IBM SPSS Custom Tables 22

IBM
Product Information

This edition applies to version 22, release 0, modification 0 of IBM SPSS Statistics and to all subsequent releases and modifications until otherwise indicated in new editions.
## Contents

### Chapter 1. Table Builder Interface  
1 Table Builder Interface  
1 Building Tables  
1 To Build a Table  
2 Stacking Variables  
2 Nesting Variables  
2 Layers  
2 Showing and Hiding Variable Names and/or Labels  
2 Summary Statistics  
3 Categories and Totals  
3 Computed Categories  
4 Tables of Variables with Shared Categories (Comperimeter Tables)  
4 Customizing the Table Builder  
5 Custom Tables: Options Tab  
5 Custom Tables: Titles Tab  
5 Custom Tables: Test Statistics Tab  

### Chapter 2. Simple Tables for Categorical Variables  
19 Simple Tables for Categorical Variables  
19 A Single Categorical Variable  
20 Percentages  
20 Totals  
21 Crosstabulation  
21 Percentages in Crosstabulations  
22 Controlling Display Format  
22 Marginal Totals  
23 Sorting and Excluding Categories  

### Chapter 3. Stacking, Nesting, and Layers with Categorical Variables  
25 Stacking Categorical Variables  
25 Stacking with Crosstabulation  
26 Nesting Categorical Variables  
26 Suppressing Variable Labels  
27 Nested Crosstabulation  
29 Layers  
30 Two Stacked Categorical Layer Variables  
30 Two Nested Categorical Layer Variables  

### Chapter 4. Totals and Subtotals for Categorical Variables  
33 Simple Total for a Single Variable  
33 What You See Is What Gets Totaled  
34 Display Position of Totals  
34 Totals for Nested Tables  
35 Layer Variable Totals  
36 Subtotals  
36 What You See Is What Gets Subtotaled  
37 Hiding Subtotaled Categories  
37 Layer Variable Subtotals  

### Chapter 5. Computed Categories for Categorical Variables  
39 Simple Computed Category  
40 Hiding Categories in a Computed Category  
40 Referencing Subtotals in a Computed Category  
42 Using Computed Categories to Display Nonexhaustive Subtotals  

### Chapter 6. Tables for Variables with Shared Categories  
45 Table of Counts  
46 Table of Percentages  
47 Totals and Category Control  
47 Nesting in Tables with Shared Categories  

### Chapter 7. Summary Statistics  
49 Summary Statistics Source Variable  
50 Summary Statistics Source for Categorical Variables  
50 Summary Statistics Source for Scale Variables  
51 Stacked Variables  
51 Custom Total Summary Statistics for Categorical Variables  
52 Displaying Category Values  

### Chapter 8. Summarizing Scale Variables  
55 Summarizing Scale Variables  
55 Stacked Scale Variables  
55 Multiple Summary Statistics  
56 Count, Valid N, and Missing Values  
57 Different Summaries for Different Variables  
58 Group Summaries in Categories  
58 Multiple Grouping Variables  
59 Nesting Categorical Variables within Scale Variables  

### Chapter 9. Test Statistics  
61 Test Statistics  
61 Tests of Independence (Chi-Square)  
61 Effects of Nesting and Stacking on Tests of Independence  
64 Comparing Column Means  
65 Effects of Nesting and Stacking on Column Means Tests  
66 Comparing Column Proportions  
66 Effects of Nesting and Stacking on Column Proportions Tests  
69 A Note on Weights and Multiple Response Sets  

### Chapter 10. Multiple Response Sets  
73 Counts, Responses, Percentages, and Totals  
75 Using Multiple Response Sets with Other Variables  

---

**iii**
Statistics Source Variable and Available Summary .......................... 76
Multiple Category Sets and Duplicate Responses ...................... 76
Significance Testing with Multiple Response Sets ..................... 77
Tests of Independence with Multiple Response Sets ................... 78
Comparing Column Means with Multiple Response Sets ............... 78

Chapter 11. Missing Values .................................................. 81
Tables without Missing Values ............................................... 81
Including Missing Values in Tables ......................................... 82

Chapter 12. Formatting and Customizing Tables .......................... 85
Formatting and Customizing Tables ......................................... 85

Summary Statistics Display Format ........................................... 85
Display Labels for Summary Statistics ...................................... 86
Column Width ................................................................. 87
Display Value for Empty Cells ............................................... 87
Display Value for Missing Statistics ........................................ 88

Chapter 13. Sample Files ....................................................... 89

Notices ............................................................................. 97
Trademarks ....................................................................... 99

Index .............................................................................. 101
Chapter 1. Table Builder Interface

Table Builder Interface

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

Building Tables

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.

Categorical variables can be either nominal or ordinal

- **Nominal.** A variable can be treated as nominal when its values represent categories with no intrinsic ranking (for example, the department of the company in which an employee works). Examples of nominal variables include region, postal code, and religious affiliation.
• *Ordinal.* A variable can be treated as ordinal when its values represent categories with some intrinsic ranking (for example, levels of service satisfaction from highly dissatisfied to highly satisfied). Examples of ordinal variables include attitude scores representing degree of satisfaction or confidence and preference rating scores.

**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. See the topic “Stacking Variables” on page 4 for more information.

**Multiple Response Sets**

Custom Tables also supports a special kind of “variable” called a **multiple response set.** Multiple response sets are not really variables in the normal sense. You cannot see them in the Data Editor, and other procedures do not 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. See the topic [Chapter 10, “Multiple Response Sets,” on page 73](#) for more information.

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

**Table 1. Measurement level icons**

<table>
<thead>
<tr>
<th>Numeric</th>
<th>String</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Icon" /></td>
<td><img src="image2" alt="Icon" /></td>
<td><img src="image3" alt="Icon" /></td>
<td><img src="image4" alt="Icon" /></td>
</tr>
</tbody>
</table>

**Table 2. Multiple response set icons**

<table>
<thead>
<tr>
<th>Multiple response set type</th>
<th>Icon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple response set, multiple categories</td>
<td><img src="image5" alt="Icon" /> <img src="image6" alt="Icon" /></td>
</tr>
<tr>
<td>Multiple response set, multiple dichotomies</td>
<td><img src="image7" alt="Icon" /> <img src="image8" alt="Icon" /></td>
</tr>
</tbody>
</table>
You can change the measurement level of a variable in the table builder by right-clicking the variable in the variable list and selecting Categorical or Scale from the pop-up menu. You can permanently change a variable's measurement level in the Variable View of the Data Editor. Variables defined as nominal or ordinal are treated as categorical by Custom Tables.

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

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

Canvas pane. You build a table by dragging and dropping variables onto the rows and columns of the canvas pane. The canvas pane displays a preview of the table that will be created. The canvas pane does not 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.
- If any variable in the active dataset contains more than 12,000 defined value labels, you cannot use the table builder to create tables. If you don’t need to include variables that exceed this limitation in your tables, you can define and apply variable sets that exclude those variables. If you need to include any variables with more than 12,000 defined values labels, you can use CTABLES command syntax to generate the tables.

To Build a Table
1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. Drag and drop one or more variables to the row and/or column areas of the canvas pane.
3. Click OK to create the table.
4. Select (click) the variable on the canvas pane.
5. Drag the variable anywhere outside the canvas pane, or press the Delete key.

   To change the measurement level of a variable:
6. Right-click the variable in the variable list (you can do this only in the variable list, not on the canvas).
7. Select Categorical or Scale from the pop-up 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
1. In the variable list, select all of the variables you want to stack, then drag and drop them together into the rows or columns of the canvas pane.
   
or
2. 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.

Table 3. Stacked categorical variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Categories</th>
<th>Summary Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable 1</td>
<td>Category 1</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>Category 2</td>
<td>456</td>
</tr>
<tr>
<td>Variable 2</td>
<td>Category 1</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>Category 2</td>
<td>456</td>
</tr>
<tr>
<td></td>
<td>Category 3</td>
<td>789</td>
</tr>
</tbody>
</table>

See the topic “Stacking Categorical Variables” on page 25 for more information.

### 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
1. Drag and drop a categorical variable into the row or column area of the canvas pane.
2. 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.

Table 4. Nested categorical variables.

<table>
<thead>
<tr>
<th>Variable 1</th>
<th>Variable 2</th>
<th>Summary Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>Category 1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Category 2</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Category 3</td>
<td>56</td>
</tr>
<tr>
<td>Category 2</td>
<td>Category 1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Category 2</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Category 3</td>
<td>56</td>
</tr>
</tbody>
</table>
Layers
You can use layers to add a dimension of depth to your tables, creating three-dimensional "cubes." Layers are similar to nesting or stacking; the primary difference is that only one layer category is visible at a time. For example, using Age category as the row variable and Gender as a layer variable produces a table in which information for males and females is displayed in different layers of the table.

To Create Layers
1. Click Layers on the Table tab in the table builder to display the Layers list.
2. Drag and drop the scale or categorical variable(s) that will define the layers into the Layers list.

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:
1. Right-click the variable in the table preview on the canvas pane.
2. Select Show Variable Label or Show Variable Name from the pop-up 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.
The summary statistics (and other options) available here depend on the measurement level of the summary statistics source variable, as displayed at the top of the dialog box. 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
1. 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
1. Select (click) the summary statistics source variable on the canvas pane of the Table tab.
2. In the Define group of the Table tab, click **Summary Statistics.**
   
   or

3. Right-click the summary statistics source variable on the canvas pane and select **Summary Statistics** from the pop-up menu.

4. 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 Display list, or you can drag and drop selected statistics from the Statistics list into the Display list.
5. Click the up or down arrows to change the display position of the currently selected summary statistic.
6. Select a display format from the Format drop-down list for the selected summary statistic.
7. Enter the number of decimals to display in the Decimals cell for the selected summary statistic.
8. Click **Apply to Selection** to include the selected summary statistics for the currently selected variables on the canvas pane.
9. 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. See the topic [*Custom Total Summary Statistics for Categorical Variables*](#) on page 9 for more information.

**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. This only differs from count if weighting is in effect.
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 do not 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 on 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. See the topic “Summary Statistics for Multiple Response Sets” on page 8 for more information.

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 computational base:

Simple percentage. Percentages 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 does not 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.

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/Row Responses % (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:

1. From the menus, choose:
   - Analyze > Tables > Custom Tables...
   - The table builder will open.
2. Drag and drop a categorical variable into the Rows or Columns area of the canvas.
3. Right-click the variable on the canvas and select Categories and Totals from the pop-up menu.
4. Click (check) the Total check box, and then click Apply.
5. Right-click the variable again on the canvas and select Summary Statistics from the pop-up menu.
6. 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:

| Table 5. Ordinal variable categories in rows, summary statistics count and mean in columns. |
|---|---|---|
| Variables | Categories | Count | Mean |
| Variable 1 | 1 Agree | 196 | |
| | 2 Neutral | 936 | |
| | 3 Disagree | 744 | |
| | Total | 1876 | 2.29 |
7. 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.

**Summary Statistics Display Formats**

The following display format options are available:

- **nnnn.** Simple numeric.

- **nnnn%.** Percentage 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 a percentage 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 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.
- Insert computed categories.
- 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.
• 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, insert computed categories, 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
1. Drag and drop a categorical variable or multiple response set onto the canvas pane.
2. Right-click the variable on the canvas pane, and select Categories and Totals from the pop-up menu.
   or
3. 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:
4. Ctrl-click each variable on the canvas pane.
   or
5. 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
6. 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:
1. Select (click) a category in the list.
2. Click the up or down arrow to move the category up or down in the list.
   or
3. Click in the Value(s) column for the category, and drag and drop it in a different position.

To Exclude Categories
1. Select (click) a category in the list.
2. Click the arrow next to the Exclude list.
   or
3. 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, cell count, or summary statistic in ascending or descending order.
1. In the Sort Categories group, click the By drop-down list and select the sort criterion you want to use: value, label, count, or summary statistic (such as mean, median, or mode). The available summary statistics for sorting depends on the summary statistics you have selected to display in the table.
2. 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
1. 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.
2. Click **Add Subtotal....**
3. In the Define Subtotal dialog box, optionally modify the subtotal label text.
4. To show only a subtotal and suppress the display of the categories that define the subtotal, select **Hide subtotaled categories from the table.**
5. Click **Continue** to add the subtotal.

Totals
1. 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.

Computed Categories

You can display categories computed from summary statistics, totals, subtotals, and/or constants. See the topic [“Computed Categories” on page 13](#) for more information.

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. See the topic [“Summary Statistics for Categorical Variables” on page 6](#) for more information.

*Note:* If you select multiple custom total statistics that are also in the body of the table and you hide the statistics labels, then the totals are resorted into the same order as in the body of the table—and since the labels aren’t displayed, you may not know what each total statistic actually represents. In general, selecting multiple statistics and hiding the statistics labels is probably not a good idea.

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 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 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.

**Computed Categories**

In addition to displaying the aggregated results of summary statistics, a table can display one or more categories computed from these aggregated results, from constant values, from subtotals and totals, or a combination of them. The results are known as computed categories or postcomputes. A computed category acts like a category in a single variable with the following similarities and differences:

- A computed category is positioned like the other categories.
- A computed category operates on the same statistics as the other categories.
- Computed categories do not affect subtotals, totals, or significance tests.
- By default, the values of computed categories use the same formatting for summary statistics as the other categories. You can override the format when defining the computed category.

Because computed categories can be used to total aggregated results, they can be similar to subtotals. However, computed categories have the following advantages over subtotals:

- Computed categories can be calculated from the results of other subtotals.
- Computed categories can overlap with each other, operating on the same (or some of the same) categories.
- Computed categories do not have to include values from all other categories above or below the computed category. That is, computed categories are not exhaustive.
- Computed categories can include values from categories that are not adjacent.

Unlike totals and subtotals, computed categories are calculated from the aggregated data rather than the original data. Therefore, the values of computed categories may not match the results of totals and subtotals. Also, because you have the option to hide source categories when defining the computed category, it may be difficult to interpret subtotals in the resulting table. If you use computed categories, it is recommended that you specify custom labels for subtotals.

**To Define a Computed Category**

Computed categories are added from the **Categories and Totals dialog box.** For information about accessing that dialog box, see the topic **"Categories and Totals on page 10"**

1. In the Categories and Totals dialog box, click **Add Category...**
2. In **Label for Computed Category**, specify a label for the computed category. You can drag categories from the Categories list to include labels for those categories.
3. Build an expression by selecting categories and/or totals and subtotals and using operators to define the computed categories. You can also type constant values (e.g., 500) to include in the expression.
4. To show only a computed category and suppress the display of the categories that define the computed category, select **Hide categories used in expression from table.**
5. Click the **Display Formats** tab to change the display format and number of decimal places for the computed category. See the topic “Display Formats for Computed Categories” for more information.

