Chapter 1

Getting started with SPSS

Introduction

The aim of this first chapter is to give you an overview of SPSS and, through this, to show you that handling and analyzing quantitative data need not be difficult. In particular, by the end of this chapter you will:

- understand what a dataset is;
- be able to open an existing dataset with SPSS and also create your own dataset;
- understand the SPSS environment and be able to navigate your way around it;
- have gained some experience of undertaking simple analyses with SPSS;
- be able to save a dataset and also the output of the analyses you have undertaken.

Understanding what a dataset is

At the heart of quantitative data analysis is the dataset. A dataset is actually just an array of numbers organized into rows and columns. Perhaps the best way to explain it is to start with a real-life example. You will see in Box 1.1 an amended and reduced version of a self-complete questionnaire used as part of the Northern Ireland Young Life and Times Survey 2005. The full questionnaire is much longer than this. All I have done here is to keep two basic questions (what sex the young person is and then what type of school they went to) and then the eight questions in the questionnaire on bullying in school. The layout for the bullying questions is almost exactly as it appears in the original questionnaire.

There are two key things to note from this questionnaire. The first is that each question is given an abbreviated name (i.e. "RSEX," "TYPESCHL," "SCLOTBUL" and so on). As will be seen shortly, each question basically equates to a single **variable** and these abbreviated names are the names used by SPSS for each variable. Many questionnaires of this type do not actually include the variable name like this but simply add them in afterwards. The other thing to note is that there is a number assigned to each of the boxes that the respondent is able to tick. These represent the **values** that each of the variables can take. Thus for the first question—"Are you male or female?"—the variable name is "RSEX" and this variable only has two values: either "1" or "2," indicating that the respondent is male or female respectively.

The actual dataset derived from this questionnaire, as it appears in SPSS, is shown in Figure 1.1. You will shortly be asked to open this dataset and explore it. However, for now it is useful just to draw out and highlight two key points from this. First, each row (i.e.

Box 1.1 Amended extract from the Northern Ireland Young Life and Times Survey 2005

RSEX
1. Are you male or female?

	✓
Male	[ ] 1
Female	[ ] 2

TYPESCHL
20. What type of school do you (did you last) attend?

	✓
Planned integrated	[ ] 1
Grammar	[ ] 2
Secondary	[ ] 3
Other	[ ] 4

The next few questions are related to bullying in school.

SCLOTBUL
22. Would you say that students at your school get bullied by other students?

A lot	[ ] 1
A little	[ ] 2
Not at all	[ ] 3
Don't know	[ ] 4

STFBULJB
23. Are there particular staff at your school whose job it is to deal with bullying?

Yes	[ ] 1	(Please go to next question 24)
No	[ ] 2	(Please go to next question 25)
Don't know	[ ] 3	(Please go to next question 25)

GOTOSTAF
24. Do you think that most people – if they were bullied – would or would not go and talk to one of these members of staff?

Would talk to them	[ ] 1
Would not talk to them	[ ] 2
It depends	[ ] 3
Don't know	[ ] 4

SCHLBUL
25. In general, do you think your school provides real help for people who are bullied or not?

Yes	[ ] 1
No	[ ] 2
Don't know	[ ] 3

UBULLSCH
26. Have you yourself ever been bullied in school?

	✓	
Yes	[ ] 1	(Please go to the next question)
No	[ ] 2	(Please go to question 28)

OFTENBUL
27. How often have you yourself been bullied at school in the last two months in school?

	✓
A lot	[ ] 1
A little	[ ] 2
Not al all	[ ] 3

UBULLOTH
28. Have you yourself ever taken part in bullying other students?

	✓	
Yes	[ ] 1	(Please go to the next question)
No	[ ] 2	(Please go to question 30)

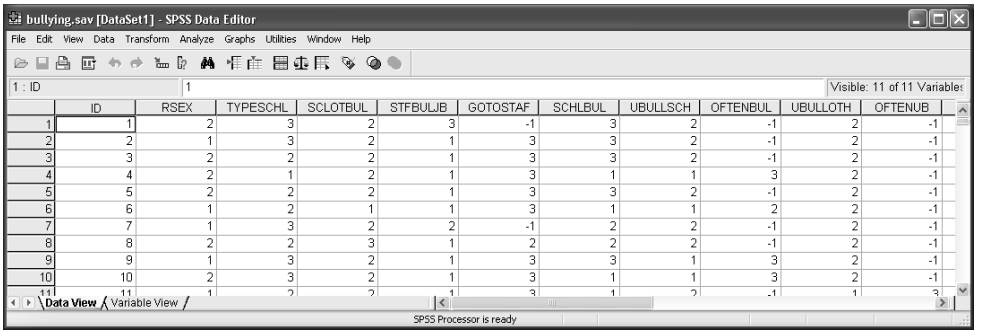
OFTENUB
29. How often have you taken part in bullying other students at school in the last two months in school?

	✓
A lot	[ ] 1
A little	[ ] 2
Not at all	[ ] 3

horizontal line) represents what is called one case. A **case** is the term used to refer to the unit of analysis. In this example each case represents one young person who completed the questionnaire. Only the first ten cases are actually visible in Figure 1.1. However, this questionnaire was completed by 819 young people and so there are a total of 819 cases (i.e. 819 rows) in this dataset and they could be viewed simply by scrolling down using the right-hand scroll-bar. The second thing to note is that each column (i.e. vertical line) represents a variable. The name of each variable can be seen at the top of each column and they correspond to and appear in the same order as the variables in the questionnaire itself (Box 1.1). The only additional variable that does not appear in the questionnaire is the first one—"ID"—which is the unique number given to each questionnaire (and thus is a unique number that can be used to identify each case).

All of the numbers that appear in the middle of the screen are basically the values for each of the variables (i.e. the values corresponding to what boxes each respondent ticked). To take the first case as an example (i.e. the first row) then reading across horizontally we can see that their unique ID number is "1," their sex ("RSEX") is coded as "2" and the type of school they attend ("TYPESCHL") is coded as "3" and so on. If we refer back to the original questionnaire we can see that these values mean that this person is therefore female and attends a secondary school. If we continue we can see that the value for the next variable ("SCLOTBUL") is "2". Again, referring back to the questionnaire we can see that this means that when asked "Would you say that students at your school get bullied by other students?" this respondent answered "A little." You can easily continue across the rest of the line to find out how she answered the remaining questions.

The only other thing that needs to be explained are the values of "-1" for the variables "GOTOSTAF" and "OFTENBUL." These are values added in afterwards by the researcher to indicate that these two questions were skipped by the respondent. To understand this, have a look again at the questionnaire in Box 1.1. For Question 23—"Are there particular staff at your school whose job it is to deal with bullying?"—the respondent is given differing instructions depending on how they answer this question. Thus if they answer "yes" (there is a particular member of staff) then they are asked to go onto the next question (Question 24) which asks how approachable that member of staff is. However, if they had answered "no" or "don't know" to Question 23, as our first respondent has, then there is no point asking them this next question and so they are instructed to skip it and jump directly to Question 25. In such circumstances, rather than



	ID	RSEX	TYPESCHL	SCLOTBUL	STFBULJB	GOTOSTAF	SCHLBUL	UBULLSCH	OFTENBUL	UBULLOTH	OFTENUB
1	1	2	3	2	3	-1	3	2	-1	2	-1
2	2	1	3	2	1	3	3	2	-1	2	-1
3	3	2	2	2	1	3	3	2	-1	2	-1
4	4	2	1	2	1	3	1	1	3	2	-1
5	5	2	2	2	1	3	3	2	-1	2	-1
6	6	1	2	1	1	3	1	1	2	2	-1
7	7	1	3	2	2	-1	2	2	-1	2	-1
8	8	2	2	3	1	2	2	2	-1	2	-1
9	9	1	3	2	1	3	3	1	3	2	-1
10	10	2	3	2	1	3	1	1	3	2	-1
11	11	1	2	2	1	3	1	2	-1	1	3

Figure 1.1 The bullying dataset as it appears in SPSS

just leaving this variable ("GOTOSTAF") blank for this respondent a special value, "-1," is typed in to indicate that this question has been legitimately skipped.

Before we get you to actually open up and explore this dataset there are three general points to draw from what you have seen so far. The first is that all data need to be translated into numeric form for SPSS to work with (with the exception of descriptive "labels" that we will get onto later). Thus, while the categories that a respondent can choose from for each question are all described in words (e.g. "A lot" or "A little"), we have had to assign them numbers so that they can be entered into the dataset. An important lesson from this is that you should, therefore, think very carefully about how you design your questionnaire to ensure that, as far as possible, the data you gather can be coded in this way. This usually means trying as far as possible to restrict yourself to using **closed questions** (i.e. questions where there are only a fixed number of response categories to choose from). There will obviously be times when you need to include **open-ended questions** (i.e. a question that is followed by a space where the respondent writes down their answer in their own words). However, you need to bear in mind that you will have to go back and translate these qualitative answers into codes at some point if you want to analyze them quantitatively.

