as grouping separator and period used as decimal indicator regardless of locale settings.

CCA, CCB, CCC, CCD, CCE. Custom currency formats. The current defined format for each custom currency is displayed in the list. These formats are defined on the Currency tab in the SPSS Options dialog box (Edit menu, Options).

General Rules and Limitations

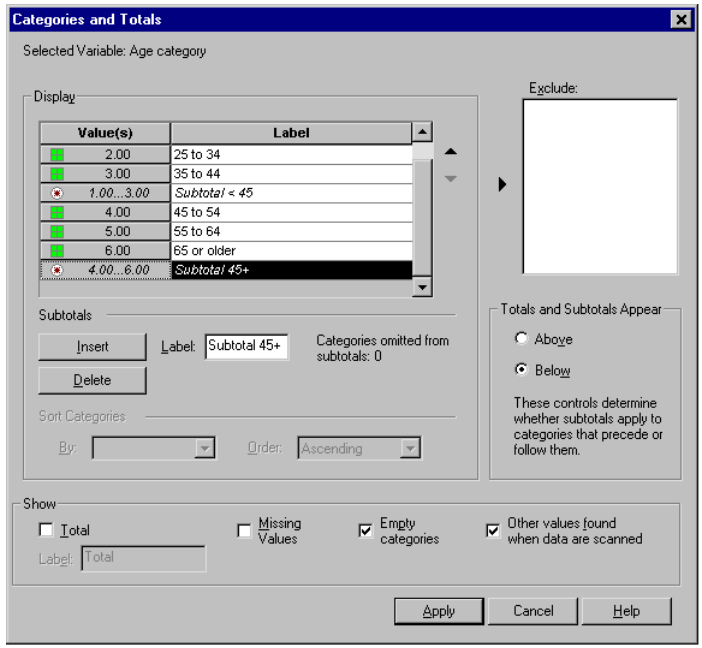
- With the exception of Auto, the number of decimals is determined by the Decimals column setting.
- With the exception of the comma, dollar, and dot formats, the decimal indicator used is the one defined for the current locale in your Windows Regional Options control panel.
- Although comma/dollar and dot will display either a comma or period respectively as the grouping separator, there is no display format available at creation time to display a grouping separator based on the current locale settings (defined in the Windows Regional Options control panel).

Categories and Totals

The Categories and Totals dialog box allows you to:

- Reorder and exclude categories.
- Insert subtotals and totals.
- Include or exclude empty categories.
- Include or exclude categories defined as containing missing values.
- Include or exclude categories that do not have defined value labels.

Figure 2-9
Categories and Totals dialog box



- This dialog box is available only for categorical variables and multiple response sets. It is not available for scale variables.
- For multiple selected variables with different categories, you cannot insert subtotals, exclude categories, or manually reorder categories. This occurs only if you select multiple variables in the canvas preview and access this dialog box for all selected variables simultaneously. You can still perform these actions for each variable separately.
- For variables with no defined value labels, you can only sort categories and insert totals.

To Access the Categories and Totals Dialog Box

- Drag and drop a categorical variable or multiple response set onto the canvas pane.
- Right-click the variable on the canvas pane, and select Categories and Totals from the pop-up context menu.

  or

- Select (click) the variable on the canvas pane, and then click Categories and Totals in the Define group on the Table tab.

You can also select multiple categorical variables in the same dimension on the canvas pane:

- Ctrl-click each variable on the canvas pane.

  or

- Click outside the table preview on the canvas pane, and then click and drag to select the area that includes the variables you want to select.

  or

- Right-click any variable in a dimension and select Select All [dimension] Variables to select all of the variables in that dimension.

To Reorder Categories

To manually reorder categories:

- Select (click) a category in the list.
- Click the up or down arrow to move the category up or down in the list.

  or

- Click in the Value(s) column for the category, and drag and drop it in a different position.

To Exclude Categories

- Select (click) a category in the list.
- Click the arrow next to the Exclude list.

  or

- Click in the Value(s) column for the category and drag and drop it anywhere outside the list.

  If you exclude any categories, any categories without defined value labels will also be excluded.

To Sort Categories

You can sort categories by data value, value label, or cell count in ascending or descending order.

- In the Sort Categories group, click the By drop-down list and select the sort criterion you want to use (value, label, or cell count).
- Click the Order drop-down list to select the sort order (ascending or descending).

  Sorting categories is not available if you have excluded any categories.

Subtotals

- Select (click) the category in the list that is the last category in the range of categories that you want to include in the subtotal.
- Click Insert. You can also modify the subtotal label text.

Totals

- Click the Total check box. You can also modify the total label text.

  If the selected variable is nested within another variable, totals will be inserted for each subtable.

Display Position for Totals and Subtotals

Totals and subtotals can be displayed above or below the categories included in each total.

- If Below is selected in the Totals and Subtotals Appear group, totals appear above each subtable, and all categories above and including the selected category (but below any preceding subtotals) are included in each subtotal.
- If Above is selected in the Totals and Subtotals Appear group, totals appear below each subtable, and all categories below and including the selected category (but above any preceding subtotals) are included in each subtotal.

Important: You should select the display position for subtotals before defining any subtotals. Changing the display position affects all subtotals (not just the currently selected subtotal), and it also *changes the categories included in the subtotals.*

Custom Total and Subtotal Summary Statistics

You can display statistics other than "totals" in the Totals and Subtotals areas of the table using the Summary Statistics dialog box. For more information, see "Summary Statistics for Categorical Variables" above.

Totals, Subtotals, and Excluded Categories

Cases from excluded categories are not included in the calculation of totals.

Missing Values, Empty Categories, and Values without Value Labels

Missing values. This controls the display of **user-missing** values, or values defined as containing missing values (for example, a code of 99 to represent "not applicable" for pregnancy in males). By default, user-missing values are excluded. Select (check) this option to include user-missing categories in tables. Although the variable may contain more than one missing value category, the table preview on the canvas will display only one generic missing value category. All defined user-missing categories will be included in the table. **System-missing values** (empty cells for numeric variables in the Data Editor) are always excluded.

Empty categories. Empty categories are categories with defined value labels but no cases in that category for a particular table or subtable. By default, empty categories are included in tables. Deselect (uncheck) this option to exclude missing categories from the table.

Other values found when data are scanned. By default, category values in the data file that do not have defined value labels are automatically included in tables. Deselect (uncheck) this option to exclude values without defined value labels from the table. If you exclude any categories with defined value labels, categories without defined value labels are also excluded.

Tables of Variables with Shared Categories (Comperimeter Tables)

Surveys often contain many questions with a common set of possible responses. You can use stacking to display these related variables in the same table, and you can display the shared response categories in the columns of the table.

To Create a Table for Multiple Variables with Shared Categories

- Drag and drop the categorical variables from the variable list into the Rows area of the canvas. The variables should be **stacked**. For more information, see "Stacking Variables" above.

- From the Category Position drop-down list, select Row labels in columns.

Figure 2-10
Stacked variables with shared response categories in columns

	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

Customizing the Table Builder

Unlike standard dialog boxes, you can change the size of the table builder in the same way that you can change the size of any standard window:

- Click and drag the top, bottom, either side, or any corner of the table builder to decrease or increase its size.

On the Table tab, you can also change the size of the variable list, the Categories list, and the canvas pane.

- Click and drag the horizontal bar between the variable list and the Categories list to make the lists longer or shorter. Moving it down makes the variable list longer and the Categories list shorter. Moving it up does the reverse.

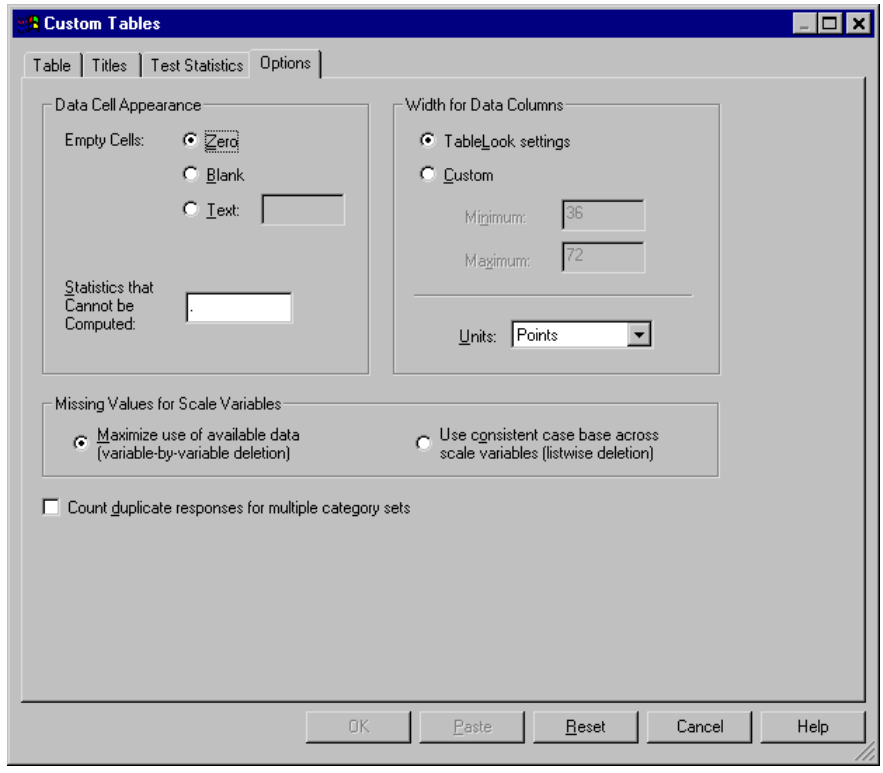
- Click and drag the vertical bar between the variable list and Categories list from the canvas pane to make the lists wider or narrower. The canvas automatically resizes to fit the remaining space.

Custom Tables: Options Tab

The Options tab allows you to:

- Specify what is displayed in empty cells and cells for which statistics cannot be computed.
- Control how missing values are handled in the computation of scale variable statistics.
- Set minimum and/or maximum data column widths.
- Control the treatment of duplicate responses in multiple category sets.

Figure 2-11
Custom Tables: Options tab



Data Cell Appearance. Controls what is displayed in empty cells and cells for which statistics cannot be computed.

- **Empty cells.** For table cells that contain no cases (cell count of 0), you can select one of three display options: zero, blank, or a text value that you specify. The text value can be up to 255 characters long.
- **Statistics that Cannot be Computed.** Text displayed if a statistic cannot be computed (for example, the mean for a category with no cases). The text value can be up to 255 characters long. The default value is a period (.).

Width for Data Columns. Controls minimum and maximum column width for data columns. This setting does not affect columns widths for row labels.

- **TableLook settings.** Uses the data column width specification from the current default TableLook. You can create your own custom default TableLook to use when new tables are created, and you can control both row label column and data column widths with a TableLook.
- **Custom.** Overrides the default TableLook settings for data column width. Specify the minimum and maximum data column widths for the table and the measurement unit: points, inches, or centimeters.

Missing Values for Scale Variables. For tables with two or more scale variables, controls the handling of missing data for scale variable statistics.

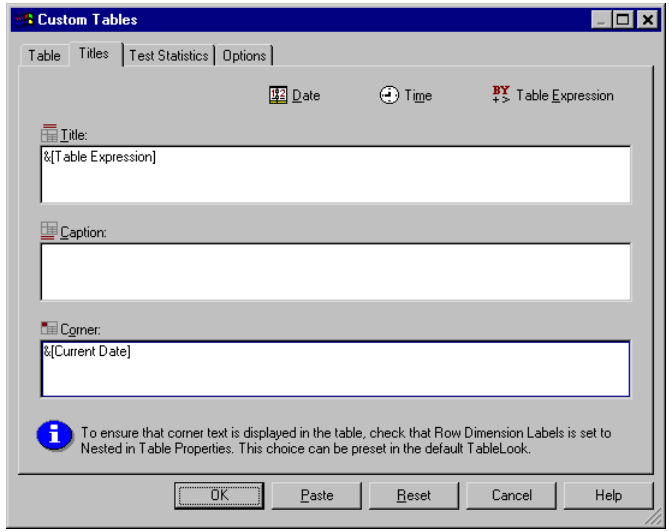
- **Maximize use of available data (variable-by-variable deletion).** All cases with valid values for each scale variable are included in summary statistics for that scale variable.
- **Use consistent case base across scale variables (listwise deletion).** Cases with missing values for any scale variables in the table are excluded from the summary statistics for all scale variables in the table.

Count duplicate responses for multiple category sets. A duplicate response is the same response for two or more variables in the multiple category set. By default, duplicate responses are not counted, but this may be a perfectly valid condition that you do want to include in the count (such as a multiple category set representing the manufacturer of the last three cars purchased by a survey respondent).

Custom Tables: Titles Tab

The Titles tab controls the display of titles, captions, and corner labels.

Figure 2-12
Custom Tables: Titles tab



Title. Text that is displayed above the table.

Caption. Text that is displayed below the table and above any footnotes.

Corner. Text that is displayed in the upper left corner of the table. Corner text is displayed only if the table contains row variables and if the pivot table row dimension label property is set to Nested. This is *not* the default TableLook setting.

You can include the following automatically generated values in the table title, caption, or corner label:

Date. Current year, month, and day displayed in a format based on your current Windows Regional Options settings.

Time. Current hour, minute, and second displayed in a format based on your current Windows Regional Options settings.

Table Expression. Variables used in the table and how they're used in the table. If a variable has a defined variable label, the label is displayed. In the generated table, the following symbols indicate how variables are used in the table:

- **+** indicates stacked variables.
- **>** indicates nesting.
- **BY** indicates crosstabulation or layers.

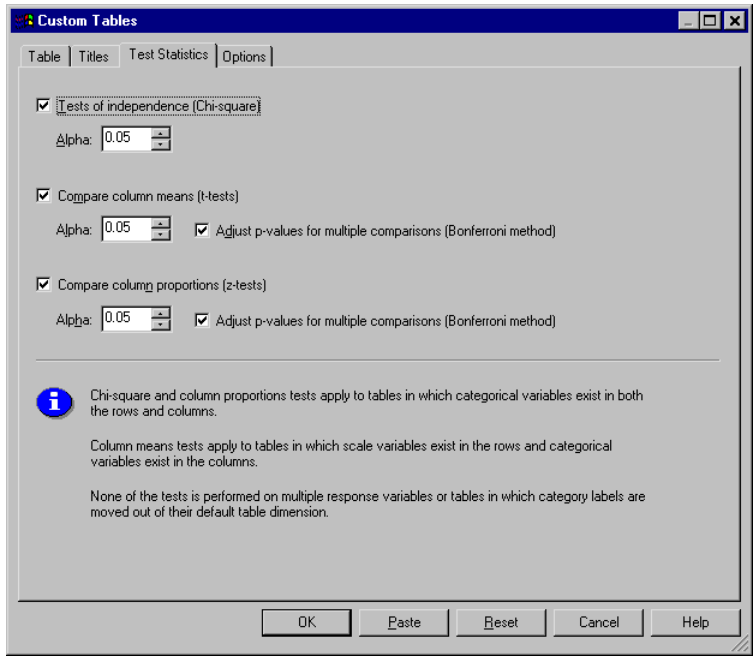
Custom Tables: Test Statistics Tab

The Test Statistics tab allows you to request various significance tests for your custom tables, including:

- Chi-square tests of independence.
- Tests of the equality of column means.
- Tests of the equality of column proportions.

These tests are not available for multiple response variables or tables in which category labels are moved out of their default table dimension.

Figure 2-13
Custom Tables: Test Statistics tab



Tests of independence (Chi-square). This option produces a chi-square test of independence for tables in which at least one category variable exists in both the rows and columns. You can also specify the alpha level of the test, which should be a value greater than 0 and less than 1.

Compare column means (t-tests). This option produces pairwise tests of the equality of column means for tables in which at least one category variable exists in the columns and at least one scale variable exists in the rows. You can select whether the *p* values of the tests are adjusted using the Bonferroni method. You can also specify the alpha level of the test, which should be a value greater than 0 and less than 1.

Compare column proportions (z-tests). This option produces pairwise tests of the equality of column proportions for tables in which at least one category variable exists in both the columns and rows. You can select whether the *p* values of the tests are adjusted using the Bonferroni method. You can also specify the alpha level of the test, which should be a value greater than 0 and less than 1.

Chapter
3

Simple Tables for Categorical Variables

Most tables you want to create will probably include at least one **categorical variable**. A categorical variable is one with a limited number of distinct values or categories (for example, gender or religion).

An icon next to each variable in the variable list identifies the variable type.


Scale


Categorical


Multiple response set, multiple categories


Multiple response set, multiple dichotomies

Custom Tables is optimized for use with categorical variables that have defined **value labels**. For more information, see "Building Tables" in the chapter *Table Builder Interface*.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav*. This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

A Single Categorical Variable

Although a table of a single categorical variable may be one of the simplest tables you can create, it may often be all you want or need.

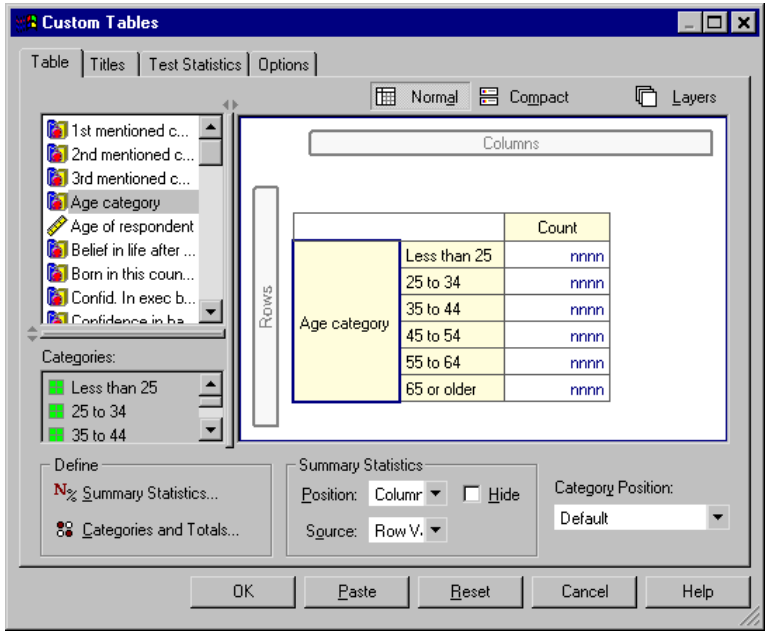
- From the menus, choose:

  Analyze
    Tables
      Custom Tables...

- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.

A preview of the table is displayed on the canvas pane. The preview doesn't display actual data values; it displays only placeholders where data will be displayed.

Figure 3-1
A single categorical variable in rows



▶ Click OK to create the table.

The table is displayed in the Viewer window.

Figure 3-2
A single categorical variable in rows

		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

In this simple table, the column heading *Count* isn't really necessary, and you can create the table without this column heading.

▶ Open the table builder again (Analyze menu, Tables, Custom Tables).

▶ Select (click) Hide for Position in the Summary Statistics group.

▶ Click OK to create the table.

Figure 3-3
Single categorical variable without summary statistics column label

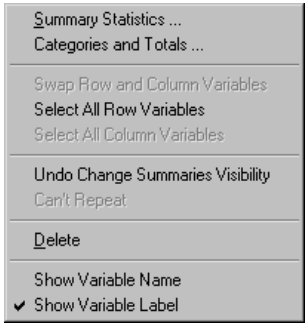
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Percentages

In addition to counts, you can also display percentages. For a simple table of a single categorical variable, if the variable is displayed in rows, you probably want to look at column percentages. Conversely, for a variable displayed in columns, you probably want to look at row percentages.

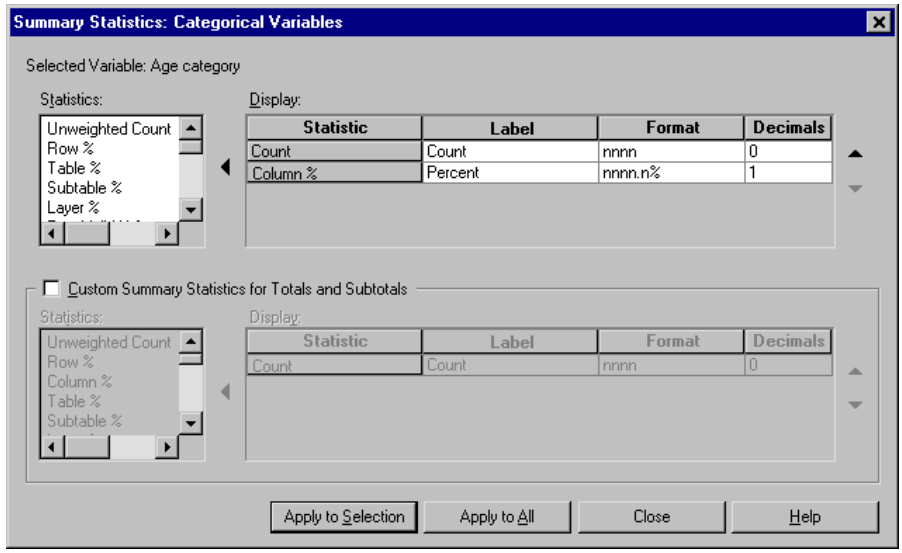
- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Deselect (uncheck) Hide for Position in the Summary Statistics group. Since this table will have two columns, you want to display the column labels so you know what each column represents.
- Right-click on *Age category* on the canvas pane and select Summary Statistics from the pop-up context menu.

Figure 3-4
Right-click context menu on canvas pane



- In the Summary Statistics dialog box, select Column % in the Statistics list and click the arrow to add it to the Display list.
- In the Label cell in the Display list, delete the default label and type Percent.

Figure 3-5
Summary Statistics Categorical Variables dialog box



- Click Apply to Selection and then click OK in the table builder to create the table.

Figure 3-6
Counts and column percentages

		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%

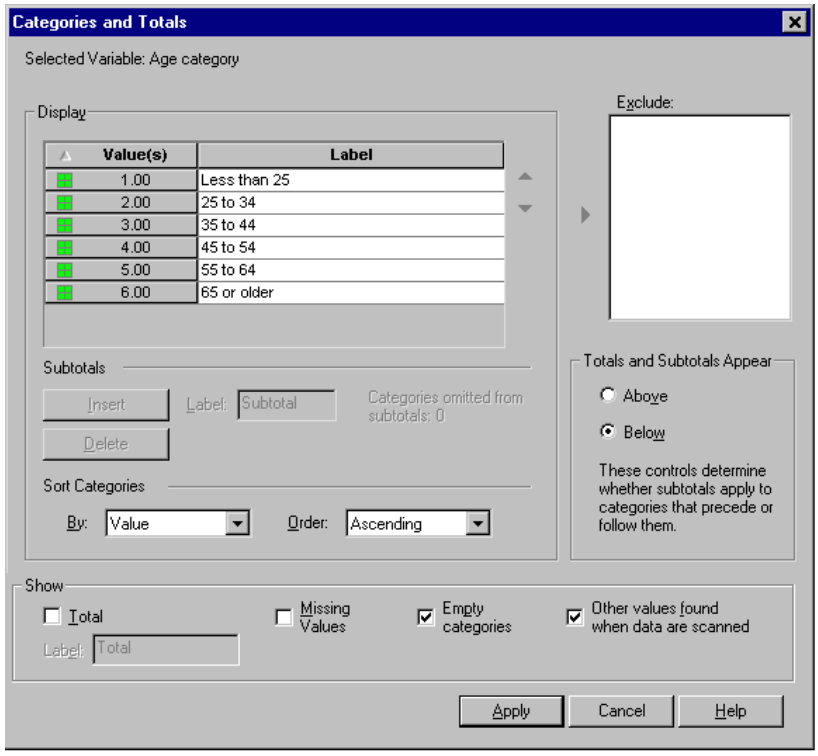
Totals

Totals are not automatically included in custom tables, but it's easy to add totals to a table.

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.

- Select (click) Total in the Categories and Totals dialog box.

Figure 3-7
Categories and Totals dialog box



- Click Apply and then click OK in the table builder to create the table.

Figure 3-8
Counts, column percentages, and totals

		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%

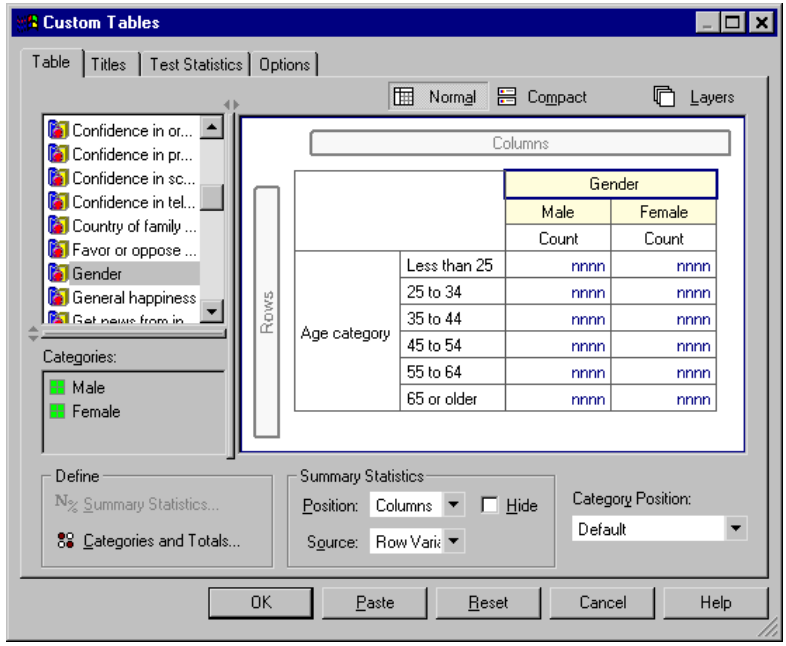
For more information, see the chapter *Totals and Subtotals for Categorical Variables*.

Crosstabulation

Crosstabulation is a basic technique for examining the relationship between two categorical variables. For example, using *Age category* as a row variable and *Gender* as a column variable, you can create a two-dimensional crosstabulation that shows the number of males and females in each age category.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to delete any previous selections in the table builder.
- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.
- Drag and drop *Gender* from the variable list to the Columns area on the canvas pane. (You may have to scroll down through the variable list to find this variable.)

Figure 3-9
Crosstabulation in table builder canvas preview



- Click OK to create the table.

Figure 3-10
Crosstabulation of Age category and Gender

		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

Percentages in Crosstabulations

In a two-dimensional crosstabulation, both row and column percentages may provide useful information.

- Open the table builder again (Analyze menu, Tables, Custom Tables).

- Right-click on *Gender* on the canvas pane.

You may notice that Summary Statistics is disabled in the pop-up context menu. This is because you can only select summary statistics for the innermost variable in the statistics source dimension. The default statistics source dimension (row or column) for categorical variables is based on the order in which you drag and drop variables onto the canvas pane. In this example, we dragged *Age category* to the rows dimension first—and since there aren't any other variables in the rows dimension, *Age category* is the statistics source variable. You can change the statistics source dimension, but in this example, you don't need to do that. For more information, see "Summary Statistics" in the chapter *Table Builder Interface*.

- Right-click on *Age category* on the canvas pane and select Summary Statistics from the pop-up context menu.

- In the Summary Statistics dialog box, select Column % in the Statistics list and click the arrow to add it to the Display list.

- Select Row % in the Statistics list and click the arrow to add it to the Display list.
- Click Apply to Selection and then click OK in the table builder to create the table.

Figure 3-11
Crosstabulation with row and column percentages

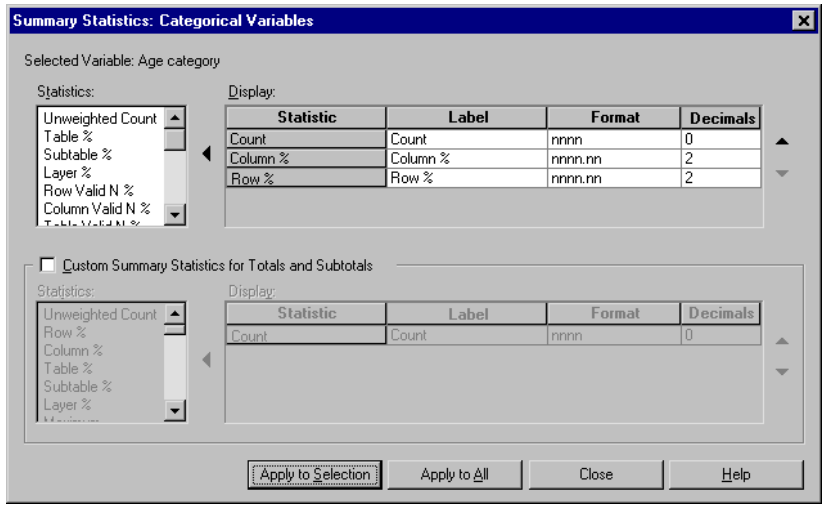
		Gender					
		Male			Female		
		Count	Column %	Row %	Count	Column %	Row %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65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

Controlling Display Format

You can control the display format, including the number of decimals displayed in summary statistics. For example, by default percentages are displayed with one decimal and a percent sign. But what if you want the cell values to show two decimals and no percent sign?

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click on *Age category* on the canvas pane and select Summary Statistics from the pop-up context menu.
- For the two selected percentage summary statistics (Column % and Row %), select nnnn.n from the Format drop-down list and type 2 in the Decimals cell for both of them.

Figure 3-12
Summary Statistics dialog box



▶ Click OK to create the table.

Figure 3-13
Formatted cell display for row and column percentages

		Gender					
		Male			Female		
		Count	Column %	Row %	Count	Column %	Row %
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

Marginal Totals

It's fairly common in crosstabulations to display **marginal totals**—totals for each row and column. Since these aren't included in Custom Tables by default, you need to explicitly add them to your tables.

▶ Open the table builder (Analyze menu, Tables, Custom Tables).

- Click Reset to delete any previous selections in the table builder.
- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.
- Drag and drop *Gender* from the variable list to the Columns area on the canvas pane. (You may have to scroll down through the variable list to find this variable.)
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select (click) Total in the Categories and Totals dialog box and then click Apply.
- Right-click on *Gender* on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select (click) Total in the Categories and Totals dialog box and then click Apply.
- Select (click) Hide for Position in the Summary Statistics group. (Since you're displaying only counts, you don't need to identify the "statistic" displayed in the data cells of the table.)
- Click OK to create the table.

Figure 3-14
Crosstabulation with marginal totals

		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

Sorting and Excluding Categories

By default, categories are displayed in the ascending order of the data values that the category value labels represent. For example, although value labels of *Less than 25, 25 to 34, 35 to 44, ...*, etc., are displayed for age categories, the actual underlying data values are 1, 2, 3, ..., etc., and it is those underlying data values that control the default display order of the categories.

You can easily change the order of the categories and also exclude categories that you don't want displayed in the table.

Sorting Categories

You can sort categories in several ways:

- Ascending or descending order of data values
- Ascending or descending order of value labels
- Ascending or descending order of cell values
- Manual rearrangement or individual categories

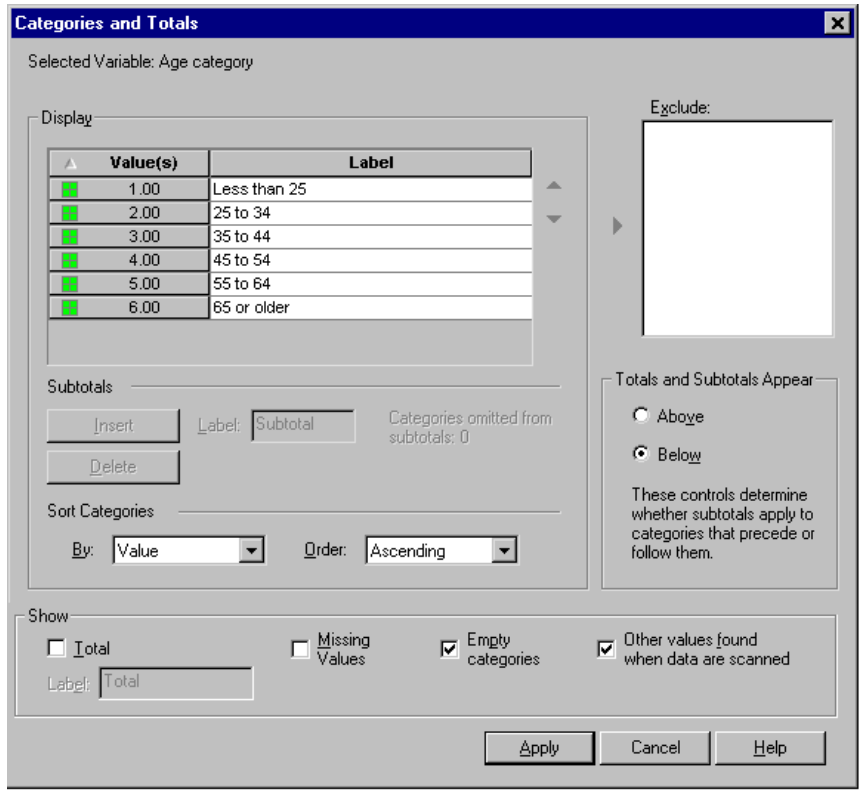
▶ Open the table builder (Analyze menu, Tables, Custom Tables).

▶ If *Age category* isn't already displayed in the Rows area on the canvas pane, drag and drop it there.

▶ Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.

Both data values and the associated value labels are displayed in the current sort order, which in this case is still ascending order of data values.

Figure 3-15
Default category order, ascending by data values



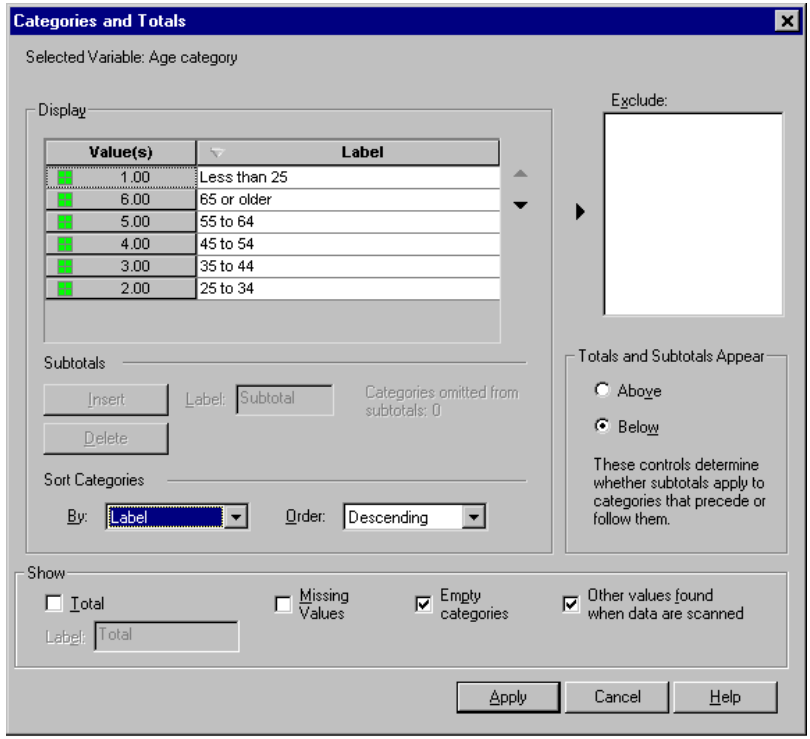
- In the Sort Categories group, select Descending from the Order drop-down list.

