

# 1 Data Input

Raw data can come in a variety of sources. One of the most common data collection methods is *social surveys* which can be administered using formal interview schedules or self-completion questionnaires such as the example in Appendix 1 (p. 57). Other forms of data collection are *proforma* that are filled out (for example, job or college applications) and *personal records* (for example, a company's personnel records, medical records or academic transcripts). The possible sources of data can be almost infinite, depending upon circumstances and the ability of the researcher to recognise a source of good data when it presents itself.

In data input to SPSS, the task is to take the information which has been collected (*raw data*), organise it and move it to a working SPSS datafile.

The process of inputting data can be broken into three stages.

- Coding
- The mechanics of transferring coded information into a SPSS file
- Configuring the data within SPSS.

## Cases and variables

For a computer package to be able to receive information in the form of data, these data must be put into a regular, predictable format.

A complete individual record is usually called a *case*. A short questionnaire is provided in Appendix 1 at the end of this module. You will need to complete this now, as it will form the basis of much of the explanation that will follow. Once you have completed it, you become a *case* and your responses to this questionnaire make up a single complete *record*. Each other person who answers the questionnaire would also become a case in the data. (You may find in some manuals that the words 'record' or 'observation' are sometimes used in the place of 'case'.)

Cases, such as survey respondents, personnel records or medical files, often correspond to individual people. Cases, however, can also refer to *organisations* or *timespans*. For instance, in household survey, the information obtained relates to the entire household rather than to individuals within the house. In this instance the case would be the household, not the individual. It is also possible to have, for example, a database of hospitals (where each case is the information held on a complete hospital) or a geographically based dataset (where information on complete cities makes up the cases). A historian might have a database where the cases are timespans – say, decades. The point is that, in each instance, the case is the information recorded on a single unit of analysis.

For each case, the specific bits of information recorded for the case are called *Variables*. In the 'Drinking questionnaire' (Appendix 1), each separate answer you gave is a response to a *variable* in the dataset. For instance, age, where you said you are X years old is a variable; and your response to Question 5, where you said you *never drink alcohol/drink rarely/drink moderately/drink frequently/drink heavily* is another variable. Your particular answers to these questions are the *values* that these variables have for you. A different person answering the

questionnaire would give responses to the same questions/variables but their values for those variables could be different. For instance their name, age, faculty and drinking habits could be different from yours.

Here are two additional examples.

- A medical admission being the *case*, with *variables* being things like the cause of admission, the GP who attended, etc.
- If a household was the *case*, *variables* could be things like the number of people living in the house, tenure of housing, household income, the type of heating in the house, etc.

These concepts (cases, variables and values) are fundamental for any kind of data analysis. Think about them one more time, and make sure you understand how they relate to one another. The questions you ask, or the types of information you record, are *variables*. Each individual answer or piece of information is a *value*. The people or things you test in the study are *cases* ... For each case, you have one value for each variable. (Norusis, 1988, p. 61)

## 'Rectangular' format

Computers keep track of the bits of information on cases by always having a variable's information (*codings*) appear in the same location. For most datasets, the data usually appear in what we can call a 'grid' of *cases* by *variables*. Each *case*, or record, makes up one complete 'line' of data and the *variables* appear as columns. We can show how this would look for the answers some other people gave to the 'Drinking questionnaire' (Figure 1.1)

Figure 1.1 Responses to the 'Drinking questionnaire' in grid format

001	Robert	1	45	1	3	1	8	3	3	2	30.33
002	Margaret	2	23	2	1	1	2	2	0	0	0.00
003	Fred	1	52	3	2	1	2	0	2	1	4.55
004	Paddy	1	36	5	3	1	16	6	10	0	15.45

Robert's responses to the questionnaire are presented in a row format. Each respondent is given a unique *identifier*. Robert for example has been allocated 001, and Margaret has been allocated 002. The third column contains details of the respondent's gender. The first case, Robert, is male which has the value 1, and the second respondent, Margaret, is female which has the value 2.

