# *SPSS Tables*<sup>™</sup> *11.5*



For more information about SPSS<sup>®</sup> 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 is a registered trademark and the other product names are the trademarks of SPSS Inc. for its proprietary computer software. No material describing such software may be produced or distributed without the written permission of the owners of the trademark and license rights in the software and the copyrights in the published materials.

The SOFTWARE and documentation are provided with RESTRICTED RIGHTS. Use, duplication, or disclosure by the Government is subject to restrictions as set forth in subdivision (c)(1)(ii) of The Rights in Technical Data and Computer Software clause at 52.227-7013. Contractor/manufacturer is SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

General notice: Other product names mentioned herein are used for identification purposes only and may be trademarks of their respective companies.

TableLook is a trademark of SPSS Inc. Windows is a registered trademark of Microsoft Corporation. DataDirect, DataDirect Connect, INTERSOLV, and SequeLink are registered trademarks of MERANT Solutions Inc. Portions of this product were created using LEADTOOLS © 1991–2000, LEAD Technologies, Inc. ALL RIGHTS RESERVED. LEAD, LEADTOOLS, and LEADVIEW are registered trademarks of LEAD Technologies, Inc. Portions of this product were based on the work of the FreeType Team (*http://www.freetype.org*).

SPSS Tables<sup>™</sup> 11.5 Copyright © 2002 by SPSS Inc. All rights reserved. Printed in the United States of America.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

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

ISBN 1-56827-302-9

# Preface

SPSS<sup>®</sup> 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

25

1	Getting Started with SPSS Tables 1
	What's New in Tables?
	Table Structure and Terminology
	Pivot Tables
	Variables and Level of Measurement
	Rows, Columns, and Cells
	Stacking
	Crosstabulation
	Nesting
	Layers
	Tables for Variables with Shared Categories
	Multiple Response Sets
	Totals and Subtotals
	Custom Summary Statistics for Totals
	Sample Data File
	Building a Table
	Opening the Custom Table Builder
	Selecting Row and Column Variables
	Inserting Totals and Subtotals
	Summarizing Scale Variables

# **2** Table Builder Interface

Building Tables												•	 26
To Build a Table													 29
Stacking Variables													 29
Nesting Variables.													 30
Layers													 32

Showing and Hiding Variable Names and/or Labels
Summary Statistics
Categories and Totals
Tables of Variables with Shared Categories
(Comperimeter Tables)
Customizing the Table Builder
Custom Tables: Options Tab
Custom Tables: Titles Tab
Custom Tables: Test Statistics Tab

# *3* Simple Tables for Categorical Variables 55

A Single Categorical Variable 56
Percentages
Totals
Crosstabulation
Percentages in Crosstabulations
Controlling Display Format
Marginal Totals
Sorting and Excluding Categories

# *4 Stacking, Nesting, and Layers with Categorical Variables*

Stacking Categorical Variables
Stacking with Crosstabulation
Nesting Categorical Variables
Suppressing Variable Labels
Nested Crosstabulation
Layers
Two Stacked Categorical Layer Variables
Two Nested Categorical Layer Variables

73

# **5** Totals and Subtotals for **Categorical Variables**

Simple Total for a Single Variable									.91
What You See Is What Gets Totaled $$									.93
Display Position of Totals					•				.95
Totals for Nested Tables									.95
Layer Variable Totals									.98
Subtotals									.99
What You See Is What Gets Subtotaled									101
Layer Variable Subtotals								•	102

# 6 Tables for Variables with **Shared Categories**

# 

# 7 Summary Statistics

# 113

103

*91* 

Summary Statistics Source Variable	4
Summary Statistics Source for Categorical Variables11	15
Summary Statistics Source for Scale Variables	8
Stacked Variables	21
Custom Total Summary Statistics for Categorical Variables 12	24
Displaying Category Values	28

# 8 Summarizing Scale Variables

Stacked Scale Variables	32
Multiple Summary Statistics	33
Count, Valid N, and Missing Values	34
Different Summaries for Different Variables 1	36
Group Summaries in Categories	39
Multiple Grouping Variables	39
Nesting Categorical Variables within Scale Variables 1	42

# Test Statistics

Tests of Independence (Chi-Square)	145
Effects of Nesting and Stacking on Tests of Independence. $\ldots$ 1	150
Comparing Column Means	152
Effects of Nesting and Stacking on Column Means Tests 1	155
Comparing Column Proportions	158
Effects of Nesting and Stacking on Column Proportions Tests 1	162

# Multiple Response Sets

# 

Defining Multiple Response Sets	168
Counts, Responses, Percentages, and Totals	170
Using Multiple Response Sets with Other Variables	174
Statistics Source Variable and Available Summary Statistics $\ . \ .$	176
Multiple Category Sets and Duplicate Responses	178

х

# **11** Missing Values

Tables without Missing Values					•	•		•		182
Including Missing Values in Tables .										184

# **12** Formatting and Customizing Tables 187

Summary Statistics Display Format
Display Labels for Summary Statistics
Column Width
Display Value for Empty Cells
Display Value for Missing Statistics
Changing the Default TableLook

# Syntax Reference

CTABLES	 
MRSETS	 

# AppendixTABLES Command Syntax Converter237

Subject Index	241
---------------	-----

# Syntax Index 247

# **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.

2

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

Figure	1-1
--------	-----

Table builder with table preview

Custom Tables						_ 0
able Titles Test Statistic:	s Options					
•			Ē	Normal 🗄	Compact	🗋 Layers
<ul> <li>✓ Highest year sc</li> <li>✓ Hours per day</li> <li>☑ How get paid la</li> </ul>			Colum	ins		
How many gran					Count	Mean
How often does		Mala	Get news from	No	nnnn	
Labor force status	Gender	Male	internet	Yes	nnnn	
News sources	Gender	Female	Get news from	No	nnnn	
Number of child			internet	Yes	nnnn	
		A great deal			nnnn	
Categories.	television	Only some		nnnn		
Married		Hardly any		nnnn		
Widowed	Hours per day	watching TV				nnnn
Separated	Highest year o	f school comple	ted			nnnn
Never married	Age of respon	dent				nnnn
📕 Missing Value	•					•
Define	Summa	ry Statistics —				
N <sub>%</sub> Summary Statistics	Positio	n: Columns		•	Hide Catego	ry Position:
S Categories and Totals	. S <u>o</u> urc	e: Row Variab	oles	~	Defaul	t
		OK	Paste	<u>R</u> ese	t Cance	l Help

# 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.

4

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

Getting Started with SPSS Tables

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

Figure 1-3 Horizontal stacking

Gen	der	Age category					
		Less than					65 or
Male	Female	25	25 to 34	35 to 44	45 to 54	55 to 64	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	Less than 25	108	134
category	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	Less than	Male	108
category	25	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 threedimensional "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

Ge	Gender Female				
Ag	je	Less than 25	134		
ca	tegory	25 to 34		351	
	35 to 44 370		370		
	Gender Age categor	Male Less than 2 Y 25 to 34 35 to 44 45 to 54 55 to 64 65 or older	:5	108 276 309 221 136 178	

# **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 a⊓y
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. 8

## Figure 1-8

Multiple response set displayed in a table

		Count	Column %
News	Get news from internet	867	41.7%
sources	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

#### Getting Started with SPSS Tables

#### 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	1 A great deal	Count	196
in television	2 Only some	Count	936
	3 Hardly a⊓y	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

Open File		?	×
Look jn: 🔂 samp	ole_files	▼ ← 🗈 💣 Ⅲ-	
🛗 hivassay.sav	🛅 property_assess.sav	🛅 smalldemo.sav	1
hourlywagedata	.sav 🛛 🧰 salesperformance.sav	🛅 smokers.sav	١.
📲 mailresponse.sa	v 🛅 satisf.sav	🛅 storebrand.sav	١.
📲 🛗 marketvalues.sa	iv 🛅 shampoo_ph.sav	📰 survey_sample.sav	١.
📑 🛗 mutualfund.sav	🧰 site.sav	🛅 tastetest.sav	
🛗 polishing. sav	🛗 siteratings.sav	🛅 telco.sav	
•			•
File <u>n</u> ame: surv	vey_sample.sav	<u>O</u> pen	
Files of type: SPS	SS (*.sav)	→ Paste	
		Cancel	

- ► 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...

Figu Ana	ure 1 <i>lyze</i>	<b>-14</b> men	u, Ta	bles									
<u>F</u> ile	<u>E</u> dit	⊻iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	<u>W</u>	indow	<u>H</u> elp			
					Re <u>p</u> or D <u>e</u> scri	ts iptive Stat	istics	+					
					<u>I</u> able:	s			<u>C</u> us	tom Ta	bles		
					Compa <u>G</u> ener Mi <u>x</u> ed <u>C</u> orrela <u>R</u> egre	are <u>M</u> eans :al Linear M Models ate :ssion	: Model	) ) ) )	<u>B</u> as <u>G</u> er <u>M</u> ul <u>T</u> ab	ic Tabl heral Ta tiple Re bles of F	es ibles isponse irequent	Tables	
					Login Classif Data F Scale Nonpa Time S Surviv Multipl Missin	rssion ear Reduction arametric 1 Series val le Respon ig <u>V</u> alue A	Fests ise nalysis	* * * * * * * * *					

This opens the custom table builder.

# Figure 1-15 Custom table builder

👷 Tables				_ 🗆 🗙
Table Titles Test Statistics	TableLook Options			
•	1 🖽	lormal I匾 Com	pact 🕞	Layers
1st mentioned c     2nd mentioned c     3rd mentioned c     3rd mentioned c     Age of respondent     Belief in life after     Bon in this coun     Confid. In exec     Confid. In exec     Categories:     Africe     American Indian     American Indian	Rows	Columns		
Define Summary Statistics Categories and Totals	Summary Statistic Position: Colum Source:	↓ Lide	Categor <u>y</u> Positi Default	on:
0	IK <u>P</u> aste	<u>R</u> eset	Cancel	Help

# 12

Getting Started with SPSS Tables

# **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

📲 Custom Tables				_ 🗆 🗙
Table Titles Test Statistics Opti	ons	Normal 📇	Compact	Lavers
1st mentioned c     2nd mentioned c     3rd mentioned c     Aqe category		Col	umns	
Age of respondent			Count	
📔 Belief in life after		Less than 25	nnnn	
Born in this coun		25 to 34	nnnn	
Confid. In exec b		35 to 44	nnnn	
	Age category	45 to 54	nnnn	
Categories:		55 to 64	nnnn	
📕 Less than 25 📃		65 or older	nnnn	
25 to 34	-		,	
Define	- Summary Statis	tics		
N <sub>%</sub> Summary Statistics	Position: Col	umr 🔻 🗖 Hi	de Category	Position:
			Default	-
55 Lategories and Lotals	Source: Ro	₩V. ▼		
OK	Paste	<u>R</u> eset	Cancel	Help

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

Security Tables					-	
Table Titles Test Statistics	Optio	ins				
			Norm <u>a</u> l 🗄	Compact	🗋 Laye	ers
Confidence in or 🔺			C	olumns		וו
Confidence in sc				Ge	nder	
Contidence in tel				Male	Female	
Eavor or oppose				Count	Count	
Gender			Less than 25	nnnn	nnnn	
🚺 General happiness 🔤	12		25 to 34	nnnn	nnnn	
Get news from in	ŝ	Ago ostogow	35 to 44	nnnn	nnnn	
Categories:		Age categoly	45 to 54	nnnn	nnnn	
			55 to 64	nnnn	nnnn	
E Fomalo			65 or older	nnnn	nnnn	
_ Define		Summary Statis	stics —			
N <sub>%</sub> Summary Statistics		Position: Col	umns 🔻 🗖	Hide Categ	gory Position:	
😌 Categories and Totals		Courses Bos	u) (mir 💌	Defa	ult	•
		oguide: No				
	ОК	Paste	<u>R</u> ese	t Can	cel He	lp
						//

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.

#### Getting Started with SPSS Tables

#### 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

Summary Statistics: Categ	orical Variables				×
Selected Variable: Age categ	jory				
Statistics:	<u>D</u> isplay:				
Unweighted Count 🔺	Statistic	Label	Format	Decimals	
Column %	Count	Count	nnnn	0	•
Table % Subtable %	Row %	Row %	nnnn.n%		_
Laver %					
□ <u>C</u> ustom Summary Stati	stics for Totals and Subtotals				
Custom Summary Stati	stics for Totals and Subtotals			11	
Custom Summary Stati Statistics: Unweighted Count	stics for Totals and Subtotals Display: Statistic	Label	Format	Decimals	
Custom Summary Stati Statistics:	stics for Totals and Subtotals Display: Statistic Count	Label	Format	Decimals 0	
Custom Summary Stati Statistics: Unweighted Count	stics for Totals and Subtotals Display Statistic Count	Label Count	Format	Decimals 0	
Custom Summary Stati     Statistics:     Unweighted Count     Row %     Column %     Table %     Subtable %	stics for Totals and Subtotals Display: Statistic Count	Label Count	Format	Decimals 0	•
Column % Table %	stics for Totals and Subtotals Display: Statistic Count	Label Count	Format nnnn	Decimals 0	•
Column % Column % Column % Column % Subtable %	stics for Totals and Subtotals Display Statistic Count	Label Count	Format nnnn	Decimals 0	•
Custom Summary Stati	stics for Totals and Subtotals Display Statistic Count	Label Count	Format	Decimals 0	•
Column & Col	stics for Totals and Subtotals Display Statistic Count Apply to Selection	Label Count	Format nrnnn Close	Decimals 0 Help	•

16

• 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

Scustom Tables					-	. 🗆 🗙
Table Titles Test Statistic	s Optic	ons				
			🛅 Norm <u>a</u> l	岩 Compact	🕒 La	yers
<ul> <li>Confidence in or</li> <li>Confidence in pr</li> </ul>				Colu	mns	
Confidence in sc					Ger	ider
Country of family				Ma	ale	
Eavor or oppose				Count	Row %	C
Gender			Less than 25	nnnn	nnnn	
General happiness	DWS		25 to 34	nnnn	nnnn	
🚺 🔂 Get neuve from in	ŭ	0 <b>-</b>	35 to 44	nnnn	nnnn	
Tatagorias:		Age categoly	45 to 54	nnnn	nnnn	
			55 to 64	nnnn	nnnn	
Male			65 or older	nnnn	nnnn	
- remaie	•					►
Define		- Summary Statis	tics			
N <sub>%</sub> Summary Statistics		Position: Col	umns 🔻 🗖	Hide Cate	gory Position:	
S Categories and Totals.		S <u>o</u> urce: Roy	w Variat 💌	Def	ault	-
	OK	Past	e <u>R</u> es	set Car	ncel H	lelp

# **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

×
E <u>x</u> clude:
Totals and Subtotals Appear C Aboye C Below These controls determine whether subtotals apply to
Categories may precede or follow them.  Dther values found when data are scanned  Cancel Help

► 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

🚜 Custom Tables					-	□ ×
Table   Titles   Test Statistic	:s) Op	otions				
	⊧	·	Normal	岩 Compact	🗋 Laye	ers
🚺 Confidence in m 🔺				Male	Female	•
Confidence in m				Count	Count	
Confidence in or			Less than 25	nnnn	nnnn	_
Confidence in or			25 to 34	nnnn	nnnn	
Confidence in sci	NS.		35 to 44	nnnn	nnnn	
Confidence in tel	8		Subtotal < 45	nnnn	nnnn	
Country of family		Age category	45 to 54	nnnn	nnnn	
E Country of origin			55 to 64	nnnn	nnnn	
€			65 or older	nnnn	nnnn	
Lategories:			Subtotal 45+	nnnn	nnnn	
Male	H-		Total	nnnn	nnnn	
Female					•	-
, _ Define		- Summary Sta	atistics			
N <sub>%</sub> Summary Statistics		Position: C	olumns 🔻 🗖	I Hide Cate	gory Position:	
00 Catanatian and Tatala				Del	ault	•
on Lategories and Lotais.		Source: H	iow vari 💌			
	OK	Past	e <u>R</u> es	et Car	ncel He	lp

• 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	Less than 25	108	44.6%	134	55.4%	
category	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%	

