Data entry

Chapter 2

Things to think about before entering data

Before you enter any data into SPSS, you need to think about what type of data you have and what you want to do with it. For example, you may have the details of a person's gender (male or female) and their preference for a certain product, or you may have collected a set of reaction times. Are you trying to find differences within your data or any relationships? Or do you just want to summarise and describe it?

What is the level of data? Is it *nominal* (where the data is labelled by the category it belongs to, such as car colour), *ordinal* (where the data defines an order, such as rating a group of vegetables on how much you like them), *interval* or *ratio* (where the data is measured on a scale of equal intervals, like a clock or a tape measure)? This needs to be considered before entering any data. If you have nominal data then you may need to use this as a 'grouping variable' when entering the data in SPSS. We can explain the use of a grouping variable by the following example. If you were going to calculate statistics by hand you would probably write the scores of males and females on a mathematics test in the following fashion:

Person number	Male	Person number	Female
1	84	6	49
2	79	7	86
3	64	8	85
4	73	9	74
5	75	10	81

Notice here that the variable 'gender' is measured on a nominal scale and the variable maths score is measured on an interval scale. Gender is our grouping variable because it defines our two groups. Later on we may wish to compare the maths scores between these two groups. We need to give the values of the grouping variable a label, so we will allocate the males an arbitrary number 1 and the females the number 2. When we do enter our data into SPSS it will look similar to the data shown below, which, as you can see, is a lot different from how you intuitively think about entering the data.

Person	Gender	Maths score
1	1	84
2	1	79
3	1	64
4	1	73
5	1	75
6	2	49
7	2	86
8	2	85
9	2	74
10	2	81

✓ Details on how to assign values to categories in SPSS are discussed later in this chapter.

A grouping variable is related to the research design. With an independent measures design (where the data in different samples come from different participants), you will have at least one grouping variable. The example above is an independent measures or unrelated design. However, if you have a repeated measures or related design (where the scores in each sample come from the same or matched participants) your data will be set up without any grouping variables and looks very much how you would set it out by hand. This is because the key rule for putting data into SPSS is that each row is a different case, so each row has all the data from a single participant. For example, two similar maths tests are undertaken, one in week one and the other in week two by each participant. The data would be entered as below.

Person	Maths score (week 1)	Maths score (week 2)
1	84	89
2	79	83
3	64	69
4	73	74
5	75	80
6	49	59
7	86	87
8	85	87
9	74	78
10	81	84

Notice that we do not use 'week' as a grouping variable here. This is because we have a repeated measures design and so want both scores from person 1 in row 1, person 2 in row 2 and so on. This is an important feature of inputting data into SPSS and we need to understand the link between grouping variables and research designs. Independent measures variables (such as gender in the first example) are used as grouping variables but repeated measures variables (such as week in the second example) are not.

Using SPSS for the first time

If you are using SPSS for the first time, you will be presented with the **What would you like to do?** window before being able to use the data editor screen. This window gives you the opportunity to run a self-help tutorial, type in data, run an existing query, create a new query using Database Wizard, open an existing data source or open an alternative type of file (an example of this could be some data that had been prepared in Microsoft Excel).


- Select **Type in Data**.
- Click on the **OK** button.

> ✓ You can tell SPSS not to show this dialogue box in the future by clicking in the box at the bottom of the window. Users of SPSS in public areas such as universities or users who share a computer may not see this window as a previous user may have already clicked in the box. If this is the case then the first active window you are presented with will be the SPSS **Data Editor**.

Data Editor

The SPSS **Data Editor** is where the data is entered in order to carry out the statistical analysis. On first view, the Data Editor looks very much like other spreadsheets. However, there are a number of differences. In SPSS the row numbers down the side of the data entry sheet refer to participants or cases. The columns are not lettered, as they would be in a spreadsheet but carry the label 'var'. This is the case until the variables have been defined.