  The sort order is now reversed.

- Select Labels from the By drop-down list.

  The categories are now sorted in descending alphabetical order of the value labels.

Figure 3-16
Descending alphabetical sort order

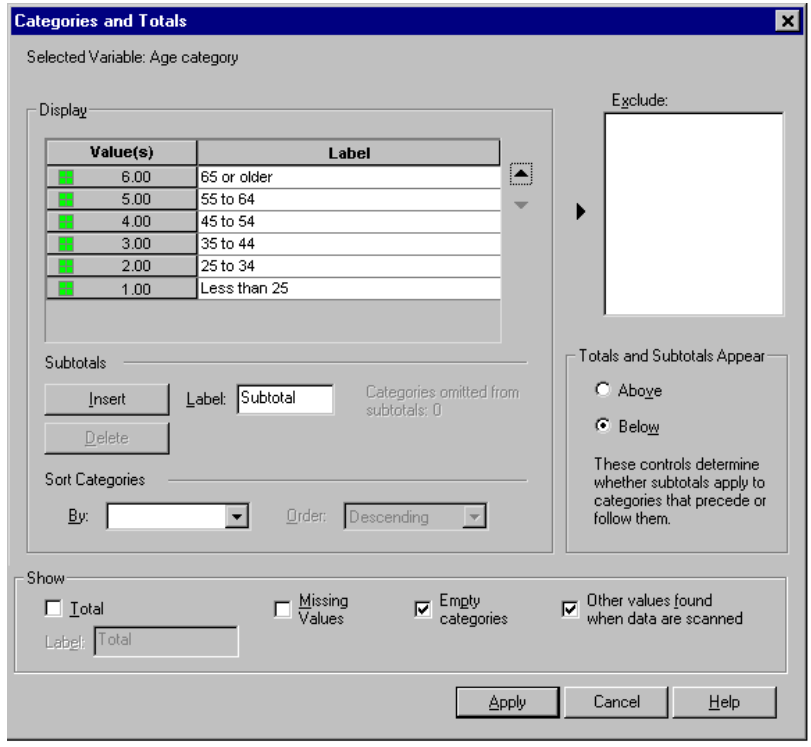


Notice that the category labeled *Less than 25* is at the top of the list. In alphabetical sorting, letters come after numbers. Since this is the only label that starts with a letter and since the list is sorted in descending (reverse) order, this category sorts to the top of the list.

If you want a particular category to appear at a different location in the list, you can easily move it.

- Click the category labeled *Less than 25* in the Label list.
- Click the down arrow to the right of the list. The category moves down one row in the list.
- Keep clicking the down arrow until the category is at the bottom of the list.

Figure 3-17
Manually arranged categories



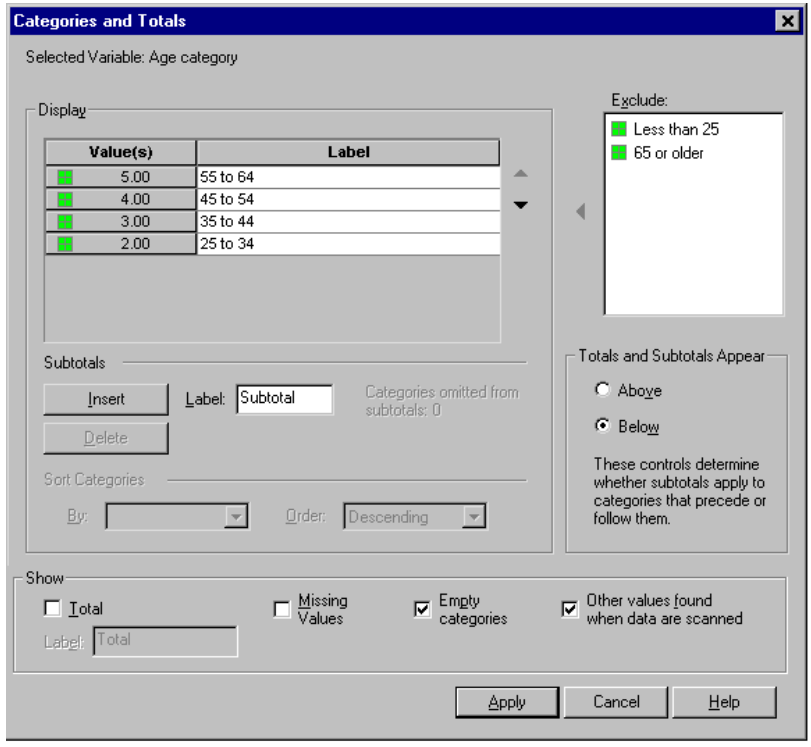
Excluding Categories

If there are some categories that you don't want to appear in the table, you can exclude them.

- Click the category labeled *Less than 25* in the Label list.
- Click the arrow key to the left of the Exclude list.
- Click the category labeled *65 or older* in the Label list.
- Click the arrow key to the left of the Exclude list again.

The two categories are moved from the Display list to the Exclude list. If you change your mind, you can easily move them back to the Display list.

Figure 3-18
Manually excluded categories in Categories and Totals dialog box



▶ Click Apply and then click OK in the table builder to create the table.

Figure 3-19
Table sorted by descending value label, some categories excluded

		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

Notice that the totals are lower than they were before the two categories were excluded. This is because totals are based on the categories included in the table. Any excluded categories are excluded from the total calculation. For more information, see the chapter *Totals and Subtotals for Categorical Variables*.

Chapter
4

Stacking, Nesting, and Layers with Categorical Variables

Stacking, nesting, and layers are all methods for displaying multiple variables in the same table. This chapter focuses on using these techniques with categorical variables, although they can also be used with scale variables.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav*. This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

Stacking Categorical Variables

Stacking can be thought of as taking separate tables and pasting them together into the same display. For example, you could display information on gender and age category in separate sections of the same table.

- From the menus, choose:

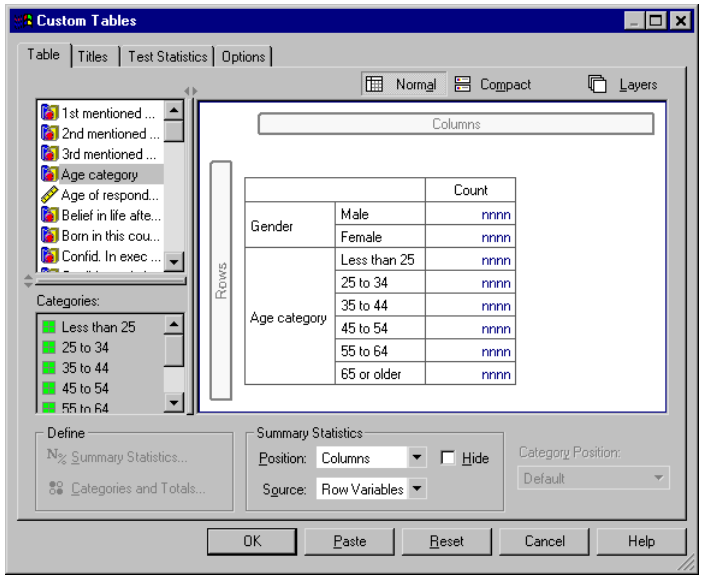
  Analyze
  Tables
  Custom Tables...

- In the table builder, drag and drop *Gender* from the variable list to the Rows area on the canvas pane.

▶ Drag and drop *Age category* from the variable list to the Rows area **below** *Gender*.

The two variable are now stacked in the row dimension.

Figure 4-1
Stacked categorical variables displayed on the canvas pane



▶ Click OK to create the table.

Figure 4-2
Table of categorical variables stacked in rows

		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

You can also stack variables in columns in a similar fashion.

Stacking with Crosstabulation

A stacked table can include other variables in other dimensions. For example, you could crosstabulate two variables stacked in the rows with a third variable displayed in the column dimension.

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- If *Age category* and *Gender* aren't already stacked in the rows, follow the directions above for stacking them.
- Drag and drop *Get news from internet* from the variable list to the Columns area on the canvas pane.
- Click OK to create the table.

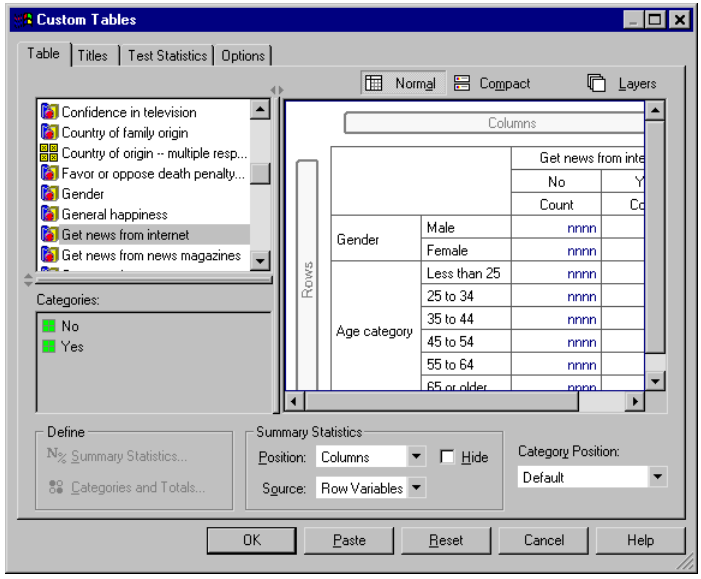
Figure 4-3
Two stacked row variables crosstabulated with a column variable

		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

Note: There a several variables with labels that start with "Get news from ...," so it may be difficult to distinguish between them in the variable list (since the labels may be too wide to be displayed completely in the variable list). There are two ways to see the entire variable label:

- Position the mouse pointer on a variable in the list to display the entire label in a pop-up ToolTip.
- Click and drag the vertical bar that separates the variable and Categories lists from the canvas pane to make the lists wider.

Figure 4-4
Variable list widened to display complete variable labels



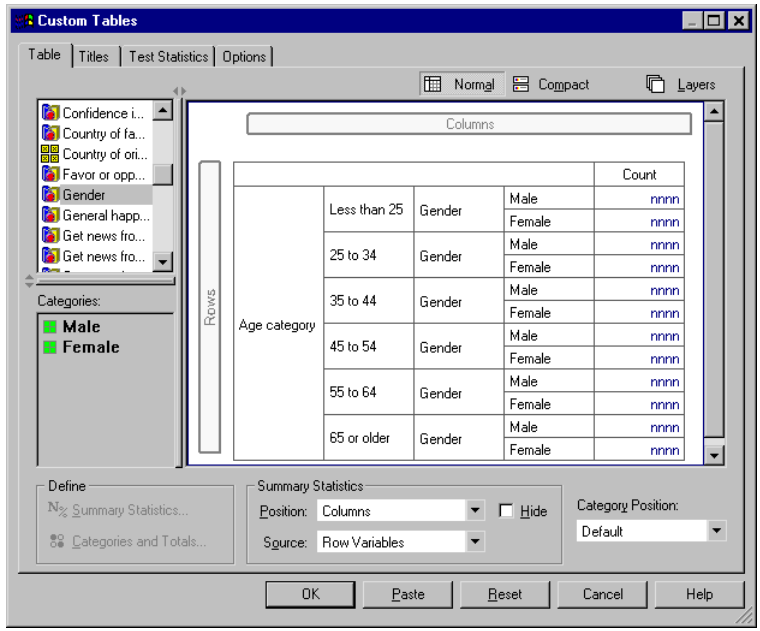
Nesting Categorical Variables

Nesting, like crosstabulation, can show the relationship between two categorical variables, except that one variable is nested within the other in the same dimension. For example, you could nest *Gender* within *Age category* in the row dimension, showing the number of males and females in each age category.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to delete any previous selections in the table builder.
- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.
- Drag and drop *Gender* from the variable list to the right of *Age category* in the Rows area.

The preview on the canvas pane now shows that the nested table will contain a single column of counts, with each cell containing the number of males or females in each age category.

Figure 4-5
Gender nested within Age category

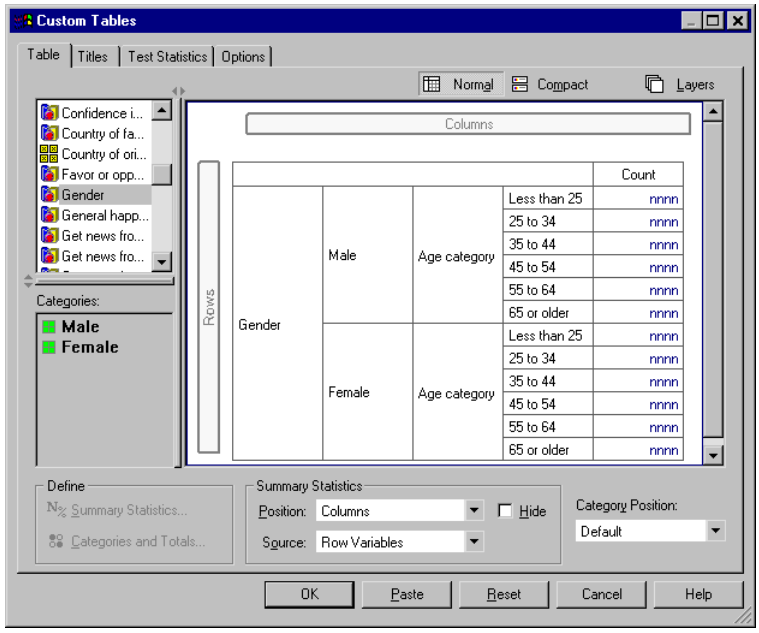


You may notice that the variable label *Gender* is displayed repeatedly, once for each age category. You can minimize this kind of repetition by placing the variable with the fewest categories at the outermost level of the nesting.

- Click the variable label *Gender* on the canvas pane.
- Drag and drop the variable as far to the left in the Rows area as you can.

Now instead of *Gender* being repeated six times, *Age category* is repeated twice. This is a less cluttered table that will produce essentially the same results.

Figure 4-6
Age category nested within Gender in table builder preview



▶ Click OK to create the table.

Figure 4-7
Table of Age category nested within Gender

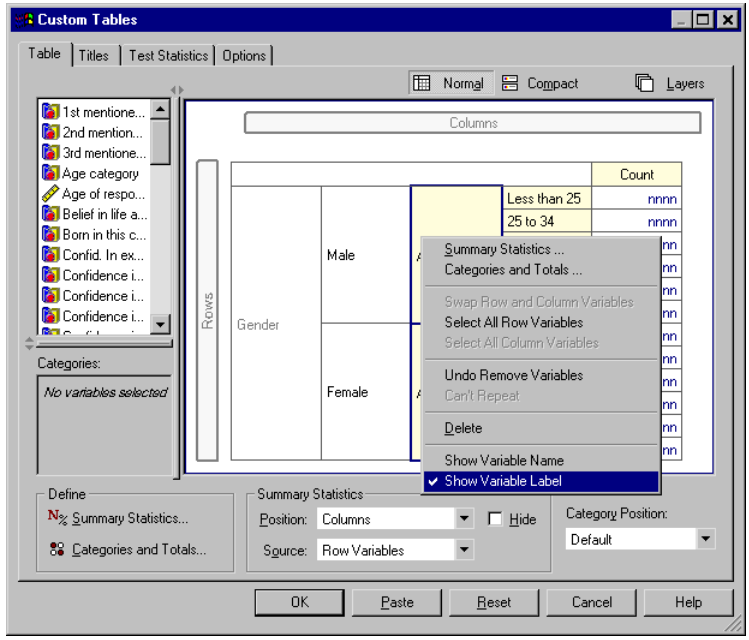
				Count
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age category	Less than 25	134
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

Suppressing Variable Labels

Another solution to redundant variable labels in nested tables is simply to suppress the display of variable names or labels. Since the value labels for both *Gender* and *Age category* are probably sufficiently descriptive without the variable labels, we can eliminate the labels for both variables.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click *Age category* on the canvas pane and deselect (uncheck) Show Variable Label on the pop-up context menu.
- Do the same for *Gender*.

Figure 4-8
Suppressing variable labels via the context menu in the table builder



The variable labels are still displayed in the table preview, but they won't be included in the table.

- Click OK to create the table.

Figure 4-9
Nested table without variable labels

		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

If you want the variable labels included with the table somewhere—without displaying them multiple times in the body of the table—you can include them in the table title or corner label.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click the Titles tab.
- Click anywhere in the Title text box.
- Click Table Expression. The text *&[Table Expression]* is displayed in the Title text box. This will generate a table title that includes the variable labels for the variables used in the table.
- Click OK to create the table.

Figure 4-10
Variable labels in table title

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

The greater than sign (>) in the title indicates that *Age category* is nested within *Gender.*

Nested Crosstabulation

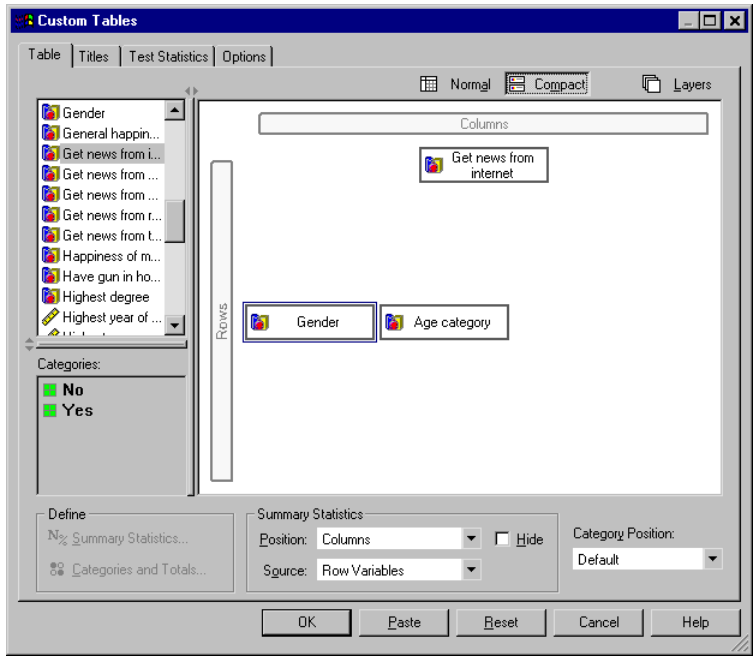
A nested table can contain other variables in other dimensions. For example, you could nest *Age category* within *Gender* in the rows and crosstabulate the nested rows with a third variable in the column dimension.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- If *Age category* isn't already nested within *Gender* in the rows, follow the directions above for nesting them.
- Drag and drop *Get news from internet* from the variable list to the Columns area on the canvas pane.

You may notice that the table is too large to display completely on the canvas pane. You can scroll up/down or right/left on the canvas pane to see more of the table preview, or:

- Click Compact in the table builder to see a compact view. This displays only the variable labels, without any information on categories or summary statistics included in the table.
- Increase the size of the table builder by clicking and dragging any of the sides or corners of the table builder.

Figure 4-11
Compact view on the canvas pane



► Click OK to create the table.

Figure 4-12
Nested crosstabulation

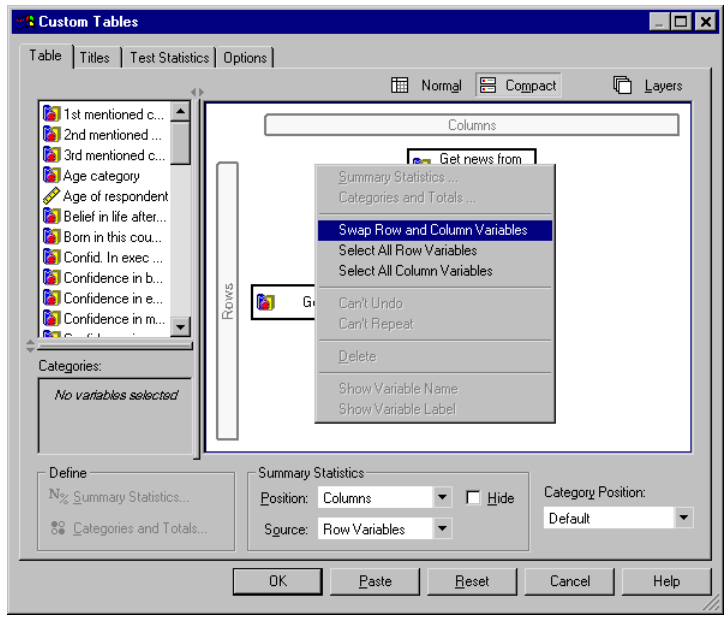
				Get news from internet	
				No	Yes
				Count	Count
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
			35 to 44	217	92
			45 to 54	169	52
			55 to 64	112	24
			65 or older	155	23
	Female	Age category	Less than 25	87	47
			25 to 34	209	142
			35 to 44	218	152
			45 to 54	177	83
			55 to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

Swapping Rows and Columns

What do you do if you spend a lot of time setting up a complex table and then decide it's absolutely perfect—except that you want to switch the orientation, putting all of the row variables in the columns and vice versa? For example, you've created a nested crosstabulation with *Age category* and *Gender* nested in the rows, but now you want these two demographic variables nested in the columns instead.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click anywhere on the canvas pane and select Swap Row and Column Variables from the pop-up context menu.

Figure 4-13
Swapping row and column variables



The row and column variables have now been switched.

Before creating the table, let's make a few modifications to make the display less cluttered.

- Select Hide to suppress the display of the summary statistics column label.

- Right-click *Gender* on the canvas pane and deselect (uncheck) Show Variable Label.
- Now click OK to create the table.

Figure 4-14
Crosstabulation with demographic variables nested in columns

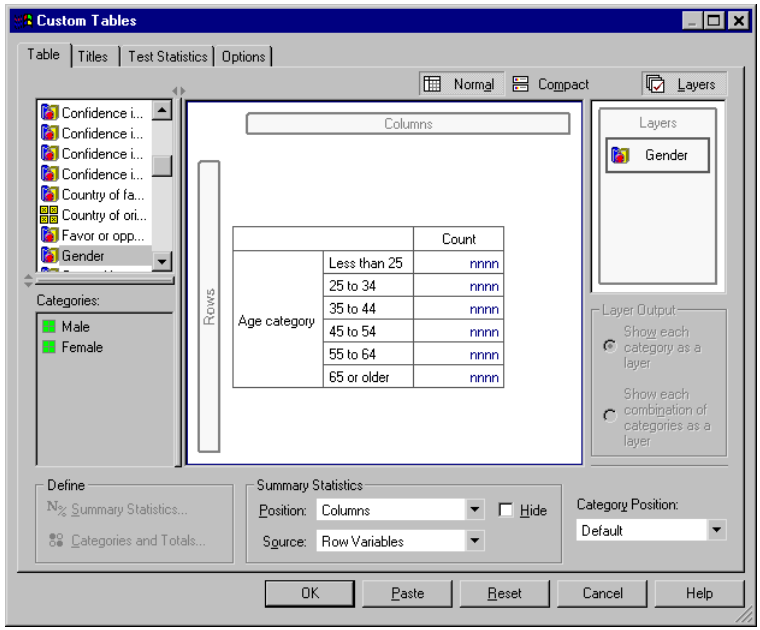
		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

Layers

You can use layers to add a dimension of depth to your tables, creating three-dimensional "cubes." Layers are, in fact, quite similar to nesting or stacking; the primary difference is that only one layer category is visible at a time. For example, using *Age category* as the row variable and *Gender* as a layer variable produces a table in which information for males and females is displayed in different layers of the table.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to delete any previous selections in the table builder.
- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.
- Click Layers at the top of the table builder to display the Layers list.
- Drag and drop *Gender* from the variable list to the Layers list.

Figure 4-15
Age category in rows, Gender in layers



At this point, you might notice that adding a layer variable has no visible effect on the preview displayed on the canvas pane. Layer variables do not affect the preview on the canvas pane unless the layer variable is the statistics source variable and you change the summary statistics.

▶ Click OK to create the table.

Figure 4-16
Simple layered table

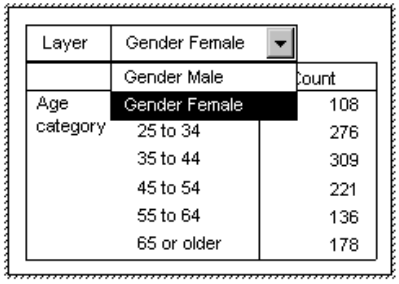
Gender Male

		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

At first glance, this table doesn't look any different than a simple table of a single categorical variable. The only difference is the presence of the label *Gender Male* at the top of the table.

- Double-click the table in the Viewer window to activate it.
- You can now see that the label *Gender Male* is actually a choice in a drop-down list.
- Click the down arrow on the drop-down list to display the whole list of layers.

Figure 4-17
List of layers in activated pivot table



In this table, there is only one other choice in the list.

- Select *Gender Female* from drop-down list.

Figure 4-18
Simple layered table with different layer displayed

Gender Female

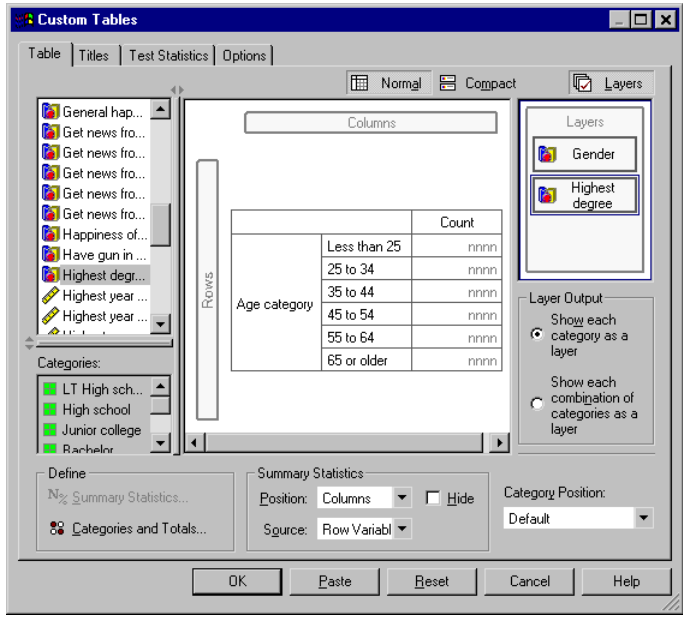
		Count
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Two Stacked Categorical Layer Variables

If you have more than one categorical variable in the layers, you can either stack or nest the layer variables. By default, layer variables are stacked. (*Note*: If you have any scale layer variables, layer variables can only be stacked.)

- Open the table builder (Analyze menu, Tables, Custom Tables).
- If you don't already have *Age category* in the rows and *Gender* in the layers, follow the directions above for creating a layered table.
- Drag and drop *Highest degree* from the variable list to the Layer list below *Gender*.

Figure 4-19
Stacked layer variables in table builder

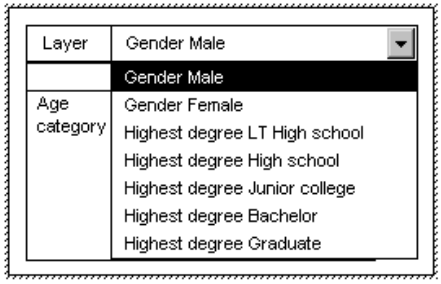


The two radio buttons below the Layer list in the Layer Output group are now activated. The default selection is Show each category as a layer. This is equivalent to stacking.

- Click OK to create the table.

- Double-click the table in the Viewer window to activate it.
- Click the down arrow on the drop-down list to display the whole list of layers.

Figure 4-20
List of stacked layers in activated pivot table



There are seven layers in the table: two layers for the two *Gender* categories and five layers for the five *Highest degree* categories. For stacked layers, the total number of layers is the sum of the number of categories for the layer variables (including any total or subtotal categories you have requested for the layer variables).

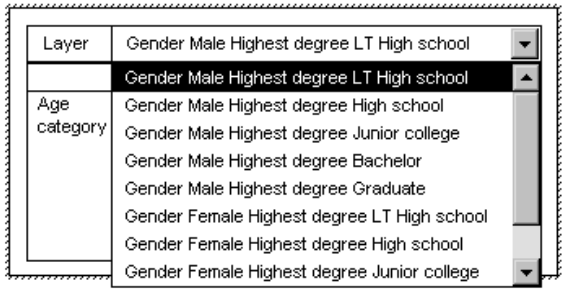
Two Nested Categorical Layer Variables

Nesting categorical layer variables creates a separate layer for each combination of layer variable categories.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- If you haven't done so already, follow the directions above for creating a table of stacked layers.
- In the Layer Output group, select Show each combination of categories as a layer. This is equivalent to nesting.
- Click OK to create the table.
- Double-click the table in the Viewer window to activate it.

- Click the down arrow on the drop-down list to display the whole list of layers.

Figure 4-21
List of nested layers in activated pivot table



There are 10 layers in the table (you have to scroll through the list to see all of them), one for each combination of *Gender* and *Highest degree*. For nested layers, the total number of layers is the *product* of the number of categories for each layer variable (in this example, $5 \times 2 = 10$).

Printing Layered Tables

By default, only the currently visible layer is printed. To print all layers of a table:

- Double-click the table in the Viewer window to activate it.
- From the Viewer window menus, choose

  Format
  Table Properties...

- Click the Printing tab.
- Select Print all layers.

You can also save this setting as part of a TableLook, including the default TableLook.

Chapter
5

Totals and Subtotals for Categorical Variables

You can include both totals and subtotals in custom tables. Totals and subtotals can be applied to categorical variables at any nesting level in any dimension—row, column, and layer.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav*. This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

Simple Total for a Single Variable

- From the menus, choose:

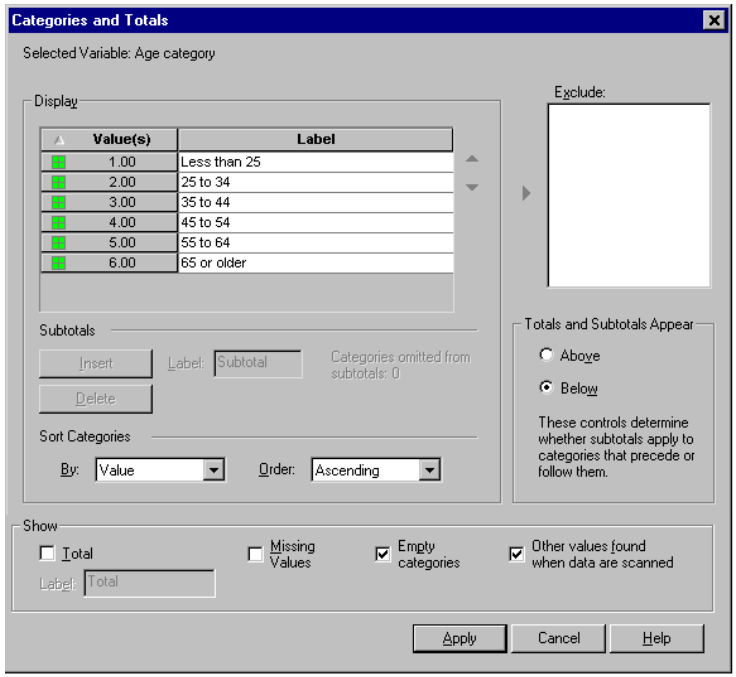
  Analyze
  Tables
  Custom Tables...

- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.

- Right-click on *Age category* on the canvas pane and select Summary Statistics from the pop-up context menu.

- In the Summary Statistics dialog box, select Column % in the Statistics list and click the arrow to add it to the Display list.
- In the Label cell in the Display list, delete the default label and type Percent.
- Click Continue.
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select (click) Total in the Categories and Totals dialog box.

Figure 5-1
Categories and Totals dialog box



- Click Apply and then click OK in the table builder to create the table.

Figure 5-2
Simple total for a single categorical variable

		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%

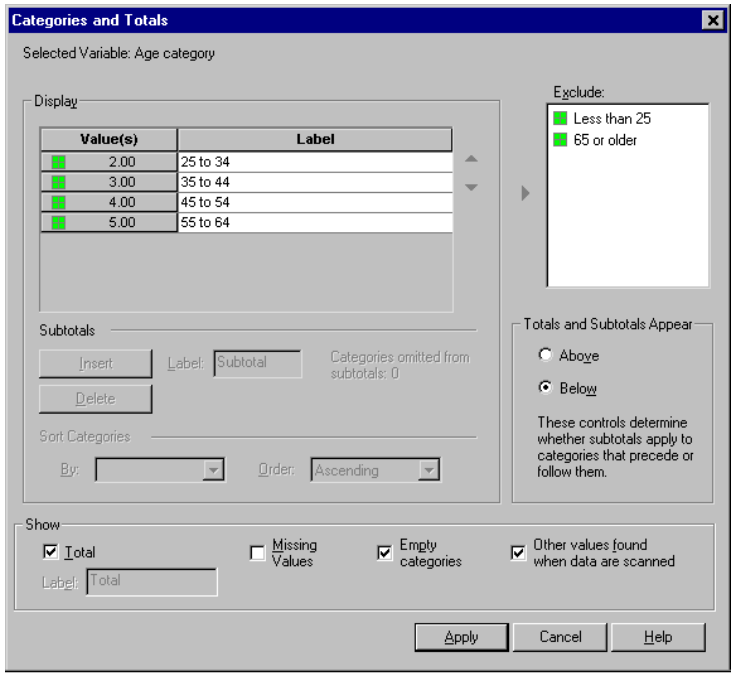
What You See Is What Gets Totaled

Totals are based on categories displayed in the table. If you choose to exclude some categories from a table, cases from those categories are not included in total calculations.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.
- Click the category labeled *Less than 25* in the Label list.
- Click the arrow key to the left of the Exclude list.
- Click the category labeled *65 or older* in the Label list.
- Click the arrow key to the left of the Exclude list again.

The two categories are moved from the Display list to the Exclude list.

Figure 5-3
Manually excluded categories in Categories and Totals dialog box



▶ Click Apply and then click OK in the table builder to create the table.

Figure 5-4
Total in table with excluded categories

		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%

The total count in this table is only 2,107, compared to 2,828 when all of the categories are included. This is because only the categories that are used in the table are included in the total. (The percentage total is still 100% because all of the percentages are based on the total number of cases used in the table, not the total number of cases in the data file.)

Display Position of Totals

By default, totals are displayed below the categories being totaled. You can change the display position of totals to show them above the categories being totaled.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.
- In the Totals and Subgroups Appear group, select Above.
- Click Apply and then click OK in the table builder to create the table.

Figure 5-5
Total displayed above totaled categories

		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%

Totals for Nested Tables

Since totals can be applied to categorical variables at any level of the nesting, you can create tables that contain group totals at multiple nesting levels.

Group Totals

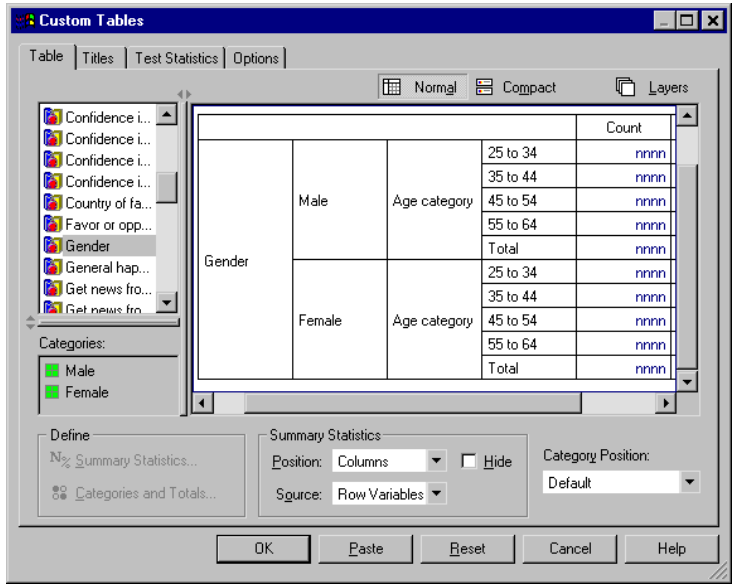
Totals for categorical variables nested within other categorical variables represent group totals.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Drag and drop *Gender* to the left of *Age category* on the canvas pane.
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.