All cases reserve 'space' for each variable. In responding to the questionnaire, respondents may not answer all the questions, thus some information may be *missing* for some variables. For example, someone answering the 'Drinking questionnaire' may not be able to remember how much they spent on alcohol last weekend or might refuse to tell how much they drank. It is important to record this data as *missing*. In some instances it may be that the question is *not relevant*. For example, if the respondent did not drink alcohol during the previous weekend (Question 6) they did not need to complete Questions 7 and 8 about types of alcohol consumed and the amount spent. In this instance a 'not relevant' code might be used for Questions 7 and 8. (We will return to this below.)

**Exercise**

As an exercise, try transferring *your* responses to the questionnaire on to 'line 005' of the grid in Figure 1.1.

**Inputting data into SPSS**

It is possible to input data into SPSS in a number of ways. Since the SPSS 'data grid' is in the form of a spreadsheet, the normal conventions for using a spreadsheet apply. This means that data can be brought into SPSS in one of four ways:

1. Imported from SPSS on another computer as an *SPSS Portable file (\*.por)*
2. Imported from *database spreadsheets* (for example, Excel, dBase, Foxpro, Access)
3. Imported from *text files* (ascii-files)
4. Entered directly onto the *SPSS Data Editor* window.

**Option 1: importing an SPSS portable file**

Often we will want to transfer a SPSS data file from one computer to another. This is accomplished by putting the data file that is to be transferred into a special form – a *SPSS*

*portable file*. The data is read from the portable file and immediately saved as a normal SPSS data file. (The reason we do this is that the 'internal architecture' of computers can differ and there is no guarantee that a SPSS data file that works fine on one computer will always work automatically on another.) Below, we have an example of importing the *BSA Crime* dataset. (Note that before you begin, you already have to have downloaded the SPSS portable file for the *Crime* dataset from the Palgrave web site onto the hard disk of your computer. Instructions on how to do this are given at the Palgrave web site (<http://www.palgrave.com/sociology/miller>) and in the Appendix to the Orientation chapter, p.30.)

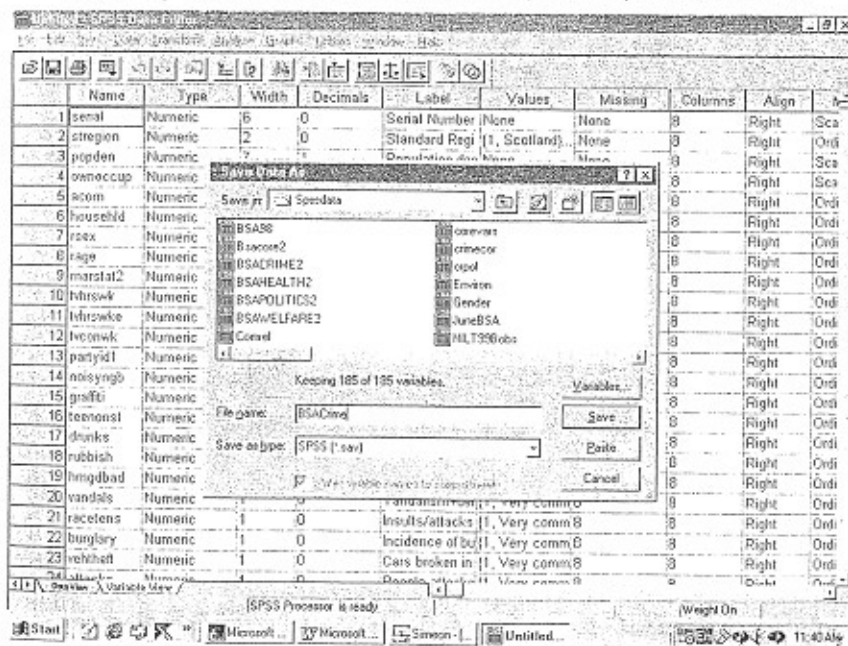
Click on the **File** menu, then on **Open** and then choose **Data...** A window like that in Figure 1.2a will open up. Go to the 'Look in' box in the *Open File* window and select the folder on your hard disk where you have saved the *BSACrime* portable file. (In our example the name of the folder is 'Spssdata'.) Open the 'pull down' menu for 'Files of type', scroll down and select 'SPSS portable (\*.por)'. Now, highlight the file named '*BSACrime.por*' (it may be the only \*.por file appearing on your computer screen). The file's name should appear in 'File name' box. Now, click on the 'Open' button and the data for the *Crime* dataset should fill the data grid.