The second key point following on from this is that while numbers have been used for convenience to represent these different categories they may not actually mean anything numerically. Thus while males are coded "1" and females "2" for the variable "RSEX" this has no significance whatsoever. It does not mean, for example, that females are twice as much as males. It could easily have been coded the other way around or using other numbers (i.e. "0" and "1" or whatever). What this means is that we need to be extremely careful in terms of understanding precisely what type of variable we are dealing with and what the values associated with each variable actually represent. This is something we will return to in the next chapter.

The third and final key point to draw from the questionnaire and dataset shown is the usefulness of what is called **precoding** your questionnaire. Precoding a questionnaire basically means including where possible the variable names and values for each response category on the questionnaire itself. Box 1.1 shows just one of the possible ways that this can be done. The benefit of doing this is in terms of helping to reduce errors that can occur when entering data into SPSS directly from the questionnaire. Imagine, for example, that you have 500 questionnaires of four pages in length and the questions are not precoded and neither are the response categories. In such circumstances it would be very easy to make mistakes typing in the data. If you can precode your questionnaire then whatever box has been ticked you can see immediately which variable it relates to and also what the value is that you need to type in.

Opening an existing dataset in SPSS

Having been introduced to the bullying dataset it is now time to open it up and explore it in a little more detail. First of all you need to download the dataset from the companion website. All the datasets featured in this book are downloadable from the website and you can follow the same routine in each case.

To download and then open the bullying dataset you should begin by going to the companion website and finding the list of datasets. Right click on the link to download the **bullying.sav** dataset. Select "Save Target As. . ." and this will open the Save As window shown in Figure 1.2. Select an appropriate place to save the dataset using the "Save in:"

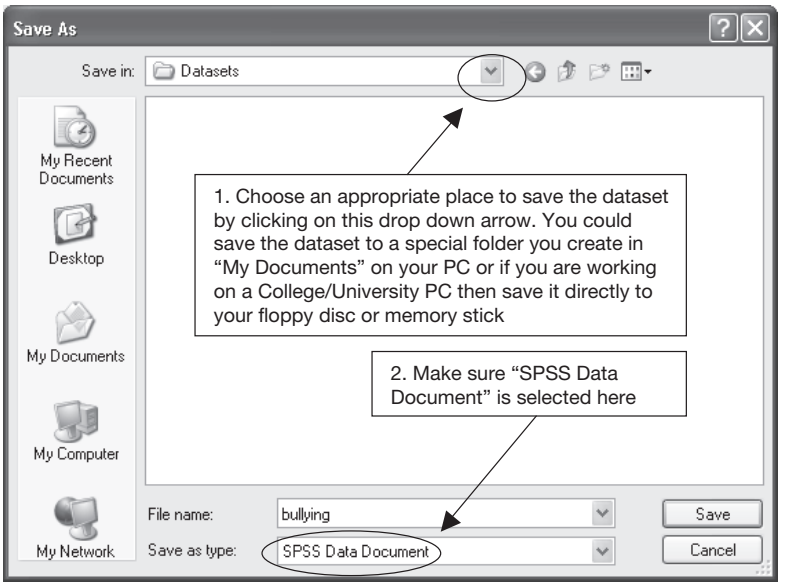


Figure 1.2 Save As window

drop-down menu as shown and also make sure you select "SPSS Data Document" using the "Save as type:" drop-down menu. When you have done this click on the "Save" button. When the download is complete another dialogue box will appear and you should click the "Close" button.

Open SPSS as you would any other program. For those of you using Windows XP operating system you can do this by clicking on the Start button located at the bottom left of your screen as shown in Figure 1.3. This will bring up the initial **SPSS 15.0 for Windows** window as shown in Figure 1.4. Make sure that the option "Open an existing data source" is selected and that "More Files. . ." is selected within this and then click "OK." This, in turn, opens up the **Open Data** window as shown in Figure 1.5. Make sure you select "SPSS (*.sav)" for "Files of type:" as shown and then use the "Look in:" drop-down menu to navigate around your PC to find the "bullying.sav" dataset where you saved it. Click on the dataset and then click on the "Open" button. You should now have the screen originally shown in Figure 1.1.

Most of the default settings for SPSS are fine and not worth changing. However, one minor change is worthwhile that will help you subsequently when you start exploring and analyzing the data. To make this change select **Edit → Options. . .** (or **SPSS → Preferences. . .** if you are using a Mac version). This will open the dialogue box shown in Figure 1.6. To help you find variables quickly and easily for analysis you should change the settings so that SPSS displays the shortened variable names. To do this, select "Display names" (or just "Names" for Mac versions) as shown and then click "OK." You will see what effect this has a little later. In addition, if you were dealing with a large dataset with a lot of variables then you may also find it useful to ask SPSS to display the variable names alphabetically. However, this is not necessary for the present book. Rather, we will stick to the default option of SPSS displaying variable names as they appear in the dataset.

Figure 1.3 Opening SPSS in Windows XP

We can now begin to explore the dataset. You will see a number of short-cut icons running across the top of the SPSS window, immediately below the top menu. If you hover your cursor over any of the icons a short description will appear explaining what the icons represent. They are all pretty self-explanatory. There are just two that are worth pointing out here. The first is the **Value labels icon** second from the right, which looks like a luggage tag as indicated in Figure 1.7. Clicking on this icon will display the value labels rather than the numbers in the dataset itself. To see what this means try clicking on the icon. You should see that the numbers in the dataset are replaced by their labels as shown in Figure 1.7.

The other icon to note is the **Variables icon** as indicated in Figure 1.8. Clicking on this icon calls up the **Variables** window as also shown. This is a useful feature that helps you to explore the variables in the dataset. Clicking on any of the variables in the left-hand pane will automatically bring up its details, including how it is coded and labeled as shown in Figure 1.8. You can now also see what the effect was of changing the default setting in SPSS so that it displays the variable names in lists as here.

Another way to explore the variables is to click on the "Variable View" tab at the bottom left of the window. Doing this results in the view as shown in Figure 1.9. As can be seen, all the variables are now listed in rows with all of the relevant details, including the description of the variable ("Label") and how it is coded ("Values"). As can be seen

