

<i>Chapter 1</i>	Introduction	1
<b>Chapter 2</b>	<b>Data entry in SPSS</b>	<b>2</b>
<i>Chapter 3</i>	Exploring data in SPSS	3
<i>Chapter 4</i>	Data handling	4
<i>Chapter 5</i>	Tests of difference for two sample designs	5
<i>Chapter 6</i>	Tests of correlation	6
<i>Chapter 7</i>	Tests for nominal data	7
<i>Chapter 8</i>	Analysis of variance	8
<i>Chapter 9</i>	Bivariate and multiple regression	9
<i>Chapter 10</i>	Analysis of covariance and multivariate analysis of variance	10
<i>Chapter 11</i>	Discriminant analysis and logistic regression	11
<i>Chapter 12</i>	Factor analysis, and reliability and dimensionality of scales	12
<i>Chapter 13</i>	Beyond the basics	13
	Glossary	Gloss
	References	Refs
	Appendix	App
	Index	Index

# Data entry in SPSS

In this chapter

- The Data Editor window
- Defining a variable in SPSS
- Entering data
- Saving a data file
- Opening a data file
- Data entry exercises
- Answers to data entry exercises



## **SPSS for Psychologists online**

Visit [www.palgrave.com/psychology/brace](http://www.palgrave.com/psychology/brace) for sample exercises and data sets

## SECTION 1: THE DATA EDITOR WINDOW

### What is the Data Editor window?

- The Data Editor is the active window when you start SPSS. It is used to record all the data we want to analyse.
- It has two views: the Variable View and the Data View.
- The Variable View allows us to name each column in the Data table and specify what sort of values the column will contain.
- The Data View contains a table with a large number of cells in rows and columns. The table can be very large with only a small part of it visible, in which case use the scroll bars on the edges of the window to move round the table.
- In psychology, we almost always enter data in the same way. Normally each row represents an individual participant and each column represents a variable.

### The arrangement of the data in the Data Editor window

The precise way that the data are entered in the Data Editor window is critical and will depend, in part, on the details of your study. If you are entering data from an experiment, then you need to consider the design employed. In an independent groups design, each participant will provide one measure of performance. In addition, you will need to indicate which of your experimental groups each participant was assigned to. Thus, the most basic independent groups design will require that you use one column of your data table to record which group your participant was in, and a second column to record that participant's score. By comparison, in a repeated measures design each participant's performance will be assessed (at least) twice. Thus you will have two measures of performance, one from each condition. You will therefore need to use two columns of your data table to record these two performance levels.



In SPSS, the word 'variable' means a column in the data table; it does not have the same meaning as it does in experimental design. For example, in a repeated measures design there is one dependent variable that is recorded across two columns of the data table.

## SECTION 2: DEFINING A VARIABLE IN SPSS

Our first job is to set up SPSS with important information about each of our variables. This process of defining the variables is described here.

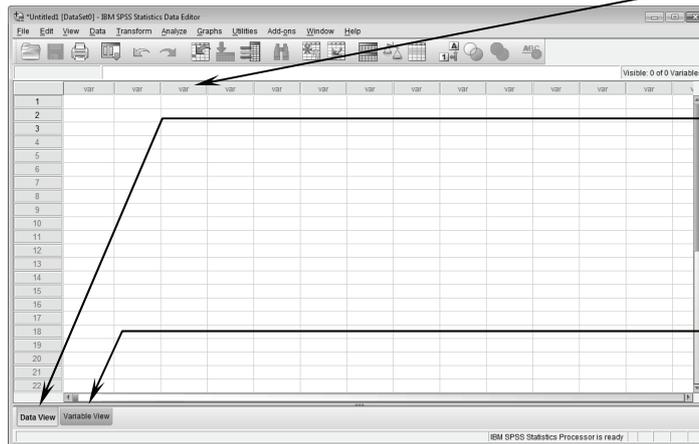
### The Data View and Variable View

If you look at the bottom left hand corner of the Data Editor window you will notice two ‘tabs’. One tab is labelled ‘Data View’ and the other is labelled ‘Variable View’. You can think of these as the index tabs for two different pages of information. When you first enter the Data Editor window, you will probably find the Variable View tab selected. If you click on the Data View tab, you will be presented with the empty data table. If you click on the Variable View tab, a different screen of information will be displayed. These two views are illustrated on the next page.

The Data View is the screen you will use when entering your data into SPSS. At present this view shows an empty data table in which each of the variables (columns) is labelled ‘var’. Before you can type your data into this data table you must set up the variables so they are ready to receive your data. SPSS needs to know the name of each of your variables; these names will be inserted at the top of the columns of the data table. In addition, you need to give SPSS other important information about each of your variables. This process of defining the variables is undertaken in the Variable View. If you click on the Variable View tab you will notice that in this view the columns are headed **Name**, **Type**, **Width**, **Decimals** etc. In the Variable View of the data table the variables are arranged down the side of the table and each column gives information about a variable. For example, in the column headed **Name** we are going to provide the name of each variable, in the **Type** column we are going to tell SPSS what type of variable this is, and so on.



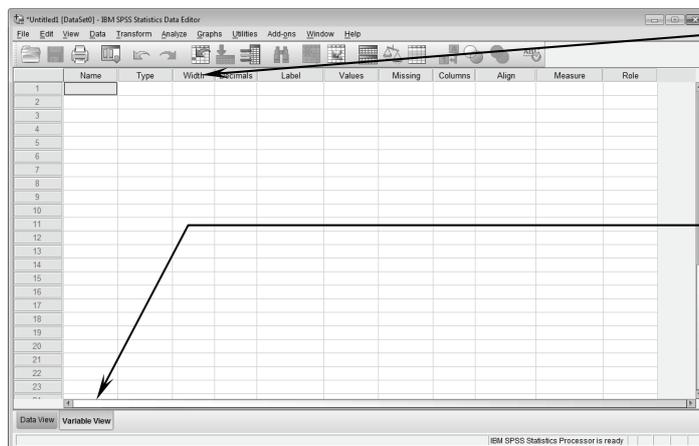
In the Data View of the Data Editor window each row of the data table represents data from one case and each column contains data from one variable. However, in the Variable View the columns and rows are used differently. In this view each row gives information about one variable. Don't let this confuse you – remember once you have set up all your variables and are ready to enter your data, you will work in the Data View where a row is a case (usually a participant) and a column is a variable.