You now need to save the imported data as a normal SPSS data file. It is quite easy to do this: click on the **File** menu again and then on **Save As...** A window like that in Figure 1.2b should appear. Click on the 'Save' button and the data grid will be saved as a normal SPSS data file (*BSACrime.sav*). Later on, when you want to analyse the *Crime* dataset, you always will call up the SPSS \*.sav file, *BSACrime.sav*.

Figure 1.2a Importing a SPSS portable file (BSACrime) using File Open Data

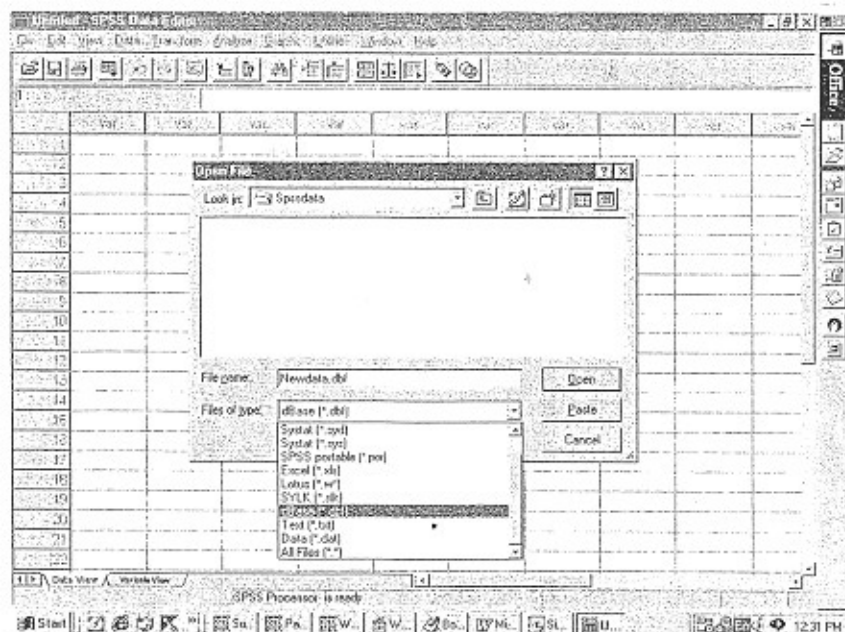


Figure 1.2b Saving an imported SPSS portable file (BSACrime) as a SPSS\*.sav file



**Option 2: importing data from spreadsheets**

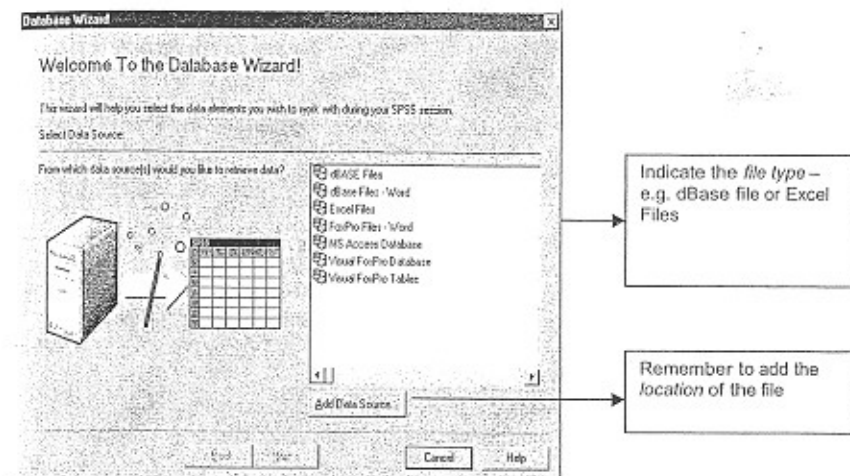
Those of you that have used a spreadsheet package probably will have noticed that the SPSS 'grid' which consists of rows of 'cases' and 'columns' of variables corresponds to the way a spreadsheet configures data. Spreadsheet users sometimes find that a convenient way to input data into SPSS is to type the data into your favourite spreadsheet programme and then import the data to SPSS. SPSS has facilities for handling data from the more commonly-used spreadsheets. It is possible to use the **File Open** option to import a spreadsheet file. Click on the **File** menu, then on **Open** and then choose **Data**. A window like that in Figure 1.2c will open up. Select the location and scroll down to select appropriate file type, then click **OK**. Here, we have selected a dBase file called 'Newdata.dbf'.

