4 The IBM SPSS Statistics environment

4.1 What will this chapter tell me?

At about 5 years old I moved from nursery to primary school. Even though my older brother (you know, Paul, 'the clever one') was already there, I was really apprehensive on my first day. My nursery school friends were all going to different schools and I was terrified about meeting new children. I arrived in my classroom, and as I'd feared, it was full of scary children. In a fairly transparent ploy to make me think that I'd be spending the next 6 years building sand castles, the teacher told me to play in the sandpit. While I was nervously trying to discover whether I could build a pile of sand high enough to bury my head in it, a boy came to join me. His name was Jonathan Land, and he was really nice. Within an hour, he was my new best friend (5-year-olds are fickle …) and I loved school. We remained close friends all through primary school. Sometimes, new environments seem scarier than they really are. This chapter introduces you to what might seem like a scary new environment: IBM SPSS Statistics. I won't lie, the SPSS environment is a more unpleasant environment in which to spend time than a sandpit, but try getting a plastic digger to do a least squares regression for you. For the purpose of this chapter, I intend to be a 5-year-old called Jonathan. Thinking like a 5-year-old comes quite naturally to me, so it should be fine. I will hold your hand, and show you how to use the diggers, excavators, grabbers, cranes, front loaders, telescopic handlers, and tractors[1] in the sandpit of IBM SPSS Statistics. In short, we're going to learn the tools of IBM SPSS Statistics, which will enable us, over subsequent chapters, to build a magical sand palace of statistics. Or thrust our faces into our computer monitor. Time will tell.

[1] Yes, I have been spending a lot of time with a vehicle-obsessed 2-year-old boy recently.

Figure 4.1 All I want for Christmas is … some tasteful wallpaper

4.2 Versions of IBM SPSS Statistics

This book is based primarily on version 25 of IBM SPSS Statistics (I generally call it SPSS for short). IBM regularly improves and updates SPSS, but this book covers only a small proportion of the functionality of SPSS, and focuses on tools that have been in the software a long time and work well. Consequently, improvements made in new versions of SPSS Statistics are unlikely to impact the contents of this book. With a bit of common sense, you can get by with a book that doesn't explicitly cover the latest version (or the version you're using). So, although this edition was written using version 25, it will happily cater for earlier versions (certainly back to version 18), and most likely for versions 26 onwards (unless IBM does a major overhaul just to keep me on my toes).

IBM SPSS Statistics comes in four flavours:[2]

[2] You can look at a detailed comparison here:
https://www.ibm.com/marketplace/spss-statistics/purchase

- **Base**: Most of the functionality covered in this book is in the base package. The exceptions are exact tests and bootstrapping, which are available only in the *premium* edition.
- **Standard**: This has everything in the base package but also covers generalized linear models (which we don't get into in this book).
- **Professional**: This has everything in the standard edition, but with missing value imputation and decision trees and forecasting (again, not covered in this text).
- **Premium**: This has everything in the professional package but also exact tests and bootstrapping (which we cover in this book), and structural equation modelling and complex sampling (which we don't cover).

There is also a subscription model where you can buy monthly access to a base package (as described above but also including bootstrapping) and, for an extra fee, add-ons for:

- **Custom tables and advanced statistics users**: is similar to the standard package above in that it adds generalized linear models. It also includes logistic regression, survival analysis, Bayesian analysis and more customization of tables.

- **Complex sampling and testing users**: adds functionality for missing data and complex sampling as well as categorical principal components analysis, multidimensional scaling, and correspondence analysis.
- **Forecasting and decision trees users**: as the name suggests, this adds functionality for forecasting and decision trees as well as neural network predictive models.

If you are subscribing, then most of the contents of this book appear in the base subscription package, with a few things (e.g., Bayesian statistics and logistic regression) requiring the advanced statistics add on.