This is the Data View of the Data Editor window.

Note that the Data View tab is highlighted. This tells you that you are looking at the Data View.

Click on the Variable View tab to change to the variable view shown below.



This is the Variable View of the Data Editor window.

Note that the Variable View tab is highlighted. This tells you that you are looking at the Variable View.

Henceforth, when we refer to ‘the Data Editor window’ without specifying which view, we will be referring to the Data View.

### Setting up your variables

If you are not already in the Variable View of the Data Editor, click on the Variable View tab now. We will now use this view to set up each of the variables we need.



An alternative way to switch from the Data View to the Variable View is to double click on the grey header (which will probably be labelled ‘var’) at the top of the column you wish to define. This will take you to the appropriate row of the Variable View.

## Variable name

The first thing we need to do is to give the variable a meaningful name. Type the name of your first variable into the first row of the **Name** column. You should choose a variable name that makes sense to you and which you are not likely to forget. Students often use the variable name ‘score’. This is not a good choice as it tells us almost nothing about the variable. Examples of more useful variable names might include ‘memscore’ (for participants’ scores in a memory experiment), ‘introver’ (a participant’s introversion score), ‘sex’ or ‘famfaces’ (the number of famous faces named by a participant). Your variable name can be any length, but we suggest you keep it fairly short so it is easy to read. Two points about variable names: they must start with a letter of the alphabet (i.e. not a number); and they cannot contain spaces or some special characters such as colons, semicolons, hyphens or commas (full stops, the @, #, \$ and \_ characters are allowed). If you enter an invalid variable name SPSS will warn you when you try to move from the **Name** column.



The underscore character ( \_ ) can be used in place of spaces in variable names. For example the name ‘Q1\_I’ might be used for the scores from Question 1 Part 1.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Sex										
2											
3											
4											

We have given the first variable the variable name ‘Sex’ because we are going to use this variable to code the sex of our participants.

Once you have entered the variable name, use either the mouse (point and click) or the tab key to move to the next column of the table. As you move the cursor, several of the other columns of the table will be filled with either words or numbers. These are the default settings for the variable ‘sex’. You can leave these settings as they are, or you can change some or all of them before moving on to define your next variable. Below we explain each of the settings and how to adjust them.

## Variable type

The second column in the Variable View table is headed **Type**. SPSS can handle variables of several different types. For example, variables can be numeric (containing numbers) or string (containing letters) or even dates. The **Type** column is used to indicate what type each variable is. The **Type** will now be set to **Numeric** (unless the default settings have been changed on your copy of SPSS).

If you want to change the variable type, click on the word **Numeric**, and then click on the button that appears in the cell. This will call up the **Variable Type** dialogue box (see below).

The screenshot shows the IBM SPSS Statistics Data Editor with the Variable Type dialog box open for the 'Sex' variable. The dialog box has the following options and settings:

- Numeric**
- Comma
- Dot
- Scientific notation
- Date
- Dollar
- Custom currency
- String
- Restricted Numeric (integer with leading zeros)

Settings for the Numeric type:

- Width: 8
- Decimal Places: 2

Buttons: OK, Cancel, Help

Information: The Numeric type honors the digit grouping setting, while the Restricted Numeric never uses digit grouping.

Annotations:

- Click in the cell and then click on this button to call up the **Variable Type** dialogue box (see below).
- Select the variable type you want from the list. The default is **Numeric** (as shown here).
- These values affect only the way the data are displayed (leave them as they are).
- Click the **OK** button to close this dialogue box.

We strongly recommend that, until you are an experienced user, you only use numeric variables. It is very easy to use numbers to represent categories and this will save you trouble later (e.g. you can use the numbers 1 and 2 rather than 'm' and 'f' to record the sex of your participants). Until you are a much more experienced SPSS user you are unlikely to need to use any of the other variable types.

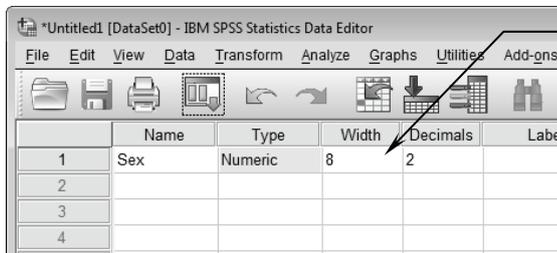


If at all possible avoid using string variables in SPSS – if you ignore this advice you will regret it later!

### Variable width and decimal places

As we saw above, the **Variable Type** dialogue box allows you to set the **Width** and **Decimal Places** of the variable (see above). Alternatively, these settings can be changed in the third and fourth columns of the Variable View (see below).

These settings adjust the number of characters before and after the decimal place used to display the variable in the Data Editor and Output Viewer windows. These settings do not affect the way the value is stored or the number of decimal places used in statistical calculations. Changing decimal places, however, does affect the number of decimal places shown in SPSS output. With numeric data the default settings are for a total **Width** of 8 with 2 **Decimal Places** (e.g. 12345.78). If you attempt to input a data value that will not fit into the width, then SPSS will round it in order to display the value. However, the value you entered is stored by SPSS and used in all calculations. One effect of this is that unless you set **Decimal Places** to zero, all values, even integers (whole numbers without decimal places) will be displayed with 2 decimal places. Thus if you enter a value of '2' in the Data Editor window, SPSS will display '2.00'. This might look a little untidy, but is of little consequence and it is probably not worth altering these settings to stop this happening.



	Name	Type	Width	Decimals	Label
1	Sex	Numeric	8	2	
2					
3					
4					

You can change the variable **Width** and number of **Decimals** places by changing these values. However, this only alters the look of the table.



You can probably leave the variable **Type**, **Width** and **Decimals** settings at their default values.