6. Click **Continue** to add the computed category.

**Display Formats for Computed Categories**

By default, a computed category uses the same display format and number of decimal places as the other categories in the variable. You can override these on the Display Formats tab in the Computed Category dialog box. The Display Formats tab lists the current summary statistics on which the computed category operates in addition to the display formats and number of decimal places for those statistics.

For each summary statistic, you can:

1. Select a display format from the Format drop-down list for the summary statistic. For a full list of display formats, see the topic “Summary Statistics Display Formats” on page 10.
2. Enter the number of decimals to display in the Decimals cell for the selected summary statistic.

**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

*Table 6. Stacked variables with shared response categories in columns*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable 1</td>
<td>12</td>
<td>34</td>
<td>56</td>
</tr>
<tr>
<td>Variable 2</td>
<td>56</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>Variable 3</td>
<td>34</td>
<td>56</td>
<td>12</td>
</tr>
</tbody>
</table>

See the topic “Chapter 6, Tables for Variables with Shared Categories,” on page 43 for more information.

**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:

1. 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.
2. 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.
3. 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.

**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).

**Hide small counts.** You can choose to hide counts that are less than a specified integer. Hidden values will be displayed as <N, where N is the specified integer. The specified integer must be greater than or equal to 2.

---

**Custom Tables: Titles Tab**

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

**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 are 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.
- Significance tests for multiple response sets and subtotals. (For information on significance testing for multiple response sets, see “Significance Testing with Multiple Response Sets” on page 77)

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

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. The table must include the mean as a summary statistic. 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. Finally, while the variance for the means test is always based on just the categories compared for multiple response tests, for ordinary categorical variables it can be estimated from just the categories compared or all categories.

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. The table must include counts or simple column percentages. 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.

Identify Significant Differences. If you select Compare column means or Compare column proportions, you can choose how to indicate significant differences.

- In a separate table. Significance tests results are displayed in a separate table. If two values are significantly different, the cell corresponding to the larger value displays a key identifying the column of the smaller value. Following is an example.

- In the main table using APA-style subscripts. The main table itself identifies significant differences with APA-style formatting using subscript letters. If two values are significantly different, those values display different subscript letters. These subscripts are not footnotes. When this option is in effect, the defined footnote style in the current TableLook is overridden and footnotes are displayed as superscript numbers. Following is an example.

For the full example that demonstrates how to create and interpret these tables, see “Comparing Column Proportions” on page 66

Note: The APA-style table includes a caption that explains how to interpret the subscripts in the table. If you specify a caption on the Titles tab, the caption specified on the Titles tab will be displayed instead of the caption that explains the subscripts.

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.
Use subtotals in place of subtotaled categories. If selected, then each subtotal replaces its categories for significance testing. Otherwise, only subtotals for which the subtotaled categories are hidden replace their categories for testing.

Include multiple response variables. If selected, tests are performed using categories of multiple response sets. Otherwise multiple response sets are ignored when performing tests.
Chapter 2. Simple Tables for Categorical Variables

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). Categorical variables can be either nominal or ordinal.

- **Nominal.** A variable can be treated as nominal when its values represent categories with no intrinsic ranking (for example, the department of the company in which an employee works). Examples of nominal variables include region, postal code, and religious affiliation.
- **Ordinal.** A variable can be treated as ordinal when its values represent categories with some intrinsic ranking (for example, levels of service satisfaction from highly dissatisfied to highly satisfied). Examples of ordinal variables include attitude scores representing degree of satisfaction or confidence and preference rating scores.

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

Table 7. Measurement level icons

<table>
<thead>
<tr>
<th>Measurement Level</th>
<th>Numeric</th>
<th>String</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (Continuous)</td>
<td><img src="icon.png" alt="Icon" /></td>
<td>n/a</td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
</tr>
<tr>
<td>Ordinal</td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
</tr>
<tr>
<td>Nominal</td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
<td><img src="icon.png" alt="Icon" /></td>
</tr>
</tbody>
</table>

Custom Tables is optimized for use with categorical variables that have defined value labels. See the topic "Building Tables" on page 1 for more information.

Sample Data File

The examples in this chapter use the data file survey_sample.sav. See the topic data files for more information.

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

A Single Categorical Variable

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

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. 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.
3. Click OK to create the table.
The table is displayed in the Viewer window.

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>242</td>
</tr>
<tr>
<td>25 to 34</td>
<td>627</td>
</tr>
<tr>
<td>35 to 44</td>
<td>679</td>
</tr>
<tr>
<td>45 to 54</td>
<td>481</td>
</tr>
<tr>
<td>55 to 64</td>
<td>326</td>
</tr>
<tr>
<td>65 or older</td>
<td>479</td>
</tr>
</tbody>
</table>

Figure 2. Single categorical variable in rows

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

4. Open the table builder again (Analyze menu, Tables, Custom Tables).
5. In the Summary Statistics group, select (click) Hide for Position.
6. Click OK to create the table.

<table>
<thead>
<tr>
<th>Age category</th>
<th>Less than 25</th>
<th>242</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 to 34</td>
<td>627</td>
<td></td>
</tr>
<tr>
<td>35 to 44</td>
<td>679</td>
<td></td>
</tr>
<tr>
<td>45 to 54</td>
<td>481</td>
<td></td>
</tr>
<tr>
<td>55 to 64</td>
<td>326</td>
<td></td>
</tr>
<tr>
<td>65 or older</td>
<td>479</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Single categorical variable without summary statistics column label

## 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.

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. In the Summary Statistics group, deselect Hide for Position. Since this table will have two columns, you want to display the column labels so you know what each column represents.
3. Right-click Age category on the canvas pane and select Summary Statistics from the pop-up menu.
4. In the Summary Statistics dialog box, select Column N % in the Statistics list and click the arrow to add it to the Display list.
5. In the Label cell in the Display list, delete the default label and type Percent.
6. Click Apply to Selection and then click OK in the table builder to create the table.

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>242</td>
<td>8.6%</td>
</tr>
<tr>
<td>25 to 34</td>
<td>627</td>
<td>22.2%</td>
</tr>
<tr>
<td>35 to 44</td>
<td>679</td>
<td>24.0%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>481</td>
<td>17.0%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>326</td>
<td>11.0%</td>
</tr>
<tr>
<td>65 or older</td>
<td>479</td>
<td>16.9%</td>
</tr>
</tbody>
</table>

Figure 4. Counts and column percentages

## Totals

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

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Right-click Age category on the canvas pane and select Categories and Totals from the pop-up menu.
3. Select (click) **Total** in the Categories and Totals dialog box.
4. Click **Apply** and then click **OK** in the table builder to create the table.

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>242</td>
<td>9.6%</td>
</tr>
<tr>
<td>25 to 34</td>
<td>827</td>
<td>22.2%</td>
</tr>
<tr>
<td>35 to 44</td>
<td>679</td>
<td>18.9%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>681</td>
<td>17.0%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>320</td>
<td>11.3%</td>
</tr>
<tr>
<td>65 or older</td>
<td>479</td>
<td>10.9%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2520</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

*Figure 5. Counts, column percentages, and totals*

See the topic [Chapter 4, “Totals and Subtotals for Categorical Variables,” on page 33](#) for more information.

**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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click **Reset** to delete any previous selections in the table builder.
3. In the table builder, drag and drop **Age category** from the variable list to the Rows area on the canvas pane.
4. 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.)
5. Click **OK** to create the table.

*Figure 6. Crosstabulation of Age category and Gender*

**Percentages in Crosstabulations**

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

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Right-click **Gender** on the canvas pane.

   You may notice that **Summary Statistics** is disabled in the pop-up menu. This is because you can select summary statistics only 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. See the topic “Summary Statistics” on page [5](#) for more information.
3. Right-click Age category on the canvas pane and select Summary Statistics from the pop-up menu.

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

5. Select Row N % in the Statistics list and click the arrow to add it to the Display list.

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

<table>
<thead>
<tr>
<th>Age category</th>
<th>Male</th>
<th></th>
<th></th>
<th>Female</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>168</td>
<td>8.8%</td>
<td>44.6%</td>
<td>134</td>
<td>8.4%</td>
<td>55.4%</td>
</tr>
<tr>
<td>25 to 34</td>
<td>276</td>
<td>22.5%</td>
<td>44.0%</td>
<td>251</td>
<td>21.8%</td>
<td>58.0%</td>
</tr>
<tr>
<td>35 to 44</td>
<td>309</td>
<td>25.2%</td>
<td>45.5%</td>
<td>370</td>
<td>23.1%</td>
<td>54.5%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>221</td>
<td>18.0%</td>
<td>45.6%</td>
<td>260</td>
<td>16.5%</td>
<td>54.1%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>136</td>
<td>11.1%</td>
<td>42.5%</td>
<td>164</td>
<td>11.5%</td>
<td>57.5%</td>
</tr>
<tr>
<td>65 or older</td>
<td>178</td>
<td>14.5%</td>
<td>37.2%</td>
<td>301</td>
<td>18.8%</td>
<td>52.6%</td>
</tr>
</tbody>
</table>

Figure 7. Crosstabulation with row and column percentages

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?

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

<table>
<thead>
<tr>
<th>Age category</th>
<th>Male</th>
<th></th>
<th></th>
<th>Female</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>168</td>
<td>9.79</td>
<td>44.63</td>
<td>134</td>
<td>8.38</td>
<td>55.37</td>
</tr>
<tr>
<td>25 to 34</td>
<td>276</td>
<td>22.46</td>
<td>44.02</td>
<td>251</td>
<td>21.94</td>
<td>55.90</td>
</tr>
<tr>
<td>35 to 44</td>
<td>309</td>
<td>25.15</td>
<td>45.81</td>
<td>370</td>
<td>23.13</td>
<td>54.49</td>
</tr>
<tr>
<td>45 to 54</td>
<td>221</td>
<td>18.03</td>
<td>45.35</td>
<td>260</td>
<td>16.25</td>
<td>54.95</td>
</tr>
<tr>
<td>55 to 64</td>
<td>136</td>
<td>11.27</td>
<td>42.50</td>
<td>164</td>
<td>11.50</td>
<td>57.50</td>
</tr>
<tr>
<td>65 or older</td>
<td>178</td>
<td>14.55</td>
<td>37.16</td>
<td>301</td>
<td>18.81</td>
<td>52.84</td>
</tr>
</tbody>
</table>

Figure 8. Formatted cell display for row and column percentages

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to delete any previous selections in the table builder.
3. In the table builder, drag and drop Age category from the variable list to the Rows area on the canvas pane.
4. 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.)
5. Right-click Age category on the canvas pane and select Categories and Totals from the pop-up menu.
6. Select (click) Total in the Categories and Totals dialog box and then click Apply.
7. Right-click Gender on the canvas pane and select Categories and Totals from the pop-up menu.
8. Select (click) Total in the Categories and Totals dialog box and then click Apply.
9. In the Summary Statistics group, select (click) **Hide** for Position. (Since you're displaying only counts, you don't need to identify the "statistic" displayed in the data cells of the table.)

10. Click **OK** to create the table.

![Table](https://example.com/table.png)

**Figure 9. Crosstabulation with marginal totals**

### 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 to be displayed in the table.

#### Sorting Categories

You can manually rearrange categories or sort categories in ascending or descending order of:

- Data values.
- Value labels.
- Cell counts.
- Summary statistics. The available summary statistics for sorting depends on the summary statistics you have selected to display in the table.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. If *Age category* isn't already displayed in the Rows area on the canvas pane, drag and drop it there.
3. Right-click *Age category* on the canvas pane and select **Categories and Totals** from the pop-up 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.
4. In the Sort Categories group, select **Descending** from the Order drop-down list.
   - The sort order is now reversed.
5. Select **Labels** from the By drop-down list.
   - The categories are now sorted in descending alphabetical order of the value labels.
   - 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.
6. Click the category labeled *Less than 25* in the Label list.
7. Click the down arrow to the right of the list. The category moves down one row in the list.
8. Keep clicking the down arrow until the category is at the bottom of the list.

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

1. Click the category labeled *Less than 25* in the Label list.
2. Click the arrow key to the left of the Exclude list.
3. Click the category labeled *65 or older* in the Label list.
4. 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.
5. Click **Apply** and then click **OK** in the table builder to create the table.

![Table sorted by descending value label, some categories excluded](image)

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. See the topic “Chapter 4, ‘Totals and Subtotals for Categorical Variables,’” on page 33 for more information.
Chapter 3. 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`. See the topic [data files](#) for more information.

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.

1. From the menus, choose:
   - **Analyze > Tables > Custom Tables...**
2. In the table builder, drag and drop `Gender` from the variable list to the Rows area on the canvas pane.
3. 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.
4. Click **OK** to create the table.

![Figure 11. Table of categorical variables stacked in rows](#)

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.

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

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to delete any previous selections in the table builder.
3. In the table builder, drag and drop Age category from the variable list to the right of Gender in the Row area on the canvas pane.
4. Drag and drop Gender from the variable list to the right of Age category in the Row 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.
   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.
5. Click the variable label Gender on the canvas pane.
6. 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.
7. Click OK to create the table.
Note: Custom Tables do not honor layered split file processing. To achieve the same result as layered split files, place the split file variables in the outermost nesting layers of the table.

**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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Right-click Age category on the canvas pane and deselect Show Variable Label on the pop-up menu.
3. Do the same for Gender.
   The variable labels are still displayed in the table preview, but they won't be included in the table.
4. Click OK to create the table.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Age</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Less than 25</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 to 34</td>
<td>276</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>35 to 44</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45 to 54</td>
<td>221</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>55 to 64</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65 or older</td>
<td>178</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 13. Table of Age category nested within Gender*

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.

5. Open the table builder (Analyze menu, Tables, Custom Tables).
6. Click the Titles tab.
7. Click anywhere in the Title text box.
8. 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.
9. Click OK to create the table.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Less than 25</td>
</tr>
<tr>
<td></td>
<td>25 to 34</td>
</tr>
<tr>
<td></td>
<td>35 to 44</td>
</tr>
<tr>
<td></td>
<td>45 to 54</td>
</tr>
<tr>
<td></td>
<td>55 to 64</td>
</tr>
<tr>
<td></td>
<td>65 or older</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Female</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 25</td>
</tr>
<tr>
<td></td>
<td>25 to 34</td>
</tr>
<tr>
<td></td>
<td>35 to 44</td>
</tr>
<tr>
<td></td>
<td>45 to 54</td>
</tr>
<tr>
<td></td>
<td>55 to 64</td>
</tr>
<tr>
<td></td>
<td>65 or older</td>
</tr>
</tbody>
</table>

*Figure 14. Nested table without variable labels*
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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. If Age category isn’t already nested within Gender in the rows, follow the directions above for nesting them.
3. 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 you can:
   - 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.
4. Click OK to create the table.

