

## 3. Looking at Data: Relationships

This chapter describes methods for examining the relationship between two quantitative variables or two categorical variables.



### 3.1 Relationships between quantitative variables

In analyzing the relationship between two quantitative variables, the first step is to construct a **scatterplot**, which allows you to examine the nature and strength of the relationship. The **correlation coefficient** is a numerical measure of the strength of the linear relationship. If the relationship is linear, then the **least squares regression line** can be used as a mathematical model to describe the relationship between the two variables.

**Example 1-1: Student Measurements** (page 5), continued: use this data set to answer the following questions. Is there a relationship between height and armspan? Is it linear? How strong is it? Can you develop a model to predict armspan from height?

#### Constructing a scatterplot

To construct a scatterplot between *height* and *armspan*, follow these steps. In this example, *height* will be on the  $x$  axis and *armspan* on the  $y$  axis, because we are interested in predicting *armspan* from *height*. We would reverse them if we were interested in predicting *height* from *armspan*.

1. Click **Graphs, Legacy Dialogs**, and then **Scatter/Dot**, and click **Define**. The “Simple Scatterplot” window in Figure 3-1 appears.
2. Click *armspan*, then click  to move *armspan* into the “Y Axis” box.
3. Click *height*, then click  to move *height* into the “X Axis” box.
4. Click **OK**.

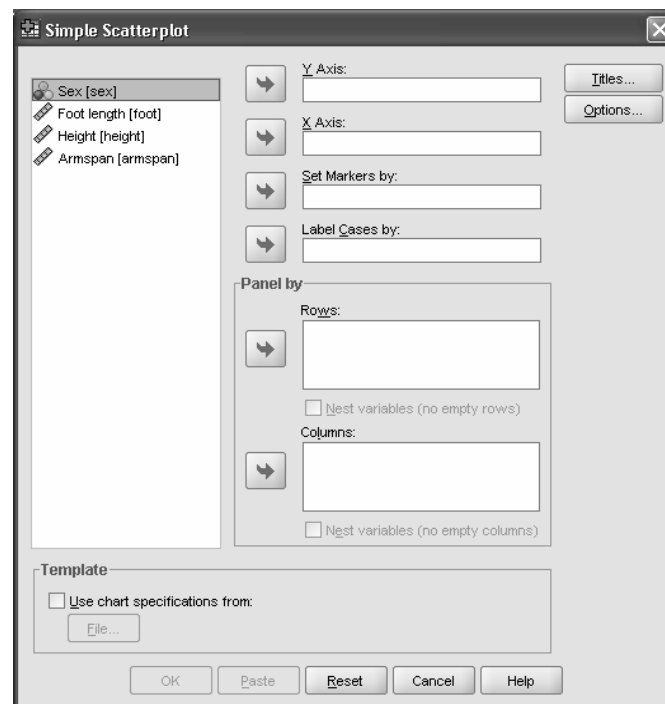


Figure 3-1

Figure 3-2 is the resulting SPSS output. The scatterplot reveals that there is a moderately strong, approximately linear, relationship between *height* and *armspan*. There is also an outlier on the upper end.

To add a categorical variable to a scatterplot (not included in the text) see **Adding a categorical variable to a scatterplot** on page 50 of this manual. To add the least squares line to the scatterplot, see **Editing a scatterplot** on page 45.

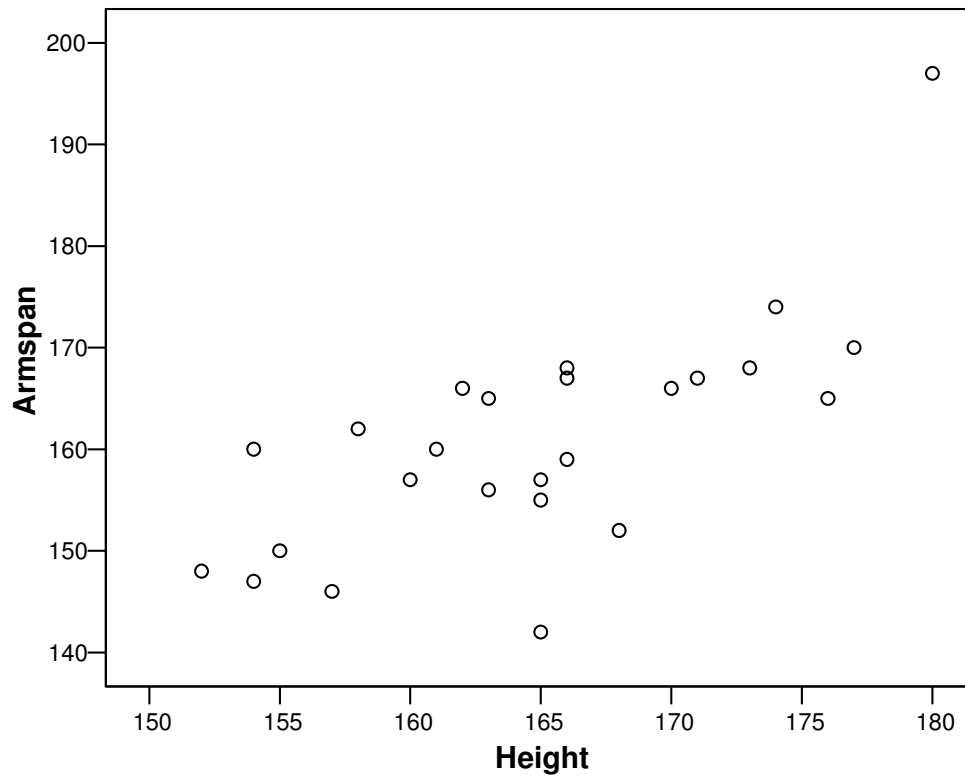


Figure 3-2

### Computing the correlation coefficient

To obtain the value of the correlation coefficient  $r$  between *height* and *armspan*, follow these steps.

1. Click **Analyze**, click **Correlate**, click **Bivariate**. The window in Figure 3-3 appears.
2. Click *height*, then click  to move *height* into the “Variables” box.
3. Click *armspan*, then click  to move *armspan* into the “Variables” box.
4. Make sure that “Pearson” has a check in front of it in the “Correlation Coefficients” box and that the other two boxes are not checked.
5. Click **OK**. The output is shown in Table 3-1.

In Table 3-1, the correlation,  $r = .718$ , appears as the entry in the Height row and Armspan column.

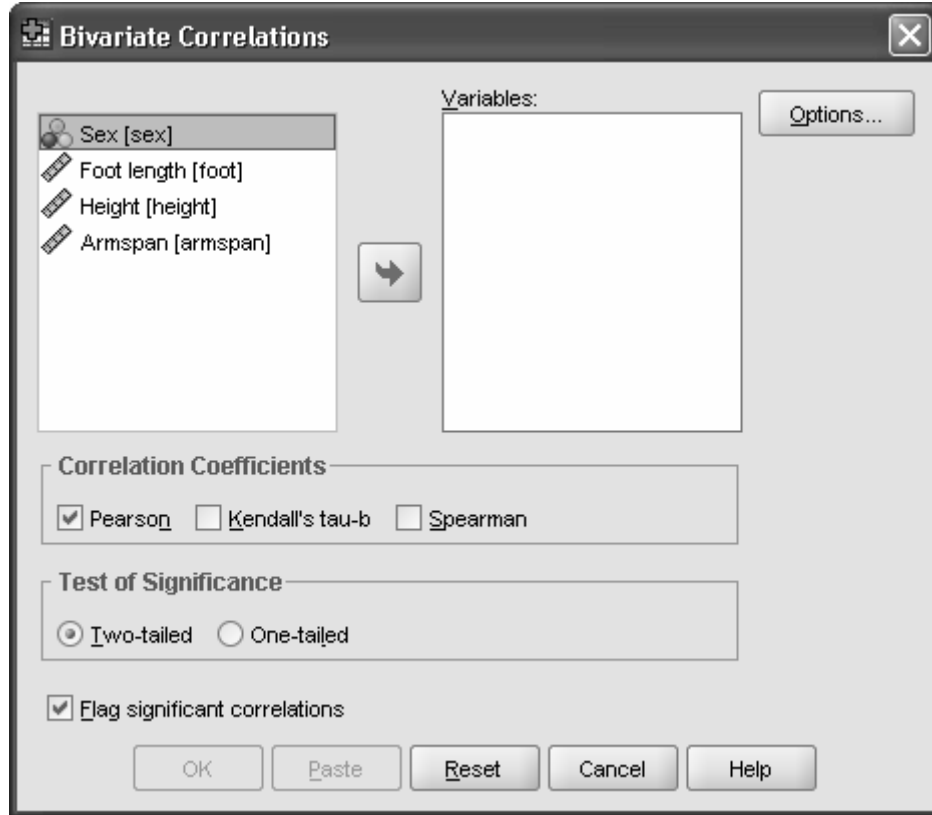


Figure 3-3

## Correlations



|         |                     | Height | Armspan |
|---------|---------------------|--------|---------|
| Height  | Pearson Correlation | 1      | .718**  |
|         | Sig. (2-tailed)     | .      | .000    |
|         | N                   | 25     | 25      |
| Armspan | Pearson Correlation | .718** | 1       |
|         | Sig. (2-tailed)     | .000   | .       |
|         | N                   | 25     | 25      |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Table 3-1

### Computing the equation of the least squares line

To obtain the equation of the least squares regression line for predicting *armspan* from *height*, follow these steps.

1. Click **Analyze**, click **Regression**, and then click **Linear**. The “Linear regression” window in Figure 3-4 appears.
2. Click *armspan*, then click  to move *armspan* into the “Dependent” box.
3. Click *height*, then click  to move *height* into the “Independent(s)” box.
4. Click **OK**.