Figure 1.2c Importing files from other spreadsheets using **File Open Data**

The default is a SPSS system data file \*.sav. Scroll down to select file type (Here, a dBase File, Newdata.dbf, has been chosen).

Some caution is needed when importing files from different applications. Check if the SPSS version you are using can read the latest version of the database application you want to import. If SPSS cannot read the latest version, save the spreadsheet data using a version which can be read by the SPSS programme you are using.

SPSS also provides the option to import database files using the Capture Wizard using the pull-down menu. Click on the **File** menu as before, then on **Open Database** and then choose **New Query** to open the capture **Database Wizard**. A window like that in Figure 1.2d will appear. The **Database Wizard** provides a set of step-by-step instructions that can be followed to introduce a database file into SPSS.

Figure 1.2d Importing spreadsheet files from other applications using **Database Wizard**

Indicate the *file type* – e.g. dBase file or Excel Files

Remember to add the *location* of the file

It is also possible to use **Copy** and **Paste** commands to import to and export from SPSS and other spreadsheet applications. To bring a distinct block of data from a spreadsheet into SPSS, simply open the spreadsheet application, and using the cursor highlight the data block (excluding any headings) and use **Copy** from the **Edit** menu to copy the data block. Once you have copied the data, go to the **Data View** window in SPSS (the SPSS data grid), 'drag' your mouse from the upper-left-hand corner of the data grid to highlight a block of cells in the grid and **Paste** the data onto the grid. **Note:** for this 'copy and paste' to work, the number of rows and columns highlighted on the data grid must *match* the number of rows and columns of data that are being taken from the spreadsheet. Once the data values are safely imported, the variables are labelled using **Variable View** (how to do this is explained below).

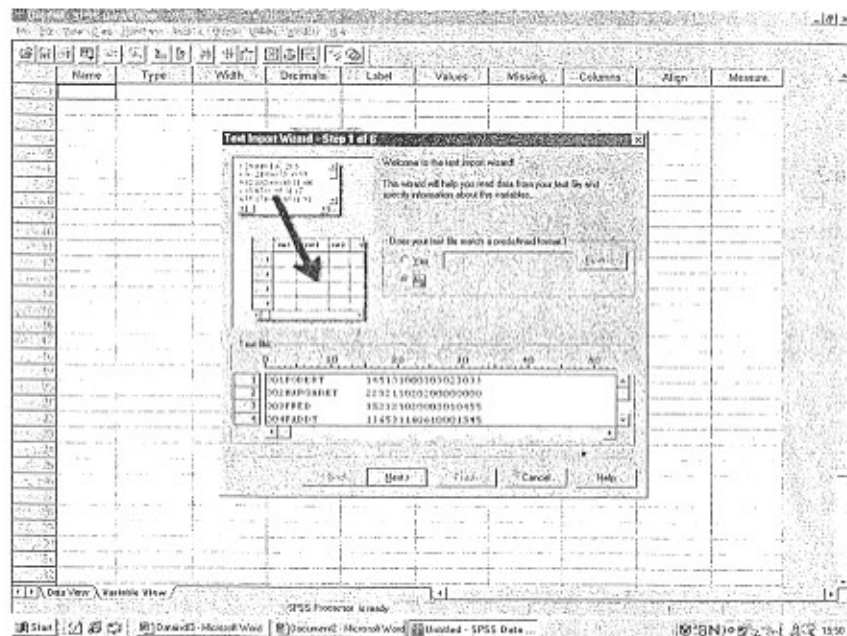
**Option 3: importing text files**

Others may have data coming from a non-spreadsheet source. These sources probably will be put into the form of an ASCII file. (Basically, here an ASCII file is a grid of rows and columns of