If you look at the bottom left-hand corner of the screen you will notice two tabs, **Data View** and **Variable View**. The **Data View** tab is the current active sheet and this is where raw data is entered. However, it is advisable to define your variables first. To make the **Variable View** sheet active click on the **Variable View** tab as shown below.
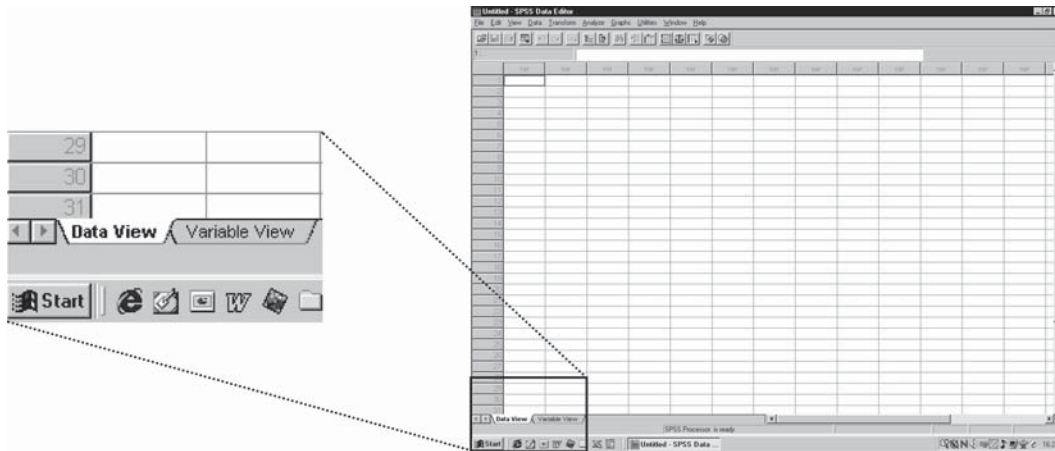
In the **Variable View** sheet the column headings have now changed from 'var' to the ones shown in the screen overleaf. The sheet is no longer a data entry sheet but the location where your variables are defined. You can switch between the **Data View** and **Variable View** sheets by clicking on the appropriate tab. An alternative way of accessing the **Variable View** sheet is to double click on 'var' in one of the columns within the **Data View** sheet. The latter way is useful if your variables are already defined and you wish to edit the one you have double clicked on as this method takes you directly to chosen variable.

✓ Many new users of SPSS get confused when switching between the **Data View** and **Variable View** sheets. This is because when in **Data View** the variables are in columns so the data should be entered and read moving down the sheet. When in **Variable View** the variables are in rows. Therefore, the variable information should be entered and read moving across the sheet.