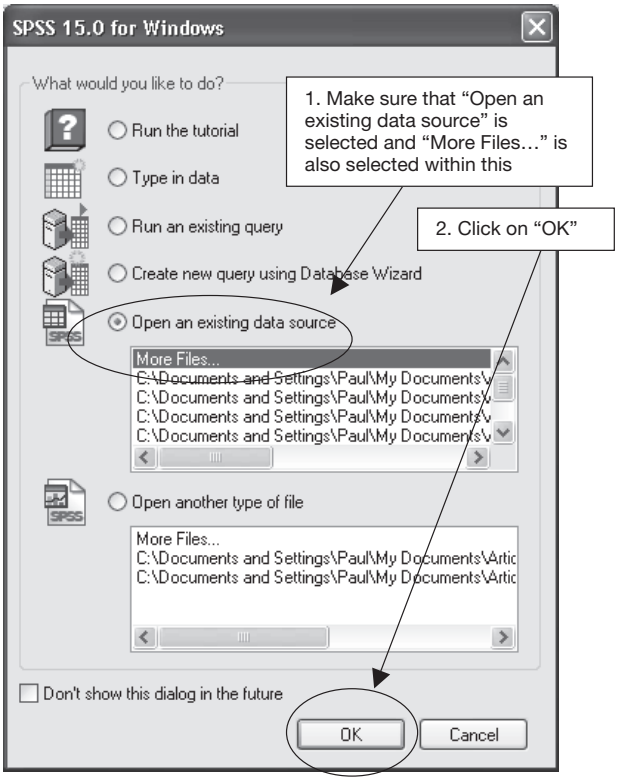


Figure 1.4
SPSS 15.0 for Windows window in SPSS

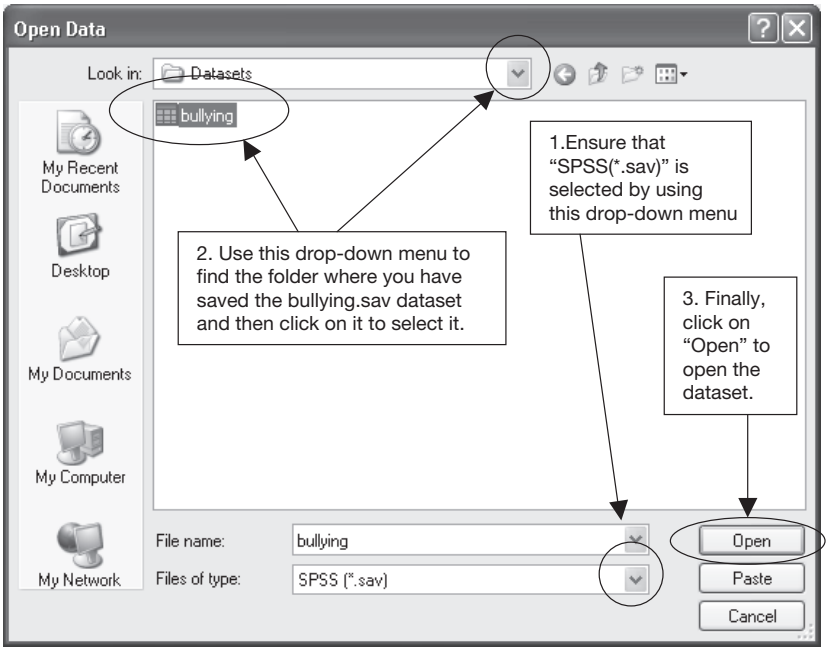


Figure 1.5 Open Data window

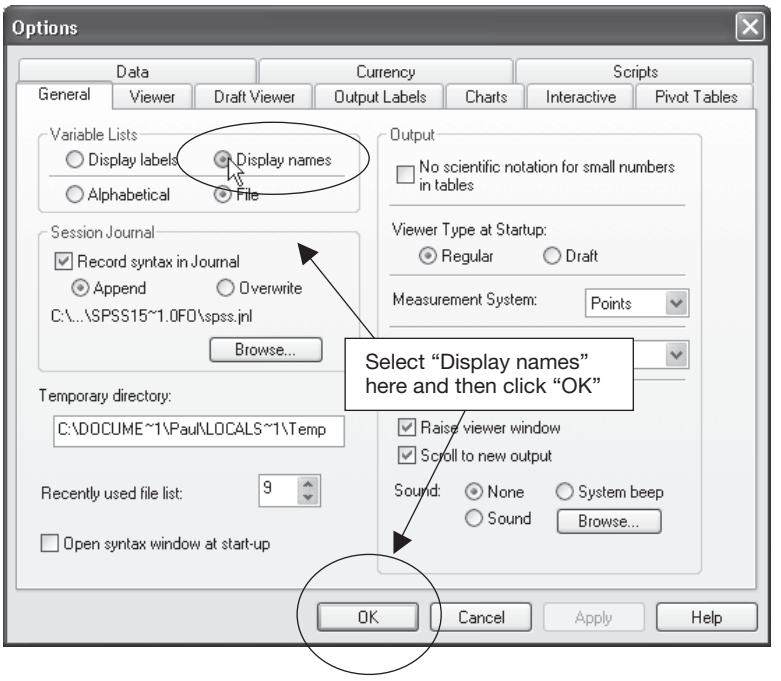


Figure 1.6 Options window in SPSS

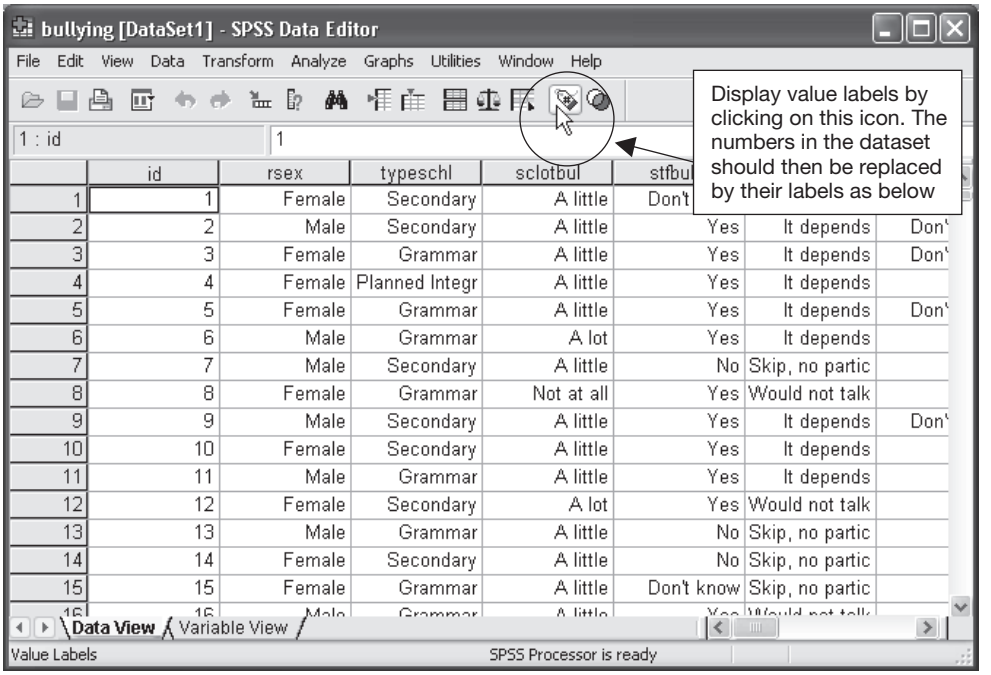


Figure 1.7 Value labels icon in SPSS

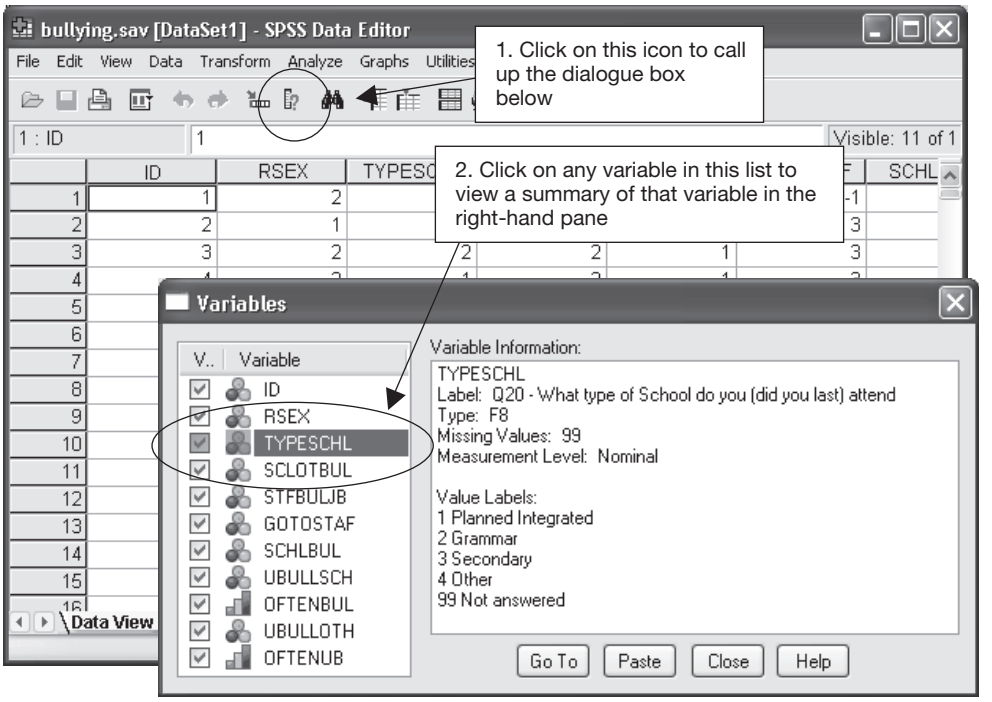


Figure 1.8 Viewing summaries of variables

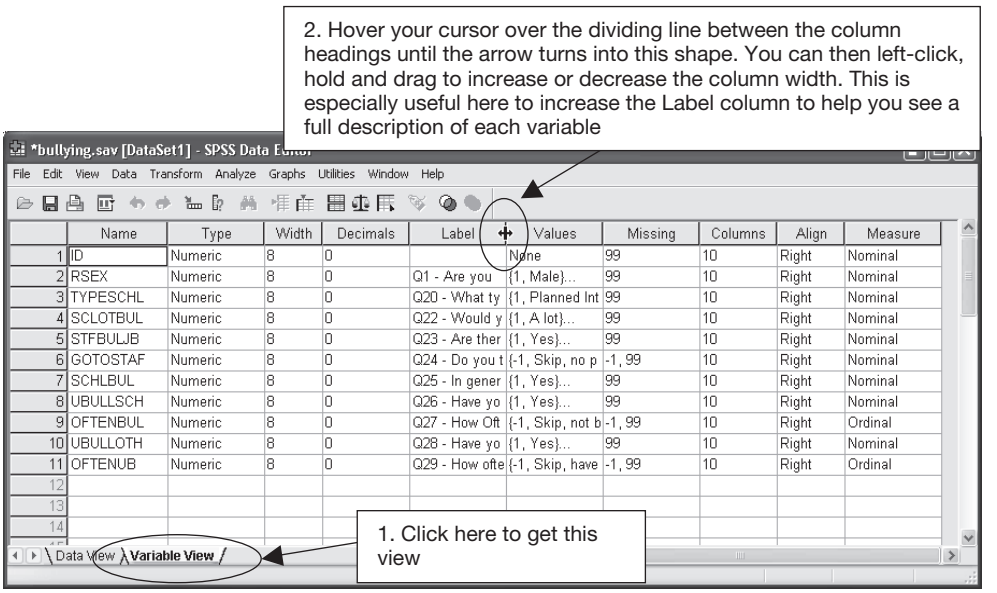


Figure 1.9 Variable View in SPSS

from Figure 1.9, only the first part of each of the variable labels is showing. This can be rectified by increasing the width of the Label column as explained in Figure 1.9.

If you want to view the list of values and how they have been coded for any of the variables then all you need to do is to click on the relevant cell in the Values column for the variable you are interested in. A small grey box then appears in the right-hand side of that cell. Clicking on this box calls up the **Value Labels** window shown in Figure 1.10. Also note in Figure 1.10 the effects of widening the Label column as suggested earlier.

We will examine more of the features of the Variable View in SPSS in the next section when we go through the procedure for creating your own dataset. For now, it would seem a shame having reached this stage if you were not given the chance just to have a little go at analyzing the data. With this in mind, let's just see what proportion of young people in the sample claimed to have been bullied in school. Suppose we want to calculate the actual numbers and percentages who answered "yes" and "no" to this question as well as display this with a bar chart. To do this we will examine the responses to Question 26 (see Box 1.1) which relates to the variable named "UBULLSCH."

To do this, select **Analyze → Descriptive Statistics → Frequencies. . .** from the top menu. This calls up the **Frequencies** window as shown in Figure 1.11. Select the variable you are interested in (i.e. "UBULLSCH") by clicking on it once. The variable should be highlighted and the arrow button in the middle of the window should become darkened. Click on the arrow button as shown to place that variable in the right-hand pane.