Getting Started with SPSS Tables

# 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

Security Custom Tables					- 🗆 🗙		
Table Titles Test Statistics Options							
4		. 🖽	Normal 믉	Compact	Layers		
1st mentioned c     2nd mentioned c     3rd mentioned c     Age catenory			Col	umns			
Age of respondent				Count			
Belief in life after			Less than 25	nnnn			
Born in this coun	kows		25 to 34	nnnn			
Confidence in ba			35 to 44	nnnn			
\$		Age category	45 to 54	nnnn			
Lategories:			55 to 64	nnnn			
Less than 25			65 or older	nnnn			
25 to 34							
Define		- Summary Statis	tics				
N <sub>%</sub> Summary Statistics		Position: Colu	umr 🔻 🗖 <u>H</u> i	de Category F	Position:		
Categories and Totale     Causes: Remit      Default							
		Source. They	Y Y .				
OK <u>P</u> aste <u>R</u> eset Cancel Help							

► 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

📲 Tables						_ 🗆 🗙
Table   Titles   Test Statistic	s   Tablel	_ook Option	s ]			
		[	Normal	層 Comp	act	Layers
Happiness of m	Rows	.ge category	Less than 25 25 to 34 35 to 44 45 to 54 55 to 64	Columns Cour Hours pe	nt er day	
variabis)			65 or older	_		_
Define		Summary Stat	istics	Hide	Category Po	sition:
<u>Categories</u> and Totals		Source: Ro	ow Vari 🔻		Default	•
	OK	Paste	<u><u>R</u>e:</u>	set	Cancel	Help

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

State Custom Tables				_ 🗆 🗙
Table Titles Test Statistics Op	tions			
0	. [1	Normal 🗄	Compact	Layers
Get news from te		C	olumns	
Highest year of s				Mean
Highest year sch		Less than 25	Hours per d	nnnn
Highest year sch		25 to 34	Hours per d	nnnn
Hours per day w	Ann antonom	35 to 44	Hours per d	nnnn
¢	Age categoly	45 to 54	Hours per d	nnnn
Categories:		55 to 64	Hours per d	nnnn
No categories (scale		65 or older	Hours per d	nnnn
Define	Summary Statis	stics	Cater	nru Position:
**% ⊇ummary Statistics	Position: Col	umns 🔻 🗋	Hide Defai	di -
S Categories and Totals	S <u>o</u> urce: Rou	w Vari 🔻	Derau	
OK	Paste	<u>R</u> eset	: Canc	el Help

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

Custom Tables				_ 🗆 ×
Table Titles Test Statistics	Opti	ons	🔲 Normal 🚍 Compact	Lavers
Get news from te Happiness of ma Have gun in home Havest degree			Columns	
Highest year of s				Mean
Highest year sch			Summary Statistics	nnnn
Highest year sch	19		Categories and Totals	nnnn
Hours per day w 💌	No.	An		nnnn
			Select All Row Variables	nnnn
			Select All Column Variables	nnnn
No categories (scale variable)			Undo Add Variables Can't Repeat	- nnnn
Define		Sur	<u>D</u> elete	
N <sub>%</sub> Summary Statistics		<u>P</u> o:	Show Variable Name	Fory Position:
S Categories and Totals		S <u>c</u> ▼	Show Variable Label	iult 👻
	ОК		<u>P</u> aste <u>R</u> eset Ca	ncel Help

► 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

Summary Statistics: Scale Variables								
Selected Variable: Hours per day	watching TV							
Statistics:	Display:							
Count 🔺	Statistic	Label	Format	Decimals				
Unweighted Count 💻	Mean	Mean	Auto		<b></b>			
Maximum	Median	Median	Auto					
Minimum					*			
Missing								
	Apply to <u>S</u> election	Apply to <u>A</u> ll	Close	<u>H</u> elp				

22

• 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

Custom Tables	cs Options	]			
Get news from te A Happiness of ma Have gun in home		c	Normal 🗄 I	Co <u>m</u> pact	Layers
Highest degree				Mean	Median
Highest year of s	I	Less than 25	Hours per d	nnnn	nnnn
Highest year sch		25 to 34	Hours per d	nnnn	nnnn
Highest year sch		35 to 44	Hours per d	nnnn	nnnn
🖉 Hours per day w 💌	category	45 to 54	Hours per d	nnnn	nnnn
Categories:		55 to 64	Hours per d	nnnn	nnnn
		65 or older	Hours per d	nnnn	nnnn
Define		ummary Statistic osition: Colum	s ns ▼	e Category F	Position:
S Categories and Totals		iource: Row V	′ari ▼	Default	1
	UK	Paste	<u>H</u> eset	Cancel	Help

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

Custom Tables						-	
Table   Titles   Test Statistic	s Option	s					
		. 6	Norma	- ∺	Compact	Ch Lay	vers
🚺 1 st mentioned c 🔺		,				 ¬	- II
2nd mentioned c		(	Columns				
3rd mentioned c							
Age category					Mean	Median	
🗿 Belief in life after		Less than 25	Hours	Sum	n nmary Statistics .		
Born in this coun Confidence in ba		25 to 34	Hours	Cate	egories and Tota	\$	
	category	35 to 44	Hours	tours Swap Row and Column Variables Select All Row Variables			
Confidence in ea		45 to 54	Hours				
Confidence in m		55 to 64	Hours	Select All Column Variables			
	<u> </u>	65 of older	Hours -	Can	't Undo		
Categories:				Can	't Repeat		
			. –	Cut			Ctrl+X
	[	Summary Statis	tics——	Cop	y		Ctrl+C
№% <u>S</u> ummary Statistics		Position: Colu	imns 🔻	Past	te		Ctrl+V
S Categories and Totals		Source: Rov	v Vari, 🔻 🔤	<u>D</u> ele	ete		Del
				Sho	w Variable Nam	э	
	OK	<u>P</u> aste		/ Sho	w Variable Labe		, ,
							-//

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

Se Custo	om Tabl	es					_ 🗆 🗙
Table	Titles	Test Statisti	ics   Options	1			
			E	2 <u>D</u> ate	🕘 Time	<b>BY</b> +> Table <u>E</u>	xpression
Li Avera by Age (	tle: age Daily Category	Number of F	lours of Telev	vision Watche	ed		
	aption:						
	orner:						
1	) Toens Nested	ure that corn in Table Pro	er text is displ perties. This c	ayed in the ta choice can be	ble, check that Ro preset in the defa	w Dimension Lab ult TableLook.	els is set to
			OK	<u>P</u> aste	<u>R</u> eset	Cancel	Help

• 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.

Chapter 2

26

# **Building Tables**

#### Figure 2-1

Custom Tables: Table tab

Custom Tables	atistics	Options					
					Norm <u>a</u> l	🗄 Compact	Layers
Highest de 🔺					Columns		
Highest ye	$\square$					Count	Percent
Highest ye			Less than 25				
How get p			25 to 34				
How many			35 to 44				
How often		Age category	Subtotal < 45	i			
🚺 Labor forc		Age categoly	45 to 54				
📔 Marital stat 🖵	12		55 to 64				
	Rov		65 or older				
Categories:			Total				
<ul> <li>Missing Values</li> <li>Very happy</li> <li>Pretty happy</li> <li>Not too happy</li> </ul>				General	Very happy		
			Male	happiness	Pretty happy		
		Gender			Not too happy		
				General	Very happy		
			Female	happiness	Pretty happy		
					Not too happy		
- Define		Summari	u Statistics				
No Summary Statistic:		Position	Columns			Hide Cate	egory Position:
S Categories and To	otals	Source	: Row Variab	les	•	De	fault 🔻
			OK	Pas	ste <u>R</u> es	et Ca	ncel Help

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.



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 rightclicking 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 Chapter 2

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	Less than 25	242
category	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	Less than	Male	108
category	25	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*.

Chapter 2

#### Layers

You can use layers to add a dimension of depth to your tables, creating threedimensional "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

G	ender Fer	nale		
A	ge	Less than 25	134	
0	ategory	25 to 34	351	
		35 to 44	370	
	Gender Age categoi	Male           Less than 25           'Y         25 to 34           35 to 44           45 to 54           55 to 64           65 or older	108 276 309 221 136 178	

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

Summary Statistics: Categor	ical Variables				×
Selected Variable: Age category	,				
Statistics:	<u>D</u> isplay:				
Row %	Statistic	Label	Format	Decimals	
Table % 📃	Count	Count	nnnn	0	
Subtable %	Column %	Column %	nnnn.n%	1	
Layer %					•
Row Valid N &					
🕞 🔽 Custom Summary Statistic	s for Totals and Subtotals				
Statistics:	Display:				
Row %	Statistic	Label	Format	Decimals	
Column % 🗖	Count	Count	nnnn	0	
Table %	Mean	Mean	Auto		
Subtable %					-
Layer %					
	Apply to <u>S</u> election	Apply to <u>A</u> ll	Close	Help	

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	1 A great deal	196	
in television	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	1 A great deal	Count	196
in television	2 Only some	Count	936
	3 Hardly a⊓y	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		×
Selected Variable: Age c	ategory	
Display		E <u>x</u> clude:
Value(s)	Label	
2.00	25 to 34	
3.00	35 to 44	
1.003.00     1.003.00	Subtotal < 45	
4.00	45 to 54	
5.00	55 to 64	
6.00	65 or older	
	Subtotal 45+	
Subtotals       Insert     I       Delete     I       Sort Categories     I       By:     I	abet Subtotal 45+ Categories omitted from subtotals: 0	Totals and Subtotals Appear C Aboye C Below These controls determine whether subtotals apply to categories that precede or follow them.
Show I Iotal Label: Total	☐ <u>Missing</u> ☐ Empty Categories	✓ Other values found when data are scanned
		Cancel <u>H</u> elp

- 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

	A great deal	Only some	Hardly a⊓y
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

Stacked variables with shared response categories in columns

Chapter 2

#### **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	
--------	---------	-------------	--

🚜 Custom Tables	
Table Titles Test Statistics Options	
Data Cell Appearance	Width for Data Columns
Empty Cells: 💿 Zero	TableLook settings
⊂ <u>B</u> lank	C <u>C</u> ustom
C Iext:	Minimum: 36
	Maximum: 72
Statistics that	
Computed:	Units: Points
Missing Values for Scale Variables	
<ul> <li>Maximize use of available data</li> <li>(variable buvariable deletion)</li> </ul>	C Use consistent case base across
(valiable by valiable deletion)	
Count duplicate responses for multiple categor	ry sets
OK	Paste <u>R</u> eset Cancel Help

**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

📲 Custom Ta	ables				- 🗆 🗙
Table Titles	Test Statistics   Op	tions			
		월 Date	🕘 Time	<b>₿</b> ¥ ↓> Table <u>E</u> xp	ression
<u> </u>					
&[Table Ex	pression]				
Caption:					
Corner:					
&[Current D	late]				
To e Nest	nsure that corner text is ed in Table Properties.	displayed in the tab This choice can be j	le, check that Row preset in the defaul	Dimension Labels i: t TableLook.	s set to
	OK.	Paste	<u>R</u> eset	Cancel	Help

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

Custom Tables	_ 🗆 🗙
Table Titles Test Statistics Options	
I [Tests of independence (Chi-square) Alpha: 0.05 ÷	
Compare column means (t-tests)	
Alpha: 0.05 🚊 🔽 Adjust p-values for multiple comparisons (Bonferroni method)	
Compare column proportions (z-tests)	
Alpha: 0.05 😤 🔽 Adjust p-values for multiple comparisons (Bonferroni method)	
Chi-square and column proportions tests apply to tables in which categorical variables exist in both the rows and columns. Column means tests apply to tables in which scale variables exist in the rows and categorical variables exist in the columns. None of the tests is performed on multiple response variables or tables in which category labels are moved out of their default table dimension.	
OK <u>P</u> aste <u>R</u> eset Cancel	Help

**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.



# 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). Chapter 3

56

## 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

Ma Custom Tables				_ 🗆 🗙
Table Titles Test Statistics Op	itions			
φ. 	. 🗇	Normal 🔡	Compact	Layers
Ist mentioned c         Image: Strategy of the st		Col	lumns	
Age of respondent			Count	
🚺 Belief in life after		Less than 25	nnnn	
Born in this coun		25 to 34	nnnn	
Confidence in ba	Age category	35 to 44	nnnn	
÷	Age category	45 to 54	nnnn	
Categories:		55 to 64	nnnn	
Less than 25		65 or older	nnnn	
25 to 34				
Define	🖵 Summary Statis	tics		
N <sub>%</sub> <u>S</u> ummary Statistics	Position: Colu	umr 🔻 🗖 <u>H</u> i	de Category I	Position:
82 Categories and Totals	S <u>o</u> urce: Rov	₩ ₩. ▼	Default	•
OK	<u>P</u> aste	<u>R</u> eset	Cancel	Help

• 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	Less than 25	242
category	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	Less than 25	242
category	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

Statistic Count Column %	Count Percent	nnnn nnnn.n%	Decimals 0 1	4
Count Column %	Count Percent	nnnn nnnn.n%	0	4
Column %	Percent	nnnn.n%	1	
is for Totals and Subtota	ls			
Display:				
Statistic	Label	Format	Decimals	
Count	Count	nnnn	10	
Count				
	s for Totals and Subtota Display: Statistic Count	s for Totals and Subtotals Displag Statistic Label Count Count	s for Totals and Subtotals Display: Statistic Label Format Count Innnn	s for Totals and Subtotals Display: Statistic Label Format Decimals Count Innnn 0

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	Less than 25	242	8.6%
category	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

🛆 Value(s)	Label		
1.00	Less than 25		
2.00	25 to 34 🗸 🗸		
3.00	35 to 44		
4.00	45 to 54		
5.00	55 to 64		
	Label: Subtotals: 0		
Delete	Laber, Jourona subtotals: 0		<ul> <li>Below</li> <li>These controls determine whether subtotals apply to</li> </ul>
Delete	subtotals: 0	,	Below These controls determine whether subtotals apply to categories that precede of follow them.
Delete	Laber: Jourovai subtotals: 0	,	Below These controls determine whether subtotals apply to categories that precede of follow them.

• 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	Less than 25	242	8.6%
category	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.)

📲 Custom Tables					_ 0	×
Table   Titles   Test Statistics	Optic	ons				
0		E	Normal 🗄	🗄 Co <u>m</u> pact	🗋 Layers	
Confidence in or			C	olumns		
Confidence in sc				Ger	nder	
Countra of family				Male	Female	
Favor or oppose				Count	Count	
Gender			Less than 25	nnnn	nnnn	
📔 General happiness 🔤	VS		25 to 34	nnnn	nnnn	
Get news from in	æ	Age category	35 to 44	nnnn	nnnn	
Categories:		Age category	45 to 54	nnnn	nnnn	
Mala			55 to 64	nnnn	nnnn	
Female			65 or older	nnnn	nnnn	
_ Define		- Summary Statis	tics			
N <sub>%</sub> Summary Statistics		Position: Col	umns 🔻 🗖	Hide Categ	ory Position:	
•• Categories and Tatala		Cause Day	u Martin 🖛	Defa	ult 🗸 👻	
oe Lategories and 1 otals		source: Rou	w vark 🔻			
	OK	<u>P</u> aste	<u>R</u> ese	t Cano	el Help	

Crosstabulation in table builder canvas preview

Figure 3-9

# Chapter 3

• Click OK to create the table.