Before creating the table, let's move the totals back below the totaled categories.

- In the Totals and Subgroups Appear group, select Below.
- Click Apply to save the setting and return to the table builder.

Figure 5-6
Age category nested within Gender in the table builder



- Click OK to create the table.

Figure 5-7
Age category totals within Gender categories

				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%

The table now displays two group totals: one for males and one for females.

Grand Totals

Totals applied to nested variables are always group totals, not grand totals. If you want totals for the entire table, you can apply totals to the variable at the outermost nesting level.

- Open the table builder again.
- Right-click on *Gender* on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select (click) Total in the Categories and Totals dialog box.
- Click Apply and then click OK in the table builder to create the table.

Figure 5-8
Grand totals for a nested table

				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%

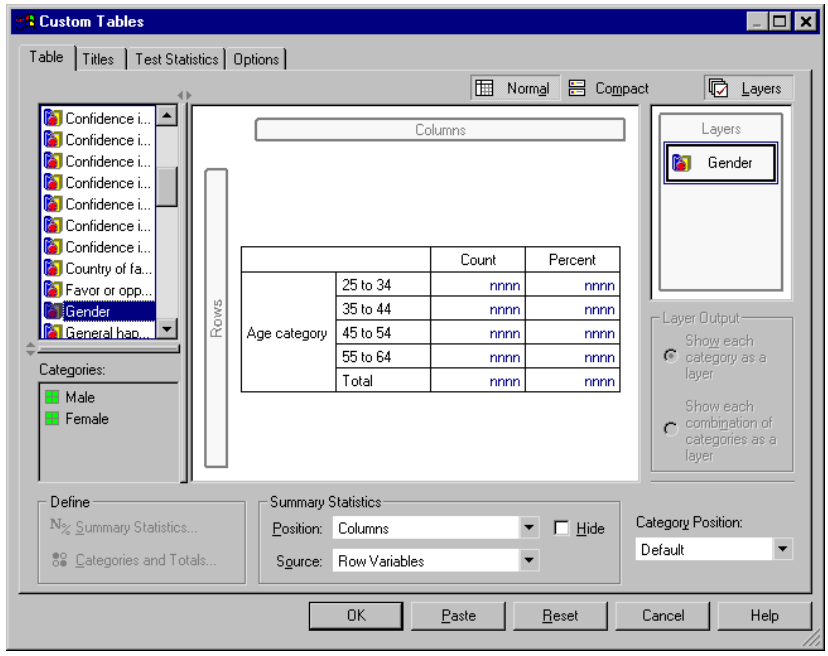
Notice that the grand total is only 2,107, not 2,828. This is because two age categories are still excluded from the table, so the cases in those categories are excluded from all totals.

Layer Variable Totals

Totals for layer variables are displayed as separate layers in the table.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Layers in the table builder to display the Layers list.
- Drag and drop *Gender* from the row area on the canvas pane to the Layers list.

Figure 5-9
Layer variable in table builder



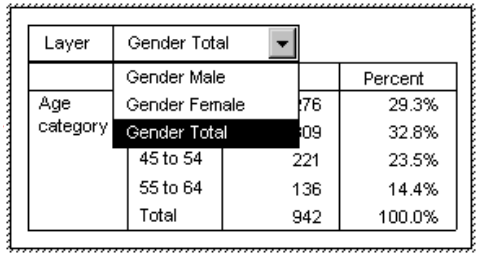
Note: Since you already specified totals for *Gender*, you don't need to do so now. Moving the variable between dimensions does not affect any of the settings for that variable.

- Click OK to create the table.

- Double-click the table in the Viewer to activate it.
- Click the down arrow in the Layer drop-down list to display a list of all the layers in the table.

  There are three layers in the table: *Gender Male*, *Gender Female*, and *Gender Total*.

Figure 5-10
Total layer in Layer list in activated pivot table



Display Position of Layer Totals

For layer variable totals, the display position (above or below) for totals determines the layer position for the totals. For example, if you specify Above for a layer variable total, the total layer is the first layer displayed.

Subtotals

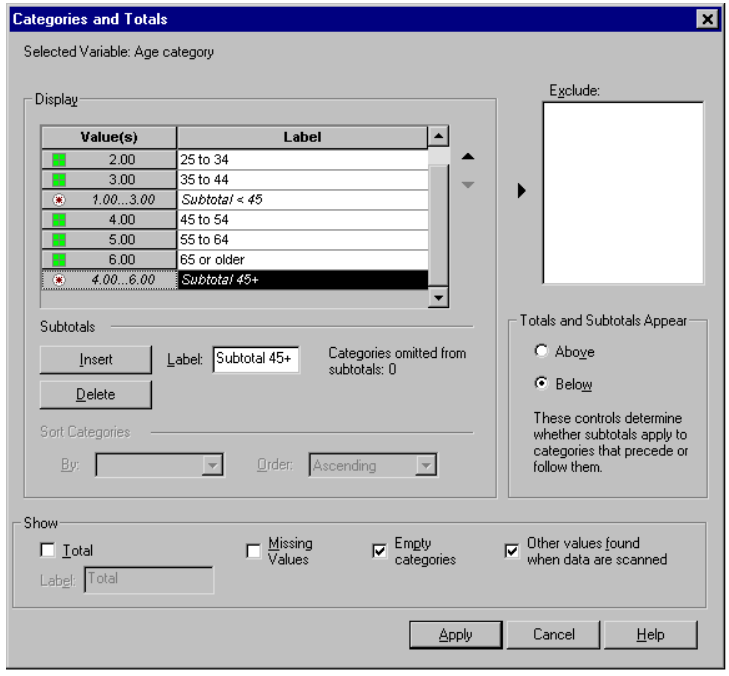
You can include subtotals for subsets of categories of a variable. For example, you could include subtotals for age categories that represent all of the respondents in the sample survey under and over age 45.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to clear any previous settings in the table builder.
- In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.

- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select 3.00 in the Value(s) list.
- In the Label text field next to the Insert button, type Subtotal < 45.
- Then click Insert.

  This inserts a row containing the subtotal for the first three age categories.
- Select 6.00 in the Value(s) list.
- In the Label text field next to the Insert button, type Subtotal 45+.
- Then click Insert.

Figure 5-11
Defining subtotals in the Categories and Totals dialog box



Important note: You should select the display position for totals and subtotals (Above or Below) before defining any subtotals. Changing the display position affects all subtotals (not just the currently selected subtotal), and it also *changes the categories included in the subtotals.*

- Click Apply and then click OK in the table builder to create the table.

Figure 5-12
Subtotals for Age category

		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

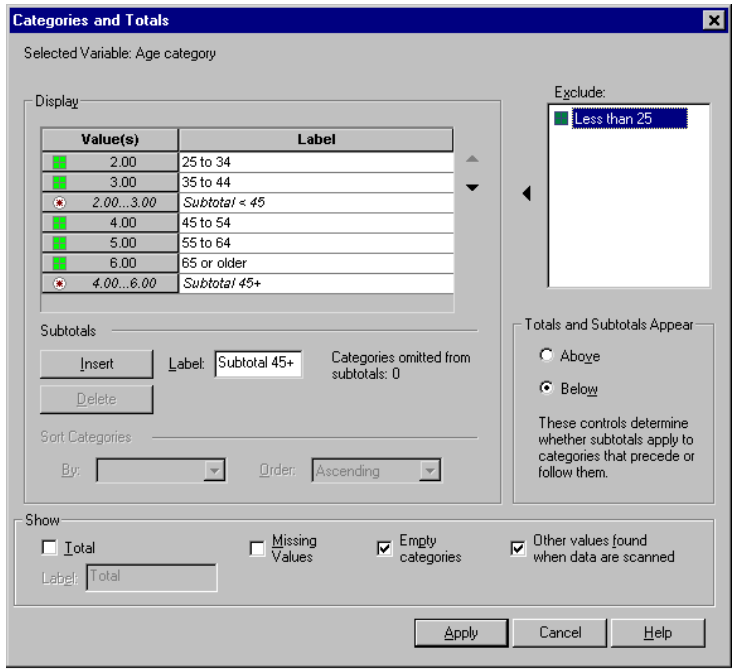
What You See Is What Gets Subtotaled

Just like totals, subtotals are based on the categories included in the table.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click on *Age category* on the canvas pane and select Categories and Totals from the pop-up context menu.

  Note that the value (not the label) displayed for the first subtotal is 1.00...3.00, indicating that the subtotal includes all of the values in the list between 1 and 3.
- Select 1.00 in the Value(s) list (or click on the label *Less than 25*).
- Click the arrow key to the left of the Exclude list.

Figure 5-13
Subtotals when categories are excluded



The first age category is now excluded, and the value displayed for the first subtotal changes to 2.00...3.00, indicating the fact that the excluded category will not be included in the subtotal because subtotals are based on the categories included in the table. Excluding a category automatically excludes it from any subtotals, so you cannot, for example, display only subtotals without the categories on which the subtotals are based.

Layer Variable Subtotals

Just like totals, subtotals for layer variables are displayed as separate layers in the table. Essentially, the subtotals are treated as categories. Each category is a separate layer in the table, and the display order of the layer categories is determined by the category order specified in the Categories and Totals dialog box, including the display position of the subtotal categories.

Chapter
6

Tables for Variables with Shared Categories

Surveys often contain many questions with a common set of possible responses. For example, our sample survey contains a number of variables concerning confidence in various public and private institutions and services, all with the same set of response categories: 1 = *A great deal*, 2 = *Only some*, and 3 = *Hardly any.* You can use stacking to display these related variables in the same table—and you can display the shared response categories in the columns of the table.

Figure 6-1
Table of variables with shared categories

	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

Note: In the previous version of Custom Tables, this was known as a "table of frequencies."

Sample Data File

The examples in this chapter use the data file *survey_sample.sav.* This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

Table of Counts

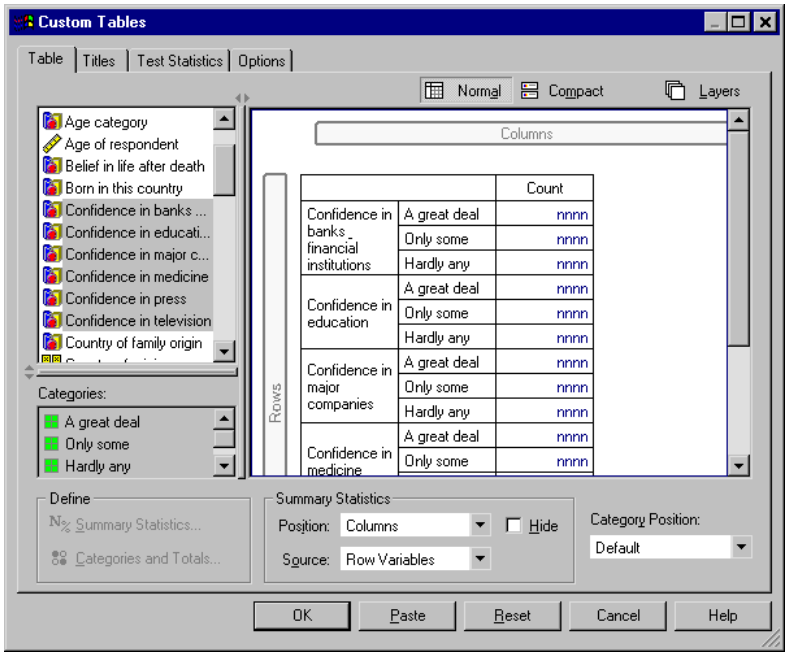
- From the menus, choose:

  Analyze
    Tables
      Custom Tables...

- In the variable list in the table builder, click *Confidence in banks...* and then Shift-click *Confidence in television* to select all of the "confidence" variables. (*Note*: This assumes that variable labels are displayed in alphabetical order, not file order, in the variable list.)

- Drag and drop the six confidence variables to the Rows area on the canvas pane.

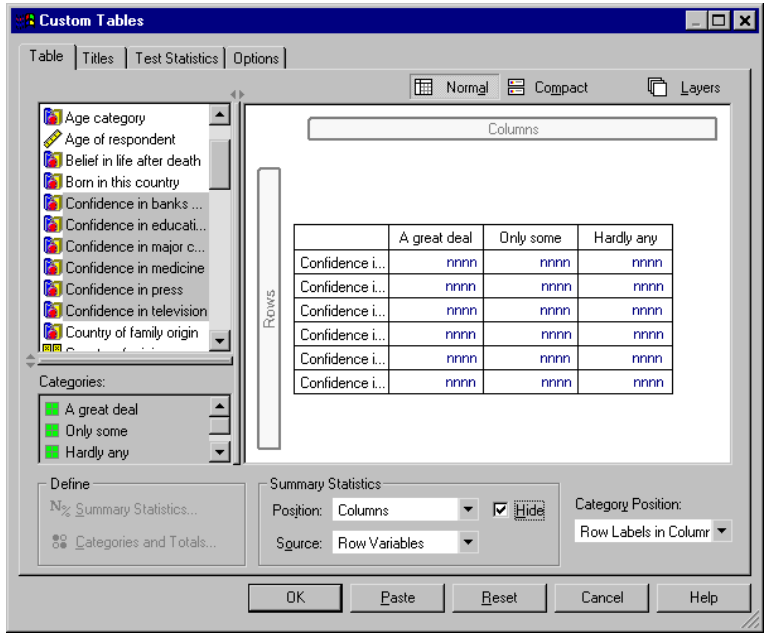
Figure 6-2
Confidence variables stacked in rows



This stacks the variables in the row dimension. By default, the category labels for each variable are also displayed in the rows, resulting in a very long, narrow table (6 variables × 3 categories = 18 rows)—but since all six variables share the same defined category labels (value labels), you can put the category labels in the column dimension.

- From the Category Position drop-down list, select Row Labels in Columns.
- Now the table has only six rows, one for each of the stacked variables, and the defined categories become columns in the table.
- Before creating the table, select (click) Hide for Position in the Summary Statistics group, since the summary statistic label *Count* isn't really necessary.

Figure 6-3
Category labels in columns



- Click OK to create the table.

Figure 6-4
Table of stacked row variables with shared category labels in columns

	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

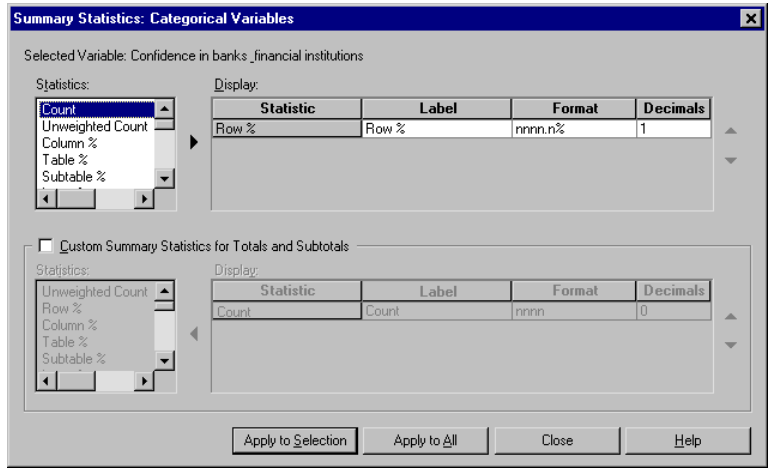
Instead of displaying the variables in the rows and categories in the columns, you could create a table with the variables stacked in the columns and the categories displayed in the rows. This might be a better choice if there were more categories than variables, whereas in our example there are more variables than categories.

Table of Percentages

For a table with variables stacked in rows and categories displayed in columns, the most meaningful (or at least easiest to understand) percentage to display is row percentages. (For a table with variables stacked in the columns and categories displayed in the rows, you would probably want column percentages.)

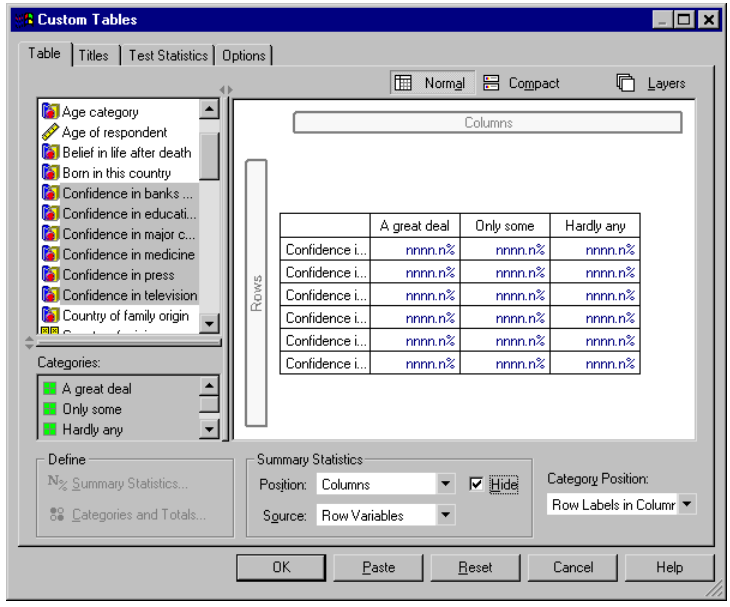
- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Right-click on any one of the confidence variables in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- Select Row % in the Statistics list and click the arrow button to move it to the Display list.
- Click on any cell in the *Count* row in the Display list and click the arrow button to move it back to the Statistics list, removing it from the Display list.

Figure 6-5
Row percentages selected in Summary Statistics dialog box



- Click Apply to All to apply the summary statistic change to all of the stacked variables in the table.

Figure 6-6
Row percentages in table preview on canvas pane



Note: If your table preview doesn't look like this figure, you probably clicked Apply to Selection instead of Apply to All, which applies the new summary statistic only to the selected variable. In this example, that would result in two columns for each category: one with count placeholders displayed for all of the other variables and one with a row percentage placeholder displayed for the selected variable. This is exactly the table that would be produced but *not* the one that we want in this example.

- Click OK to create the table.

Figure 6-7
Table of row percentages for variables stacked in rows, categories displayed in columns

	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%

Note: You can include any number of summary statistics in a table of variables with shared categories. Our examples show only one at a time to keep them simple.

Totals and Category Control

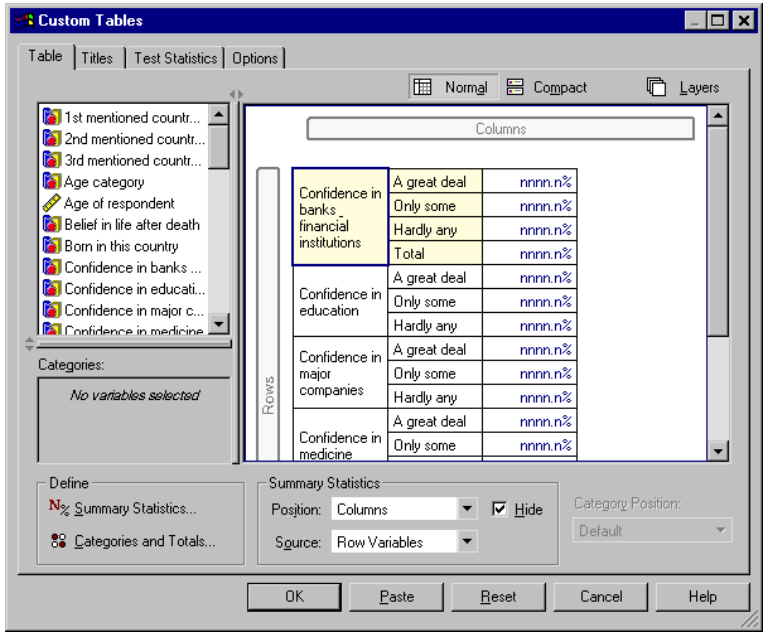
You can create tables with categories in the opposite dimension from the variables only if all of the variables in the table have the same categories, displayed in the same order. This includes totals, subtotals, and any other category adjustments you make. This means that any modifications you make in the Categories and Totals dialog box must be made for all variables in the table that share the categories.

- Open the table builder again (Analyze menu, Tables, Custom Tables).

- Right-click on the first confidence variable in the table preview on the canvas pane and select Categories and Totals from the pop-up context menu.

▶ Select (check) Total in the Categories and Totals dialog box and then click Apply.

Figure 6-8
Probably not the results you want



The first thing you'll probably notice is that the category labels have moved from the columns back to the rows. You may also notice that the Category Position control is now disabled. This is because the variables no longer share the exact same set of "categories." One of the variables now has a total category.

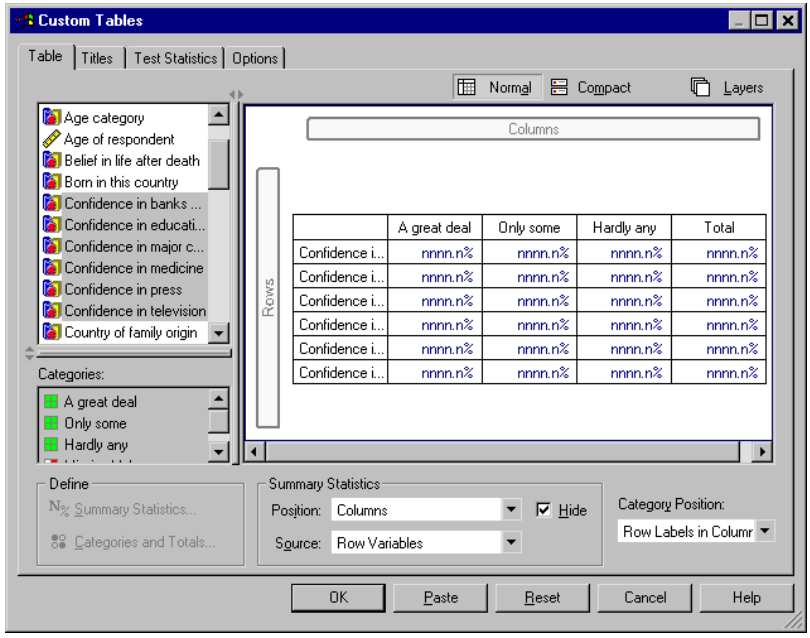
▶ Right-click any one of the confidence variables on the canvas pane and select Select All Row Variables from the pop-up context menu—or Ctrl-click each stacked variable on the canvas pane until they are all selected (you may have to scroll down the pane or expand the table builder window).

▶ Click Categories and Totals in the Define group.

▶ If Total isn't already selected (checked) in the Categories and Totals dialog box, select it now and then click Apply.

- The Category Position drop-down list should be enabled again, since now all of the variables have the additional total category, so you can now select Row Labels in Columns.

Figure 6-9
Categories and totals in columns



- Click OK to create the table.

Figure 6-10
Table of row percentages for variables stacked in rows, categories and totals displayed in columns

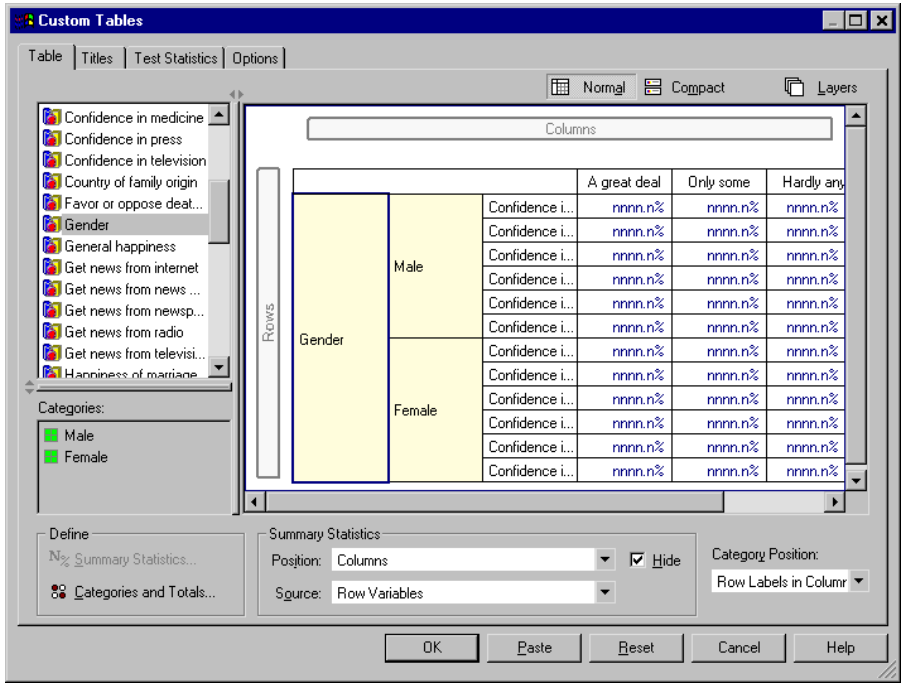
	A great deal	Only some	Hardly any	Total
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%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

Nesting in Tables with Shared Categories

In nested tables, the stacked variables with the shared categories must be at the innermost nesting level of their dimension if you want to display the category labels in the opposite dimension.

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Drag and drop *Gender* from the variable list to the left side of the Rows area.

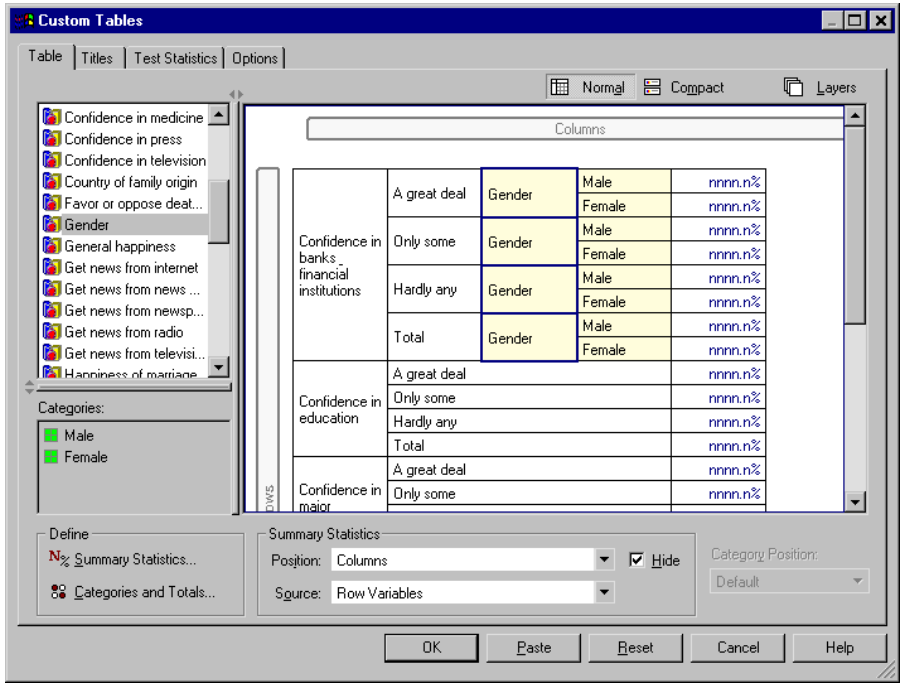
Figure 6-11
Nested variables with shared categories



The stacked variables with shared categories are now nested within gender categories in the table preview.

- Now drag and drop *Gender* to the right of one of the stacked confidence variables in the table preview.

Figure 6-12
Another example of results you probably do not want



Once again, the category labels have reverted back to the row dimension, and the Category Position control is disabled. You now have one stacked variable that also has *Gender* nested within it, while the other stacked variables contain no nested variables. You could add *Gender* as a nested variable to each of the stacked variables, but then moving row labels to columns would result in the category labels for *Gender* being displayed in the columns, not the category labels for the stacked variables with the shared categories. This is because *Gender* would now be the innermost nested variable, and changing the category position always applies to the innermost nested variable.

Chapter 7

Summary Statistics

Summary statistics include everything from simple counts for categorical variables to measures of dispersion, such as the standard error of the mean for scale variables. It does *not* include significance tests available on the Test Statistics tab in the Custom Tables dialog box. Significance tests are covered in the chapter *Test Statistics*.

Summary statistics for categorical variables and multiple response sets include counts and a wide variety of percentage calculations, including:

- Row percentages
- Column percentages
- Subtable percentages
- Table percentages
- Valid N percentages

In addition to the summary statistics available for categorical variables, summary statistics for scale variables and custom total summaries for categorical variables include:

- Mean
- Median
- Percentiles
- Sum
- Standard deviation
- Range
- Minimum and maximum values

Additional summary statistics are available for multiple response sets. For more information, see “Counts, Responses, Percentages, and Totals” in the chapter *Multiple Response Sets*. For a complete list of summary statistics see “Summary Statistics” in the chapter *Table Builder Interface*.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav*. This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

Summary Statistics Source Variable

Available summary statistics depend on the measurement level of the summary statistics source variable. The source of summary statistics (the variable on which the summary statistics are based) is determined by:

- **Measurement level**. If a table (or a table section in a stacked table) contains a scale variable, summary statistics are based on the scale variable.
- **Variable selection order**. The default statistics source dimension (row or column) for categorical variables is based on the order in which you drag and drop variables onto the canvas pane. For example, if you drag a variable to the rows area first, the row dimension is the default statistics source dimension.
- **Nesting**. For categorical variables, summary statistics are based on the innermost variable in the statistics source dimension.

A stacked table may have multiple summary statistics source variables (both scale and categorical), but each table section has only one summary statistics source.

Summary Statistics Source for Categorical Variables

- From the menus, choose:

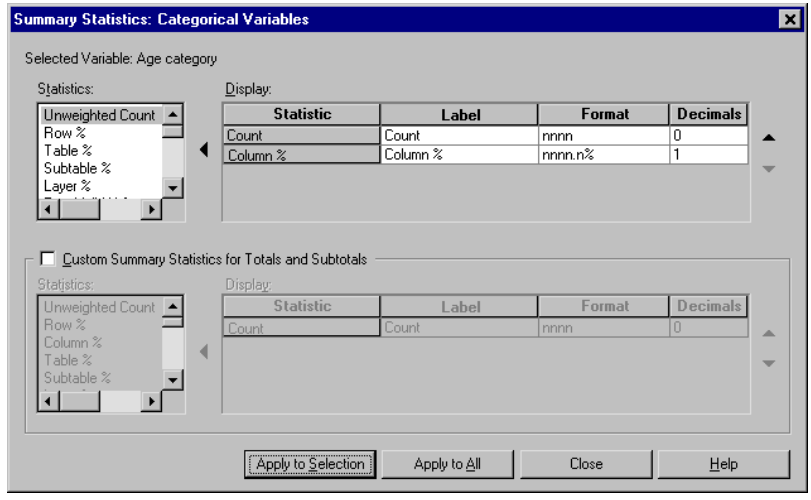
  Analyze
    Tables
      Custom Tables...

- In the table builder, drag and drop *Age category* from the variable list into the Rows area of the canvas pane.

- Right-click on *Age category* on the canvas pane and select Summary Statistics from the pop-up context menu. (Since this is the only variable in the table, it is the statistics source variable.)

- In the Summary Statistics dialog box, select *Column %* in the Statistics list and click the arrow to add it to the Display list.

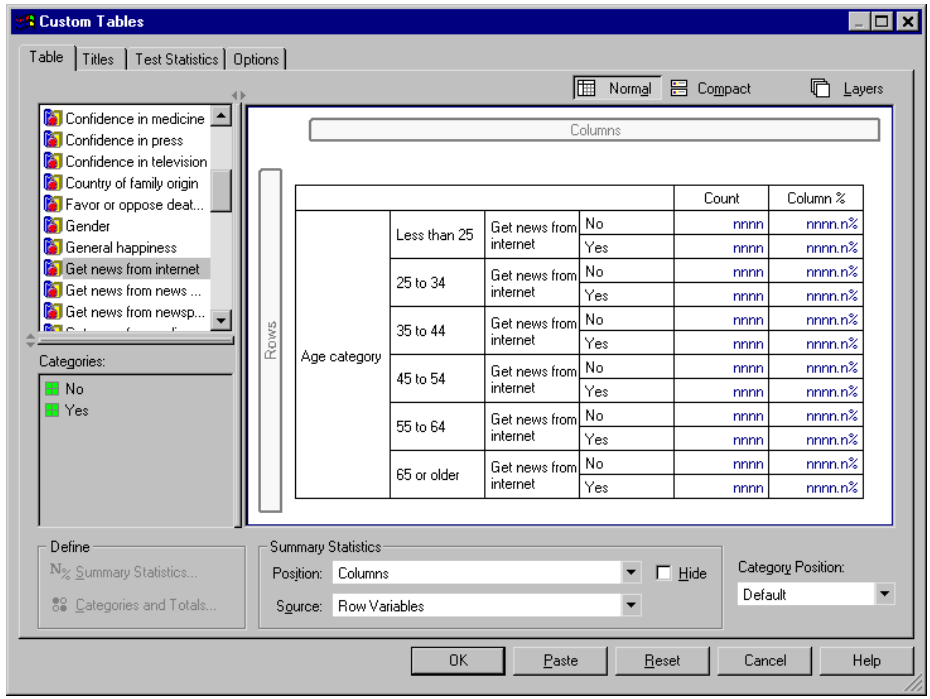
Figure 7-1
Summary Statistics dialog box for categorical variables



- Click Apply to Selection.

- In the table builder, drag and drop *Get news from internet* to the right of *Age category* on the canvas pane.

Figure 7-2
Nested categorical variables

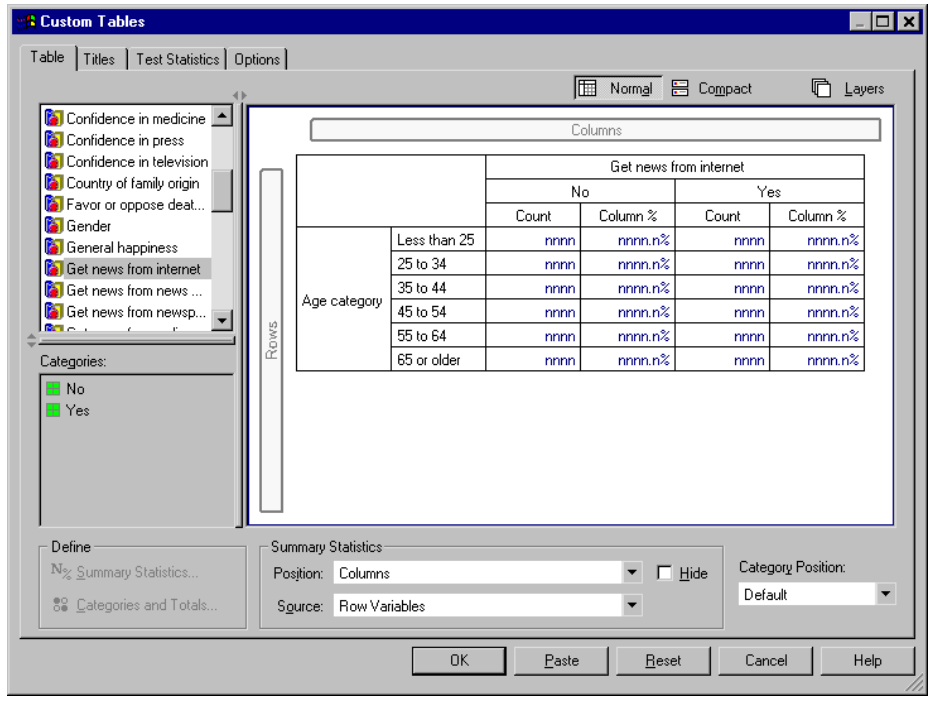


- Right-click on *Age category* on the canvas pane again. The Summary Statistics item on the context menu is now disabled, because *Age category* is not the innermost nested variable in the statistics source dimension.