Data View

Variable View
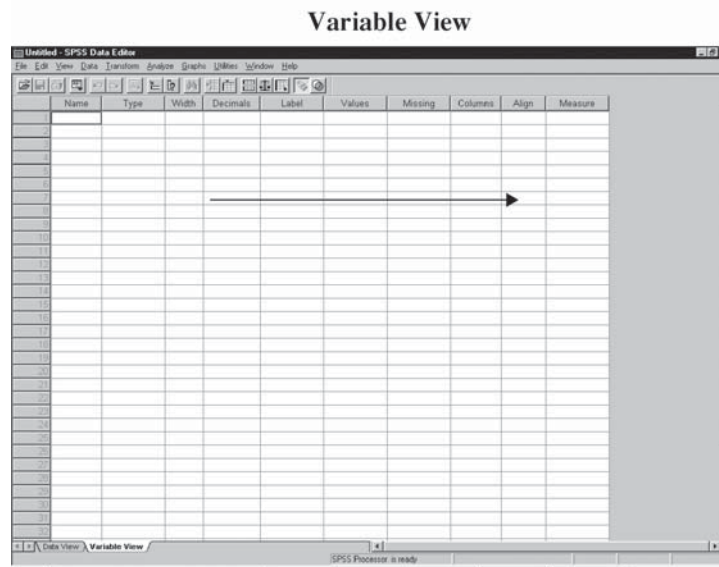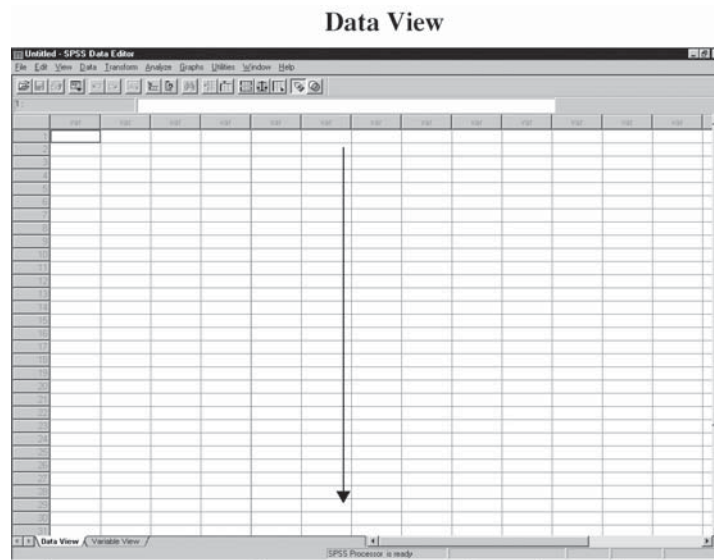
Now we are going to look at a scenario and consider how each variable should be set up.

Scenario

A teacher wanted to see whether there was a difference in the amount of time boys and girls spent on their homework. He asked ten boys and ten girls aged 9 years to make a note of the number of minutes they spent studying over a 2-week period.

Setting up variables
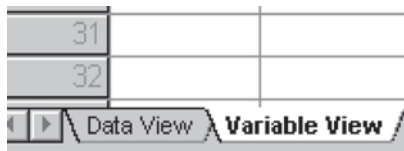
- Click on the **Variable View** tab in the data editor window.
- Click in the **Name** column on row **1**.
- Enter the name 'gender' and press the return key on your keyboard.

✓ You will now notice that SPSS has automatically filled in information in the cells for the remaining columns. This information will not necessarily be correct so you need to check each cell and add information or correct it if needed, as discussed below.

SPSS has a few rules about the name of your variable. For example, it cannot exceed eight characters. Therefore, sometimes it might be more appropriate to call your variable an acronym of the name you really wanted. Some other rules for variable names are that you cannot use blank spaces or special characters; the name must begin with a letter and cannot end with a full stop. Each variable name must be unique to that variable and you should try to avoid ending the name with an underscore.

Now we will briefly describe the remaining columns within the **Variable View**.

Within the **Type** column you need to identify what type of variable you have. If you click inside the **Type** cell in row 1 for gender, you will see that the default is **Numeric**. To change the variable type:
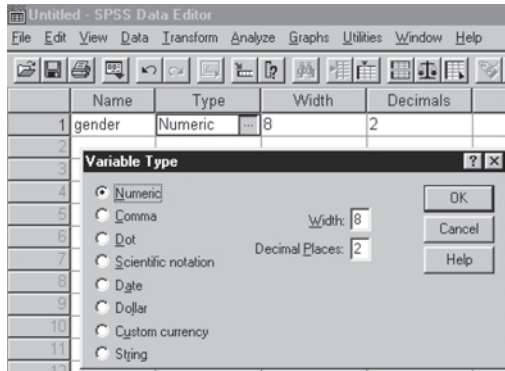
- Click on the grey box that will appear next to the word **Numeric**. You will notice a variety of options.
- If you have a variable that is a currency or a date then it is advisable to use one of these options. **String** should only be used if you want to identify a participant or case by name. When you have selected this option, you can change the number of characters to more than the default of eight if you wish.