**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.

---

**Figure 15. Variable labels in table title**

**Figure 16. Nested crosstabulation**

**Figure 16. Nested crosstabulation**
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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Right-click anywhere on the canvas pane and select Swap Row and Column Variables from the pop-up menu.
   The row and column variables have now been switched.
3. Select Hide to suppress the display of the summary statistics column label.
4. Right-click Gender on the canvas pane and deselect Show Variable Label.
5. Now click OK to create the table.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age category</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Less than 25</td>
<td>25 to 34</td>
</tr>
<tr>
<td></td>
<td>35 to 44</td>
<td>45 to 54</td>
</tr>
<tr>
<td></td>
<td>55 or older</td>
<td>55 or older</td>
</tr>
<tr>
<td></td>
<td>Less than 34</td>
<td>34 to 44</td>
</tr>
<tr>
<td></td>
<td>44 to 54</td>
<td>54 to 64</td>
</tr>
<tr>
<td></td>
<td>64 or older</td>
<td>64 or older</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>69</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>197</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>87</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>218</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>261</td>
</tr>
<tr>
<td>No</td>
<td>49</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>122</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>152</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>44</td>
</tr>
</tbody>
</table>

Figure 17. Crosstabulation with demographic variables nested in columns

Layers

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

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

<table>
<thead>
<tr>
<th>Gender Male</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Category</td>
<td></td>
</tr>
<tr>
<td>Less than 25</td>
<td>108</td>
</tr>
<tr>
<td>25 to 34</td>
<td>276</td>
</tr>
<tr>
<td>35 to 44</td>
<td>338</td>
</tr>
<tr>
<td>45 to 54</td>
<td>221</td>
</tr>
<tr>
<td>55 to 64</td>
<td>136</td>
</tr>
<tr>
<td>65 or older</td>
<td>178</td>
</tr>
</tbody>
</table>

Figure 18. Simple layered table

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.

7. Double-click the table in the Viewer window to activate it.
8. You can now see that the label Gender Male is actually a choice in a drop-down list.
9. Click the down arrow on the drop-down list to display the whole list of layers.
In this table, there is only one other choice in the list.

10. Select Gender Female from the drop-down list.

<table>
<thead>
<tr>
<th>Gender Female</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age category</td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>134</td>
</tr>
<tr>
<td>25 to 34</td>
<td>351</td>
</tr>
<tr>
<td>35 to 44</td>
<td>370</td>
</tr>
<tr>
<td>45 to 54</td>
<td>264</td>
</tr>
<tr>
<td>55 to 64</td>
<td>164</td>
</tr>
<tr>
<td>65 or older</td>
<td>321</td>
</tr>
</tbody>
</table>

Figure 19. Simple layered table with different layer displayed

**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.*)

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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.
3. Drag and drop Highest degree from the variable list to the Layer list below Gender.
   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.
4. Click OK to create the table.
5. Double-click the table in the Viewer window to activate it.
6. Click the down arrow on the drop-down list to display the whole list of layers.

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. If you haven’t done so already, follow the directions above for creating a table of stacked layers.
3. In the Layer Output group, select **Show each combination of categories as a layer**. This is equivalent to nesting.
4. Click OK to create the table.
5. Double-click the table in the Viewer window to activate it.
6. Click the down arrow on the drop-down list to display the whole list of layers.

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 x 2 = 10).

**Printing Layered Tables**

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

1. Double-click the table in the Viewer window to activate it.
2. From the Viewer window menus, choose:
Format > Table Properties...

3. Click the Printing tab.
4. Select Print all layers.

You can also save this setting as part of a TableLook, including the default TableLook.
Chapter 4. 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, or layer.

Sample Data File

The examples in this chapter use the data file `survey_sample.sav`. See the topic [data files](#) for more information.

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**

1. From the menus, choose:
   - Analyze > Tables > Custom Tables...
2. In the table builder, drag and drop *Age category* from the variable list to the Rows area on the canvas pane.
3. Right-click *Age category* on the canvas pane and choose Summary Statistics from the pop-up menu.
4. In the Summary Statistics dialog box, select Column N % in the Statistics list and click the arrow to add it to the Display list.
5. In the Label cell in the Display list, delete the default label and type Percent.
6. Click Apply to Selection.
7. Right-click *Age category* on the canvas pane and choose Categories and Totals from the pop-up menu.
8. Select (click) Total in the Categories and Totals dialog box.
9. Click Apply and then click OK in the table builder to create the table.

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>242</td>
<td>9.9%</td>
</tr>
<tr>
<td>25 to 34</td>
<td>627</td>
<td>22.2%</td>
</tr>
<tr>
<td>35 to 44</td>
<td>679</td>
<td>24.9%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>481</td>
<td>17.0%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>320</td>
<td>11.3%</td>
</tr>
<tr>
<td>65 or older</td>
<td>479</td>
<td>17.8%</td>
</tr>
<tr>
<td>Total</td>
<td>2020</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

*Figure 20. Simple total for a single categorical variable*

**What You See Is What Gets Toted**

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Right-click *Age category* on the canvas pane and choose Categories and Totals from the pop-up menu.
3. Click the category labeled Less than 25 in the Label list.
4. Click the arrow key to the left of the Exclude list.
5. Click the category labeled 65 or older in the Label list.
6. 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.

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

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 to 34</td>
<td>637</td>
<td>20.8%</td>
</tr>
<tr>
<td>35 to 44</td>
<td>673</td>
<td>22.2%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>481</td>
<td>22.0%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>320</td>
<td>15.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2,107</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

*Figure 21. Total in table with excluded categories*

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

**Display Position of Totals**

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

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

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 to 34</td>
<td>637</td>
<td>20.8%</td>
</tr>
<tr>
<td>35 to 44</td>
<td>673</td>
<td>22.2%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>481</td>
<td>22.0%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>320</td>
<td>15.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2,107</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

*Figure 22. Total displayed above totaled categories*

**Totals for Nested Tables**

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

**Group Totals**

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

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Drag and drop **Gender** to the left of **Age category** on the canvas pane.
3. Right-click **Age category** on the canvas pane and choose **Categories and Totals** from the pop-up menu.
   Before creating the table, let's move the totals back below the totaled categories.
4. In the Totals and Subgroups Appear group, select **Below categories to which they apply**.
5. Click **Apply** to save the setting and return to the table builder.
6. Click **OK** to create the table.
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.
1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Right-click Gender on the canvas pane and choose Categories and Totals from the pop-up menu.
3. Select (click) Total in the Categories and Totals dialog box.
4. Click Apply and then click OK in the table builder to create the table.

Layer Variable Totals

Totals for layer variables are displayed as separate layers in the table.
1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Layers in the table builder to display the Layers list.
3. Drag and drop Gender from the row area on the canvas pane to the Layers list.
   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.
4. Click OK to create the table.
5. Double-click the table in the Viewer to activate it.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>25 to 34</td>
<td>276</td>
<td>20.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35 to 44</td>
<td>309</td>
<td>22.9%</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>45 to 54</td>
<td>221</td>
<td>16.6%</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>55 to 64</td>
<td>136</td>
<td>10.0%</td>
</tr>
<tr>
<td>Female</td>
<td>Total</td>
<td></td>
<td>942</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>25 to 34</td>
<td></td>
<td>276</td>
<td>20.3%</td>
</tr>
<tr>
<td></td>
<td>35 to 44</td>
<td></td>
<td>309</td>
<td>22.9%</td>
</tr>
<tr>
<td></td>
<td>45 to 54</td>
<td></td>
<td>221</td>
<td>16.6%</td>
</tr>
<tr>
<td></td>
<td>55 to 64</td>
<td></td>
<td>136</td>
<td>10.0%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>942</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Age category</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>25 to 34</td>
<td>276</td>
<td>20.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35 to 44</td>
<td>309</td>
<td>22.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45 to 54</td>
<td>221</td>
<td>16.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55 to 64</td>
<td>136</td>
<td>10.0%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>942</td>
<td>100.0%</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>25 to 34</td>
<td>276</td>
<td>20.3%</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>35 to 44</td>
<td>309</td>
<td>22.9%</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>45 to 54</td>
<td>221</td>
<td>16.6%</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>55 to 64</td>
<td>136</td>
<td>10.0%</td>
</tr>
<tr>
<td>Female</td>
<td>Total</td>
<td></td>
<td>942</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Figure 23. Age category totals within Gender categories

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.
1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Right-click Gender on the canvas pane and choose Categories and Totals from the pop-up menu.
3. Select (click) Total in the Categories and Totals dialog box.
4. Click Apply and then click OK in the table builder to create the table.

Layer Variable Totals

Totals for layer variables are displayed as separate layers in the table.
1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Layers in the table builder to display the Layers list.
3. Drag and drop Gender from the row area on the canvas pane to the Layers list.
   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.
4. Click OK to create the table.
5. Double-click the table in the Viewer to activate it.
6. 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.

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 categories to which they apply 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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to clear any previous settings in the table builder.
3. In the table builder, drag and drop Age category from the variable list to the Rows area on the canvas pane.
4. Right-click Age category on the canvas pane and choose Categories and Totals from the pop-up menu.
5. Select 3.00 in the Value(s) list.
6. Click Add Subtotal to display the Define Subtotal dialog box.
7. In the Label text field, type Subtotal < 45.
8. Then click Continue.

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

9. Select 6.00 in the Value(s) list.
10. Click Add Subtotal to display the Define Subtotal dialog box.
11. In the Label text field, type Subtotal 45+.
12. Then click Continue.

*Important note:* You should select the display position for totals and subtotals (Above categories to which they apply or Below categories to which they apply) 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.

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

<table>
<thead>
<tr>
<th>Age category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25</td>
<td>242</td>
</tr>
<tr>
<td>25 to 34</td>
<td>627</td>
</tr>
<tr>
<td>35 to 44</td>
<td>673</td>
</tr>
<tr>
<td>45 to 54</td>
<td>1,540</td>
</tr>
<tr>
<td>55 to 64</td>
<td>300</td>
</tr>
<tr>
<td>65 or older</td>
<td>479</td>
</tr>
<tr>
<td>Subtotal 45+</td>
<td>1,280</td>
</tr>
</tbody>
</table>

*Figure 25. Subtotals for Age category*

**What You See Is What Gets Subtotaled**

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

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Right-click Age category on the canvas pane and choose Categories and Totals from the pop-up menu.
Note: 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.

3. Select 1.00 in the Value(s) list (or click the label Less than 25).
4. Click the arrow key to the left of the Exclude list.

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.

**Hiding Subtotaled Categories**

You can suppress the display of the categories that define a subtotal and display only the subtotal, effectively "collapsing" categories without affecting the underlying data.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to clear any previous settings in the table builder.
3. In the table builder, drag and drop Age category from the variable list to the Rows area on the canvas pane.
4. Right-click Age category on the canvas pane and choose Categories and Totals from the pop-up menu.
5. Select 3.00 in the Value(s) list.
6. Click Add Subtotal to display the Define Subtotal dialog box.
7. In the Label text field, type Less than 45.
8. Select (check) Hide subtotaled categories from the table.
9. Then click Continue.
   
   This inserts a row containing the subtotal for the first three age categories.
10. Select 6.00 in the Value(s) list.
11. Click Add Subtotal to display the Define Subtotal dialog box.
12. In the Label text field, type 45 or older.
14. Then click Continue.
15. To include a total with the subtotals, select (check) Total in the Show group.
16. Click Apply.
   
   The canvas reflects the fact that subtotals will be displayed but the categories that define the subtotals will be excluded.
17. Click OK to produce the table.

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 45</td>
<td>1280</td>
</tr>
<tr>
<td>45 or older</td>
<td>1280</td>
</tr>
<tr>
<td>Total</td>
<td>2560</td>
</tr>
</tbody>
</table>

*Figure 26. Table displaying only subtotals and totals*

**Layer Variable Subtotals**

Just like totals, subtotals for layer variables are displayed as separate layers in the table. Essentially, the subtotals are treated as categories. Each category is a separate layer in the table, and the display order of the layer categories is determined by the category order specified in the Categories and Totals dialog box, including the display position of the subtotal categories.
Chapter 5. Computed Categories for Categorical Variables

You can include computed categories in custom tables. These are new categories that are calculated from categories of the same variable at any nesting level in any dimension–row, column, or layer. For example, you could include a computed category that shows the difference between two categories.

Sample Data File

The examples in this chapter use the data file `survey_sample.sav`. See the topic data files for more information.

Simple Computed Category

1. From the menus, choose:
   `Analyze > Tables > Custom Tables...`
2. In the table builder, drag and drop `Age category` from the variable list to the Rows area on the canvas pane.
3. Right-click `Age category` on the canvas pane and choose `Categories and Totals` from the pop-up menu.
4. Select `3.00` in the Value(s) list.
5. Click `Add Category` to display the Define Compute Category dialog box.
6. In the Label for Computed Category text field, type `Less than 45`.
7. Select `Less than 25 (1.00)` in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box. `[1]` is displayed in the expression.
8. Click the plus (+) operator button in the dialog box (or press the + key on the keyboard).
9. Select `25 to 34 (2.00)` in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
10. Click the plus (+) operator button in the dialog box (or press the + key on the keyboard).
11. Select `35 to 44 (3.00)` in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
12. Then click `Continue`.
   This inserts a row containing the subtotal for the first three age categories.
13. Select `5.00` in the Value(s) list.
14. Click `Add Subtotal` to display the Define Subtotal dialog box.
15. In the Label text field, type `Less than 65`.
16. Then click `Continue`.
   This inserts a row containing the subtotal for the first five age categories.
17. Click `Apply` and then click `OK` in the table builder to create the table.
The table includes a computed category (Less than 45) and a subtotal (Less than 65). The subtotal includes categories also included in the computed category. You could not create the same table with subtotals alone, because subtotals cannot share the same categories.

**Hiding Categories in a Computed Category**

As with subtotals, you can suppress the display of the categories that are used in a computed category’s expression and display only the computed category itself. The following example builds on the previous one.