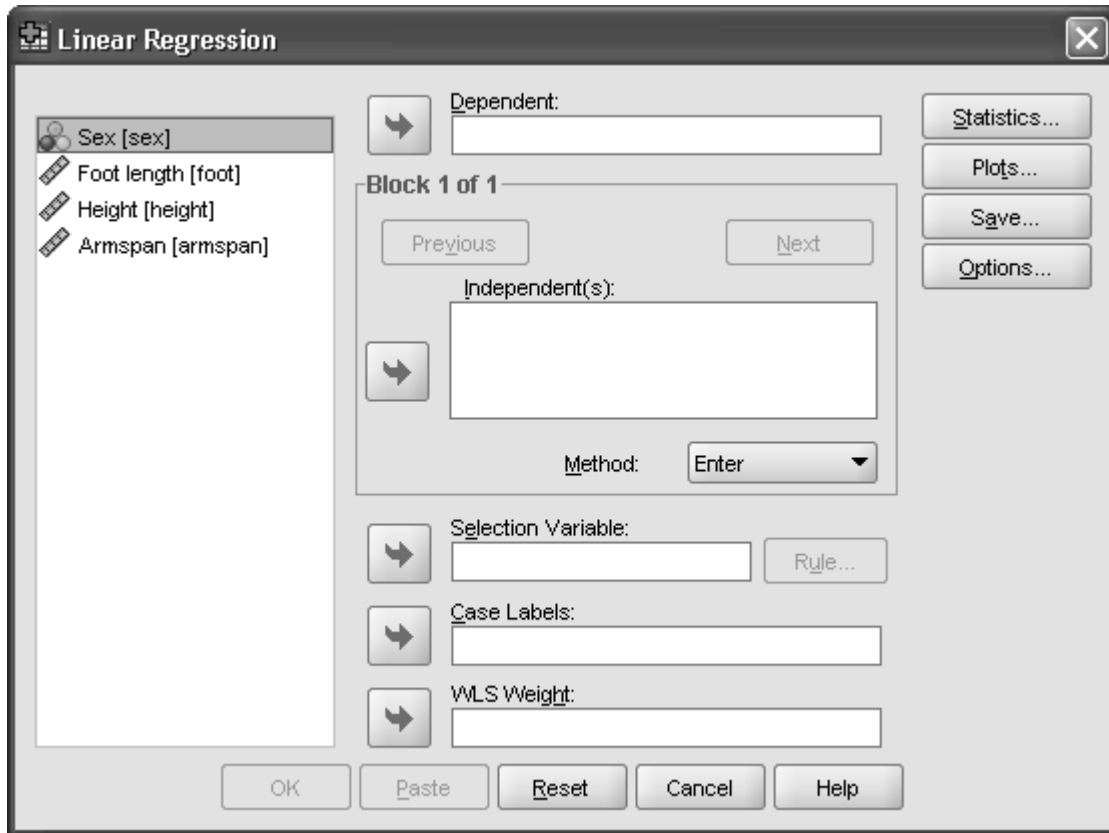


Figure 3-4

The SPSS output consists of four tables: **Variables Entered/Removed**, **Model Summary**, **ANOVA**, and **Coefficients**. The only two we will use in this class are the **Model Summary** and **Coefficients** tables, which are displayed as Table 3-2 and Table 3-3.

Model Summary

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | .718 <sup>a</sup> | .515     | .494              | 8.019                      |

a. Predictors: (Constant), Height

Table 3-2

Coefficients<sup>a</sup>

| Model |            | Unstandardized Coefficients |            | Standardized Coefficients | t     | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
|       |            | B                           | Std. Error | Beta                      |       |      |
| 1     | (Constant) | -14.378                     | 35.499     |                           | -.405 | .689 |
|       | Height     | 1.064                       | .215       | .718                      | 4.944 | .000 |

a. Dependent Variable: Armspan

Table 3-3

The **Model Summary** table gives  $R^2 = .515$  (page 209 in the text); this means that 51.5% of the variability in *armspan* is explained by the linear relationship with *height* (48.5% of the variability is therefore not explained by the linear relationship with height). From the **Coefficients** table, we obtain the equation of the least squares line as



$$\mathit{armspan} = -14.378 + 1.064 \mathit{height}$$

The variable *armspan* in the above equation is actually the *predicted armspan* because the equation does not give the actual arm span for an individual student, but only the predicted value using the least squares regression line. To add the least squares line to the scatterplot, see Adding the least squares line (SPSS versions 12.0 - 16.0) on page 45 or Adding the least squares line (SPSS version 11.5 or earlier) on page 47.

### Computing residuals

A **residual** is the difference between an observed value of the response variable and the value predicted by the regression line:  $\mathit{residual} = \mathit{actual} - \mathit{predicted}$ . A **residual plot** is a scatterplot with the residuals on the  $y$ -axis and the explanatory variable on the  $x$ -axis. This plot is used to detect patterns in the residuals.

The residuals can be generated when the least squares line is computed; see **Computing the equation of the least squares line**, page 41. The whole sequence of commands will be repeated here.

1. Click **Analyze**, click **Regression**, and then click **Linear**. The “Linear regression” window in Figure 3-4 (page 42) appears.
2. Click *armspan*, then click  to move *armspan* into the “Dependent” box.
3. Click *height*, then click  to move *height* into the “Independent(s)” box.
4. Click **Save**. The “Linear Regression: Save” box in Figure 3-5 appears.
5. Click **Unstandardized** in the “Residuals” at the top of the window. A check should appear in the box in front of **Unstandardized**. (Predicted values can also be saved by clicking **Unstandardized** in the “Predicted Values” area on the left side.)
6. Click **Continue**, then click **OK**.

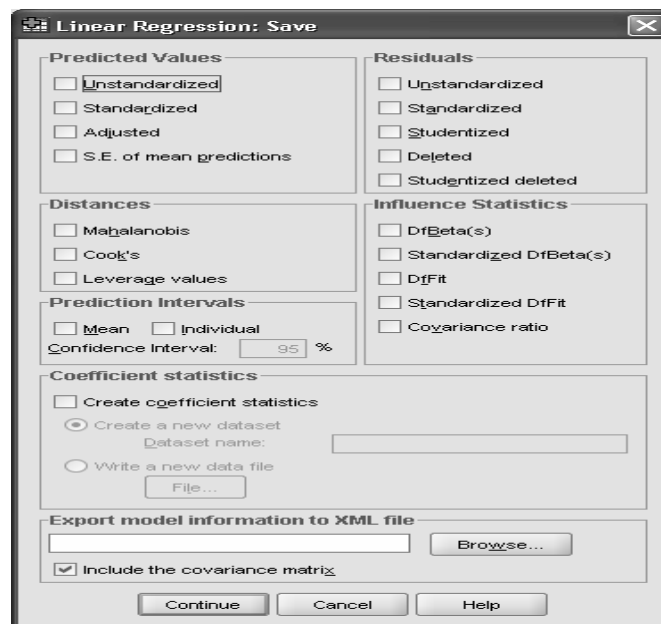
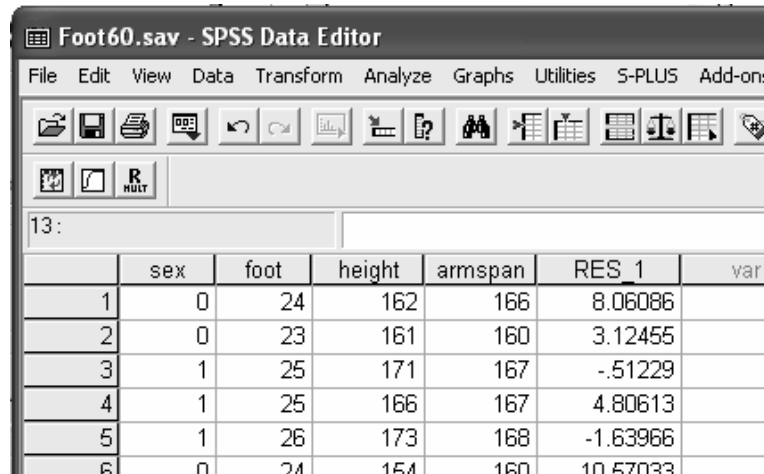


Figure 3-5

The residuals for this data set have been generated, saved, and added to the Data Editor, as shown in Figure 3-6. The default name of the residuals is *RES\_1*. You can change the name as you would change the name of any other variable; see Section 1.4. You can also compute descriptive statistics for the residuals and create a residual plot (Activity 28-3 on page 582 of the text) as described in the next section.



|   | sex | foot | height | armspan | RES_1    | var |
|---|-----|------|--------|---------|----------|-----|
| 1 | 0   | 24   | 162    | 166     | 8.06086  |     |
| 2 | 0   | 23   | 161    | 160     | 3.12455  |     |
| 3 | 1   | 25   | 171    | 167     | -.51229  |     |
| 4 | 1   | 25   | 166    | 167     | 4.80613  |     |
| 5 | 1   | 26   | 173    | 168     | -1.63966 |     |
| 6 | 0   | 24   | 154    | 160     | 10.57033 |     |

Figure 3-6

### Creating a residual plot

A residual plot is a scatterplot of the residuals from a linear regression model plotted against the explanatory variable. The residuals for the regression must first be saved, as described in the previous section. Then follow the steps for creating a scatterplot (**Constructing a scatterplot**, page 39), using the variable *Unstandardized residual (RES\_1)* for the y-axis and the explanatory variable for the x-axis as shown in the scatterplot window in Figure 3-7. A horizontal line at 0 can be added to the residual plot with the Chart Editor; see the next section: **Editing a scatterplot: Adding a reference line** (SPSS versions 12.0 - 16.0), page 46.