✓ The **String** option will limit the type of numerical data analysis that can be carried out with that variable, so use it with caution. If you are in doubt, leaving the default option of **Numeric** will generally suffice for most types of data analysis. In the case of our scenario, we would leave the variable type as **Numeric**.

The **Width** column allows us to specify the total number of characters required for the column. The default setting is for eight characters to be displayed and input into each cell. There is no need to change this option unless you will be using more than eight characters.

The **Decimals** option allows you to specify the number of decimal places for your variable. For example, if you are using whole numbers then you may wish to change the decimals to 0. SPSS always defaults to two decimal places.

Many users of SPSS underestimate the importance of the **Label** cell. You may already have expressed frustration at only being able to use eight characters for the variable name. This is where you can give your variable a longer label with spaces, as this label will appear on output tables instead of the shortened variable name.

Also, if you have used acronyms or short titles when naming your variables, SPSS will show the full label assigned to that variable when the mouse pointer is held over the variable name in **Data View**.
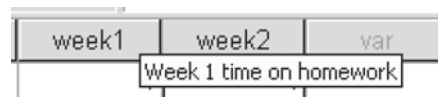
As stated earlier, if we have a grouping variable we need to assign each category of that variable to a number because SPSS will only recognise numerical data.
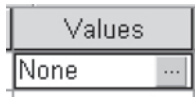
Values
None ...

- Click within the **Values** column along the row of gender and a grey box will become apparent.
- Click on the grey box, the **Value Labels** window appears.

In our scenario we have the grouping variable 'gender'. We have decided to call 'males' number 1 and 'females' number 2.

- Enter '1' into the **Value** box and then either press the tab key or click the mouse to arrive at the **Value Label** box. Type in 'male'.
- Click on the **Add** button and you will see **1.00 = "male"** appear in the box below. Then follow the same procedure, entering '2' in the **Value** box, for females.
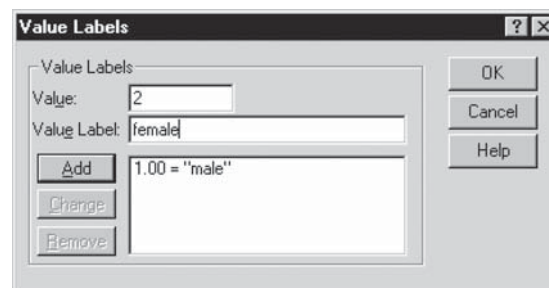

✓ When you come to enter your raw data you will need to enter the numeric values assigned to your two groups. It is therefore important to remember which way round you assigned these labels.

As default, SPSS considers that you have no missing values. However, within your data there may be various reasons why you have missing values. The **Missing** column enables you, under the **Discrete Missing Values** radio button, to assign up to three values to be counted as missing data. SPSS will then ignore that participant or case during the analysis. A good example of this might be that a participant did not want to answer a question rather than could not as this may be useful information for your research. It is also possible to identify a range of missing numbers using the **Range plus one optional discrete missing value** radio button.

✓ When assigning missing values always ensure that you choose a value outside your data range.

- If the 'gender' variable above did have missing values, the range is only between 1 and 2 so assigning the value of 99 as shown here would be fine.
- Once missing values have been assigned, click on **OK** to continue.
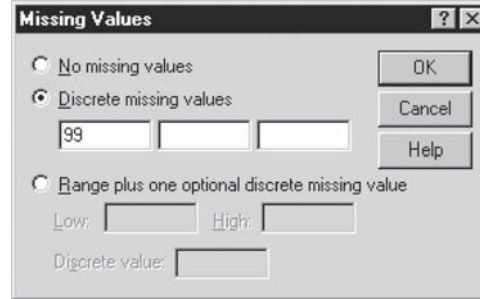
The **Columns** feature need only be used if you have more than eight characters in a column. For example, if you were using very large numbers or string variables with large names, the width of the column which is shown in the **Data View** sheet can be adjusted. Alternatively, you may wish to reduce the width of column if using small numbers to make a larger viewable area on your computer screen. If you decide to do this, you should make sure your variable name is still identifiable in the **Data View** sheet.