1. From the menus, choose: **Analyze > Tables > Custom Tables...**
2. Right-click **Age category** on the canvas pane and choose **Categories and Totals** from the pop-up menu.
3. Select the **Less than 45** computed category in the Value(s) list.
4. Click **Edit** to display the Define Compute Category dialog box.
5. Select **Hide categories used in expression from table**.
6. Then click **Continue**.
7. Select the **Less than 65** subtotal in the Value(s) list.
8. Click **Edit** to display the Define Subtotal dialog box.
9. Select **Hide subtotaled categories from the table**.
10. Then click **Continue**.
11. Click **Apply** and then click **OK** in the table builder to create the table.

![Figure 27. Computed category with subtotal](image1)

Like the previous example, the table includes a computed category and a subtotal. But in this case the categories in each are hidden so that only these totals are shown.

**Referencing Subtotals in a Computed Category**

You can include subtotals in a computed category’s expression.

1. From the menus, choose: **Analyze > Tables > Custom Tables...**
2. Click **Reset** to clear any previous settings in the table builder.
3. In the table builder, drag and drop **Labor force status** from the variable list into the Rows area of the canvas pane.

![Figure 28. Computed category with subtotal and hidden categories](image2)
4. Drag and drop Marital status from the variable list into the Columns area.

5. Right-click Labor force status on the canvas pane and choose Categories and Totals from the pop-up menu.

6. Select 2 in the Value(s) list.
7. Click Add Subtotal to display the Define Subtotal dialog box.
8. In the Label text field, type Working.
9. Select Hide subtotaled categories from the table.
10. Then click Continue.
    
    This inserts a row containing the subtotal for the first two working status categories.

11. Select 8 in the Value(s) list.
12. Click Add Subtotal to display the Define Subtotal dialog box.
13. In the Label text field, type Not Working.
15. Then click Continue.
    
    This inserts a row containing the subtotal for the other working status categories.

16. Select the Not Working subtotal in the Value(s) list.
17. Click Add Category to display the Define Compute Category dialog box.
18. In the Label for Computed Category text field, type Working / Not Working.
19. Select Working (Working #1) in the Totals and Subtotals list and click the arrow button to copy it to the Expression for Computed Category text box.
20. Click the division (/) operator button in the dialog box (or press the / key on the keyboard).
21. Select Not Working (Not Working #2) in the Totals and Subtotals list and click the arrow button to copy it to the Expression for Computed Category text box.

    By default, the computed category uses the same format as the variable's statistic, which is Count in this case. Because we want to show decimal places resulting from the division in the computed category's expression and the default format for Count does not include decimal places, we need to change the format.

22. Click the Display Formats tab.
23. Change the Decimals setting for Count to 2.
24. Then click Continue.
25. Click Apply and then click OK in the table builder to create the table.

<table>
<thead>
<tr>
<th>Marital status</th>
<th>Married</th>
<th>Widowed</th>
<th>Divorced</th>
<th>Separated</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
</tr>
<tr>
<td>Labor force status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working</td>
<td>516</td>
<td>64</td>
<td>230</td>
<td>87</td>
<td>494</td>
</tr>
<tr>
<td>Not Working</td>
<td>429</td>
<td>219</td>
<td>116</td>
<td>26</td>
<td>166</td>
</tr>
<tr>
<td>Working / Not Working</td>
<td>214</td>
<td>29</td>
<td>2.81</td>
<td>2.98</td>
<td>2.92</td>
</tr>
</tbody>
</table>

Figure 29. Computed category showing ratio of subtotals

The table includes two subtotals and a computed category. The computed category shows the ratio of the subtotals so that you can easily compare the groups represented by each subtotal. There's a much lower ratio of working to not working widowed respondents compared to the other groups. Also, there is a slightly lower ratio of married respondents, perhaps resulting from spouses who leave the workforce to stay home with a child.
Using Computed Categories to Display Nonexhaustive Subtotals

Subtotals are exhaustive. That is, all subtotals in a table include all values above or below their positions in the table. Computed categories, on the other hand, are not exhaustive and allow you to sum a mix of categories in a table.

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. Click Reset to clear any previous settings in the table builder.
3. In the table builder, drag and drop Think of self as liberal or conservative from the variable list into the Rows area of the canvas pane.
4. Right-click Think of self as liberal or conservative on the canvas pane and choose Categories and Totals from the pop-up context menu.
5. Select 3 in the Value(s) list.
6. Click Add Category to display the Define Computed Category dialog box.
7. In the Label for Computed Category text field, type Liberal Subtotal. Note that there are four spaces before the text. These spaces are used for indentation in the resulting table.
8. Select Extremely liberal (1) in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
9. Click the plus (+) operator button in the dialog box (or press the + key on the keyboard).
10. Select Liberal (2) in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
11. Click the plus (+) operator button in the dialog box (or press the + key on the keyboard).
12. Select Slightly liberal (3) in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
13. Click Continue.
   This inserts a row containing the subtotal for the liberal categories.
14. Select 7 in the Value(s) list.
15. Click Add Category to display the Define Computed Category dialog box.
16. In the Label for Computed Category text field, type Conservative Subtotal. Note that there are four spaces before the text. These spaces are used for indentation in the resulting table.
17. Select Slight conservative (5) in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
18. Click the plus (+) operator button in the dialog box (or press the + key on the keyboard).
19. Select Conservative (6) in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
20. Click the plus (+) operator button in the dialog box (or press the + key on the keyboard).
21. Select Extremely conservative (7) in the Categories list and click the arrow button to copy it to the Expression for Computed Category text box.
22. Click Continue.
   This inserts a row containing the subtotal for the conservative categories.
23. Click Apply and then click OK in the table builder to create the table.
The table includes two computed categories that do not include all the categories displayed in the table. The Moderate category is not included in either computed category. You cannot create the same table with subtotals because subtotals are exhaustive.

Figure 30. Computed categories displaying nonexhaustive subtotals

The table includes two computed categories that do not include all the categories displayed in the table. The Moderate category is not included in either computed category. You cannot create the same table with subtotals because subtotals are exhaustive.
Chapter 6. Tables for Variables with Shared Categories

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

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. See the topic data files for more information.

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

Table of Counts

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. 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.)
3. Drag and drop the six confidence variables to the Rows area on the canvas pane.
   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 x 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.
4. 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.
5. Before creating the table, select (click) Hide for Position in the Summary Statistics group, since the summary statistic label Count isn’t really necessary.
6. Click OK to create the table.
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.)

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Right-click any one of the confidence variables in the table preview on the canvas pane and choose **Summary Statistics** from the pop-up menu.
3. Select **Row N %** in the Statistics list and click the arrow button to move it to the Display list.
4. Click 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.
5. Click **Apply to All** to apply the summary statistic change to all of the stacked variables in the table.
   
   *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.
6. Click **OK** to create the table.

![Figure 32. Table of stacked row variables with shared category labels in columns](image)

### Figure 32. Table of stacked row variables with shared category labels in columns

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.

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Right-click the first confidence variable in the table preview on the canvas pane and choose Categories and Totals from the pop-up menu.
3. Select (check) Total in the Categories and Totals dialog box and then click Apply.
   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.
4. Right-click any one of the confidence variables on the canvas pane and select Select All Row Variables from the pop-up 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).
5. Click Categories and Totals in the Define group.
6. If Total isn’t already selected (checked) in the Categories and Totals dialog box, select it now and then click Apply.
7. The Category Position drop-down list should be enabled again, since now all of the variables have the additional total category, so select Row Labels in Columns.
8. Click OK to create the table.

<table>
<thead>
<tr>
<th></th>
<th>A great deal</th>
<th>Only some</th>
<th>Hardly any</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence in banks &amp; financial institutions</td>
<td>25.3%</td>
<td>57.3%</td>
<td>16.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Confidence in education</td>
<td>27.2%</td>
<td>56.1%</td>
<td>16.7%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Confidence in major companies</td>
<td>27.5%</td>
<td>59.2%</td>
<td>13.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Confidence in movie tickets</td>
<td>45.0%</td>
<td>46.1%</td>
<td>8.9%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Confidence in press</td>
<td>9.5%</td>
<td>47.2%</td>
<td>43.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Confidence in television</td>
<td>19.4%</td>
<td>49.9%</td>
<td>39.7%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Figure 34. Table of row percentages for variables stacked in rows, categories and totals displayed in columns

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.

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Drag and drop Gender from the variable list to the left side of the Rows area.
   The stacked variables with shared categories are now nested within gender categories in the table preview.
3. Now drag and drop Gender to the right of one of the stacked confidence variables in the table preview.

Once again, the category labels have reverted to the row dimension, and the Category Position control is disabled. You now have one stacked variable that also has Gender nested within it, while the other stacked variables contain no nested variables. You could add Gender as a nested variable to each of the stacked variables, but then moving row labels to columns would result in the category labels for Gender being displayed in the columns, not the category labels for the stacked variables with the shared
categories. This is because *Gender* would now be the innermost nested variable, and changing the category position always applies to the innermost nested variable.
Chapter 7. Summary Statistics

Summary statistics include everything from simple counts for categorical variables to measures of
dispersion, such as the standard error of the mean for scale variables. It does not include significance tests
available on the Test Statistics tab in the Custom Tables dialog box. See the topic “Test Statistics” on page
61 for more information.

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. A complete list of summary
statistics is also available. See the topic “Summary Statistics” on page 5 for more information.

Sample Data File

The examples in this chapter use the data file survey_sample.sav. See the topic data files for more
information.

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**

1. From the menus, choose:
   - Analyze > Tables > Custom Tables...
2. In the table builder, drag and drop Age category from the variable list into the Rows area of the canvas pane.
3. Right-click Age category on the canvas pane and select Summary Statistics from the pop-up menu. (Since this is the only variable in the table, it is the statistics source variable.)
4. In the Summary Statistics dialog box, select Column N % in the Statistics list and click the arrow to add it to the Display list.
5. Click Apply to Selection.
6. In the table builder, drag and drop Get news from internet to the right of Age category on the canvas pane.
7. Right-click Age category on the canvas pane again. The Summary Statistics item on the pop-up menu is now disabled because Age category is not the innermost nested variable in the statistics source dimension.
8. Right-click Get news from internet on the canvas pane. The Summary Statistics item is enabled because it is now the summary statistics source variable, since 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.)
9. Drag and drop Get news from internet from the Rows area on the canvas pane into the Columns area.
10. Right-click Get news from internet on the canvas pane again. The Summary Statistics item on the pop-up 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**

1. 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.
   - The first thing you may notice is that the Count and Column N % summaries have been replaced with Mean—and if you right-click 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.
2. Drag and drop Hours per day watching TV from the Rows area into the Columns area above Get news from internet.
3. Right-click Hours per day watching TV and select Summary Statistics from the pop-up menu. (It's still the statistics source variable even when you move it to a different dimension.)
4. 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.)
5. In the Decimals cell, type 2.
6. Click Apply to Selection.
   - The table preview on the canvas pane now shows that the mean values will be displayed with two decimals.
7. Click **OK** to create the table.

### 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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click **Reset** to clear any previous settings in the table builder.
3. 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.)
4. Drag and drop the five news variables into the Rows area of the canvas pane.
   - The five news variables are stacked in the row dimension.
5. Click **Get news from newspapers** on the canvas pane so that only that variable is selected.
6. Now right-click **Get news from internet** and select **Summary Statistics** from the pop-up menu.
7. In the Summary Statistics dialog box, select **Column N %** 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.)
8. Then click **Apply to Selection**.
   - 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.
9. Right-click **Get news from newspapers** on the canvas pane and select **Summary Statistics** from the pop-up menu.
10. In the Summary Statistics dialog box, select **Column N %** from the Statistics list and click the arrow to add it to the Display list.
11. Then click **Apply to All**.

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

### Custom Total Summary Statistics for Categorical Variables

For categorical statistics source variables, you can include custom total summary statistics that are different from the statistics displayed for the categories of the variable. For example, for an ordinal variable, you could display percentages for each category and the mean or median for the custom total summary statistic.
1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click **Reset** to clear any previous settings in the table builder.
3. Click **Confidence in press** in the variable list, and then Ctrl-click **Confidence in TV** to select both variables.
4. Drag and drop the two variables into the Rows area of the canvas pane. This stacks the two variables in the row dimension.
5. Right-click either variable on the canvas pane and select **Select All Row Variables** from the pop-up menu. (They may both already be selected, but we want to make sure.)
6. Right-click the variable again and select **Categories and Totals** from the pop-up menu.
7. 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.
8. Right-click either variable on the canvas pane and select **Summary Statistics** from the pop-up menu.
9. 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.
10. Click **Column N %** in the Statistics list and click the arrow key to move it to the Display list.
11. Click (check) **Custom Summary Statistics for Totals and Subtotals**.
12. 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.
13. Click **Mean** in the custom summary Statistics list and click the arrow to move it to the custom summary Display list.
14. Click the **Format** cell for the mean in the Display list and select **nnnn** from the drop-down list of formats. (You may have to scroll up the list to find this choice.)
15. In the Decimals cell, type 2.
16. Click **Apply to All** to apply these settings to both variables in the table.
   - 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.
17. In the table builder, in the Summary Statistics group, select **Rows** from the Position drop-down list.
   - This moves all the summary statistics to the row dimension, displaying all summary statistics in a single column in the table.
18. Click **OK** to create the table.

![Table Preview](image)

**Figure 36. Categorical variables with custom total summary statistics**

**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.
1. From the menus, choose:
   
   **Edit > Options...**

2. In the Options dialog box, click the **Output Labels** tab.
3. In the Pivot Table Labeling group, select **Values and Labels** from the **Variable values in labels shown as** drop-down list.
4. Click **OK** to save this setting.
5. Open the table builder (Analyze menu, Tables, Custom Tables) and click **OK** to create the table again.

<table>
<thead>
<tr>
<th>Confidence in press</th>
<th>1 A great deal</th>
<th>Column %</th>
<th>2 Only some</th>
<th>Column %</th>
<th>3 Hardly any</th>
<th>Column %</th>
<th>Total</th>
<th>Column %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence in television</td>
<td>1 A great deal</td>
<td>Column %</td>
<td>2 Only some</td>
<td>Column %</td>
<td>3 Hardly any</td>
<td>Column %</td>
<td>Mean</td>
<td>2.34</td>
</tr>
</tbody>
</table>

**Figure 37. Values and labels displayed for variable categories**

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:

6. From the menus, choose:
   
   **Edit > Options...**

7. In the Options dialog box, click the **Output Labels** tab.
8. In the Pivot Table Labeling group, select **Labels** from the **Variable values in labels shown as** drop-down list.
9. Click **OK** to save this setting.
Chapter 8. Summarizing Scale Variables

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

See the topic “Summary Statistics for Scale Variables and Categorical Custom Totals” on page 8 for more information.