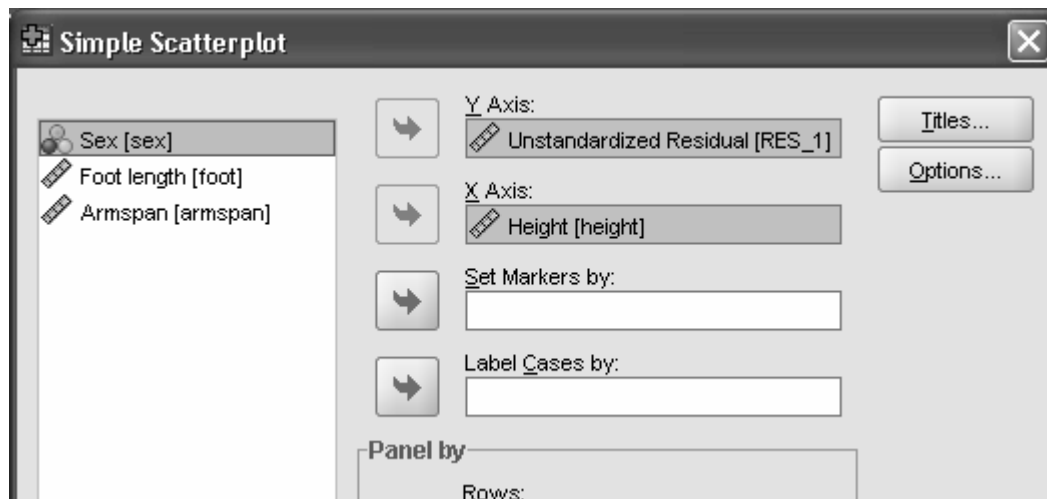



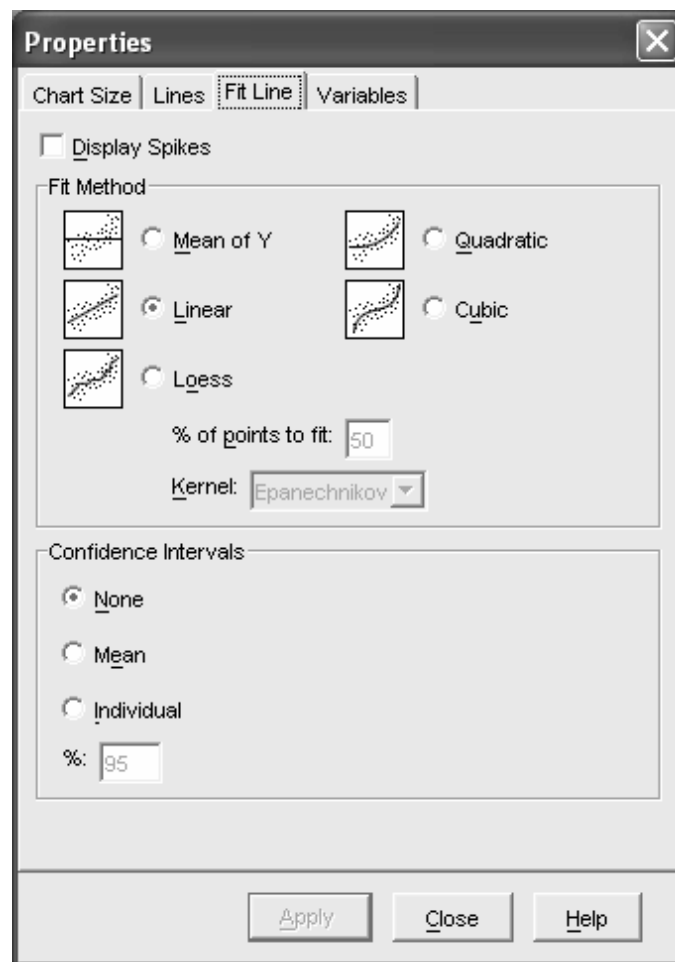
Figure 3-7

### Editing a scatterplot

Many aspects of a scatterplot can be changed in the SPSS Chart Editor. First, generate the scatterplot as described previously (page 39). Double-click on the scatterplot in the output window to start the Chart Editor. Then follow the directions below depending on what you want to do, and depending on which version of SPSS you are using.

### Adding the least squares line (SPSS versions 12.0 - 16.0)

1. In the Chart Editor, place the arrow exactly on one of the points in the scatterplot and click. The points will become highlighted.
2. Click the **Add fit line at Total** icon  in the toolbar. The SPSS window in Figure 3-8 appears.
3. The **Linear** fit method should be selected by default. If not, select it and click **Apply**.
4. Click **Close**.



**Figure 3-8**

### **Adding a reference line (SPSS versions 12.0 - 16.0)**

Horizontal and vertical reference lines can be added to a scatterplot. For example, in a residual plot, it is often desired to add a horizontal line at 0. To do this, create a residual plot (see **Creating a residual plot**, page 44), double-click on the residual plot to start Chart Editor, and follow these steps:

1. Click **Options**, and then click **Y Axis Reference Line**. The SPSS window in Figure 3-9 appears.
2. The **Reference Line** window should be selected with a default **Y Axis Position** of 0. If not, change the **Y Axis Position** to 0, since this is where we want to add a line to the residual plot, and click **Apply**.
3. Click **Close**.





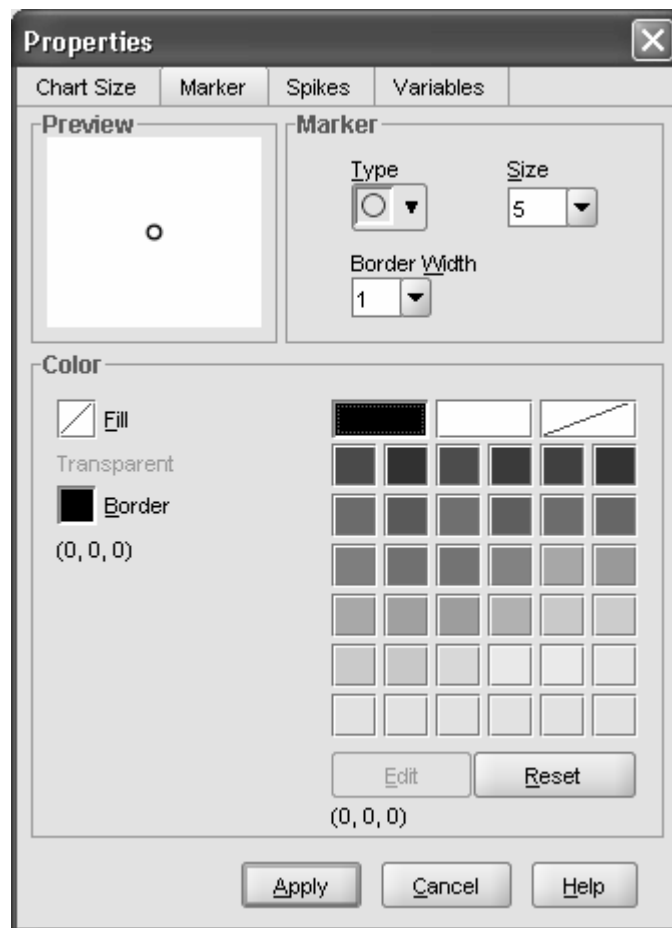
**Figure 3-9**

### **Changing the size, color, and shape of the points (SPSS versions 12.0 - 16.0)**

To change the size, color, or shape characteristics of the plotted points, follow these steps once you are in the Chart Editor (double-click on the scatterplot in the output window to start Chart Editor).

1. In the Chart Editor, place the arrow exactly on one of the points in the scatterplot and click. The points will become highlighted.

2. Click **Edit** and click **Properties**, or click the **Show Properties Window** icon  in the toolbar. The SPSS window in Figure 3-10 appears.
3. Change the color of the plotting symbol by selecting the color desired in the color grid. The **Border** window on the left side will be filled in with the selected color.
4. Change the plotting symbol by clicking  under **Type** and selecting the symbol desired.
5. Change the size of the plotting symbol by choosing from the list of sizes in the upper right corner of the **Marker** window. Generally, smaller symbols are better for larger numbers of points.
6. When all changes are completed, click **Apply**.
7. Click **Close**.



**Figure 3-10**

### **Changing axis labels and positioning (SPSS versions 12.0 - 16.0)**

To change axis labels and positioning, follow the instructions under **Editing histograms**, page 26.

### **Adding the least squares line (SPSS version 11.5 or earlier)**

1. In the Chart Editor, click **Chart**, then click **Options**. The “Scatterplot Options” window appears as in Figure 3-11.
2. Click **Total** in the “Fit Line” box so that a check appears in the box in front of **Total**.

3. Click **Fit Options**. The window in Figure 3-12 appears.
4. The “Fit Method” defaults to “Linear Regression” which is what is desired.
5. Click **Continue**.
6. Click **OK**.

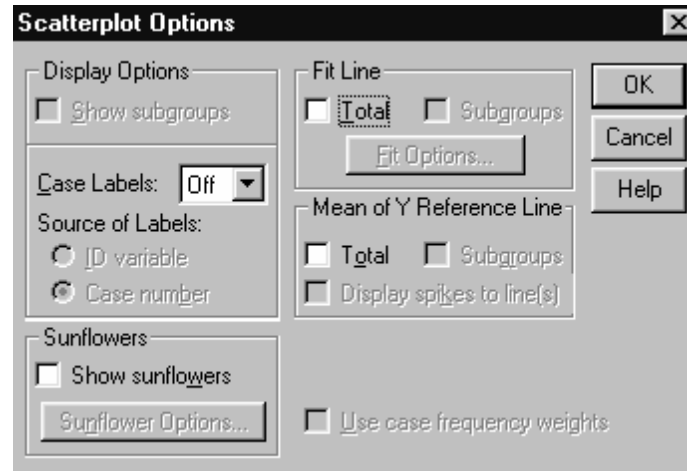


Figure 3-11

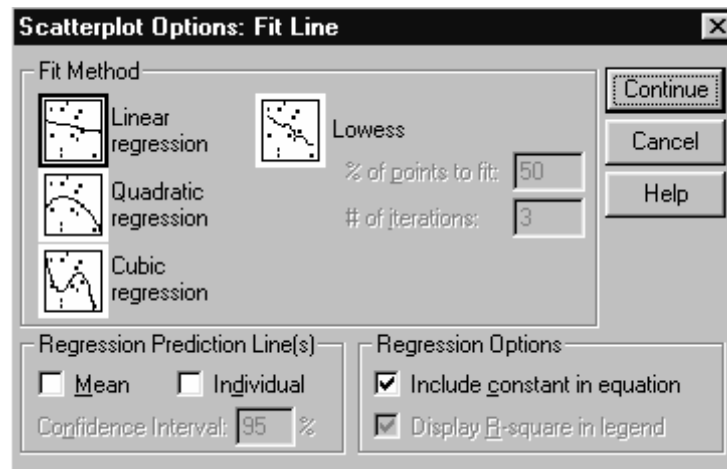


Figure 3-12

### **Adding a reference line (SPSS version 11.5 or earlier)**

Horizontal and vertical reference lines can be added to a scatterplot. For example, in a residual plot, it is often desired to add a horizontal line at 0. To do this, create the residual plot (see **Creating a residual plot**, page 44), double-click on the residual plot to start Chart Editor, and follow these steps:

1. Click **Chart** and click **Reference Line**. The SPSS window in Figure 3-13 appears.
2. Click “Y scale” so that it is selected. Then click **OK**. The SPSS window in Figure 3-14 appears.
3. The default position of 0 is where we want to add a line, so click **Add**. A “0” should appear in the large box. Other reference lines can be added by entering a position in the “Position of Line(s)” box and clicking **Add** after each one.
4. Click **OK**. A horizontal line at 0 should appear on the scatterplot. Its color can be changed by clicking on it and editing its color.

