EXERCISE 4

Correcting and preparing your data

This Exercise explores the data in your saved file of merged data (see Exercise 3), consisting of the responses of 335 people (including yourself) to a questionnaire.

Opening SPSS

Open SPSS in the usual way, selecting the data file *Merged Questionnaire Data* which was saved in the previous Exercise. Ensure that the value labels (e.g. Female) are visible in **Data View** (if not, choose **Value Labels** in the **View** drop-down menu or click the **Labels** icon in the toolbar).

Describing categorical data: Obtaining a frequency distribution

Use **Analyze➔Descriptive Statistics➔Frequencies…** procedure described in **4.3.1** to obtain a frequency listing for the variable *Smoker*. In the **Frequencies** dialog box, click **Charts...** . In the **Frequencies: Charts** dialog box, select **Bar Chart(s)**.

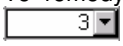
Inspect the frequency table in the **SPSS Viewer**. Is the information in the table what you expected? Before taking any steps to remedy the situation, inspect the bar chart as well.

The bar chart

You will notice immediately that, although the variable *Smoker* was supposed to consist only of Yes and No responses, the horizontal axis of the bar chart also shows a bar for 3. There is obviously an error in the data set. Look at the frequency table again. It shows that the 335 cases that were processed included an entry of *3*. There is also one missing value labelled *System.* In **Data View**, this will be represented by a full stop. Return to **Data View** by clicking the name of your merged file in the **Task Bar** at the foot of the screen.

In **Data View**, you will see that for the variable *Smoker*, Case 10 has a *3* and Case 14 has no value. The *3* in Case 10 should obviously be a *2*, since there is no entry in *NpDay*. In Case 14, the person is recorded as smoking 5 cigarettes per day, so the missing value should be replaced by *1* (Yes). Such transcription errors are common when one is preparing large data sets, which is why it is so important to screen your data before carrying out any analysis. Sometimes it is more convenient to find suspicious values by highlighting the appropriate variable in **Data View** and selecting **Edit➔Find…**.

You then enter the suspect value (in this case *3*) in the **Find what** box and click **Find Next**.

To remedy the two transcription errors that you have found, click *3* for Case 10 to get 

, click the arrow and select *No* from the choice of options. Do the same for Case 14, but select *Yes* from the choice of options.

Save the corrected data file, using the **Save As** item within the **File** drop-down menu, to a new file name *Questionnaire Data (corrected)* so that you do not confuse it with the uncorrected data file *Merged Questionnaire Data*.

Now re-run the **Frequencies** procedure and notice the differences in the output. Your data-screening operation has detected and rectified two errors in the original data set.

Obtaining a bar chart from the Graphs menu

You can obtain a bar chart directly, without any additional statistics, by selecting **Graphs➔Chart Builder…** and selecting the **Simple Bar** from the gallery of **Bar** to obtain the **Simple Bar** preview. Click and drag the variable name *Smoker* to the *X-Axis* box. Click **OK** to obtain the bar chart.

Editing a bar chart

Now try to edit the bar chart in the **Viewer**. (There will be more on editing graphs in Chapter 5.) Initially, bar charts (and other graphics) appear in colour on the screen. A coloured screen image, however, does not print well in black and white. To make the image suitable for black-and-white printing, some editing is necessary. Proceed as follows.

- Double-click anywhere on the bar chart to open the **Chart Editor** window. To edit any part of the figure, you must select that part of the screen figure and double-click it to open the editing dialog box. At the same time, the item(s) will show a purple colour or appear within a purple frame. So double-click one of the bars to see the **Properties** dialog box or alternatively right click to open the **Properties** dialog box.
- Click the **Fill & Border** tab at the top. Click **Fill** and select the desired colour (e.g. a light grey). If you click **Apply**, you can preview the result in the chart and change to another colour if desired. Once you are satisfied with the change, click **Close**. You can also change the fill pattern by clicking the **Pattern** box at the bottom left of the **Color** panel.

It is possible to control many other features of charts and graphs with the **Chart Editor**. For example, by double-clicking an axis, a dialog box will appear enabling you to label the axis and position the label either centrally or to right or left (use the **Justification** selection). You can also change other aspects of the screen figure, such as the spacing of bars and boxes in graphs. (Select the **Bar Options** tab in the **Properties** dialog box.)

There are many other adjustments that can be made; but the way forward is to try some more editing yourself.

When you have finished editing the graph, return to **SPSS Viewer** by clicking  in the top right-hand corner. (You can also leave the Chart Editor by choosing **File** and **Close**.) To save your edited chart, ensure that it has a box around it; if not, click anywhere within the bar chart and a box will appear. Then select **File➔Save** to obtain a directory dialog box for selecting the disk drive and folder for the file.

Try printing out your chart, following the instructions in Section 3.5.

Describing categorical data: Cross-tabulation

The next part of the Exercise is to produce some contingency tables, using the **Crosstabs** procedure (Section 4.3.1)**.** A cross-tabulation is a table showing the frequency of observations in each combination of two categorical variables. Cross-tabulate the *Sex* and *Faculty* of the cases in your merged data set as follows:

Choose **Analyze➔Descriptive Statistics➔Crosstabs…** to open the **Crosstabs** dialog box. Enter one of the variables into the **Row(s)** box by clicking its name and then on  . Enter the other variable into the **Column(s)** box. Click **OK**.

From an inspection of the output answer the following question:

- **How many females are in the Faculty of Medicine?**

You can re-arrange this table by using the **Pivot** procedure (see Section 3.2.2). Double-click on the table so that a hashed box surrounds it. Select the **Pivot** drop-down menu, then **Pivoting Trays**. Experiment with the data by interchanging the variables among the **Layer**, **Column** and **Row** borders. (Do this by clicking and dragging the variables between the borders in the **Pivoting Trays1** box.)

If you want to save the cross-tabulation output, click the second sub-table containing the cross-tabulation and then **Save**. Complete the dialog box.

Finishing the session

Close down SPSS and any other windows before logging out.