- Right click on *Get news from internet* on the canvas pane. The Summary Statistics item is enabled because this is now the summary statistics source variable, because it is the innermost nested variable in the statistics source dimension. (Since the table has only one dimension—rows—it is the statistics source dimension.)

- Drag and drop *Get news from internet* from the Rows area on the canvas pane into the Columns area.

Figure 7-3
Crosstabulated categorical variables



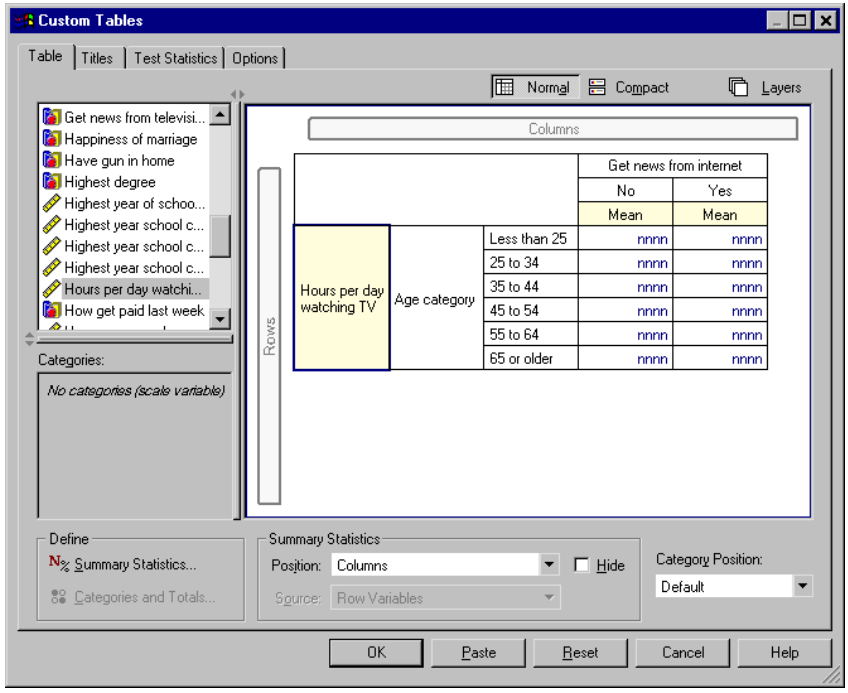
- Right-click on *Get news from internet* on the canvas pane again. The Summary Statistics item on the pop-up context menu is now disabled, because the variable is no longer in the statistics source dimension.

Age category is once again the statistics source variable, because the default statistics source dimension for categorical variables is the first dimension where you put variables when creating the table. In this example, the first thing we did was put variables in the row dimension. Thus, the row dimension is the default statistics source dimension; and since *Age category* is now the only variable in that dimension, it is the statistics source variable.

Summary Statistics Source for Scale Variables

- Drag and drop the scale variable *Hours per day watching TV* to the left of *Age category* in the Rows area of the canvas pane.

Figure 7-4
Crosstabulation with scale summary statistics variable



The first thing you may notice is that the *Count* and *Column %* summaries have been replaced with *Mean*—and if you right-click on *Hours per day watching TV* on the canvas pane, you'll see that it is now the summary statistics source variable. For a table with a scale variable, the scale variable is always the statistics source variable regardless of its nesting level or dimension, and the default summary statistic for scale variables is the mean.

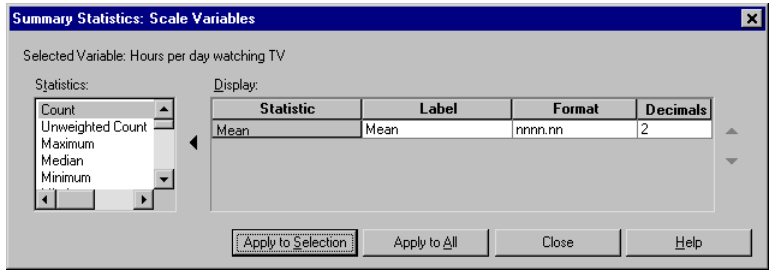
- Drag and drop *Hours per day watching TV* from the Rows area into the Columns area above *Get news from internet*.

- Right-click on *Hours per day watching TV* and select Summary Statistics from the pop-up context menu. (It's still the statistics source variable even when you move it to a different dimension.)

- In the Summary Statistics dialog box, click the Format cell for the mean in the Display list and select nnnn from the Format drop-down list. (You may have to scroll up the list to find this choice.)

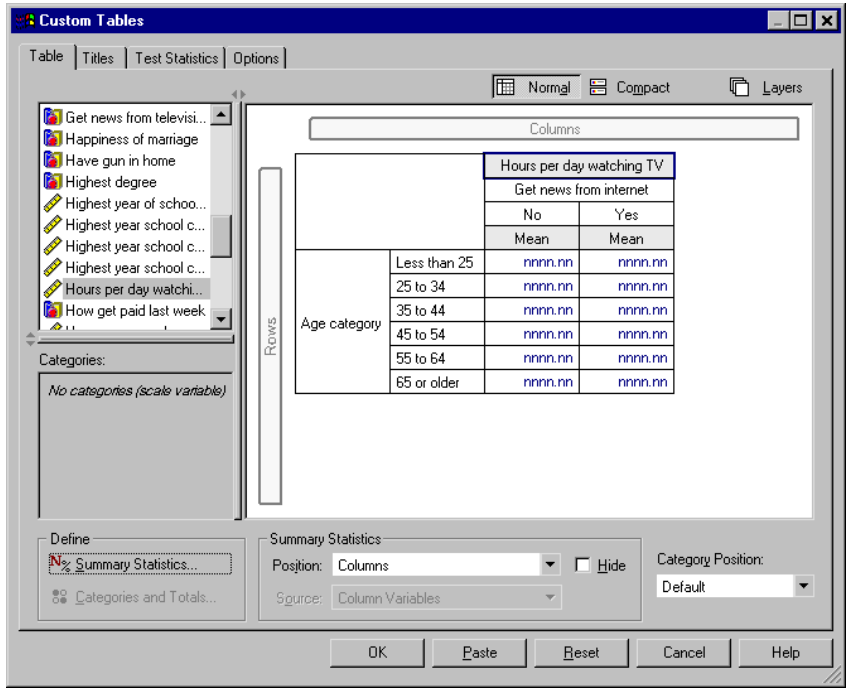
- In the Decimals cell, type 2.

Figure 7-5
Summary Statistics dialog box for scale variables



- Click Apply to Selection.

Figure 7-6
Scale summary statistic with two decimals



The table preview on the canvas pane now shows that the mean values will be displayed with two decimals.

- Click OK to create the table.

Figure 7-7
Scale variable summarized within crosstabulated categorical variables

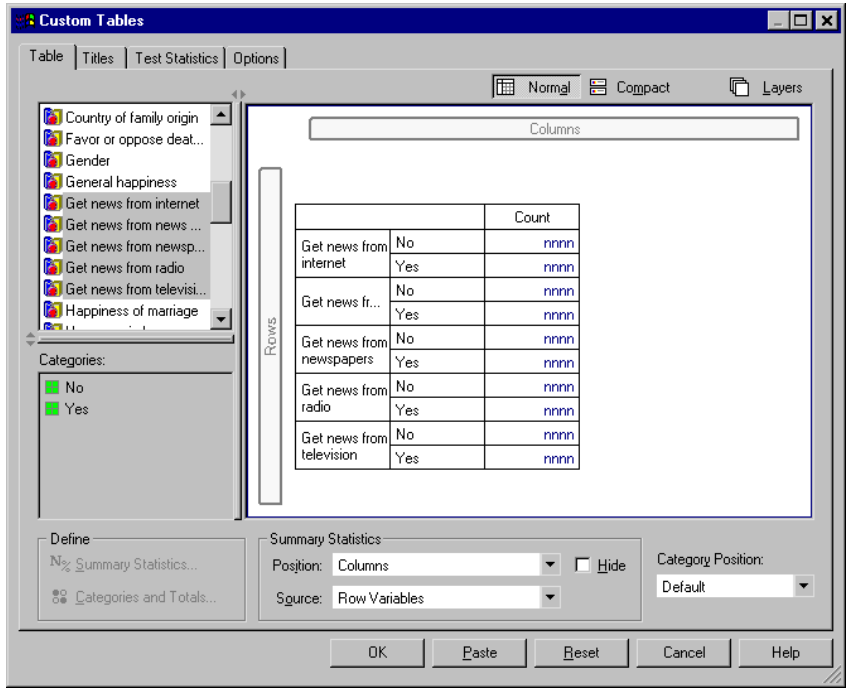
		Hours per day watching TV	
		Get news from internet	
		No	Yes
		Mean	Mean
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

Stacked Variables

Since a stacked table can contain multiple statistics source variables, and you can specify different summary statistics for each of those statistics source variables, there are a few special considerations for specifying summary statistics in stacked tables.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to clear any previous settings in the table builder.
- Click *Get news from internet* in the variable list and then shift-click *Get news from television* in the variable list to select all of the "news" variables. (*Note*: This assumes that variable labels are displayed in alphabetical order, not file order, in the variable list.)
- Drag and drop the five news variables into the Rows area of the canvas pane.

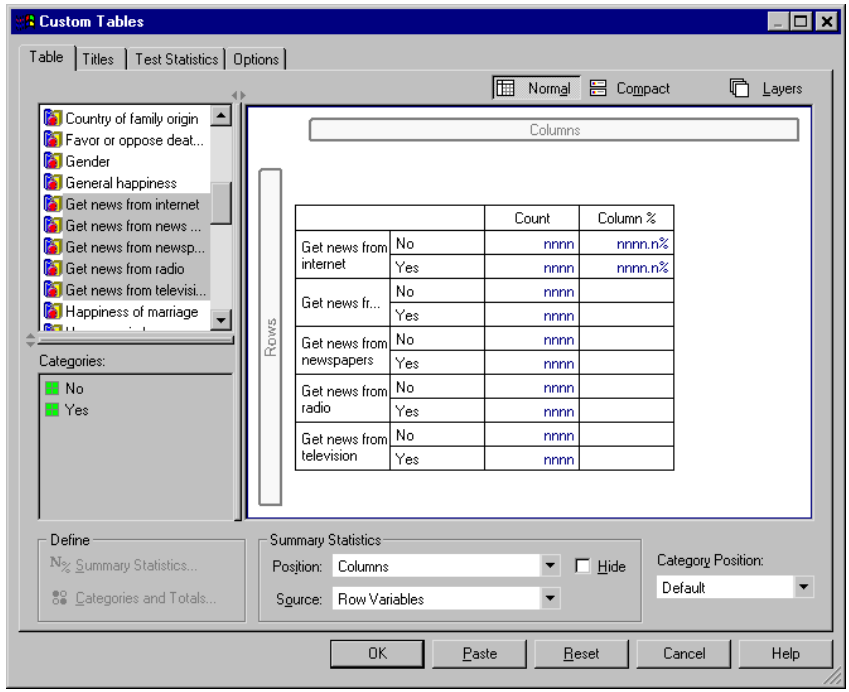
Figure 7-8
News variables stacked in rows



The five news variables are stacked in the row dimension.

- Click *Get news from internet* on the canvas pane so that only that variable is selected.
- Now right-click *Get news from internet* and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select *Column%* from the Statistics list and click the arrow to add it to the Display list. (You can use the arrow to move selected statistics from the Statistics list into the Display list, or you can drag and drop selected statistics from the Statistics list into the Display list.)
- Then click Apply to Selection.

Figure 7-9
Additional statistic applied to one variable in a stacked table

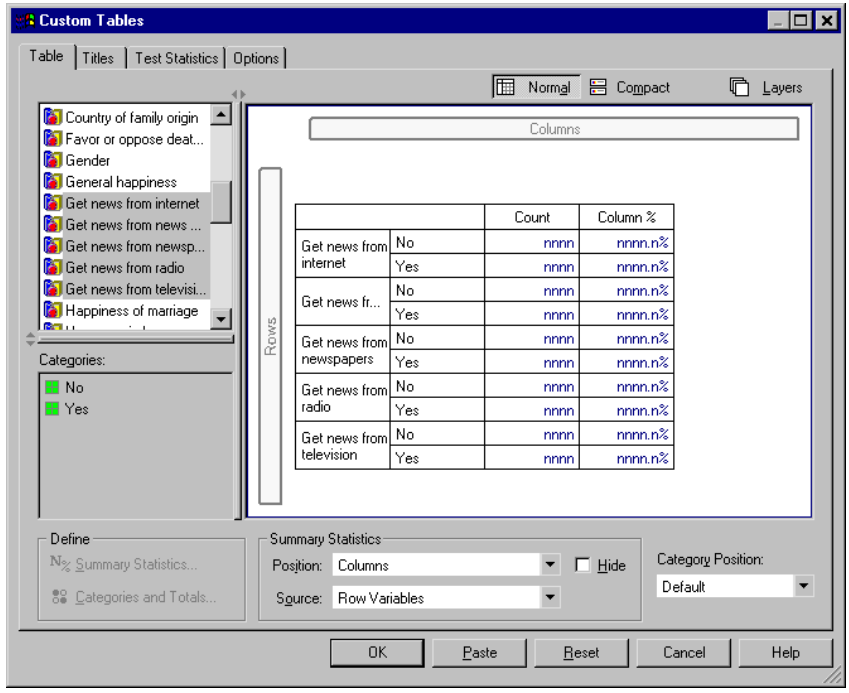


A column is added for column percentages—but the table preview on the canvas pane indicates that column percentages will be displayed for only one variable. This is because in a stacked table there are multiple statistics source variables, and each one

can have different summary statistics. In this example, however, we want to display the same summary statistics for all variables.

- Right-click *Get news from newspapers* on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select *Column%* from the Statistics list and click the arrow to add it to the Display list.
- Then click Apply to All.

Figure 7-10
Additional statistic applied to all variables in a stacked table



Now the table preview indicates that column percentages will be displayed for all of the stacked variables.

Custom Total Summary Statistics for Categorical Variables

For categorical statistics source variables, you can include custom total summary statistics that are different from the statistics displayed for the categories of the variable. For example, for an ordinal variable, you could display percentages for each category and the mean or median for the custom total summary statistic.

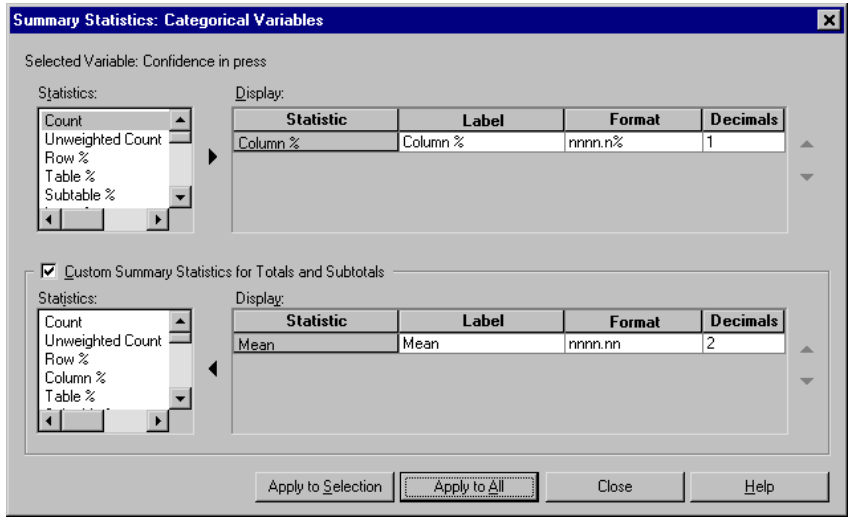
- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to clear any previous settings in the table builder.
- Click *Confidence in press* in the variable list, and then Ctrl-click *Confidence in TV* to select both variables.
- Drag and drop the two variables into the Rows area of the canvas pane. This stacks the two variables in the row dimension.
- Right-click either variable on the canvas pane and select Select All Row Variables from the pop-up context menu. (They may both already be selected, but we want to make sure.)
- Right-click the variable again and select Categories and Totals from the pop-up context menu.
- In the Categories and Totals dialog box, click (check) Total, and then click Apply.

  The table preview on the canvas pane now displays a total row for both variables. In order to display custom total summary statistics, totals and/or subtotals must be specified for the table.

- Right-click either variable on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, click *Count* in the Display list and click the arrow to move it to the Statistics list, removing it from the Display list.

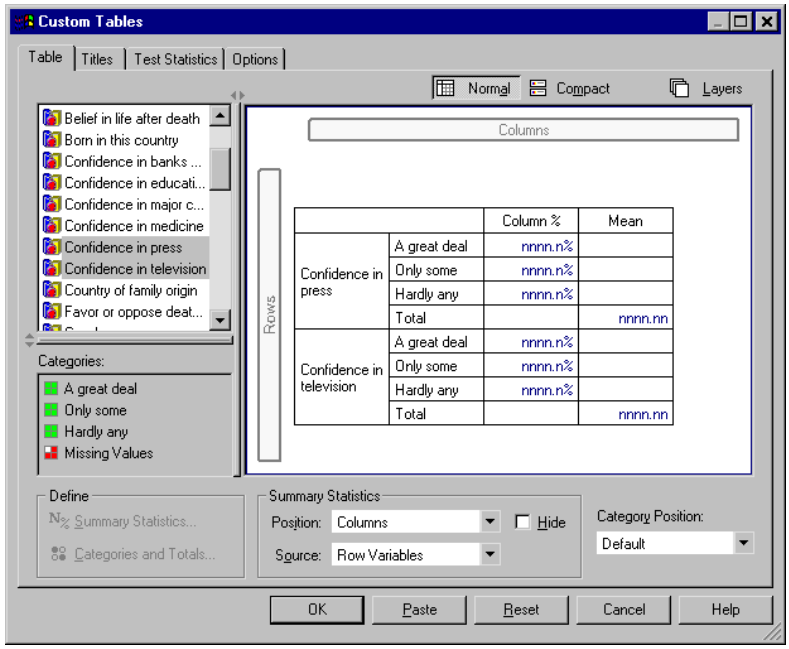
- Click *Column%* in the Statistics list and click the arrow key to move it to the Display list.
- Click (check) Custom Summary Statistics for Totals and Subtotals.
- Click *Count* in the custom summary Display list and click the arrow to move it to the custom summary Statistics list, removing it from the Display list.
- Click *Mean* in the custom summary Statistics list and click the arrow to move it to the custom summary Display list.
- Click the Format cell for the mean in the Display list and select nnnn from the drop-down list of formats. (You may have to scroll up the list to find this choice.)
- In the Decimals cell, type 2.

Figure 7-11
Selecting custom summary statistics for totals



- Click Apply to All to apply these settings to both variables in the table.

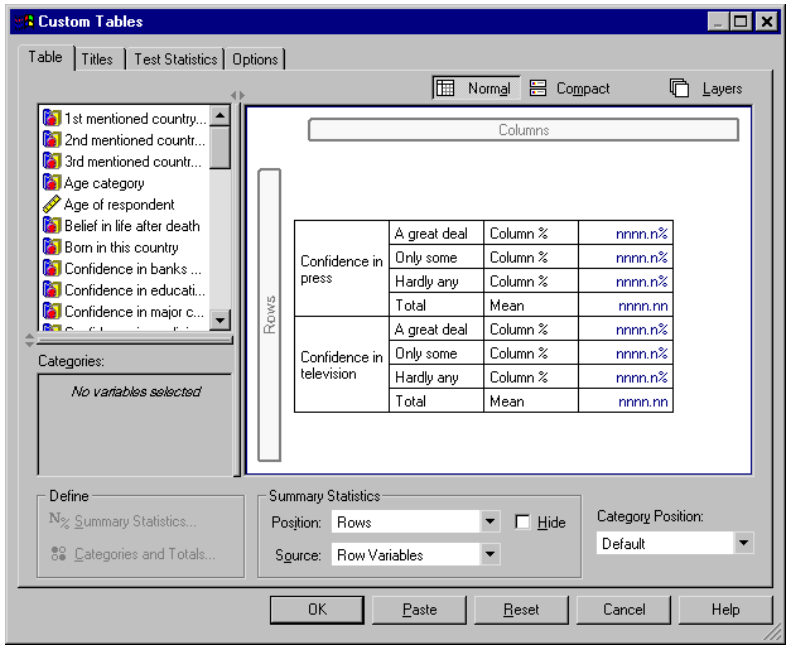
Figure 7-12
Custom total summary statistics for row variables displayed in columns



A new column has been added for the custom total summary statistic, which may not be what you want, since the preview on the canvas pane clearly indicates that this will result in a table with many empty cells.

- In the table builder, in the Summary Statistics group, select Rows from the Position drop-down list.

Figure 7-13
Summary statistics for row variables displayed in rows



This moves all the summary statistics to the row dimension, displaying all summary statistics in a single column in the table.

- Click OK to create the table.

Figure 7-14
Categorical variables with custom total summary statistics

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

Displaying Category Values

There's only one small problem with the preceding table—it may be hard to interpret the mean value without knowing the underlying category values on which it is based. Is a mean of 2.34 somewhere between *A great deal* and *Only some*—or is it somewhere between *Only some* and *Hardly any*?

Although we can't address this problem directly in Custom Tables, we can address it in a more general way.

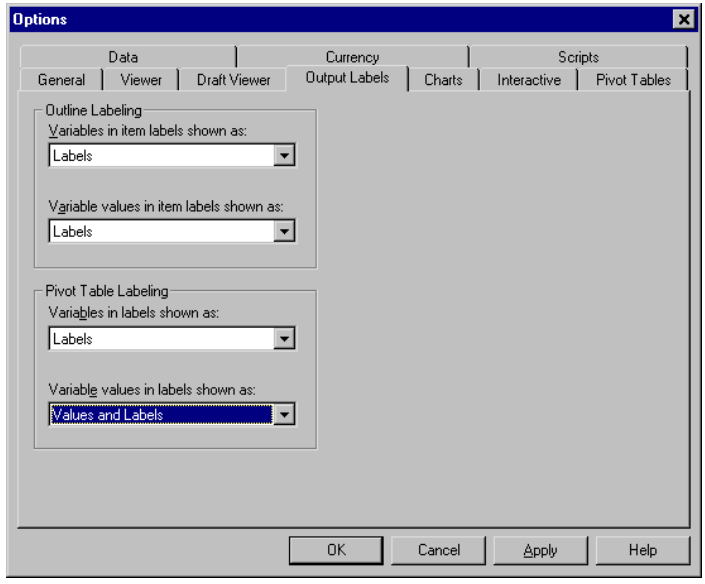
- From the menus, choose:

  Edit
  Options...

- In the Options dialog box, click the Output Labels tab.

- In the Pivot Table Labeling group, select Values and Labels from the Variable values in labels shown as drop-down list.

Figure 7-15
Output labeling options



- Click OK to save this setting.

- Open the table builder (Analyze menu, Tables, Custom Tables) and click OK to create the table again.

Figure 7-16
Values and labels displayed for variable categories

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

The category values make it clear that a mean of 2.34 is somewhere between *Only some* and *Hardly any.* Displaying the category values in the table makes it much easier to interpret the value of custom total summary statistics, such as the mean.

This display setting is a global setting that affects all pivot table output from all procedures and persists across sessions until you change it. To change the setting back to display only value labels:

- From the menus, choose:

Edit
  Options...

- In the Options dialog box, click the Output Labels tab.

- In the Pivot Table Labeling group, select Labels from the Variable values in labels shown as drop-down list.

- Click OK to save this setting.

Chapter
8

Summarizing Scale Variables

A wide range of summary statistics are available for scale variables. In addition to the counts and percentages available for categorical variables, summary statistics for scale variables also include:

- Mean
- Median
- Percentiles
- Sum
- Standard deviation
- Range
- Minimum and maximum values

For more information, see “Summary Statistics for Scale Variables and Categorical Custom Totals” in the chapter *Table Builder Interface*.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav.* This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

Stacked Scale Variables

You can summarize multiple scale variables in the same table by stacking them in the table.

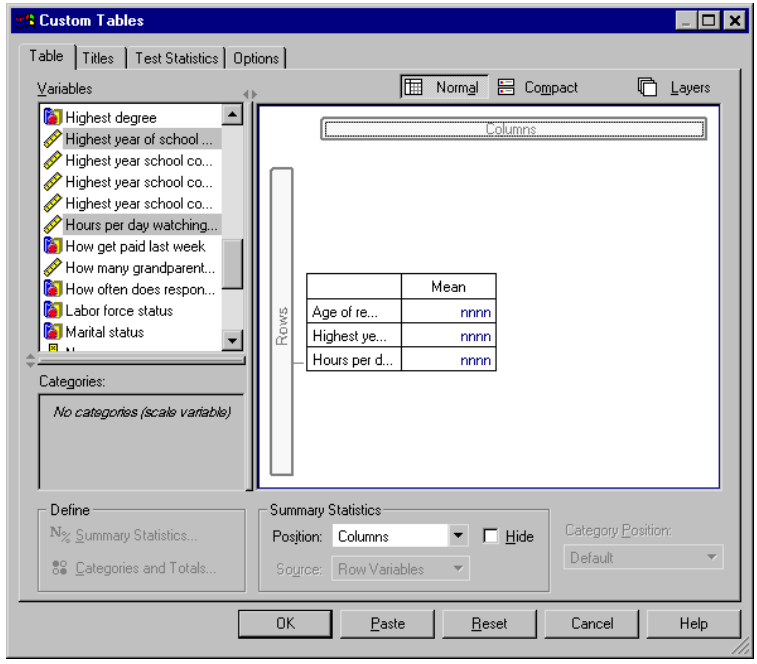
- From the menus, choose:

  Analyze
    Tables
      Custom Tables...

- In the table builder, click *Age of respondent* in the variable list, Ctrl-click *Highest year of school completed*, and Ctrl-click *Hours per day watching TV* to select all three variables.

- Drag and drop the three selected variables to the Rows area of the canvas pane.

Figure 8-1
Stacked scale variables in table builder



The three variables are stacked in the row dimension. Since all three variables are scale variables, no categories are displayed, and the default summary statistic is the mean.

- Click OK to create the table.

Figure 8-2
Table of mean values of stacked scale variables

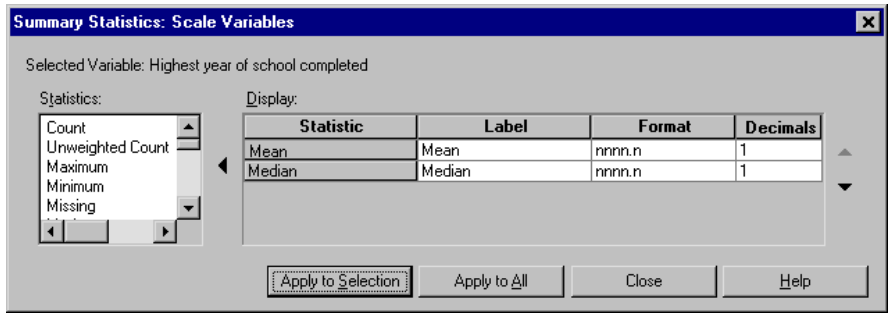
	Mean
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

Multiple Summary Statistics

By default, the mean is displayed for scale variables; however, you can choose other summary statistics for scale variables, and you can display more than one summary statistic.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click any one of the three scale variables in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select *Median* in the Statistics list and click the arrow to add it to the Display list. (You can use the arrow to move selected statistics from the Statistics list to the Display list, or you can drag and drop selected statistics from the Statistics list into the Display list.)
- Click the Format cell for the median in the Display list and select nnnn from the drop-down list of formats.
- In the Decimals cell, type 1.
- Make the same changes for the mean in the Display list.

Figure 8-3
Mean and median selected in Summary Statistics dialog box



- Click Apply to All to apply these changes to all three scale variables.
- Click OK in the table builder to create the table.

Figure 8-4
Mean and median displayed in table of stacked scale variables

	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

Count, Valid N, and Missing Values

It is often useful to display the number of cases used to compute summary statistics, such as the mean, and you might assume (not unreasonably) that the summary statistic *Count* would provide that information. However, this will not give you an accurate case base if there are any missing values. To obtain an accurate case base, use *Valid N*.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click any one of the three scale variables in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Count in the Statistics list and click the arrow to add it to the Display list.

- Then select Valid N in the Statistics list and click the arrow to add it to the Display list.
- Click Apply to All to apply these changes to all three scale variables.
- Click OK in the table builder to create the table.

Figure 8-5
Count versus Valid N

	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

For all three variable, *Count* is the same: 2,832. Not coincidentally, this is the total number of cases in the data file. Since the scale variables aren't nested within any categorical variables, *Count* simply represents the total number of cases in the data file.

Valid N, on the other hand, is different for each variable and differs quite a lot from *Count* for *Hours per day watching TV*. This is because there is a large number of **missing values** for this variable—that is, cases with no value recorded for this variable or values defined as representing missing data (such as a code of 99 to represent "not applicable" for pregnancy in males).

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click any one of the three scale variables in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Valid N in the Display list and click the arrow key to move it back to the Statistics list, removing it from the Display list.
- Select Count in the Display list and click the arrow key to move it back to the Statistics list, removing it from the Display list.
- Select Missing in the Statistics list and click the arrow key to add it to the Display list.
- Click Apply to All to apply these changes to all three scale variables.
- Click OK in the table builder to create the table.

Figure 8-6
Number of missing values displayed in table of scale summary statistics

	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

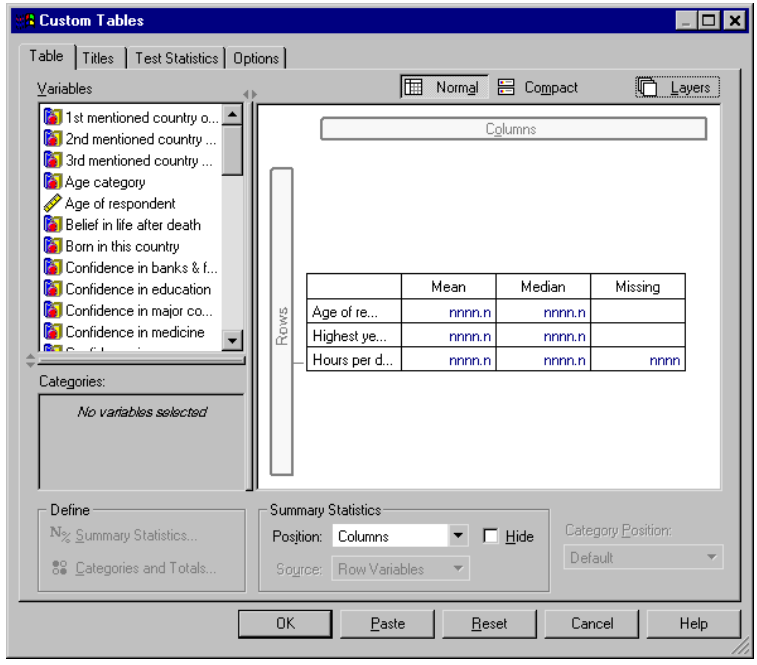
The table now displays the number of missing values for each scale variable. This makes it quite apparent that *Hours per day watching TV* has a large number of missing values, whereas the other two variables have very few. This may be a factor to consider before putting a great deal of faith in the summary values for that variable.

Different Summaries for Different Variables

In addition to displaying multiple summary statistics, you can display different summary statistics for different scale variables in a stacked table. For example, the previous table revealed that only one of the three variables has a large number of missing values; so you might want to show the number of missing values for only that one variable.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click *Age of respondent* in the table preview on the canvas pane, and then Ctrl-click *Highest year of school completed* to select both variables.
- Right-click either of the two selected variables and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Missing in the Display list and click the arrow key to move it back to the Statistics list, removing it from the Display list.
- Click Apply to Selection to apply the change to only the two selected variables.

Figure 8-7
Table preview for different summary statistics for different variables



The placeholders in the data cells of the table indicate that the number of missing values will be displayed only for *Hours per day watching TV.*

▶ Click OK to create the table.

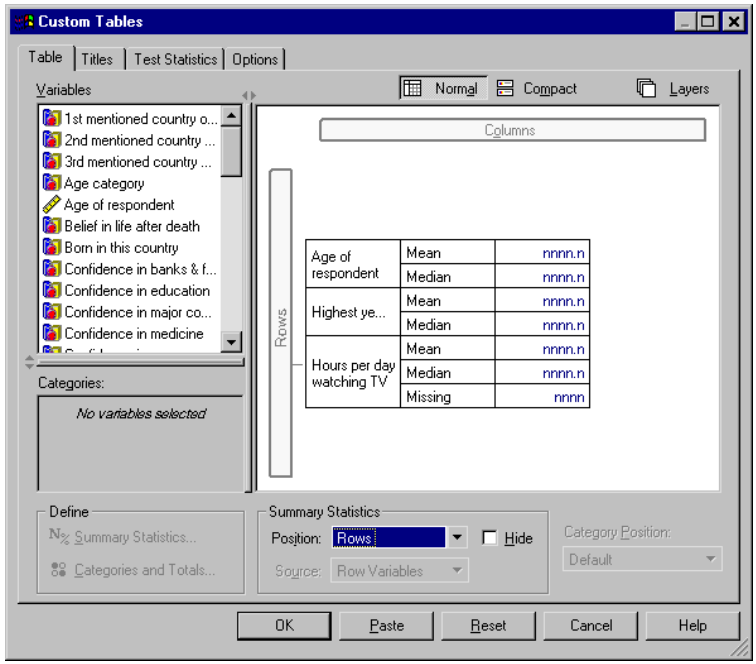
Figure 8-8
Table of different summary statistics for different variables

	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

Although this table provides the information that we want, the layout may make it difficult to interpret the table. Somebody reading the table might think that the blank cells in the *Missing* column indicate zero missing values for those variables.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- In the Summary Statistics group in the table builder, select Rows from the Position drop-down list.

Figure 8-9
Moving summary statistics from the column dimension to the row dimension



- Click OK to create the table.

Figure 8-10
Summary statistics and variables both displayed in the row dimension

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

Now it's clear that the table reports the number of missing values for only one variable.

Group Summaries in Categories

You can use categorical variables as grouping variables to display scale variable summaries within groups defined by the categories of the categorical variable.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Drag and drop *Gender* from the variable list into the Columns area of the canvas pane.

  If you right-click *Gender* in the table preview on the canvas pane, you will see that Summary Statistics is disabled on the pop-up context menu. This is because in a table with scale variables, the scale variables are always the statistics source variables.

- Click OK to create the table.

Figure 8-11
Grouped scale summaries using a categorical column variable

		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

This table makes it easy to compare the averages (mean and median) for males and females, and it clearly shows that there isn't much difference between them—which may not be terribly interesting but might be useful information.