#### Figure 3-10

Crosstabulation of Age category and Gender

		Gender		
		Male	Female	
		Count	Count	
Age	Less than 25	108	134	
category	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	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
category	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

Summary Statistics: Catego	orical Variables				2
Selected Variable: Age catego	ory				
Statistics:	<u>D</u> isplay:				
Unweighted Count 🔺	Statistic	Label	Format	Decimals	
Table %	Count	Count	nnnn	0	
Subtable %	Column %	Column %	nnnn.nn	2	
Layer %	Row %	Row %	nnnn.nn	2	
Custom Summary Statist	tics for Totals and Subtotals Display:	s ————			
Unweighted Count 🔺	Statistic	Label	Format	Decimals	
Row %	Count	Count	nnnn	0	
Column % Table %	•				-
Layer %					
	Apply to Selection	Apply to <u>A</u> ll	Close	<u>H</u> elp	

• 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	Less than 25	108	8.79	44.63	134	8.38	55.37
category	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	Less than 25	108	134	242
category	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	3-1	5
--------	-----	---

andc(a)			
1.00	Less than 25		
2.00	25 to 34		
3.00	35 to 44		
4.00	45 to 54		
5.00	55 to 64		
6.00	65 or older		
Delete	subtotals: 0	Belo <u>w</u>	
Delete t Categories — By: Value	subtotals: 0	Below     These controls de     whether subtotals     categories that pre     follow them.	etermir apply ecede

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

	s			
ected Variable: Age	category			
				Evolude:
isplay				
Value(s)	$\nabla$	Label		
1.00	Less than 25			
6.00	65 or older		-	A     A
5.00	55 to 64			
4.00	45 to 54			
3.00	35 to 44			
2.00	25 to 34			
Insert Delete	Label: Subtotal	subtotals: 0	omitted from	Below
Sort Categories –				whether subtotals apply to categories that precede of
Sort Categories – <u>By</u> : Label	<b></b>	भः Descending	•	whether subtotals apply to categories that precede of follow them.
Sort Categories	<b></b>	er: Descending		whether subtotals apply to categories that precede of follow them.
Sort Categories - <u>By</u> : <mark>Label</mark> ow <u>I</u> otal	Orde	er: Descending	Empty	whether subtotals apply to categories that precede or follow them.
Sort Categories - By: Label w- Jotal abgl: Total	Orde	er: Descending ssing slues I	Empty categories	<ul> <li>Other values found when data are scanned</li> </ul>

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

Categories and Totals	;		×
Selected Variable: Age	category		
- Display			E <u>x</u> clude:
Value(s)	Label		
6.00	65 or older		
5.00	55 to 64		
4.00	45 to 54		•
3.00	35 to 44		
2.00	25 to 34	_	
1.00	Less than 25		
Subtotals	Labet Subtotal Categories om subtotals: 0	itted from	Totals and Subtotals Appear     C Aboye     Below     These controls determine     whether subtotals apply to     categories that precede or     follow them.
Show Iotal Label: Total	☐ <u>Missing</u> Im Em Values Im Cat	ipty tegories	Other values (ound when data are scanned
	[	<u>A</u> pply	Cancel <u>H</u> elp

# **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

Categories and Tota	ls		×
Selected Variable: Age	e category		
			Exclude:
Display-			Less than 25
Value(s)	Label		65 or older
5.00	55 to 64 🔺		
4.00	45 to 54 🗸 🗸		
3.00	35 to 44		
2.00	25 to 34		
Subtotals <u>Insert</u> <u>D</u> elete	Labet Subtotal Categories omitted from subtotals: 0	- Tota C G	als and Subtotals Appear Abo <u>v</u> e Belo <u>w</u> bese controls determine
Sort Categories -	✓ Order: Descending ✓	r w c fc	hether subtotals apply to ategories that precede or illow them.
Label: Total	☐ <u>Missing</u> ☑ Empty Values ☑ Categories	₪ Ot wł	her values <u>(</u> ound ien data are scanned
			ancel <u>H</u> elp

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

Simple Tables for Categorical Variables

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*.

# 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

📲 Custom Tables						_ 🗆 🗙
Table Titles Test Statistic	s   Opt	ions				
•			Norm	al 🗄 Comp	act 🗋	Layers
Ist mentioned         1st mentioned         1st mentioned         1st mentioned         1st mentioned         1st mentioned				Columns		
Age of respond				Count		
👸 Belief in life afte		Gender	Male	nnnn		
Born in this cou		Gender	Female	nnnn		
Confid. In exec	12		Less than 25	nnnn		
¢	Ro	Áge category	25 to 34	nnnn		
Lategories:			35 to 44	nnnn		
Less than 25			45 to 54	nnnn		
20 to 34			55 to 64	nnnn		
45 to 54			65 or older	nnnn	]	
55 to 64						
Define		Summary Stat	istics		Category Position	
**% Summary Statistics		Position: Co	olumns 💌	I <u>H</u> ide	Default	-
S <u>C</u> ategories and Totals		S <u>o</u> urce: Ro	ow Variables 🔻		Derduit	
		OK I	Paste	<u>R</u> eset	Cancel	Help

• Click OK to create the table.

# Figure 4-2

Table of categorical variables stacked in rows

		Count
Gender	Male	1232
	Female	1600
Age	Less than 25	242
category	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

		Get news fr	rom internet
		No	Yes
		Count	Count
Gender	Male	873	359
	Female	1092	508
Age	Less than 25	146	96
category	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

Two stacked row variables crosstabulated with a column variable

*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

& Custom Tables						-	□ >
Table Titles Test Statistics Options							
0			Norn	nal 🗄 Comp	act 🕻	Laye	rs
Confidence in television				Colu	umns		-
Country of origin multiple resp		ıГ			Get news	from inte	
Favor or oppose death penalty					No	Y	
General happiness					Count	Cd	
Get news from internet		l le	Sender	Male	nnnr	1	
👔 Get news from news magazines 🖵		ĽĽ	Gender	Female	nnnr	n in the second se	
	0.WS			Less than 25	nnnr	n i	
Categories:	č			25 to 34	nnnr	n in the	
No.				35 to 44	nnnr	n in the second s	
Yes		<sup>e</sup>	Age category	45 to 54	nnnr	n l	
				55 to 64	nnnr	n l	
				65 or older	nppr		-
						<u> </u>	
C Define Sum	mary S	Statis	tics				
N <sub>%</sub> Summary Statistics Posi	tion:	Colu	umns 🔻	<u>∏</u> <u>H</u> ide	Category Pos	ition:	
PO Categories and Totals		D			Default		•
oe Lategories and Fotals 5 <u>o</u> t	nce:	HO	w variables 💌				
OK		P	aste	<u>R</u> eset	Cancel	Hel	p

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.



📲 Custom Tables							. 🗆 🗙												
Table Titles Test Statis	tics   O	ptions		E Namel	Coursed	ē.													
					E Lompact	, ų La	yers												
Confidence i				Columns															
Country of ra																			
Favor or opp	$\square$					Count													
🚺 Gender			1 11 25	o .	Male	nnnn													
🚺 General happ			Less than 25	Gender	Female	nnnn													
Get news fro			25 to 34	Gender	Male	nnnn													
			201004	Gender	Female	nnnn													
Categories:	Rows	Rows	N2		35 to 44	Gender	Male	nnnn											
Male			Ane category			Female	nnnn												
Female																45 to 54	Gender	Male	nnnn
					Female	nnnn													
			55 to 64 Ge	Gender	Male	nnnn													
				Female	nnnn														
			65 or older	Gender	Male	nnnn													
	_				remale	nnnn	-												
_ Define		Summary SI	tatistics																
N <sub>%</sub> Summary Statistics		Position:	Columns	<b>▼</b> [	<u>H</u> ide Ca	tegory Position:													
SS <u>C</u> ategories and Total		S <u>o</u> urce:	Row Variables	•	D	efault	-												
		OK	Pas	te <u>B</u> e	eset C	ancel H	telp												

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

Custom Tables							_ 🗆 🗙	
Table Titles Test Statistics Options								
				Norm <u>a</u> l	E Compact	6	Layers	
Confidence i				Columns			ר≜ר	
Country of fa							-	
Favor or opp	$\square$					Count	1	
🗿 Gender					Less than 25	nnnr		
General happ					25 to 34	nnnr		
Get news fro			Male	Åge getegoru	35 to 44	nnnr	<u> </u>	
					Age categoly	45 to 54	nnnr	
Categories:	WS				55 to 64	nnnr		
Male	2	Gender			65 or older	nnnr		
Female			Female	Age category	Less than 25	nnnr	<u> </u>	
					25 to 34	nnnr		
					35 to 44	nnnr		
					45 to 54	nnnr		
					55 to 64	nnnr		
					65 or older	nnnr	•	
Define		Summary 9	itatistics					
N <sub>%</sub> Summary Statistics		Position:	Columns	<b>▼</b> [	<u>H</u> ide <sup>Ca</sup>	ategory Position:		
88 Categories and Tota		S <u>o</u> urce:	Row Variables	•	D	efault	-	
		OK	Pa	ste <u>R</u> e	eset C	Cancel	Help	

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	Less than 25	108
		category	25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age	Less than 25	134
		category	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.

Chapter 4

► 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.

#### Stacking, Nesting, and Layers with Categorical Variables

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			

**Figure 4-10** Variable labels in table title

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

Custom Tables	
Table   Titles   Test Statistic	s Options
	Mormal Compact C Layers
Gender General happin Get news from i Get news from i Get news from r Happiness of m	Columns Get news from internet
Define	Summary Statistics
™% <u>S</u> ummary Statistics	Position: Columns
88 Categories and Totals	Source: Row Variables
	OK <u>Paste</u> <u>R</u> eset Cancel Help

• Click OK to create the table.

# Figure 4-12

Nested crosstabulation

				Get news fr	om internet
				No	Yes
				Count	Count
Gender	Male	Age	Less than 25	59	49
		category	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	Less than 25	87	47
		category	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

Custom Tables			_ 🗆 🗡
Table Titles Test Statistics	Options		
•		🛅 Normal 🔚 Com	pact 🗋 Layers
1st mentioned c     2nd mentioned     3rd mentioned     3rd mentioned     Age category     Age of respondent     Belief in life after     Born in this cou     Confidence in b     Confidence in m     Confidence in m     Categories:     No variables selected	Bows	Columns Columns Column Statistics Categories and Totals Categories and Totals Select All Row Variables Select All Column Variables Can't Undo Can't Repeat Delete Show Variable Name Show Variable Label	
Define N <sub>%</sub> Summary Statistics % Categories and Totals	Summary S Position: S <u>o</u> urce:	tatistics Columns T Hide Row Variables T	Category Position: Default 🔹
	OK	Paste Reset	Cancel Help

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.

#### Chapter 4

- ▶ 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	е			Female					
			Age cate	egory					Age ca	ategory			
		Less	25 to	35 to	45 to	55 to	65 or	Less	25 to	35 to	45 to	55 to	65 or
		than 25	34	44	54	64	older	than	34	44	54	64	older
Get news from	No	59	159	217	169	112	155	87	209	218	177	140	261
internet	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 threedimensional "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.

#### 84

## Figure 4-15

Age category in rows, Gender in layers

<b>** Custom Tables</b>						_ 🗆 🗙
Table   Titles   Test Statis	tics   0	lptions ]				
1				Mormal F	3 Compact	Layers
Confidence i			Colur	nns		Layers
Gender I			Less than 25	Dount		
¢		Age category	25 to 34	nnnn		
Categories:	L Rowe		35 to 44	nnnn		- Layer Output
Male			45 to 54	nnnn		Sho <u>w</u> each
E Female			55 to 64	nnnn		<ul> <li>category as a laver</li> </ul>
			65 or older	nnn		Show each combination of categories as a layer
Define		Summary S	tatistics			
N <sub>%</sub> Summary Statistics 88 Categories and Total	s	<u>P</u> osition: S <u>o</u> urce:	Columns Row Variables	•	Hide Cal	egory Position: sfault 🔹
		OK	Pas	te <u>R</u> esel	t Ca	ancel Help

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	Less than 25	108
category	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

Layer	Gender Female	-	
	Gender Male		Count
Age	Gender Female		108
category	25 to 34		276
	35 to 44		309
	45 to 54		221
	55 to 64		136
	65 or older		178

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	Less than 25	134
category	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 Custom Tables \_ 🗆 🗙 Table Titles Test Statistics Options Normal 🗄 Compact 🔽 Layers 🚺 General hap... Layers 🛐 Get news fro.. 🛐 Get news fro.. 🛐 Gender 🛐 Get news fro.. Highest 🛐 Get news fro... 1 degree 🗿 Get news fro... Count Happiness of... Less than 25 nnnr 🚺 Have gun in .. 25 to 34 nnnr 👔 Highest degr... Rows 35 to 44 🔗 Highest year . nnnr Layer Output Age category 45 to 54 🔗 Highest year .. nnnr Sho<u>w</u> each -æ 55 to 64 category as a nnnr laver 65 or older Categories: nnnn Show each 📕 LT Hiah sch... combination of High school categories as a Junior college laver -Bachelor Summary Statistics Define Category Position: Position: Columns ▼ □ Hide Default Categories and Totals. Source: Row Variabl 🔻 0K <u>P</u>aste <u>R</u>eset Cancel Help

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

	Layer	Gender Male Highest degree LT High school	-
		Gender Male Highest degree LT High school	
	Age	Gender Male Highest degree High school	
	category	Gender Male Highest degree Junior college	
		Gender Male Highest degree Bachelor	
		Gender Male Highest degree Graduate	
		Gender Female Highest degree LT High school	
		Gender Female Highest degree High school	
Ļ	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Gender Female Highest degree Junior college	<b>•</b>

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

Categories and Totals			×
Selected Variable: Age of	category		
			Fueluder
Display			
△ Value(s)	Label		
1.00	Less than 25		
2.00	25 to 34		
3.00	35 to 44		
4.00	45 to 54		
5.00	55 to 64		
6.00	65 or older		
Insert Delete Sort Categories — By: Value	Label: Subtotal Categories omitted from subtotals: 0	T M c fc	Above Below hese controls determine hether subtotals apply to ategories that precede or plow them.
Show <b>I</b> otal Label: Total	☐ Missing	₩ W	her values <u>f</u> ound nen data are scanned
	Арру	C	ancel <u>H</u> elp

• 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	Less than 25	242	8.6%
category	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

gories and Tota acted Variable: Age	ls : category	
		E <u>x</u> clude:
shidă		Less than 25
Value(s)	Label	5 or older
2.00	25 to 34	
4.00	45 to 54	
5.00	55 to 64	
Insert Delete iort Categories - By:	Label: Subtotal Categories omitted from subtotals: 0	<ul> <li>Below</li> <li>These controls determin whether subtotals apply categories that precede follow them.</li> </ul>
ow ✓ <u>I</u> otal abet: Total	☐ <u>Missing</u> ☐ Emply Values ☐ Categories	☑ Other values <u>f</u> ound when data are scanned
100 <u>0</u> 0		

• 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	25 to 34	627	29.8%
category	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.)

Totals and Subtotals for Categorical Variables

# **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.

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

# **Totals for Nested Tables**

Figure 5-5

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

📲 Custom Tables					_ 🗆
Table   Titles   Test Stati	stics Options				
•		ſ	Normal B	Compact	Layers
🚺 Confidence i 🔺					Count
Confidence i				25 to 34	nnnn
Confidence i				35 to 44	nnnn
Country of fa		Male	Age category	45 to 54	nnnn
Favor or opp				55 to 64	nnnn
🚺 Gender	Gender			Total	nnnn
General hap	Gender			25 to 34	nnnn
Get news fro				35 to 44	nnnn
\$		Female	Age category	45 to 54	nnnn
Categories:				55 to 64	nnnn
📕 Male				Total	nnnn
E Female	•				
r Define ───	Sun	nmary Statistics			
N <sub>% Summary Statistics.</sub>	. <u>P</u> os	sition: Columns	· ·	Hide Categ	jory Position:
82 Categories and Tot.		urce: RomVa	riables 💌	Defa	ult 💌
	<u> </u>	urce. Now va			
	OK	<u>P</u> aste	<u>R</u> ese	t Cano	cel Help

• Click OK to create the table.

#### Figure 5-7

Age category totals within Gender categories

				Count	Percent
Gender	Male	Age	25 to 34	276	29.3%
		category	35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age	25 to 34	351	30.1%
		category	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.

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

Figure 5-8 Grand totals for a nested table

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.

×

<b>Figure 5-9</b> Layer variable in ta	able	builder				
Se Custom Tables						_ 0
Table   Titles   Test Statis	tics   1	Options ]				
0				🛄 Norr	mal 🗄 Comp	oact 🔽 Layers
Confidence i         Confidence i			Co	lumns		Layers
Confidence i				Count	Percent	
Favor or opp			25 to 34	nnnn	nnnn	
Gender	W5		35 to 44	nnnn	nnnn	– Laver Output – – – –
General hap	ŭ	Age category	45 to 54	nnnn	nnnn	Show each
Categories:			55 to 64	nnnn	nnnn	Category as a
			Total	nnnn	nnnn	
Female						Show each combi <u>n</u> ation of categories as a layer
_ Define		Summary 9	Statistics			
N <sub>%</sub> <u>S</u> ummary Statistics		Position:	Columns		r <u>∏</u> <u>H</u> ide	Category Position:
SS <u>C</u> ategories and Total	s	S <u>o</u> urce:	Row Variables	-	·	Default 💌
			OK	Paste	<u>R</u> eset	Cancel Help

*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

Layer	Gender Total	-		
	Gender Male			Percent
Age	Gender Female	:76	}	29.3%
category	Gender Total	:09	)	32.8%
	45 to 54	221		23.5%
	55 to 64	136	3	14.4%
	Total	942	2	100.0%

## **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
--

ispla <u>y</u>		Exclude:
Value(s)	Label	
2.00	25 to 34	· •
3.00	35 to 44	- L
€ 1.003.00	Subtotal < 45	
4.00	45 to 54	
5.00	55 to 64	
6.00	65 or older	
¥.006.00	Subtotal 45+	
Subtotals	Label: Subtotal 45+ Categories omitted fr subtotals: 0	Totals and Subtotals Appe
Delete	Order; Ascending	These controls determin whether subtotals apply categories that precede follow them.
w Total abel: Total	☐ <u>Missing</u> ☐ Empty Values ☐ Categorie	s Other values found when data are scanned

*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	Less than 25	242
category	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

Categories and Totals	\$	×
Selected Variable: Age	category	
		Fueluder
Display		
Value(s)	Label	Less trian 20
2.00	25 to 34	
3.00	35 to 44	
€ 2.003.00	Subtotal < 45	•
4.00	45 to 54	
5.00	55 to 64	
6.00	65 or older	
€ 4.006.00	Subtotal 45+	
Insert Delete Sort Categories — By:	Labet: Subtotal 45+ Categories omitted from subtotals: 0	Above     Below     These controls determine     whether subtotals apply to     categories that precede or     follow them.
Show I Iotal Label: Total	☐ <u>Missing</u> ☐ Emply Categories	☑ Other values <u>found</u> when data are scanned
		Cancel <u>H</u> elp

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.