4.3 Windows, Mac OS and Linux

SPSS Statistics works on Windows, Mac OS, and Linux (and Unix-based operating systems such as IBM AIX, HP-UX, and Solaris). SPSS Statistics is built on a program called Java, which means that the Windows, Mac OS and Linux versions differ very little (if at all). They look a bit different, but only in the way that, say, Mac OS looks different from Windows anyway.[3] I have taken the screenshots from Windows because that's the operating system that most readers will use, but you can use this book if you have a Mac (or Linux). In fact, I wrote this book using a Mac.

[3] You can get the Mac OS version to display itself like the Windows version, but I have no idea why you'd want to do that.

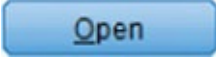**Figure 4.2** The start-up window of IBM SPSS

4.4 Getting started

SPSS mainly uses two windows: the **data editor** (this is where you input your data and carry out statistical functions) and the **viewer** (this is where the results of any analysis appear). You can also activate the **syntax editor** window (see Section 4.10), which is for entering text commands (rather than using dialog boxes). Most beginners ignore the syntax window and click merrily away with their mouse, but using syntax does open up additional functions and can save time in the long run. Strange people who enjoy statistics can find numerous uses for syntax and dribble excitedly when discussing it. At times I'll force you to use syntax, but only because I wish to drown in my own saliva.

When SPSS loads, the start-up window in Figure 4.2 appears. At the top left is a box labelled *New Files*, where you can select to open an empty data editor window, or begin a database query (something not covered in this book). Underneath, in the box labelled *Recent Files*, there will appear a list of any SPSS data files (on the current computer) on which you've recently worked. If you want to open an existing file, select it from the list and then click . If you want to open a file that isn't in the list, select and click to open a window for browsing to the file you want (see Section

). The dialog box also has an overview of what's new in this release and contains links to tutorials and support, and a link to the online developer community. If you don't want this dialog to appear when SPSS starts up, then select . ☑ **D̲on't show this dialog in the future**

Figure 4.3 The SPSS Data Editor
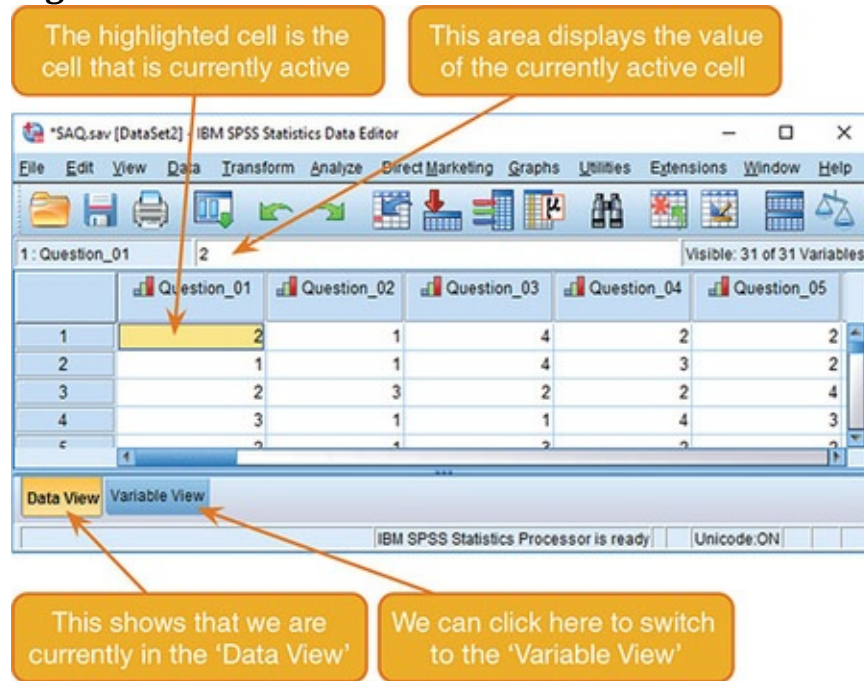
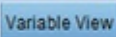## 4.5 The data editor