Multiple Grouping Variables

You can subdivide the groups further by nesting and/or using both row and column categorical grouping variables.

- Open the table builder (Analyze menu, Tables, Custom Tables).

- Drag and drop *Get news from internet* from the variable list to the far left side of the Rows area of the canvas pane. Make sure to position it so that all three scale variables are nested within it, not just one of them.

Figure 8-12
Correct: All three scale variables nested within the categorical variable

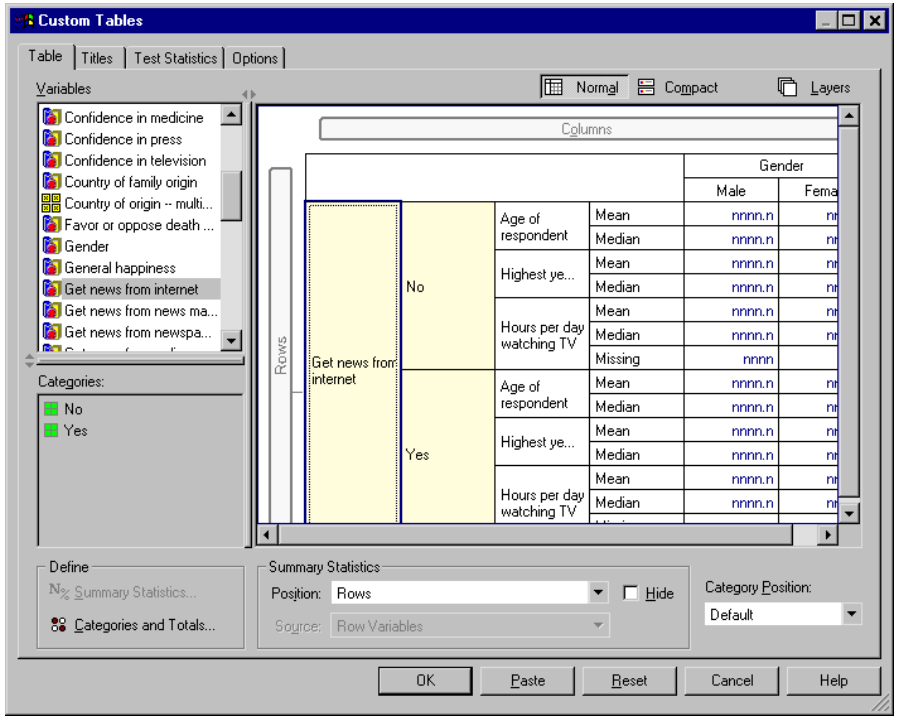
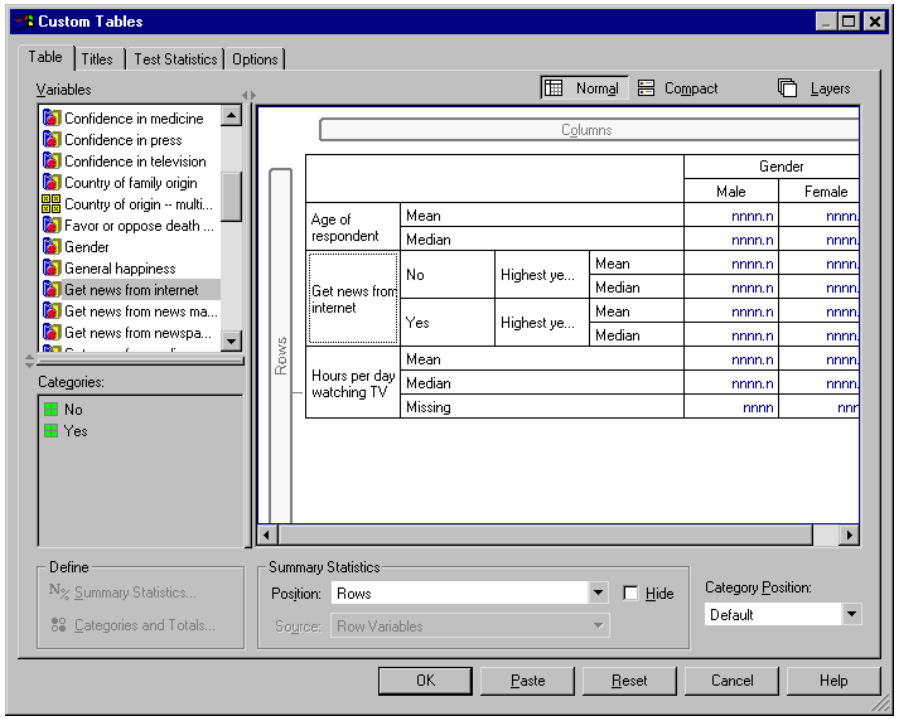


Figure 8-13
Wrong: Only one scale variable nested within the categorical variable



Although there may be times when you want something like the second example above, it's not what we want in this case.

▶ Click OK to create the table.

Figure 8-14
Scale summaries grouped by categorical row and column variables

				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
			Missing	213	282
	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

Nesting Categorical Variables within Scale Variables

Although the above table may provide the information we want, it may not provide it in the easiest format to interpret. For example, you can compare the average age of men who use the internet to get news and those who don't—but it would be easier to do if the values were next to each other rather than separated by half the table. Swapping the positions of the two row variables, nesting the categorical grouping variable within the three scale variables might improve the table. With scale variables, nesting level has no effect on the statistics source variable. The scale variable is always the statistics source variable regardless of nesting level.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click *Age of respondent* in the table preview on the canvas pane, Ctrl-click *Highest year of school completed*, and Ctrl-click *Hours per day watching TV* to select all three scale variables.
- Drag and drop the three scale variables on the far left side of the Rows area, nesting the categorical variable *Get news from internet* within each of the three scale variables.

▶ Click OK to create the table.

Figure 8-15
Categorical row variable nested within stacked scale variables

				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

The choice of nesting order depends on the relationships or comparisons you want to emphasize in the table. Changing the nesting order of the scale variables doesn't change the summary statistics values; it changes only their relative positions in the table.

Chapter
9

Test Statistics

Three different tests of significance are available for studying the relationship between row and column variables. This chapter discusses the output of each of these tests, with special attention to the effects of nesting and stacking. For more information, see the chapter *Stacking, Nesting, and Layers with Categorical Variables*.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav*. This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

Tests of Independence (Chi-Square)

The chi-square test of independence is used to determine whether there is a relationship between two categorical variables. For example, you may want to determine whether *Labor force status* is related to *Marital status*.

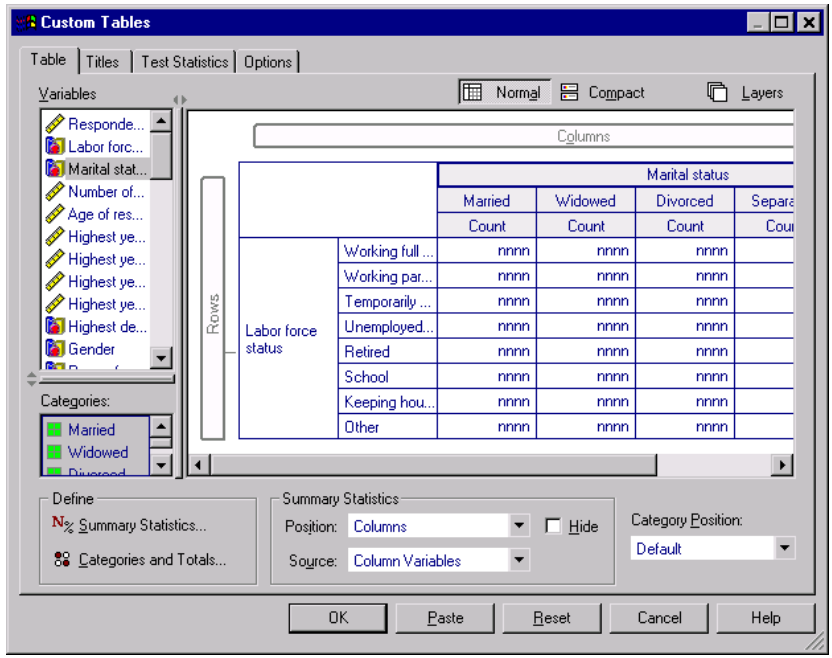
- From the menus, choose:

  Analyze
  Tables
  Custom Tables...

- In the table builder, drag and drop *Labor force status* from the variable list into the Rows area of the canvas pane.

▶ Drag and drop *Marital status* from the variable list into the Columns area.

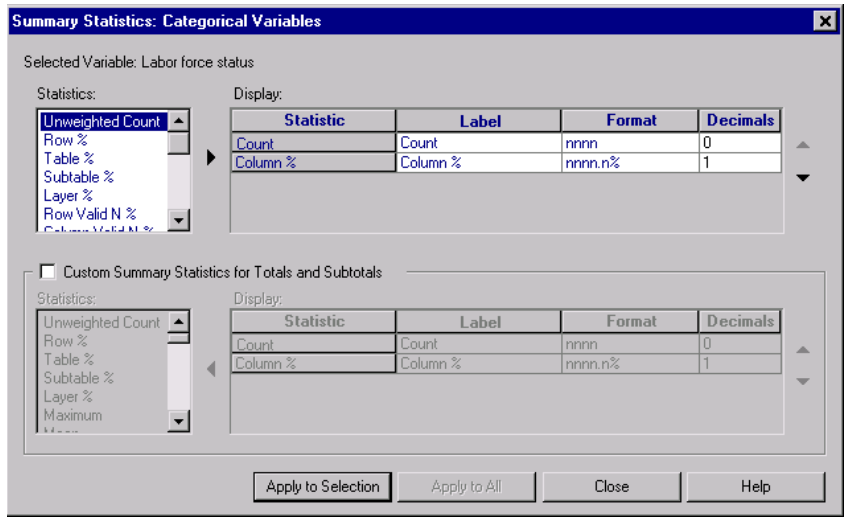
Figure 9-1
Variables displayed on canvas pane



▶ Select Rows as the position for the summary statistics.

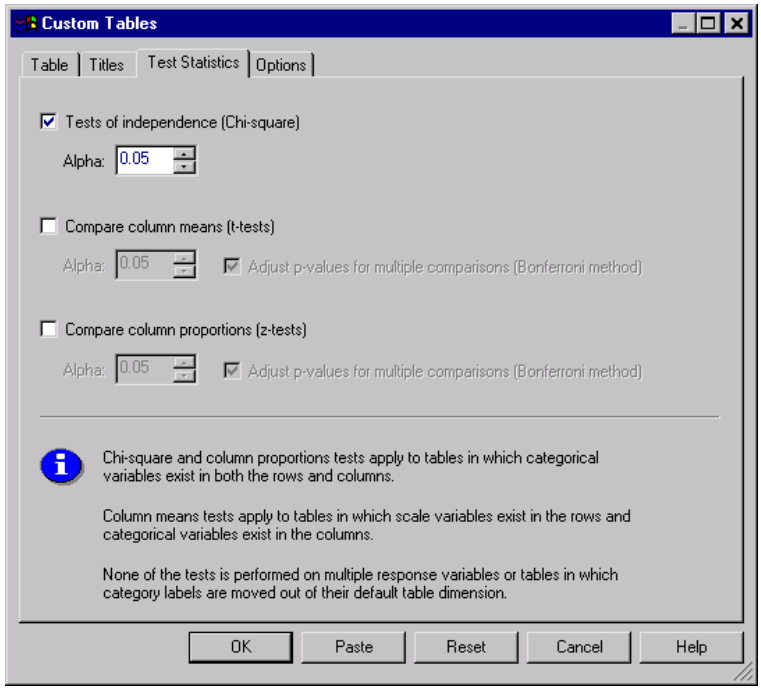
- Select *Labor force status* and click Summary Statistics in the Define group.

Figure 9-2
Summary Statistics dialog box



- Select Column % in the Statistics list and add it to the Display list.
- Click Apply to Selection.
- In the Custom Tables dialog box, click the Test Statistics tab.

Figure 9-3
Test Statistics tab with the Tests of independence (chi-square) selected



- Select Tests of independence (Chi-square).
- Click OK to create the table and obtain the chi-square test.

Figure 9-4
Labor force status by Marital status

			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%

This table is a crosstabulation of *Labor force status* by *Marital status*, with counts and column proportions shown as the summary statistics. Column proportions are computed so that they sum to 100% down each column. If these two variables are unrelated, then in each row the proportions should be similar across columns. There appear to be differences in the proportions, but you can check the chi-square test to be sure.

Figure 9-5
Pearson's chi-square test

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

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

The test of independence hypothesizes that *Labor force status* and *Marital status* are unrelated—that is, that the column proportions are the same across columns, and any observed discrepancies are due to chance variation. The chi-square statistic measures the

overall discrepancy between the observed cell counts and the counts you would expect if the column proportions were the same across columns. A larger chi-square statistic indicates a greater discrepancy between the observed and expected cell counts—greater evidence that the column proportions are not equal, that the hypothesis of independence is incorrect, and, therefore, that *Labor force status* and *Marital status* are related.

The computed chi-square statistic has a value of 729.242. In order to determine whether this is enough evidence to reject the hypothesis of independence, the significance value of the statistic is computed. The significance value is the probability that a random variate drawn from a chi-square distribution with 28 degrees of freedom is greater than 729.242. Since this value is less than the alpha level specified on the Test Statistics tab, you can reject the hypothesis of independence at the 0.05 level. Thus, *Labor force status* and *Marital status* are in fact related.

Effects of Nesting and Stacking on Tests of Independence

The rule for tests of independence is as follows: a separate test is performed for each innermost subtable. To see how nesting affects the tests, consider the previous example, but with *Marital status* nested within levels of *Gender.*

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Drag and drop *Gender* from the variable list into the Columns area of the canvas pane **above** *Marital status*.
- Click OK to create the table.

Figure 9-6
Pearson's chi-square test

		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.

With *Marital status* nested within levels of *Gender*, two tests are performed—one for each level of *Gender*. The significance value for each test indicates that you can reject the hypothesis of independence between *Marital status* and *Labor force status* for both males and females. However, the table notes that more than 20% of each table's cells have expected counts of less than 5, and the minimum expected cell count is less than 1. These notes indicate that the assumptions of the chi-square test may not be met by these tables, and so the results of the tests are suspect. The notes in this figure are numbered, rather than lettered, because this figure is formatted using the SPSS Doc TableLook instead of the default.

Note: The footnotes may be cut off from view by the cell boundaries. You can make them visible by changing the alignment of these cells in the Cell Properties dialog box.

To see how stacking affects the tests:

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Drag and drop *Highest degree* from the variable list into the Rows area **below** *Labor force status*.
- Click OK to create the table.

Figure 9-7
Pearson's chi-square test

		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
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.

With *Highest degree* stacked with *Labor force status*, four tests are performed—a test of the independence of *Marital status* and *Labor force status*, and a test of *Marital status* and *Highest degree* for each level of *Gender*. The test results for *Marital status* and *Labor force status* are the same as before. The test results for *Marital status* and *Highest degree* indicate these variables are not independent.

Comparing Column Means

The column means tests are used to determine whether there is a relationship between a categorical variable in the Columns and a continuous variable in the Rows. Moreover, you can use the test results to determine the relative ordering of categories of the categorical variable in terms of the mean value of the continuous variable. For example, you may want to determine whether *Hours per day watching TV* is related to *Get news from newspapers*.

- From the menus, choose:

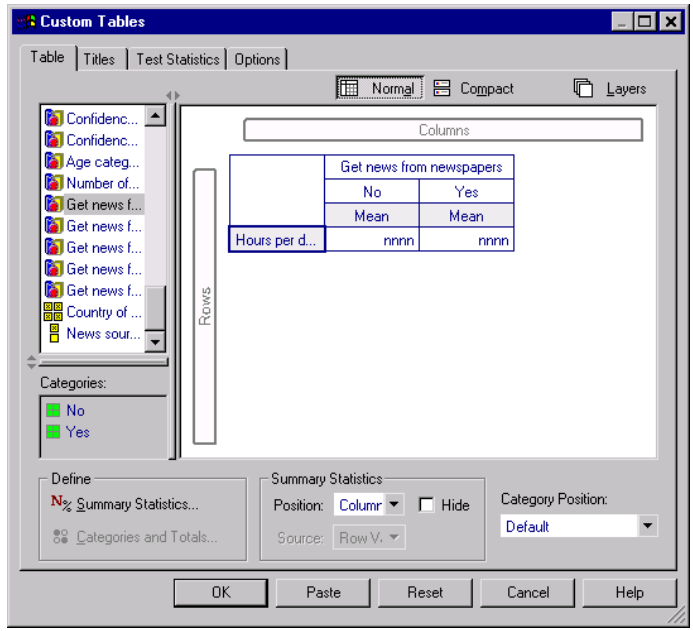
  Analyze
  Tables
  Custom Tables...

- Click Reset to restore the default settings to all tabs.

- In the table builder, drag and drop *Hours per day watching TV* from the variable list into the Rows area of the canvas pane.

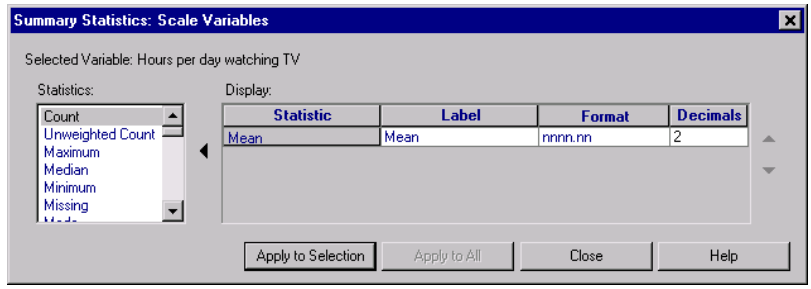
- Drag and drop *Get news from newspapers* from the variable list into the Columns area.

Figure 9-8
Variables displayed on canvas pane



▶ Select *Hours per day watching TV* and click Summary Statistics in the Define group.

Figure 9-9
Summary Statistics dialog box

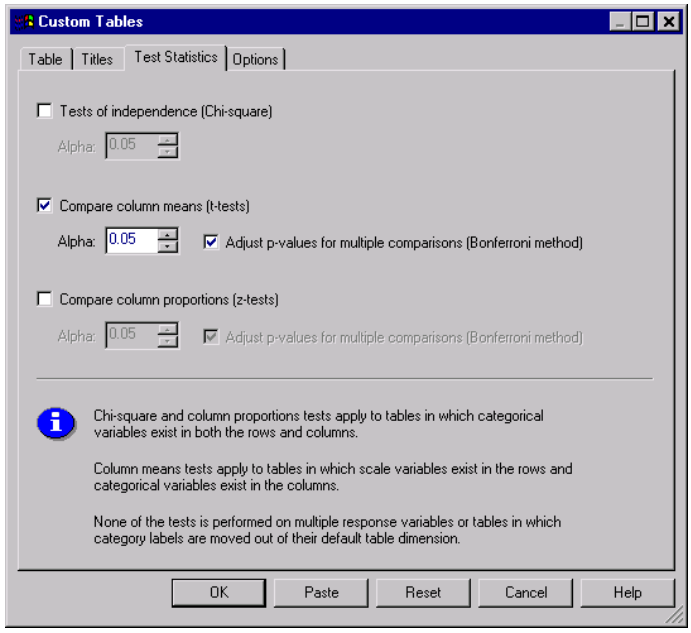


▶ Select nnnn as the format.

▶ Select 2 as the number of decimals to display. Notice that this causes the format to now read nnnn.nn.

- Click Apply to Selection.
- In the Custom Tables dialog box, click the Test Statistics tab.

Figure 9-10
Test Statistics tab with Compare column means (t tests) selected



- Select Compare column means (t-tests).
- Click OK to create the table and obtain the column means tests.

Figure 9-11
Get news from newspapers by Hours per day watching TV

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

This table shows the mean *Hours per day watching TV* for people who do and do not get their news from newspapers. The observed difference in these means suggests that

people who do not get their news from newspapers spend approximately 0.18 more hours watching TV than people who do get their news from newspapers. To see whether this difference is due to chance variation, check the column means tests.

Figure 9-12
Comparisons of column means

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

The column means test table assigns a letter key to each category of the column variable. For *Get news from newspapers*, the category *No* is assigned the letter A, and *Yes* is assigned the letter B. For each pair of columns, the column means are compared using a *t* test. Since there are only two columns, only one test is performed. For each significant pair, the key of the category with the smaller mean is placed under the category with larger mean. Since no keys are reported in the cells of the table, this means that the column means are not statistically different.

Effects of Nesting and Stacking on Column Means Tests

The rule for column means tests is as follows: a separate set of pairwise tests is performed for each innermost subtable. To see how nesting affects the tests, consider the previous example, but with *Hours per day watching TV* nested within levels of *Labor force status*.

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Drag and drop *Labor force status* from the variable list into the Rows area of the canvas pane.
- Click OK to create the table.

Figure 9-13
Comparisons of column means

			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		

With *Hours watching TV* nested within levels of *Labor force status*, seven sets of column means tests are performed: one for each level of *Labor force status*. The same letter keys are assigned to the categories of *Get news from newspapers*. For respondents *working full time*, the B key appears in A's column. This means that for full-time employees, the mean value of *Hours per day watching TV* is lower for people who get their news from newspapers. No other keys appear in the columns, so you can conclude that there are no other statistically significant differences in the column means.

Bonferroni adjustments. When multiple tests are performed, the Bonferroni adjustment is applied to column means tests to ensure that the alpha level (or false positive rate) specified on the Test Statistics tab applies to each *set* of tests. Thus, in this table, no Bonferroni adjustments were applied, because although seven sets of tests are performed, within each set only one pair of columns is compared.

To see how stacking affects the tests:

- Open the table builder again (Analyze menu, Tables, Custom Tables).

- Drag and drop *Get news from internet* from the variable list into the Columns area *to the left* of *Get news from newspapers*.

- Click OK to create the table.

Figure 9-14
Comparisons of column means

			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			

With *Get news from internet* stacked with *Get news from newspapers*, 14 sets of column means tests are performed—one for each level of *Labor force status* for *Get news from internet* and *Get news from newspapers*. Again, no Bonferroni adjustments are applied, because within each set, only one pair of columns is compared. The tests for *Get news from newspapers* are the same as before. For *Get news from internet*, the category *No* is assigned the letter A and *Yes* is assigned the letter B. The B key is reported in the A column for each set of column means tests except for those respondents *temporarily not working*. This means that the mean value of *Hours per day watching TV* is lower for people who get their news from the Internet than for people who do not get their news from newspapers. No keys are reported for the *temporarily not working* set; thus, the column means are not statistically different for these respondents.

Comparing Column Proportions

The column proportions tests are used to determine the relative ordering of categories of the Columns categorical variable in terms of the category proportions of the Rows categorical variable. For example, after using a chi-square test to find that *Labor force status* and *Marital status* are not independent, you may want to see which rows and columns are responsible for this relationship.

- From the menus, choose:

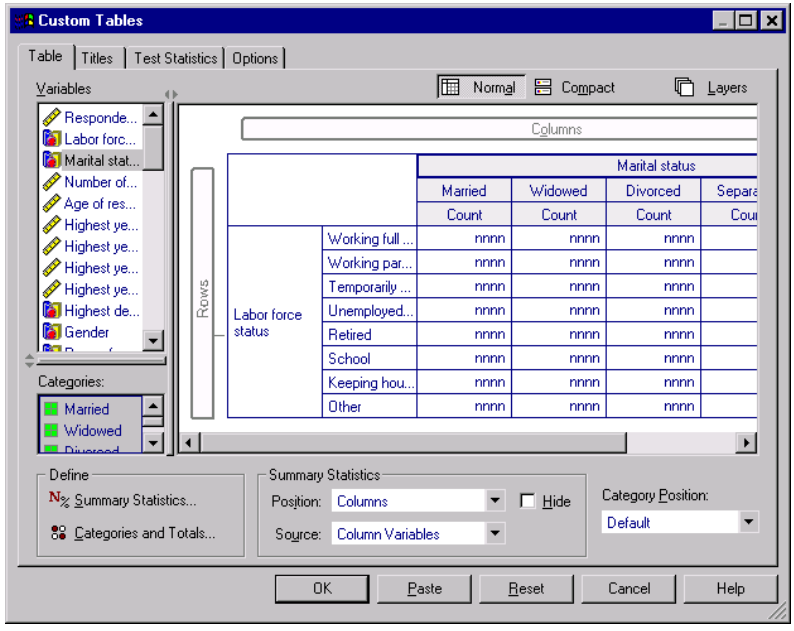
  Analyze
  Tables
  Custom Tables...

- Click Reset to restore the default settings to all tabs.

- In the table builder, drag and drop *Labor force status* from the variable list into the Rows area of the canvas pane.

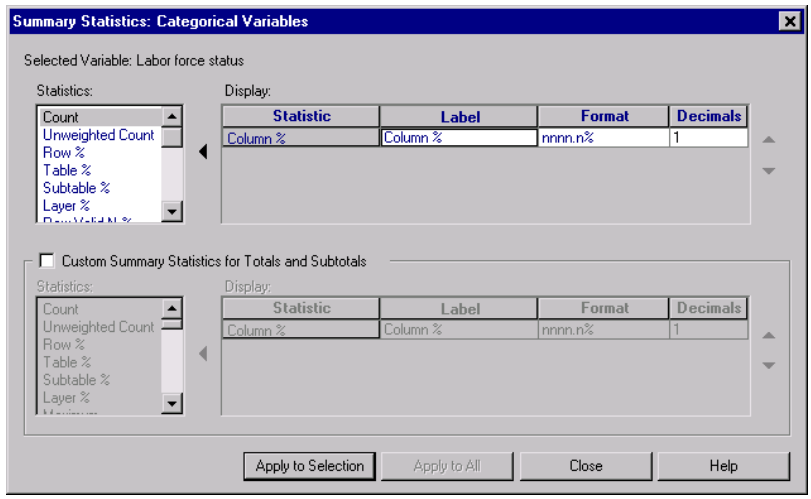
- Drag and drop *Marital status* from the variable list into the Columns area.

Figure 9-15
Variables displayed on canvas pane



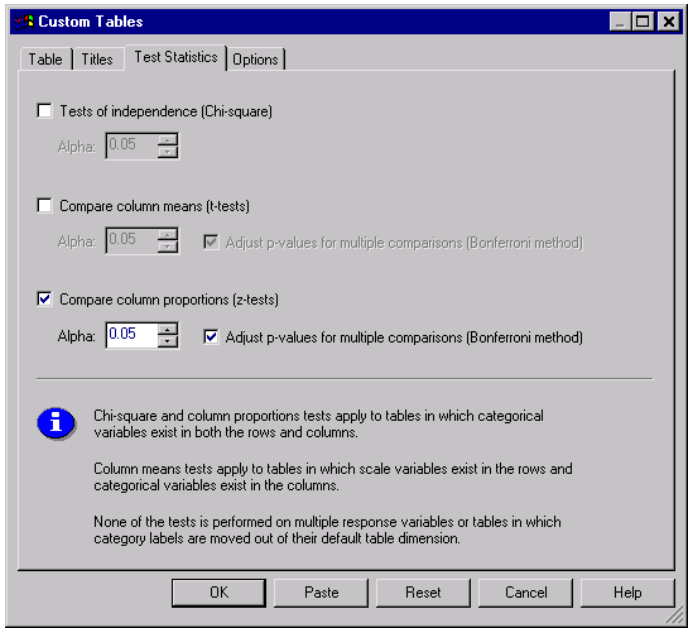
- Select *Labor force status* and click Summary Statistics in the Define group.

Figure 9-16
Summary Statistics dialog box



- Select Column % in the Statistics list and add it to the Display list.
- Deselect Count from the Display list.
- Click Apply to Selection.
- In the Custom Tables dialog box, click the Test Statistics tab.

Figure 9-17
Test Statistics tab with Compare column proportions (z tests) selected



- Select Compare column proportions (z-tests).
- Click OK to create the table and obtain the column proportions tests.

Figure 9-18
Labor force status by Marital status

		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%

This table is a crosstabulation of *Labor force status* by *Marital status*, with column proportions shown as the summary statistic.

Figure 9-19
Comparisons of column proportions

		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					
	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					

The column proportions test table assigns a letter key to each category of the column variables. For *Marital status*, the category *Married* is assigned the letter A, *Widowed* is assigned the letter B, and so on, through the category *Never married*, which is assigned the letter E. For each pair of columns, the column proportions are compared using a *z* test. Seven sets of column proportions tests are performed, one for each level of *Labor force status*. Since there are 5 levels of *Marital status*, 10 pairs of columns $((5 \times 4)/2)$ are compared in each set of tests, and Bonferroni adjustments are used to adjust the significance values. For each significant pair, the key of the smaller category is placed under the category with the larger proportion.

For the set of tests associated with *Working full time*, the B key appears in each of the other columns. Also, the A key appears in C's column. No other keys are reported in other columns. Thus, you can conclude that the proportion of divorced persons who are working full time is greater than the proportion of married persons working full time, which in turn is greater than the proportion of widowers working full time. The proportions of people who are separated or never married and working full time cannot be differentiated from people who are divorced or married and working full time, but these proportions are greater than the proportion of widowers working full time.

For the tests associated with *Working part time* or *School*, the A, B, and C keys appear in E's column. No other keys are reported in other columns. Thus, the proportions of people who have never been married and are in school or are working

part time are greater than the proportions of married, widowed, or divorced people who are in school or working part time.

For the tests associated with *Temporarily not working* or with *Other* labor status, no other keys are reported in any columns. Thus, there is no discernible difference in the proportions of married, widowed, divorced, separated, or never-married people who are temporarily not working or are in an otherwise uncategorized employment situation.

The tests associated with *Retired* show that the proportion of widowers who are retired is greater than the proportions of all other marital categories who are retired. Moreover, the proportions of married or divorced people who are retired is greater than the proportion of never-married persons who are retired.

There are greater proportions of people married, widowed, or separated and *keeping house* than proportions of people divorced or never married and keeping house.

The proportion of people who have never been married and are *unemployed, laid off* is higher than the proportions of people who are married or widowed and unemployed. Also, note that the *Separated* column is marked with a ".", which indicates that the observed proportion of separated people in the *Unemployed, laid off* row is either 0 or 1, and therefore no comparisons can be made using that column for unemployed respondents.

Effects of Nesting and Stacking on Column Proportions Tests

The rule for column proportions tests is as follows: a separate set of pairwise tests is performed for each innermost subtable. To see how nesting affects the tests, consider the previous example, but with *Labor force status* nested within levels of *Gender.*

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Drag and drop *Gender* from the variable list into the Rows area of the canvas pane.
- Click OK to create the table.

Figure 9-20
Comparisons of column proportions

				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					

With *Labor force status* nested within levels of *Gender*, 14 sets of column proportions tests are performed—one for each level of *Labor force status* for each level of *Gender.* The same letter keys are assigned to the categories of *Marital status*.

There are a couple of things to note about the table results:

- With more tests, there are more columns with zero column proportion. They are most common among separated respondents and widowed males.
- The column differences previously seen among respondents *keeping house* seems to be entirely due to females.

To see how stacking affects the tests:

- Open the table builder again (Analyze menu, Tables, Custom Tables).
- Drag and drop *Highest degree* from the variable list into the Rows area **below** *Gender*.
- Click OK to create the table.

Figure 9-21
Comparisons of column proportions

				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					
Highest degree	LT High school				A C E			
	High school							
	Junior college			B		B		B
	Bachelor			B				B
	Graduate			B				

With *Highest degree* stacked with *Gender*, 19 sets of column means tests are performed—the 14 previously discussed plus one for each level of *Highest degree*. The same letter keys are assigned to the categories of *Marital status*.

There are a few things to note about the table results:

- The test results for the 14 previously run sets of tests are the same.
- People who have *LT high school* degree are more common among widowers than among married, divorced, or never-married respondents.
- People with some post-high school education tend to be more common among those people who are married, divorced, and never married than among widowers.

Chapter
10

Multiple Response Sets

Custom Tables supports a special kind of “variable” called a **multiple response set**. Multiple response sets aren’t really “variables” in the normal sense. You can’t see them in the Data Editor, and other procedures don’t recognize them. Multiple response sets use multiple variables to record responses to questions where the respondent can give more than one answer. Multiple response sets are treated like categorical variables, and most of the things you can do with categorical variables you can also do with multiple response sets.

Multiple response sets are constructed from multiple variables in the data file. A multiple response set is a special construct within an SPSS-format data file. You can define and save multiple sets in an SPSS-format data file, but you cannot import or export multiple response sets from/to other file formats. (You can copy multiple response sets from other SPSS data files using Copy Data Properties on the Data menu in the Data Editor window.)

Note: Custom Tables does not support significance testing for tables that contain multiple response sets.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav.* This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

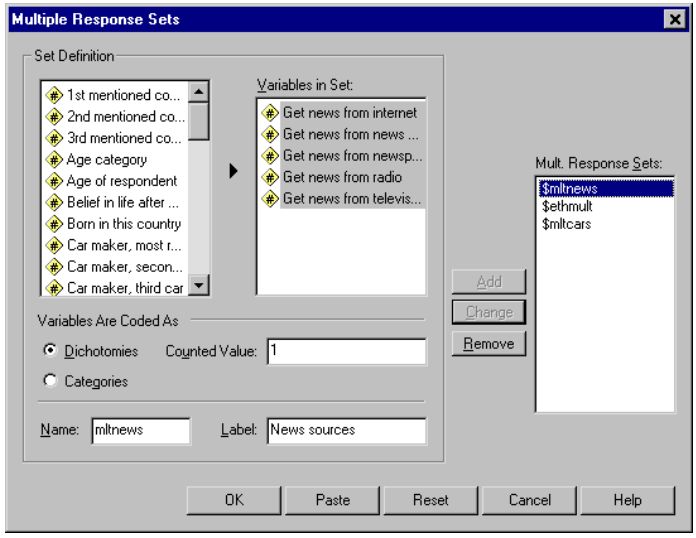
Defining Multiple Response Sets

To define multiple response sets:

- From the menus, choose:

  Analyze
  Tables
  Multiple Response Sets...

Figure 10-1
Multiple Response Sets dialog box



- Select two or more variables.
- If your variables are coded as dichotomies, indicate which value you want to have counted.
- Enter a unique name for each multiple response set. The name can be up to seven characters long. A dollar sign is automatically added to the beginning of the set name.
- Enter a descriptive label for the set. (This is optional.)
- Click Add to add the multiple response set to the list of defined sets.

Dichotomies

A multiple dichotomy set typically consists of multiple dichotomous variables: variables with only two possible values of a yes/no, present/absent, checked/not checked nature. Although the variables may not be strictly dichotomous, all of the variables in the set are coded the same way and the Counted Value represents the positive/present/checked condition.

For example, a survey asks the question, "Which of the following sources do you rely on for news?" and provides five possible responses. The respondent can indicate multiple choices by checking a box next to each choice. The five responses become five variables in the data file, coded 0 for *No* (not checked) and 1 for *Yes* (checked). In the multiple dichotomy set, the Counted Value is 1.

The sample data file already has three defined multiple response sets. *$mltnews* is a multiple dichotomy set.

- Select (click) *$mltnews* in the Mult. Response Sets list.

  This displays the variables and settings used to define this multiple response set.