### Variable label

The fifth column in the Variable View table is headed **Label**. This column is used to enter a variable label.

A variable label is simply a phrase that is associated with the variable name and which helps you to remember what data this variable contains. If you have called a variable something like 'Sex', then you probably do not need to be reminded about what it is describing. If, however, you have a large number of variables, then variable labels can be very useful. For example, if you are entering the data from a questionnaire, you might have a variable named 'q3relbef'. In this case a variable label might be invaluable, as it could remind you that this variable coded the responses to question 3 on your questionnaire which asked about religious belief. You can type in any phrase using any characters that you like, but it is best

to keep it fairly short. SPSS will not try to interpret this label; it will simply insert it into the output next to the appropriate variable name when you perform any analysis. It is also worth remembering that when you have to select variables for inclusion in an analysis, SPSS will list them by these variable labels, not the names. This is another reason to keep the labels short and meaningful

To add a variable label, type it in to the column **Label**.



Variable labels are included in the output produced by SPSS. Although they are not essential, they act as a reminder about the variables and can be very helpful when you are interpreting the output. We recommend you take the time to use them whenever appropriate.

### Value labels

A value label is a label assigned to a particular value of a variable. You are most likely to use value labels for nominal or categorical variables. For example, we might want to use value labels to remind ourselves that, when entering values for the religion of our respondents, we used the codes: 1 = Buddhist; 2 = Christian; 3 = Hindu; 4 = Muslim; 5 = Other; 0 = Atheist.

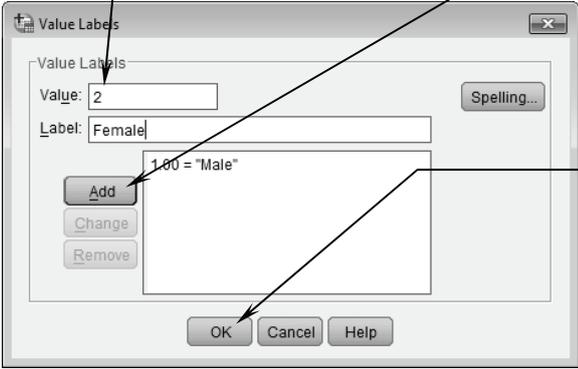
A second use for value labels is with a grouping or independent variable. For example, you might want to compare the reaction time of participants who were tested under one of several different doses of alcohol. You could use a value label to remind yourself that group 1 received no alcohol, group 2 received 1 unit of alcohol and group 3 received 2 units. Value labels will be inserted into the SPSS output to remind you what these values mean.

Value labels are entered using the **Values** column of the Variable View table. At present this column will probably contain the word **None**. Click the mouse on this cell, or use the tab key to move to this cell. As you do so a button will appear at the right hand side of the cell. Click on this button to call up the **Value Labels** dialogue box (see below).

Note that we have used a variable label – not really necessary in this case but usually useful.

	Name	Type	Width	Decimals	Label	Values	Missing	Column
1	Sex	Numeric	8	2	Sex of Participant	None	None	8
2								

To add value labels, click in the **Values** cell, then click on this button. This will call up the **Value Labels** dialogue box (see next page).



1. If you are using value labels, enter the value into the **Value** box – then enter the label for this value into the **Label** box.

2. Then click on the **Add** button to add this value label to the list of labels for this variable. Repeat these steps to add additional values and labels.

3. When you have added all the values and labels for the variable, click on the **OK** button to close the dialogue box and return to the Variable View table.



Value labels can be a great help when interpreting SPSS printout. Although they are not essential we recommend that you use them when appropriate. It would not be appropriate to add value labels to some variables. For example, you would not want to add a label to every possible value of a continuous variable such as reaction time. A good rule-of-thumb is that you should add value labels to all nominal variables and should consider adding them to ordinal variables. They are unlikely to be needed for interval or ratio variables.

### Missing values

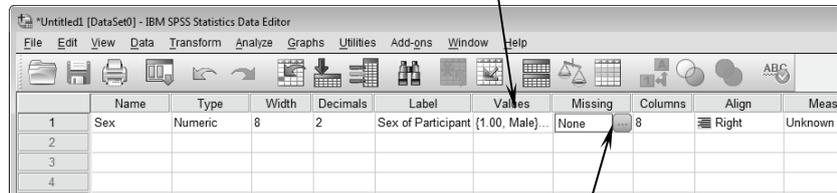
Sometimes you will not have a complete set of data. For example, some participants might decline to tell you their religion or their age, or you might lose or be unable to collect data from some participants (e.g. as the result of equipment failure). These gaps in the data table are known as missing values.

When we have a missing value we need to be able to tell SPSS that we do not have valid data for this participant on this variable. We do this by choosing a value that cannot normally occur for this variable. In the religion example above, we might choose to code religion as 9 when the participant does not state their religion. Thus, 9 is the missing value for the variable religion. The missing value can be different for each variable, but does not have to be. The important thing is that this value cannot normally occur for this variable. For age you could use 99 (unless you are testing very old people). Alternatively, you can use a negative number (e.g. -9) assuming that negative values cannot occur for the variable that you have measured.

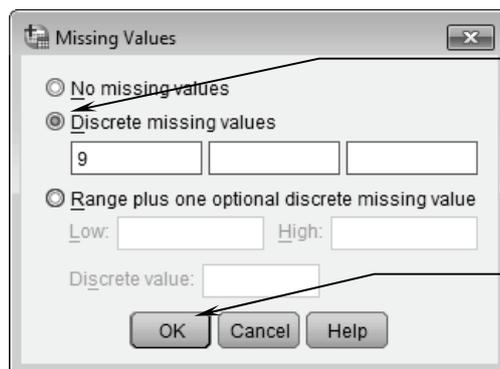
Before you specify any missing values, the cell in the Missing column of the Variable View table will contain the word None. To specify a missing value click in the Missing column of the Variable View table. A button will appear at the

right hand end of the cell. Click on this button to call up the **Missing Values** dialogue box (see below)

Note we can see the value labels we added in the previous step.



Click on the button in the **Missing** cell to call up the **Missing Values** dialogue box (see below).