raw numbers laid out in a similar manner to a spreadsheet, but without any of the labelling that a spreadsheet grid would have. ASCII files are a form of data that is basic to many different types of computer programmes and hence is used as a common means of transferring information between them.) SPSS can handle data from an ASCII file quite happily.

Open the main **File** menu and select **Read Text Data** in order to open a Text Import Wizard as shown in Figure 1.2e. The Text Import Wizard provides a set of step-by-step instructions that can be followed to introduce an ASCII file if data into SPSS.

Figure 1.2e Text Import Wizard



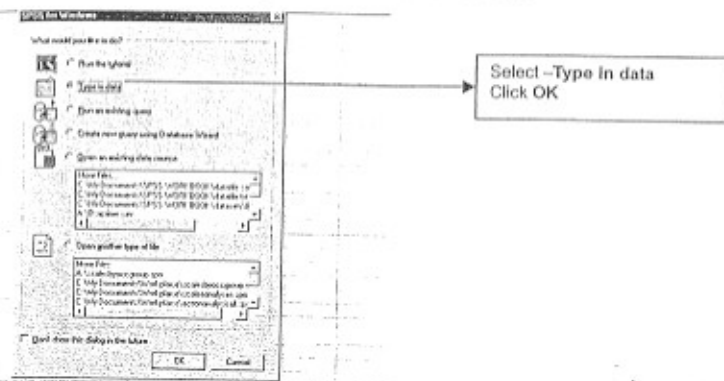
#### Option 4: creating a new SPSS data file

There are two stages to creating a new data file in SPSS. Enter the data into the Data View window, and specify the name, type and other characteristics of each variable. New data can be typed directly into the grid on the SPSS Data View window of the Data Editor.

When you first open SPSS the dialog box offers a range of options. A new file can be opened by selecting **Type in data** option from the dialog box and click on **OK** (Figure 1.2f). This will open the Data View window of the Data Editor.

Alternatively, a new data file can be opened using the pull-down **File** menu, choosing **New** and then **Data**.

Figure 1.2f Dialog box to create a new data file



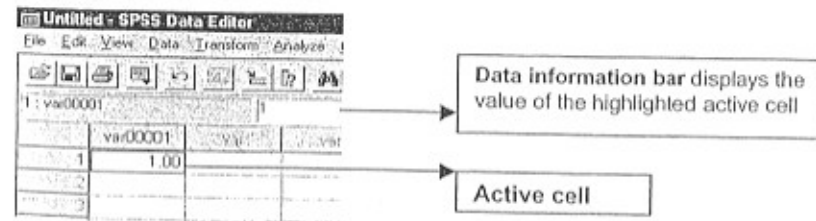
Both procedures open an empty file on the Data Editor window. Remember that the Data Editor in SPSS V10 has two formats:

- **Data View**, which is the spreadsheet for *entering* the data (the default format)
- **Variable View**, for *naming* and *specifying* the characteristics of variables.

To change between the **Data View** and **Variable View** formats, click on the **tab** label on the bottom-left-hand side of the window.

To enter data using the **Data View** format on the Data Editor window, bring the cursor to the upper-left-hand corner of the grid (row 1, column 1). Click once, this will highlight the active cell where the data will be entered. Type in the first value. Once you have entered the data, press the **Enter** key (to move down the column) or the **Tab** key (to move across the row), and the next cell will be highlighted. As you enter data into the cells the value will appear in the data entry information bar above the column headings (Figure 1.2g).