Sample Data File

The examples in this chapter use the data file survey_sample.sav. See the topic data files for more information.

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

Stacked Scale Variables

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

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. 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.
3. Drag and drop the three selected variables to the Rows area of the canvas pane.
   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.
4. Click OK to create the table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of respondent</td>
<td>45</td>
</tr>
<tr>
<td>Highest year of school completed</td>
<td>13</td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 38. Table of mean values of stacked scale variables

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.

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

3. 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.)

4. Click the **Format** cell for the median in the Display list and select **nnnn** from the drop-down list of formats.

5. In the Decimals cell, type 1.

6. Make the same changes for the mean in the Display list.

7. Click **Apply to All** to apply these changes to all three scale variables.

8. Click **OK** in the table builder to create the table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of respondent</td>
<td>45.6</td>
<td>42.0</td>
</tr>
<tr>
<td>Highest year of school completed</td>
<td>13.3</td>
<td>13.0</td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>2.9</td>
<td>2.0</td>
</tr>
</tbody>
</table>

*Figure 39. Mean and median displayed in table of stacked scale variables*

### 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**.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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 menu.
3. In the Summary Statistics dialog box, select **Count** in the Statistics list and click the arrow to add it to the Display list.
4. Then select **Valid N** in the Statistics list and click the arrow to add it to the Display list.
5. Click **Apply to All** to apply these changes to all three scale variables.
6. Click **OK** in the table builder to create the table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Count</th>
<th>Valid N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of respondent</td>
<td>45.6</td>
<td>42.0</td>
<td>2832</td>
<td>2828</td>
</tr>
<tr>
<td>Highest year of school completed</td>
<td>13.3</td>
<td>13.0</td>
<td>2832</td>
<td>2830</td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>2.9</td>
<td>2.0</td>
<td>2832</td>
<td>2337</td>
</tr>
</tbody>
</table>

*Figure 40. Count versus Valid N*

For all three variables, **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).

7. Open the table builder (Analyze menu, Tables, Custom Tables).
8. 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 menu.
9. 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.

10. 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.

11. Select **Missing** in the Statistics list and click the arrow key to add it to the Display list.

12. Click **Apply to All** to apply these changes to all three scale variables.

13. Click **OK** in the table builder to create the table.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of respondent</td>
<td>45.6</td>
<td>42.0</td>
<td>4</td>
</tr>
<tr>
<td>Highest year of school completed</td>
<td>13.3</td>
<td>13.0</td>
<td>12</td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>2.9</td>
<td>2.0</td>
<td>495</td>
</tr>
</tbody>
</table>

*Figure 41. Number of missing values displayed in table of scale summary statistics*

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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.
3. Right-click either of the two selected variables and select **Summary Statistics** from the pop-up menu.
4. 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.
5. Click **Apply to Selection** to apply the change to only the two selected variables.

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

6. Click **OK** to create the table.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of respondent</td>
<td>45.6</td>
<td>42.0</td>
<td></td>
</tr>
<tr>
<td>Highest year of school completed</td>
<td>13.3</td>
<td>13.0</td>
<td></td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>2.9</td>
<td>2.0</td>
<td>495</td>
</tr>
</tbody>
</table>

*Figure 42. Table of different summary statistics for different variables*

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.

7. Open the table builder (Analyze menu, Tables, Custom Tables).
8. In the Summary Statistics group in the table builder, select **Rows** from the Position drop-down list.
9. Click **OK** to create the table.
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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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 menu. This is because in a table with scale variables, the scale variables are always the statistics source variables.
3. Click OK to create the table.

![Figure 43. Summary statistics and variables both displayed in the row dimension](image)

[Table: Group Summaries in Categories]

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Age of respondent</td>
<td>45.5</td>
</tr>
<tr>
<td>Highest year of school completed</td>
<td>13.0</td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>2.9</td>
</tr>
<tr>
<td>Missing</td>
<td>213</td>
</tr>
</tbody>
</table>

Figure 43. Summary statistics and variables both displayed in the row dimension

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

**Multiple Grouping Variables**

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

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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.
   - Although there may be times when you want something like the second example above, it's not what we want in this case.
3. Click OK to create the table.

![Figure 44. Grouped scale summaries using a categorical column variable](image)
Nesting Categorical Variables within Scale Variables

Although the above table may provide the information you 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. Swapping the positions of the two row variables and 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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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.
3. Drag and drop the three scale variables onto the far left side of the Rows area, nesting the categorical variable Get news from internet within each of the three scale variables.
4. Click OK to create the table.

The choice of nesting order depends on the relationships or comparisons that you want to emphasize in the table. Changing the nesting order of the scale variables doesn’t change the summary statistics values; it changes only their relative positions in the table.
Chapter 9. Test Statistics

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. See the topic Chapter 3, “Stacking, Nesting, and Layers with Categorical Variables,” on page 25 for more information.

Sample Data File

The examples in this chapter use the data file survey_sample.sav. See the topic data files for more information.

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.

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. In the table builder, drag and drop Labor force status from the variable list into the Rows area of the canvas pane.
3. Drag and drop Marital status from the variable list into the Columns area.
4. Select Rows as the position for the summary statistics.
5. Select Labor force status and click Summary Statistics in the Define group.
6. Select Column N % in the Statistics list and add it to the Display list.
7. Click Apply to Selection.
8. In the Custom Tables dialog box, click the Test Statistics tab.
9. Select Tests of independence (Chi-square).
10. Click OK to create the table and obtain the chi-square test.
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.

<table>
<thead>
<tr>
<th>Labor force status</th>
<th>Married</th>
<th>Widowed</th>
<th>Divorced</th>
<th>Separated</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working full time</td>
<td>Count</td>
<td>778</td>
<td>44</td>
<td>205</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>57.6%</td>
<td>15.5%</td>
<td>66.1%</td>
<td>62.4%</td>
</tr>
<tr>
<td>Working part-time</td>
<td>Count</td>
<td>130</td>
<td>20</td>
<td>35</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>10.3%</td>
<td>15.5%</td>
<td>26.5%</td>
<td>7.1%</td>
</tr>
<tr>
<td>Temporarily not working</td>
<td>Count</td>
<td>23</td>
<td>2</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>1.7%</td>
<td>0.6%</td>
<td>7.3%</td>
<td>5.4%</td>
</tr>
<tr>
<td>Unemployed, lost of</td>
<td>Count</td>
<td>13</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>1.0%</td>
<td>0.0%</td>
<td>7.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Retired</td>
<td>Count</td>
<td>160</td>
<td>150</td>
<td>53</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>12.5%</td>
<td>11.6%</td>
<td>4.2%</td>
<td>0.4%</td>
</tr>
<tr>
<td>School</td>
<td>Count</td>
<td>9</td>
<td>1</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>0.7%</td>
<td>0.1%</td>
<td>6.1%</td>
<td>2.2%</td>
</tr>
<tr>
<td>Keeping house</td>
<td>Count</td>
<td>200</td>
<td>50</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>14.9%</td>
<td>3.4%</td>
<td>5.0%</td>
<td>4.6%</td>
</tr>
<tr>
<td>Other</td>
<td>Count</td>
<td>16</td>
<td>8</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>1.2%</td>
<td>0.6%</td>
<td>2.7%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Figure 47. Labor force status by Marital status

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.

1. Open the table builder again (Analyze menu, Tables, Custom Tables).
2. Drag and drop Gender from the variable list into the Columns area of the canvas pane above Marital status.

3. Click OK to create the table.

<table>
<thead>
<tr>
<th>Marital status</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square</td>
<td>245.637</td>
<td>542.599</td>
</tr>
<tr>
<td>df</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Sig</td>
<td>.000*</td>
<td>.000*</td>
</tr>
</tbody>
</table>

* 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.

Figure 49. Pearson's chi-square test

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, so the results of the tests are suspect.

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:

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

<table>
<thead>
<tr>
<th>Lowest degree</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital status</td>
<td>245.637</td>
<td>542.599</td>
</tr>
<tr>
<td>df</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Sig</td>
<td>.000*</td>
<td>.000*</td>
</tr>
</tbody>
</table>

* 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.

Figure 50. Pearson's chi-square test

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.

1. From the menus, choose: Analyze > Tables > Custom Tables...
2. Click Reset to restore the default settings to all tabs.
3. In the table builder, drag and drop Hours per day watching TV from the variable list into the Rows area of the canvas pane.
4. Drag and drop Get news from newspapers from the variable list into the Columns area.
5. Select Hours per day watching TV and click Summary Statistics in the Define group.
6. Select nnnn as the format.
7. Select 2 as the number of decimals to display. Notice that this causes the format to now read nnnn.nn.
8. Click Apply to Selection.
9. In the Custom Tables dialog box, click the Test Statistics tab.
10. Select Compare column means (t-tests).
11. Click OK to create the table and obtain the column means tests.

<table>
<thead>
<tr>
<th>Get news from newspapers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>2.62</td>
</tr>
<tr>
<td>Yes</td>
<td>2.74</td>
</tr>
</tbody>
</table>

Figure 51. Get news from newspapers by Hours per day watching TV

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.

<table>
<thead>
<tr>
<th>Get news from newspapers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>(a)</td>
</tr>
<tr>
<td>Yes</td>
<td>(b)</td>
</tr>
</tbody>
</table>

Figure 52. Comparisons of column means

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.
Significance Results in APA-style Notation

If you do not want the significance results in a separate table, you can choose to display them in the main table. Significance results are identified using an APA-style notation with subscript letters. Complete the previous steps for comparing column means, but make the following change on the Test Statistics tab:

1. In the Identify Significant Differences area, select In the main table using APA-style subscripts.
2. Click OK to create the table and obtain the column means tests using APA-style notation.

<table>
<thead>
<tr>
<th>Get news from newspapers</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Hours per day watching TV</td>
<td>2.97*</td>
<td>2.74*</td>
</tr>
</tbody>
</table>

Figure 53. Comparisons of column means using APA-style notation

The column means test table assigns a subscript letter to the categories of the column variable. For each pair of columns, the column means are compared using a $t$ test. If a pair of values is significantly different, the values have different subscript letters assigned to them. Since there are only two columns, only one test is performed. Because the column means in this example share the same subscript letter, 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.

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

<table>
<thead>
<tr>
<th>Get news from newspapers</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
</tr>
<tr>
<td>Labor force status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working full time</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Working part-time</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Temporarily not working</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Unemployed, laid off</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Retired</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>School</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Keeping house</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Other</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
</tbody>
</table>

Figure 54. Comparisons of column means

With Hours per day 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 the
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:

4. Open the table builder again (Analyze menu, Tables, Custom Tables).
5. Drag and drop Get news from internet from the variable list into the Columns area to the left of Get news from newspapers.
6. Click **OK** to create the table.

<table>
<thead>
<tr>
<th>Labor force status</th>
<th>Get news from internet</th>
<th>Get news from newspapers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working full time</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Working part-time</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Temporarily not working</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Unemployed, laid off</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Retired</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>School</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Keeping house</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
<tr>
<td>Other</td>
<td>Hours per day watching TV</td>
<td>Hours per day watching TV</td>
</tr>
</tbody>
</table>

Figure 55. Comparisons of column means

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.

1. From the menus, choose:
   **Analyze > Tables > Custom Tables**...
2. Click **Reset** to restore the default settings to all tabs.
3. In the table builder, drag and drop Labor force status from the variable list into the Rows area of the canvas pane.

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

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

6. Select Column N % in the Statistics list and add it to the Display list.

7. Deselect Count from the Display list.

8. Click Apply to Selection.

9. In the Custom Tables dialog box, click the Test Statistics tab.

10. Select Compare column proportions (z-tests).

11. Click OK to create the table and obtain the column proportions tests.

<table>
<thead>
<tr>
<th>Labor force status</th>
<th>Married</th>
<th>Widowed</th>
<th>Divorced</th>
<th>Separated</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column %</td>
<td>Column %</td>
<td>Column %</td>
<td>Column %</td>
<td>Column %</td>
</tr>
<tr>
<td>Working full-time</td>
<td>57.0%</td>
<td>15.5%</td>
<td>8.1%</td>
<td>82.4%</td>
<td>59.1%</td>
</tr>
<tr>
<td>Working part-time</td>
<td>10.3%</td>
<td>7.1%</td>
<td>7.0%</td>
<td>9.7%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Temporarily not-working</td>
<td>1.7%</td>
<td>1.7%</td>
<td>2.0%</td>
<td>1.1%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Unemployed, laid off</td>
<td>0.0%</td>
<td>1.1%</td>
<td>2.2%</td>
<td>0.8%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Retired</td>
<td>12.5%</td>
<td>53.0%</td>
<td>11.9%</td>
<td>6.5%</td>
<td>2.6%</td>
</tr>
<tr>
<td>School</td>
<td>2.7%</td>
<td>1.4%</td>
<td>1.6%</td>
<td>2.2%</td>
<td>9.0%</td>
</tr>
<tr>
<td>Keeping house</td>
<td>14.3%</td>
<td>19.4%</td>
<td>5.6%</td>
<td>14.0%</td>
<td>5.3%</td>
</tr>
<tr>
<td>Other</td>
<td>2.2%</td>
<td>2.8%</td>
<td>2.7%</td>
<td>4.3%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

Figure 56. Labor force status by Marital status

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

<table>
<thead>
<tr>
<th>Labor force status</th>
<th>Married</th>
<th>Widowed</th>
<th>Divorced</th>
<th>Separated</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N)</td>
<td>(E)</td>
<td>(C)</td>
<td>(D)</td>
<td>(E)</td>
</tr>
<tr>
<td>Working full-time</td>
<td>E</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>ABC</td>
</tr>
<tr>
<td>Working part-time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporarily not-working</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed, laid off</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>E</td>
<td>A C D E</td>
<td>E</td>
<td></td>
<td>A B</td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keeping house</td>
<td>C E</td>
<td>C E</td>
<td></td>
<td></td>
<td>A B C</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 57. Comparisons of column proportions

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 five levels of Marital status, (5*4)/2 = 10 pairs of columns 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 the C column. No other keys are reported in other columns. Thus, you can conclude that the proportion of divorced persons who are working full time is greater than the proportion of married persons working full time, which in turn is greater than the proportion of widowers working full time. The proportions of people who are separated or never married and working full time cannot be differentiated from people who are divorced or married and working full time, but these proportions are greater than the proportion of widowers working full time.