To generate a bar chart as well as a frequency table you need to click on the "Charts . . ." button as also shown in Figure 1.11. This calls up the additional **Frequencies: Charts** window. Select "Bar charts" and then "Percentages" as indicated. Once done, click the

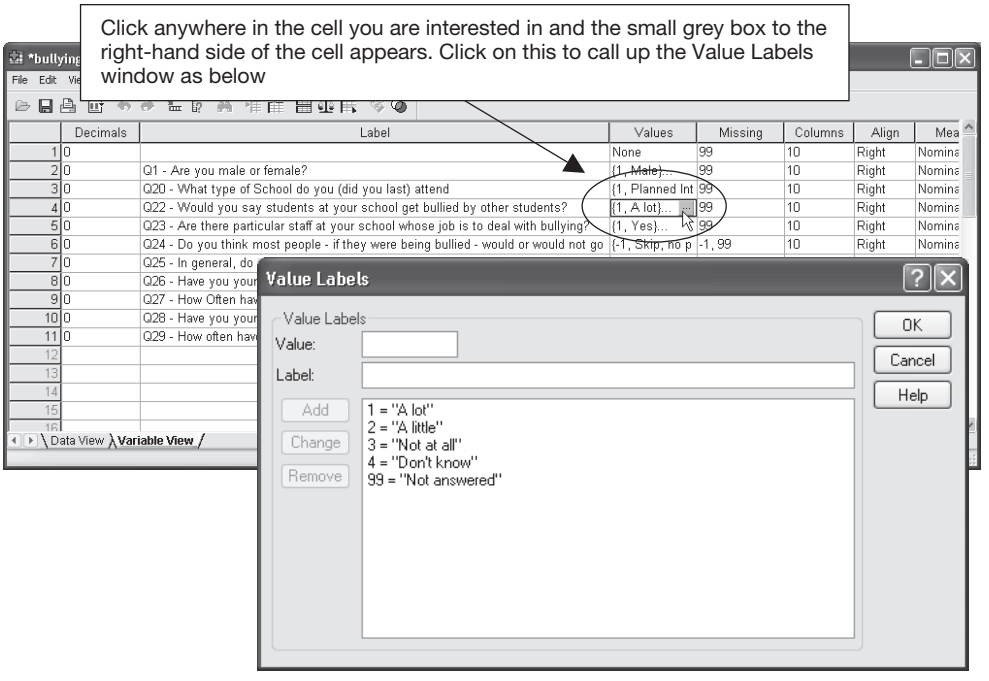


Figure 1.10 Value Labels window in SPSS

Continue button. Finally, click OK in the main **Frequencies** window. This will produce a new **Output** window with the results as shown in Figure 1.12.

As can be seen from the main frequency tables, 16 respondents (2.0 percent) did not answer the question. Of the rest, 30.4 percent claimed to have been bullied in school. While we have only done this to give you a bit of a taster of what SPSS can do, I cannot resist making two quick substantive points about this output. First, there is actually a problem of validity here in terms of what constitutes "being bullied" at school. Unfortunately, and as can be seen from the original questionnaire (Box 1.1), no definition of bullying was actually offered to respondents before being asked to answer this question. With no guidance, it is likely that different respondents will have different things in mind when they think of bullying. Therefore, we need to be careful in interpreting this statistic of 30.4 percent. The second point relates to the bar chart. I encouraged you to do this just so that you can see how easy it is with SPSS to generate charts like this. However, one of the points I will be making in Chapters 3 and 4 is that we need to think about the use of charts carefully. In cases where there are only two categories, as here, it is almost always unnecessary to include a bar chart to illustrate the findings. Simply stating the percentage figures of those who responded "yes" and "no" would be sufficient.

To continue analyzing the data all you need to do is to minimize this output window and continue working with the main dataset. Each time you do a further bit of analysis, the results are added to the ones already created in the output window. As you go on, therefore, this output window keeps a record of the results of all of your analysis.