> ✓ If some of the characters are no longer viewable in the **Data View**, holding the mouse pointer over the variable name will show the full variable label, providing you have entered one as explained previously.

Align can be used if you wish to realign the way that the data is viewed in the SPSS **Data View** sheet. The options to choose from are left, right and centre. Using this option has no bearing on your statistical analysis but is cosmetic in terms of viewing your raw data.

The **Measure** column is where you identify which level of data for each variable is. In our example 'gender' is nominal data so **Nominal** should be selected from the drop-down list.
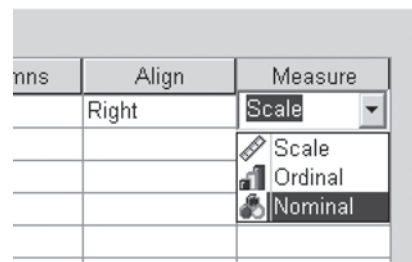
> ✓ SPSS groups interval and ratio data as **Scale**. This is because the same statistical tests are applied to both interval and ratio levels of data.

The **Variable View** sheet below is an example of what you should see if you have set up the variables for the homework scenario described above.

- Once the variables have been set up and labelled, click on the **Data View** tab to enter your raw data.
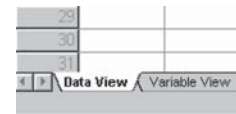
Entering data

In our scenario we obtained the homework times for ten males and ten females. We have decided arbitrarily to call each male number 1 and each female number 2. Therefore, we must enter in the first column a number 1 in the first 10 rows and a number 2 in the rows 11 to 20 as shown below.
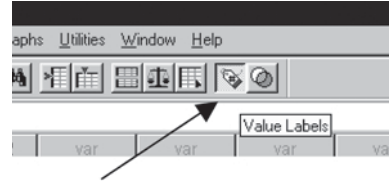

- So that we can see the labels that we have given to these values click on the **View** drop-down menu and select **Value Labels**.
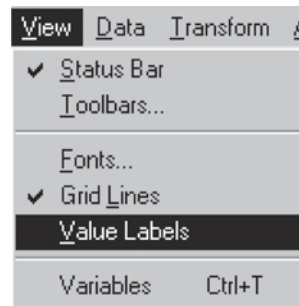
- Alternatively, click on the **Value Labels** short-cut icon.
- The labels we entered in the **Variable View** screen are now shown.

> ✓ To revert back to the numerical values either click the short-cut icon or go to the **View** drop-down menu and select **Value Labels** again.

After completing our first column we can then go on to enter the homework times for weeks 1 and 2.

Grouping variable

You must assign a separate name for each condition when setting up the SPSS Data Editor

Independent measures

	gender	week1	week2
1	male	45.00	51.00
2	male	56.00	53.00
3	male	67.00	65.00
4	male	123.00	105.00
5	male	89.00	98.00
6	male	99.00	96.00
7	male	115.00	122.00
8	male	136.00	125.00
9	male	102.00	100.00
10	male	39.00	39.00
11	female	124.00	136.00
12	female	99.00	87.00
13	female	135.00	115.00
14	female	145.00	156.00
15	female	89.00	91.00
16	female	97.00	100.00
17	female	112.00	104.00
18	female	85.00	79.00
19	female	156.00	142.00
20	female	130.00	119.00