1. To include up to three different missing values click on this circle (so that it becomes filled) then enter your missing value(s) in the box(es).

2. Click on the **OK** button to close the dialogue box and return to the Variable View table.

SPSS allows you to specify the missing values in several ways:

1. **No missing values:** This is the default setting for this dialogue box. If this option is selected, SPSS will treat all values for this variable as valid.
2. **Discrete missing values:** This option allows you to enter up to three discrete values. For example, 7, 9 and 11 could all be set as missing values by selecting this option and entering the values in the three boxes. If you have only one missing value, enter it into the first of the three boxes (as we've done above).
3. **Range plus one optional discrete missing value:** This option allows you to indicate that a range of values is being used as missing values. For example, selecting this option and entering the values 7 and 11 in the **Low** and **High** value boxes would instruct SPSS to treat the values 7, 8, 9, 10 and 11 as missing values. If, in addition to this range of values, the value 0 was typed

into the **Discrete value** box, then SPSS would treat the values 7, 8, 9, 10, 11 and 0 as missing.

In practice we rarely need more than one missing value for a variable (occasionally you might want more than one – for example you might wish to distinguish between an unanswered question and an illegible answer as both are missing values). You will therefore almost always want to enter your missing value into the first of the **Discrete missing values** boxes. To do this, simply click on the circle next to the words **Discrete missing values** and then enter your missing value into the first of the three boxes. Now click on the **OK** button to return to the Variable View table.



The **Missing Values** dialogue box does not allow you to label the missing values. Once you have entered them, however, you can label them in the **Value Labels** dialogue box if you wish. For example you could add labels to show that 9 = unanswered; 10 = illegible.

### Column format

The next column of the Variable View table is labelled **Columns**. This entry in the table is used to specify the width of the column that the variable occupies in the Data View table of the Data Editor window. You can leave this value at its default setting unless you want to change the appearance of the Data View table. You may, for example, want to fit more columns onto the screen in order to see more variables without having to scroll. In this case you could reduce the width of each column. To adjust the settings, click on the cell and then use the up and down buttons that will appear at the right-hand end of the cell to adjust the value. You can look at the effect of the change you have made by switching to the Data View.

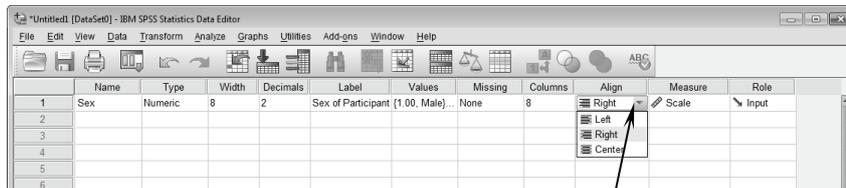


You can also change the width of a column by dragging it with the mouse. Switch to the Data View and then move the mouse to the top row of the table and hover over the border between two columns. The mouse pointer will change to a double headed arrow. You can now click and drag to adjust the width of the column. The changes you make here will be reflected in the Variable View settings.

### Column alignment

The column of the Variable View labelled **Align** allows you to specify the alignment of the text within the cells of the Data View table. This setting has no effect on the operation of SPSS and only changes the appearance of the Data View table. The default setting is right alignment in which the decimal points of the values in the column are lined up. In left alignment the values are flush to the left-hand end of the cell. In centre alignment the values are centred in the cell (and thus the decimal points will not necessarily line up).

If you wish to change the Column Alignment, click in the **Align** cell and then click on the menu button that will appear in the cell and select the required alignment from the drop-down list (see below).

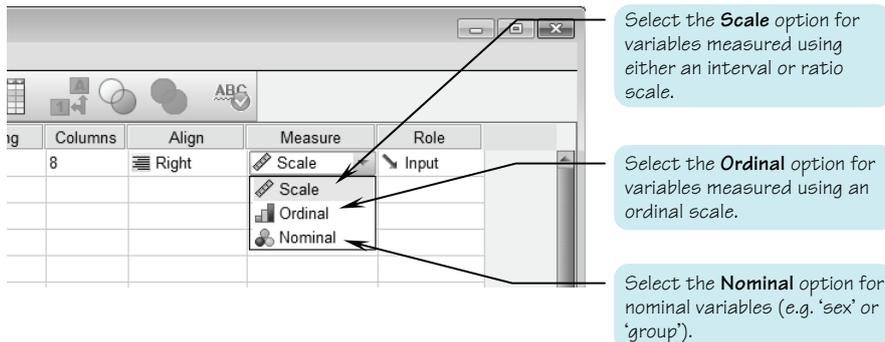


To change the column alignment, click on this button and select from the drop-down list.

## Measure

The next column of the Variable View table is labelled **Measure**. This column is used to specify the level of measurement for the variable. SPSS offers three options: **Nominal**, **Ordinal** and **Scale**. Psychologists usually distinguish four levels of measurement: nominal, ordinal, interval and ratio (see Chapter 1, Section 1). SPSS does not distinguish between interval and ratio data and uses the term **Scale** to describe a variable measured using either of these levels of measurement.

To set the measurement option, click in the **Measure** cell of the Variable View table and then click on the button that appears in the cell and select from the drop-down list (see below). The relevant icons will appear in the SPSS dialogue boxes as a reminder of the level of measurement of this variable.



Select the **Scale** option for variables measured using either an interval or ratio scale.

Select the **Ordinal** option for variables measured using an ordinal scale.

Select the **Nominal** option for nominal variables (e.g. 'sex' or 'group').