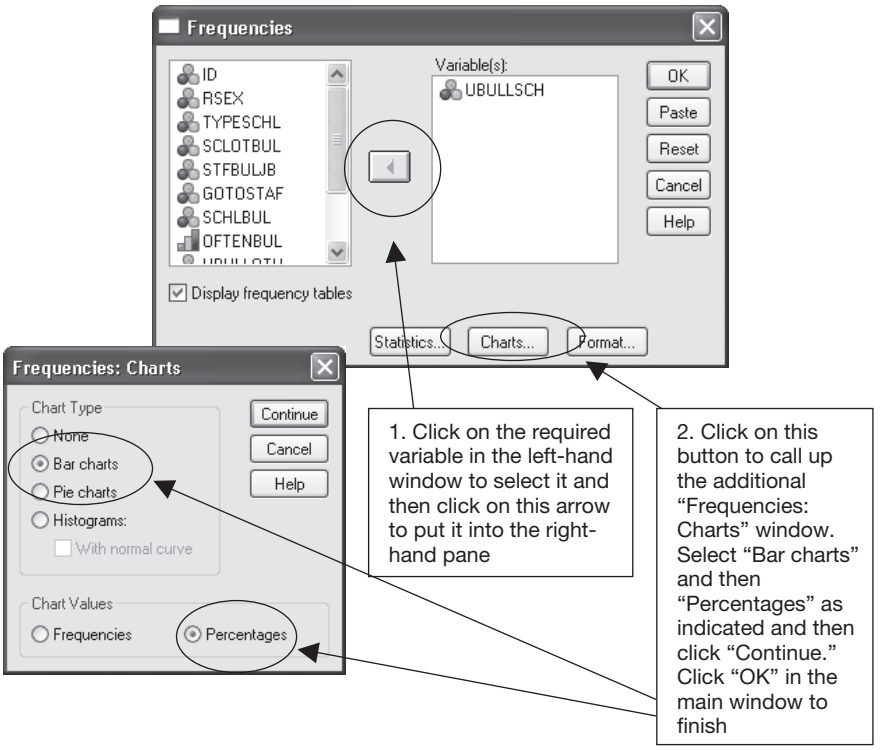


Figure 1.11 Frequencies window in SPSS

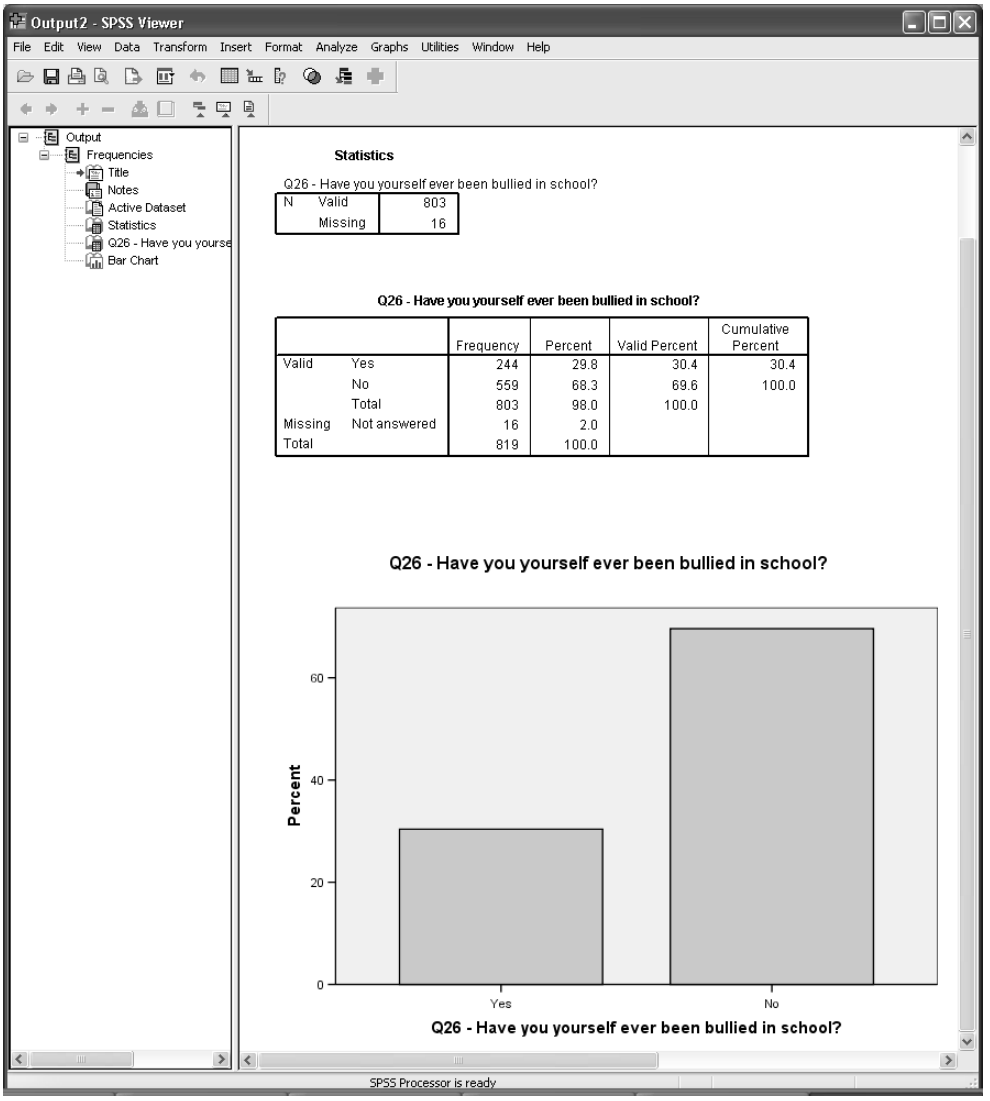


Figure 1.12 SPSS Output window

EXERCISE 1.1

Explore the data a little further. See if you can calculate the frequencies and percentage responses for the other questions asked. Try generating a few more bar charts. Also, have a go at creating a pie chart. You will learn a lot from SPSS by simply experimenting with data in this way.

When finished, you can save this output as a separate file. To do this, simply choose **File → Save As. . .** from the top menu of the **Output** window. You will need to give the file a name (give it something short but meaningful so that you will be able to distinguish between the many different output files you will tend to accumulate) and then choose a place to save it. The file will be saved with the suffix ".spo" that indicates that it is an output file.

Finally, whenever you come to analyze the data properly you are bound to want to make some changes to the dataset itself. You should always remember, therefore, to save the main dataset before closing it down. This can be done either by just clicking on the Save Icon contained within the list of icons towards the top of the main **Data View** window (resembling a floppy disc) or by selecting **File → Save As. . .** from the top menu.

Creating your own dataset

So far in this chapter you have been introduced to the SPSS environment and shown how to open, analyze and save an existing dataset. In this final section we will look at how you would create your own dataset from scratch. The dataset we will create this time involves basic educational and economic indicators for 20 countries as shown in Table 1.1. All of these data are freely available from the website of the United Nations Statistics Division—http://unstats.un.org—which provides an excellent example of how easy it is for you to collect quantitative data for secondary analysis. In fact, if you are a student there is no reason why you could not use data such as these as the basis for your dissertation.

EXERCISE 1.2

To start with, have a quick look at the website to see where these figures were taken from. See what other social, economic and educational indicators are available.

The 20 countries in Table 1.1 were chosen randomly from all those for whom information on illiteracy rates were available. As can be seen in Table 1.1, alongside male and female illiteracy rates, two further indicators have been included—school life expectancy (which is the total number of years of schooling that a child in that country can expect to receive on average) and per capita Gross Domestic Product (which is a measure of the value of the total output of economic goods and services produced in that country per head of population). Further details on each of these measures, together with an outline of their limitations, are provided on the UN Statistics Division website.

The dataset to be entered into SPSS will look just as it does in Table 1.1 but with an additional first column representing a unique "ID" number for each country. To begin creating the dataset we need to open SPSS as outlined earlier, selecting "Type in data" in the initial **SPSS 15.0 for Windows** window (see Figure 1.4) this time. Before typing in any numbers we begin by defining the variables. The following explanation is for SPSS Version 10.0 and above. For an outline of how to define variables for Version 9.0 and earlier see Appendix 1. To begin with, we click on the "Variable View" tab as shown in Figure 1.13. What we will do now is work our way across the columns to type in the information required.

Table 1.1 Educational and economic indicators for 20 countries

Country	*Continent*	*Estimated adult (15+) illiteracy rates (%)*		*School life expectancy (years)*	*Per capita GDP (US$)*
		Males	*Females*		
Argentina	Americas	3.0	3.0	16	3,375
Benin	Africa	45.2	74.5	7	521
Burundi	Africa	33.2	48.1	5	86
Chile	Americas	4.2	4.4	14	4,523
Dominican Republic	Americas	12.0	12.7	12	2,408
El Salvador	Americas	17.6	22.9	11	2,302
Ghana	Africa	37.1	54.3	7	354
Hungary	Europe	0.6	0.7	15	8,384
Iran	Asia	16.5	29.6	11	2,079
Laos	Asia	23.0	39.1	9	361
Malta	Europe	8.2	6.6	14	11,790
Mauritania	Africa	48.5	68.7	7	381
Morocco	Africa	36.7	61.7	9	1,463
Namibia	Africa	16.2	17.2	12	2,307
Senegal	Africa	43.9	71.5	6	641
Sierra Leone	Africa	60.2	79.5	7	197
Swaziland	Africa	19.6	21.9	10	1,653
Macedonia	Europe	1.8	5.9	12	2,225
United Arab Emirates	Asia	24.4	19.3	11	22,130
Uruguay	Americas	2.7	1.9	15	3,274

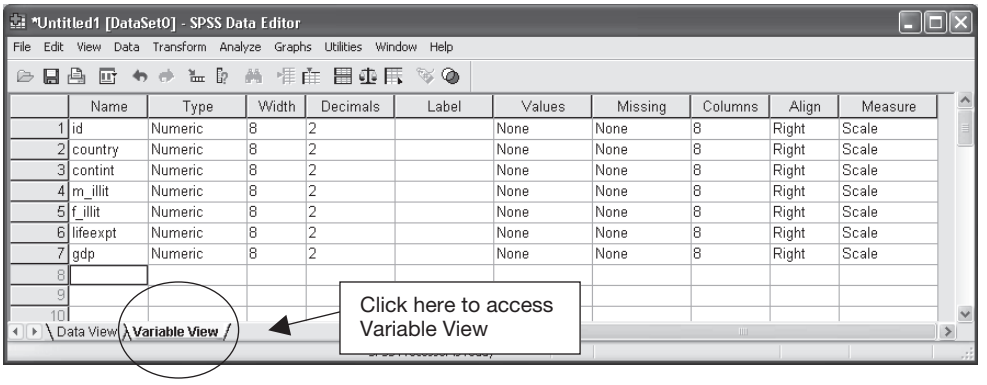


Figure 1.13 Defining variables in SPSS

The first column, headed "Name," is where you type in a shortened variable name for each of the variables. This has already been done in Figure 1.13 for all the variables that have simply been entered in the order they appear in Table 1.1 (but including the "id" variable first). To start entering the names simply click in the first cell (top left) and type in "id" and then click in the second cell below and type in the name of the second variable ("country") and so on. You will see I have used shortened names for each of the variables. From SPSS Version 14.0 onwards you can actually use names up to 64 characters.