Figure 1.2g Data entry directly to Data View window of the Data Editor



Remember data from an individual questionnaire (case) are entered in *rows*, and responses to a particular question are entered in *columns* (Figure 1.2h).

Figure 1.2h Questionnaire data entered directly to Data View window of the Data Editor

Each row represents a respondent or a case

Each column represents a response to a question – a variable

	var00001	var00002	var00003	var00004	var00005
1	1	Robert	1	45	1
2	2	Margaret	2	32	2
3	3	Fred	3	52	3
4	4	Paddy	1	36	6

### Editing data

It is possible to *edit* or *change* data in an existing data file by locating the data you wish to edit, clicking once on the cell, and replacing the old value in the cell by typing in a new value. You might want to edit data if you discover *errors* or *mistakes* in the existing data. Data editing can be done at any time, not just when a new dataset is first being created.

### Entering new variables or cases

It is possible to enter new variables or cases to a data file. To insert a new case (for instance, a new respondent to a survey), simply select a cell in the row above which you want to insert the new row. Go to the **Data** main menu and click to open it. Click on **Insert Case** or click on the **Insert Case** button on the toolbar. This creates a new row, into which the data can be entered. Alternatively, go to the empty rows at the bottom of the grid and enter the data onto the first empty row.

To insert a new variable (or column), select any cell in the column to the right of which you want the new column. Open the **Data** main menu and click on **Insert Variable** or click on the **Insert Variable** button on the toolbar. This creates a new variable, into which new data can be entered. Again, alternatively you can go to the empty columns at the far-right-hand side of the grid and enter the data into the first empty column.

### Deleting a variable or case

It is possible to delete a variable or case. To delete a case or row click on the case number on the shaded column in the left side of the file. This will highlight the entire row, then open the **Edit** main menu and click **Clear** or press the **Del** key. (In Figure 1.3a, we are deleting the variable Margaret.)

Figure 1.3a Deleting a case

Click on case number of the case you want to delete – this highlights the entire row. Click on Del key.

	id	name	sex	age	faculty	drink	wand	total	beer	wine	spirits	money
1	1	Robert	Male	45	Social S	3	Yes	0	0	0	0	33,000
2	2	Margaret	Female	32	never dr	Yes	0	2	0	2	0	55,000
3	3	Fred	Male	52	drinks int	Yes	0	2	0	0	0	55,000
4	4	Paddy	Male	36	5	3	Yes	1	6	0	6	45,000

To delete a variable or column click on the variable label on the shaded row (Figure 1.3b). This will highlight the entire variable, open the **Edit** main menu and delete the column by clicking **Clear** or press the **Del** key. (Here, we are deleting the variable age.)

Figure 1.3b Deleting a variable

Click on the variable name – this highlights the column. Click on Del key.

	id	name	sex	age	faculty	drink	wand	total	beer	wine	spirits	money
1	1	Robert	Male	45	Social S	3	Yes	0	0	0	0	33,000
2	2	Margaret	Female	32	never dr	Yes	0	2	0	2	0	55,000
3	3	Fred	Male	52	drinks int	Yes	0	2	0	0	0	55,000
4	4	Paddy	Male	36	5	3	Yes	1	6	0	6	45,000

### Saving new files

It is important to *name* and *save* the new datafile you have created. Do not wait until you have entered all the data, as computer malfunctions or power cuts can result in lost work. To save the file you must assign a *unique name* to the file which identifies the data which the file contains. For example, we might assign the name *drinksurvey* to the file which contains the data from the survey. Files can be saved using by opening the **File** main menu and selecting **Save As**. Then use the dialog box to select the directory or disk (floppy disk) in which you want to save the data. Enter the new name of the file in the **File Name Box** (In Figure 1.3c, *drinksurvey*), and click **OK**.

SPSS will save the data file as an SPSS data file (\*.sav) as it is at the point in time when you save the data. Later on, if you make additional changes to the data file that you also want to be saved, you will want to save the newer version of the data file again. (See the Orientation chapter, p. 25, for more details about saving files.)