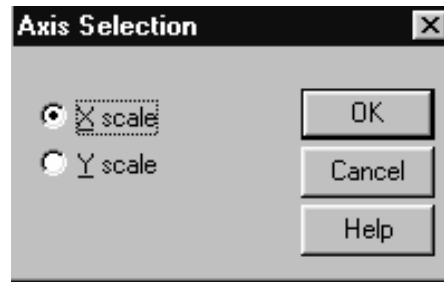


Figure 3-13

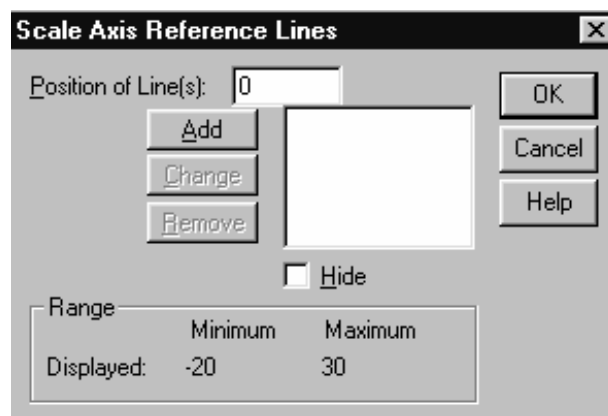




Figure 3-14

### **Changing the size, color, and shape of the points (SPSS version 11.5 or earlier)**

The default size and color of the points in a scatterplot can make them difficult to see when the graph is printed. To change the characteristics of the plotted points, follow these steps once you are in the Chart Editor (double-click on the scatterplot in the output window to start Chart Editor).

1. In Chart Editor, place the arrow exactly on one of the points in the scatterplot and click. The points will become highlighted (sometimes it will appear that only some of the points are highlighted, but that's OK).
2. Click  in the toolbar. The SPSS window in Figure 3-15 appears.
3. Select black from the color palette, then click **Apply**.
4. Click **Close**.
5. To also change the plotting size and symbol, click  in the toolbar. The SPSS window in Figure 3-16 appears.
6. Change the plotting symbol by clicking on one of the symbols. Open circles (top row) and filled circles (3<sup>rd</sup> row) are preferred; open circles are preferred to filled circles if any of the points overlap.
7. Change the size of the plotting symbol by choosing from the options in the lower right. **Small** is a good size for open circles unless you have many points, in which case you might want to use **Tiny**. **Tiny** is a good size for filled circles.
8. Click **Apply All**.
9. Click **Close**.

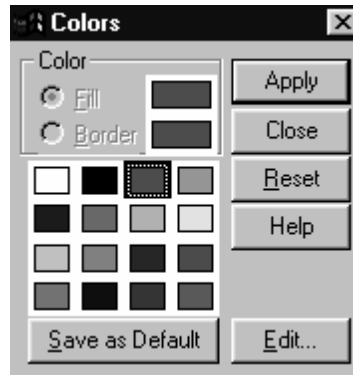


Figure 3-15

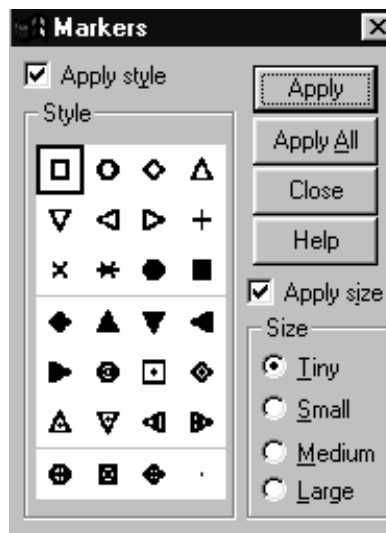





Figure 3-16

### Changing axis labels and positioning (SPSS versions 11.5 or earlier)

To change axis labels and positioning, click **Chart**, then click **Axis**. Follow the instructions under **Editing histograms**, page 26.


### Adding a categorical variable to a scatterplot

A scatterplot can be modified to display the points for each category of a categorical variable with a different symbol or color. For example, in **Example 1-1: Student Measurements** on page 5, suppose it is desired to examine the relationship between *height* and *armspan* for males and females separately. To create a scatterplot with males and females plotted in different colors, follow these steps.



1. Click **Graph**, **Legacy Dialogs**, and **Scatter/Dot**, and click **Define**. The “Simple Scatterplot” window in Figure 3-1 (page 39) appears.
2. Click *armspan*, then click  to move *armspan* into the “Y Axis” box.
3. Click *height*, then click  to move *height* into the “X Axis” box.
4. Click *sex*, then click  to move *sex* into the “Set Markers by” box.
5. Click **OK**.

The resulting scatterplot will have males and females plotted in different colors with a legend on the side to indicate which color represents which gender. For printing on a black-and-white printer, we want to display males and females with different plotting symbols, rather than different colors. To do this, first create the scatterplot as above, then follow these steps.

**For SPSS versions 12.0 - 16.0:**

1. Double-click on the scatterplot to start the Chart Editor.
2. Click on the plotting symbol for males in the legend to the right of the scatterplot. Only the male points will be highlighted.
3. Change the color of the male points to black and the plotting symbol to a “+” by clicking  on the toolbar and following the instructions in Editing a scatterplot: Changing the size, color, and shape of the points (SPSS versions 12.0 - 16.0)(page 46).
4. Repeat steps 2-3 for the female points except, of course, choose a different plotting symbol. The plotting symbols for the two groups should be as distinctive as possible. For example, a small square and a small circle are not easily distinguished. However, an open circle and a “+” are easily distinguished.

**For SPSS versions 11.5 or earlier:**


1. Double-click on the scatterplot to start the Chart Editor.
2. Click on one of the points which is for a male (or click on the plotting symbol for males in the legend to the right of the scatterplot). Only the male points will be highlighted.
3. Change the color of the male points to black by clicking  on the toolbar and following the instructions in Editing a scatterplot: Changing the size, color, and shape of the points (SPSS version 11.5 or earlier) (page 49).
4. Keep the male points highlighted and change the plotting symbol for these points by clicking  on the toolbar and following the instructions in Editing a scatterplot: Changing the size, color, and shape of the points (SPSS version 11.5 or earlier) (page 49).
5. Repeat steps 2-4 for the female points except, of course, choose a different plotting symbol. The plotting symbols for the two groups should be as distinctive as possible. For example, a small square and a small circle are not easily distinguished. However, an open circle and a “+” are easily distinguished.

**Adding separate regression lines**

Separate least squares regression lines for each group can also be added to the plot. Follow these steps.



**For SPSS version 15.0:**

1. Double-click on the scatterplot to start Chart Editor.
2. Plot the two least squares regression lines for males and females by clicking the **Add fit line at**


**Subgroups** icon  on the toolbar, and make sure that the option **Linear** is selected. Click **Close**.

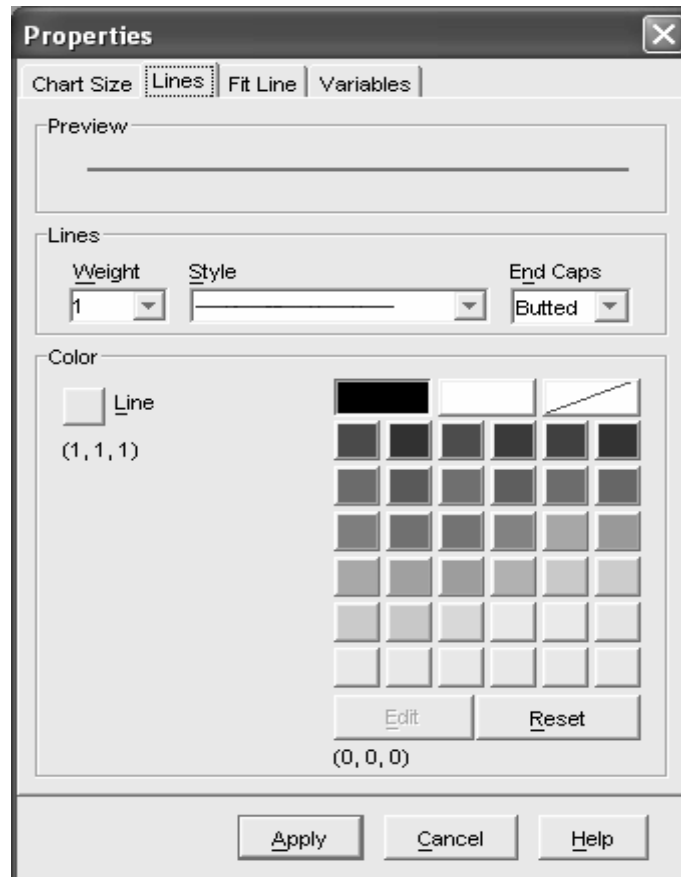
**For SPSS versions 12.0 - 14.0:**

1. Double-click on the scatterplot to start Chart Editor.
2. Click on the plotting symbol for males in the legend to the right of the scatterplot. Only the male points will be highlighted.

3. Plot the least squares regression line for the males by clicking the **Add fit line at Subgroups** icon  on the toolbar, and make sure that the option **Linear** is selected. Click **Close**.
4. Click on the plotting symbol for females in the legend to the right of the scatterplot. Only the female points will be highlighted.
5. Plot the least squares regression line for the females by clicking the **Add fit line at Subgroups** icon  on the toolbar, and make sure that the option **Linear** is selected. Click **Close**.

Separate regression lines in different colors for each group will appear. To make the lines have different appearance (such as dashed and solid) rather than different colors for printing on a black-and-white printer, follow these steps.

1. Place the tip of the arrow exactly on one of the regression lines in the legend and click. The selected line will be highlighted. If it is not highlighted, then you did not successfully select it.
2. Click the **Show Properties Window** icon  on the toolbar and make sure the **Lines** tab is selected. The window in Figure 3-17 appears.
3. Change the color of the line to black by clicking on the black rectangle in the color grid.
4. Change the line style to a dashed line by clicking on the dashed line in the “Style” box.
5. Click **Apply**, then click **Close**.
6. Repeat step 3 for the other group. Its line style can be left as the default solid line.
7. Click **File**, then **Close** to exit the Chart Editor.





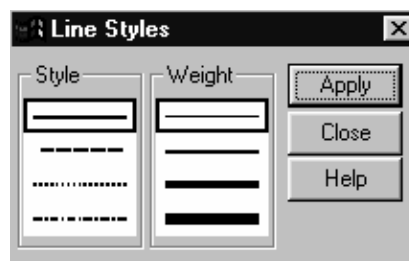
**Figure 3-17**

**For SPSS versions 11.5 or earlier:**

1. Double-click on the scatterplot to start the Chart Editor.
2. Click **Chart**, then click **Options**. The window in Figure 3-11 appears.
3. Click **Subgroups** in the “Fit Line” box so that a check appears in the box in front of **Subgroups**. Be sure that there is not a check in front of **Total** (click on it if there is to unselect it).
4. Click **Fit Options**. The window in Figure 3-12 appears.
5. The “Fit Method” defaults to “Linear Regression” which is what is desired.
6. Click **Continue**.
7. Click **OK**.