However, as earlier versions restricted variable names to just eight characters I have just developed the habit of working to this and, because some of you may have earlier versions of SPSS, we will just stick to this practice. In terms of the names you use, you can choose any name you want but it is obviously useful to pick something that can help remind you of what that variable actually represents. All variable names have to begin with a letter and some characters are not allowed; SPSS will politely tell you if you try to use one of these!

As you type in each variable name and press return (or click in the next cell below) you will see that the remaining cells in that row are automatically filled with the default settings for variables. Once you have typed all the names in, you need to amend the information in these remaining cells as necessary for each variable, which is what we will do next.

The second column, headed "Type," simply defines what type of variable we are dealing with. In most cases this will be numeric (which is the default setting). As the name suggests, a **numeric variable** is one where all the values are represented by numbers. However, we can see in our dataset that there is at least one variable that is not numeric (i.e. "country") but actually consists of words. This is what is known within SPSS as a **string variable**. To amend numeric to string for the variable "country," click anywhere in the cell with the word "Numeric" in it and then click on the small grey square that will appear in the right-hand side of that cell, as shown in Figure 1.14. This opens the **Variable Type** window as also shown in Figure 1.14. Select "String" and enter the maximum number of characters for this variable. This should obviously be at least as much as the longest name included within this variable (in this case the "United Arab Emirates"). All the other variables will be numeric and so no other amendments are required in this column. (You may be thinking why do we not define the variable "Continent" as a string variable as well? This will become clear shortly.)

The next column, headed "Width," specifies the maximum number of digits or characters for each variable. As can be seen, for the variable "country" this has changed to 20 (characters) as we have just specified. For the other numeric variables, the default is 8 (digits) and this tends to be more than sufficient and therefore need not be changed.

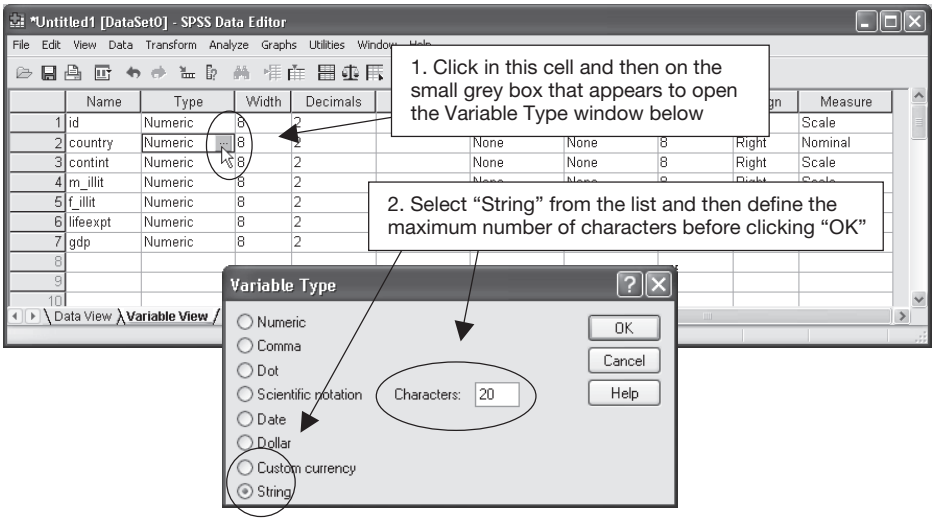


Figure 1.14 Variable Type window in SPSS

As the heading suggests, the next column—"Decimals"—specifies the number of decimal places that are displayed in SPSS for each variable. It is clearly no longer relevant to the string variable "country" and, hence, there is a "0" entered for that variable and this cannot be altered. For the remaining variables, to alter the number of decimal places simply click in the cell required and type in the required number. Alternatively you could use the "up" and "down" arrows that appear to the right-hand side of the cell when you click in it to increase or decrease the number as required. For the remaining variables, it can be seen from Table 1.1 that the male and female illiteracy rates are reported to just one decimal place while the remaining variables have no decimal places (including the variables "id" and "continent"). You should therefore alter the number of decimal places accordingly so that they appear as they do in Figure 1.17.

The next column, headed "Label," includes a proper description of each variable that will appear in any tables or charts you produce with SPSS. It is, therefore, worth naming these properly. To add in the variable names, simply click in the relevant cell and type in the name. You will find that the column width increases automatically as you type. The names you enter should be as shown in Figure 1.17.

The "Values" column is used to specify value labels where necessary. In this particular case this only needs to be done for the variable "continent." No labels are needed for the last four variables as the numbers actually represent real numbers (this will all be explained properly in the next chapter). Similarly, there is no need to add labels for "id" or for "country." As for "continent" it can be seen that there are four possible values, corresponding to the four continents. When typing in the data into SPSS, rather than making this a string variable, it is easier to give each continent a number and type that number in. Which number we assign to each continent is arbitrary. For now we can just code it as follows: Africa = 1; Americas = 2; Asia = 3; and Europe = 4.

What we need to do in this Values column is to set the labels for each of the four numbers. To do this we click in the cell corresponding to the variable "continent" and then click on the small grey box that appears to the right-hand side of the cell. This calls up the **Value Labels** window shown in Figure 1.15. For each value, type in the value and then its corresponding label in the fields indicated and click "Add." Repeat the process for the others and then click "OK" when finished.

The next column, headed Missing, is where you define any missing values for that variable. Missing values are simply used to indicate that data are missing for that variable. We saw an example of this when examining the bullying dataset earlier in this chapter (see Figure 1.9). In this case two missing values were actually defined. The first, as explained earlier was "-1" to indicate that the respondent had legitimately skipped that question because it did not apply to them. However, there was also a second missing value—"99"—that was used to indicate that the respondent did not answer a question they were expected to.

For missing values you can always just leave the cell empty, in which case SPSS will just insert a period—"."—to indicate nothing has been entered. However, the problem with just leaving it blank like this is that you will never know in the future whether it is blank because there are missing data or because of a mistake typing the data in. As such it is always better to choose a value that you can use to type in to indicate that data are missing. In choosing a value it is good practice to choose the same one for all variables and also one that is clearly different from the actual values used (i.e. "99" in this case) so that if it inadvertently gets included in any analysis it should stand out as peculiar. Some people specify a range of missing values (maybe "97," "98" and "99") to indicate different reasons for the data being missing.

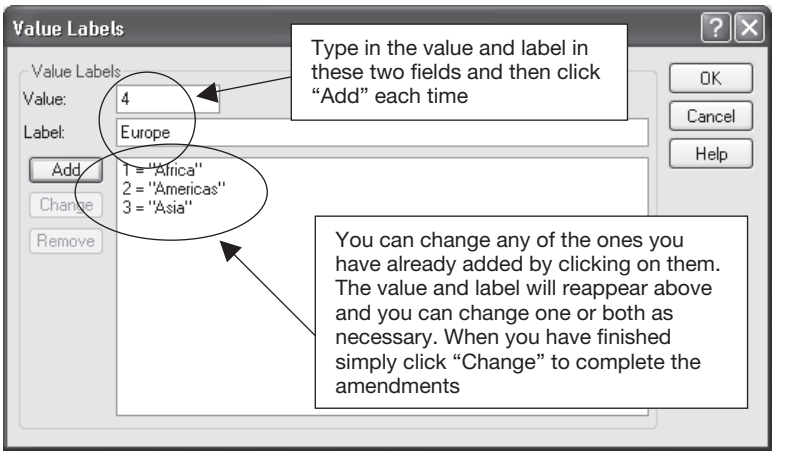


Figure 1.15 Value Labels window in SPSS

The key thing you need to do is to define these values as missing in SPSS so that they are counted as missing and not then included in any subsequent analysis. In this present case there are no missing values and so this is not necessary. However, to do this simply click in the relevant cell within the column headed "Missing" and then click on the grey box that appears to the right-hand side. This would call up the **Missing Values** window shown in Figure 1.16. As an example, if we were creating the bullying dataset then we would select "Discrete missing values" and then type in "-1" and "99" as shown and then click "OK."