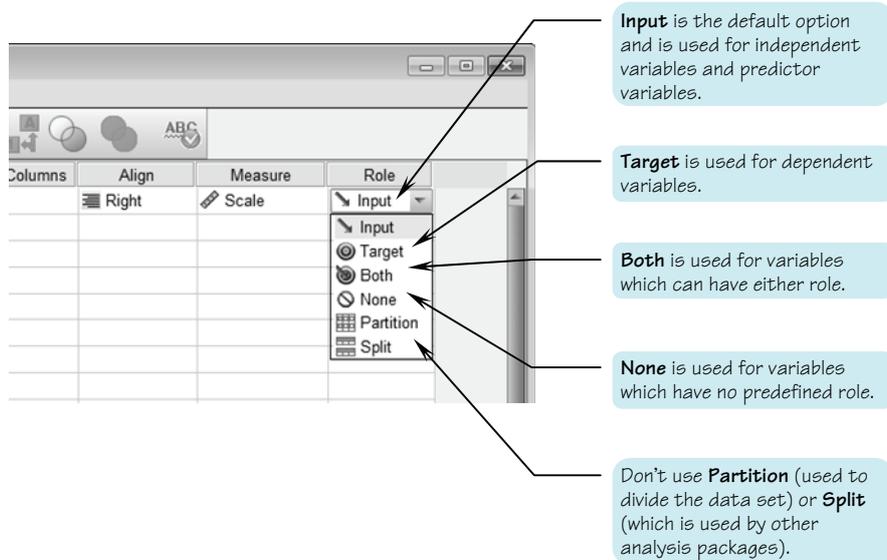


If you open a data file created using an earlier version of SPSS, the Measure will be set for you – variables with value labels will be set as **Nominal**, while variables with only a small number of values will be set as **Ordinal**. All other variables will be set as **Scale**.

## Role

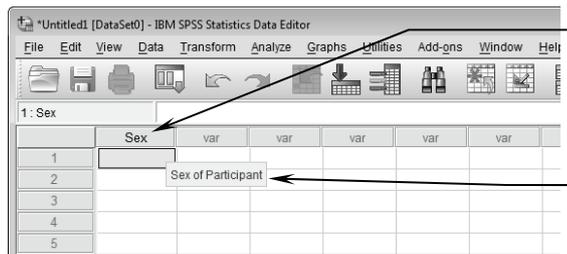
The final column of the Variable View table is called **Role**. This is a recent addition to SPSS, and is intended to help users who are undertaking complex analyses. The idea is that you can identify a group of variables as having a particular role in your analysis. For example you might have several variables which are going to be used as dependent variables and others which will be independent variables. If you indicate this through the **Role** setting, these variables will be automatically assigned when you come to undertake some kinds of analyses. The four relevant **Role** options are **Input** (used for independent variables), **Target** (used for dependent variables), **Both** (used for variables which may take on either role) and **None** (for variables with no role set).

In practice this is not likely to be useful for the novice user and, we recommend that you leave **Role** at its default setting of **Input**.



Once you have completed the definition of your first variable, switch to the Data View by clicking on the Data View tab at the bottom left-hand corner of the table. You will now see the name of your new variable appear at the top of the appropriate column of the Data Editor window (see below). If you changed the column width and/or alignment you will see the effect of these changes.

Now switch back to the Variable View and repeat this process for each of the variables required for your data file.



The new variable name appears at the top of the column. This column is now ready to accept data.

Note that if you move the mouse pointer over the variable name this pop-up displays the variable label.



Remember, for most variables you can accept the default settings. In practice all you need to do is to enter a variable name, set the level of measurement and, if appropriate, add variable labels and value labels.

### Copying variable settings

It is easy to copy the settings from one variable and ‘paste’ these on to one or more new variables.

Suppose, for example, that you have administered a questionnaire that contains 20 items. Each item consists of a printed statement to which the participant is asked to respond by choosing from one of several options such as ‘Strongly Disagree’, ‘Disagree’, ‘Neither Agree or Disagree’, ‘Agree’, and ‘Strongly Agree’. In our SPSS data table, each question will be represented by a variable, which we might call Q1, Q2 etc. For each of these variables it would be useful to enter the value labels 1 = ‘Strongly Disagree’, 2 = ‘Disagree’ etc. This would be rather time consuming. However, if we enter these value labels for the first variable, we can then move the cursor to the **Values** cell of the Variable View table and select **C**opy from the **E**dit menu (or right click and select **C**opy). If we now click in the cell (or select the range of cells) we want to copy these labels to, and select **P**aste from the **E**dit menu (or right click and select **P**aste), the value labels will be copied to all the selected cells.

## SECTION 3: ENTERING DATA

### A first data entry exercise

As a data entry exercise, we will enter the data from a very simple study in which we have recorded the sex (coded as 1 = male, 2 = female), the age and the memory score (number of words recalled from a list of 20).

Before we can enter these data, we need to define the three variables to be used (see the previous section for details of how to define a variable). Remember that as sex is a nominal variable, we should use value labels to remind ourselves what the values 1 and 2 represent.

Once the three variables have been defined we can begin entering the data. You can copy the data for the first five participants from the screen-shot shown below.

Click on the top left-hand cell of the table (ensure that you are at the top left hand corner of the window by checking the scroll bars). This cell will become highlighted (it will have a bold border). Any number you now type will appear in the bar above the variable names at the top of the window. If you press the <Enter> key or the <Tab> key, or use the mouse or cursor keys (up, down, left, and right arrows) to move to another cell, this number will be inserted into the cell.

### Moving around the Data Editor window



Check that you are in Data View before trying to enter data.

	Sex	Age	MemScore	var
1	1.00	23.00	12.00	
2	1.00	32.00	8.00	
3	1.00	25.00	11.00	
4	2.00	18.00	9.00	
5	2.00	21.00	17	
6	2.00	21.00		
7				

This shows that a value is currently being entered for the fifth participant in the column.

This value represents the memory score for the fifth participant. As you type in the number it appears both here and in the cell that was highlighted.

These are the row or case numbers. Normally you can think of them as participant numbers.

You can use either the mouse or the cursor keys to move round the data table. Alternatively, you can press the <Enter> key to move down to the next

participant for the current variable, or the <Tab> key to move across to the next variable for the current participant.

It is best to enter the data one participant at a time, working across the data table. For example, you should enter the sex, age and memory score for the first participant in row one, then for the second participant in row two, and so on. If you enter the data a column at a time working down the columns (e.g. the sex of all the participants first, then their ages etc.), you may make a mistake and such an error is likely to result in the data from one participant being assigned to another participant.