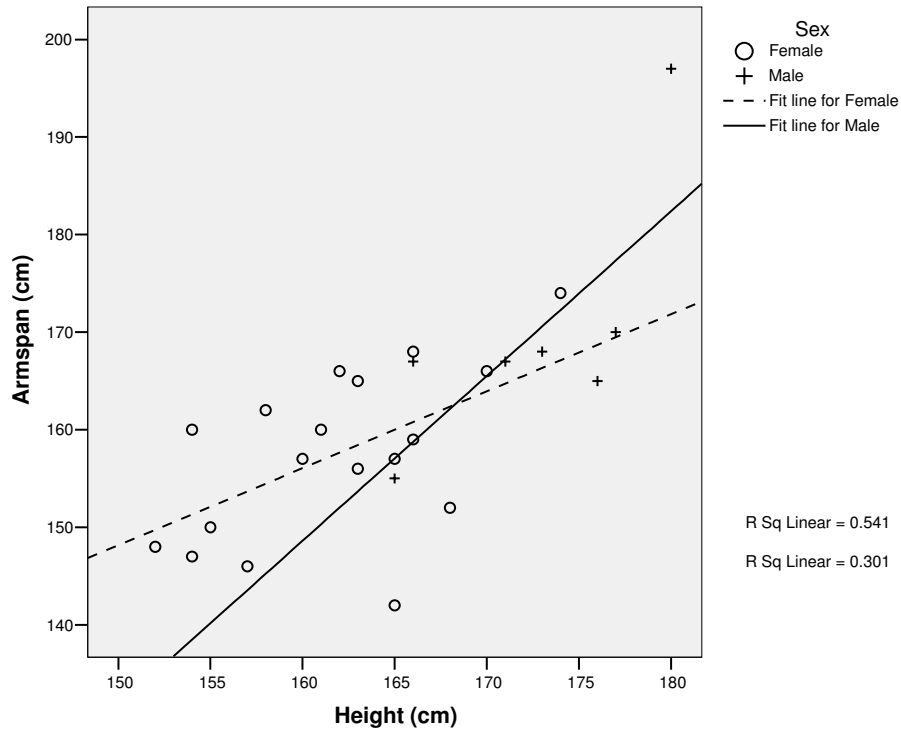
Separate regression lines in different colors for each group will appear. They will also be added to the legend. To make the lines have a different appearance (such as dashed and solid) rather than different colors for printing on a black-and-white printer, follow these steps.

1. Place the tip of the arrow exactly on one of the regression lines (either in the plot or in the legend) and click. Three small black squares will appear along the regression line if it has been selected. If the squares do not appear then you did not successfully select it.
2. Click  on the toolbar and follow the instructions in *Editing a scatterplot: Changing the size, color, and shape of the points (SPSS version 11.5 or earlier) (page 49)*, to change the color of the line to black.
3. Click  on the toolbar. The window in Figure 3-18 appears.
4. Change the line style to a dashed line by clicking on the dashed line in the “Style” box.
5. Click **Apply**, then click **Close**.
6. Repeat steps 1-2 for the other group. You do not need to change its line style from the default solid line.
7. Click **File**, then **Close** to exit the Chart Editor.



**Figure 3-18**

The final scatterplot, which also includes editing of the axis labels, appears in Figure 3-19.



**Figure 3-19**

### 3.2 Relationships between categorical variables

The relationship between two categorical variables can be examined numerically through a two-way table, also called a contingency table or a cross-tabulation.



**Example 3-1: Student Travels.** Twenty-three students in a statistics class were asked to record the numbers of states and nations visited, whether the student had visited Walt Disney World and whether the student had visited Europe. We wish to examine the relationship between the two categorical variables: Walt Disney World (*wdw*) and Europe (*europe*). These variables were coded as 0 for No and 1 for Yes; value labels (page 9) were added to indicate this. In addition, variable labels (page 9) were added (“Walt Disney World” for *wdw* and “Europe” for *europe*). A subset of this data set is displayed as entered into SPSS in Figure 3-20.

|   | states | nations | wdw | europe | v8 |
|---|--------|---------|-----|--------|----|
| 1 | 11     | 3       | 0   | 0      |    |
| 2 | 11     | 2       | 0   | 0      |    |
| 3 | 35     | 2       | 1   | 0      |    |
| 4 | 33     | 13      | 1   | 1      |    |
| 5 | 26     | 4       | 1   | 0      |    |
| 6 | 28     | 1       | 0   | 0      |    |
| 7 | 35     | 2       | 1   | 0      |    |

Figure 3-20

### Creating a two-way table

To create a two-way table of *wdw* and *europe*, follow these steps.

1. Click **Analyze**, then click **Descriptive Statistics**, then click **Crosstabs**. The window “Crosstabs” in Figure 3-21 appears.
2. Click on *wdw*, then click  to move *wdw* to the “Row(s)” box.
3. Click on *europe*, then click  to move *europe* to the “Column(s)” box.
4. Click the “Cells” button at the right of the “Crosstabs” window. The “Crosstabs: Cell Display” window in Figure 3-22 appears.
5. Click **Row** in the “Percentages” box so that a check appears in the box in front of **Row**.
6. Click **Continue**.
7. Click **OK**.

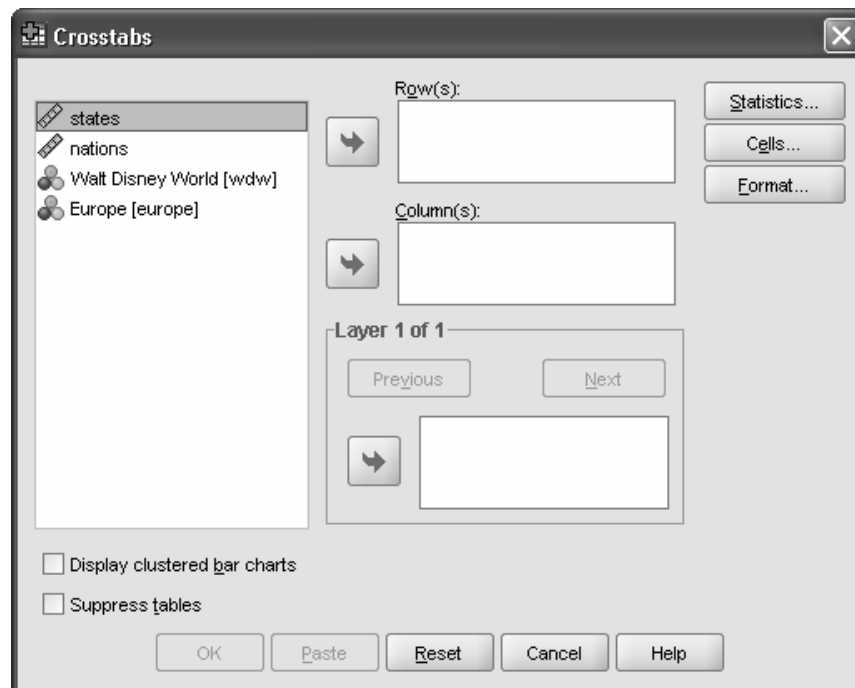


Figure 3-21

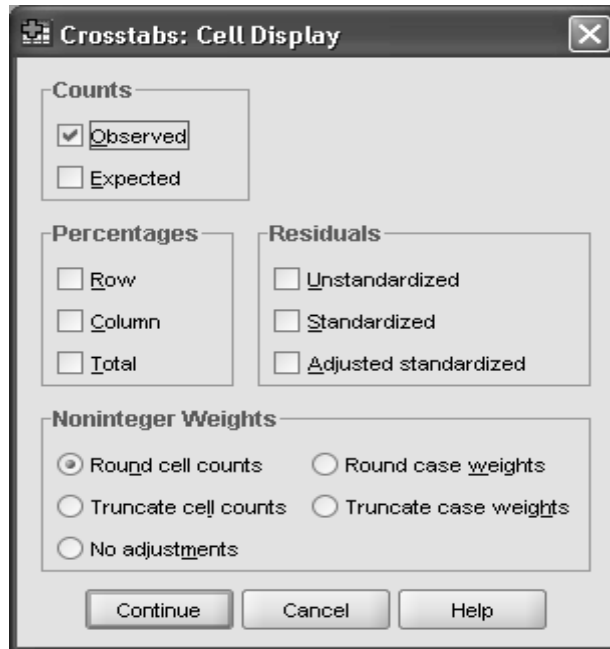


Figure 3-22

The output is displayed in Table 3-4. Note that the variable labels and value labels appear in the row and column headings. The crosstabulation shows the cell counts, the marginal totals, and row percentages. It indicates that 23.1% of those students who had not been to Disney World had been to Europe, while 20.0% of those who had been to Disney World had been to Europe. Since these two percentages are almost equal, this indicates that there is little association between having been to Disney World and having been to Europe.

Walt Disney World \* Europe Crosstabulation

|                   |                            |                            | Europe |        | Total  |
|-------------------|----------------------------|----------------------------|--------|--------|--------|
|                   |                            |                            | No     | Yes    |        |
| Walt Disney World | No                         | Count                      | 10     | 3      | 13     |
|                   |                            | % within Walt Disney World | 76.9%  | 23.1%  | 100.0% |
|                   | Yes                        | Count                      | 8      | 2      | 10     |
|                   |                            | % within Walt Disney World | 80.0%  | 20.0%  | 100.0% |
| Total             | Count                      | 18                         | 5      | 23     |        |
|                   | % within Walt Disney World | 78.3%                      | 21.7%  | 100.0% |        |

Table 3-4

### Entering count data

Sometimes, it is desired to directly enter the counts from a two-way table into SPSS.