The next two columns, headed "Columns" and "Align," simply define the width of the columns as they appear in **Data View** as well as the alignment of the data in each cell. As a default setting, SPSS tends to align numeric data to the right and string variables to the left and you will notice that since you have defined "contint" as a string variable then it has automatically aligned it to the left. Generally, the only time you may want to amend these from the default settings is in relation to string variables such as "contint" where you may wish to widen the column width so that all of the names are visible in **Data View**. To do this you simply need to click in the relevant cell and change the value to 20.

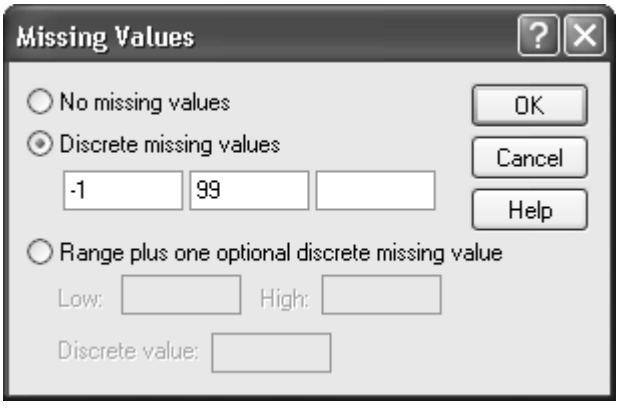


Figure 1.16 Missing Values window in SPSS

The final column, headed "Measure," specifies the type of variable we are dealing with. This is something that will be covered in detail in the next chapter. For now, just trust me that "id," "country" and "contint" are all nominal variables and the rest are all scale variables. To set these simply click on each of the cells, in turn, and then use the drop-down menu that will appear to the right of that cell as shown in Figure 1.17. It is very important to define these properly so that you can use the Chart Builder function to be described in Chapter 4. Once you have done all of this you have finished defining the variables and your **Variable View** should look as it does in Figure 1.17.

You are now ready to enter the data in **Data View**. To do this click on the Data View tab (bottom left of the main SPSS window) and then simply type in the data as carefully as possible from Table 1.1. Once completed it should appear as it does in Figure 1.18. When you have got to this stage the first thing you should do is to save it as the dataset: "international.sav" in the same way as was explained earlier for the bullying.sav dataset.

This is a dataset we will return to in later chapters so keep it safe. A copy of it is also downloadable from the companion website (although I have waited until now to tell you this to make sure you get some experience of creating a dataset!). However, it would be wrong to come this far without at least having a little look at the data. We will do two quick bits of analysis here. First, let's explore the variable school life expectancy ("lifeexpt"). To do this select **Analyze → Descriptive Statistics → Explore. . .** from the top menu. This will open up the **Explore** window shown in Figure 1.19. Click on "lifeexpt" in the list on the left-hand field and then click on the arrow button as shown in Figure 1.19 to place it in the "Dependent List:" field. In this particular case we will just look at the statistics for this variable so select this as an option as shown. Click "OK" to finish.

The main output that appears from this is shown in Output 1.1. As can be seen, the average (mean) school life expectancy for the 20 countries is 10.5 years. It can also be seen that the minimum school life expectancy among the countries in the sample is 5 and the maximum is 16. You will also see a number of other statistics presented. Don't worry—we will cover what most of these mean in the next chapter. For now it is just worth checking whether you gained the same results as shown in Output 1.1. If not then this is probably because you made an error typing in the data.

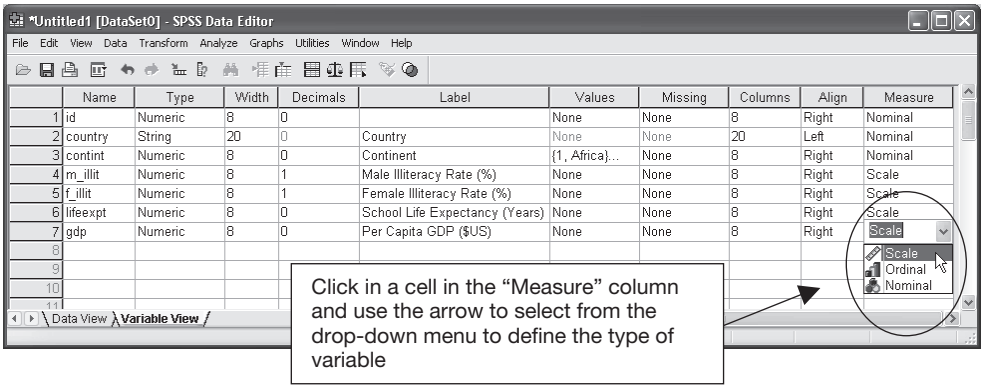
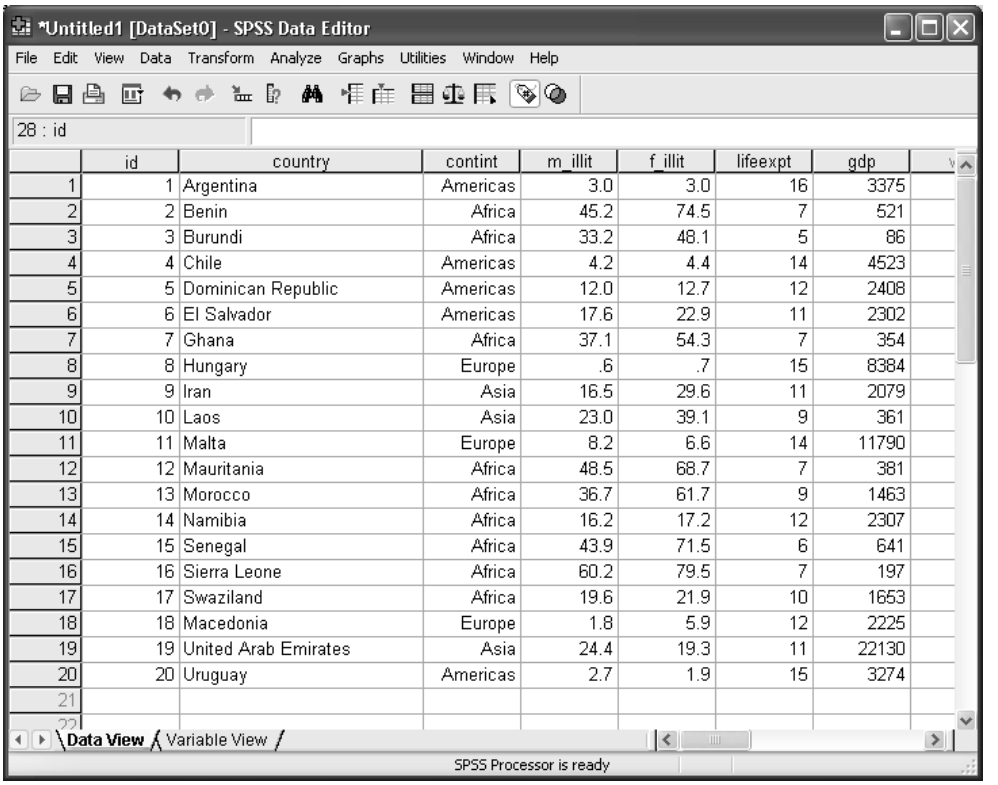


Figure 1.17 Completed Variable View for the international.sav dataset



	id	country	contint	m_illit	f_illit	lifeexpt	gdp
1	1	Argentina	Americas	3.0	3.0	16	3375
2	2	Benin	Africa	45.2	74.5	7	521
3	3	Burundi	Africa	33.2	48.1	5	86
4	4	Chile	Americas	4.2	4.4	14	4523
5	5	Dominican Republic	Americas	12.0	12.7	12	2408
6	6	El Salvador	Americas	17.6	22.9	11	2302
7	7	Ghana	Africa	37.1	54.3	7	354
8	8	Hungary	Europe	.6	.7	15	8384
9	9	Iran	Asia	16.5	29.6	11	2079
10	10	Laos	Asia	23.0	39.1	9	361
11	11	Malta	Europe	8.2	6.6	14	11790
12	12	Mauritania	Africa	48.5	68.7	7	381
13	13	Morocco	Africa	36.7	61.7	9	1463
14	14	Namibia	Africa	16.2	17.2	12	2307
15	15	Senegal	Africa	43.9	71.5	6	641
16	16	Sierra Leone	Africa	60.2	79.5	7	197
17	17	Swaziland	Africa	19.6	21.9	10	1653
18	18	Macedonia	Europe	1.8	5.9	12	2225
19	19	United Arab Emirates	Asia	24.4	19.3	11	22130
20	20	Uruguay	Americas	2.7	1.9	15	3274
21							