Saving your data

Saving in SPSS is similar to other applications. When you save your data, SPSS saves the information on both the **Data Editor** and **Variable View**.
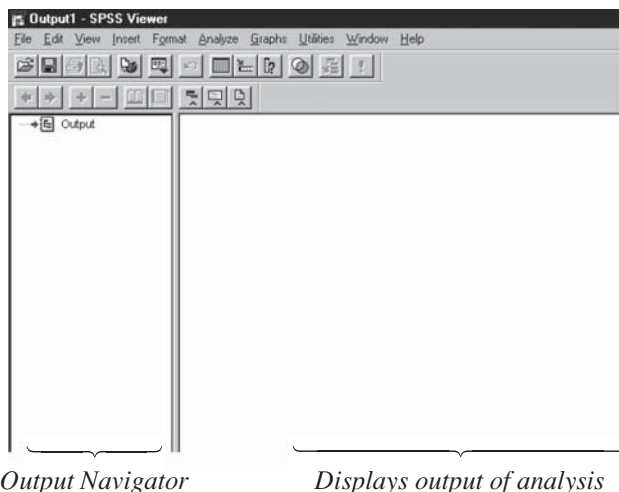
- Go to the **File** drop-down menu and select **Save As**.
- Locate where you want to save your data; we have chosen the $3\frac{1}{2}$ disk.
- Name the file; we have called ours 'homework data'.
- You will notice that the **Save as type** is called a **.sav** file.
- Then click on **Save**.
- At the top left-hand side of the **Data View** screen you will notice the title of your file.
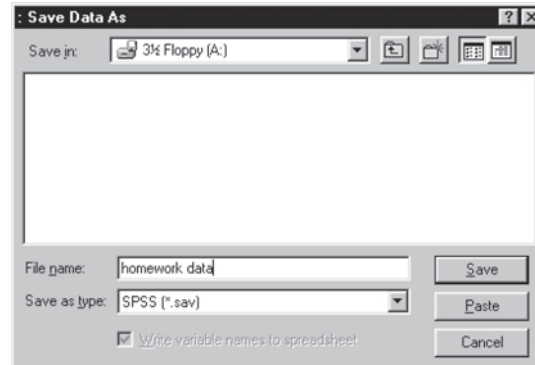
✓ It is a good idea to save your work regularly. As with other applications you can now just click on the save icon.

The SPSS output file

When you undertake any analysis, SPSS switches to the **SPSS viewer** screen where results tables and graphs are output. So far no tests have been run to produce any output, so the screen shown below is only for illustration.

Output Navigator   *Displays output of analysis*

✓ The **SPSS Viewer** file has some similar drop-down menus to the **Data Editor** but some functions can only be performed in the **Data Editor**.

The Output Navigator displays the output in an outline view. You are able to hide, collapse and move items within this screen.

✓   The **SPSS Viewer** and the **Data Editor** screens contain separate files. There-fore, when using the **Save** command each of these must be saved separately.


Saving your output

- Go to the **File** drop-down menu and select **Save As**.
- Locate where you want to save your data, we have chosen the 3½ disk.
- Name your file, we have called ours 'homework results'.
- You will notice that the output **Save as type** is called a **.spo** file.
- Then click on **Save**.
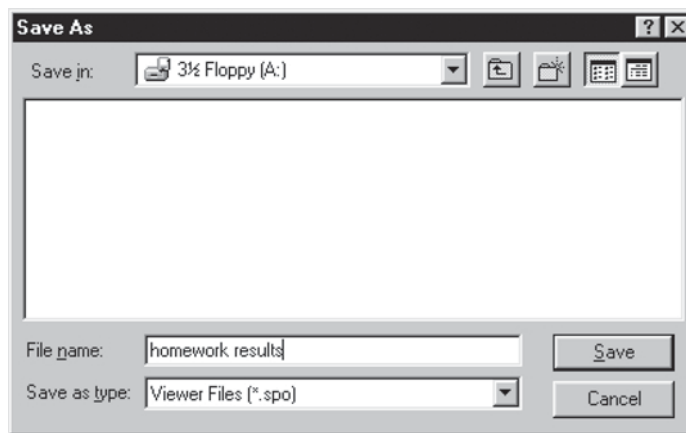
At the top left-hand side of the **SPSS Viewer** screen you will notice that the title of your file changes from Output1 to homework results.