Once you have entered all your data into the data table, you should carefully check that you have entered it correctly. Cross-checking the data file against the original record of the data is a very important stage in the process of analysis. Either cross check the original records against the data on the screen, or against a printout of the data (see Chapter 13, Section 4, for details of how to print a copy of your data).



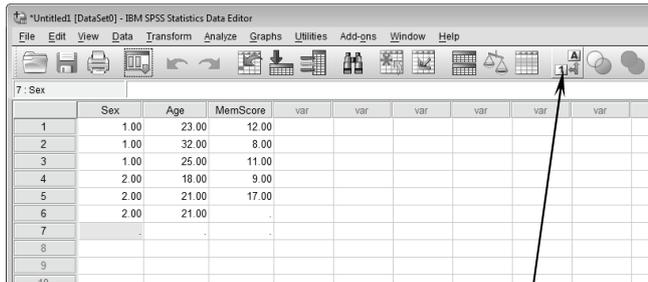
You may accidentally enter an empty row of data, which will appear as a row of cells filled with dots. If this has happened it is worth taking the time to remove the blank line(s) as SPSS will interpret each blank line as a participant for whom you have no data. Thus SPSS will tell you it has more cases than you expect. To delete the blank case, click on the case number associated with the extra row; the case will become highlighted. Now either press the delete key, or right click and select **Clear**.

Sometimes new SPSS users panic that they have ‘lost’ their data because they cannot see it on the screen. This is often because the data have scrolled out of view. Check that the scroll bars are set to the top left-hand corner of the window.

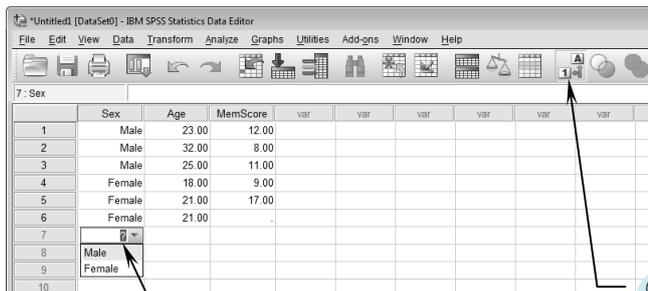
### The value labels button

If you have assigned value labels to one or more of your variables, you can choose whether you want SPSS to display the values you enter, or the labels associated with the values. For example, in this file, we have assigned the value labels ‘Male’ and ‘Female’ to the values 1 and 2 of the variable ‘Sex’. SPSS can either display the values (i.e. the numerals ‘1’ or ‘2’) or the labels ‘Male’ or ‘Female’. Clicking on the **Value Labels** button on the toolbar of the Data Editor window will toggle between these two display states (see below). This option affects only the way the data are displayed in the Data Editor window, and not the way they are entered or analysed.

Furthermore, if you select to display the value labels rather than the data values, then SPSS will help with the data entry process by providing a drop down list of possible values. See below to see how this works.



Click on the **Value Labels** button to toggle between displaying the values entered (as shown here) and the associated labels (see below).



Once the **Value Labels** button is depressed the values '1' and '2' are replaced by the associated labels 'Male' and 'Female' (as shown here).

In this mode if you double click on a cell, SPSS will offer you a drop down list of the value labels associated with this variable.

When you have finished entering your data and have carefully checked it you should save a copy of the data file. We describe how to save the data file in the next section.

 It is possible to have several data files open at a time; however, this can become confusing so we recommend that you work with only one data file until you have become familiar with SPSS.

## SECTION 4: SAVING A DATA FILE

You will have spent a lot of time entering your data, so remember to save the data file as soon as you have checked it carefully. If you are entering a large amount of data it is a good idea to save the file every few minutes.

### To save the data to a file

Click on the menu item **F**ile at the top of the screen. Now click on either **S**ave or **S**ave **A**s.

The screenshot shows the File menu in SPSS with the following options: New, Open, Open Database, Read Text Data..., Close (Ctrl+F4), Save (Ctrl+S), Save As..., Save All Data, Export to Database..., Mark File Read Only, Rename Dataset..., Display Data File Information, Cache Data..., Stop Processor (Ctrl+Period), Switch Server..., Repository, Print Preview, Print.. (Ctrl+P), Recently Used Data, Recently Used Files, and Exit. The background shows a data table with columns 'score' and 'var' and rows containing values like 12.00, 8.00, 11.00, 9.00, 17.00, and 13.00.

SPSS uses the file name 'Untitled1' for a file that has not been saved. Once you save the file your new name will appear here.

If there are two unsaved data files open at the same time, the second will be called 'Untitled2'.

Click on **S**ave to save the data file. If the file has been saved previously it will be re-saved using the same name. If not, you will be prompted for a name.

Click on **S**ave **A**s if the file has been saved before but you now want to save it under a different name.

Click on **S**ave **A**ll **D**ata if you have more than one file open, and you want to save them all.

Select **S**ave to resave the file using the existing name. The resaved file will replace the old version. If the file has not been saved previously, or if you click on **S**ave **A**s, you will be presented with the **Save Data As** dialogue box (see below).

Type the name for the file into the **File name:** box. The file name you choose should be reminiscent of the study from which the data originated (for example, 'DataEntryPractice'). You should not use a full stop in the file name and should not attach a suffix to the file name. By default SPSS will attach the suffix '.sav' to

any name you enter. Do not change this suffix, or SPSS might not recognise the file as a data file. Check which drive and which directory the file is going to be saved to, before you click the **Save** button. You may want to save the file to a different drive, or to a disk or USB stick.