# 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).

104

## **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

Section Tables					_ <b>D</b> ×
Table Titles Test Statistics Or	otions				
1			III Norma		act 🗖 Lauers
Age category				Columns	
🗿 Belief in life after death	_				
Born in this country				Count	
Confidence in banks		Confidence in	A great deal	nnnn	
Confidence in educati		financial	Only some	nnnn	
Confidence in major c		institutions	Hardly any	nnnn	- 11
Confidence in press		Confidence in	A great deal	nnnn	-
🚺 Confidence in television		education	Only some	nnnn	- 11
Country of family origin			Hardly any	nnnn	
		Confidence in	A great deal	nnnn	-
Categories:	0WS	major companies	Unly some	nnnn	-
📕 A great deal 📃 🔺	ŭ		Hardly any	nnnn	-
💶 Only some 📃		Confidence in	A great deal	nnnn	
📕 Hardly any 🔄 🔄		medicine	Unly some	nnnn	
Define	_ Sur	nmary Statistics-			
№ <sub>%</sub> <u>S</u> ummary Statistics	Po	sition: Columns	•	<u> </u>	Category Position:
SS Categories and Totals	Se	urce: Row Var	iables 💌		Default 💌
	(	эк <u>е</u>	aste	<u>R</u> eset	Cancel Help

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  $\times$  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

Section Tables						_ 🗆 🗙
Table Titles Test Statistics 0	otions					
			Norma	al 🗄 Compa	ct 🗋	Layers
Age category Age of respondent Belief in life after death Born in this country Confidence in banks	Π			Columns		
Confidence in educati			A great deal	Only some	Hardly any	
Confidence in medicine		Confidence i	nnnn	nnnn	nnnn	
🚺 Confidence in press	y)	Confidence i	nnnn	nnnn	nnnn	
Confidence in television	Sow	Confidence i	nnnn	nnnn	nnnn	
Country of family origin	1	Confidence i	nnnn	nnnn	nnnn	
÷		Confidence i	nnnn	nnnn	nnnn	
Categories:		Confidence i	nnnn	nnnn	nnnn	
A great deal Only some Hardly any						
Define	_ ⊂ Sur	mmary Statistics-				
N <sub>%</sub> <u>S</u> ummary Statistics	Po	sition: Columns	•	🔽 Hide 🛛	Category Positio	n:
S Categories and Totals	Sg	gurce: Row Var	iables 💌		Row Labels in (	Columr 🔻
	(	эк <u>р</u>	aste	<u>R</u> eset	Cancel	Help

• Click OK to create the table.

106

## 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* 

Summary Statistics: Categori	cal Variables				×
Selected Variable: Confidence in	hanks financial institution	\$			
Statistics:	Display:	·			
Count	Statistic	Label	Format	Decimals	
Unweighted Count	Row %	Row %	nnnn.n%	1	
Column %					-
Subtable % 🚽					
Custom Summary Statistics	s for Totals and Subtotals				
Statistics:	Display:				
Unweighted Count 🔺	Statistic	Label	Format	Decimals	
Row %	Count	Count	nnnn	0	
Table %					_
Subtable %					Ť
			- 1		_
	Apply to Selection	Apply to <u>All</u>	Close	<u>H</u> elp	

 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

Custom Tables							_ 🗆 🗙	
Table Titles Test Statistics 0	otions	1						
0				🖽 Norma	l 🗄 Compac	a 🗋	Layers	
Age category Age of respondent Belief in life after death Born in this country Confidence in banks					Columns			
Confidence in educati		Γ		A great deal	Only some	Hardly any		
Confidence in major c		f	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%		
Confidence in press	0	Confidence i Confidence i Confidence i		nnnn.n%	nnnn.n%	nnnn.n%		
Confidence in television	XOX.			nnnn.n%	nnnn.n%	nnnn.n%		
Country of family origin	1			nnnn.n%	nnnn.n%	nnnn.n%		
			Confidence i	nnnn.n%	nnnn.n%	nnnn.n%		
Categories:		L	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%		
A great deal								
Define	Su	umi	mary Statistics -					
N <sub>%</sub> Summary Statistics	P	osj	tion: Columns	-	₩ Hide C	Category Position	r:	
SS <u>C</u> ategories and Totals	Source: Row Variables							
		0	K P	aste <u>F</u>	<u>]</u> eset	Cancel	Help	

*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

💏 Custom Tables					
Table Titles Test Statistics 0	otions				
1			Morma	al 🔗 Compa	ct 🗖 Lavers
<ul> <li>1st mentioned countr</li> <li>2nd mentioned countr</li> <li>3rd mentioned countr</li> </ul>			(	Columns	
Age category	$\square$	Could a see in	A great deal	nnnn.n%	
Age of respondent		banks_	Only some	nnnn.n%	
Belief in life after death		financial	Hardly any	nnnn.n%	
Born in this country		institutions	Total	nnnn.n%	
Confidence in Banks		Confidence in education	A great deal	nnnn.n%	
Confidence in major c			Only some	nnnn.n%	
			Hardly any	nnnn.n%	
÷		Confidence in	A great deal	nnnn.n%	
Lategories:	50	major	Only some	nnnn.n%	
No variables selected	NO.	companies	Hardly any	nnnn.n%	
	14		A great deal	nnnn.n%	
		Medicine	Only some	nnnn.n%	-
- Define	⊢Sur	nmary Statistics-			
N <sub>%</sub> Summary Statistics	Po	sition: Columns	•	✓ Hide	Category Position:
8 Categories and Totals	Sg	jurce: Row Va	iables 🔻		Default 👻
	(	эк <u>е</u>	aste	<u>R</u> eset	Cancel Help

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

Custom Tables						_ 🗆 🗡	
Table Titles Test Statistics 0	ptions )						
			Ē	Normal 🔡	Compact	Lavers	
Age category Age of respondent Belief in life after death Born in this country	Π			Columns			
Confidence in educati			A great deal	Only some	Hardly any	Total	
Confidence in major c		Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	nnnn.n%	
Confidence in medicine	10	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	nnnn.n%	
Confidence in press     Confidence in television	MOX	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	nnnn.n%	
Country of family origin		Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	nnnn.n%	
¢		Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	nnnn.n%	
Categories:		Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	nnnn.n%	
A great deal	•						
Define           N <sub>%</sub> Summary Statistics           S           Categories and Totals	Sur Po Sg	Summary Statistics Position: Columns Sgurce: Row Variables					
		ОК	Paste	<u>R</u> eset	Cancel	Help	

• 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 a⊓y	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.

Nested variables with shared categories

💼 Custom Tables							_ [	⊐ ×				
Table Titles Test Statistics Op	tions											
<u>م</u>					Norm <u>a</u> l 🔒 I	Co <u>m</u> pact	Layer:	s				
🚺 Confidence in medicine 📥		Columns						-				
Confidence in press												
Country of family origin					A great deal	Only some	Hardly any					
Favor or oppose deat				Confidence i	nnnn.n%	nnnn.n%	nnnn.n%					
🚺 Gender		Male Gender		Confidence i	nnnn.n%	nnnn.n%	nnnn.n%					
General happiness			Mala	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%					
Get news from news			Gender		ľ	IV	Male	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%	
Get news from newsp	12				Confidence i	nnnn.n%	nnnn.n%	nnnn.n%				
🚺 Get news from radio	ŝ				Confidence i	nnnn.n%	nnnn.n%	nnnn.n%				
Get news from televisi			Confidence i	nnnn.n%	nnnn.n%	nnnn.n%						
Happiness of marriage				Confidence i	nnnn.n%	nnnn.n%	nnnn.n%					
Categories:			Female	Confidence i	nnnn.n%	nnnn.n%	nnnn.n%					
📕 Male				Confidence I	nnnn.n%	nnnn.n%	nnnn.n%					
📕 Female				Confidence i	nnnn.n%	nnnn.n%	nnnn.n%					
	-			Confidence I	1000.026	1000.02%	101011.026	-				
J JL	•						•					
Define	_ Sur	nmary Statistics					-					
№ <sub>%</sub> <u>S</u> ummary Statistics	Po	sition: Column	s		▼ ▼ Hid	e Category	Position:					
SS Categories and Totals	Sg	urce: Row Variables						-				
			OK	<u>P</u> aste	<u>R</u> eset	Cancel	Help	]				

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 6-11

112

#### Figure 6-12

Another example of results you probably do not want

Custom Tables							- 🗆 🗙		
Table   Titles   Test Statistics   Op	tions								
				Ē	Normal 🔡	Compact	Lavers		
						<u></u>			
Confidence in medicine		Columns							
Confidence in television									
Country of family origin	$\square$		A		Male	nnnn.n%			
🛐 Favor or oppose deat			A great deal	Gender	Female	nnnn.n%			
🚺 Gender		Canedanaa in	Outo energy	<b>C</b> 1	Male	nnnn.n%			
General happiness		banks_	in Loniy some	Gender	Female	nnnn.n%			
Get news from internet		financial	Hardlu anu	Gandar	Male	nnnn.n%			
Get news from newsp		insuluions	rialuy ary	denuel	Female	nnnn.n%			
🐻 Get news from radio			Total	Gender	Male	nnnn.n%			
🛐 Get news from televisi			1000	dondor		nnnn.n%			
Happiness of matriage			A great deal			nnnn.n%			
Categories:		Confidence in	Only some			nnnn.n%			
Male		education	Hardly any			nnnn.n%			
Female			Total			nnnn.n%			
			A great deal			nnnn.n%			
	3M.0	maior	Only some			nnnn.n%	-		
- Define	⊢ Su	mmarv Statistics-							
N <sub>%</sub> Summary Statistics	Po	sition: Columns			▼ I⊽ Hid	e Category F	<sup>b</sup> osition:		
		ejden. Columno				Default	~		
55 Lategories and Totals	Sg	purce: Row Var	iables		•				
				1 m					
			OK	<u>P</u> aste	<u>R</u> eset	Cancel	Help		

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.

## **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*

Summary Statistics: Categor	ical Variables				×
Selected Variable: Age categor					
	х 				
Statistics:	<u>D</u> isplay:				
Unweighted Count	Statistic	Label	Format	Decimals	
Row %	Count	Count	nnnn	0	<b></b>
Lable %	Column %	Column %	nnnn.n%	1	_
Laver %					×
L     L     Custom Summary Statistic	cs for Totals and Subtotals				
Statistics:	Displa <u>y</u> :				
Unweighted Count 🔺	Statistic	Label	Format	Decimals	
Row %	Count	Count	nnnn	0	
Lolumn %					
Subtable %					•
	Applu to Selection		Close	Halo	1
	Apply to Selection 1		CI028	Teih	

- 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.

116

#### Figure 7-2

Nested categorical variables

Custom Tables							_ [	]		
able Titles Test Statistics O	ptions ]									
				Γ	Normal	Compact	Lavers			
Confidence in medicine										
Confidence in medicine		Columns								
Confidence in television										
Eavor or oppose deat						Count	Column %			
👸 Gender			Less them 25	Get news from	No	nnnn	nnnn.n%			
🛐 General happiness			Less than 20	internet	Yes	nnnn	nnnn.n%			
Get news from internet			25 to 34	Get news from	No	nnnn	nnnn.n%			
Get news from news			2010/04	internet	Yes	nnnn	nnnn.n%			
Get news from newsp	5	Age category	35 to 44 45 to 54	Get news from	No	nnnn	nnnn.n%			
	Į Š			internet	Yes	nnnn	nnnn.n%			
Categories:				45 to 54	Get news from	No	nnnn	nnnn.n%		
No No				internet	Yes	nnnn	nnnn.n%			
Yes 🗧				Get news from	No	nnnn	nnnn.n%			
				internet	Yes	nnnn	nnnn.n%			
			65 or older	Get news from	No	nnnn	nnnn.n%			
				internet	Yes	nnnn	nnnn.n%			
- Define	⊢ Sur	mmary Statistics-								
N <sub>%</sub> Summary Statistics	Po	sition: Columns	:		- <b>-</b>	Hide Categ	gory Position:			
••			de la la c			- Defa	ault	•		
oo Lategories and Lotals	50	purce: Row Va	riadies		•					
			OK	<u>P</u> aste	<u>B</u> ese	et Can	cel Help			

- 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

Custom Tables							-		
able Titles Test Statistics O	otions )								
1 1 1				F	Normal F	3 Compact	🖒 Lave	rs	
Confidence in medicine				ј <del>.</del>					
Confidence in press		Columns							
🐻 Confidence in television 🔄	_				Get news fr	om internet			
Country of family origin				N	0	Ye	es		
Favor or oppose deat				Count	Column %	Count	Column %		
General happiness			Less than 25	nnnn	nnnn.n%	nnnn	nnnn.n%		
Get news from internet		Age category	25 to 34	nnnn	nnnn.n%	nnnn	nnnn.n%		
🚺 Get news from news			35 to 44	nnnn	nnnn.n%	nnnn	nnnn.n%		
📔 Get news from newsp 🔪			45 to 54	nnnn	nnnn.n%	nnnn	nnnn.n%		
	DWS		55 to 64	nnnn	nnnn.n%	nnnn	nnnn.n%		
Categories:	æ		65 or older	nnnn	nnnn.n%	nnnn	nnnn.n%		
No Yes									
Define	⊢Sur	mmary Statistics-							
N <sub>%</sub> Summary Statistics	Po	sition: Columns	:		•	Hide Categ	gory Position:		
S Categories and Totals	Sgurce: Row Variables								
			ОК	<u>P</u> aste	<u>R</u> ese	t Can	cel Hel	р	

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

Custom Tables						_ [	
able   Titles   Test Statistics   Op	otions ]						
				Morm <u>a</u> l	Co <u>m</u> pact	Layer:	s
Get news from televisi				Columns	1		]
🛐 Have gun in home					Get news fr	om internet	
Highest degree					No	Yes	
Highest year of school					Mean	Mean	
Highest year school c				Less than 25	nnnn	nnnn	
Highest year school c				25 to 34	nnnn	nnnn	
🔗 Hours per day watchi		Hours per day watching TV	ay Age category	35 to 44	nnnn	nnnn	
How get paid last week	12			45 to 54	nnnn	nnnn	
	No.K			55 to 64	nnnn	nnnn	
Categories:				65 or older	nnnn	nnnn	
Define	Sun	nmary Statistics-					
N <sub>%</sub> Summary Statistics	Pos	sition: Columns		<b>▼</b> [	<u>H</u> ide Cat	egory Position:	
SS <u>C</u> ategories and Totals	So	urce: Row Va	iables	~	De	fault	•
		OK	Pa	ste <u>B</u> e	eset Ca	ncel Help	5

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

Summary Statistics: Scale Variables										
Selected Variable: Hours per d	ay watching TV									
Statistics:	<u>D</u> isplay:									
Count 🔺	Statistic	Label	Format	Decimals						
Unweighted Count 💻	Mean	Mean	nnnn.nn	2						
Maximum					-					
Minimum										
	Apply to Selection	Apply to <u>A</u> ll	Close	<u>H</u> elp						

► Click Apply to Selection.