**Example 3-2: Age and Political Interest.** Data were collected on the *Age* (grouped into 3 categories) and *Political Interest* (3 categories) for 1265 American adults. The table is given below as **Table 3-5**. The goal is to enter this data set into SPSS and reproduce this table.

|              | Age group |       |       |       |
|--------------|-----------|-------|-------|-------|
| Interest     | 18-35     | 36-55 | 56-94 | Total |
| not much     | 146       | 146   | 89    | 381   |
| somewhat     | 192       | 260   | 154   | 606   |
| very much    | 47        | 125   | 106   | 278   |
| <b>Total</b> | 385       | 531   | 349   | 1265  |

Table 3-5

In the previous example (**Example 3-1: Student Travels**, page 54), the counts for the *wdw* by *europa* crosstabulation were generated by SPSS from an SPSS data set with one row per observation. There are a total of 1265 cases in Table 3-5, so it would be necessary to enter 1265 rows of data to create the raw data set. Fortunately, in SPSS, there is an easier way to enter a data set such as this. The key is to define three variables: *age*, *interest*, and a variable called *count* which contains the count for each cell. Then use the *weighting* option in SPSS to analyze the data. Detailed instructions for entering and analyzing the data follow.

Figure 3-23 shows how the data were entered in the SPSS Data Editor. See Section 1.4, page 5, for instructions on entering data. The variables *age* and *interest* should be set as nominal variables and *count* as a scale variable in SPSS. Note that the row and column totals should not be entered. Note also that the variables *age* and *interest* were entered with numerical values representing the different categories. **Value labels** were then used to attach descriptive labels to the different categories. The “Define Labels” window for *age* is shown in Figure 3-24 and the one for *interest* is shown in Figure 3-25; see Section 1.9 (page 9) for instructions on entering value labels.

|   | age | interest | count |
|---|-----|----------|-------|
| 1 | 1   | 1        | 146   |
| 2 | 1   | 2        | 192   |
| 3 | 1   | 3        | 47    |
| 4 | 2   | 1        | 146   |
| 5 | 2   | 2        | 260   |
| 6 | 2   | 3        | 125   |
| 7 | 3   | 1        | 89    |
| 8 | 3   | 2        | 154   |
| 9 | 3   | 3        | 106   |

Figure 3-23

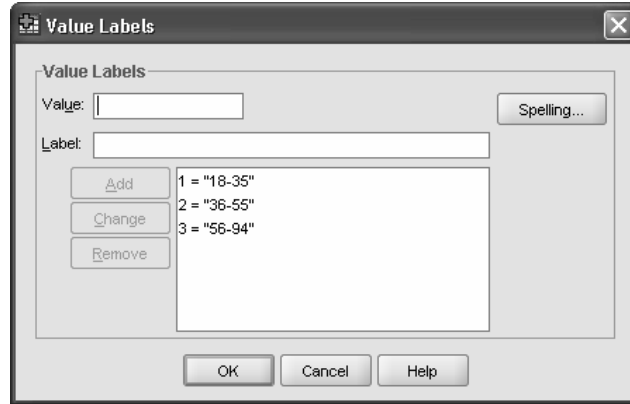


Figure 3-24

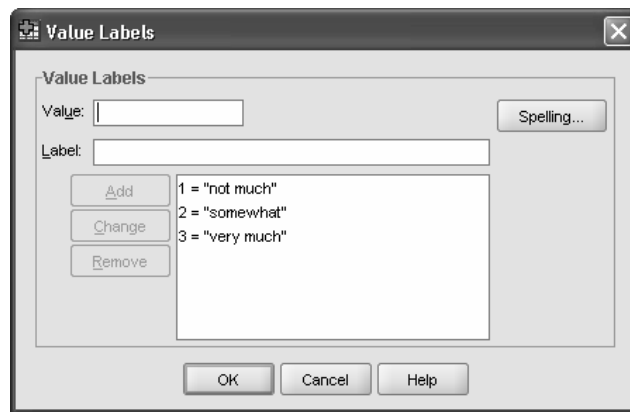



Figure 3-25

Before doing a crosstabulation, it is necessary to activate the weighting option. Follow these steps to activate the weighting option.

1. Click **Data**, and click **Weight Cases**. The “Weight Cases” window in Figure 3-26 appears.
2. Click **Weight Cases by**.
3. Click *count*, then click  to move *count* into the “Frequency Variable” box.
4. Click **OK**.

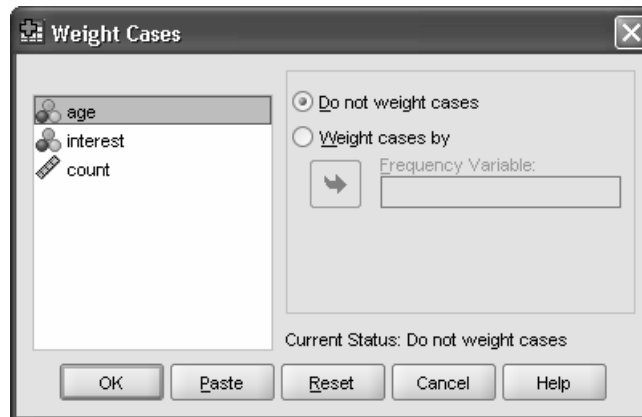


Figure 3-26

Now that the weighting option has been activated, you can have SPSS generate a crosstabulation by following the steps in **Creating a two-way table** on page 55; the “Crosstabs” window should be set up as shown in Figure 3-27. The result, with row percents, is shown in Table 3-6. It shows a tendency for older people to have more interest in politics. We can also produce a segmented or clustered bar graph of this table (see next section).

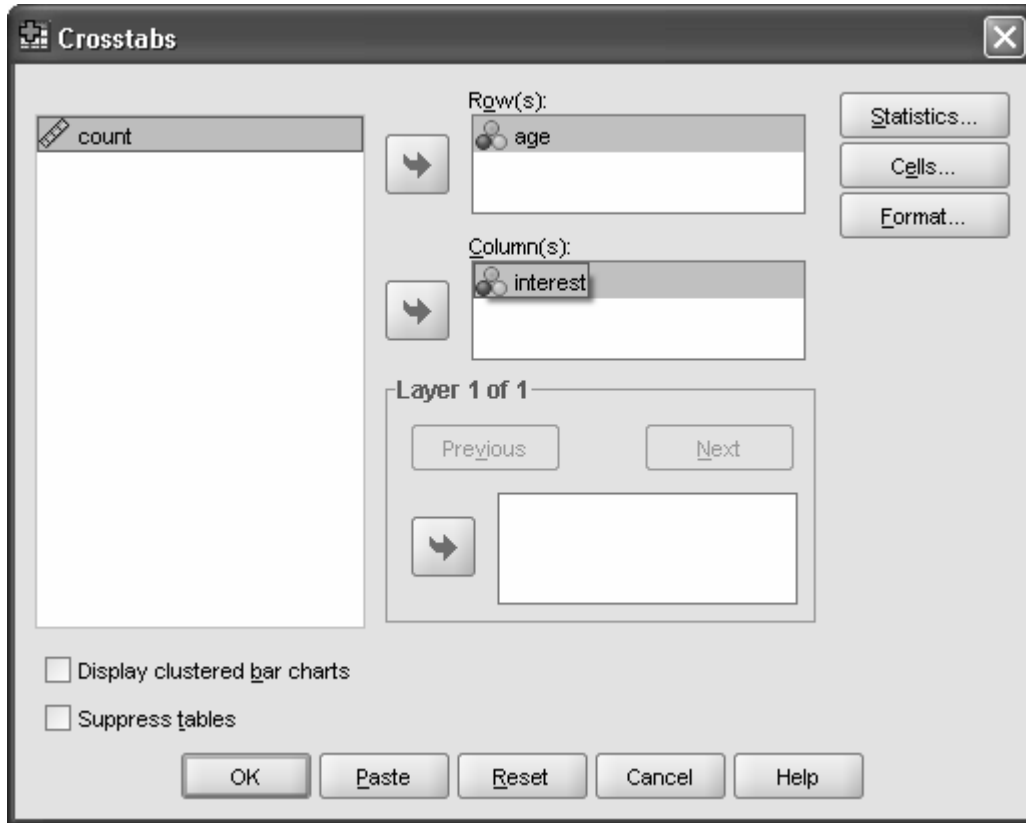


Figure 3-27

AGE \* INTEREST Crosstabulation

|       |              |              | INTEREST |          |           | Total  |
|-------|--------------|--------------|----------|----------|-----------|--------|
|       |              |              | not much | somewhat | very much |        |
| AGE   | 18-35        | Count        | 146      | 192      | 47        | 385    |
|       |              | % within AGE | 37.9%    | 49.9%    | 12.2%     | 100.0% |
|       | 36-55        | Count        | 146      | 260      | 125       | 531    |
|       |              | % within AGE | 27.5%    | 49.0%    | 23.5%     | 100.0% |
|       | 56-94        | Count        | 89       | 154      | 106       | 349    |
|       |              | % within AGE | 25.5%    | 44.1%    | 30.4%     | 100.0% |
| Total | Count        | 381          | 606      | 278      | 1265      |        |
|       | % within AGE | 30.1%        | 47.9%    | 22.0%    | 100.0%    |        |



Table 3-6

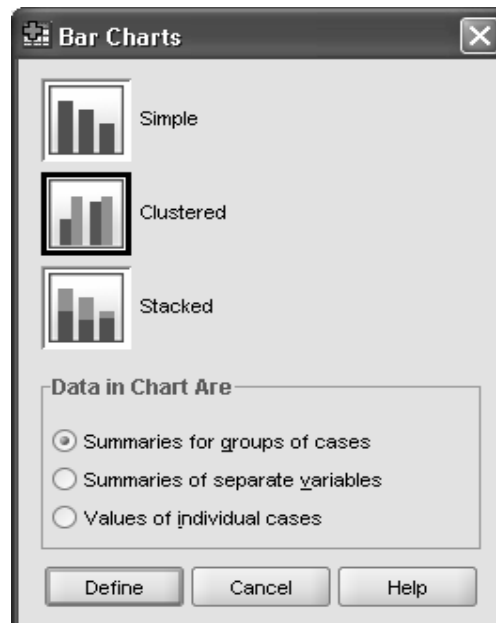
### Clustered and segmented bar graphs

Clustered and segmented bar graphs (or charts) are two different graphical ways to represent the conditional distribution of one categorical variable (call it the response variable) given each level of another categorical variable (call it the explanatory variable). Conditional distributions are discussed on pages 95-96 of the text. Clustered bar graphs are not discussed in the text but segmented bar graphs are discussed on pages 96-97 of the text. Both are a compact way of displaying the conditional distributions represented in a two-way table. In a clustered bar graph, the bar graphs representing the conditional distributions of the response variable given the levels of the explanatory variable are clustered together in one graph (see Figure 3-32, page 63). In a segmented bar graph (called a *stacked* bar chart in SPSS), there is a bar for each level of the explanatory variable. Each bar has the same height (100%), and is divided into segments which represent the distribution of the response variable for that level of the explanatory variable (see the example in Figure 3-37, page 66). Clustered and segmented bar graphs both display the same thing in different ways. Some people prefer the segmented and some people prefer the clustered; it's a matter of individual preference.