Unsurprisingly, the data editor window is where you enter and view data (). At the top of this window (or the top of the screen on a Mac) is a menu bar like ones you've probably seen in other programs. As I am sure you're aware, you can navigate menus by using your mouse/trackpad to move the on-screen arrow to the menu you want and pressing (*clicking*) the left mouse button once. The click will reveal a list of menu items in a list, which again you can click

using the mouse. In SPSS if a menu item is followed by a ▶ then clicking on it will reveal another list of options (a *submenu*) to the right of that menu item; if it doesn't then clicking on it will activate a window known as a *dialog box*. Any window in which you have to provide information or a response (i.e., 'have a dialog' with the computer) is a dialog box. When referring to selecting items in a menu, I will use the menu item names connected by arrows to indicate moving down items or through submenus. For example, if I were to say that you should

select the *Save As …* option in the *File* menu, you will see *File* ![icon] *Save As …* The data editor has a **data view** and a **variable view**. The data view is for entering data, and the variable view is for defining characteristics of the variables within the data editor. To switch between the views, select one of the tabs at the bottom of the data editor (Data View Variable View); the highlighted tab indicates which view you're in (although it's obvious). Let's look at some features of the data editor that are consistent in both views. First, the menus. Some letters are underlined within menu items in Windows, which tells you the *keyboard shortcut* for accessing that item. With practice these shortcuts are faster than using the mouse. In Windows, menu items can be activated by simultaneously pressing *Alt* on the keyboard and the underlined letter. So, to

access the *File* ![icon] *Save As …* menu item you would simultaneously press *Alt* and F on the keyboard to activate the *File* menu, then, keeping your finger on the *Alt* key, press A. In Mac OS, keyboard shortcuts are listed in the menus, for

example, you can save a file by simultaneously pressing ⌘ and S (I denote

these shortcuts as ⌘ + S). Below is a brief reference guide to each of the menus:

- *File* This menu contains all the options that you expect to find in *File* menus: you can save data, graphs or output, open previously saved files and print graphs, data or output.
- *Edit* This menu contains edit functions for the data editor. For example, it is possible to *cut* and *paste* blocks of numbers from one part of the data editor to another (which is handy when you realize that you've entered lots of numbers in the wrong place). You can insert a new variable into the data editor (i.e., add a column) using ![Insert Variable icon], and add a new row of data between two existing rows using ![Insert Cases icon]. Other useful options for large data sets are the ability to skip to a particular row (![Go to Case icon]) or column (![Go to Variable icon]) in the data editor. Finally, although for most people the default preferences are fine, you can change them by selecting . ![Options icon]
- *View* This menu deals with system specifications such as whether you have grid lines on the data editor, or whether you display value labels (exactly

what value labels are will become clear later).

- *Data* This menu is all about manipulating the data in the data editor. Some of the functions we'll use are the ability to split the file (⊞ **Split File...**) by a grouping variable (see Section 6.10.4), to run analyses on only a selected sample of cases (⊞ **Select Cases...**), to weight cases by a variable (⚖ **Weight Cases...**) which is useful for frequency data (Chapter 19), and to convert the data from wide format to long or vice versa (⊞ **Restructure...**) which we'll use in Chapter 12.

- *Transform* This menu contains items relating to manipulating variables in the data editor. For example, if you have a variable that uses numbers to code groups of cases then you might want to switch these codes around by changing the variable itself (⊞ Recode into Same Variables...) or creating a new variable (⊞ Recode into Different Variables...); see SPSS Tip 11.2. You can also create new variables from existing ones (e.g., you might want a variable that is the sum of 10 existing variables) using the *compute* function (⊞ Compute Variable...); see Section 6.12.6.