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

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

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

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

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

**Significance Results in APA-style Notation**

If you do not want the significance results in a separate table, you can choose to display them in the main table. Significance results are identified using an APA-style notation with subscript letters. Complete the previous steps for comparing column proportions, but make the following change on the Test Statistics tab:

1. In the Identify Significant Differences area, select **In the main table using APA-style subscripts**.
2. Click **OK** to create the table and obtain the column means tests using APA-style notation.

<table>
<thead>
<tr>
<th>Labor force status</th>
<th>Married</th>
<th>Widowed</th>
<th>Divorced</th>
<th>Separated</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column N %</td>
<td>Column N %</td>
<td>Column N %</td>
<td>Column N %</td>
<td>Column N %</td>
</tr>
<tr>
<td>Working full time</td>
<td>57.0%</td>
<td>15.5%</td>
<td>66.1%</td>
<td>82.4%</td>
<td>59.1%</td>
</tr>
<tr>
<td>Working part time</td>
<td>10.3%</td>
<td>7.1%</td>
<td>7.0%</td>
<td>8.7%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Temporarily not working</td>
<td>1.7%</td>
<td>7.7%</td>
<td>2.0%</td>
<td>1.1%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Unemployed, laid off</td>
<td>1.0%</td>
<td>1.1%</td>
<td>2.2%</td>
<td>0.0%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Retired</td>
<td>12.5%</td>
<td>53.0%</td>
<td>11.9%</td>
<td>6.5%</td>
<td>7.5%</td>
</tr>
<tr>
<td>School</td>
<td>7.3%</td>
<td>4.3%</td>
<td>1.6%</td>
<td>2.2%</td>
<td>6.6%</td>
</tr>
<tr>
<td>Keeping house</td>
<td>14.9%</td>
<td>13.4%</td>
<td>5.6%</td>
<td>14.0%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Other</td>
<td>1.2%</td>
<td>2.9%</td>
<td>2.7%</td>
<td>4.3%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

*Figure 58. Comparisons of column proportions using APA-style notation*
The column proportions test table assigns a subscript letter to the categories of the column variable. For each pair of columns, the column proportions are compared using a z test. If a pair of values is significantly different, the values have different subscript letters assigned to them.

For the set of tests associated with Working full time, the Widowed category has a subscript letter not used in the other columns, and the Separated and Never Married categories share the same two subscripts. Considering these subscript letters and the actual proportions shown in the table, you can make the same analysis as demonstrated in the previous example with separate tables. Thus, you can conclude that the proportion of divorced persons who are working full time is greater than the proportion of married persons working full time, which in turn is greater than the proportion of widowers working full time. The proportions of people who are separated or never married and working full time cannot be differentiated from people who are divorced or married and working full time, but these proportions are greater than the proportion of widowers working full time. The rest of the analysis from the previous example applies.

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.

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

<table>
<thead>
<tr>
<th></th>
<th>Marital status</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Married</td>
<td>Widowed</td>
<td>Divorced</td>
<td>Separated</td>
<td>Never</td>
</tr>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
<td>married</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td><strong>Male</strong></td>
<td><strong>Labor</strong></td>
<td><strong>status</strong></td>
<td><strong>Working</strong></td>
<td><strong>time</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>force</td>
<td></td>
<td>full-time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>status</td>
<td></td>
<td>working</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>working</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>part-time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>temporarily not working</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unemployed, laid off</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retired</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>School</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Keeping house</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Working full time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Working part-time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>temporarily not working</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unemployed, laid off</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retired</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>School</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Keeping house</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

Figure 59. Comparisons of column proportions

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

<table>
<thead>
<tr>
<th>Marital status</th>
<th>Married</th>
<th>Widowed</th>
<th>Divorced</th>
<th>Separated</th>
<th>Never married</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
<td>(E)</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Labor force status</td>
<td>Working full time</td>
<td>Working part time</td>
<td>Temporarily not working</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Labor force status</td>
<td>Working full time</td>
<td>Working part time</td>
<td>Temporarily not working</td>
</tr>
<tr>
<td>Highest degree</td>
<td>LT High school</td>
<td>High school</td>
<td>Junior college</td>
<td>Bachelor</td>
<td>Graduate</td>
</tr>
</tbody>
</table>

Figure 60. Comparisons of column proportions

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 less than a 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.
A Note on Weights and Multiple Response Sets

Case weights are always based on counts, not responses, even when one of the variables is a multiple response variable.
Chapter 10. Multiple Response Sets

Custom Tables and the Chart Builder support 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 a data file. You can define and save multiple response sets in IBM SPSS Statistics data files, but you cannot import or export multiple response sets from/to other file formats. You can copy multiple response sets from other IBM SPSS Statistics data files using Copy Data Properties, which is accessed from the Data menu in the Data Editor window.

Sample Data File

The examples in this chapter use the data file survey_sample.sav. See the topic data files for more information.

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

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.

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. 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 61. Multiple dichotomy set icon

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

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.

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

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 N % 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 N % 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?

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Right-click News sources in the table preview on the canvas pane and select Summary Statistics from the pop-up menu.
3. 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.
4. Click Apply to Selection, and then click OK to create the table.

<table>
<thead>
<tr>
<th>News sources</th>
<th>Count</th>
<th>Column N % Responses</th>
<th>Column Responses %</th>
<th>Column Responses % (Base: Count)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Get news from internet</td>
<td>667</td>
<td>41.7%</td>
<td>667</td>
<td>41.7%</td>
</tr>
<tr>
<td>Get news from radio</td>
<td>581</td>
<td>29.5%</td>
<td>551</td>
<td>29.5%</td>
</tr>
<tr>
<td>Get news from television</td>
<td>1077</td>
<td>51.3%</td>
<td>1077</td>
<td>51.3%</td>
</tr>
<tr>
<td>Get news from news magazine</td>
<td>294</td>
<td>14.1%</td>
<td>294</td>
<td>14.1%</td>
</tr>
<tr>
<td>Get news from newspapers</td>
<td>805</td>
<td>38.7%</td>
<td>805</td>
<td>38.7%</td>
</tr>
<tr>
<td>Total</td>
<td>2081</td>
<td>100.0%</td>
<td>2081</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Figure 64. Column response percentages with count as the percentage base

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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.
3. Right-click Gender in the table preview on the canvas pane and deselect Show Variable Label on the pop-up menu.
4. 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).
5. Click OK to create the table.
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.

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

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to clear any previous settings.
3. 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.
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).

4. Right-click *Car maker, most recent cars* in the table preview on the canvas pane and select **Categories and Totals** from the pop-up menu.

5. Select (click) **Total** in the Categories and Totals dialog box, and then click **Apply**.

6. Right-click *Car maker, most recent cars* again and select **Summary Statistics** from the pop-up menu.

7. In the Summary Statistics dialog box, select **Responses** in the Statistics list and click the arrow to add it to the Display list.

8. Click **Apply to Selection**, and then click **OK** to create the table.

<table>
<thead>
<tr>
<th>Car maker, most recent cars</th>
<th>Count</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>1530</td>
<td>1530</td>
</tr>
<tr>
<td>Japanese</td>
<td>1327</td>
<td>1327</td>
</tr>
<tr>
<td>Korean</td>
<td>695</td>
<td>695</td>
</tr>
<tr>
<td>German</td>
<td>693</td>
<td>693</td>
</tr>
<tr>
<td>Swedish</td>
<td>360</td>
<td>360</td>
</tr>
<tr>
<td>Other</td>
<td>343</td>
<td>343</td>
</tr>
<tr>
<td>Total</td>
<td>2652</td>
<td>2652</td>
</tr>
</tbody>
</table>

Figure 67. Multiple category set: Counts and responses without duplicates

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.

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

10. Click the **Options** tab.

11. Click (check) **Count duplicate responses for multiple category sets**.

12. Click **OK** to create the table.

<table>
<thead>
<tr>
<th>Car maker, most recent cars</th>
<th>Count</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>1530</td>
<td>2730</td>
</tr>
<tr>
<td>Japanese</td>
<td>1327</td>
<td>1717</td>
</tr>
<tr>
<td>Korean</td>
<td>695</td>
<td>783</td>
</tr>
<tr>
<td>German</td>
<td>693</td>
<td>754</td>
</tr>
<tr>
<td>Swedish</td>
<td>360</td>
<td>383</td>
</tr>
<tr>
<td>Other</td>
<td>343</td>
<td>359</td>
</tr>
<tr>
<td>Total</td>
<td>2652</td>
<td>6770</td>
</tr>
</tbody>
</table>

Figure 68. Multiple category set with duplicate responses included

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.

**Significance Testing with Multiple Response Sets**

You can use multiple response sets in significance tests in essentially the same way you would use categorical variables.

- For tests of independence (chi-square) or comparing column proportions (z-tests), tests are performed on counts, and Count must be one of the summary statistics displayed in the table.
• For multiple category sets, tests comparing column proportions or column means (t-tests) are not performed if Count duplicate responses for multiple category sets is selected on the Options tab. See the topic "Custom Tables: Options Tab" on page 14 for more information.

**Tests of Independence with Multiple Response Sets**

This example creates a crosstabulation of a categorical variable and a multiple response set and performs a chi-square test of independence on the crosstabulation.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to clear any previous settings.
3. Drag and drop News sources (this is the descriptive label for the multiple dichotomy set $mltnews) from the variable list into the Columns area of the canvas pane.
4. Drag and drop Gender from the variable list into the Rows area of the canvas pane.
5. Click the Test Statistics tab.
7. If it is not already selected, select Include multiple response variables in test.
8. Click OK to run the procedure.

<table>
<thead>
<tr>
<th>Gender</th>
<th>News sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.285</td>
</tr>
<tr>
<td>df</td>
<td>5</td>
</tr>
<tr>
<td>Sig</td>
<td>.068</td>
</tr>
</tbody>
</table>

Figure 69. Chi-square results

The significance level of 0.068 for the chi-square test indicates that males and females probably do not differ significantly in their choices of news sources (assuming you use a significance value of 0.05 or lower as your criterion for determining statistical significance).

**Comparing Column Means with Multiple Response Sets**

This example calculates means of a scale variable within categories defined by a multiple response set and compares each category mean to every other category mean for significant differences.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click Reset to clear any previous settings.
3. Drag and drop News sources (this is the descriptive label for the multiple dichotomy set $mltnews) from the variable list into the Columns area of the canvas pane.
4. Drag and drop Age of respondent into the Rows area of the canvas pane.
5. Click the Test Statistics tab.
7. If it is not already selected, select Include multiple response variables in test.
8. Click OK to run the procedure.
Each category of the multiple response set is identified by a letter (A, B, C, D, E), and for each category for which the mean of another category is both lower and differs significantly from the mean of that category, the letter representing the category with the lower mean is displayed.

- *Get news from newspapers* (A) has the highest mean age, and all other category means differ significantly from it.
- *Get news from television* (C) has the next highest mean age, and all remaining category means (B, D, and E) differ significantly from it. (C also differs significantly from A, as previously indicated.)
- The mean ages for *Get news from magazine* (B), *Get news from radio* (D), and *Get news from internet* (E) do not differ significantly from each other.
Chapter 11. Missing Values

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

There are two kinds of missing values in IBM SPSS Statistics:

- **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.

There are 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*. See the topic data files for more information. This is a very simple, completely artificial data file, with only one variable and ten cases, designed to illustrate basic concepts about missing values.

Tables without Missing Values

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

1. From the menus, choose:
   - **Analyze > Tables > Custom Tables...**
2. 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.
3. Right-click the variable on the canvas pane and select **Categories and Totals** from the pop-up menu.
4. Click (check) **Total** in the Categories and Totals dialog box, and then click **Apply**.
5. Right-click *Variable with missing values* in the table preview on the canvas pane again and select **Summary Statistics** from the pop-up menu.
6. In the Summary Statistics dialog box, select **Column N %** in the Statistics list and click the arrow to add it to the Display list.
7. Click **Apply to Selection**.
   - 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.
8. Click **OK** to create the table.
Including Missing Values in Tables

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Right-click Variable with missing values in the table preview on the canvas pane and select Categories and Totals from the pop-up menu.
3. Click (check) Missing Values in the Categories and Totals dialog box, and then click Apply.
   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.
4. Right-click Variable with missing values in the table preview on the canvas pane again and select Summary Statistics from the pop-up menu.
5. In the Summary Statistics dialog box, click (check) Custom Summary Statistics for Totals and Subtotals.
6. Select Valid N in the custom summary Statistics list and click the arrow to add it to the Display list.
7. Do the same for Total N.
8. Click Apply to Selection, and then click OK in the table builder to create the table.

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.

9. Open the table builder (Analyze menu, Tables, Custom Tables).
10. Right-click Variable with missing values in the table preview on the canvas pane and select Summary Statistics from the pop-up menu.
11. 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.
12. Do the same for Column Total N %.
13. You can also add them both to the list of custom summary statistics for totals and subtotals.
14. Click **Apply to Selection**, and then click **OK** to create the table.

<table>
<thead>
<tr>
<th>Variable with missing values</th>
<th>Count</th>
<th>Column N %</th>
<th>Column Valid N %</th>
<th>Column Total N %</th>
<th>Valid N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>2</td>
<td>22.2%</td>
<td>26.6%</td>
<td>20.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>3</td>
<td>33.3%</td>
<td>42.5%</td>
<td>30.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>2</td>
<td>22.2%</td>
<td>28.6%</td>
<td>20.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Don’t know</td>
<td>1</td>
<td>11.1%</td>
<td>5.0%</td>
<td>10.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not applicable</td>
<td>1</td>
<td>11.1%</td>
<td>0.0%</td>
<td>10.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

*Figure 73. Table with missing values and valid and total percentages*

- **Column N %** is the percentage in each category based on the number of non-missing and user-missing values (since user-missing values have been explicitly included in the table).
- **Column Valid N %** is the percentage in each category based on only the valid, non-missing cases. These values are the same as the column percentages were in the original table that did not include user-missing values.
- **Column Total N %** is the percentage in each category based on all cases, including both user-missing and system-missing. If you add up the individual category percentages in this category, you'll see that they add up to only 90%, because one case out of the total of 10 cases (10%) has the system-missing value. Although this case is included in the base for the percentage calculations, no category is provided in the table for cases with system-missing values.
Chapter 12. Formatting and Customizing Tables

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. See the topic data files for more information.

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.