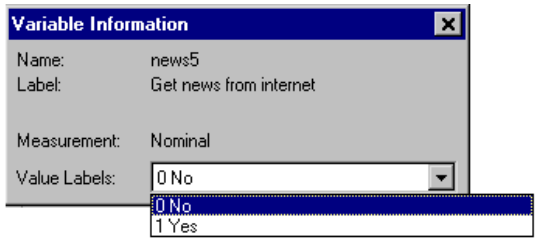
  - The Variables in Set list displays the five variables used to construct the multiple response set.
  - The Variables Are Coded As group indicates that the variables are dichotomous.
  - The Counted Value is 1.

- Select (click) one of the variables in the Variables in Set list.

- Right-click the variable and select Variable Information from the pop-up context menu.

- In the Variable Information window, click the arrow on the Value Labels drop-down list to display the entire list of defined value labels.

Figure 10-2
Variable information for multiple dichotomy source variable



The value labels indicate that the variable is a dichotomy with values of 0 and 1, representing *No* and *Yes*, respectively. All five variables in the list are coded the same way, and the value of 1 (the code for *Yes*) is the counted value for the multiple dichotomy set.

Categories

A multiple category set consists of multiple variables, all coded the same way, often with many possible response categories. For example, a survey item states, "Name up to three nationalities that best describe your ethnic heritage." There may be hundreds of possible responses, but for coding purposes the list is limited to the 40 most common nationalities, with everything else relegated to an "other" category. In the data file, the three choices become three variables, each with 41 categories (40 coded nationalities and one "other" category).

In the sample data file, *$ethmult* and *$mltcars* are multiple category sets.

Basic Rules for Multiple Response Sets

- All variables in the set should be coded the same way.
- Value labels should be used consistently. If one variable has defined value labels, all of the variables should have the same value assigned to the same value labels.
- For multiple dichotomy sets, any defined variable labels for variables in the set should be unique. Two or more variables in the set should not have the same variable label.

Counts, Responses, Percentages, and Totals

All of the summary statistics available for categorical variables are also available for multiple response sets. Some additional statistics are also available for multiple response sets.

▶ From the menus, choose:

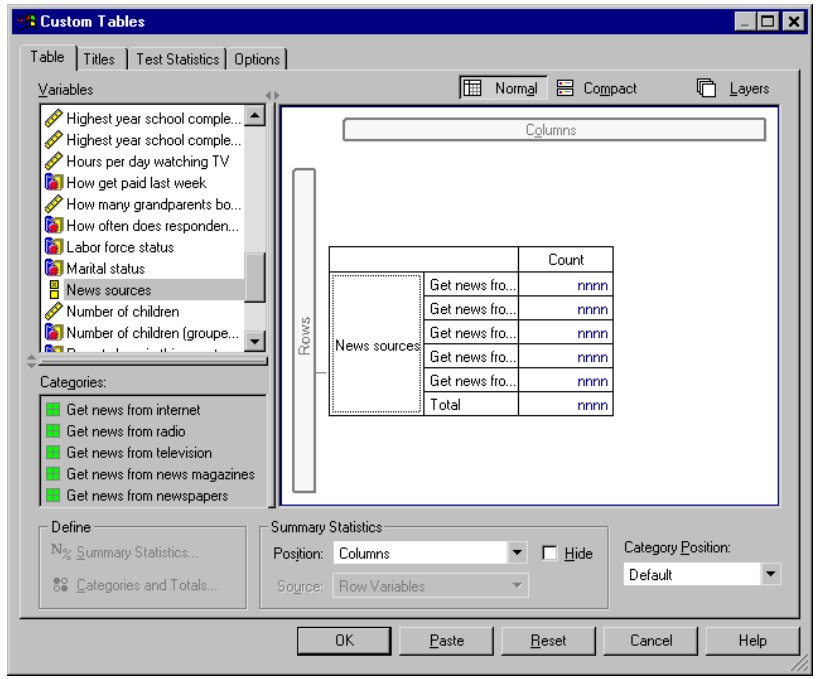
Analyze
  Tables
    Custom Tables...

- Drag and drop *News sources* (this is the descriptive label for the multiple response set *$mltnews*) from the variable list into the Rows area of the canvas pane.

The icon next to the "variable" in the variable list identifies it as a multiple dichotomy set.

Figure 10-3
Multiple dichotomy set icon

Figure 10-4
Multiple dichotomy set displayed in table preview



For a multiple dichotomy set, each "category" is, in fact, a separate variable, and the category labels are the variable labels (or variable names for variables without defined variable labels). In this example, the counts that will be displayed represent the number of cases with a *Yes* response for each variable in the set.

- Right-click *News sources* in the table preview on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select (click) Total in the Categories and Totals dialog box, and then click Apply.
- Right-click *News sources* again and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Column % in the Statistics list and click the arrow to add it to the Display list.
- Click Apply to Selection, and then click OK to create the table.

Figure 10-5
Multiple dichotomy counts and column percentages

		Count	Column %
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%

Totals That Don't Add Up

If you look at the numbers in the table, you may notice that there is a fairly large discrepancy between the "totals" and the values that are supposedly being totaled—specifically, the totals appear to be much lower than they should be. This is because the count for each "category" in the table is the number of cases with a value of 1 (a *Yes* response) for that variable, and the total number of *Yes* responses for all five variables in the multiple dichotomy set might easily exceed the total number of cases in the data file.

The total "count," however, is the total number of cases with a *Yes* response for at least one variable in the set, which can never exceed the total number of cases in the data file. In this example, the total count of 2,081 is almost 800 lower than the total number of cases in the data file. If none of these variables have missing values, this means that almost 800 survey respondents indicated that they don't get news from any of those sources. The total count is the base for the column percentages; so the column percentages in this example sum to more than the 100% displayed for the total column percentage.

Totals That Do Add Up

While "count" is typically a fairly unambiguous term, the above example demonstrates how it could be confusing in the context of totals for multiple response sets, for which *responses* is often the summary statistic you really want.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click *News sources* in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Responses in the Statistics list and click the arrow to add it to the Display list.
- Select Column Responses % in the Statistics list and click the arrow to add it to the Display list.
- Click Apply to Selection, and then click OK to create the table.

Figure 10-6
Multiple dichotomy responses and column response percentages

		Count	Column %	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%

For each "category" in the multiple dichotomy set, *Responses* is identical to *Count*—and this will always be the case for multiple dichotomy sets. The totals, however, are very different. The total number of responses is 3,594—over 1,500 more than the total count and over 700 more than the total number of cases in the data file.

For percentages, the totals for *Column %* and *Column Responses %* are both 100%—but the percentages for each category in the multiple dichotomy set are much lower for column response percentages. This is because the percentage base for column response percentages is the total number of responses, which in this case is 3,594, resulting in much lower percentages than the column percentage base of 2,081.

Percentage Totals Greater Than 100%

Both column percentages and column response percentages yield total percentages of 100% even though, in our example, the individual values in the *Column %* column clearly sum to greater than 100%. So, what if you want to show percentages based on total count rather than total responses but also want the "total" percentage to accurately reflect the sum of the individual category percentages?

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click *News sources* in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Column Responses % (Base: Count) in the Statistics list and click the arrow to add it to the Display list.
- Click Apply to Selection, and then click OK to create the table.

Figure 10-7
Column response percentages with count as the percentage base

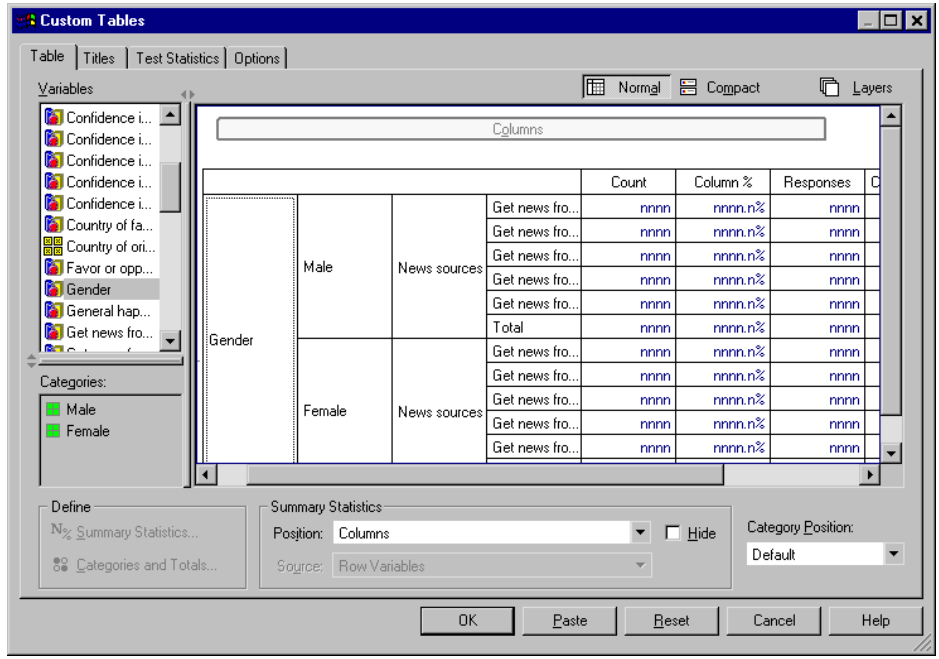
		Count	Column %	Responses	Column Responses %	Column Responses % (Base: Count)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news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

Using Multiple Response Sets with Other Variables

In general, you can use multiple response sets just like categorical variables. For example, you can crosstabulate a multiple response set with a categorical variable or nest a multiple response set within a categorical variable.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Drag and drop *Gender* from the variable list to the left side of the Rows area on the preview pane, nesting the multiple response set *News sources* within gender categories.

Figure 10-8
Table preview of nested multiple response set



- Right-click *Gender* in the table preview on the canvas pane and deselect (uncheck) Show Variable Label on the pop-up context menu.
- Do the same for *News sources*.

  This will remove the columns with the variable labels from the table (since they aren't really necessary in this case).
- Click OK to create the table.

Figure 10-9
Multiple response set nested within a categorical variable

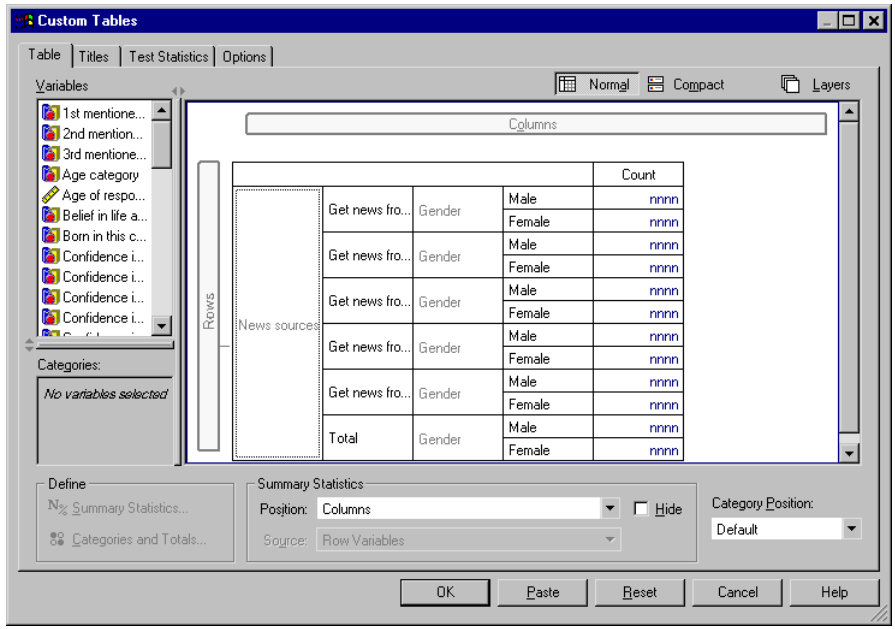
		Count	Column %	Responses	Column Responses %	Column Responses % (Base: Count)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	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%

Statistics Source Variable and Available Summary Statistics

In the absence of a scale variable in a table, categorical variables and multiple response sets are treated the same way regarding the statistics source variable: the innermost nested variable in the statistics source dimension is the statistics source variable. Since there are some summary statistics that can be assigned only to multiple response sets, this means that the multiple response set must be the innermost nested variable in the statistics source dimension if you want any of the special multiple response summary statistics.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- In the table preview on the canvas pane, drag and drop *News sources* to the left of *Gender*, changing the nesting order.

Figure 10-10
Categorical variable nested within multiple response set



All of the special multiple response summary statistics—responses, column response percentages—are removed from the table preview, because the categorical variable *Gender* is now the innermost nested variable and therefore the statistics source variable.

Luckily, the table builder "remembers" these settings. If you move *News sources* back to its previous position, nested within *Gender*, all of the response-related summary statistics are restored to the table preview.

Multiple Category Sets and Duplicate Responses

Multiple category sets provide one feature not available for multiple dichotomy sets: the ability to count duplicate responses. In many cases, duplicate responses in multiple category sets probably represent coding errors. For example, for a survey question such as "What three countries do you think make the best cars?" a response of *Sweden, Germany, and Sweden* probably isn't valid.

In other cases, however, duplicate responses may be perfectly valid. For example, if the question were "Where were your last three cars made?" a response of *Sweden, Germany, and Sweden* makes perfect sense.

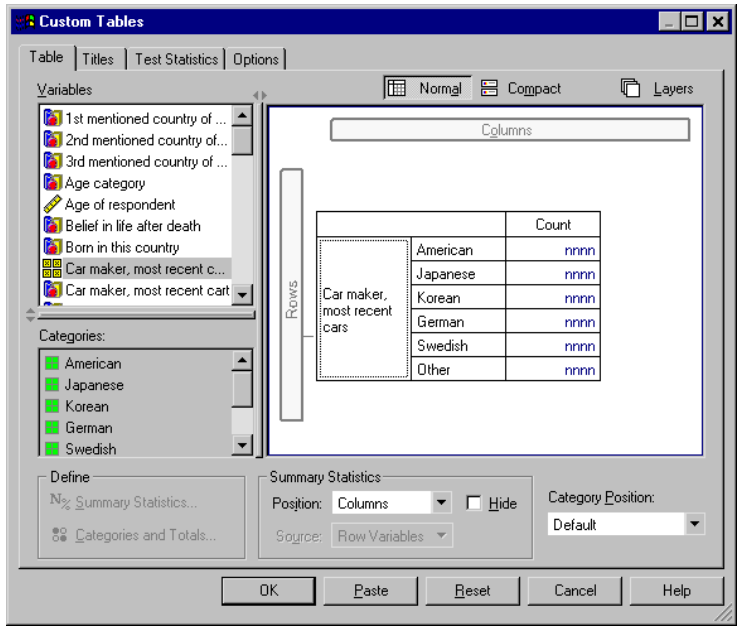
Custom Tables provides a choice for duplicate responses in multiple category sets. By default, duplicate responses are not counted, but you can request that they be included.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to clear any previous settings.
- Drag and drop *Car maker, most recent cars* from the variable list into the Rows area of the canvas pane.

The icon next to the "variable" in the variable list identifies it as a multiple category set.

Figure 10-11
Multiple category set icon

Figure 10-12
Multiple category set in table builder preview



For multiple category sets, the categories displayed represent the common set of defined value labels for all of the variables in the set (whereas for multiple dichotomy sets, the "categories" are actually the variable labels for each variable in the set).

- Right-click *Car maker, most recent cars* in the table preview on the canvas pane and select Categories and Totals from the pop-up context menu.
- Select (click) Total in the Categories and Totals dialog box, and then click Apply.
- Right-click *Car maker, most recent cars* again and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, select Responses in the Statistics list and click the arrow to add it to the Display list.
- Click Apply to Selection, and then click OK to create the table.

Figure 10-13
Multiple category set: Counts and responses without duplicates

		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

By default, duplicate responses are not counted; so in this table, the values for each category in the *Count* and *Responses* columns are identical. Only the totals differ.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click the Options tab.
- Click (check) Count duplicate responses for multiple category sets.
- Click OK to create the table.

Figure 10-14
Multiple category set with duplicate responses included

		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

In this table, there is quite a noticeable difference between the values in the *Count* and *Responses* columns, particularly for American cars, indicating that many respondents have owned multiple American cars.

Chapter
11

Missing Values

Many data files contain a certain amount of missing data. A wide variety of factors can result in missing data. For example, survey respondents may not answer every question, certain variables may not be applicable to some cases, and coding errors may result in some values being thrown out.

There are two kinds of missing values in SPSS:

- **User-missing.** Values defined as containing missing data. Value labels can be assigned to these values to identify why the data are missing (such as a code of 99 and a value label of *Not Applicable* for pregnancy in males).
- **System-missing.** If no value is present for a numeric variable, it is assigned the system-missing value. This is indicated by a period in the Data View of the Data Editor.

SPSS offers a number of facilities that can help to compensate for the effects of missing data and even analyze patterns in missing data. This chapter, however, has a much simpler goal: to describe how Custom Tables handles missing data and how missing data affect the computation of summary statistics.

Sample Data File

The examples in this chapter use the data file *missing_values.sav.* This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed. This is a very simple, completely artificial data file, with only one variable and ten cases, designed to illustrate basic concepts about missing values.

Tables without Missing Values

By default, user-missing categories are not displayed in custom tables (and system-missing values are never displayed).

- From the menus, choose:

  Analyze
    Tables
      Custom Tables...

- In the table builder, drag and drop *Variable with missing values* (the only variable in the file) from the variable list into the Rows area of the canvas pane.

- Right-click the variable on the canvas pane and select Categories and Totals from the pop-up context menu.

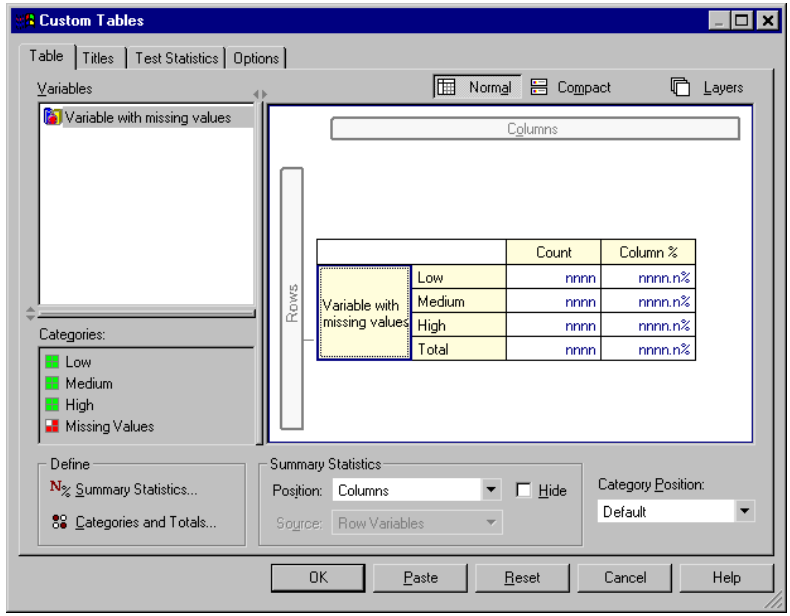
- Click (check) Total in the Categories and Totals dialog box, and then click Apply.

- Right-click *Variable with missing values* in the table preview on the canvas pane again and select Summary Statistics from the pop-up context menu.

- In the Summary Statistics dialog box, select Column % in the Statistics list and click the arrow to add it to the Display list.

- Click Apply to Selection.

Figure 11-1
Table preview without missing values



You may notice a slight discrepancy between the categories displayed in the table preview on the canvas pane and the categories displayed in the Categories list (below the variable list on the left side of the table builder). The Categories list contains a category labeled *Missing Values* that isn't included in the table preview because missing value categories are excluded by default. Since "values" is plural in the label, this indicates that the variable has two or more user-missing categories.

▶ Click OK to create the table.

Figure 11-2
Table without missing values

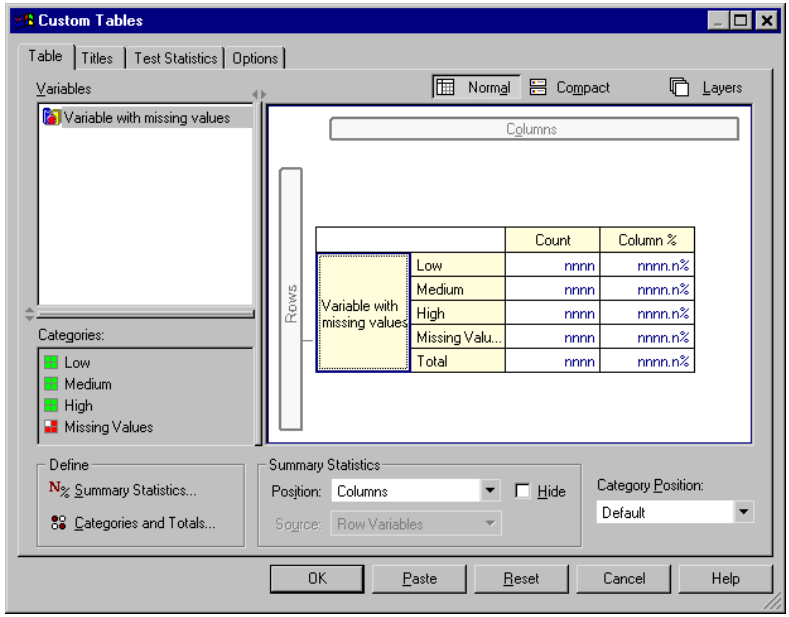
		Count	Column %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	Total	7	100.0%

Everything in this table is perfectly fine. The category values add up to the totals, and the percentages accurately reflect the values you'd get using the total count as the percentage base (for example, $3/7 = 0.429$, or 42.9%). The total count, however, is not the total number of cases in the data file; it's the total number of cases with **non-missing** values, or cases that don't have user-missing or system-missing values for that variable.

Including Missing Values in Tables

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click *Variable with missing values* in the table preview on the canvas pane and select Categories and Totals from the pop-up context menu.
- Click (check) Missing Values in the Categories and Totals dialog box, and then click Apply.

Figure 11-3
Table preview with missing values category displayed



Now the table preview includes a *Missing Values* category. Although the table preview displays only one category for missing values, all user-missing categories will be displayed in the table.

- Right-click *Variable with missing values* in the table preview on the canvas pane again and select Summary Statistics from the pop-up context menu.
- In the Summary Statistics dialog box, click (check) Custom Summary Statistics for Totals and Subtotals.
- Select Valid N in the custom summary Statistics list and click the arrow to add it to the Display list.
- Do the same for Total N.
- Click Apply to Selection, and then click OK in the table builder to create the table.

Figure 11-4
Table with missing values

		Count	Column %	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

The two defined user-missing categories—*Don't know* and *Not applicable*—are now displayed in the table, and the total count is now 9 instead of 7, reflecting the addition of the two cases with user-missing values (one in each user-missing category). The column percentages are also different now, because they are based on the number of non-missing and user-missing values. Only system-missing values are not included in the percentage calculation.

Valid N shows the total number of non-missing cases (7), and *Total N* shows the total number of cases, including both user-missing and system-missing. The total number of cases is 10, one more than the count of non-missing and user-missing values displayed as the total in the *Count* column. This is because there's one case with a system-missing value.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Right-click *Variable with missing values* in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- Select Column Valid N% in the top Statistics list (not the custom summaries for totals and subtotals) and click the arrow to add it to the Display list.
- Do the same for Column Total N%.
- You can also add them both to the list of custom summary statistics for totals and subtotals.
- Click Apply to Selection, and then click OK to create the table.

Figure 11-5
Table with missing values and valid and total percentages

		Count	Column %	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

- *Column %* is the percentage in each category based on the number of non-missing and user-missing values (since user-missing values have been explicitly included in the table).
- *Column Valid N %* is the percentage in each category based on only the valid, non-missing cases. These values are the same as the column percentages were in the original table that did not include user-missing values.
- *Column Total N %* is the percentage in each category based on all cases, including both user-missing and system-missing. If you add up the individual category percentages in this category, you'll see that they add up to only 90%, because one case out of the total of 10 cases (10%) has the system-missing value. Although this case is included in the base for the percentage calculations, no category is provided in the table for cases with system-missing values.

Chapter
12

Formatting and Customizing Tables

Custom Tables provides the ability to control a number of table-formatting properties as part of the table-building process, including:

- Display format and labels for summary statistics
- Minimum and maximum data column width
- Text or value displayed in empty cells

These settings persist within the table builder interface (until you change them, reset the table builder settings, or open a different data file), enabling you to create multiple tables with the same formatting properties without manually editing the tables after creating them. You can also save these formatting settings, along with all of the other table parameters, using the Paste button in the table builder interface to paste command syntax into a syntax window, which you can then save as a file.

You can also change many formatting properties of tables after they have been created, using all of the formatting capabilities available in the Viewer for pivot tables. This chapter, however, focuses on controlling table formatting properties before the table is created. For more information on pivot tables, use the Index tab in the Help system and type pivot tables as the keyword.

Sample Data File

The examples in this chapter use the data file *survey_sample.sav.* This file is located in the *tutorial\sample_files* folder within the folder in which SPSS is installed.

All examples provided here display variable labels in dialog boxes, sorted in alphabetical order. Variable list display properties are set on the General tab in the Options dialog box (Edit menu, Options).

Summary Statistics Display Format

Custom Tables attempts to apply relatively intelligent default formats to summary statistics, but there will probably be times when you want to override these defaults.

- From the menus, choose:

  Analyze
    Tables
      Custom Tables...

- In the table builder, drag and drop *Age category* from the variable list into the Rows area on the canvas pane.

- Drag and drop *Confidence in television* below *Age category* in the Rows area, stacking the two variables in the row dimension.

- Right-click *Age category* in the table preview on the canvas pane and select Select All Row Variables from the pop-up context menu.

- Right-click *Age category* again and select Categories and Totals from the pop-up context menu.

- In the Categories and Totals dialog box, select (check) Total and then click Apply.

- Right-click either variable in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.

- Select Column % in the Statistics list and click the arrow key to add it to the Display list.

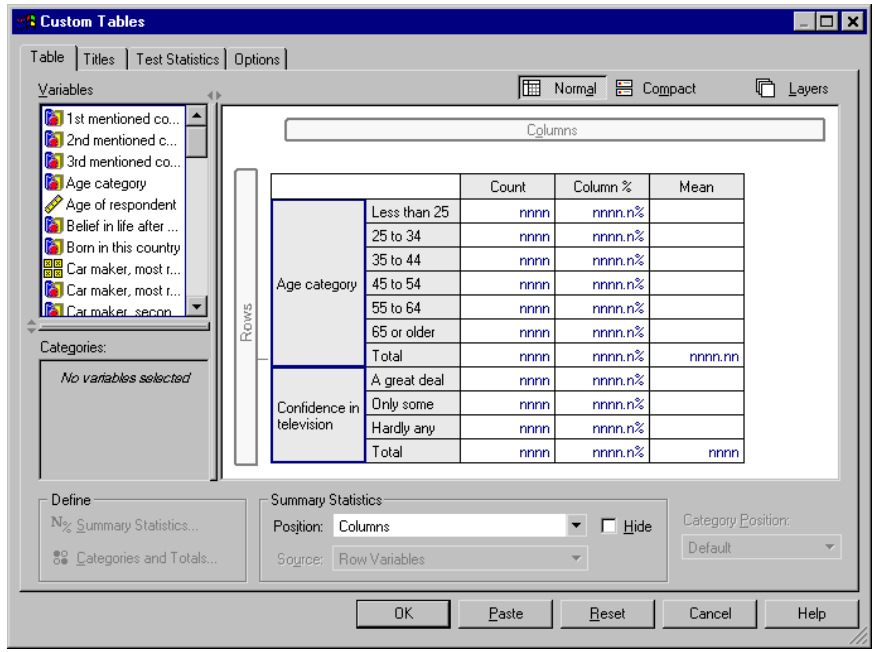
- Select (check) Custom Summary Statistics for Totals and Subtotals.

- In the Statistics list for custom summary statistics, select Column % and click the arrow to add it to the Display list.

- Do the same for Mean.

- Then click Apply to All.

Figure 12-1
Default display formats in table preview

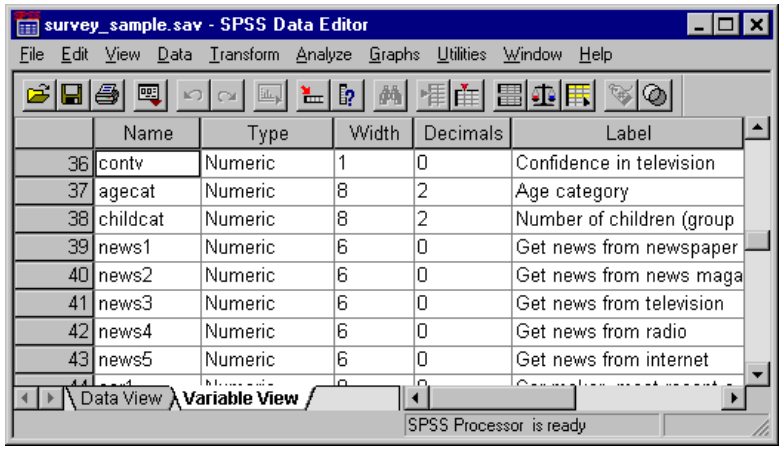


The placeholder values in the table preview reflect the default format for each summary statistic.

- For counts, the default display format is *nnnn*—integer values with no decimal places.
- For percentages, the default display format is *nnnn.n%*—numbers with a single decimal place and a percent sign after the value.
- For the mean, the default display format is *different* for the two variables.

For summary statistics that aren't some form of count (including Valid N and Total N) or percentage, the default display format is the display format defined for the variable in the Data Editor. If you look at the variables in Variable View in the Data Editor, you will see that *Age category* (variable *agecat*) is defined as having two decimal positions, while *Confidence in television* (variable *contv*) is defined as having zero decimal positions.

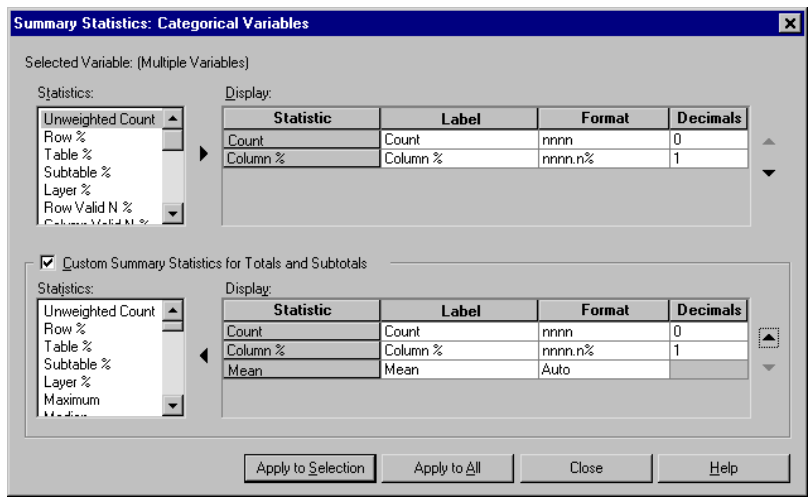
Figure 12-2
Variable View in the Data Editor



This is one of those cases where the default format probably isn't the format you want, since it would probably be better if both mean values displayed the same number of decimals.

- Right-click either variable in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.

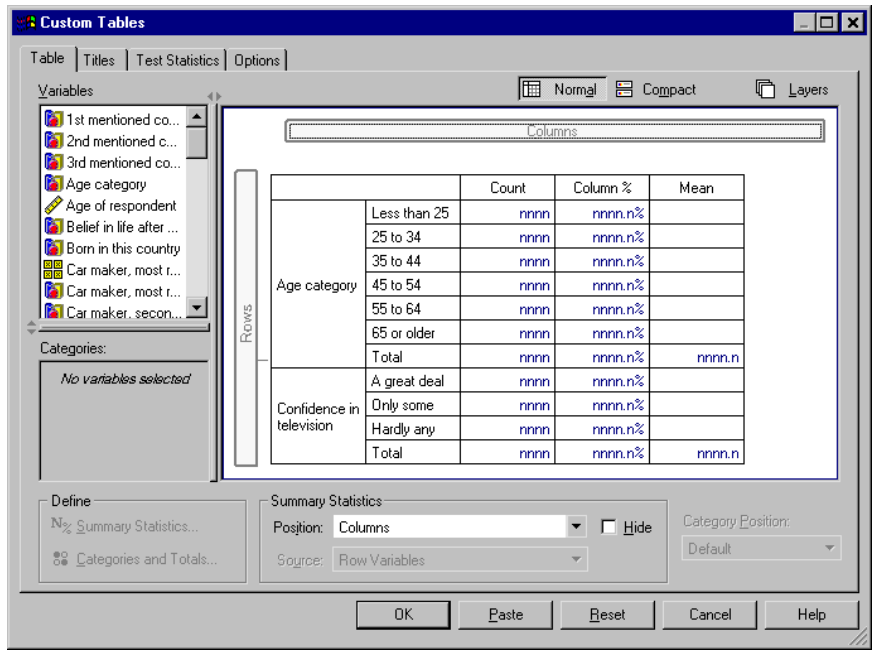
Figure 12-3
Summary Statistics dialog box



For the mean, the Format cell in the Display list indicates that the format is *Auto*, which means that the defined display format for the variable will be used, and the Decimals cell is disabled. In order to specify the number of decimals, you first need to select a different format.

- In the custom summary statistics Display list, click the Format cell for the mean, and select nnnn from the drop-down list of formats.
- In the Decimals cell, enter a value of 1.
- Then click Apply to All to apply this setting to both variables.

Figure 12-4
Table preview with user-specified summary statistics display formats



Now the table preview indicates that both mean values will be displayed with one decimal position. (You could go ahead and create this table now—but you might find the “mean” value for *Age category* a little difficult to interpret, since the actual numeric codes for this variable range only from 1 to 6.)

Display Labels for Summary Statistics

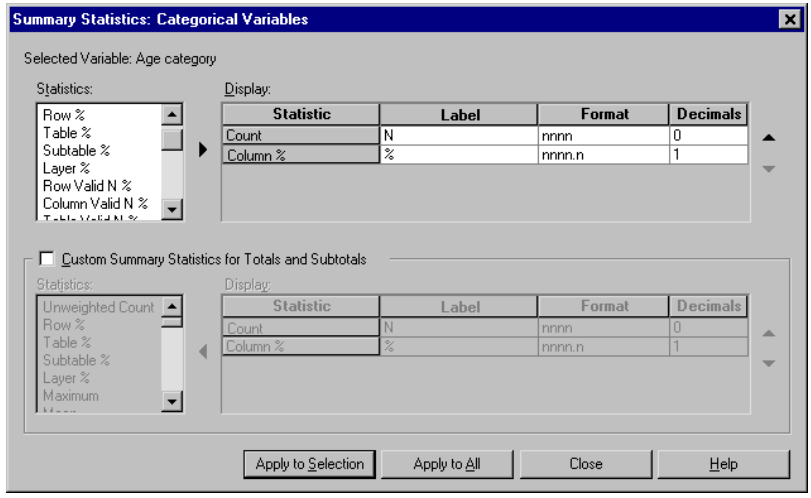
In addition to the display formats for summary statistics, you can also control the descriptive labels for each summary statistic.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click Reset to clear any previous settings in the table builder.
- In the table builder, drag and drop *Age category* from the variable list into the Rows area on the canvas pane.
- Drag and drop *How get paid last week* from the variable list into the Columns area on the canvas pane.
- Right-click *Age category* in the table preview on the canvas pane and select Summary Statistics from the pop-up context menu.
- Select Column % in the Statistics list and click the arrow key to add it to the Display list.
- Double-click anywhere in the word *Column* in the Label cell in the Display list to edit the contents of the cell. Delete the word *Column* from the label, changing the label to simply *%*.
- Edit the Label cell for *Count* in the same way, changing the label to simply *N*.