- *Analyze* The fun begins here, because the statistical procedures lurk in this menu. Below is a rundown of the bits of the statistics menu that we'll use in this book:
  - *Descriptive Statistics* We'll use this for conducting descriptive statistics (mean, mode, median, etc.), frequencies and general data exploration. We'll use *Crosstabs…* for exploring frequency data and performing tests such as chi-square, Fisher's exact test and Cohen's kappa (Chapter 19).
  - *Compare Means* We'll use this menu for *t*-tests (related and unrelated – Chapter 10) and one-way independent ANOVA (Chapter 12).
  - *General Linear Model* This menu is for linear models involving categorical predictors, typically experimental designs in which you have manipulated a predictor variable using different cases (independent design), the same cases (repeated measures deign) or a combination of these (mixed designs). It also caters for multiple outcome variables, such as in multivariate analysis of variance (MANOVA) – see Chapters 13–17.
  - *Mixed Models* We'll use this menu in Chapter 21 to fit a multilevel linear model and growth curve.
  - *Correlate* It doesn't take a genius to work out that this is where measures of correlation hang out, including bivariate correlations such as Pearson's *r*, Spearman's rho (ρ) and Kendall's tau (τ) and partial correlations (see Chapter 8).

- *Regression* There are a variety of regression techniques available in SPSS, including simple linear regression, multiple linear regression ([Chapter 9](#)) and logistic regression ([Chapter 20](#)).
  - *Loglinear* Loglinear analysis is hiding in this menu, waiting for you, and ready to pounce like a tarantula from its burrow ([Chapter 19](#)).
  - *Dimension Reduction* You'll find factor analysis here ([Chapter 19](#)).
  - *Scale* We'll use this menu for reliability analysis in [Chapter 18](#).
  - *Nonparametric Tests* Although, in general, I'm not a fan of these tests, in [Chapter 7](#) I prostitute my principles to cover the Mann–Whitney test, the Kruskal–Wallis test, Wilcoxon's test and Friedman's ANOVA.
- *Graphs* This menu is used to access the Chart Builder (discussed in [Chapter 5](#)), which is your gateway to, among others, bar charts, histograms, scatterplots, box–whisker plots, pie charts and error bar graphs.
- *Utilities* There's plenty of useful stuff here, but we don't get into it. I will mention that 📝 Data File Comments… is useful for writing notes about the data file to remind yourself of important details that you might forget (where the data come from, the date they were collected and so on).
- *Extensions (formerly Add-ons)* Use this menu to access other IBM software that augments SPSS Statistics. For example, *IBM SPSS Sample Power* computes the sample size required for studies and power statistics (see [Section 2.9.7](#)), and if you have the premium version you'll find IBM SPSS AMOS listed here, which is software for structural equation modelling. Because most people won't have these add-ons (including me) I'm not going to discuss them in the book. We'll also use the *Utilities* submenu to install custom dialog boxes (▦ Install Custom Dialog (Compatibility mode)) later in this chapter.[4]
- *Window* This menu allows you to switch from window to window. So, if you're looking at the output and you wish to switch back to your data sheet, you can do so using this menu. There are icons to shortcut most of the options in this menu, so it isn't particularly useful.
- *Help* Use this menu to access extensive searchable help files.

[4] In version 23 of IBM SPSS Statistics, this function can be found in *Utilities*

Custom Dialogs ….

SPSS Tip 4.1 Save time and avoid RSI

By default, when you go to open a file, SPSS looks in the directory in which it is stored, which is usually not where you store your data and output. So, you waste time navigating your computer trying to find your data. If you use SPSS as much as I do then this has two consequences: (1) all those seconds have added up to weeks navigating my computer when I could have been doing something useful like playing my drum kit; (2) I have increased my chances of getting RSI in my wrists, and if I'm going to get RSI in my wrists I can think of more enjoyable ways to achieve it than navigating my computer (drumming again, obviously). Luckily, we can avoid wrist death by using *Edit* ▶ ☑ Optio<u>n</u>s... to open the *Options* dialog box ([Figure 4.4](#)) and selecting the 'File Locations' tab.