Figure 1.18 Data View for international.sav dataset (with Value Labels showing)

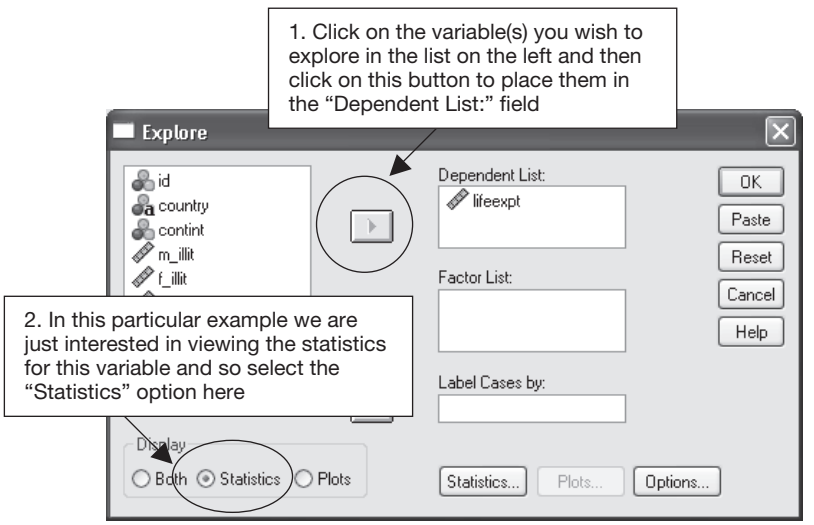


Figure 1.19 Explore window in SPSS

Output 1.1

Descriptives

			Statistic	Std. Error
School Life Expectancy (Years)	Mean		10.50	.738
	95% Confidence Interval for Mean	Lower Bound	8.96	
		Upper Bound	12.04	
	5% Trimmed Mean		10.50	
	Median		11.00	
	Variance		10.895	
	Std. Deviation		3.301	
	Minimum		5	
	Maximum		16	
	Range		11	
	Interquartile Range		7	
	Skewness		.015	.512
	Kurtosis		–1.113	.992

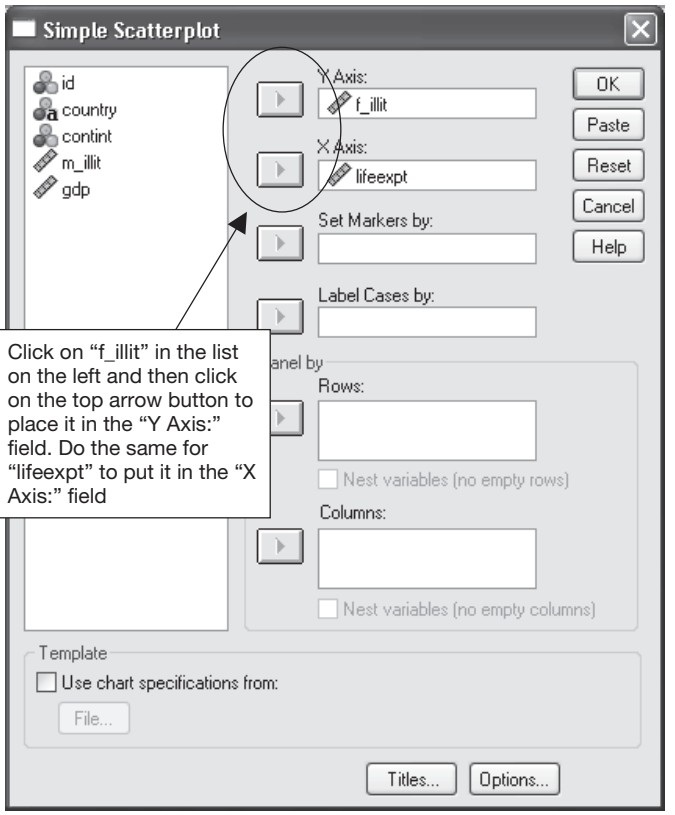


Figure 1.20 Simple Scatterplot window in SPSS

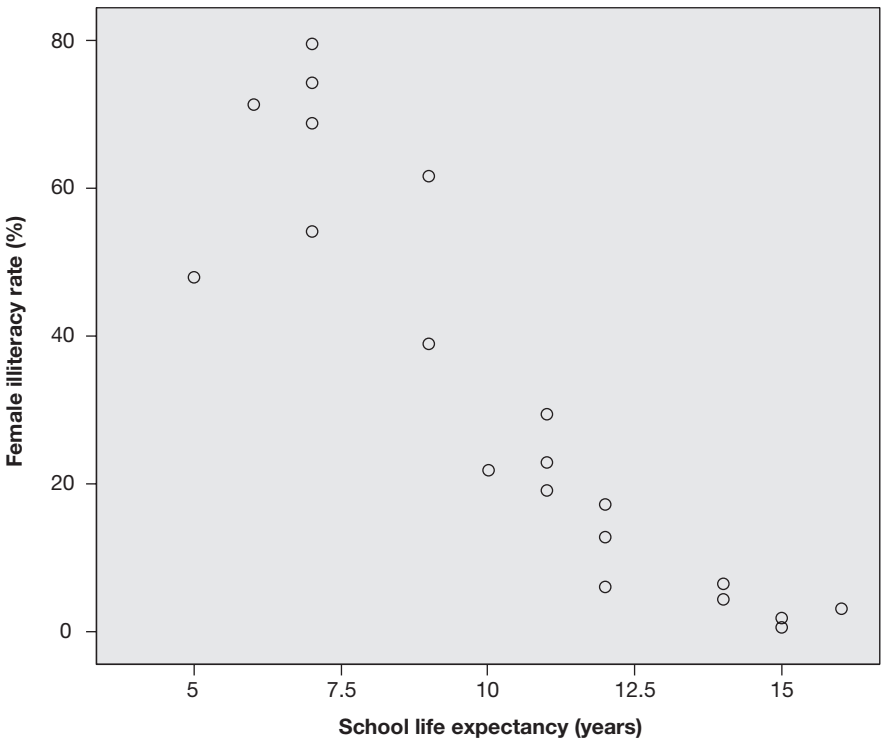


Figure 1.21 Scatterplot showing the relationship between school life expectancy and female adult illiteracy rates in 20 countries

Finally, let us have a look at what relationship exists between school life expectancy and levels of female illiteracy. What we would expect is that the higher the school life expectancy is for a country, the lower the levels of female illiteracy. We can see if this is the case using what is known as a scatterplot. To do this select **Graphs → Legacy Dialogues → Scatter/Dot. . .** from the main menu (for Version 14.0 and earlier you just need to select **Graphs → Scatter/Dot. . .**). This opens an initial Scatter/Dot window where you should select "Simple Scatter" and then click "Define." This opens the **Simple Scatterplot** window as shown in Figure 1.20. Click on the variable "f_illit" in the list on the left-hand side and then click on the top arrow button to place it into the "Y Axis:" field. Next, click on the variable "lifeexpt" in the list on the left-hand side and then click on the second arrow button to place it into the "X Axis:" field. Click "OK" to finish the procedure.

Figure 1.21 shows the resultant scatterplot that appears. As can be seen, there are 20 dots, each representing a specific country. There would certainly appear to be a relatively strong trend as was predicted with female illiteracy rates dropping as school life expectancy increases.

EXERCISE 1.3

Use the Explore function to generate summary statistics for male and female illiteracy rates (you should find that the mean rates are 22.73 and 32.175 respectively) and per capita GDP (mean is 3522.70). If you did not get these statistics go back and check whether you made an error in entering the data (and reflect on how easy it is to make mistakes doing this!). Try generating some more scatterplots between some of the other variables. For example, is there a relationship between male and female illiteracy rates? Also, is there a relationship between per capita GDP and male illiteracy? Check your methods and results on the companion website.

Conclusions

The main purpose of this chapter has been to give you an overview of the SPSS environment and the whole process that is involved in creating and analyzing datasets. I am hoping by this point that you have come to see that quantitative data analysis need not be difficult. If you have been following the exercises in this chapter you have, after all, calculated percentages as well as a range of summary statistics including: means, medians, standard deviations and interquartile ranges (even though you probably don't yet know what half of these mean—how clever is that?). In addition you have generated bar chartsand scatterplots. By the way, all of these charts can be adapted in terms of style (as we shall see in Chapter 4) and then copied and pasted directly into your research report or dissertation.

Having shown you that you can do quantitative data analysis and thus hopefully given you a little more confidence, it is now time to look in a little more detail at the meaning of some of the things you have just done.