## Figure 7-6

Scale summary statistic with two decimals

Custom Tables						_ 🗆 >
Table Titles Test Statistics 0	otions					
				Normal	🗄 Compact	Layers
Get news from televisi 🔺				Columns		
📔 Have gun in home				Hours per day	y watching TV	1
Highest degree				Get news f	rom internet	
Highest year of schoo				No	Yes	]
Highest year school c				Mean	Mean	
Highest year school c			Less than 25	nnnn.nn	nnnn.nn	
Hours per day watchi			25 to 34	nnnn.nn	nnnn.nn	
📔 How get paid last week 🖵	10		35 to 44	nnnn.nn	nnnn.nn	
	ŇQ	Age category	45 to 54	nnnn.nn	nnnn.nn	
Categories:	С <u>с</u>		55 to 64	nnnn.nn	nnnn.nn	
No categories (scale variable)			65 or older	nnnn.nn	nnnn.nn	
Define	Sun	nmary Statistics-				
N% Summary Statistics	Pos	sition: Columns		<b>▼</b> [	<u>H</u> ide Ca	itegory Position:
SS Categories and Totals	S <u>o</u>	urce: Column'	Variables	~	D	efault 💌
		OK	<u>P</u> as	te <u>B</u> e	eset C	ancel Help

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 fr	om internet		
		No	Yes		
		Mean	Mean		
Age	Less than 25	3.54	2.12		
category	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

Custom Tables							_ 🗆 🗙	
Table Titles Test Statistics Op	otions							
0				Norm <u>a</u> l	🗄 Compac	t L	<u>Layers</u>	
Country of family origin Favor or oppose deat Gender General happiness	Π			Columns				
Get news from internet				Count				
Get news from newsp		Get news from	No	nnnn				
📔 Get news from radio		internet	Yes	nnnn				
🚺 Get news from televisi		Got nows fr	No	nnnn				
Happiness of marriage	μ,	12	20 CHOME IIII	Yes	nnnn			
¢	No.	Get news from	No	nnnn				
Categories:		newspapers	Yes	nnnn				
No		Get news from	No	nnnn				
Yes		radio	Yes	nnnn				
		Get news from	No	nnnn				
		television	Yes	nnnn				
Define	Sun	nmary Statistics-						
N <sub>%</sub> Summary Statistics	Pos	ition: Columns		▼ [	<u>H</u> ide C	ategory Posit	ion:	
S Categories and Totals	So	urce: Row Var	iables	•		Default	•	
		OK	Pas	ste <u>R</u> e	set	Cancel	Help	

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 popup 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

💏 Custom Tables						_ 🗆 🗙			
Table Titles Test Statistics Op	otions								
0				Norm <u>a</u> l	🗄 Compact	Layers			
Country of family origin Favor or oppose deat Gender General happiness	Π			Columns					
Get news from news				Count	Column %				
Get news from newsp		Get news from	No	nnnn	nnnn.n%				
🗿 Get news from radio		internet	Yes	nnnn	nnnn.n%				
🚺 Get news from televisi	εŋ.	15			Get news fr	No	nnnn		
Happiness of marriage			doctions i	Yes	nnnn				
÷	NO2	Get news from	No	nnnn					
Categories:		newspapers	Yes	nnnn					
No No		Get news from	No	nnnn					
Yes		radio	Yes	nnnn					
		Get news from	No	nnnn					
		television	Yes	nnnn					
_ Define	Sun	nmary Statistics-							
N% Summary Statistics	Pos	sition: Columns		- F	<u>H</u> ide Cal	tegory Position:			
S Categories and Totals	S <u>o</u>	urce: Row Var	iables	•	De	efault 💌			
		OK	<u>P</u> a:	ste <u>R</u> e	set Ca	ancel Help			

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

👷 Custom Tables						_ 🗆 🗙	
Table   Titles   Test Statistics   Op	tions						
				Normal	😑 Compact		
Country of family origin				Jum			
Eavor or oppose deat				Columns			
Gender	_						
🛐 General happiness 🔤 🔤	$\square$						
Get news from internet				Count	Column %		
Get news from news		Got nouse from	No	nnnn	nnnn n%		
Get news from radio		internet	Yes	nnnn	nnnn.n%		
Get news from televisi			No	nnnn	nnnn.n%		
🚺 Happiness of marriage 🔤		10	10	Get news fr	Yes	nnnn	nnnn.n%
¢	WO.	Get news from	No	nnnn	nnnn.n%		
Categories:	L.	newspapers	Yes	nnnn	nnnn.n%		
💶 No		Get news from	No	nnnn	nnnn.n%		
E Yes		radio	Yes	nnnn	nnnn.n%		
		Get news from	No	nnnn	nnnn.n%		
		television	Yes	nnnn	nnnn.n%		
l l l							
Define	_ Sun	nmary Statistics-					
™ <sub>%</sub> <u>S</u> ummary Statistics	Pos	sition: Columns			<u>H</u> ide	tegory Position:	
SS Categories and Totals	So	urce: Row Var	iables	•	De	efault 💌	
		OK	Pa	ste Re	set Ca	ancel Help	

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

124

## **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 dropdown 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

Summary Statistics: Categori	cal Variables				×
Selected Variable: Confidence in	press				
S <u>t</u> atistics:	Display:				
Count 🔺	Statistic	Label	Format	Decimals	
Unweighted Count 🗖	Column %	Column %	nnnn.n%	1	
How %					-
Subtable %					
🕞 🔽 Custom Summary Statistics	s for Totals and Subtotals				
Statistics:	Display:				
Count 🔺	Statistic	Label	Format	Decimals	
Unweighted Count -	Mean	Mean	nnnn.nn	2	
Column %					_
Table %					Ť
		·····			- 1
	Apply to Selection		LIOSE	Help	

• 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

🚜 Custom Tables						_ 🗆 🗙	
Table Titles Test Statistics Op	tions						
0			🖽 No	ormal 🗄 Com	pact (	Layers	
Belief in life after death	Π			Columns			
Confidence in medicine				Column %	Mean		
👔 Confidence in press			A great deal	nnnn.n%			
🚺 Confidence in television		Confidence in	Only some	nnnn.n%			
Country of family origin	2	o press	Hardly any	nnnn.n%			
Favor or oppose deat	NOX	YOX.	MOX	Total		nnnn.nn	
¢ <u> </u>			A great deal	nnnn.n%			
Categories:		Confidence in	Only some	nnnn.n%			
📕 A great deal		television	Hardly any	nnnn.n%			
Only some			Total		nnnn.nn		
Hardly any Missing Values							
Define N <sub>%</sub> Summary Statistics % Categories and Totals	-Sum Pos S <u>o</u> r	mary Statistics- ition: Columns urce: Row Var	iables	→  → <u>H</u> ide	Category Pos Default	ition: T	
		OK	Paste	<u>R</u> eset	Cancel	Help	

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

📲 Custom Tables						- 🗆 🗙
Table Titles Test Statistics Op	tions					
1st mentioned country			III N	Columns	opact D	
2nd mentioned countr     3rd mentioned countr     Age category     Age of respondent	$\square$					
Belief in life after death			A great deal	Column %	nnnn.n%	
Confidence in banks		Confidence in	Only some	Column %	nnnn.n%	
Confidence in educati	pre	press	Hardly any	Column %	nnnn.n%	
🚺 Confidence in major c 🔤	0W5		Total	Mean	nnnn.nn	
	æ	ž	A great deal	Column %	nnnn.n%	
Categories:		Confidence in	Only some	Column %	nnnn.n%	
		television	Hardly any	Column %	nnnn.n%	
NO Vanabios seiscieu			Total	Mean	nnnn.nn	
Define	_ Sun	nmary Statistics-			1	
№ <sub>%</sub> <u>S</u> ummary Statistics	Pos	sition: Rows		▼ <u>∏</u> ide	Category Positio	on:
SS Categories and Totals	So	urce: Row Var	iables	•	Default	•
		OK	<u>P</u> aste	<u>R</u> eset	Cancel	Help

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

Options			×
Data	Currency	) Se	cripts
General Viewer Draft Viewer	Output Labels	Charts Interactive	Pivot Tables
Outline Labeling         Variables in item labels shown as:         Labels         Variable values in item labels shown as:         Labels         Pivot Table Labeling         Variables in labels shown as:         Labels         Variables in labels shown as:         Labels         Variable values in labels shown as:         Variable values in labels shown as:         Variable values in labels			
	ОК С	ancel <u>Apply</u>	Help

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



## 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).

132

## **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.

Stacked	scale	variables	in	table	builder

<b>#8</b> Custom Tables	- 🗆 -
Custom Tables         Table       Titles       Test Statistics       Op         Variables       Highest degree       Image: Comparison of the statistics       Op         Highest year of school       Highest year of school co       Image: Highest year school co       Image: Highest year school co         Highest year school co       Highest year school co       Image: Highest year school co       Image: Highest year school co         Hours per day watching       How get paid last week       Image: How get paid last week       Image: How get paid last week         How often does respont       Image: How often does respont       Image: How get paid last week       Image: How often does respont         Image: Labor force status       Image: How often does respont       Image: How often does respont       Image: How often does respont         Image: Labor force status       Image: How often does respont       Image: How often does respont       Image: How often does respont         Image: Labor force status       Image: How often does respont       Image: How often does respont       Image: How often does respont         Image: Labor force status       Image: How often does respont       Image: How often does respont       Image: How often does respont         Image: Labor force status       Image: How often does respont       Image: How often does	tions
N <sub>%</sub> Summary Statistics	Sournery Statistics         Position:       Category Eosition:         Source:       Row Variables
	OK Paste Reset Cancel Help

Figure 8-1
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 dropdown list of formats.
- ► In the Decimals cell, type 1.
- Make the same changes for the mean in the Display list.

#### Figure 8-3

Maan and	modion	aplantad	in Cum	man Cta	tiation	lialaa	hav
iviean anu	Ineulan	Selecteu	iii Suiii	illaly Sta	1151165 6	iiai0g	DUX

Summary Statistics: Scale	e Va	riables				×
Selected Variable: Highest y	ear c	f school completed				
Statistics:		<u>D</u> isplay:				
Count 🔺		Statistic	Label	Format	Decimals	
Unweighted Count 💻		Mean	Mean	nnnn.n	1	
Maximum		Median	Median	nnnn.n	1	
Minimum Missing						•
Missing V						
		Apply to Selection	Apply to <u>All</u>	Close	<u>H</u> elp	

- 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.

136

#### 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

Custom Tables					-	. 🗆 🗙
Table Titles Test Statistics Opt	ions	ļ	🔝 Norm <u>a</u> l	E Compact	C La	yers
<ul> <li>Ist mentioned country</li> <li>2nd mentioned country</li> <li>3rd mentioned country</li> <li>Age category</li> <li>Age of respondent</li> <li>Belief in life after death</li> <li>Born in this country</li> <li>Confidence in banks &amp; f</li> </ul>			<u> </u>	<u>o</u> lumns		
Confidence in education			Mean	Median	Missing	
Confidence in major co	(OWS	Age of re Highest up	nnnn.n	nnnn.n		
	<u> </u>	- Hours per d	nnnn.n	nnnn.n	nnnn	
Categories: No variables selected						
Define	Sum	mary Statistics —		Cote	aoru Position:	
<sup>14</sup> % <u>S</u> ummary Statistics 8€ <u>C</u> ategories and Totals	Posj So <u>u</u>	tion: Columns (rce: Row Varial	▼ 「	Hide Def	gory <u>Costtor</u> t: ault	Y
	OK	<u>P</u> aste	e <u>R</u> es	et Car	ncel H	lelp

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

👫 Custom Tables					- 🗆 🗙
Table   Titles   Test Statistics   Opti	ons				
Variables			Norm <u>a</u> l	岩 Compact	Layers
1st mentioned country     2nd mentioned country     3rd mentioned country     Age category     Age of respondent     Belief in life after death			C	<u>o</u> lumns	
Born in this country		Age of	Mean	nnnn.n	
Confidence in banks & f		respondent	Median	nnnn.n	
Confidence in major co	10	Highest ve	Mean	nnnn.n	
🚺 Confidence in medicine 🔤	NOX		Median	nnnn.n	
		Hours per dau	Mean	nnnn.n	
Categories:		watching TV	Median	nnnn.n	
No variables selected			Missing	nnnn	
Define	Sum	mary Statistics —		Coto	range Regition:
™% <u>S</u> ummary Statistics	Pos	ition: Rows		Hide Cate	goly Fosition:
88 <u>C</u> ategories and Totals		urce: Row Varia	bles 🔻	Det	auit
	OK	Paste	e <u>R</u> es	et Car	ncel Help

• 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	Mean	13.3
completed	Median	13.0
Hours per day	Mean	2.9
watching TV	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 sum	maries us	sing a categ	gorical colur		
		Gender			
		Male	Female		
Age of respondent	Mean	44.6	46.3		
	Median	42.0	43.0		
Highest year of school	Mean	13.4	13.2		
completed	Median	13.0	13.0		
Hours per day	Mean	2.8	2.9		
watching TV	Median	2.0	2.0		

Missing

variable

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.

282

213

#### 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

Custom Tables								_ 0															
able   Titles   Test Statistics   O	ptions																						
<u>/</u> ariables					III N	orm <u>a</u> l 🔒 Co <u>r</u>	npact 🛙	<u>Layers</u>															
Confidence in medicine           Image: Confidence in press					C <u>o</u> lu	mns		<b>^</b>															
Confidence in television							Gen	der															
Country of family origin		L.					Male	Fema															
Country of origin multi					Age of	Mean	nnnn.n	nt															
👩 Favoi oi oppose deatri			Get news from internet		respondent	Median	nnnn.n	nr															
General happiness					Linkertur	Mean	nnnn.n	nr															
Get news from internet					No	Highest ye	Median	nnnn.n	nr														
👔 Get news from news ma							Mean	nnnn.n	nr														
🚺 Get news from newspa 🦕		0			Hours per day	Median	nnnn.n	nr															
				Get news from internet	Get news from internet	Get news from internet	Get news from internet	Get news from internet	Get news from internet	Get news from internet	Get news from internet	Get news from		waterning i v	Missing	nnnn							
Categories:	ll l'											Age of	Age of	Mean	nnnn.n	nr							
No 🗧																				respondent	Median	nnnn.n	nı
📕 Yes										Highest up	Mean	nnnn.n	n										
				Yes	riignest ye	Median	nnnn.n	nr															
						Mean	nnnn.n	nr															
					Hours per day watching TV	Median	nnnn.n	nr															
			li i			L.e																	
- Define		Imm	aru Statistico																				
N <sub>%</sub> <u>S</u> ummary Statistics	P	osįti	on: Rows			▼ □ <u>H</u> ide	Category <u>P</u> os	ition:															
S Categories and Totals	s	io <u>u</u> n	ce: Row Varia	bles		Ŧ	Default	•															
				OK	Paste	<u>R</u> eset	Cancel	Help															

#### Figure 8-13

Wrong: Only one scale variable nested within the categorical variable

Custom Tables							_ 🗖
Table   Titles   Test Statistics   Op	otions ]						
<u>V</u> ariables	Þ			I N	lorm <u>a</u> l 🔚 Co <u>r</u>	npact [[	<u>Layers</u>
Confidence in medicine				C <u>o</u> lu	umns		
Confidence in television						Gen	der
Country of family origin						Male	Female
Enver or oppose death		Age of	Mean			nnnn.n	nnnn.
Gender		respondent	Median			nnnn.n	nnnn.
General happiness			No	Hiskestus	Mean	nnnn.n	nnnn.
🗿 Get news from internet		Get news from	NU	Highest ye	Median	nnnn.n	nnnn.
📔 Get news from news ma		internet	internet Yes Highes	Hisbootus	Mean	nnnn.n	nnnn.
🚺 Get news from newspa 🧉	ψ			Highest ye	Median	nnnn.n	nnnn.
÷			Mean			nnnn.n	nnnn.
Categories:		Hours per day	Median			nnnn.n	nnnn.
No No		matering 11	Missing			nnnn	nnr
Define	Sum	mary Statistics —					
N <sub>%</sub> Summary Statistics	Posj	tion: Rows			▼ <u>∏</u> <u>H</u> ide	Category Posi	ition:
S Categories and Totals	Soy	rce: Row Varia	bles		-	Default	•
			OK	Paste	<u>R</u> eset	Cancel	Help

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 s	summaries	grouped b	y categorical	' row and	column	variables
		0	,			