In this dialog box we can select the folder in which SPSS will initially look for data files and other files. For example, I keep my data files in a single folder called, rather unimaginatively, 'Data'. In the dialog box in [Figure 4.4](#) I have clicked on Bro<u>w</u>se... and then navigated to my data folder. SPSS will now use this as the default location when I open files, and my wrists are spared the indignity of RSI. You can also select the option for SPSS to use the *Last folder used*, in which case SPSS remembers where you were last time it was loaded and uses that folder as the default location when you open or save files.
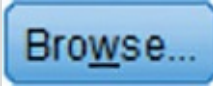
Figure 4.4 The *Options* dialog box
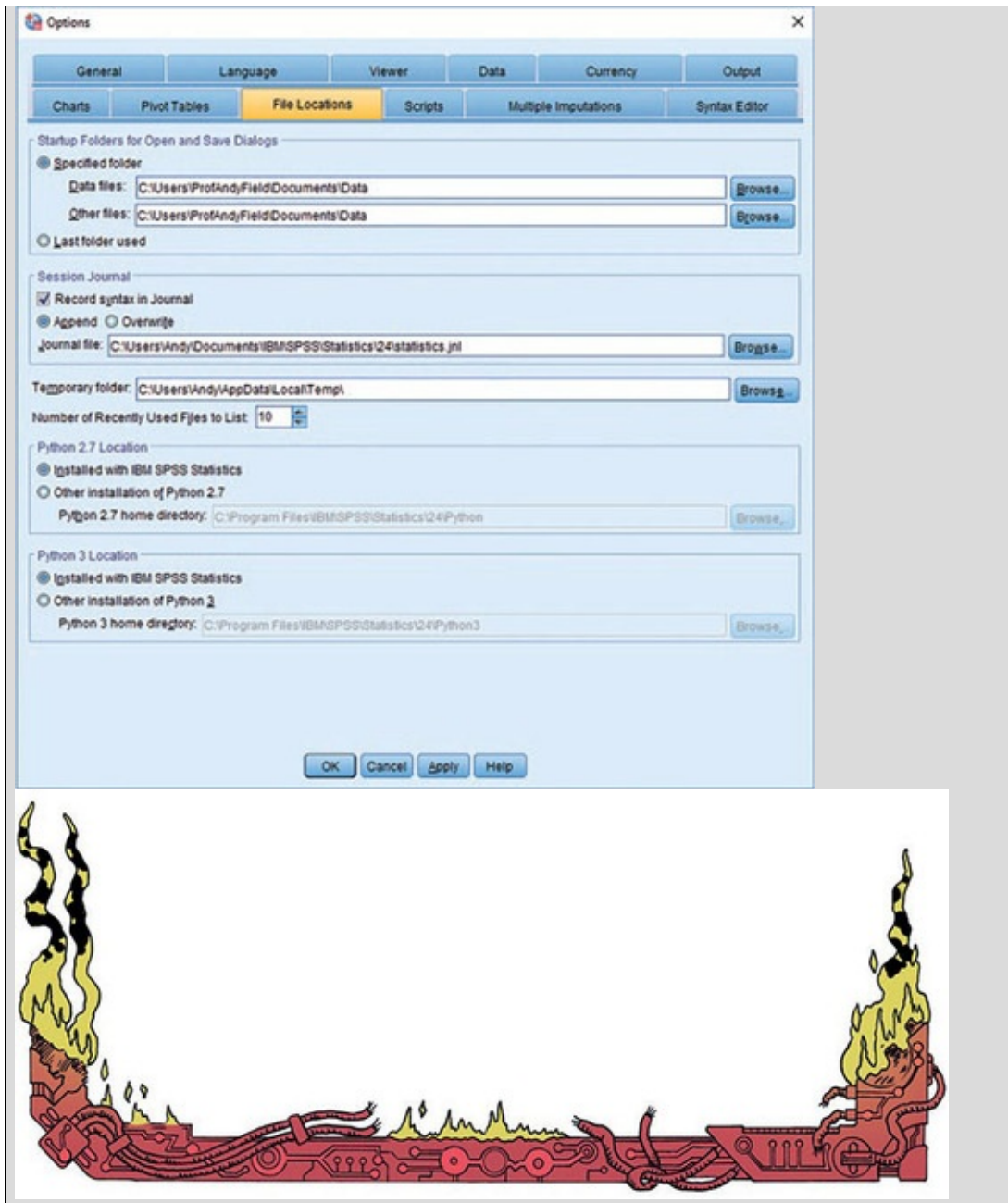
At the top of the data editor window are a set of *icons* (see Figure 4.3) that are shortcuts to frequently used facilities in the menus. Using the icons saves you time. Below is a brief list of these icons and their functions.