  While we're here, let's change the format of the Column % statistic to remove the unnecessary percent sign (since the column label indicates that the column contains percentages).

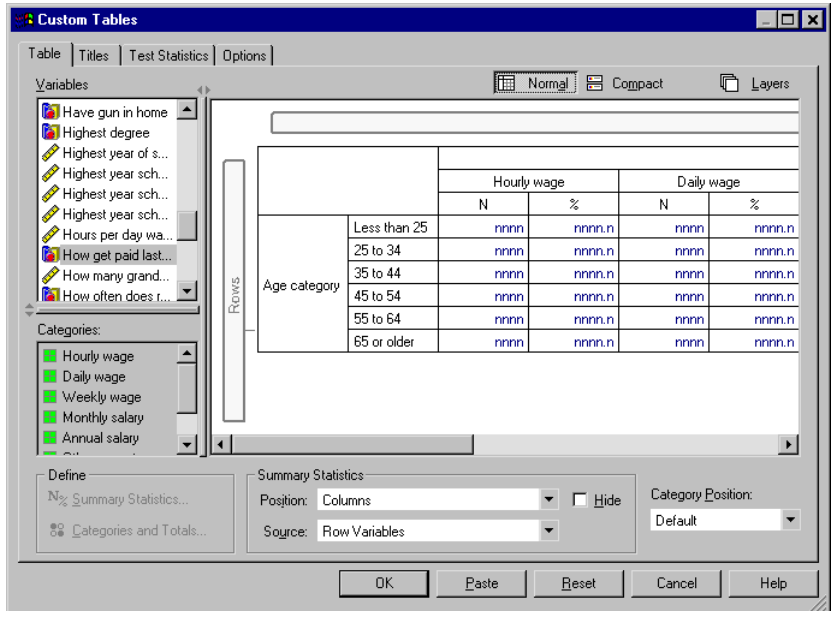
- Click the Format cell for *Column %* and select nnnn.n from the drop-down list of formats.

Figure 12-5
Summary Statistics dialog box with modified labels and formats



▶ Then click Apply to Selection.

Figure 12-6
Table preview with modified summary statistics labels



The table preview displays the modified display format and the modified labels.

- Click OK to create the table.

Figure 12-7
Table with modified summary statistics labels

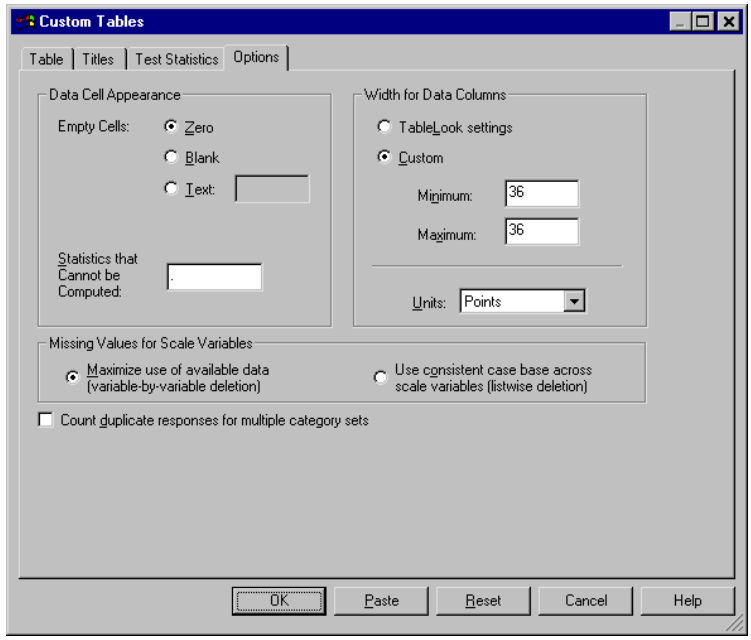
		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	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

Column Width

You may have noticed that the table in the above example is rather wide. One solution to this problem would be to simply swap the row and column variables. Another solution is to make the columns narrower, since they seem to be much wider than necessary. (In fact, the reason we shortened the summary statistics labels was so that we could make the columns narrower.)

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click the Options tab.
- In the Width for Data Columns group, select Custom.
- For the Maximum, type 36. (Make sure that the Units setting is Points.)

Figure 12-8
Custom Tables, Options tab



▶ Click OK to create the table.

Figure 12-9
Table with reduced column widths

		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

Now the table is much more compact.

Display Value for Empty Cells

By default, a 0 is displayed in empty cells (cells that contain no cases). You can instead display nothing in these cells (leave them blank) or specify a text string to display in empty cells.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click the Options tab.
- In the Data Cell Appearance group, for Empty Cells select Text and type None.
- Click OK to create the table.

Figure 12-10
Table with "None" displayed in empty cells

		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

Now the four empty cells in the table display the text *None* instead of a value of 0.

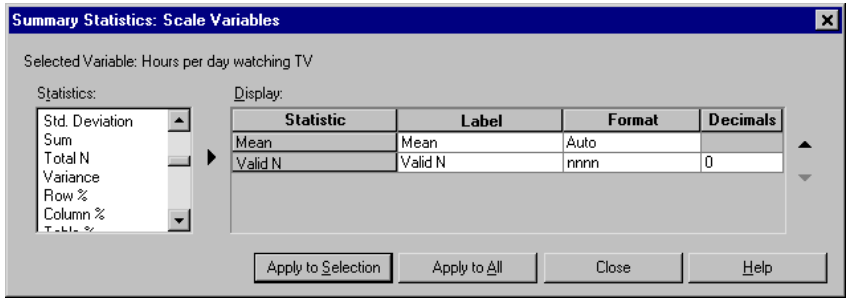
Display Value for Missing Statistics

If a statistic cannot be computed, the default display value is a period (.), which is the symbol used to indicate the system-missing value. This is different from an "empty" cell, and therefore the display value for missing statistics is controlled separately from the display value for cells that contain no cases.

- Open the table builder (Analyze menu, Tables, Custom Tables).
- Drag and drop *Hours per day watching TV* from the variable list to the top of the Columns area on the canvas, above *How get paid last week*.

  Since *Hours per day watching TV* is a scale variable, it automatically becomes the statistics source variable and the summary statistic changes to the mean.
- Right-click *Hours per day watching TV* in the table preview in the canvas pane and select Summary Statistics from the pop-up context menu.
- Select Valid N in the Statistics list and click the arrow key to add it to the Display list.

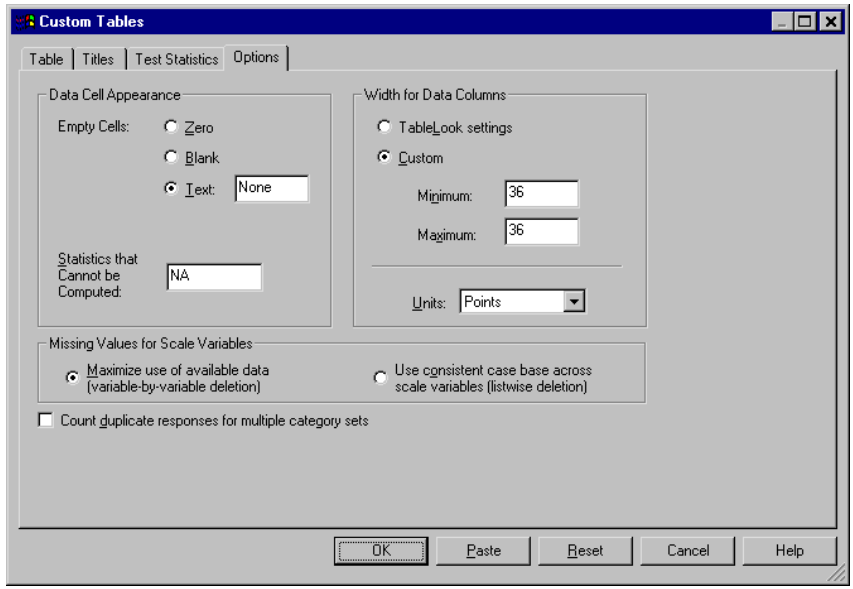
Figure 12-11
Summary Statistics dialog box for scale variables



- Click Apply to Selection.
- Click the Options tab.

- In the text field for Statistics that Cannot be Computed, type NA.

Figure 12-12
Changing the display value for statistics that can't be computed



- Click OK to create the table.

Figure 12-13
Table with "NA" displayed for missing statistics

		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

The text *NA* is displayed for the mean in three cells in the table. In each case, the corresponding *Valid N* value explains why: there are no cases with which to compute the mean.

You may, however, notice what appears to be slight discrepancy—one of those three Valid N values is displayed as a 0, not the label *None* that is supposed to be displayed in cells with no cases. This is because although there are no valid cases to use to compute the mean, the category isn't really empty. If you go back to the original table with just the two categorical variables, you will see that there are, in fact, three cases in this crosstabulated category. There are no valid cases, however, because all three have missing values for the scale variable *Hours per day watching TV*.

Changing the Default TableLook

Many of the display properties of pivot tables can be controlled with TableLooks. A wide variety of predefined TableLooks are available, and you can control the default TableLook that determines the display properties applied to pivot tables when they are created.

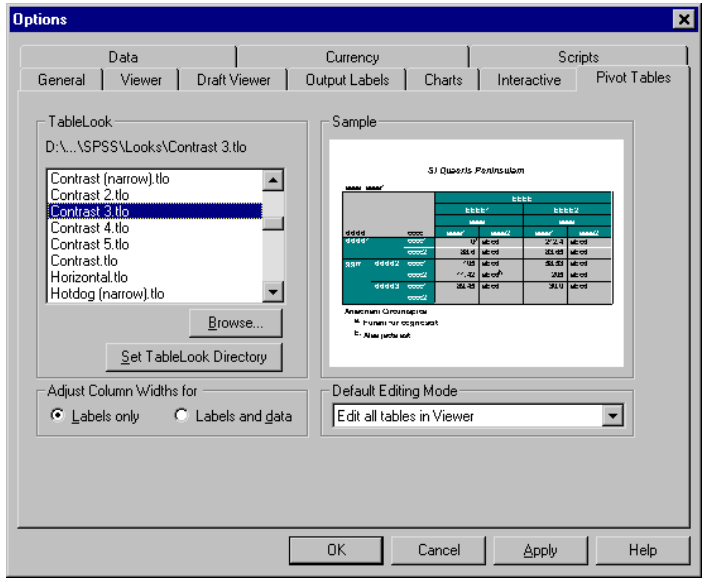
- From the menus, choose:

  Edit
  Options...

- Click the Pivot Tables tab.

- From the list of TableLooks, select *Contrast 3.tlo*.

Figure 12-14
Changing the default TableLook



- Click OK.
- Open the table builder (Analyze menu, Tables, Custom Tables).
- Click OK to create the table.

Figure 12-15
New default TableLook applied to a newly created table

		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

Every table you create will use this TableLook until you specify a different default TableLook. You can also create your own TableLooks and apply different TableLooks to tables you've already created. For more information, use the Index tab in the Help system and type TableLooks as the keyword.

Syntax Reference

CTABLES

Note: Square brackets used in the CTABLES syntax chart are required parts of the syntax and are not used to indicate optional elements. All subcommands except /TABLE are optional.

```
CTABLES

 /FORMAT MINCOLWIDTH={DEFAULT}  MAXCOLWIDTH={DEFAULT}
                     {value  }              {value  }

         UNITS={POINTS}   EMPTY= {ZERO   }   MISSING= {'.'    }
               {INCHES}          {BLANK  }            {'chars'}
               {CM    }          {'chars'}

 /VLABELS VARIABLES= varlist

         DISPLAY= {DEFAULT}
                  {NAME   }
                  {LABEL  }
                  {BOTH   }
                  {NONE   }

 /MRSETS COUNTDUPLICATES= {NO }
                          {YES}

 /SMISSING {VARIABLE}
           {LISTWISE}

 /TABLE  rows BY columns BY layers

 /SLABELS POSITION= {COLUMN}  VISIBLE= {YES}
                    {ROW   }          {NO }
                    {LAYER }

 /CLABELS {AUTO                  }
          {ROWLABELS= {OPPOSITE} }
                     {LAYER   }
          {COLLABELS= {OPPOSITE} }
                     {LAYER   }

 /CATEGORIES  VARIABLES= varlist

    { [value, value, value...]                                  }
    { ORDER= {A} KEY= {VALUE           } MISSING= {EXCLUDE} }
             {D}      {LABEL           }          {INCLUDE}
                      {summary(varname)}

     TOTAL= {NO  }  LABEL= "label" POSITION= {AFTER } EMPTY= {INCLUDE}
            {YES }                           {BEFORE}        {EXCLUDE}
```

Explicit value lists can include `SUBTOTAL='label', MISSING,` and `OTHERNM`.

```
 /TITLES  CAPTION= ['text' 'text'...]
          CORNER=  ['text' 'text'...]
          TITLE=   ['text' 'text'...]
          Text can contain the symbols  )DATE  )TIME  )TABLE

 /SIGTEST TYPE= CHISQUARE  ALPHA= {0.05              }
                                  {significance level}

 /COMPARETEST TYPE= {PROP}  ALPHA= {0.05              }
                    {MEAN}         {significance level}

          ADJUST= {BONFERRONI}  ORIGIN=COLUMN
                  {NONE      }
```

Row, column, and layer elements each have the general form

```
varname {[C]} [summary 'label' format...] {+}  varname ...
        {[S]}                             {>}
```

When nesting (>) and concatenation (+) are combined, as in `a + b > c`, nesting occurs before concatenation; parentheses can be used to change precedence, as in `(a + b) > c`.

Summary functions available for all variables: `COUNT ROWPCT.COUNT COLPCT.COUNT TABLEPCT.COUNT SUBTABLEPCT.COUNT LAYERPCT.COUNT LAYERROWPCT.COUNT LAYERCOLPCT.COUNT ROWPCT.VALIDN COLPCT.VALIDN TABLEPCT.VALIDN SUBTABLEPCT.VALIDN LAYERPCT.VALIDN LAYERROWPCT.VALIDN LAYERCOLPCT.VALIDN ROWPCT.TOTALN COLPCT.TOTALN TABLEPCT.TOTALN SUBTABLEPCT.TOTALN LAYERPCT.TOTALN LAYERROWPCT.TOTALN LAYERCOLPCT.TOTALN`

Summary functions available for scale variables and for totals, and subtotals of numeric variables: `MAXIMUM MEAN MEDIAN MINIMUM MISSING MODE PTILE RANGE SEMEAN STDDEV SUM TOTALN VALIDN VARIANCE ROWPCT.SUM COLPCT.SUM TABLEPCT.SUM SUBTABLEPCT.SUM LAYERPCT.SUM LAYERROWPCT.SUM LAYERCOLPCT.SUM`

Summary functions available for multiple response variables and their totals: `RESPONSES ROWPCT.RESPONSES COLPCT.RESPONSES TABLEPCT.RESPONSES SUBTABLEPCT.RESPONSES LAYERPCT.RESPONSES LAYERROWPCT.RESPONSES LAYERCOLPCT.RESPONSES ROWPCT.RESPONSES.COUNT COLPCT.RESPONSES.COUNT TABLEPCT.RESPONSES.COUNT SUBTABLEPCT.RESPONSES.COUNT LAYERPCT.RESPONSES.COUNT LAYERROWPCT.RESPONSES.COUNT LAYERCOLPCT.RESPONSES.COUNT ROWPCT.COUNT.RESPONSES COLPCT.COUNT.RESPONSES TABLEPCT.COUNT.RESPONSES SUBTABLEPCT.COUNT.RESPONSES LAYERPCT.COUNT.RESPONSES LAYERROWPCT. COUNT.RESPONSES LAYERCOLPCT.COUNT.RESPONSES`

For unweighted summaries, prefix `U` to a function name, as in `UCOUNT`.

Formats for summaries: `COMMAw.d DOLLARw.d Fw.d NEGPARENw.d NEQUALw.d PARENw.d PCTw.d PCTPARENw.d DOTw.d CCA...CCEw.d Nw.d Ew.d and all DATE formats`

Examples

```
CTABLES /TABLE POLVIEWS [COLPCT] BY AGECAT.
```

```
CTABLES  /TABLE $MLTNEWS [COUNT COLPCT] BY SEX
  /SLABELS VISIBLE=NO
  /CATEGORIES VARIABLES=SEX TOTAL=YES.
```

```
CTABLES /TABLE (CONFINAN + CONBUS + CONBUS + CONEDUC
  + CONPRESS + CONMEDIC)[COUNT ROWPCT]
  /CLABELS ROWLABELS=OPPOSITE.
```

Overview

The Custom Tables procedure produces tables in one, two, or three dimensions and provides a great deal of flexibility for organizing and displaying the contents.

- In each dimension (row, column, and layer), you can stack multiple variables to concatenate tables and nest variables to create subtables. See the `TABLE` subcommand.
- You can let Custom Tables determine summary statistics according to the measurement level in the dictionary, or you can assign one or more summaries to specific variables and override the measurement level without altering the dictionary. See the `TABLE` subcommand.

- You can create multiple response sets with the MRSETS command and use them like ordinary categorical variables in a table expression. You can control the percentage base by choosing an appropriate summary function, and you can control with the MRSETS subcommand whether duplicate responses from a single respondent are counted.
- You can assign totals to categorical variables at different nesting levels to create subtable and table totals, and you can assign subtotals across subsets of the values of a variable. See the CATEGORIES subcommand.
- You can determine on a per-variable basis which categories to display in the table, including whether to display missing values and empty categories for which variable labels exist. You can also sort categories by name, label, or the value of a summary function. See the CATEGORIES subcommand.
- You can specify whether to show or hide summary and category labels and where to position the labels. For variable labels, you can specify whether to show labels, names, both, or neither. See the SLABELS, CLABELS, and VLABELS subcommands.
- You can request chi-square tests and pairwise comparisons of column proportions and means. See the SIGTEST and COMPARETEST subcommands.
- You can assign custom titles and captions (see the TITLES subcommand) and control what displays for empty cells and those for which a summary function cannot be computed. See the FORMAT subcommand.
- CTABLES ignores SPLIT FILE requests if layered splits (compare groups in the graphical user interface) are requested. You can compare groups by using the split variables at the highest nesting level for row variables. See the TABLE subcommand for nesting variables.

Syntax Conventions

- The basic specification is a TABLE subcommand with at least one variable in one dimension. Multiple TABLE subcommands can be included in one CTABLES command.
- The global subcommands FORMAT, VLABELS, MRSETS, and SMISSING must precede the first TABLE subcommand and can be named in any order.
- The local subcommands SLABELS, CLABELS, CATEGORIES, TITLES, SIGTEST, and COMPARETEST follow the TABLE subcommand in any order and refer to the immediately preceding table expression.
- In general, if subcommands are repeated, their specifications are merged. The last value of each specified attribute is honored.
- Equals signs shown in the syntax charts are required.
- Square brackets shown in the syntax charts are required.
- All keywords except summary function names, attribute values, and explicit category list keywords can be truncated to as few as three characters. Function names must be spelled in full.
- The slash before all subcommands, including the first, is required.

Example

```
CTABLES /TABLE POLVIEWS [COLPCT] BY AGECAT.
```

		Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
		Column %	Column %	Column %	Column %	Column %	Column %
Think of self as liberal or conservative	Extremely liberal	4.5%	2.5%	2.1%	2.4%	1.3%	2.2%
	Liberal	18.8%	15.7%	14.6%	11.3%	10.5%	9.4%
	Slightly liberal	13.5%	14.2%	13.2%	15.4%	10.5%	10.5%
	Moderate	36.8%	37.1%	32.7%	37.2%	39.3%	38.8%
	Slightly conservative	14.3%	14.9%	19.3%	15.0%	18.4%	13.4%
	Conservative	11.7%	13.0%	14.6%	15.4%	16.4%	21.2%
	Extremely conservative	.4%	2.7%	3.5%	3.3%	3.6%	4.5%

- *POLVIEWS* defines the rows and *AGECAT* defines the columns. Column percentages are requested, overriding the default COUNT function.

Example

```
CTABLES  /TABLE $MLTNEWS [COUNT COLPCT] BY SEX
  /SLABELS VISIBLE=NO
  /CATEGORIES VARIABLES=SEX TOTAL=YES.
```

		Gender					
		Male		Female		Total	
News sources	Get news from internet	359	40.1%	508	42.9%	867	41.7%
	Get news from radio	233	26.0%	318	26.8%	551	26.5%
	Get news from television	451	50.3%	626	52.8%	1077	51.8%
	Get news from news magazines	121	13.5%	173	14.6%	294	14.1%
	Get news from newspapers	375	41.9%	430	36.3%	805	38.7%

- *$MLTNEWS* is a multiple response set.
- The COLPCT function uses the number of respondents as the percentage base, so each cell shows the percentage of males or females who gave each response and the sum of percentage for each column is greater than 100.
- Summary labels are hidden.
- The CATEGORIES subcommand creates a total for both sexes.

Example

```
CTABLES /TABLE (CONFINAN + CONBUS + CONBUS + CONEDUC
  + CONPRESS + CONMEDIC)[COUNT ROWPCT]
  /CLABELS ROWLABELS=OPPOSITE.
```

	A great deal		Only some		Hardly any	
	Count	Row %	Count	Row %	Count	Row %
Confidence in banks & financial institutions	490	26.3%	1068	57.3%	306	16.4%
Confidence in major companies	500	27.5%	1078	59.2%	243	13.3%
Confidence in major companies	500	27.5%	1078	59.2%	243	13.3%
Confidence in education	511	27.2%	1055	56.1%	315	16.7%
Confidence in press	176	9.5%	878	47.2%	808	43.4%
Confidence in medicine	844	45.0%	864	46.1%	167	8.9%

- The six confidence variables all have the same categories with the same value labels for each.
- The CLABELS subcommand moves the category labels to the columns.

TABLE Subcommand

The TABLE subcommand specifies the structure of the table, including the variables and summary functions that define each dimension. It has the general form

```
/TABLE  rows BY columns BY layers
```

The minimum specification for a row, column, or layer is a variable name. You can specify one or more dimensions.

Variable Types

The variables used in a table expression can be category variables, scale variables, or multiple response sets. Multiple response sets are defined by the MRSETS command in the SPSS Base and always begin with a $. Custom Tables uses the measurement level in the dictionary for the active data file to identify category and scale variables. You can override the default variable type for numeric variables by placing [C] or [S] after the variable name. Thus, to treat the category variable *HAPPY* as a scale variable and obtain a mean, you would specify

```
/TABLE HAPPY [S].
```

Category Variables and Multiple Response Sets

Category variables define one cell per value. See the CATEGORIES subcommand for ways of controlling how categories are displayed. Multiple response sets also define one cell per value.

Example:

```
CTABLES /TABLE HAPPY.
```

		Count
General happiness	Very happy	891
	Pretty happy	1575
	Not too happy	340

- The counts for *HAPPY* are in the rows.

Example:

```
CTABLES /TABLE BY HAPPY.
```

General happiness		
Very happy	Pretty happy	Not too happy
Count	Count	Count
891	1575	340

- The counts for *HAPPY* are in the columns.

Example:

```
CTABLES /TABLE BY BY HAPPY
```

General happiness Very happy

Count
891

- The counts for *HAPPY* are in layers.

Stacking and Nesting

Stacking (or concatenating) variables creates multiple logical tables within a single table structure.

Example:

```
CTABLES /TABLE HAPPY + HAPMAR BY CHILDCAT.
```

		Number of children (grouped categories)			
		None	1-2	3-4	5 or more
		Count	Count	Count	Count
General happiness	Very happy	197	412	221	59
	Pretty happy	499	662	314	97
	Not too happy	98	136	79	27
Happiness of marriage	Very happy	111	462	232	49
	Pretty happy	51	238	133	22
	Not too happy	5	18	10	4

- The output contains two tables: one for general happiness by number of children and one for happiness in marriage by number of children. Except for missing values, all of the cases in the data appear in both tables.

Nesting variables creates hierarchical tables.

Example:

```
CTABLES /TABLE SEX > HAPMAR BY CHILDCAT.
```

				Number of children (grouped categories)			
				None	1-2	3-4	5 or more
				Count	Count	Count	Count
Gender	Male	Happiness of marriage	Very happy	48	216	102	30
			Pretty happy	25	110	58	11
			Not too happy	3	7	4	1
	Female	Happiness of marriage	Very happy	63	246	130	19
			Pretty happy	26	128	75	11
			Not too happy	2	11	6	3

- The output contains one table with a subtable for each value of *SEX*. The same subtables would result from the table expression HAPMAR BY CHILDCAT BY SEX, but the subtables would appear in separate layers.

Stacking and nesting can be combined. When they are, by default, nesting takes precedence over stacking. You can use parentheses to alter the order of operations.

Example:

```
CTABLES /TABLE (HAPPY + HAPMAR) > SEX.
```

				Count
General happiness	Very happy	Gender	Male	373
			Female	518
	Pretty happy	Gender	Male	712
			Female	863
	Not too happy	Gender	Male	133
			Female	207
Happiness of marriage	Very happy	Gender	Male	396
			Female	459
	Pretty happy	Gender	Male	205
			Female	240
	Not too happy	Gender	Male	15
			Female	22

- The output contains two tables. Without the parentheses, the first table, for general happiness, would not have separate rows for male and female.

Scale Variables

Scale variables, such as age in years or population of towns, do not define multiple cells within a table. The table expression /TABLE AGE creates a table with one cell containing the mean of *AGE* across all cases in the data. You can use nesting and/or dimensions to display summary statistics for scale variables within categories. The nature of scale variables prevents their being arranged hierarchically. Therefore:

- A scale variable cannot be nested under another scale variable.
- Scale variables can be used in only one dimension.

Example:

```
CTABLES /TABLE AGE > HAPPY BY SEX.
```

			Gender	
			Male	Female
			Mean	Mean
Age of respondent	General happiness	Very happy	47	47
		Pretty happy	44	45
		Not too happy	43	47

Specifying Summaries

You can specify one or more summary functions for variables in any one dimension. For category variables, summaries can be specified only for the variables at the lowest nesting level. Thus, in the table expression

```
/TABLE SEX > (HAPPY + HAPMAR) BY AGECAT
```

you can assign summaries to *HAPPY* and *HAPMAR* or to *AGECAT*, but not to both and not to *SEX*.

If a scale variable appears in a dimension, that becomes the statistics dimension, and all statistics must be specified for that dimension. A scale variable need not be at the lowest level of nesting. Thus, the following is a valid specification:

```
CTABLES /TABLE AGE [MINIMUM, MAXIMUM, MEAN] > SEX > HAPPY.
```

A multiple response variable also need not be at the lowest level of nesting. The following is a valid specification:

```
CTABLES /TABLE $MLTCARS [COUNT, RESPONSES] > SEX.
```

However, if two multiple response variables are nested, as in $MULTCARS > $MULTNEWS, summaries can be requested only for the one at the innermost nesting level (in this case, *$MULTNEWS*).

The general form for a summary specification is

```
[summary 'label' format, ..., summary 'label' format]
```

- The specification follows the variable name in the table expression. You can apply a summary specification to multiple variables by enclosing them in parentheses. The following

specifications are equivalent:

```
/TABLE SEX [COUNT] + HAPPY [COUNT, COLPCT]
/TABLE (SEX + HAPPY [COLPCT])[COUNT]
```

- The brackets are required even if only one summary is specified.
- Commas are optional.
- Label and format are both optional; defaults are used if they are not specified.
- If totals or subtotals are defined for a variable (on the CATEGORIES subcommand), by default, the same functions specified for the variable are used for the totals. You can use the keyword TOTALS within the summary specification to specify different summary functions for the totals and subtotals. The specification then has the form `[summary 'label' format ... TOTALS [summary 'label' format...]]`. You must still specify TOTAL=YES on the CATEGORIES subcommand to see the totals.
- Summaries that are available for category variables are also available for scale variables and multiple response sets. Functions specific to scale variables and to multiple response sets are also available.
- If case weighting is in effect, summaries are calculated taking into account the current WEIGHT value. To obtain unweighted summaries, prefix a U to the function name, as in UCOUNT. Unweighted functions are not available where weighting would not apply, as in the MINIMUM and MAXIMUM functions.

Example:

```
CTABLES /TABLE SEX > HAPMAR [COLPCT] BY CHILDCAT.
```

				Number of children (grouped categories)			
				None	1-2	3-4	5 or more
				Column %	Column %	Column %	Column %
Gender	Male	Happiness of marriage	Very happy	63.2%	64.9%	62.2%	71.4%
			Pretty happy	32.9%	33.0%	35.4%	26.2%
			Not too happy	3.9%	2.1%	2.4%	2.4%
	Female	Happiness of marriage	Very happy	69.2%	63.9%	61.6%	57.6%
			Pretty happy	28.6%	33.2%	35.5%	33.3%
			Not too happy	2.2%	2.9%	2.8%	9.1%

Example:

```
CTABLES /TABLE AGECAT > TVHOURS [MEAN F5.2,
   STDDEV 'Standard Deviation' F5.2, PTILE 90 '90th Percentile'].
```

			Mean	Standard Deviation	90th Percentile
Age category	Less than 25	Hours per day watching TV	2.85	2.03	5
	25 to 34	Hours per day watching TV	2.78	2.37	5
	35 to 44	Hours per day watching TV	2.56	2.11	5
	45 to 54	Hours per day watching TV	2.58	1.97	5
	55 to 64	Hours per day watching TV	3.02	2.22	6
	65 or older	Hours per day watching TV	3.58	2.50	6

- Each summary function for the row variable appears by default in a column.
- Labels for standard deviation and the 90th percentile override the defaults.
- Because *TVHOURS* is recorded in whole hours and has an integer print format, the default general print formats for mean and standard deviation would also be integer, so overrides are specified.

Table 1 Summary Functions: All Variables

Function	Description	Default Label*	Default Format
COUNT	Number of cases in each category. This is the default for categorical and multiple response variables.	Count	Count
ROWPCT.COUNT	Row percentage based on cell counts. Computed within subtable.	Row %	Percent
COLPCT.COUNT	Column percentage based on cell counts. Computed within subtable.	Column %	Percent
TABLEPCT.COUNT	Table percentage based on cell counts.	Table %	Percent
SUBTABLEPCT.COUNT	Subtable percentage based on cell counts.	Subtable %	Percent
LAYERPCT.COUNT	Layer percentage based on cell counts. Same as table percentage if no layers are defined.	Layer %	Percent
LAYERROWPCT.COUNT	Row percentage based on cell counts. Percentages sum to 100% across the entire row (that is, across subtables).	Layer Row %	Percent
LAYERCOLPCT.COUNT	Column percentage based on cell counts. Percentages sum to 100% across the entire column (that is, across subtables).	Layer Column %	Percent
ROWPCT.VALIDN	Row percentage based on valid count.	Row Valid N %	Percent
COLPCT.VALIDN	Column percentage based on valid count.	Column Valid N %	Percent
TABLEPCT.VALIDN	Table percentage based on valid count.	Table Valid N %	Percent
SUBTABLEPCT.VALIDN	Subtable percentage based on valid count.	Subtable Valid N %	Percent

Table 1 Summary Functions: All Variables (Continued)

Function	Description	Default Label	Default Format
LAYERPCT.VALIDN	Layer percentage based on valid count.	Layer Valid N %	Percent
LAYERROWPCT. VALIDN	Row percentage based on valid count. Percentages sum to 100% across the entire row.	Layer Row Valid N %	Percent
LAYERCOLPCT. VALIDN	Column percentage based on valid count. Percentages sum to 100% across the entire column.	Layer Column Valid N %	Percent
ROWPCT.TOTALN	Row percentage based on total count, including user- and system-missing values.	Row Total N %	Percent
COLPCT.TOTALN	Column percentage based on total count, including user- and system-missing values.	Column Total N %	Percent
TABLEPCT.TOTALN	Table percentage based on total count, including user- and system-missing values.	Table Total N %	Percent
SUBTABLEPCT.TOTALN	Subtable percentage based on total count, including user- and system-missing values.	Subtable Total N %	Percent
LAYERPCT.TOTALN	Layer percentage based on total count, including user- and system-missing values.	Layer Total N %	Percent
LAYERROWPCT. TOTALN	Row percentage based on total count, including user- and system-missing values. Percentages sum to 100% across the entire row.	Layer Row Total N %	Percent
LAYERCOLPCT. TOTALN	Column percentage based on total count, including user- and system-missing values. Percentages sum to 100% across the entire column.	Layer Column Total N %	Percent

* This is the default on a U.S.-English system.

The .COUNT suffix can be omitted from percentages based on cell counts. Thus, ROWPCT is equivalent to ROWPCT.COUNT.