The screenshot shows the 'Save Data As' dialog box in SPSS. The 'Look in:' field is set to 'DATA SETS'. The file list contains various SPSS data files, including '1-Way Between-Subjects Anova.sav', '1-Way Within-Subjects Anova.sav', '2-Way Between-Subjects Anova.sav', '2-Way Within-Subjects Anova.sav', '3-Way Mixed Anova.sav', 'adoption survey.sav', 'ancova.sav', 'Chapter 10 data set.sav', 'Chapter 11 Dataset.sav', 'Chi-Square.sav', 'Factor Analysis.sav', 'independent t-test.sav', 'Kruskal-Wallis.sav', 'Mann-Whitney.sav', 'manova.sav', 'mcnemar.sav', 'multiple regression.sav', and 'paired t-test.sav'. The 'File name:' field contains 'DataEntryPractice'. The 'Save as type:' dropdown is set to 'SPSS Statistics (\*.sav)'. There are checkboxes for 'Write variable names to spreadsheet', 'Save value labels where defined instead of data values', and 'Save value labels into a .sas file'. Buttons for 'Variables...', 'Save', 'Paste', 'Cancel', and 'Help' are visible. A 'Store File To Repository...' button is at the bottom.

We are currently in the folder called 'Data Sets'.

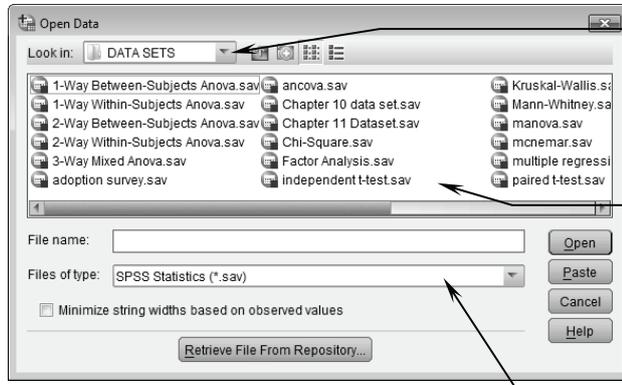
These buttons allow you to move around in folders, create new folders, and change the way the file information is displayed.

Click here and then type your chosen file name.

By default the file will be saved as an SPSS data file with the suffix '.sav' added to the end of the file name. Don't change this unless you are sure you want to save in some other format (such as Excel).



5. Select the appropriate folder from the drop down list in the box labelled **Look in**. The buttons next to this box can be used to move up one level in the directory, or to make a new folder, or to change the way the way the file information is displayed.
6. Open the file either by selecting it from the list and clicking **Open**, or by double-clicking the file name.



5. Move to the correct folder either using this drop-down list or by selecting the folder from the list below.

6. Click on the name of the file and then click the **Open** button. Alternatively double click the name of the file to open.

You can choose which type of file to show in the box above. By default only SPSS data files will be shown.

SPSS can read and write files of various formats, including Excel spreadsheet files. When you are more experienced you may like to try opening Excel files. You can do this by selecting **Excel** from the drop down list in the **Files of type:** box.



If the file you are looking for has a suffix other than '.sav', SPSS will not recognise it as a data file and will not display it in the dialogue box. If you can't find the file you are looking for, and think that it may have been saved with some other file name suffix, click on the button at the right hand end of the **Files of type:** box and select 'All files (\*.\*)' from the list of file types offered. All the files in the current directory, regardless of type or suffix name will now be displayed in the dialogue box. If you find that your data file was saved with some other suffix, use Windows Folders to make a copy and change the suffix to '.sav'.

## SECTION 6: DATA ENTRY EXERCISES

In this section, we are going to practise entering data from two different types of experimental design. Later in this chapter and in subsequent chapters we will use these data files to demonstrate other procedures. Take the time to complete these exercises, as they will help you to appreciate the way that the design employed in a study influences the shape of the data file. When you have completed these two data files, compare them to the ones shown in the next section.

### Data from an independent groups design

As explained in Chapter 1, in the independent groups design we are comparing the performance of two or more groups of different participants. In the example below, we have used this design to investigate the effect of a mnemonic instruction given to a group of participants before they were asked to learn a total of 20 words. The dependent variable was the number of words correctly recalled.

#### RODENTS IN SPACE: A SIMPLE MEMORY EXPERIMENT

Twenty-one first-year undergraduates participated in a simple memory experiment designed to investigate the effect of a mnemonic strategy upon memory for paired words. The participants were randomly divided into two groups. All participants were given 2 minutes to memorise a list of 20 words presented in pairs. All the participants were told to memorise the words, but those in the mnemonic instruction group were advised to try to form a mental image to link the two words in a pair (e.g. for the word pair ROCKET – HAMSTER a participant might imagine a small furry rodent being fired off into outer space). The participants in the other group, the non-mnemonic group, were not given this instruction. After learning the words for 2 minutes the participants were then required to complete some simple mental arithmetic problems for 2 minutes. Finally all participants tried to recall the words in any order. The number of words correctly recalled was recorded. The data are summarised below.

Memory scores (out of 20) for the mnemonic instruction group:

20, 18, 14, 18, 17, 11, 20, 18, 20, 19, 20

Memory scores (out of 20) for the non-mnemonic group:

10, 20, 12, 9, 14, 15, 16, 14, 19, 12

Using these data attempt to do the following:

1. Set up an SPSS data file to record these data. Give appropriate names to the variables you are using.
2. Apply value and variable labels where appropriate.
3. Enter and check the data, then save the file using an appropriate file name.

4. Ensure that you can re-open the file.
5. Compare the data file you have constructed to the one illustrated in the next section of this chapter.

### Data from a repeated measures design

As you will remember, in the repeated measures design, every participant is exposed to each condition and thus contributes a data point from each level of the independent variable. This will be reflected in the structure of the data file, which will have a column for each level of the independent variable. In the example below we have used this design to investigate mental representation.

#### COMPARING MENTAL IMAGES

If you ask someone the question 'How many windows are there in the front of your home?' most people will report that they attempt to answer the question by 'inspecting' a mental image of their house. Does this mean that we store information in the form of mental images? Some psychologists think not, arguing that information is actually stored in a more abstract form and that our experience of inspecting mental images is illusory. However, several lines of evidence support the idea that we are able to manipulate information utilising a form of representation that shares many qualities with mental images. This experiment is modeled on one such line of evidence.