Use this icon to open a previously saved file (if you are in the data editor, SPSS assumes you want to open a data file; if you are in the output viewer, it will offer to open a viewer file).

Use this icon to save files. It will save the file you are currently working on (be it data, output or syntax). If the file hasn't already been saved it will produce the *Save Data As* dialog box.

Use this icon for printing whatever you are currently working on (either the data editor or the output). The exact print options will depend on your printer. By default, SPSS prints everything in the output window, so a useful way to save trees is to print only a selection of the output (see SPSS Tip 4.5).

Clicking on this icon activates a list of the last 12 dialog boxes that were used; select any box from the list to reactivate the dialog box. This icon is a useful shortcut if you need to repeat parts of an analysis.

The big arrow on this icon implies to me that clicking it activates a miniaturizing ray that shrinks you before sucking you into a cell in the data editor, where you will spend the rest of your days cage-fighting decimal points. It turns out my intuition is wrong, though, and this icon opens the 'Case' tab of the *Go To* dialog box, which enables you to go to a specific case (row) in the data editor. This shortcut is useful for large data files. For example, if we were analysing a survey with 3000 respondents, and wanted to look at participant 2407's responses, rather than tediously scrolling down the data editor to find row 2407 we could click this icon, enter 2407 in the response box and

click ([Figure 4.5](#), left).

 As well as data files with huge numbers of cases, you sometimes have ones with huge numbers of variables. Like the previous icon, clicking this one opens the *Go To* dialog box but in the 'Variable' tab, which enables you to go to a specific variable (column) in the data editor. For example, the data file we use in [Chapter 18](#) (**SAQ.sav**) contains 23 variables and each variable represents a question on a questionnaire and is named accordingly. If we wanted to go to Question 15, rather than getting wrist cramp by scrolling across the data editor to find the column containing the data for Question 15, we could click this icon, scroll down the variable list to Question 15 and click

 ([Figure 4.5](#), right).

Figure 4.5 The *Go To* dialog boxes for a case (left) and a variable (right)
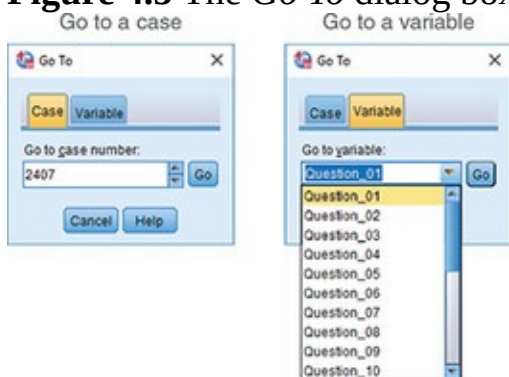
 Clicking on this icon opens a dialog box that shows you the variables in the data editor on the left and summary information about the selected variable on the right. [Figure 4.6](#) shows the dialog box for the same data

file that we discussed for the previous icon. I have selected the first variable in the list on the left, and on the right we see the variable name (Question_01), the label (Statistics makes me cry), the measurement level (ordinal), and the value labels (e.g., the number 1 represents the response of 'strongly agree').