Table 2 Summary Functions: Scale Variables, Totals, and Subtotals

Function	Description	Default Label	Default Format
MAXIMUM	Largest value.	Maximum	General
MEAN	Arithmetic mean. The default for scale variables.	Mean	General
MEDIAN	50^{th} percentile.	Median	General
MINIMUM	Smallest value.	Minimum	General
MISSING	Count of missing values (both user- and system-missing).	Missing	General
MODE	Most frequent value. If there is a tie, the smallest value is shown.	Mode	General
PTILE	Percentile. Takes a numeric value between 0 and 100 as a required parameter. PTILE is computed the same way as APTILE in SPSS Tables. Note that in SPSS Tables, the default percentile method was HPTILE.	Percentile ####.##	General
RANGE	Difference between maximum and minimum values.	Range	General
SEMEAN	Standard error of the mean.	Std Error of Mean	General
STDDEV	Standard deviation.	Std Deviation	General
SUM	Sum of values.	Sum	General
TOTALN	Count of nonmissing, user-missing, and system-missing values. The count excludes valid values hidden via the CATEGORIES subcommand.	Total N	Count
VALIDN	Count of nonmissing values.	Valid N	Count
VARIANCE	Variance.	Variance	General
ROWPCT.SUM	Row percentage based on sums.	Row Sum %	Percent
COLPCT.SUM	Column percentage based on sums.	Column Sum %	Percent
TABLEPCT.SUM	Table percentage based on sums.	Table Sum %	Percent
SUBTABLEPCT.SUM	Subtable percentage based on sums.	Subtable Sum %	Percent
LAYERPCT.SUM	Layer percentage based on sums.	Layer Sum %	Percent

Table 2 Summary Functions: Scale Variables, Totals, and Subtotals (Continued)

Function	Description	Default Label	Default Format
LAYERROWPCT. SUM	Row percentage based on sums. Percentages sum to 100% across the entire row.	Layer Row Sum %	Percent
LAYERCOLPCT. SUM	Column percentage based on sums. Percentages sum to 100% across the entire column.	Layer Column Sum %	Percent

Table 3 Summary Functions: Multiple Response Sets

Function	Description	Default Label	Default Format
RESPONSES	Count of responses.	Responses	Count
ROWPCT.RESPONSES	Row percentage based on responses. Total number of responses is the denominator.	Row Responses %	Percent
COLPCT.RESPONSES	Column percentage based on responses. Total number of responses is the denominator.	Column Responses %	Percent
TABLEPCT.RESPONSES	Table percentage based on responses. Total number of responses is the denominator.	Table Responses %	Percent
SUBTABLEPCT.RESPONSES	Subtable percentage based on responses. Total number of responses is the denominator.	Subtable Responses %	Percent
LAYERPCT.RESPONSES	Layer percentage based on responses. Total number of responses is the denominator.	Layer Responses %	Percent
LAYERROWPCT.RESPONSES	Row percentage based on responses. Total number of responses is the denominator. Percentages sum to 100% across the entire row (that is, across subtables).	Layer Row Responses %	Percent
LAYERCOLPCT. RESPONSES	Column percentage based on responses. Total number of responses is the denominator. Percentages sum to 100% across the entire column (that is, across subtables).	Layer Column Responses %	Percent

Table 3 Summary Functions: Multiple Response Sets (Continued)

Function	Description	Default Label	Default Format
ROWPCT.RESPONSES.COUNT	Row percentage: responses are the numerator and total count is the denominator.	Row Responses % (Base: Count)	Percent
COLPCT.RESPONSES.COUNT	Column percentage: responses are the numerator and total count is the denominator.	Column Responses % (Base: Count)	Percent
TABLEPCT.RESPONSES. COUNT	Table percentage: responses are the numerator and total count is the denominator.	Table Responses % (Base: Count)	Percent
SUBTABLEPCT.RESPONSES. COUNT	Subtable percentage: responses are the numerator and total count is the denominator.	Subtable Responses % (Base: Count)	Percent
LAYERPCT. RESPONSES.COUNT	Layer percentage: responses are the numerator and total count is the denominator.	Layer Responses % (Base: Count)	Percent
LAYERROWPCT.RESPONSES. COUNT	Row percentage: responses are the numerator and total count is the denominator. Percentages sum to 100% across the entire row (that is, across subtables).	Layer Row Responses % (Base: Count)	Percent
LAYERCOLPCT.RESPONSES. COUNT	Column percentage: responses are the numerator and total count is the denominator. Percentages sum to 100% across the entire column (that is, across subtables).	Layer Column Responses % (Base: Count)	Percent
ROWPCT.COUNT.RESPONSES	Row percentage: count is the numerator and total responses are the denominator.	Row Count % (Base: Responses)	Percent
COLPCT.COUNT.RESPONSES	Column percentage: count is the numerator and total responses are the denominator.	Column Count % (Base: Responses)	Percent
TABLEPCT.COUNT. RESPONSES	Table percentage: count is the numerator and total responses are the denominator.	Table Count % (Base: Responses)	Percent

Table 3 Summary Functions: Multiple Response Sets (Continued)

Function	Description	Default Label	Default Format
SUBTABLEPCT.COUNT. RESPONSES	Subtable percentage: count is the numerator and total responses are the denominator.	Subtable Count % (Base: Responses)	Percent
LAYERPCT.COUNT. RESPONSES	Layer percentage: count is the numerator and total responses are the denominator.	Layer Count % (Base: Responses)	Percent
LAYERROWPCT.COUNT. RESPONSES	Row percentage: count is the numerator and total responses are the denominator. Percentages sum to 100% across the entire row (that is, across subtables).	Layer Row Count % (Base: Responses)	Percent
LAYERCOLPCT.COUNT. RESPONSES	Row percentage: count is the numerator and total responses are the denominator. Percentages sum to 100% across the entire column (that is, across subtables).	Layer Column Count % (Base: Responses)	Percent

Formats for Summaries

A default format is assigned to each summary function:

Count The value is expressed in F (standard numeric) format with 0 decimal places. If you have fractional weights and want a count that reflects those weights, use F format with appropriate decimal places.

Percent The value is expressed with one decimal place and a percent symbol.

General The value is expressed in the variable's print format.

These default formats are internal to CTABLES and cannot be used in TABLE expressions. To override the default formats, use any of the print formats available in the SPSS Base except Z, PBHEX, and HEX, or the additional formats described in Table 4.

Table 4 Additional Formats for Summaries

Format	Description	Example
NEGPARENw.d	Parentheses appear around negative numbers.	–1234.567 formatted as NEGPAREN9.2 yields (1234.57).
NEQUALw.d	"N=" precedes the number.	1234.567 formatted as NEQUAL9.2 yields N=1234.57.
PARENw.d	The number is parenthesized.	1234.567 formatted as PAREN8.2 yields (1234.57).
PCTPARENw.d	A percent symbol follows the value, which is parenthesized.	1234.567 formatted as PCTPAREN10.2 yields (1234.57%).

Missing Values in Summaries

Table 5 presents the rules for including cases in a table for VALIDN, COUNT, and TOTALN functions when values are included or excluded explicitly through an explicit category list or implicitly through inclusion or exclusion of user-missing values.

Table 5 Inclusion/Exclusion of Values in Summaries

Variable and Value Type	VALIDN	COUNT	TOTALN
Categorical Variable: shown valid value Multiple Dichotomy Set: at least one "true" value Multiple Category Set: at least one shown valid value Scale Variable: valid value	Include	Include	Include
Categorical Variable: included user-missing value Multiple Category Set: all values are included user-missing Scale Variable: user-missing or system-missing	Exclude	Include	Include
Categorical Variable: excluded user-missing or system-missing Multiple Dichotomy Set: all values are "false" Multiple Category Set: all values are excluded user-missing, system-missing, or excluded valid, but at least one value is not excluded valid	Exclude	Exclude	Include
Categorical Variable: excluded valid value Multiple Dichotomy Set: all values are excluded valid values	Exclude	Exclude	Exclude

SLABELS Subcommand

The SLABELS subcommand controls the position of summary statistics in the table and whether summary labels are shown.

```
/SLABELS POSITION= {COLUMN}  VISIBLE= {YES}
                   {ROW   }          {NO }
                   {LAYER }
```

By default, summaries appear in the columns and labels are visible.

Example: Summary Label Positioning

```
CTABLES /TABLE NEWS [COUNT COLPCT].
```

		Count	Column %
How often does respondent read newspaper	Every day	805	43.0%
	Few times a week	420	22.5%
	Once a week	294	15.7%
	Less than once a week	202	10.8%
	Never	149	8.0%

```
CTABLES /TABLE NEWS [COUNT COLPCT]
  /SLABELS POSITION=ROW VISIBLE=NO.
```

How often does respondent read newspaper	Every day	805 43.0%
	Few times a week	420 22.5%
	Once a week	294 15.7%
	Less than once a week	202 10.8%
	Never	149 8.0%

CLABELS Subcommand

The CLABELS subcommand controls the location of category labels.

```
/CLABELS {AUTO                    }
         {ROWLABELS= {OPPOSITE} }
                     {LAYER   }
         {COLLABELS= {OPPOSITE} }
                     {LAYER   }
```

By default, category labels are nested under the variables to which they belong. Category labels for row and column variables can be moved to the opposite dimension or to the layers. If labels exist in both dimensions, only one dimension, row labels or column labels, can be moved; they cannot be swapped.

Example:

```
CTABLES
  /TABLE (CONFINAN + CONEDUC + CONBUS + CONMEDIC + CONPRESS + CONTV )
```

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

- Six variables are stacked in the rows, and their category labels are stacked under them.

```
CTABLES
  /TABLE (CONFINAN + CONEDUC + CONBUS + CONMEDIC + CONPRESS + CONTV )
  /SLABELS VISIBLE=NO   /CLABELS ROWLABELS=OPPOSITE
```

	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

- The category labels are moved to the columns. Where variables are stacked, as in this example, the value labels for all of them must be exactly the same to allow for this format. Additionally, all must have the same category specifications, and data-dependent sorting is not allowed.

CATEGORIES Subcommand

The CATEGORIES subcommand controls the order of categories in the rows and columns of the table, the showing and hiding of ordinary and user-missing values, and the computation of totals and subtotals.

```
/CATEGORIES  VARIABLES= varlist

    { [value, value, value...]                                  }
    { ORDER= {A} KEY= {VALUE           } MISSING= {EXCLUDE} }
             {D}      {LABEL           }          {INCLUDE}
                      {summary(varname)}

     TOTAL= {NO  }  LABEL= "label" POSITION= {AFTER } EMPTY= {INCLUDE}
            {YES }                           {BEFORE}        {EXCLUDE}
```

The minimum specification is a variable list and one of the following: a category specification, TOTAL specification, or EMPTY specification. The variable list can be a list of variables or the keyword ALL, which refers to all category variables in the table expression. ALL cannot be used with the explicit category list.

Explicit Category Specification

The explicit category specification is a bracketed list of data values or value ranges in the order in which they are to be displayed in the table. Values not included in the list are excluded from the table. This form allows for subtotals and showing or hiding of specific values (both ordinary and user-missing).

- The list can include both ordinary and user-missing values but not the system-missing value (.).
- Values are optionally separated by commas.
- String and date values must be quoted. Date values must be consistent with the variable's print format.
- The LO, THRU, and HI keywords can be used in the value list to refer to a range of categories. LO and HI can be used only as part of a range specification.
- The MISSING keyword can be used to refer to all user-missing values.
- The OTHERNM keyword can be used to refer to all nonmissing values not explicitly named in the list. It can go anywhere within the list. The values to which it refers appear in ascending order.
- If a value is repeated in the list, the last instance is honored. Thus, for a variable *RATING* with integer values 1 through 5, the following specifications are equal:

```
/CATEGORIES VARIABLES = RATING [1,2,4,5,3]
/CATEGORIES VARIABLES = RATING [1 THRU 5,3]
/CATEGORIES VARIABLES = RATING [OTHERNM,3]
```

- For a multiple dichotomy set, you can order the variables in the set by using the names of the variables in the set. The variable names are not enclosed in quotes.

- The SUBTOTAL keyword is used within a category list to request subtotals for a variable. The position of a subtotal within the list determines where it will appear in the table and the categories to which it applies. By default, a subtotal applies to all values that precede it up to the next subtotal. If POSITION=BEFORE is specified (see Totals on p. 225), subtotals apply to the categories that follow them in the list. Hierarchical and overlapping subtotals are not supported. You can specify a label for a subtotal by placing it in quotes immediately following the SUBTOTAL keyword and an equals sign, as illustrated in the following example.

Example:

```
CTABLES  /TABLE AGECAT
   /CATEGORIES VARIABLES=AGECAT [1, 2, 3, SUBTOTAL='Subtotal < 45',
   4, 5, 6, SUBTOTAL='Subtotal 45+'].
```

		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

Implicit Category Specification

The implicit list allows you to sort the categories and to show or hide user-missing values without having to enumerate the values. It also provides for data-dependent sorting. If you do not supply an explicit value list, you can use the following keywords:

ORDER *The sorting order*. You can select A (the default) for ascending order, or D for descending order.

KEY *The sort key.* You can select VALUE (the default) to sort by the values, or LABEL to sort by the value labels. When values are sorted by label, any unlabeled values appear after the labeled values in the table. You can also specify a summary function for data-dependent sorting.

MISSING *Whether user-missing values are included*. You can specify EXCLUDE (the default) or INCLUDE. System-missing values are never included.

Data-Dependent Sorting. The following conventions and limitations apply to sorting using a summary function as the key:

- The sort function must be a summary function supported in CTABLES. The PTILE, MODE, and MEDIAN functions cannot be used.
- The sort function must be used in the table. The exception to this is COUNT. You can sort by COUNT even if counts do not appear in the table.
- Data-dependent sorting is not available if category labels are repositioned using the CLABELS subcommand.

- Summary functions available only for scale variables require that you give the variable name in parentheses, as in MEAN(AGE). Other functions, such as COUNT, do not require a variable name, but you can supply one to restrict the sort.
- When a variable name is given and multiple logical tables are created through stacking, the entire table is sorted based on the first logical table that includes the categorical variable being sorted and the variable specified in the key.
- When a table contains more than one dimension, the sort is based on the distribution of the key within the categories of the sorted variable without regard to the contents of the other dimensions. Thus, given the table
  `CTABLES /TABLE A BY B + C /CAT VAR=A ORDER=A KEY=COUNT(A),`
  the rows are sorted according to the counts for the categories of *A* without regard to the values of *B* and *C*. If there are no missing values in the other dimension, the result is the same as sorting on the totals for that dimension, in this case *B* or *C*. If the other dimension has an unbalanced pattern of missing values, the sorting may give unexpected results; however, the result is unaffected by differences in the pattern for *B* and *C*.
- If the sort variable is crossed with stacked category variables, the first table in the stack determines the sort order.
- To ensure that the categories are sorted the same way in each layer of the pivot table, layer variables are ignored for the purpose of sorting.

Example:

```
CTABLES
  /TABLE CAR1 BY AGECAT
  /CATEGORIES VARIABLES=AGECAT TOTAL=YES
  /CATEGORIES VARIABLES=CAR1 ORDER=D KEY=COUNT.
```

		Age category						
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Total
		Count	Count	Count	Count	Count	Count	Count
Car maker, most recent car	American	99	267	293	214	140	215	1228
	Japanese	73	136	140	107	66	104	626
	German	18	91	69	63	36	61	338
	Korean	23	77	88	45	35	50	318
	Swedish	18	32	46	20	24	25	165
	Other	11	24	43	32	19	24	153

- The first CATEGORIES subcommand requests a total across all age categories.
- The second CATEGORIES subcommand requests a sort of the categories of *CAR1* in descending order using COUNT as the key. The categories of *CAR1* are sorted according to the total counts.

Example:

```
CTABLES
  /TABLE AGE [MEAN F5.1] > CAR1 BY SEX
  /CATEGORIES VARIABLES=SEX TOTAL=YES
  /CATEGORIES VARIABLES=CAR1 KEY=MEAN(AGE).
```

			Gender		
			Male	Female	Total
			Mean	Mean	Mean
Age of respondent	Car maker, most recent car	Swedish	42.6	45.6	44.3
		Japanese	43.5	45.5	44.7
		Korean	43.4	46.2	45.0
		American	45.3	46.5	45.9
		German	44.3	47.6	46.2
		Other	48.6	46.4	47.3

- The first CATEGORIES subcommand requests a total across the values of *SEX*.
- The second CATEGORIES subcommand requests that the categories of *CAR1* be sorted according to the mean of *AGE*. The categories are sorted according to the total means for both sexes, and that would be the case if the totals were not shown in the table.

Totals

A total can be specified for any category variable regardless of its level of nesting within a dimension. Totals can be requested in more than one dimension. The following options are available:

TOTAL *Whether to display a total for a variable.* You can specify TOTAL=NO (the default) or TOTAL=YES.

LABEL *The label for the total.* The specification is a quoted string.

POSITION *Whether a total comes after or before the categories of the variable being totaled.* You can specify AFTER (the default) or BEFORE. POSITION also determines whether subtotals specified in an explicit list of categories apply to the categories that precede them (AFTER) or follow them (BEFORE).

Scale variables cannot be totaled directly. To obtain a total or subtotals for a scale variable, request the total or subtotals for the category variable within whose categories the summaries for the scale variable appear.

Example:

```
CTABLES /TABLE AGECAT
  /CATEGORIES VARIABLES=AGECAT TOTAL=YES LABEL='Total Respondents'.
```

		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
	Total Respondents	2828

Example:

```
CTABLES /TABLE AGE [MEAN 'Average' F5.1] > SEX
  /CATEGORIES VARIABLES=SEX TOTAL=YES LABEL='Combined'.
```

			Average
Age of respondent	Gender	Male	44.6
		Female	46.3
		Combined	45.6

- The summary function for *AGE* appears in cells determined by the values of *SEX*. The total is requested for *SEX* to obtain the average age across both sexes.

Empty Categories

Empty categories are those for which no cases appear in the data. For an explicit category list, this includes all explicitly named values and all labeled values implied by THRU, OTHERNM, or MISSING. For an implicit category list, this includes all values for which value labels exist.

EMPTY *Whether to show categories whose count is zero.* You can specify EMPTY=INCLUDE (the default) or EMPTY=EXCLUDE.

TITLES Subcommand: Titles, Captions, and Corner Text

The TITLES subcommand specifies table annotations. If the subcommand is used, a title, caption, or corner text must be specified. No caption, title, or corner text is displayed by default.

```
/TITLES  CAPTION= ['text' 'text'...]
         CORNER=  ['text' 'text'...]
         TITLE=   ['text' 'text'...]
```

CAPTION *Caption lines.* The caption appears below the table. Multiple lines can be specified. Each line must be quoted.

CORNER *Corner text.* Corner text appears in the corner cell of the table, above row titles and next to column titles. Multiple lines can be specified. Each line must be quoted.

Pivot tables show all corner text that fits in the corner cell. The specified text is ignored if the table has no corner cell.

The system default TableLook uses the corner area for display of row dimension labels. To display CTABLES corner text, the Row Dimension Labels setting in Table Properties should be set to Nested. This choice can be preset in the default TableLook.

TITLE *Title text.* The title appears above the table. Multiple lines can be specified. Each line must be quoted.

The following symbols can be used within any caption, corner text, or title line. Each must be specified using an opening right parenthesis and all uppercase letters.

)DATE *Current date.* Displays a locale-appropriate date stamp that includes the year, month, and day.

)TIME *Current time.* Displays a locale-appropriate time stamp.

)TABLE *Table description.* Inserts a description of the table, which consists of the table expression stripped of measurement levels, statistics specifications, and "/TABLE." If variable labels are available, they are used instead of variable names in the table expression.

Example:

```
CTABLES /VLABELS VARIABLES=SEX HAPMAR DISPLAY=NONE
  /TABLE SEX > HAPMAR BY CHILDCAT [COLPCT]
  /SLABELS VISIBLE=NO
  /TITLE TITLE = 'Marital Happiness for Men and Women '+
              'by Number of Children'
         CAPTION= 'Report created at )TIME on )DATE' ')TABLE'.
```

Marital Happiness for Men and Women by Number of Children

		Number of children (grouped categories)			
		None	1-2	3-4	5 or more
Male	Very happy	63.2%	64.9%	62.2%	71.4%
	Pretty happy	32.9%	33.0%	35.4%	26.2%
	Not too happy	3.9%	2.1%	2.4%	2.4%
Female	Very happy	69.2%	63.9%	61.6%	57.6%
	Pretty happy	28.6%	33.2%	35.5%	33.3%
	Not too happy	2.2%	2.9%	2.8%	9.1%

Report created at 08:33:53 AM on 08/26/2002
Gender > Happiness of marriage BY Number of children (grouped categories)

- The VLABELS subcommand suppresses the display of variable labels for *SEX* and *HAPMAR*.
- The SLABELS subcommand suppresses the default label for the summary function.

- The TITLE specification on the TITLE subcommand uses the standard SPSS convention to break a single string across input lines.
- The CAPTION specification uses the)DATE,)TIME, and)TABLE keywords to print the date, time, and a description of the table structure.

Significance Testing

Custom Tables can perform the chi-square test of independence and pairwise comparisons of column proportions for tables that contain at least one category variable in both the rows and the columns, and pairwise comparisons of column means for tables that contain at least one summary variable in the rows and one category variable in the columns.

Chi-Square Tests: SIGTEST Subcommand

```
/SIGTEST TYPE= CHISQUARE  ALPHA= {0.05             }
                                 {significance level}
```

The SIGTEST subcommand has the following specifications:

TYPE T*he type of significance test*. The specification is required. The only current choice is CHISQUARE.

ALPHA *The significance level for the test*. The specification must be greater than 0 and less than 1. The default is 0.05.

Example:

```
CTABLES /TABLE AGECAT BY MARITAL
  /CATEGORIES VARIABLES=AGECAT MARITAL TOTAL=YES
  /SIGTEST TYPE=CHISQUARE.
```

		Marital status					
		Married	Widowed	Divorced	Separated	Never married	Total
		Count	Count	Count	Count	Count	Count
Age category	Less than 25	37	1	5	5	194	242
	25 to 34	271	13	63	16	263	626
	35 to 44	379	11	129	44	116	679
	45 to 54	275	18	123	13	52	481
	55 to 64	186	31	76	7	20	320
	65 or older	197	209	48	8	17	479
	Total	1345	283	444	93	662	2827

Pearson Chi-Square Tests

		Marital status
Age category	Chi-square	1473.381
	df	20
	Sig.	.000*

Results are based on nonempty rows and columns in each innermost subtable.

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

Pairwise Comparisons of Proportions and Means: COMPARETEST Subcommand

```
/COMPARETEST TYPE= {PROP}  ALPHA= {0.05              }
                  {MEAN}         {significance level}

         ADJUST= {BONFERRONI}  ORIGIN=COLUMN
                 {NONE      }
```

The SIGTEST subcommand has the following specifications:

TYPE T*he type of pairwise comparison*. The specification is required. To compare proportions when the test variable in the rows is categorical, choose PROP. To compare means when the test variable in the rows is scale, choose MEAN.

ALPHA *The significance level for the test*. The specification must be greater than 0 and less than 1. The default is 0.05.

ADJUST *The method for adjusting* p *values for multiple comparisons*. Valid options are NONE and BONFERRONI. If ADJUST is not specified, the Bonferroni correction is used.

ORIGIN *The direction of the comparison.* This specification will determine whether column means (proportions) or row means (proportions) are being compared. In SPSS 11.5, only COLUMN is supported.

Example:

```
CTABLES /TABLE AGECAT BY MARITAL
  /CATEGORIES VARIABLES=AGECAT MARITAL TOTAL=YES
  /COMPARETEST TYPE=PROP ALPHA=.01.
```

Comparisons of Column Proportions[a]

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Age category	Less than 25				B	A B C D
	25 to 34	B		B	B	A B C D
	35 to 44	B E		B E	A B C E	B
	45 to 54	B E		B E		
	55 to 64	E	E	E		
	65 or older	E	A C D E	E		

Results are based on two-sided tests with significance level .01. For each significant pair, the key of the category with the smaller column proportion appears under the category with the larger column proportion.

a. Tests are adjusted for all pairwise comparisons within each innermost subtable using the Bonferroni correction.

- The table of counts is identical to that shown in the example for chi-square above.
- The comparison output shows a number of predictable pairs for marital status among different age groups that are significant at the 0.01 level specified with ALPHA in the command.

Example:

```
CTABLES /TABLE AGE > SEX BY MARITAL
  /CATEGORIES VARIABLES=SEX TOTAL=YES
  /COMPARETEST TYPE=MEAN.
```

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
			Mean	Mean	Mean	Mean	Mean
Age of respondent	Gender	Male	49	66	48	44	32
		Female	45	70	48	41	32
		Total	47	70	48	42	32

Comparisons of Column Means[a]

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
			(A)	(B)	(C)	(D)	(E)
Age of respondent	Gender	Male	E	A C D E	E	E	
		Female	E	A C D E	D E	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.

a. Tests are adjusted for all pairwise comparisons within each innermost subtable using the Bonferroni correction.

FORMAT Subcommand

```
/FORMAT MINCOLWIDTH={DEFAULT}  MAXCOLWIDTH={DEFAULT}
                    {value  }              {value  }

        UNITS={POINTS}   EMPTY= {ZERO   }   MISSING= {'.'    }
              {INCHES}          {BLANK  }            {'chars'}
              {CM    }          {'chars'}
```

The FORMAT subcommand controls the appearance of the table. At least one of the following attributes must be specified: MINCOLWIDTH, MAXCOLWIDTH, UNITS, EMPTY, or MISSING.

MINCOLWIDTH *The minimum width of columns in the table.* This includes the main tables as well as any tables of significance tests. DEFAULT honors the column labels setting in the current TableLook. The value must be less than or equal to the setting for MAXCOLWIDTH.

MAXCOLWIDTH *The maximum width of columns in the table.* This includes the main tables as well as any tables of significance tests. DEFAULT honors column labels setting in the current TableLook. The value must be greater than or equal to the setting for MINCOLWIDTH.

UNITS *The measurement system for column width values.* The default is POINTS. You can also specify INCHES or CM (centimeters). UNITS is ignored unless MINCOLWIDTH or MAXCOLWIDTH is specified.

EMPTY *Fill characters used when a count or percentage is zero.* ZERO (the default) displays a 0 using the format for the cell statistic. BLANK leaves the statistic blank. You can also specify a quoted character string. If the string is too wide for the cell, the text is truncated.

If FORMAT EMPTY=BLANK, there will be no visible difference between cells that have a count of 0 and cells for which no statistics are defined.

MISSING *Fill characters used when a cell statistic cannot be computed.* This specification applies to non-empty cells for which a statistic, such as standard deviation, cannot be computed. The default is a period (.). You can specify a quoted string. If the string is too wide for the cell, the text is truncated.

VLABELS Subcommand

```
/VLABELS VARIABLES= varlist

        DISPLAY= {DEFAULT}
                 {NAME   }
                 {LABEL  }
                 {BOTH   }
                 {NONE   }
```

By default, the display of variable labels is controlled by the TVARS specification on the SET command in the SPSS Base system. The VLABELS subcommand allows you to show a name, label, or both for each table variable. The minimum specification is a variable list and a DISPLAY specification. To give different specifications for different variables, use multiple VLABELS subcommands.

VARIABLES *The variables to which the subcommand applies.* You can use ALL or *VARNAME* TO *VARNAME*, which refers to the order of variables in the current active data file. If a specified variable does not appear in a table, VLABELS is ignored for that variable.

DISPLAY *Whether the variable's name, label, both, or neither is shown in the table.* DEFAULT honors the SET TVARS setting. NAME shows the variable name only. LABEL shows the variable label only. BOTH shows the variable name and label. NONE hides the name and label.

SMISSING Subcommand

```
/SMISSING {VARIABLE}
          {LISTWISE}
```

If more than one scale variable is included in a table, you can control whether cases that are missing on one are included in summaries for which they have valid values.

VARIABLE *Exclude cases variable by variable.* A case is included in summaries for each scale variable for which it has a valid value regardless of whether it has missing values for other scale variables in the table.

LISTWISE *Exclude cases that are missing on any scale variable in the table.* This ensures that summaries for all scale variables in the table are based on the same set of cases.

Listwise deletion applies on a per-table basis. Thus, given the specification

```
/TABLE (AGE [MEAN,COUNT]>SEX) + (AGE+CHILDS)[MEAN,COUNT] > HAPPY
```

all cases with valid values for *AGE* will be used in the AGE > SEX table regardless of whether they have missing values for *CHILDS* (assuming that they also have valid values for *SEX*).

MRSETS Subcommand

```
/MRSETS COUNTDUPLICATES= {NO }
                         {YES}
```

For multiple response sets that combine multiple category variables, a respondent can select the same response for more than one of the variables. Typically, only one response is desired. For example, if *$MAGS* combines *MAG1* to *MAG5* to record which magazines a respondent reads regularly, if a respondent indicated the same magazine for *MAG1* and *MAG2*, you would not want to count that magazine twice. However, if *$CARS* combines *CAR1* to *CAR5* to indicate which cars a respondent owns now, and a respondent owns two cars of the same make, you might want to count both responses. The MRSETS subcommand allows you to specify whether duplicates are counted. By default, duplicates are not counted.

The MRSETS specification applies only to RESPONSES and percentages based on RESPONSES. It does not affect counts, which always ignore duplicates.

MRSETS

```
MRSETS

 /MDGROUP NAME= setname  LABEL= 'label'
    VARIABLES= varlist
    VALUE= {value  }
           {'chars'}

 /MCGROUP NAME= setname  LABEL= 'label'
    VARIABLES= varlist

 /DELETE NAME= {[setlist]}
               {ALL      }

 /DISPLAY NAME= {[setlist]}
                {ALL      }
```

The set name must begin with a $ and follow SPSS variable naming conventions.

Square brackets shown in the DELETE and DISPLAY subcommands are required if one or more set names is specified, but not with the keyword ALL.

Example

```
MRSETS
  /MDGROUP NAME=$mltnews LABEL='News sources'
   VARIABLES=news5 news4 news3 news2 news1
   VALUE=1
  /DISPLAY NAME=[$mltnews].

MRSETS
  /MCGROUP NAME=$mltcars
   LABEL='Car maker, most recent car'
   VARIABLES=car1 car2 car3
  /DISPLAY NAME=[$mltcars].
```

Overview

The MRSETS command defines and manages multiple response sets. The set definitions are saved in the SPSS data file, so they are available whenever the file is in use. Two types can be defined:

- Multiple dichotomy (MD) groups combine variables such that each variable becomes a category in the group. For example, take five variables that ask for *yes*/*no* responses to the questions:

  Do you get news from the Internet?
  Do you get news from the radio?
  Do you get news from television?
  Do you get news from news magazines?
  Do you get news from newspapers?

These variables are coded 1 for *yes* and 0 for *no*. A multiple dichotomy group combines the five variables into a single variable with five categories in which a respondent could be counted zero to five times, depending on how many of the five elementary variables contain a 1 for that respondent. It is not required that the elementary variables be dichotomous. If the five elementary variables had the values 1 for *regularly*, 2 for *occasionally*, and 3 for *never*, it would still be possible to create a multiple dichotomy group that counts the variables with 1's and ignores the other responses.

- Multiple category (MC) groups combine variables that have identical categories. For example, suppose that instead of having five *yes*/*no* questions for the five news sources, there are three variables, each coded 1 = *Internet*, 2 = *radio*, 3 = *television*, 4 = *magazines*, and 5 = *newspapers*. For each variable, a respondent could select one of these values. In a multiple category group based on these variables, a respondent could be counted zero to three times, once for each variable for which he or she selected a news source. For this sort of multiple response group, it is important that all of the source variables have the same set of values and value labels and the same missing values.

The MRSETS command also allows you to delete sets and to display information about the sets in the data file.

Syntax Conventions

The following conventions apply to the MRSETS command:

- All subcommands are optional, but at least one must be specified.
- Subcommands can be issued more than once in any order.
- Within a subcommand, attributes can be specified in any order. If an attribute is specified more than once, the last instance is honored.
- Equals signs are required where shown in the syntax diagram.
- Square brackets are required where shown in the syntax diagram.
- The TO convention and the ALL keyword are honored in variable lists.

MDGROUP Subcommand

```
/MDGROUP NAME= setname  LABEL= 'label'
    VARIABLES= varlist
    VALUE= {value  }
           {'chars'}
```

The MDGROUP subcommand defines or modifies a multiple dichotomy set. A name, variable list, and value must be specified. Optionally, a label can be specified for the set.

NAME *The name of the multiple dichotomy set.* The name must follow SPSS variable naming conventions and begin with a $. If the name refers to an existing set, the set definition is overwritten.

LABEL *The label for the set.* The label must be quoted and cannot be wider than the SPSS limit for variable labels. By default, the set is unlabeled.

VARIABLES *The list of elementary variables that define the set.* Variables must be of the same type (numeric or string). At least two variables must be specified.

VALUE *The value that indicates presence of a response.* This is also referred to as the "counted" value. If the set type is numeric, the counted value must be an integer. If the set type is string, the counted value, after trimming trailing blanks, cannot be wider than the narrowest elementary variable.

Elementary variables need not have variable labels, but because variable labels are used as value labels for categories of the MD variable, a warning is issued if two or more variables of an MD set have the same variable label. A warning is also issued if two or more elementary variables use different labels for the counted value—for example, if it is labeled *Yes* for Q1 and *No* for Q2. When checking for label conflicts, case is ignored.

MCGROUP Subcommand

```
/MCGROUP NAME= setname  LABEL= 'label'
    VARIABLES= varlist
```

The MCGROUP subcommand defines or modifies a multiple category group. A name and variable list must be specified. Optionally, a label can be specified for the set.

NAME *The name of the multiple category set.* The name must follow SPSS variable naming conventions and begin with a $. If the name refers to an existing set, the set definition is overwritten.

LABEL *The label for the set.* The label must be quoted and cannot be wider than the SPSS limit for variable labels. By default, the set is unlabeled.

VARIABLES *The list of elementary variables that define the set.* Variables must be of the same type (numeric or string). At least two variables must be specified.

The elementary variables need not have value labels, but a warning is issued if two or more elementary variables have different labels for the same value. When checking for label conflicts, case is ignored.

DELETE Subcommand

```
/DELETE NAME= {[setlist]}
              {ALL      }
```

The DELETE subcommand deletes one or more set definitions. If one or more set names is given, the list must be enclosed in square brackets. ALL can be used to delete all sets; it is not enclosed in brackets.

DISPLAY Subcommand

```
/DISPLAY NAME= {[setlist]}
               {ALL      }
```

The DISPLAY subcommand creates a table of information about one or more sets. If one or more set names is given, the list must be enclosed in square brackets. ALL can be used to refer to all sets; it is not enclosed in brackets.

TABLES Command Syntax Converter

If you have command syntax files that contain TABLES syntax that you want to convert to CTABLES syntax, a simple utility program is provided to help you get started with the conversion process. There are, however, significant differences between TABLES and CTABLES functionality, and it is likely that you will find that the utility program cannot convert some of your TABLES syntax jobs or may generate CTABLES syntax that produces tables that do not closely resemble the original tables produced by the TABLES command. In most cases, you can edit the converted syntax to produce a table closely resembling the original.

The utility program is designed to:

- Create a new syntax file from an existing syntax file. The original syntax file is not altered.
- Convert only TABLES commands in the syntax file. Other commands in the file are not altered.
- Retain the original TABLES syntax in commented form.
- Identify the beginning and end of each conversion block with comments.
- Identify TABLES syntax commands that could not be converted.
- Convert command syntax files that follow either interactive or production mode syntax rules.

The utility program may convert TABLES commands incorrectly under some circumstances, including TABLES commands that contain:

- Parenthesized variable names with the initial letters “sta” or “lab” in the TABLES subcommand if the variable is parenthesized by itself. For example, var1 by (statvar) by (labvar). These will be interpreted as the (STATISTICS) and (LABELS) keywords.
- SORT subcommands that use the abbreviations A or D to indicate ascending or descending sort order. These will be interpreted as variable names.

The utility program cannot convert TABLES commands that contain:

- Syntax errors.
- OBSERVATION subcommands that refer to a range of variables using the TO keyword (for example, var01 TO var05).
- String literals broken into segments separated by plus signs (for example, TITLE "My" + "Title").
- Macro calls that, in the absence of macro expansion, would be invalid TABLES syntax. Since the converter does not expand the macro calls, it treats them as if they were simply part of the standard TABLES syntax.

The utility program will not convert TABLES commands contained in macros. All macros are unaffected by the conversion process.

Using the Conversion Utility Program

The conversion utility program, *syntaxconverter.exe*, is installed in the same directory as SPSS. It is designed to run from a command prompt. The general form of the command is:

```
[SPSS install location]\syntaxconverter.exe [path]\inputfilename.sps [path]\outputfilename.sps
```

If any directory names contain spaces, enclose the entire path and filename in quotes, as in:

```
"c:\program files\spss\syntaxconverter.exe" c:\myfiles\oldfile.sps "c:\new files\newfile.sps"
```

Interactive versus Production Mode Command Syntax Rules

The conversion utility program can convert command files that use interactive or production mode syntax rules.

Interactive. The interactive syntax rules are:

- Each command begins on a new line.
- Each command ends with a period (.).

Production mode. The SPSS Production Facility and commands in files accessed via the INCLUDE command in a different command file use production mode syntax rules:

- Each command must begin in the first column of a new line.
- Continuation lines must be indented at least one space.
- The period at the end of the command is optional.

If your command files use production mode syntax rules and don't contain periods at the end of each command, you need to include the command line switch -b (or /b) when you run *syntaxconverter.exe*, as in:

```
"c:\program files\spss\syntaxconverter.exe" -b c:\myfiles\oldfile.sps c:\myfiles\newfile.sps
```

Subject Index

Syntax Index