

## EXERCISE 7

# Recoding data; selecting cases; line graph

---

### Aim

This Exercise shows you how to recode data, select cases and draw a line graph.

### Opening SPSS

Open SPSS with the data file *Metric Data* saved in an earlier Exercise.

In order to make the graphics involving Faculty simpler, we suggest that the 'Other' (with value 4) Faculty category is eliminated because the file does not contain any case with that category unless you declared yourself as belonging to that category. There are several ways of doing this: here we recommend clicking on the **Variable View** tab of the data file, clicking the options for **Values** for the variable *Faculty*, highlighting 4 = 'Other' in the **Value Labels** dialog box and clicking the **Remove** button. Click **OK** to return to the data file.


### Recoding data

Sometimes you may wish to recode values or categories within a variable (e.g. you might want to combine more than one value or category into a single new value or category). Suppose that you are not particularly interested in whether people are doing a MSc degree or a PhD degree, but just want to know whether they are postgraduates. You can change the data set to give you this information, either within the original variable, *status*, or by creating a new variable containing the recoded information.

In this session you are going to use a new variable, since this retains the original variable *status* for checking that the recoding has been done correctly. It also maintains the original values in the data set.

Use the **Recode** (Section 4.4.3) procedure to recode the status codes *MSc Postgrad* and *PhD Postgrad* (i.e. categories 2 and 3) into a new category 1 and the codes *Undergrad* and *Other* (i.e. categories 1 and 4) into a new category 2.

You will need to follow the section carefully. The **Recode** procedure creates a new variable which you are asked to name: we suggest *NewStatus*. To do this, you will have to choose the **Recode into Different Variables** option within the **Recode** procedure.

- Choose **Transform** → **Recode into Different Variables...** to open the **Recode into Different Variables** dialog box.
- Highlight *Status* and click  to transfer it into the **Input Variable** → **Output Variable** box.
- Type *NewStatus* (remember no space between *New* and *Status*) in the **Name** box within the **Output Variable** box and click **Change**. The new variable name *NewStatus* will now appear alongside *Status*.
- You might also type *New Status* (a space is allowed here) into the **Output Variable Label** box as a label for the new variable *NewStatus*.

- Click **Old and New Values** and then fill in the corresponding values in the **Value** box of **Old Value** and in the **Value** box of **New Value**, clicking **Add** each time. The following should then appear in the right-hand box: 1 → 2, 2 → 1, 3 → 1, 4 → 2.
- Finally click **Continue** and **OK**.

When you have followed this procedure, check that you have the new variable at the far right of your data set. Now you should clarify the values by adding suitable labels. To do this, switch to **Variable View** and then click **None** in the cell in the **Values** column for the row of the new variable *NewStatus*.

When the grey box with three dots appears, click it to open the **Values Labels** dialog box. Complete this box in the usual way by assigning *Postgrad* to value 1 and *Others* to value 2, and finally click **OK**. To see whether this has worked, switch to **Data View** and check the data for the new variable *NewStatus*.

Save the data file again.

Now use the **Visual Binning** procedure to recode the heights of people as Tall, Medium or Short.

- Click **Visual Binning...** in the **Transform** drop-down menu, transfer the variable *Height in metres* to the **Variables to Bin** box by highlighting *Height in metres* and clicking on the arrow. Click **Continue**.
- When the **Visual Binning** dialog box re-appears, click the variable name *Height in metres* in the **Scanned Variable List** box. A histogram will now appear on the right.
- Name the new variable in the **Binned Variable** box as *HeightBin*.
- Enter the following values and labels into **Value** and **Label** cells: 1.70 Short; 1.80 Medium; 2.10 Tall. By accepting the default **Included (<=)** in **Upper Endpoints**, Short is defined as 1.70m or less, Medium as 1.71m to 1.80m, and Tall as 1.81 to 2.10m. The value of 2.10m is arbitrary; any value greater than the tallest person would suffice. Re-read the end of Section 4.4.3 for extra help.
- Click **OK** and note the information that the procedure will create 1 variable. Click **OK** within the information panel.

Check your data to make certain that they have all been classified in the manner that you planned.

## Pie chart

Produce a pie chart with a title showing what percentages of the cases are tall, medium or short. Since there are a few people whose height is not known, use **Select Cases...** (Section 3.1.1) in the **Data** menu to select cases with a height, say, of more than 1 metre by clicking the radio button **If condition is satisfied** and then entering *Metres > 1* in the panel. Edit the pie chart to show the percentage for each slice (see Exercise 6 if you need to refresh your memory).

- **Edit the pie chart to prepare it for black-and-white printing. Print the pie chart.**

## Select cases

It is also useful to be able to select the cases you want to analyse. Suppose, for example, that you wished to consider only the data relating to females. Use **Select Cases** (Section 3.3.1) to specify that only the female cases will be analysed.

Now suppose that, since smoking is said to suppress appetite, you wanted to see whether female smokers were lighter in weight than non-smokers. Use the **Compare Means** (Section 4.3.2) procedure to do this. Remember the **Dependent** variable will be *Kilos* and the **Independent** variable *Smoker*.

- **Print the Report table produced. Note that as a result of the Select Cases procedure you have just followed, this table will apply to the female respondents only.**
- **Are there any differences between the smokers and the non-smokers? Comment briefly on any differences you find. (When you think about this, bear in mind the difference in size between the smoking and non-smoking groups.)**


## Line graph

A **line graph** is suitable when there is an interval or ordinal scale for one of the variables with not more than about ten values. When the scale is nominal, a bar chart is preferable. Now that you have an ordinal scale of height with three values in the variable *HeightBin*, you can draw a line graph of *Sex* against *HeightBin*.

First, however, the selection of females in the previous section must be reversed by returning to the **Select Cases** dialog box and clicking the **All cases** radio button.

- Choose **Graphs**→**Chart Builder**→**Line** and click the Multiple Line picture to highlight it. Drag it to the **Chart preview** above.
- Insert the variable *HeightBin* into the **X-Axis** box and *Sex* into the **Set pattern** box.
- Select the **Statistic** directory in the **Element Properties**, click **Percentage (?)** and then click **Apply**.
- Click **OK** to plot the lines.

A two-line graph should then appear, one line for Male and one line for Female, with the points on the abscissa labelled Short, Medium and Tall. Click the chart to enter **Chart Editor** and delete (**Binned**) from the x-axis label by clicking on it so that it is highlighted with a purple border and then move the cursor in and delete the appropriate section. In order to differentiate the sexes clearly in the printed graph, change one of the lines to a discontinuous line by clicking on one of the lines in the Sex of Student legend box and then selecting a different **Style** from the choice in the **Lines** dialog box. You can change the colour within the same dialog box. Click **Apply** and **Close**. You may also wish to show the markers for the different

categories, in which case, select the icon  from the toolbar (**Show Line Markers**). If you wish to change the style of the markers, double-click on one of them to open the **Properties** dialog box. There the type, size and colour can be altered.

## Finishing the session

Close down SPSS and any other windows before logging out of the computer.