FAQ

Whenever I try to enter a variable name I get the following alert message, what am I doing wrong?
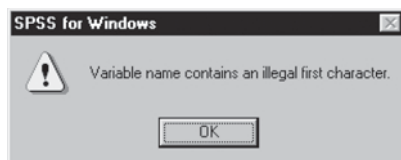
*An illegal first character has been entered. SPSS will not let you enter a number, hyphen or other non-letter-based character as the first character of a variable name. SPSS will also not allow you to enter certain characters as part of the variable name, such as * ! ?*

I have entered my variable name, but the grey bit with 'var' now has 'var0000', I think I have done something wrong.

Yes, you have done something wrong but it only takes a minute to correct it. All you have done is typed in your variable name in the Data View screen rather than in the Variable View screen. One way to correct this is to click on the 'var0000'. This will highlight the column black and then press delete on the keyboard, this will delete the whole column. Now, switch to Variable View and type the variable name in the correct place.

I have tried to assign value labels but this message has appeared, what should I do next?
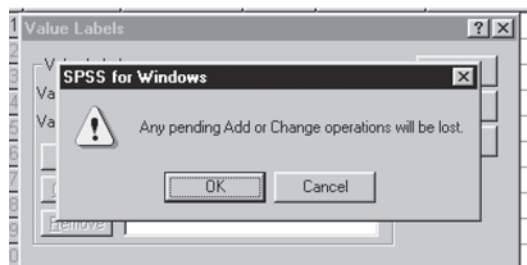

You have added your numerical value and the label but selected OK instead of selecting Add and then OK. Press the Cancel command and then press Add and OK.

I have saved both of my data and output files as 'homework'. When I want to open them again, how will I know which is which?

*The data and the output file names have a different ending. With the Data Editor the file name ends with .sav and the **SPSS viewer** the file name ends with .spo. Also if you are opening the files from your disk you can see a difference in the*

icons. The data file looks like a miniature spreadsheet, whereas the output file looks like a small bar chart.
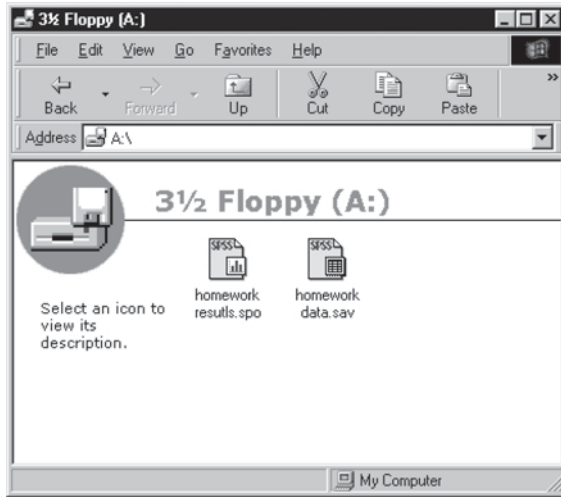
I have entered my data as shown below, and know that I want to find differences between males and females but have no idea how I am going to do it.

	males	females
1	25.00	29.00
2	23.00	28.00
3	31.00	21.00
4	41.00	26.00
5	27.00	39.00

You have entered your data incorrectly, remember that gender is nominal data and therefore needs a grouping variable. Refer to the beginning of this chapter on how to enter data.

I have two grouping variables: gender and eye colour. I have used the values 1 and 2 for my gender variable, do I need to assign different numbers for the different eye colours; I have three colours?

Not necessarily, whichever numbers you assign are unique to that individual variable. You can use the numbers 1, 2 and 3 for the eye colour or completely different numbers. As long as you give each number a label in each variable's Value Label section you will be fine.