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

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

# **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.

146

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

#### Figure 9-1

Variables displayed on canvas pane

Custom Tables							_ 🗆 >
Table   Titles   Test St	atistics	Options					
⊻ariables				Norma	I 🗄 Compac	nt 🗋	Layers
Responde					C <u>o</u> lumns		
🚺 Marital stat						Marital status	
Number of				Married	Widowed	Divorced	Separa
Age of res				Count	Count	Count	Cour
Highest ve			Working full	nnnn	nnnn	nnnn	
Highest ye			Working par	nnnn	nnnn	nnnn	
Highest ye	W5		Temporarily	nnnn	nnnn	nnnn	
🚺 Highest de	2	Labor force	Unemployed	nnnn	nnnn	nnnn	
🚺 Gender 🔍 🖵		status	Retired	nnnn	nnnn	nnnn	
¢			School	nnnn	nnnn	nnnn	
Categories:			Keeping hou	nnnn	nnnn	nnnn	
📕 Married 🔺			Other	nnnn	nnnn	nnnn	
Widowed	•						•
- Define		Summ	hary Statistics				
N <sub>%</sub> Summary Statistic	:s	Positi	on: Columns	-	T Hide	Category <u>P</u> osition	n:
88 <u>C</u> ategories and T	otals	Sour	ce: Column Varial	oles 💌		Default	•
			ОК <u>Р</u>	aste <u>f</u>	<u>R</u> eset	Cancel	Help

• 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

Summary Statistics: Catego	orical Variables				×
Selected Variable: Labor force	e status				
Statistics:	Display:				
Unweighted Count	Statistic	Label	Format	Decimals	
Row %	Count	Count	nnnn	0	
Table %	Column %	Column %	nnnn.n%	1	
Subtable %					•
Layer & Row Malid M %					
Calumn Walked M %					
🕞 🗖 Custom Summary Statis	tics for Totals and Subtotals	:			
Statistics:	Display:				
Unweighted Count 🔺	Statistic	Label	Format	Decimals	
Row % 🗖	Count	Count	nnnn	0	
Table %	Column %	Column %	nnnn.n%	1	
Subtable %	`				-
Layer &					
		1			
	Apply to Selection	Apply to All	Llose	Help	

- 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

Security Custom Tables	_ 🗆 🗙
Table Titles Test Statistics Options	
I Tests of independence (Chi-square) Alpha: 0.05	
Compare column means (t-tests)	
Alpha: 0.00 = Adjust p-values for multiple comparisons (Bonferroni method)	
Compare column proportions (z-tests)	
Alpha: 0.05 📩 🔽 Adjust p-values for multiple comparisons (Bonferroni method)	
Chi-square and column proportions tests apply to tables in which categorical variables exist in both the rows and columns.	
Column means tests apply to tables in which scale variables exist in the rows and categorical variables exist in the columns.	
None of the tests is performed on multiple response variables or tables in which category labels are moved out of their default table dimension.	
OK Paste Reset Cancel	Help

- ► 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	Working full	Count	778	44	295	58	392	
force	time	Column %	57.8%	15.5%	66.1%	62.4%	59.1%	
Status	Working	Count	138	20	35	9	102	
	part-time	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,	Count	13	3	10	0	32	
	laid off	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	Count	200	55	25	13	35	
	house	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	Chi-square	729.242
status	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 <sup>*,1,2</sup>	.000 <sup>*,1,2</sup>		

<sup>\*.</sup> The Chi-square statistic is significant at the 0.05 level.

2. The minimum expected cell count in this sub-table is less than one.

More than 20% of cells in this sub-table have expected cell counts less than 5.

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	-1-
	- 1 -
Male Fema	ale
Marital status Marital s	status
Labor force Chi-square 246.637 54	2.589
status df 28	28
Sig000 <sup>*,1,2</sup> .0	00*.1.2
Highest Chi-square 43.844 10	5.506
degree df 16	16
Sig000*	.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

Custom Tables						_ 🗆 🗙
Table Titles Test S	tatistics	Options				
	•		Normal	🛛 🗄 Co <u>m</u> pact	• D	Layers
Confidenc				Columns		
Age categ			Get news from	m newspapers	1	
Get news f			No	Yes		
🚺 Get news f		Harris and	Mean	Mean		
Get news f		Hours per a	nnnn	nnn	]	
Get news f	ú.					
Example Country of	Sow					
News sour						
¢						
Define		Summary	Statistics			
N <sub>%</sub> <u>S</u> ummary Statisti	DS	Position:	Columr 💌	THide C	ategory Positio	n:
Se Categories and Totals Source: Row V						•
			. 1			
	U	K Pa:	ste R	eset l	Lancel	Неір

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

#### Figure 9-9

Summary Statistics dialog box

Summary Statistics: Scale Variables							
Selected Variable: Hours per day watching TV							
Statistics:	Display:						
Count 🔺	Statistic	Label	Format	Decimals			
Unweighted Count -	Mean	Mean	nnnn.nn	2			
Maximum •					_		
Minimum					Ť		
Missing							
111-1-	1						
	Apply to Selection	Apply to All	Close	Help			

- ► 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

🕮 Custom Tables 📃 🗆 🗙						
Table Titles Test Statistics Options						
Tests of independence (Chi-square)						
Compare column means (I-tests)						
Alpha: 0.05 🔹 🔽 Adjust p-values for multiple comparisons (Bonferroni method)						
Compare column proportions (z-tests)     Alpha: 0.05 * Adjust p-values for multiple comparisons (Bonferroni method)						
Chi-square and column proportions tests apply to tables in which categorical variables exist in both the rows and columns. Column means tests apply to tables in which scale variables exist in the rows and categorical variables exist in the columns.						
None of the tests is performed on multiple response variables or tables in which category labels are moved out of their default table dimension.						
OK Paste Reset Cancel Help						

- ► 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 new newsp	/s from apers
			No	Yes
			(A)	(B)
Labor force	Working full time	Hours per day watching TV	в	
status	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).

#### 156

- 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 new inter	's from net	Get new newsp	/s from apers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force	Working full time	Hours per day watching TV	в		в	
status	Working part-time	Hours per day watching TV	в			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	в			
	Retired	Hours per day watching TV	в			
	School	Hours per day watching TV	в			
	Keeping house	Hours per day watching TV	в			
	Other	Hours per day watching TV	в			

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.

📲 Custom Tables								_ 🗆 🗙
Table   Titles   Test St	atist	ics	Options					
<u>V</u> ariables					Norma	_ E Compac		Layers
Responde						C <u>o</u> lumns		$\equiv$
🚺 Marital stat	6	_					Marital status	
Windows of Number of					Married	Widowed	Divorced	Separa
Age or res					Count	Count	Count	Cour
Highest ye				Working full	nnnn	nnnn	nnnn	
🔗 Highest ye				Working par	nnnn	nnnn	nnnn	
🔗 Highest ye		M2		Temporarily	nnnn	nnnn	nnnn	
🚺 Highest de	d	2	Labor force	Unemployed	nnnn	nnnn	nnnn	
🚺 Gender 🚽		┢	. status	Retired	nnnn	nnnn	nnnn	
÷				School	nnnn	nnnn	nnnn	
Categories:				Keeping hou	nnnn	nnnn	nnnn	
📕 Married 🔺				Other	nnnn	nnnn	nnnn	
Widowed	•							•
_ Define			Summar	y Statistics ——				
N <sub>%</sub> Summary Statistic	:s		Position	Columns	•	<u>∏</u> <u>H</u> ide <sup>(</sup>	Category <u>P</u> osition	n:
•Cotogorios and T	otali			Colores Verial			Default	-
oe Categories and t	otan	s	Source	Column Variat	JIES			
				<u>р</u>	aste <u>F</u>	leset	Cancel	Help

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

Summary Statistics: Catego	rical Variables				×
Selected Variable: Labor force	status				
Statistics:	Displau				
Count 🔺	Statistic	Label	Format	Decimals	
Unweighted Count	Column %	Column %	nnnn.n%	1	
Row %					-
Subtable %					
Layer %					
🛛 🔽 🗖 Custom Summary Statist	ics for Totals and Subtotals				
Statistics:	Display:				
Count	Statistic	Label	Format	Decimals	
Bow 2	Column %	Column %	nnnn.n%	1	
Table %					-
Subtable %					
Layer %					
	Apply to Selection	Apply to All	Close	Help	1
	Apply to belection	- 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10	0,000	Trop	

- 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

🚜 Custom Tables	
Table Titles Test Statistics Options	
Tests of independence (Chi-square)	
Alpha: 0.05	
Compare eduran means () tests)	
Alpha: 0.05 👘 🗹 Adjust p-values for multiple comparisons (Bonferroni method)	
Compare column proportions (starts)	
Alpha: 0.05 🚔 🔽 Adjust p-values for multiple comparisons (Bonferroni method)	
	_
Chi-square and column proportions tests apply to tables in which categorical	
variables exist in both the rows and columns.	
Column means tests apply to tables in which scale variables exist in the rows and categorical variables exist in the columns	
Categorical variables exist in the columns.	
None of the tests is performed on multiple response variables or tables in which category labels are moved out of their default table dimension.	
OK Paste Reset Cancel H	elp
	-//,

- ► 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	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%		
force status	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	Working full time	в		AВ	в	в	
force	Working part-time					АВС	
olaido	Temporarily not working						
	Unemployed, laid off					AВ	
	Retired	E	ACDE	E			
	School					АВС	
	Keeping house	CE	CE		CE		
	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, 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

162

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	в		в	в	в
			Working part-time					А
			Temporarily not working					
			Unemployed, laid off					А
			Retired	E	ACDE	E		
			School					AC
			Keeping house					
			Other				А	
	Female	Labor force status	Working full time	в		AB	в	в
			Working part-time	в				в
			Temporarily not working					
			Unemployed, laid off					А
			Retired	E	ACDE	E		
			School					ABC
			Keeping house	CE	CE		С	
			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	Working full time	в		в	B	в
		force status	Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	ACDE	E		
			School					AC
			Keeping house					
			Other				А	
	Female	Labor	Working full time	в		АВ	в	в
		force	Working part-time	в				в
		olaido	Temporarily not working					
			Unemployed, laid off					A
			Retired	E	ACDE	E		
			School					ABC
			Keeping house	CE	CE		с	
			Other					
Highest	Highest LT High school				ACE			
degree	High school							
	Junior college			в		в		в
	Bachelor	в				в		
	Graduate	в						

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.

# 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

Multiple Response Sets 🗙
Set Definition <ul> <li>Ist mentioned co</li> <li>2nd mentioned co</li> <li>Get news from internet</li> <li>Get news from news</li> <li>Get news from televis</li> <li>Get news from televis</li> <li>Get news from televis</li> <li>Get news from televis</li> </ul> Mult. Response Set:           Substrained co <ul> <li>Get news from reado</li> <li>Get news from televis</li> <li>Get news from televis</li> <li>Get news from televis</li> <li>Car maker, most r</li> <li>Car maker, third car</li> </ul> <ul> <li>Add</li> <li>Change</li> <li>Remove</li> <li>Remove</li> <li>Mame: mitnews</li> <li>Labet</li> <li>News sources</li> </ul> <ul> <li>Add</li> <li>Change</li> <li>Remove</li> <li>Remove</li> </ul>
OK Paste Reset Cancel Help

- ► 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

Variable Infor	nation 🗙	
Name:	news5	
Label:	Get news from internet	
Measurement:	Nominal	
Value Labels:	0 No 🗸	
	0 No	
	1 Yes	

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

Section Tables							_ 🗆 🗙
Table Titles Test Statistics Option	s						
⊻ariables	, i				Norr	mal 🗄 Comp	pact 🗋 Layers
<ul> <li>Highest year school comple</li> <li>Highest year school comple</li> <li>Hours per day watching TV</li> <li>How get paid last week</li> <li>How many grandparents bo</li> <li>How often does responden</li> <li>How often does responden</li> <li>Marital status</li> </ul>			[			C <u>o</u> lumns	
■ News sources ✓ Number of children Image: Number of children (groupe		0.W5			Get news fro Get news fro Get news fro	nnnn nnnn nnnn	
¢		æ		News sources	Get news fro	nnnn	
Categories:					Get news fro Total	nnnn	
Get news from internet     Get news from radio     Get news from television     Get news from news magazines     Get news from newspapers							
Define     P       N% Summary Statistics     P       SS Categories and Totals     P	Sun Po: So	nmar sitior gros	ny 9 n: s:	Statistics Columns Row Variables	•	r <u>H</u> ide	Category <u>P</u> osition: Default
				ОК	Paste	<u>R</u> eset	Cancel Help

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	Get news from internet	867	41.7%
sources	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	Get news from internet	867	41.7%	. 867	. 24.1%
sources	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

						Column
					Column	Responses %
		Count	Column %	Responses	Responses %	(Base: Count)
News	Get news from internet	867	41.7%	867	24.1%	41.7%
sources	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

<b>6 Custom Tables</b> Table Titles Test Statis	tics Options						_	_ 🔲
Variables 🔶					Norm <u>a</u> l	岩 Co <u>m</u> pact		ayers
Confidence i				C <u>o</u> lumns				<b>^</b>
Confidence i					Count	Column %	Responses	d
🚺 Confidence i				Get news fro	nnnn	nnnn.n%	nnnn	
Country of fa				Get news fro	nnnn	nnnn.n%	nnnn	
Eavor or opp		Male	News sources	Get news fro	nnnn	nnnn.n%	nnnn	
Gender		Maio	INCONS SOURCES	Get news fro	nnnn	nnnn.n%	nnnn	
🗿 General hap				Get news fro	nnnn	nnnn.n%	nnnn	
👔 Get news fro 🝟	Gender			Total	nnnn	nnnn.n%	nnnn	
÷				Get news fro	nnnn	nnnn.n%	nnnn	
Categories:				Get news fro	nnnn	nnnn.n%	nnnn	
📕 Male		Female	News sources	Get news fro	nnnn	nnnn.n%	nnnn	yers
📘 Female				Get news fro	nnnn	nnnn.n%	nnnn	-
				Get news fro	nnnn	nnnn.n%	nnnn	
Define	▲	nmary Statistics-						•
N <sub>% Summary Statistics</sub>	s Sc	sition: Columns	riables		▼ [	i <u>H</u> ide Cat De	egory <u>P</u> osition: fault	•
			OK	Pasi	te <u>R</u> es	et Ca	ncel	Help

- 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.

176

### Figure 10-9

Multiple response set nested within a categorical variable

					Column	Column Responses %
		Count	Column %	Responses	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.

Custom Tables								_ 🗆 🗙	
Table   Titles   Test Stati	stics	Options							
<u>V</u> ariables						Normal 🔡 Cor	npact	🗅 Layers 💧	
1st mentione					C <u>o</u> lumns				
Age category						Count			
Age of respo		(	C		Male	nnnn			
🚺 Belief in life a			Liet news fro	Gender	Female	nnnn			
Born in this c			Cot nouse fro		Male	nnnn			
Confidence i			det news ito	Gender	Female	nnnn			
Confidence i			Get news fro	Candar	Male	nnnn			
🚺 Confidence i 🔄		NO2	News sources	dot nome no	Gendel	Female	nnnn		
			Get news fro	Gender	Male	nnnn			
Categories:			dottione lio.	Grondor	Female	nnnn			
No variables selected			Get news fro	Gender	Male	nnnn			
				Grondor	Female	nnnn			
			Total	Gender	Male	nnnn			
					Female	nnnn		-	
Define		- Summary S	tatistics						
N <sub>%</sub> Summary Statistics		Position:	Columns			▼	Category Po	sition:	
88 <u>C</u> ategories and Tota	als	So <u>u</u> rce:	Row Variables			~	Default	•	
				OK	Paste	Reset	Cancel	Help	

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

Custom Tables					_ 🗆 🗙
Table Titles Test Statistics Optio	ns				
<u>V</u> ariables		同	Norm <u>a</u> l 🗄	Co <u>m</u> pact	Layers
1st mentioned country of     2nd mentioned country of     3rd mentioned country of     Age category     Age of respondent	Π		C <u>o</u> lu	imns	
Belief in life after death				Count	
Born in this country			American	nnnn	
Car maker, most recent c	is a		Japanese	nnnn	
Lar maker, most recent cart	NOX	Car maker,	Korean	nnnn	
Categories:	l"L	cars	German	nnnn	
			Swedish	nnnn	
Japanese			Other	nnnn	
Korean					
E German					
J Swedish	Cummo	ru Statistica			
Nex Summary Statistics	Positio	ry Statistics		Le Category	Position:
5 - 2 Juninary Statistics	Fosjuor	i. Columns		Default	-
5 <u>Categories and Totals</u>	Soyrce	e: Row Variab	es 🔻		
C	)K	<u>P</u> aste	<u>R</u> eset	Cancel	

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	American	1938	1938
maker,	Japanese	1327	1327
most	Korean	695	695
cars	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	American	1938	2797
maker,	Japanese	1327	1717
most	Korean	695	760
cars	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.