1. From the menus, choose:
   Analyze > Tables > Custom Tables...
2. In the table builder, drag and drop Age category from the variable list into the Rows area on the canvas pane.
3. Drag and drop Confidence in television below Age category in the Rows area, stacking the two variables in the row dimension.
4. Right-click Age category in the table preview on the canvas pane and select Select All Row Variables from the pop-up menu.
5. Right-click Age category again and select Categories and Totals from the pop-up menu.
6. In the Categories and Totals dialog box, select (check) Total and then click Apply.
7. Right-click either variable in the table preview on the canvas pane and select Summary Statistics from the pop-up menu.
8. Select Column N % in the Statistics list and click the arrow key to add it to the Display list.
10. In the Statistics list for custom summary statistics, select Column N % and click the arrow to add it to the Display list.
11. Do the same for **Mean**.
12. Then click **Apply to All**.
   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 percentage 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.

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.

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

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

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

---

**Display Labels for Summary Statistics**

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

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click **Reset** to clear any previous settings in the table builder.
3. In the table builder, drag and drop **Age category** from the variable list into the Rows area on the canvas pane.
4. Drag and drop **How get paid last week** from the variable list into the Columns area on the canvas pane.
5. Right-click **Age category** in the table preview on the canvas pane and select **Summary Statistics** from the pop-up context menu.
6. Select **Column N %** in the Statistics list and click the arrow key to add it to the Display list.
7. 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 **%**.
8. 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 N % statistic to remove the unnecessary percentage sign (since the column label indicates that the column contains percentages).
9. Click the Format cell for **Column N %** and select **nnnn.n** from the drop-down list of formats.
10. Then click **Apply to Selection**.
    The table preview displays the modified display format and the modified labels.
11. Click **OK** to create the table.
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.)

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. Click the Options tab.
3. In the Width for Data Columns group, select Custom.
4. For the Maximum, type 36. (Make sure that the Units setting is Points.)
5. Click OK to create the table.

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.

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

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.

1. Open the table builder (Analyze menu, Tables, Custom Tables).
2. 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.
3. 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.
4. Select Valid N in the Statistics list and click the arrow key to add it to the Display list.
5. Click Apply to Selection.
6. Click the Options tab.
7. In the text field for Statistics that Cannot be Computed, type NA.
8. Click OK to create the table.

<table>
<thead>
<tr>
<th>Age category</th>
<th>Less than 25</th>
<th>25 to 34</th>
<th>35 to 44</th>
<th>45 to 54</th>
<th>55 to 64</th>
<th>65 or older</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3</td>
<td>13</td>
<td>13</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Valid N</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Mean</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Valid N</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 77. Table with "NA" displayed for missing statistics

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 a slight discrepancy—one of those three Valid N values is displayed as a 0, rather than 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.
Chapter 13. Sample Files

The sample files installed with the product can be found in the Samples subdirectory of the installation directory. There is a separate folder within the Samples subdirectory for each of the following languages: English, French, German, Italian, Japanese, Korean, Polish, Russian, Simplified Chinese, Spanish, and Traditional Chinese.

Not all sample files are available in all languages. If a sample file is not available in a language, that language folder contains an English version of the sample file.

Descriptions

Following are brief descriptions of the sample files used in various examples throughout the documentation.

- **accidents.sav.** This is a hypothetical data file that concerns an insurance company that is studying age and gender risk factors for automobile accidents in a given region. Each case corresponds to a cross-classification of age category and gender.

- **adl.sav.** This is a hypothetical data file that concerns efforts to determine the benefits of a proposed type of therapy for stroke patients. Physicians randomly assigned female stroke patients to one of two groups. The first received the standard physical therapy, and the second received an additional emotional therapy. Three months following the treatments, each patient's abilities to perform common activities of daily life were scored as ordinal variables.

- **advert.sav.** This is a hypothetical data file that concerns a retailer’s efforts to examine the relationship between money spent on advertising and the resulting sales. To this end, they have collected past sales figures and the associated advertising costs.

- **aflatoxin.sav.** This is a hypothetical data file that concerns the testing of corn crops for aflatoxin, a poison whose concentration varies widely between and within crop yields. A grain processor has received 16 samples from each of 8 crop yields and measured the aflatoxin levels in parts per billion (PPB).

- **anorectic.sav.** While working toward a standardized symptomatology of anorectic/bulimic behavior, researchers ¹ made a study of 55 adolescents with known eating disorders. Each patient was seen four times over four years, for a total of 220 observations. At each observation, the patients were scored for each of 16 symptoms. Symptom scores are missing for patient 71 at time 2, patient 76 at time 2, and patient 47 at time 3, leaving 217 valid observations.

- **bankloan.sav.** This is a hypothetical data file that concerns a bank’s efforts to reduce the rate of loan defaults. The file contains financial and demographic information on 850 past and prospective customers. The first 700 cases are customers who were previously given loans. The last 150 cases are prospective customers that the bank needs to classify as good or bad credit risks.

- **bankloan_binning.sav.** This is a hypothetical data file containing financial and demographic information on 5,000 past customers.

- **behavior.sav.** In a classic example ², 52 students were asked to rate the combinations of 15 situations and 15 behaviors on a 10-point scale ranging from 0=”extremely appropriate” to 9=”extremely inappropriate.” Averaged over individuals, the values are taken as dissimilarities.

- **behavior_ini.sav.** This data file contains an initial configuration for a two-dimensional solution for behavior.sav.

---


© Copyright IBM Corporation 1989, 2013
• **brakes.sav.** This is a hypothetical data file that concerns quality control at a factory that produces disc brakes for high-performance automobiles. The data file contains diameter measurements of 16 discs from each of 8 production machines. The target diameter for the brakes is 322 millimeters.

• **breakfast.sav.** In a classic study, 21 Wharton School MBA students and their spouses were asked to rank 15 breakfast items in order of preference with 1 = "most preferred" to 15 = "least preferred." Their preferences were recorded under six different scenarios, from "Overall preference" to "Snack, with beverage only."

• **breakfast-overall.sav.** This data file contains the breakfast item preferences for the first scenario, "Overall preference," only.

• **broadband_1.sav.** This is a hypothetical data file containing the number of subscribers, by region, to a national broadband service. The data file contains monthly subscriber numbers for 85 regions over a four-year period.

• **broadband_2.sav.** This data file is identical to broadband_1.sav but contains data for three additional months.

• **car_insurance_claims.sav.** A dataset presented and analyzed elsewhere concerns damage claims for cars. The average claim amount can be modeled as having a gamma distribution, using an inverse link function to relate the mean of the dependent variable to a linear combination of the policyholder age, vehicle type, and vehicle age. The number of claims filed can be used as a scaling weight.

• **car_sales.sav.** This data file contains hypothetical sales estimates, list prices, and physical specifications for various makes and models of vehicles. The list prices and physical specifications were obtained alternately from edmunds.com and manufacturer sites.

• **car_sales_uprepared.sav.** This is a modified version of car_sales.sav that does not include any transformed versions of the fields.

• **carpet.sav.** In a popular example, a company interested in marketing a new carpet cleaner wants to examine the influence of five factors on consumer preference—package design, brand name, price, a Good Housekeeping seal, and a money-back guarantee. There are three factor levels for package design, each one differing in the location of the applicator brush; three brand names (K2R, Glory, and Bissell); three price levels; and two levels (either no or yes) for each of the last two factors. Ten consumers rank 22 profiles defined by these factors. The variable Preference contains the rank of the average rankings for each profile. Low rankings correspond to high preference. This variable reflects an overall measure of preference for each profile.

• **carpet_prefs.sav.** This data file is based on the same example as described for carpet.sav, but it contains the actual rankings collected from each of the 10 consumers. The consumers were asked to rank the 22 product profiles from the most to the least preferred. The variables PREF1 through PREF22 contain the identifiers of the associated profiles, as defined in carpet_plan.sav.

• **catalog.sav.** This data file contains hypothetical monthly sales figures for three products sold by a catalog company. Data for five possible predictor variables are also included.

• **catalog_seasfac.sav.** This data file is the same as catalog.sav except for the addition of a set of seasonal factors calculated from the Seasonal Decomposition procedure along with the accompanying date variables.

• **cellular.sav.** This is a hypothetical data file that concerns a cellular phone company’s efforts to reduce churn. Churn propensity scores are applied to accounts, ranging from 0 to 100. Accounts scoring 50 or above may be looking to change providers.

• **ceramics.sav.** This is a hypothetical data file that concerns a manufacturer’s efforts to determine whether a new premium alloy has a greater heat resistance than a standard alloy. Each case represents a separate test of one of the alloys; the heat at which the bearing failed is recorded.

---


• **cereal.sav.** This is a hypothetical data file that concerns a poll of 880 people about their breakfast preferences, also noting their age, gender, marital status, and whether or not they have an active lifestyle (based on whether they exercise at least twice a week). Each case represents a separate respondent.

• **clothing_defects.sav.** This is a hypothetical data file that concerns the quality control process at a clothing factory. From each lot produced at the factory, the inspectors take a sample of clothes and count the number of clothes that are unacceptable.

• **coffee.sav.** This data file pertains to perceived images of six iced-coffee brands. For each of 23 iced-coffee image attributes, people selected all brands that were described by the attribute. The six brands are denoted AA, BB, CC, DD, EE, and FF to preserve confidentiality.

• **contacts.sav.** This is a hypothetical data file that concerns the contact lists for a group of corporate computer sales representatives. Each contact is categorized by the department of the company in which they work and their company ranks. Also recorded are the amount of the last sale made, the time since the last sale, and the size of the contact’s company.

• **creditpromo.sav.** This is a hypothetical data file that concerns a department store’s efforts to evaluate the effectiveness of a recent credit card promotion. To this end, 500 cardholders were randomly selected. Half received an ad promoting a reduced interest rate on purchases made over the next three months. Half received a standard seasonal ad.

• **customer_dbase.sav.** This is a hypothetical data file that concerns a company’s efforts to use the information in its data warehouse to make special offers to customers who are most likely to reply. A subset of the customer base was selected at random and given the special offers, and their responses were recorded.

• **customer_information.sav.** A hypothetical data file containing customer mailing information, such as name and address.

• **customer_subset.sav.** A subset of 80 cases from customer_dbase.sav.

• **debate.sav.** This is a hypothetical data file that concerns paired responses to a survey from attendees of a political debate before and after the debate. Each case corresponds to a separate respondent.

• **debate_aggregate.sav.** This is a hypothetical data file that aggregates the responses in debate.sav. Each case corresponds to a cross-classification of preference before and after the debate.

• **demo.sav.** This is a hypothetical data file that concerns a purchased customer database, for the purpose of mailing monthly offers. Whether or not the customer responded to the offer is recorded, along with various demographic information.

• **demo_cs_1.sav.** This is a hypothetical data file that concerns the first step of a company’s efforts to compile a database of survey information. Each case corresponds to a different city, and the region, province, district, and city identification are recorded.

• **demo_cs_2.sav.** This is a hypothetical data file that concerns the second step of a company’s efforts to compile a database of survey information. Each case corresponds to a different household unit from cities selected in the first step, and the region, province, district, city, subdivision, and unit identification are recorded. The sampling information from the first two stages of the design is also included.

• **demo_cs.sav.** This is a hypothetical data file that contains survey information collected using a complex sampling design. Each case corresponds to a different household unit, and various demographic and sampling information is recorded.

• **dmdata.sav.** This is a hypothetical data file that contains demographic and purchasing information for a direct marketing company. dmdata2.sav contains information for a subset of contacts that received a test mailing, and dmdata3.sav contains information on the remaining contacts who did not receive the test mailing.

---

• **dietstudy.sav.** This hypothetical data file contains the results of a study of the “Stillman diet” 7. Each case corresponds to a separate subject and records his or her pre- and post-diet weights in pounds and triglyceride levels in mg/100 ml.

• **dvdplayer.sav.** This is a hypothetical data file that concerns the development of a new DVD player. Using a prototype, the marketing team has collected focus group data. Each case corresponds to a separate surveyed user and records some demographic information about them and their responses to questions about the prototype.

• **german_credit.sav.** This data file is taken from the "German credit" dataset in the Repository of Machine Learning Databases 8 at the University of California, Irvine.

• **grocery_1month.sav.** This hypothetical data file is the grocery_coupons.sav data file with the weekly purchases "rolled-up" so that each case corresponds to a separate customer. Some of the variables that changed weekly disappear as a result, and the amount spent recorded is now the sum of the amounts spent during the four weeks of the study.

• **grocery_coupons.sav.** This is a hypothetical data file that contains survey data collected by a grocery store chain interested in the purchasing habits of their customers. Each customer is followed for four weeks, and each case corresponds to a separate customer-week and records information about where and how the customer shops, including how much was spent on groceries during that week.

• **guttman.sav.** Bell 9 presented a table to illustrate possible social groups. Guttman 10 used a portion of this table, in which five variables describing such things as social interaction, feelings of belonging to a group, physical proximity of members, and formality of the relationship were crossed with seven theoretical social groups, including crowds (for example, people at a football game), audiences (for example, people at a theater or classroom lecture), public (for example, newspaper or television audiences), mobs (like a crowd but with much more intense interaction), primary groups (intimate), secondary groups (voluntary), and the modern community (loose confederation resulting from close physical proximity and a need for specialized services).

• **health_funding.sav.** This is a hypothetical data file that contains data on health care funding (amount per 100 population), disease rates (rate per 10,000 population), and visits to health care providers (rate per 10,000 population). Each case represents a different city.

• **hivassay.sav.** This is a hypothetical data file that concerns the efforts of a pharmaceutical lab to develop a rapid assay for detecting HIV infection. The results of the assay are eight deepening shades of red, with deeper shades indicating greater likelihood of infection. A laboratory trial was conducted on 2,000 blood samples, half of which were infected with HIV and half of which were clean.

• **hourlywagedata.sav.** This is a hypothetical data file that concerns the hourly wages of nurses from office and hospital positions and with varying levels of experience.

• **insurance_claims.sav.** This is a hypothetical data file that concerns an insurance company that wants to build a model for flagging suspicious, potentially fraudulent claims. Each case represents a separate claim.

• **insure.sav.** This is a hypothetical data file that concerns an insurance company that is studying the risk factors that indicate whether a client will have to make a claim on a 10-year term life insurance contract. Each case in the data file represents a pair of contracts, one of which recorded a claim and the other didn’t, matched on age and gender.

• **judges.sav.** This is a hypothetical data file that concerns the scores given by trained judges (plus one enthusiast) to 300 gymnastics performances. Each row represents a separate performance; the judges viewed the same performances.