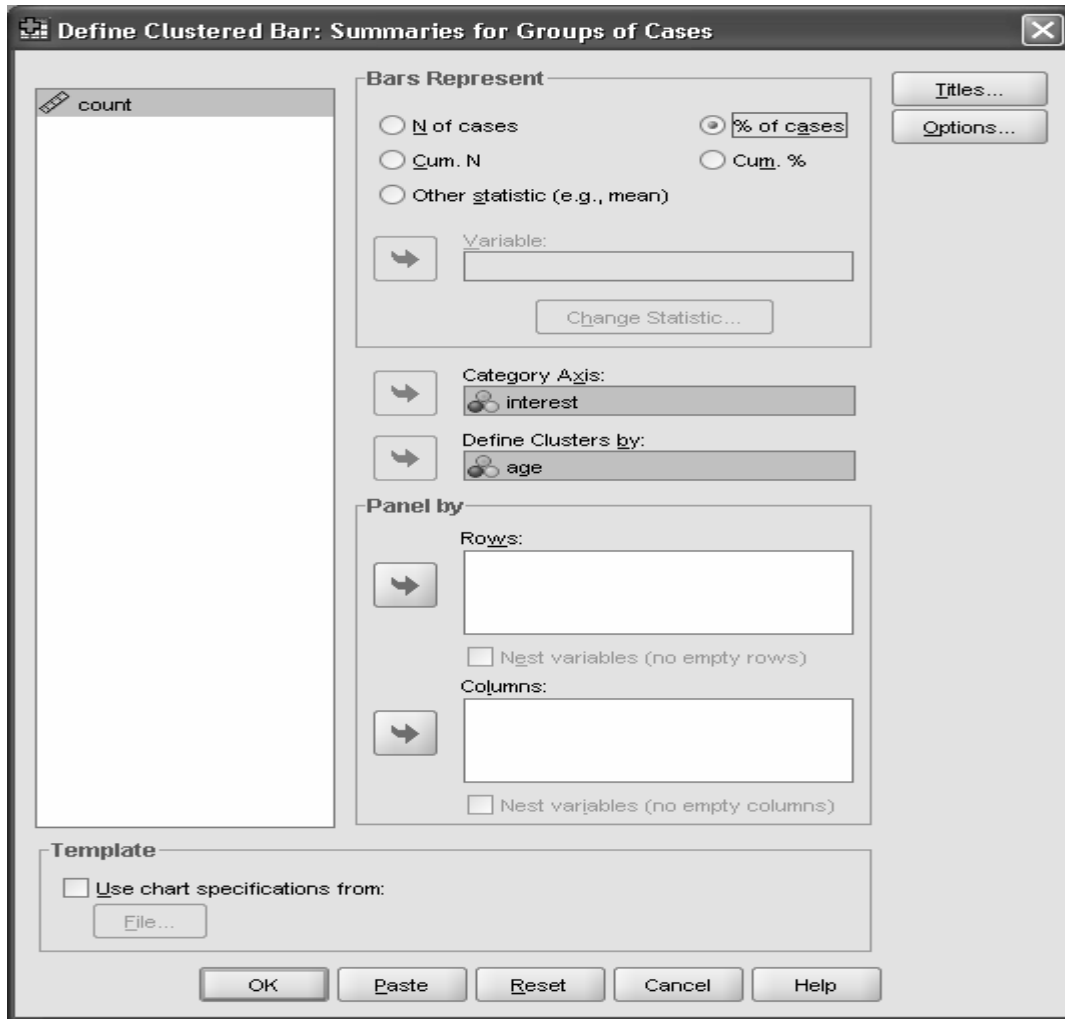
To illustrate the construction of these graphs in SPSS, suppose that in **Example 3-2: Age and Political Interest** (page 56), we wish to create clustered and segmented bar graphs of the conditional distribution of *interest* given the levels of *age*. Be sure that the **Weight cases** option has been activated since this data set has been entered with counts (see **Entering count data**, page 56).

To create a clustered bar graph of *interest* given *age*, follow these steps.

1. Click **Graphs, Legacy Dialogs**, then click **Bar**. The window in Figure 3-28 appears.
2. Click on "Clustered". "Clustered" should be highlighted as it is in Figure 3-28.
3. Click **Define**. Figure 3-29 shows how this window will appear after the next three steps.
4. Click *interest*, then click  to move *interest* into the "Category Axis" box.
5. Click *age*, then click  to move *age* into the "Define Clusters by" box.
6. Click on "% of cases" under "Bars Represent" so that this option is selected. The window will now appear as in Figure 3-29.
7. Click **OK**. A clustered bar graph appears in the output window.




**Figure 3-28**

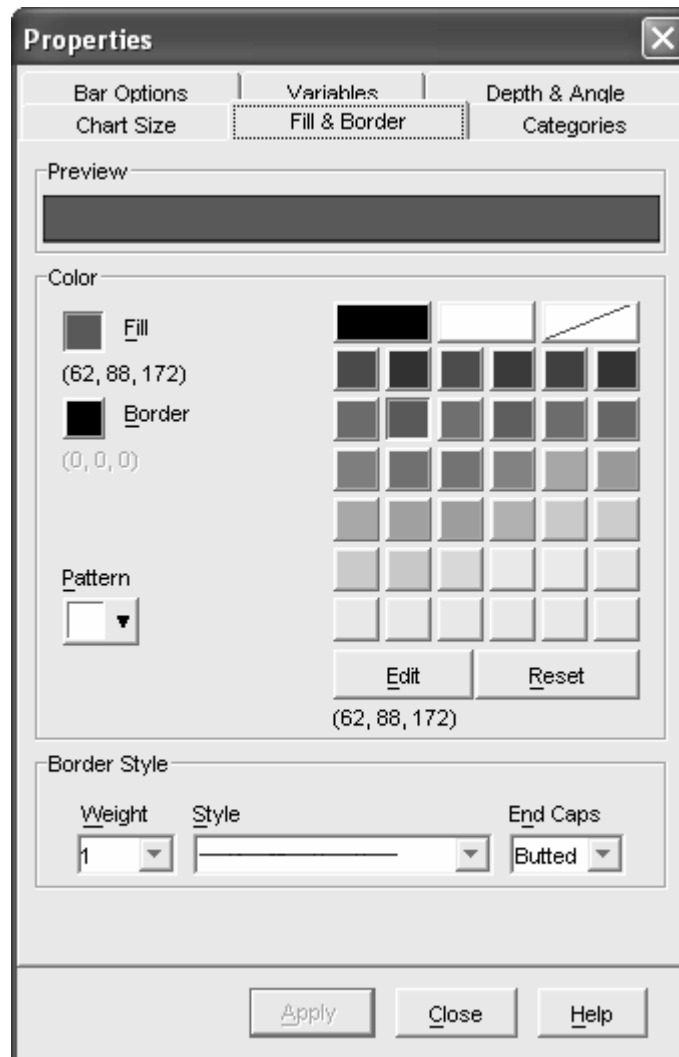


**Figure 3-29**

The clustered bar graph has bars of three different colors to represent the three different conditional distributions. To change the colors to black, white and gray for printing on a black-and-white printer, continue with the steps below.


**For SPSS versions 12.0 - 16.0:**

8. Double-click on the bar graph to start the Chart Editor.
9. Click on the blue square in the age legend to the right of the bar chart. All the blue bars will be highlighted.
10. Click the **Show Properties Window** icon  on the toolbar. Make sure the **Fill & Border** header at the top of the window is selected. The window in Figure 3-30 appears.
11. Click on the black rectangle in the color grid to change the Fill color to black, then click **Apply**. The blue bars will change to black.
12. Click **Close**.
13. Repeat steps 9-12 for the green bars to change them to white.
14. Repeat steps 9-12 for the light brown bars to change them to gray.
15. Click **File**, then click **Close** to exit the Chart Editor.



**Figure 3-30**

**For SPSS versions 11.5 or earlier** (where the bar colors are red, green, & blue):

8. Double-click on the bar graph to start the Chart Editor.
9. Click on one of the red bar segments. All the red bars will be highlighted.
10. Click  on the toolbar. The window in Figure 3-31 appears. Make sure that **Fill** is selected.
11. Click on the black color square, then click **Apply**. The red bars will change to black.
12. Without closing the “Colors” window, click on one of the green bars in the bar chart. All the green segments will be highlighted.
13. Click on the white color square in the “Colors” window, and then click **Apply**. The green segments will change to white.
14. Click on one of the blue bars in the bar chart. All the blue segments will be highlighted.
15. Click on the light gray color squares in the “Colors” window, then click **Apply**. The blue segments will change to gray.
16. Click **Close** in the “Colors” window.
17. Click **File**, then click **Close** to exit the Chart Editor.

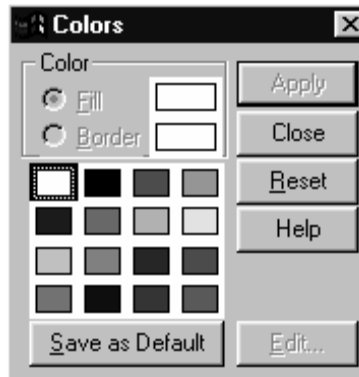


Figure 3-31

The final graph appears in Figure 3-32. The black bars represent the conditional distribution of *interest* for the 18-35 age group, the white bars for the 36-55 age group and the gray bars for the 56-94 age group. From this chart, we can clearly see that the proportion of those who are “not much” interested in politics decreases and the proportion of those who are “very much” interested in politics increases as age increases.

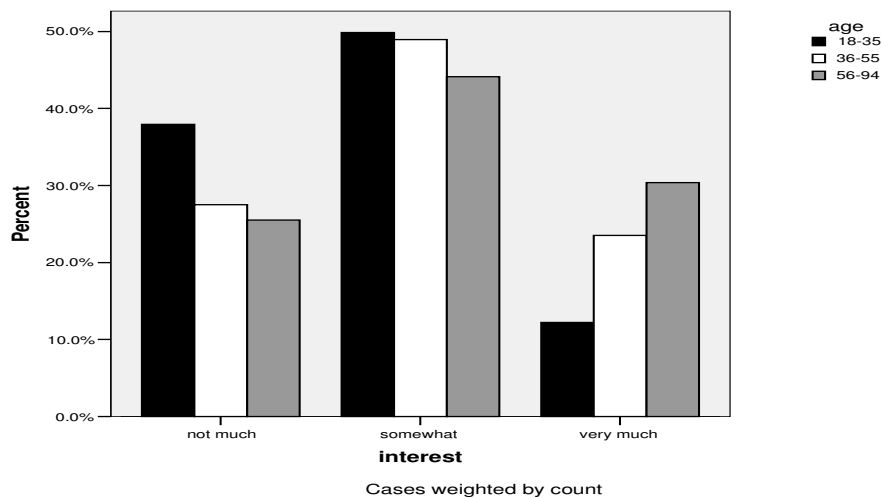




Figure 3-32

To create a segmented bar graph of *interest* given *age*, first be sure that the **Weight cases** option has been activated since this data set has been entered with counts (see Entering count data, page 56). Then follow these steps.