Imagine you were asked to decide whether or not a lion was bigger than a wolf. You could make your decision by recalling information about size that was represented in some abstract form. Alternatively, you could form a mental image of these two animals standing side-by-side and decide which was the taller. If you adopted the mental imagery approach, then you might expect the decision to take longer when the two animals were of a similar size than when they were of very different sizes. If the decision were based on a more abstract form of representation, then you would expect the relative size of the animals to have no impact on the speed of the decision. Thus, if it takes longer to compare the size of two similar sized animals than two dissimilar sized animals, this will offer some support for the idea that these decisions are based on the manipulation of image-like forms of mental representation.

In our experiment each of 16 participants undertook 20 trials. In each trial the participant was presented with a pair of animal names and had to decide as quickly as possible which of the animals was the largest. The dependent variable was the time taken to make this decision (in milliseconds). For half of the trials the difference in size between the two animals was large (e.g. mosquito – elephant) and for the other half of the trials the difference in size was small (e.g. horse – zebra). In the data table below we have recorded the mean decision time (in milliseconds) for the large size difference trials and for the small size difference trials.

**DATA**

Participant	Large diff.	Small diff.
1	936	878
2	923	1005
3	896	1010
4	1241	1365
5	1278	1422
6	871	1198
7	1360	1576
8	733	896
9	941	1573
10	1077	1261
11	1438	2237
12	1099	1325
13	1253	1591
14	1930	2742
15	1260	1357
16	1271	1963

1. Set up an SPSS data file to record these data. Give appropriate names to the variables you are using.
2. Apply value and variable labels where appropriate.
3. Enter and check the data, then save the file to disk using an appropriate file name.
4. Ensure that you can re-open the file.
5. Compare the data file you have constructed to the one illustrated in the next section of this chapter.

## SECTION 7: ANSWERS TO DATA ENTRY EXERCISES

## The data file for an independent groups design

Below is a screen-shot of the data file we constructed for this simple memory experiment. Your data table might not look identical, but should have the same basic characteristics. Note that there are two critical variables in this design. The first is a nominal variable (or grouping variable) that we have used to record whether the participant was in the mnemonic or in the non-mnemonic group. Thus it indicates the level of the IV (or factor). The other critical variable is a ratio variable and has been used to record the dependent variable, the number of words each participant recalled. In addition to these two variables we have also included a variable called 'Participant' which assigns a number to each participant. This is good practice. If you have the **Value Labels** button (on the tool bar) depressed, then the 'Condition' column will display value labels rather than values (i.e. mnemonic or non-mnemonic rather than 1 or 2).

We have called this variable 'NumWords'; we have left the variable width as 8 with 2 decimal places.

We have used the name 'Condition' for this variable. Because it's a nominal variable we have used value labels to show that the value 1 indicates the participant was in the mnemonic condition, and the value 2 that they were in the non-mnemonic condition (see screen-shot of the value labels dialogue box below). We have set the variable width to 1 (with no decimal places) and the column width to 5 (neither of these changes are essential).

For this screen-shot we did not have the **Value Labels** button depressed – if it is depressed on your system, then your screen will show the value labels rather than the values.

We have included a variable called 'Participant' which gives each participant a number. This is not essential but is good practice – we will explain why later.

These are the value labels we used for the variable 'Condition'.

	Participant	Condition	NumWords
1	1.00	1	20.00
2	2.00	1	18.00
3	3.00	1	14.00
4	4.00	1	18.00
5	5.00	1	17.00
6	6.00	1	11.00
7	7.00	1	20.00
8	8.00	1	18.00
9	9.00	1	20.00
10	10.00	1	19.00
11	11.00	1	20.00
12	12.00	2	10.00
13	13.00	2	20.00
14	14.00	2	12.00
15	15.00	2	9.00
			14.00
			15.00
			16.00
			14.00
			14.00
			14.00



**Remember:** the data file constructed for an experiment that employed an independent groups design will always require a nominal variable that is used to indicate the condition under which each participant was tested.

### The data file for a repeated measures design

Below is a screen-shot of the data file we constructed to record the data from our mental imagery experiment. Your data table might not look identical, but should have the same basic characteristics. As with the independent groups design, there are a total of three variables, but in this case two are used to record the performance of the participants when tested under a particular condition. As this is a repeated measures design each participant was tested under both conditions, so we have two data points for each participant. In this design there is no need for a nominal variable.

Compare this data file to the one on the previous page. Make sure that you understand why these two files have a different structure

	PartNum	LargeDif	SmallDif
1	1.00	936.00	874.00
2	2.00	923.00	1005.00
3	3.00	896.00	1010.00
4	4.00	1241.00	1365.00
5	5.00	1278.00	1422.00
6	6.00	871.00	1198.00
7	7.00	1360.00	1576.00
8	8.00	733.00	896.00
9	9.00	941.00	1573.00
10	10.00	1077.00	1261.00
11	11.00	1438.00	2237.00
12	12.00	1099.00	1325.00
13	13.00	1253.00	1591.00
14	14.00	1930.00	2742.00
15	15.00	1260.00	1387.00
16	16.00	1271.00	1963.00
17			
18			

We have called this variable 'SmallDif' and are using it to record the participants' mean decision times on the small-difference trials. As this is a ratio variable, we have not assigned any value labels. We have left the variable width and the column width at their default values.

We have called this variable 'LargeDif' and are using it to record the participants' mean decision times on the large-difference trials. All other variable settings are left at their default values.

We have used the variable name 'PartNum' for the participant number variable. This variable isn't essential but its good practice to include.

Participant 16, for example, had a mean decision time of 1271 milliseconds for the large difference trials and 1963 milliseconds for the small difference trials.



**Remember:** in a data file constructed for a repeated measures design there must be a variable for each condition.

## SUMMARY

- In this chapter we have shown you how to create a data file in SPSS.
- In Sections 1 and 2 we explained the different parts of the data window and showed how to define a variable.
- In Section 3 we walked you through the process of setting up a data file, and Sections 4 and 5 showed how to save and open a data file.
- Section 6 provided two data entry exercises to highlight differences between the data files used to code the data from independent groups and repeated measures designs.
- Using these data files you will learn about exploring data in Chapter 3.
- For guidance on how to incorporate your data file in a report, or print it, see Chapter 13.