# **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.

182

### **Tables without Missing Values**

By default, user-missing categories are not displayed in custom tables (and systemmissing 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

📲 Custom Tables								_ 🗆 🗙
Table Titles Test Statistics Optio	ns							
<u>V</u> ariables					Norm	al 🗄 Compa	act 🗋	Layers
Pariable with missing values						C <u>o</u> lumns		
		L.				Count	Column %	1
					Low	nnnn	nnnn.n%	
A	Distra		Variable w	iith	Medium	nnnn	nnnn.n%	
Categories:	L L L	L	missing va	lues	High	nnnn	nnnn.n%	
Low		L.			Total	nnnn	nnnn.n%	
<ul> <li>Medium</li> <li>High</li> <li>Missing Values</li> </ul>								
Define <mark>N<sub>%</sub> S</mark> ummary Statistics	Sum Pos	nmai sitior	ny Statistics n: Columna	s	•	□ <u>H</u> ide	Category <u>P</u> ositio	in:
55 Categories and Totals	So	urce	e: Row Va	ariabl	es 🔻			
		C	IK	E	aste	<u>R</u> eset	Cancel	Help

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	Low	2	28.6%
	Medium	3	42.9%
values	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.

Table preview with missing values category displayed

Section Tables						_ 🗆 🗙				
Table Titles Test Statistics Ontic	nsl									
Variables			Norma	- B Compa	a 🖻	Lavers				
Variable with missing values	<b></b>		juu rooma		~ ~	20,000				
				C <u>o</u> lumns						
				Count	Column %					
			Low	nnnn	nnnn.n%					
	WS	9 - 1 I - M	Medium	nnnn	nnnn.n%					
\$	8	Variable with missing values	High	nnnn	nnnn.n%					
Categories:		-	Missing Valu	nnnn	nnnn.n%					
Low			Total	nnnn	nnnn.n%					
Medium High										
Missing Values										
Define	Summ	ary Statistics	_		Potogory Positio					
™% <u>S</u> ummary Statistics	Pos <u>i</u> ti	on: Columns	•	<u>H</u> ide						
SC Categories and Totals	Source: Row Variables									
OK <u>P</u> aste <u>R</u> eset Cancel Help										
						///				

184

Figure 11-3

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	Low	2	22.2%		
with	Medium	3	33.3%		
missing	High	2	22.2%		
values	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	Low	2	22.2%	28.6%	20.0%		
with	Medium	3	33.3%	42.9%	30.0%		
missing values	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, nonmissing 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.

# 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

📽 Custom Tables							- 🗆 🗙			
Table Titles Test Statistics	Ontio	]								
Vedebles	l obgo	101				omeaat				
1st mentioned co     1     2nd mentioned c	Cglumns									
3rd mentioned co.										
Age category	$\square$			Count	Column %	Mean				
Age of respondent			Less than 25	nnnn	nnnn n%	in our				
🚺 Belief in life after			25 to 34	nnnn	nnnn.n%					
Born in this country	Rows		35 to 44	nnnn	nnnn.n%					
Car maker, most r Car maker, most r Car maker, secon		Age category	45 to 54	nnnn	nnnn.n%					
			55 to 64	nnnn	nnnn.n%					
\$			65 or older	nnnn	nnnn.n%					
Categories:			Total	nnnn	nnnn.n%	nnnn.nn				
No variables selected			A great deal	nnnn	nnnn.n%					
		Confidence in	Only some	nnnn	nnnn.n%					
		television	Hardly any	nnnn	nnnn.n%					
			Total	nnnn	nnnn.n%	nnnn				
]										
Define		Summary Statis	tics				losition			
™% <u>S</u> ummary Statistics		Position: Colu	imns		▼ ☐ <u>H</u> ide		osidon.			
SS Categories and Totals	88 <u>C</u> ategories and Totals				~	Default	· · · · ·			
		Γ	OK	Paste	<u>R</u> eset	Cancel	Help			
		_								

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	1	2-2
rigule		Z-7

Variable View in the Data Editor

🛅 surve	y_sample.sav	/ - SPSS Data E	ditor		_ 🗆	×				
<u>F</u> ile <u>E</u> di	: <u>V</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u> naly	/ze <u>G</u> raph	is <u>U</u> tilities <u>V</u>	<u> </u>					
▆▆▆⊵▫▫▣▙⊵м≝⋭≣₫₨♥⊘										
	Name	Туре	Width	Decimals	Label					
38	contv	Numeric	1	0	Confidence in television					
37	agecat	Numeric	8	2	Age category					
38	childcat	Numeric	8	2	Number of children (group					
39	news1	Numeric	6	0	Get news from newspaper					
40	news2	Numeric	6	0	Get news from news maga					
41	news3	Numeric	6	0	Get news from television					
42	news4	Numeric	6	0	Get news from radio					
43	news5	Numeric	6	0	Get news from internet					
	Data View $\lambda v$	ariable View /	ال ما	•						
				SPSS Process	or is ready	11.				

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

ummary Statistics: Cate <u>c</u>	goric	al Variables				
Selected Variable: (Multiple V	/aria/	bles)				
Chatiatian		Diselaw				
Statistics:		Display:	-		1 1	
Unweighted Count		Statistic	Label	Format	Decimals	
Row %		Count	Count	nnnn	0	
l able %		Column %	_ Column %	nnnn.n%	1	
Subtable %						•
Dave Valid N. %						
- Custon Summaru Stati	istics	for Totals and Subtotals				
Flow Valid N & ▼	istics	for Totals and Subtotals Display:				
Column Valid N &      Column Valid N &	istics	for Totals and Subtotals Display: Statistic	Label	Format	Decimals	
Column Valid N %      ColumN %      ColumN %      ColumN %	istics	for Totals and Subtotals Display: Statistic Count	Label	Format	Decimals	
For wall of X     ▼       Image: Construction of the state o	istics	for Totals and Subtotals Displa <u>y: <b>Statistic</b> Count Column %</u>	Label	nnnn nnnn.%	Decimals	
Forward N &     ▼       Custom Summary Statistics:     Unweighted Count       Bow %     ▲       Table %     Subtable %	istics	for Totals and Subtotals Display: Count Column % Mean	Label Count Column % Mean	nnnn n% Auto	<b>Decimals</b> 0 1	•
Now Valid N &     ▼       Image: Construction of the state o	istics	for Totals and Subtotals Display: Statistic Count Column % Mean	Label Count Column % Mean	Format nnnn nnnn.n% Auto	<b>Decimals</b> 0 1	•
For wall of X     ✓       ✓     Custom Summary Statistics:       Unweighted Count ▲       Row %       Table %       Subtable %       Layer %       Maximum	istics	for Totals and Subtotals Display Statistic Count Column % Mean	Label Count Column % Mean	Format nnnn nnnn.n% Auto	Decimals 0 1	•
For wall of X     ▼       Image: Constraint of the state of	istics	for Totals and Subtotals Display: Statistic Count Column % Mean	Label Count Column % Mean	Format nnnn nnnn.n% Auto	Decimals 0 1	•
For wall of X     ▼       Image: Constraint of the second	istics	for Totals and Subtotals Displa <u>y: Statistic</u> Count Column % Mean	Label Count Column % Mean	Format nnnn nnnn.n% Auto	Decimals 0 1	•

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

Security Tables							_ 🗆 🗙	
Table Titles Test Statistics [	Ontion	sl						
Variablee		*1			Normal 🚍 D	ompact		
				jum i		ompace		
2nd mentioned co	Columns							
3rd mentioned co								
📓 Age category	T [			Count	Column %	Mean		
Age of respondent			Less than 25	nnnn	nnnn.n%			
Belief in life after			25 to 34	nnnn	nnnn.n%			
Car maker, most r			35 to 44	nnnn	nnnn.n%			
📓 Car maker, most r		Age category	45 to 54	nnnn	nnnn.n%			
🚺 📔 Car maker. secon 🔟	WS		55 to 64	nnnn	nnnn.n%			
⊊	2		65 or older	nnnn	nnnn.n%			
			Total	nnnn	nnnn.n%	nnnn.n		
No variables selected			A great deal	nnnn	nnnn.n%			
		Confidence in	Only some	nnnn	nnnn.n%			
		leievision	Hardly any	nnnn	nnnn.n%			
			lotal	nnnn	nnnn.n%	nnnn.n		
Define	9	Summary Statist	ics			_		
N <sub>%</sub> Summary Statistics		Position: Colu	mns		▼ □ Hide	Category E	osition:	
PO Colorestics and Table						Default		
of Lategories and Lotals		Source: How	Variablés					
		Г	01	n. 1	<b>D</b> 1		1	
		L	UK	Paste	<u>H</u> eset	Cancel	Help	

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.)

192

## **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

Summary Statistics: Catego	orical Variables				×
Selected Variable: Age catego	ory				
Statistics:	<u>D</u> isplay:				
Row %	Statistic	Label	Format	Decimals	
Table %	Count	N	nnnn	0	
Subtable % —	Column %	%	nnnn.n	1	
Layer %					
Flow Valid N &					
🕞 🗖 🖸 Custom Summary Statis	tics for Totals and Subtotals				
Statistics:	Display:				
Unweighted Count	Statistic	Label	Format	Decimals	
Row % 🗖	Count	N	nnnn	0	
Table %	Column %	%	nnnn.n	1	
Subtable %	`				
Layer %					
Maximum 💽					
		1 (			
	Apply to Selection	Apply to <u>A</u> ll	Close	Help	
					÷.,

► Then click Apply to Selection.

### Figure 12-6

Table preview with modified summary statistics labels

Custom Tables							_ 🗆 🗙
Table Titles Tract Statistics	Ontio	]					
rabio [ miles ] rescordustics]	option	18					e,
					Normai : 🚍 Lo	ompact	ų Layers
🚺 Have gun in home 📩							
Highest degree							
Highest year of s							
Highest year sch				Hourly	wage	Daily w	age
Highest year sch				N	%	N	%
Hours per day wa			Less than 25	nnnn	nnnn.n	nnnn	nnnn.n
🗿 How get paid last			25 to 34	nnnn	nnnn.n	nnnn	nnnn.n
🔗 How many grand	12	Ane category	35 to 44	nnnn	nnnn.n	nnnn	nnnn.n
🛛 📓 How often does r 🔳	XDW1	Age calegoly	45 to 54	nnnn	nnnn.n	nnnn	nnnn.n
Categories:	1°L		55 to 64	nnnn	nnnn.n	nnnn	nnnn.n
			65 or older	nnnn	nnnn.n	nnnn	nnnn.n
Dailu wage							
Weekly wage							
Monthly salary							
📕 Annual salary	4						
	<u> </u>						
Define		Summary Statis	tics				
№% <u>S</u> ummary Statistics		Position: Colu	imns		▼ ☐ <u>H</u> ide	Category Po	sidon:
SS Categories and Totals		Source: Row	/Variables		•	Default	•
			OK	Paste	<u>R</u> eset	Cancel	Help
		-					

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	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
category	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.)

### 194

### Formatting and Customizing Tables

### Figure 12-8

Custom Tables, Options tab

Section Tables	
Table Titles Test Statistics Options	
Data Cell Appearance	Width for Data Columns
Empty Cells:	C TableLook settings
C <u>B</u> lank	
C <u>I</u> ext:	Minimum: 36
	Maximum: 36
Statistics that Cannot be Computed:	Units: Points
Missing Values for Scale Variables C Maximize use of available data (variable-by-variable deletion)	C Use consistent case base across scale variables (listwise deletion)
Count duplicate responses for multiple categor	y sets
	Paste Reset Cancel Help

• 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	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7	
category	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	Less than 25	91	14.0	None	None	12	9.7	3	2.0	7	3.1	14	7.7	
category	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.

ummary Statistics	: Scale V	ariables				
Selected Variable: H	ours per daj	y watching TV				
Statistics:		<u>D</u> isplay:				
Std. Deviation	<b></b>	Statistic	Label	Format	Decimals	
Sum		Mean	Mean	Auto		
Total N	<b>&gt;</b>	Valid N	Valid N	nnnn	0	
Variance Row % Column %	<b>.</b>					-
		Apply to <u>S</u> election	Apply to <u>All</u>	Close	<u>H</u> elp	

Summary Statistics dialog box for scale variables

- Click Apply to Selection.
- Click the Options tab.

Figure 12-11

▶ 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

🗱 Custom Tables		_ 🗆 🗙
Table Titles Test Statistics Options		
Data Cell Appearance	Width for Data Columns	
Empty Cells: C Zero	C TableLook settings	
C <u>B</u> lank		
	Minimum: 36	
	Ma <u>x</u> imum: 36	
Statistics that Cannot be NA Computed:	Units: Points	
Missing Values for Scale Variables		
<ul> <li>Maximize use of available data (variable-by-variable deletion)</li> </ul>	<ul> <li>Use consistent case base across scale variables (listwise deletion)</li> </ul>	
Count duplicate responses for multiple categor	ny sets	
	OK         Paste         Reset         Cancel	Help

• Click OK to create the table.

### Figure 12-13

Table with "NA" displayed for missing statistics

						Hou	rs per day	/ watching	ITV				
			How get paid last week										
		Hourly	wage	Daily	wage	Weekly wage		Monthly salary		Annual salary		Other pay rate	
			Valid		Valid		Valid		Valid		Valid		Valid
		Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
Age	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8
category	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

ptions			
Data   General Viewer Draft Viewer	Currency Dutput Labels Charts	Scripts Interactive Pivot Table	es
TableLook D:\\SPSS\Looks\Contrast 3.tlo Contrast (narrow),tlo Contrast 2.tlo Contrast 4.tlo Contrast 5.tlo Contrast 5.tlo Contrast 5.tlo Contrast 5.tlo Contrast 10 Horizontal.tlo Horizontal.tlo Horizontal.tlo Browse Set TableLook Directory	Sample SJ Querts SJ SJ S	Penjitsuban	
Adjust Column Widths for C Labels only C Labels and data	Default Editing Mode	×	
	OK Cancel	Apply Help	