1. Click **Graphs, Legacy Dialogs**, then click **Bar**. The window in Figure 3-33 appears.
2. Click on “Stacked”. “Stacked” should be highlighted as it is in Figure 3-33.
3. Click **Define**. Figure 3-34 shows how this window will appear after the next two steps.
4. Click *age*, then click  to move *age* into the “Category Axis” box.
5. Click *interest*, then click  to move *interest* into the “Define Stacks by” box. The window will now appear as in Figure 3-34.
6. Click **OK**. A segmented bar graph appears in the output window (not shown here). But we are not done yet. Read on.

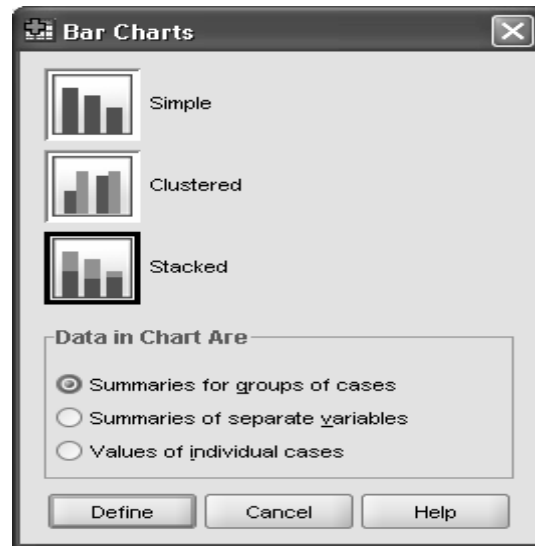


Figure 3-33

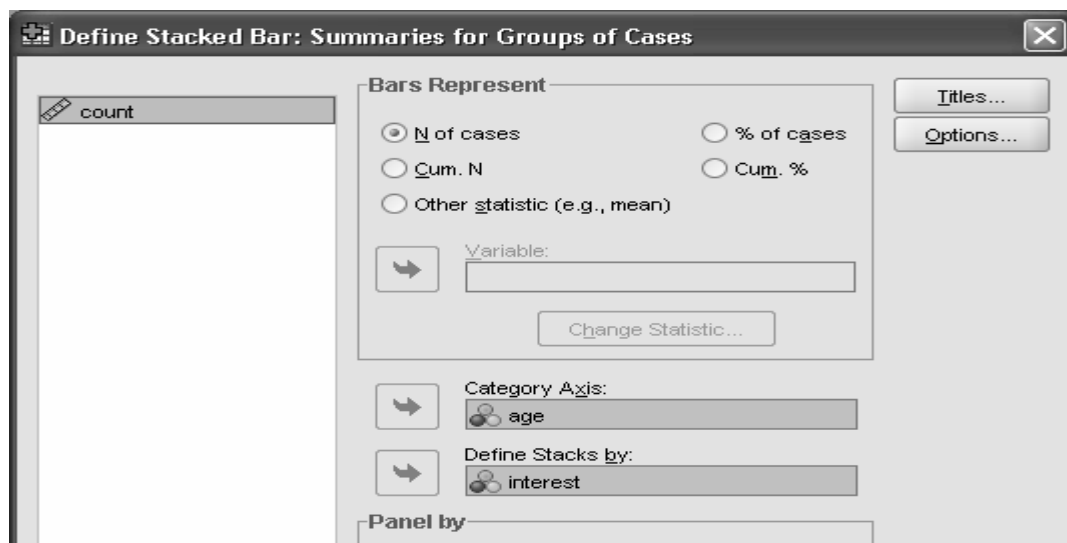



Figure 3-34


We're not done yet, because the bar graph that appears in the output window has bars of different heights. We want all the bars to be the same height so that we can compare the percentage of cases in each  $y$  category across  $x$  categories. Also, the bar graph has colors to identify the segments of the bars and you need to change these if you want to print this on a black-and-white printer. Both these aspects can be changed in the Chart Editor. To make the heights of the bars all the same, follow these steps.

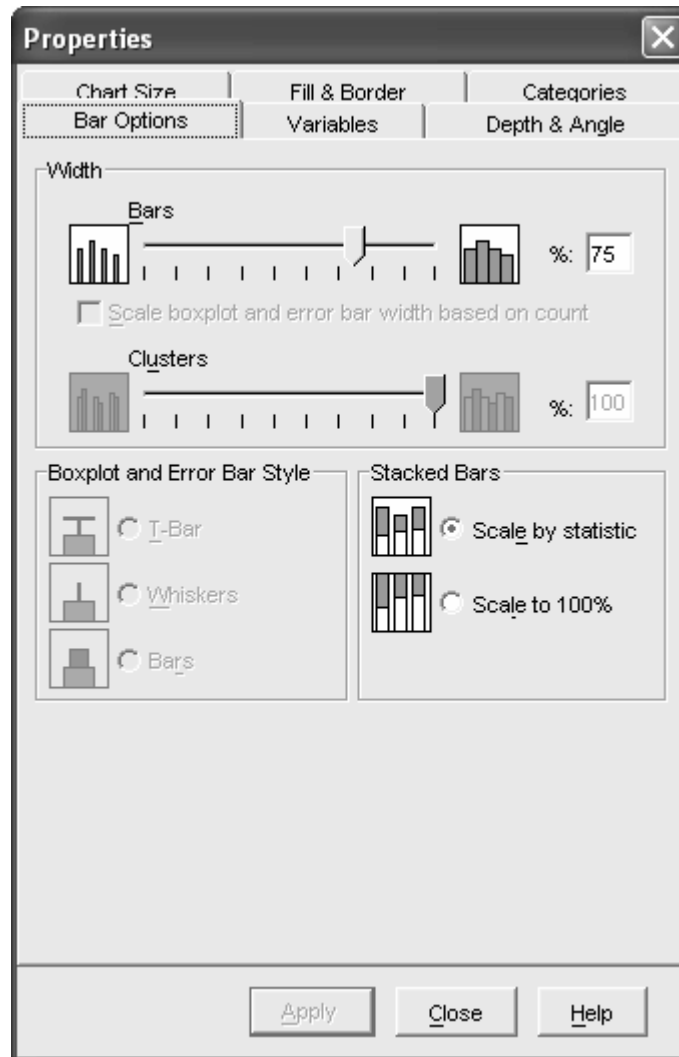
**For SPSS version 15.0:**

7. Click on the **Scale to 100%** icon  on the toolbar.

**For SPSS version 13.0 or 12.0:**


7. Double-click on the bar graph to start the Chart Editor.

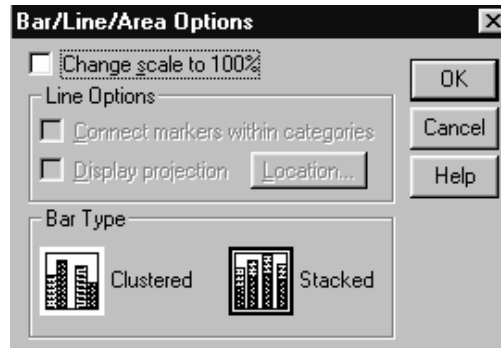
8. Click on one of the bar segments. All the bar segments will be highlighted.
9. Click the **Show Properties Window** icon  on the toolbar. Make sure the **Bar Options** header at the top of the window is selected. The window in Figure 3-35 appears.
10. Click “Scale to 100%” in the “Stacked Bars” window and click **Apply**.
11. Click **Close**.



**Figure 3-35**

**For SPSS versions 11.5 or earlier:**

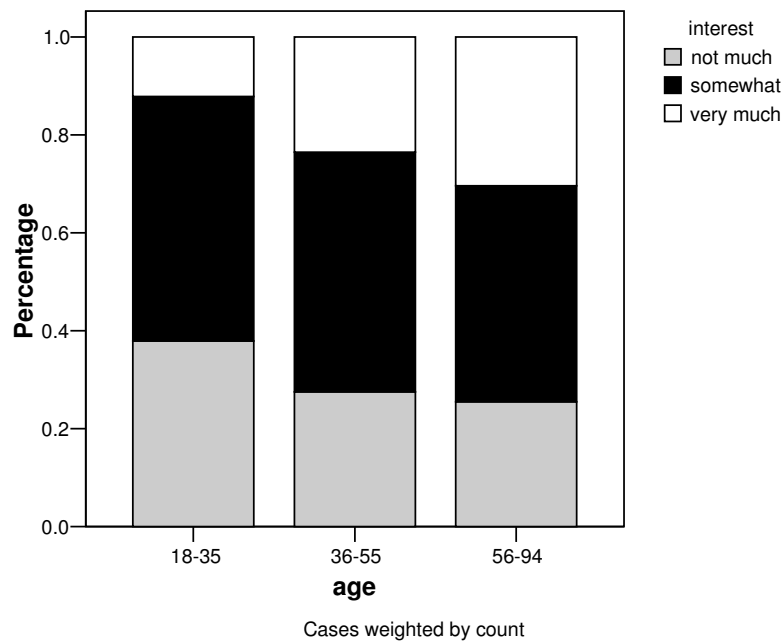
7. Double-click on the bar graph to start the Chart Editor.
8. Click  on the toolbar. The window “Bar/Line/Area Options” in Figure 3-36 appears.
9. Click **Change scale to 100%** so that a check appears in the box in front of this option.
10. Click **OK**.



**Figure 3-36**

The vertical axis now represents the *percent* of cases rather than the number of cases. Unfortunately, SPSS does not change the label on the vertical axis to reflect this: the label is still “Count”. Use the Chart Editor to change the vertical axis label.

The bars are now all the same height. To change the colors of the bars, use the Chart Editor and follow the steps given above for clustered bar charts. Figure 3-37 shows the graph after changing the bar segment colors to gray, black, and white. Again, we see that interest in politics increases as age increases.



**Figure 3-37**