---


- **kinship_dat.sav.** Rosenberg and Kim \(^{11}\) set out to analyze 15 kinship terms (aunt, brother, cousin, daughter, father, granddaughter, grandfather, grandmother, grandson, mother, nephew, niece, sister, son, uncle). They asked four groups of college students (two female, two male) to sort these terms on the basis of similarities. Two groups (one female, one male) were asked to sort twice, with the second sorting based on a different criterion from the first sort. Thus, a total of six “sources” were obtained. Each source corresponds to a 15 x 15 proximity matrix, whose cells are equal to the number of people in a source minus the number of times the objects were partitioned together in that source.

- **kinship_ini.sav.** This data file contains an initial configuration for a three-dimensional solution for *kinship_dat.sav.*

- **kinship_var.sav.** This data file contains independent variables *gender,* *generation,* and *degree* (of separation) that can be used to interpret the dimensions of a solution for *kinship_dat.sav.* Specifically, they can be used to restrict the space of the solution to a linear combination of these variables.

- **marketvalues.sav.** This data file concerns home sales in a new housing development in Algonquin, Ill., during the years from 1999–2000. These sales are a matter of public record.


- **ozone.sav.** The data include 330 observations on six meteorological variables for predicting ozone concentration from the remaining variables. Previous researchers \(^{12, 13}\), among others found nonlinearities among these variables, which hinder standard regression approaches.

- **pain_medication.sav.** This hypothetical data file contains the results of a clinical trial for anti-inflammatory medication for treating chronic arthritic pain. Of particular interest is the time it takes for the drug to take effect and how it compares to an existing medication.

- **patient_loms.sav.** This hypothetical data file contains the treatment records of patients who were admitted to the hospital for suspected myocardial infarction (MI, or “heart attack”). Each case corresponds to a separate patient and records many variables related to their hospital stay.

- **patlos_sample.sav.** This hypothetical data file contains the treatment records of a sample of patients who received thrombolytics during treatment for myocardial infarction (MI, or “heart attack”). Each case corresponds to a separate patient and records many variables related to their hospital stay.

- **poll_cs.sav.** This is a hypothetical data file that concerns pollsters’ efforts to determine the level of public support for a bill before the legislature. The cases correspond to registered voters. Each case records the county, township, and neighborhood in which the voter lives.

- **poll_cs_sample.sav.** This hypothetical data file contains a sample of the voters listed in *poll_cs.sav.* The sample was taken according to the design specified in the *poll.csplan* plan file, and this data file records the inclusion probabilities and sample weights. Note, however, that because the sampling plan makes use of a probability-proportional-to-size (PPS) method, there is also a file containing the joint selection probabilities (*poll_jointprob.sav*). The additional variables corresponding to voter demographics and their opinion on the proposed bill were collected and added the data file after the sample as taken.

- **property_assess.sav.** This is a hypothetical data file that concerns a county assessor’s efforts to keep property value assessments up to date on limited resources. The cases correspond to properties sold in the county in the past year. Each case in the data file records the township in which the property lies, the assessor who last visited the property, the time since that assessment, the valuation made at that time, and the sale value of the property.

---


• **property_assess_cs.sav.** This is a hypothetical data file that concerns a state assessor's efforts to keep property value assessments up to date on limited resources. The cases correspond to properties in the state. Each case in the data file records the county, township, and neighborhood in which the property lies, the time since the last assessment, and the valuation made at that time.

• **property_assess_cs_sample.sav.** This hypothetical data file contains a sample of the properties listed in property_assess_cs.sav. The sample was taken according to the design specified in the property_assess.csplan plan file, and this data file records the inclusion probabilities and sample weights. The additional variable Current value was collected and added to the data file after the sample was taken.

• **recidivism.sav.** This is a hypothetical data file that concerns a government law enforcement agency’s efforts to understand recidivism rates in their area of jurisdiction. Each case corresponds to a previous offender and records their demographic information, some details of their first crime, and then the time until their second arrest, if it occurred within two years of the first arrest.

• **recidivism_cs_sample.sav.** This is a hypothetical data file that concerns a government law enforcement agency’s efforts to understand recidivism rates in their area of jurisdiction. Each case corresponds to a previous offender, released from their first arrest during the month of June, 2003, and records their demographic information, some details of their first crime, and the data of their second arrest, if it occurred by the end of June, 2006. Offenders were selected from sampled departments according to the sampling plan specified in recidivism_cs.csplan; because it makes use of a probability-proportional-to-size (PPS) method, there is also a file containing the joint selection probabilities (recidivism_cs_jointprob.sav).

• **rfm_transactions.sav.** A hypothetical data file containing purchase transaction data, including date of purchase, item(s) purchased, and monetary amount of each transaction.

• **salesperformance.sav.** This is a hypothetical data file that concerns the evaluation of two new sales training courses. Sixty employees, divided into three groups, all receive standard training. In addition, group 2 gets technical training; group 3, a hands-on tutorial. Each employee was tested at the end of the training course and their score recorded. Each case in the data file represents a separate trainee and records the group to which they were assigned and the score they received on the exam.

• **satisf.sav.** This is a hypothetical data file that concerns a satisfaction survey conducted by a retail company at 4 store locations. 582 customers were surveyed in all, and each case represents the responses from a single customer.

• **screws.sav.** This data file contains information on the characteristics of screws, bolts, nuts, and tacks.

• **shampoo_ph.sav.** This is a hypothetical data file that concerns the quality control at a factory for hair products. At regular time intervals, six separate output batches are measured and their pH recorded. The target range is 4.5–5.5.

• **ships.sav.** A dataset presented and analyzed elsewhere that concerns damage to cargo ships caused by waves. The incident counts can be modeled as occurring at a Poisson rate given the ship type, construction period, and service period. The aggregate months of service for each cell of the table formed by the cross-classification of factors provides values for the exposure to risk.

• **site.sav.** This is a hypothetical data file that concerns a company's efforts to choose new sites for their expanding business. They have hired two consultants to separately evaluate the sites, who, in addition to an extended report, summarized each site as a "good," "fair," or "poor" prospect.

• **smokers.sav.** This data file is abstracted from the 1998 National Household Survey of Drug Abuse and is a probability sample of American households. [http://dx.doi.org/10.3886/ICPSR02934](http://dx.doi.org/10.3886/ICPSR02934) Thus, the first step in an analysis of this data file should be to weight the data to reflect population trends.

• **stocks.sav** This hypothetical data file contains stocks prices and volume for one year.

• **stroke_clean.sav.** This hypothetical data file contains the state of a medical database after it has been cleaned using procedures in the Data Preparation option.

---


• **stroke_invalid.sav.** This hypothetical data file contains the initial state of a medical database and contains several data entry errors.

• **stroke_survival.** This hypothetical data file concerns survival times for patients exiting a rehabilitation program post-ischemic stroke face a number of challenges. Post-stroke, the occurrence of myocardial infarction, ischemic stroke, or hemorrhagic stroke is noted and the time of the event recorded. The sample is left-truncated because it only includes patients who survived through the end of the rehabilitation program administered post-stroke.

• **stroke_valid.sav.** This hypothetical data file contains the state of a medical database after the values have been checked using the Validate Data procedure. It still contains potentially anomalous cases.

• **survey_sample.sav.** This data file contains survey data, including demographic data and various attitude measures. It is based on a subset of variables from the 1998 NORC General Social Survey, although some data values have been modified and additional fictitious variables have been added for demonstration purposes.

• **telco.sav.** This is a hypothetical data file that concerns a telecommunications company’s efforts to reduce churn in their customer base. Each case corresponds to a separate customer and records various demographic and service usage information.

• **telco_extra.sav.** This data file is similar to the telco.sav data file, but the "tenure" and log-transformed customer spending variables have been removed and replaced by standardized log-transformed customer spending variables.

• **telco_missing.sav.** This data file is a subset of the telco.sav data file, but some of the demographic data values have been replaced with missing values.

• **testmarket.sav.** This hypothetical data file concerns a fast food chain’s plans to add a new item to its menu. There are three possible campaigns for promoting the new product, so the new item is introduced at locations in several randomly selected markets. A different promotion is used at each location, and the weekly sales of the new item are recorded for the first four weeks. Each case corresponds to a separate location-week.

• **testmarket_1month.sav.** This hypothetical data file is the testmarket.sav data file with the weekly sales "rolled-up" so that each case corresponds to a separate location. Some of the variables that changed weekly disappear as a result, and the sales recorded is now the sum of the sales during the four weeks of the study.

• **tree_car.sav.** This is a hypothetical data file containing demographic and vehicle purchase price data.

• **tree_credit.sav.** This is a hypothetical data file containing demographic and bank loan history data.

• **tree_missing_data.sav** This is a hypothetical data file containing demographic and bank loan history data with a large number of missing values.

• **tree_score_car.sav.** This is a hypothetical data file containing demographic and vehicle purchase price data.

• **tree_textdata.sav.** A simple data file with only two variables intended primarily to show the default state of variables prior to assignment of measurement level and value labels.

• **tv-survey.sav.** This is a hypothetical data file that concerns a survey conducted by a TV studio that is considering whether to extend the run of a successful program. 906 respondents were asked whether they would watch the program under various conditions. Each row represents a separate respondent; each column is a separate condition.

• **ulcer_recurrence.sav.** This file contains partial information from a study designed to compare the efficacy of two therapies for preventing the recurrence of ulcers. It provides a good example of interval-censored data and has been presented and analyzed elsewhere. ¹⁶

• **ulcer_recurrence_recoded.sav.** This file reorganizes the information in ulcer_recurrence.sav to allow you model the event probability for each interval of the study rather than simply the end-of-study event probability. It has been presented and analyzed elsewhere ¹⁷.

---


• **verd1985.sav.** This data file concerns a survey. The responses of 15 subjects to 8 variables were recorded. The variables of interest are divided into three sets. Set 1 includes age and marital, set 2 includes pet and news, and set 3 includes music and live. Pet is scaled as multiple nominal and age is scaled as ordinal; all of the other variables are scaled as single nominal.

• **virus.sav.** This is a hypothetical data file that concerns the efforts of an Internet service provider (ISP) to determine the effects of a virus on its networks. They have tracked the (approximate) percentage of infected e-mail traffic on its networks over time, from the moment of discovery until the threat was contained.

• **wheeze_steubenville.sav.** This is a subset from a longitudinal study of the health effects of air pollution on children. The data contain repeated binary measures of the wheezing status for children from Steubenville, Ohio, at ages 7, 8, 9 and 10 years, along with a fixed recording of whether or not the mother was a smoker during the first year of the study.

• **workprog.sav.** This is a hypothetical data file that concerns a government works program that tries to place disadvantaged people into better jobs. A sample of potential program participants were followed, some of whom were randomly selected for enrollment in the program, while others were not. Each case represents a separate program participant.

• **worldsales.sav** This hypothetical data file contains sales revenue by continent and product.

---


Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user’s responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.
Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group  
ATTN: Licensing  
200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs.

© Copyright IBM Corp. _enter the year or years_. All rights reserved.
Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at ‘Copyright and trademark information’ at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Index

C
captions  Custom Tables 15
collapsing categories  Custom Tables 37
column means statistics  custom tables 64
column proportions statistics  custom tables 66
column width  
controlling in custom tables  14, 87
comperimeter tables  14, 45
computed categories  
Custom Tables 13, 39
display formats 14
from subtotals 40
hiding categories in expression 40
controlling number of decimals displayed  22
corner labels  
Custom Tables 15
count vs. valid N 56
crosstabulation  
Custom Tables 21
custom tables  
split file processing 4
Custom Tables
 captions 15
categorical variables 1
changing labels for summary statistics 20
changing measurement level 1
changing summary statistics dimension 9
collapsing categories 37
column width 14
compact view 28
comperimeter tables 14, 45
computed categories 10, 13, 39
controlling number of decimals displayed 5
corner labels 15
crosstabulation 21
custom totals 9
display formats 5
empty cells 14
excluding categories 10, 23
excluding categories
Custom Tables 10, 23
excluded categories  Custom Tables 10, 23

G
group totals 34
grouped summaries
scale variables 58

H
hiding statistics labels in custom tables 19
labels
changing label text for summary statistics 86
layer variables
Custom Tables 29, 30
nesting layer variables 30
printing layered tables 30
stacking layer variables 30

M
maximum  
Custom Tables 8
mean 55
Custom Tables 8
mean-frequency tables 9, 51
measurement level changing in custom tables 1
median 55
Custom Tables 8
minimum  
Custom Tables 8
missing values 56, 81
effect on percentage calculations 82
including in custom tables 82
mode
Custom Tables 8
multiple response sets 73
duplicate responses in multiple category sets 14
percentages 8
significance testing 71, 78

N
nesting variables  
Custom Tables 26, 28
scale variables 59
omitting categories
  Custom Tables 23

percentages
  in custom tables 6, 7, 20, 21
  missing values 82
  multiple response sets 8
post-computed categories
  Custom Tables 13, 39
printing tables with layers 30

range
  Custom Tables 8
reordering categories
  Custom Tables 10

sample files
  location 89
scale variables
  grouped summaries 58
  multiple summary statistics 55
  nesting 59
  stacking 55
  summaries grouped by row and column categorical variables 58
  summary statistics 55
significance tests
  Custom Tables 16
  multiple response sets 78
sorting categories
  Custom Tables 23
split file processing
  custom tables 4
stacking variables
  Custom Tables 25
  different summary statistics for different variables 57
  multiple summary statistics source variables 51
  scale variables 55
  stacking layer variables 30
standard deviation
  Custom Tables 8
statistics
  custom total summary statistics 51
  stacked tables 51
  summary statistics 49
subgroup totals 34
subtotals 36
  Custom Tables 10, 33
  hiding subtotaled categories 37
sum
  Custom Tables 8
summary statistics (continued)
  display format 85
  source dimension 49
  source variable 49
  stacked tables 51
summary statistics source variable
  scale variables 59
system-missing values 81

T
  table of frequencies
    Custom Tables 14, 45
  tables
    Custom Tables 1
  test statistics
    Custom Tables 16, 61
time
  including current time in custom tables 15
titles
  Custom Tables 15
total N 82
totals
  Custom Tables 10, 20, 33
  display position 34
  excluded categories 33
  group totals 34
  layers 35
  marginal totals for custom tables 22
  nested tables 34
user-missing values 81

V
  valid N 56, 82
  Custom Tables 8
values
  displaying category labels and values 52
  values and value labels 52
variable labels
  suppressing display in custom tables 5
variance
  Custom Tables 8