- ► 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	s per day	/watchin	ig TV					
			How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate		
			Valid		Valid		Valid		Valid		Valid		Valid	
		Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	
Age	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8	
category	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
                           EMPTY=
        UNITS={POINTS}
                                   {ZERO
                                               MISSING=
                                                         {'.' }
{'chars'}
                INCHES
                                    BLANK
                                    'chars'
                [CM
/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 }
                                                                 {INCLUDE}
                                                         EMPTY=
                                                BEFORE }
            YES }
                                                                 {EXCLUDE }
    Explicit value lists can include SUBTOTAL='label', MISSING, and OTHERNM.
/TITLES CAPTION= ['text' 'text'...]
         CORNER=
                   ['text' 'text'...]
['text' 'text'...]
         TITLE=
         Text can contain the symbols ) DATE ) TIME ) TABLE
/SIGTEST TYPE= CHISQUARE ALPHA= {0.05
                                    {significance level}
                             ALPHA= {0.05
/COMPARETEST TYPE=
                     {PROP}
                     {MEAN }
                                     {significance level]
                                 ORIGIN=COLUMN
         ADJUST= {BONFERRONI }
                  NONE
```

Row, column, and layer elements each have the general form

varname {[C]} [summary `label' format...] {+} varname ... {:}

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 LAYERPOT.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: MAXINUM MEAN MEDIAN MINIMUM MISSING MODE PILLE RANGE SEMEAN STDDEV SUM TOTALN VALIDN VARIANCE ROWPCT.SUM COLPCT.SUM TABLEPCT.SUM SUBTABLEPCT.SUM LAYERFORT.SUM LAYERFORMET.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.COUNT SUBTABLEPCT.RESPONSES.COUNT COLPCT.RESPONSES.COUNT TABLEPCT.RESPONSES.COUNT LAYERCOLPCT.RESPONSES.COUNT LAYERPCT.RESPONSES.COUNT LAYERCOLPCT.RESPONSES.COUNT ROWPCT.COUNT.RESPONSES COLPCT.COUNT.RESPONSES TABLEPCT.COUNT.RESPONSES SUBTABLEPCT.COUNT.RESPONSES LAYERROWPCT.COUNT.RESPONSES LAYERPONSES LAYERROWPCT.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 sub-command 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.

		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	Extremely liberal	4.5%	2.5%	2.1%	2.4%	1.3%	2.2%	
liberal or	Liberal	18.8%	15.7%	14.6%	11.3%	10.5%	9.4%	
conservative	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%	

CTABLES /TABLE POLVIEWS [COLPCT] BY AGECAT.

• *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					
			ale	Fen	nale	То	tal
News	Get news from internet	359	40.1%	508	42.9%	867	41.7%
sources	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.

+ CONPRESS + CONMEDIC)[COUNT ROWPCT] /CLABELS ROWLABELS=OPPOSITE.							
	A grea	t deal	Only s	some	Hardh	y 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.

CTABLES / TABLE (CONFINAN + CONBUS + CONBUS + CONEDUC

## **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.

CTABLES /TABLE HAPPY.

		Count
General	Very happy	891
happiness	Pretty happy	1575
	Not too happy	340

• The counts for HAPPY are in the rows.

#### Example:

CTABLES /TABLE BY HAPPY.

General happiness					
Very	Pretty	Not too			
happy	happy	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



• 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	Very happy	197	412	221	59	
happiness	Pretty happy	499	662	314	97	
	Not too happy	98	136	79	27	
Happiness	Very happy	111	462	232	49	
of marriage	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.

			Num	ber of children (	grouped catego	ries)	
				None	1-2	3-4	5 or more
				Count	Count	Count	Count
Gender	Male	Happiness	Very happy	48	216	102	30
	of marriage	Pretty happy	25	110	58	11	
			Not too happy	3	7	4	1
	Female	Happiness	Very happy	63	246	130	19
		of marriage	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	Very happy	Gender	Male	373
happiness			Female	518
	Pretty happy	Gender	Male	712
			Female	863
	Not too	Gender	Male	133
	happy		Female	207
Happiness	Very happy	Gender	Male	396
of marriage			Female	459
	Pretty happy	Gender	Male	205
			Female	240
	Not too	Gender	Male	15
	happy		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	General	Very happy	47	47
respondent	happiness	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:**

			Num	ber of children (	grouped catego	ries)	
				None	1-2	3-4	5 or more
				Column %	Column %	Column %	Column %
Gender	Male	Happiness	Very happy	63.2%	64.9%	62.2%	71.4%
		of marriage	Pretty happy	32.9%	33.0%	35.4%	26.2%
			Not too happy	3.9%	2.1%	2.4%	2.4%
	Female	Happiness	Very happy	69.2%	63.9%	61.6%	57.6%
	of marriag	of marriage	Pretty happy	28.6%	33.2%	35.5%	33.3%
			Not too happy	2.2%	2.9%	2.8%	9.1%

CTABLES /TABLE SEX > HAPMAR [COLPCT] BY CHILDCAT.

#### **Example:**

CTABLES /TABLE AGECAT > TVHOURS [MEAN F5.2, STDDEV 'Standard Deviation' F5.2, PTILE 90 '90th Percentile'].

				Standard	
			Mean	Deviation	90th Percentile
Age	Less than 25	Hours per day watching TV	2.85	2.03	5
category	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.

Function	Description	Default Label <sup>*</sup>	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

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

Table 1 Summary Functions: All Variables (Continued)

\* 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.

Function	Description	Default Label	Default Format
MAXIMUM	Largest value.	Maximum	General
MEAN	Arithmetic mean. The default for scale variables.	Mean	General
MEDIAN	50 <sup>th</sup> 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

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 2
 Summary Functions: Scale Variables, Totals, and Subtotals (Continued)

## 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

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

## Table 3 Summary Functions: Multiple Response Sets (Continued)

#### **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.

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%).

Table 4 Additional Formats for Summaries

## **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	Include	Include	Include
Scale Variable: valid value			
Categorical Variable: included user-missing value			
Multiple Category Set: all values are included user-missing	Exclude	Include	Include
Scale Variable: user-missing or system-missing			
Categorical Variable: excluded user-missing or system-missing			
Multiple Dichotomy Set: all values are "false"	Exclude	Exclude	Include
Multiple Category Set: all values are excluded user-missing, system-missing, or excluded valid, but at least one value is not excluded valid			
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	Every day	805	43.0%
respondent read	Few times a week	420	22.5%
newspaper	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	Every day	805
respondent read		43.0%
newspaper	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.

CTABLES

/TABLE (CONFINAN + CONEDUC + CONBUS + CONMEDIC + CONPRESS + CONTV )

		Count
Confidence in banks	A great deal	490
& financial institutions	Only some	1068
	Hardly any	306
Confidence in	A great deal	511
education	Only some	1055
	Hardly any	315
Confidence in major	A great deal	500
companies	Only some	1078
	Hardly any	243
Confidence in	A great deal	844
medicine	Only some	864
	Hardly any	167
Confidence in press	A great deal	176
	Only some	878
	Hardly any	808
Confidence in	A great deal	196
television	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=N0 /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 }
TOTAL= { **NO** } LABEL= "label" POSITION= { **AFTER** } EMPTY= { **INCLUDE** }

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	Less than 25	242
category	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.

```
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	American	99	267	293	214	140	215	1228
maker,	Japanese	73	136	140	107	66	104	626
most	German	18	91	69	63	36	61	338
car	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.

/TABLE AGE [MEAN F5.1] > CAR1 BY SEX /CATEGORIES VARIABLES=SEX TOTAL=YES /CATEGORIES VARIABLES=CAR1 KEY=MEAN(AGE).						
Gender						
		Female	Total			
	_	_	Mean	Mean	Mean	
Age of	Car	Swedish	42.6	45.6	44.3	
respondent	maker,	Japanese	43.5	45.5	44.7	
	most	Korean	43.4	46.2	45.0	
	recent		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	<i>Whether to display a total for a variable.</i> 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

*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.

```
CTABLES /TABLE AGECAT /CATEGORIES VARIABLES=AGECAT TOTAL=YES LABEL='Total Respondents'.
```

		Count
Age	Less than 25	242
category	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	Gender	Male	44.6
respondent		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)				ries)
		None	None 1-2 3-4 5		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** The 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	Less than 25	37	1	5	5	194	242
category	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	Chi-square	1473.381
category	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 {
NONE }
ORIGIN=COLUMN

The SIGTEST subcommand has the following specifications:

- **TYPE** The 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.
```

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Age	Less than 25				В	ABCD
category	25 to 34	в		в	в	ABCD
	35 to 44	BE		BE	ABCE	В
	45 to 54	BE		BE		
	55 to 64	E	E	E		
	65 or older	E	ACDE	E		

#### Comparisons of Column Proportions

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.

 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.

#### 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	Gender	Male	49	66	48	44	32
respondent		Female	45	70	48	41	32
		Total	47	70	48	42	32

#### Comparisons of Column Means

					Marital status		
							Never
			Married	Widowed	Divorced	Separated	married
			(A)	(B)	(C)	(D)	(E)
Age of respondent	Gender	Male	E	ACDE	E	E	
		Female	E	ACDE	DE	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	MINCOLV	VIDTH={I v}	<b>DEFAULT</b> } value }	MAXCOLWID	TH={ <b>DEFAULT</b> } {value}	
	UNITS=	POINTS INCHES CM	} EMPTY= }	{ <b>ZERO</b> {BLANK {'chars'	} MISSING= }	$\left\{ {}^{\prime} \cdot \cdot \\ {}^{\prime} \cdot chars \prime \right\}$

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	VARIABLE	S= varlist
I	DISPLAY=	{ <b>DEFAULT</b> } {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

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

Bonferroni correction custom tables comparisons, 229

captions custom tables, 51, 226 category labels positioning in custom tables, 220 category specification explicit, in custom tables, 222 implicit, in custom tables, 223 category variables custom tables, 207 chi-square statistics custom tables, 145 chi-square test custom tables, 228 column means statistics custom tables, 152 column proportions statistics custom tables, 158 column width controlling in custom tables, 48, 194 custom tables, 230 comperimeter tables, 47, 103 concatenation custom tables, 208 controlling number of decimals displayed, 63 corner labels custom tables, 51 corner text custom tables, 227 count vs. valid N, 134 crosstabulation custom tables, 61

custom tables captions, 51 categorical variables, 26 category label positioning, 220 category variables, 207 changing labels for summary statistics, 57 changing measurement level, 26 changing summary statistics dimension, 39 column width, 48, 230 compact view, 81 comperimeter tables, 47, 103 concatenation, 208 controlling number of decimals displayed, 33 corner labels, 51 counting duplicate responses, 232 crosstabulation. 61 custom totals. 39 display formats, 33 empty cells, 48, 231 excluding categories, 42, 66 excluding valid values, 222 hiding statistics labels, 56 how to build a table, 29 layer variables, 84, 87, 88 listwise deletion, 231 marginal totals, 64 mean-frequency tables, 39 missing summaries, 231 missing values, 231 missing values exclusion for scale summaries, 48 multiple category sets, 48 multiple response functions, 215 multiple response sets, 26, 168, 207, 232 nesting, 208 nesting layer variables, 88 nesting variables, 76, 81 overview, 204 percentage functions, 212 percentages, 36, 37, 57, 62

Subject Index

percentages for multiple response sets, 37 printing layered tables, 89 reordering categories, 42 row vs. column percentages, 57 scale variable functions, 213 scale variables, 26, 210 showing and hiding variable names and labels, 33 simple tables for categorical variables, 56 sorting categories, 66 stacking, 208 stacking variables, 73, 75 statistics source dimension, 62 subtotals, 42, 91 summary functions, 212 summary label positioning, 220 summary statistics, 36, 37, 38 summary statistics display formats, 41 swapping row and column variables, 83 syntax conventions, 205 table expression, 207 table of frequencies, 47, 103 tables of variables with shared categories, 47, 103 test statistics, 52 titles, 51 totals, 42, 59, 91 totals in tables with excluded categories, 66 value labels for categorical variables, 26 variable labels, 231 variable types, 207 custom total summary statistics, 124

dates custom tables titles, 227 including current date in custom tables, 51 decimals controlling number of decimals displayed in custom tables, 33, 63, 188 deleting categories custom tables, 42, 66 different summary statistics for different variables stacked tables, 136 display formats, 63 summary statistics in custom tables, 41, 188

displaying category values, 128

duplicate responses counting in custom tables, 232 multiple category sets, 178

empty categories excluding in custom tables, 226 including in custom tables, 226 empty cells display format in custom tables, 231 displayed value in custom tables, 48, 196 excluding categories custom tables, 42, 66 explicit category specification custom tables, 222

formats for summary functions custom tables, 218

group totals, 95 grouped summaries scale variables, 139

hiding statistics labels in custom tables, 56

implicit category specification custom tables, 223

## labels

changing label text for summary statistics, 192 positioning category labels in custom tables, 220 positioning summary labels in custom tables, 220 layer variables custom tables, 84, 87, 88 nesting layer variables, 88 printing layered tables, 89 stacking layer variables, 87 listwise deletion custom tables, 231

maximum custom tables, 38 mean, 132 custom tables. 38 mean-frequency tables, 39, 124 means pairwise comparisons in custom tables, 229 measurement level changing in custom tables, 26 median. 133 custom tables. 38 minimum custom tables, 38 missing summary custom tables, 231 missing values, 134 custom tables, 219, 223, 231 effect on percentage calculations, 184 including in custom tables, 184 mode custom tables, 38 multiple category group, defined, 234 multiple dichotomy group, defined, 233 multiple response sets counts vs. responses, 170 custom tables, 210, 232 defining, 168 duplicate responses in multiple category sets, 48, 178 functions in custom tables, 215 multiple categories, 168 multiple dichotomies, 168 percentages, 37, 170 responses, 170 summary statistics source variable, 176 totals, 170

nesting custom tables, 208 nesting variables custom tables, 76, 81 scale variables, 142 omitting categories custom tables, 66 pairwise comparisons comparison of means in custom tables, 229 comparison of proportions in custom tables, 229 custom tables, 229 percentage functions custom tables, 212 percentages custom tables, 36, 37, 57, 62 missing values, 184 multiple response sets, 37 percentiles custom tables, 214 position of totals custom tables, 225 printing tables with layers, 89 proportions pairwise comparisons in custom tables, 229

range custom tables, 38 reordering categories custom tables, 42

scale variables custom tables, 210 functions in custom tables, 213 grouped summaries, 139 Subject Index

multiple summary statistics, 133 nesting, 142 stacking, 132 summaries grouped by row and column categorical variables, 139 summary statistics, 132 totaling in custom tables, 225 significance tests custom tables, 52 sorting custom tables, 223 sorting categories custom tables, 66 stacking custom tables, 208 multiple summary statistics source variables, 121 stacking variables custom tables, 73, 75 different summary statistics for different variables, 136 scale variables, 132 stacking layer variables, 87 standard deviation custom tables, 38 statistics custom total summary statistics, 124 stacked tables, 121 subgroup totals, 95 subtotals, 99 custom tables, 42, 91, 223 sum custom tables, 38 summaries custom tables, 210 summary labels custom tables, 220 summary statistics changing label text, 192 custom total summary statistics, 124 different summaries for different variables in stacked tables, 136 display format, 188 multiple response sets, 170 source dimension, 114 source variable, 114

stacked tables, 121 statistics, 114 summary statistics source variable multiple response sets, 176 scale variables, 142 table description custom tables titles, 227 table of frequencies custom tables, 47, 103 TableLooks changing the default TableLook, 199 tables custom tables, 26 test statistics custom tables, 52 time including current time in custom tables, 51 time stamp custom tables titles, 227 titles custom tables, 51, 226 total N, 184 totals custom tables, 42, 59, 91, 225 display position, 95 excluded categories, 93 group totals, 95 layers, 98 marginal totals for custom tables, 64 nested tables, 95

unweighted functions custom tables, 211

valid N, 134, 184 custom tables, 38 valid values excluding in custom tables, 222

245

values displaying category labels and values, 128 values and value labels, 128 variable labels custom tables, 231 suppressing display in custom tables, 33 variable types custom tables, 207 variance custom tables, 38

## Syntax Index

BONFERRONI (keyword) CTABLES command, 229 CAPTION (keyword) CTABLES command, 226 CATEGORIES subcommand CTABLES command, 222 CHISQUARE (keyword) CTABLES command, 228 CLABELS (command) CATEGORIES subcommand, 222 CLABELS (subcommand) CTABLES command, 220 COMPARETEST (subcommand) CTABLES command, 229 CORNER (keyword) CTABLES command, 227 corner text CTABLES command, 227 COUNTDUPLICATES (keyword) CTABLES command, 232 CTABLES (command), 203 CAPTION keyword, 226 caption lines, 226 CHISQUARE keyword, 228 CLABELS subcommand, 220 COMPARETEST subcommand, 229 CORNER keyword, 227 corner text, 227 **COUNTDUPLICATES** keyword, 232 )DATE keyword, 227 dates in titles, 227 empty categories, 226 empty cell format, 230 EMPTY keyword, 226, 230 explicit category specification, 222 FORMAT subcommand, 230 formats for summaries, 218 implicit category specification, 223

MISSING keyword, 223, 231 missing values, 219, 223 MRSETS subcommand, 232 ORDER keyword, 223 overview, 204 position of totals, 225 scale variable totals, 225 SIGTEST subcommand, 228 SLABELS subcommand, 220 SMISSING subcommand, 231 sorting categories, 223 split file processing, 205 subtotals, 223 summary functions, 213 summary functions for multiple response sets, 215 summary functions for scale variables, 213 summary specifications, 210 syntax conventions, 205 table description in titles, 227 )TABLE keyword, 227 TABLE subcommand, 207 )TIME keyword, 227 TITLE keyword, 227 TITLES subcommand, 226 TOTAL keyword, 225 totals, 225 unweighted functions, 211 variable types, 207 VLABELS subcommand, 231

)DATE (keyword) CTABLES command, 227

EMPTY (keyword) CTABLES command, 230 explicit category specification in CTABLES command, 222 Syntax Index

FORMAT (subcommand) CTABLES command, 230 implicit category specification CTABLES command, 223 MCGROUP (subcommand) MRSETS command, 235 MDGROUP (subcommand) MRSETS command, 234 MISSING (keyword) CTABLES command, 223, 231 MRSETS (command), 233 **DELETE** subcommand, 235 **DISPLAY** subcommand, 236 MCGROUP subcommand, 235 MDGROUP subcommand, 234 syntax conventions, 234 MRSETS (subcommand) CTABLES command, 232

ORDER (keyword) CTABLES command, 223 SIGTEST (subcommand) CTABLES command, 228 SLABELS (subcommand) CTABLES command, 220 SMISSING (subcommand) CTABLES command, 231 sorting categories CTABLES command, 223 )TABLE (keyword) CTABLES command, 227 TABLE subcommand CTABLES command, 207 )TIME keyword CTABLES command, 227

TITLES (subcommand) CTABLES command, 226

TOTAL (keyword) CTABLES command, 225

VLABELS (subcommand) CTABLES command, 231