**Figure 4.6** Dialog box for the *Variables* icons
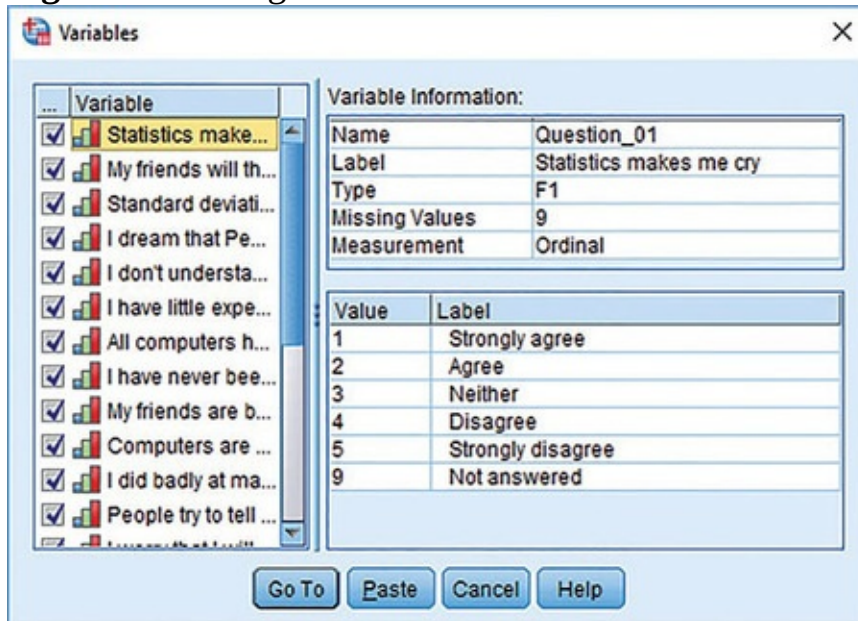
If you select a variable (column) in the data editor by clicking on the name of the variable (at the top of the column) so that the column is highlighted, then clicking this icon will produce a table of descriptive statistics for that variable in the viewer window. To get descriptive statistics for multiple variables hold down *Ctrl* as you click at the top of the columns you want to summarize to highlight them, then click the icon.

I initially thought that this icon would allow me to spy on my neighbours, but this shining diamond of excitement was snatched cruelly from me as I discovered that it enables me to search for words or numbers in the data editor or viewer. In the data editor, clicking this icon initiates a search within the variable (column) that is currently active. This shortcut is useful if you realize

from plotting the data that you have made an error, for example typed 20.02 instead of 2.02 (see Section 5.4), and you need to find the error – in this case by searching for 20.02 within the relevant variable and replacing it with 2.02 (Figure 4.7).

Figure 4.7 The *Find and Replace* dialog box
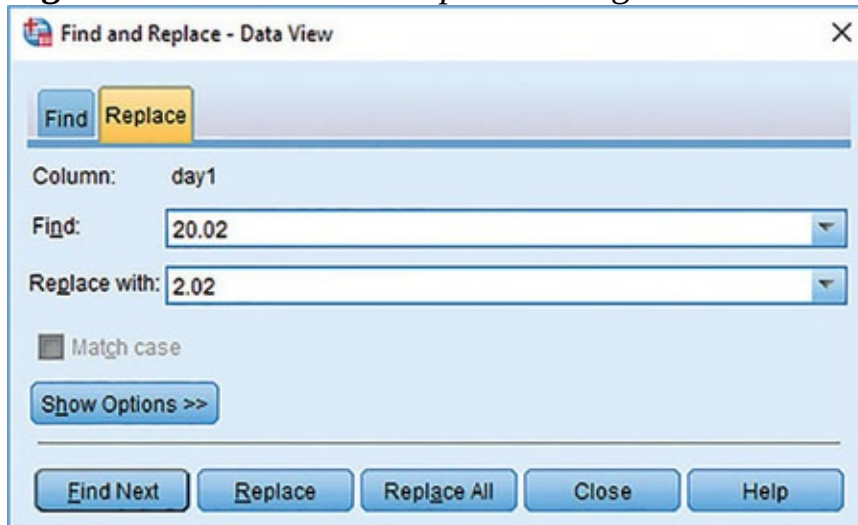
Clicking on this icon inserts a new case in the data editor (it creates a blank row at the point that is currently highlighted in the data editor).

Clicking on this icon creates a new variable to the left of the variable that is currently active (to activate a variable click the name at the top of the column).
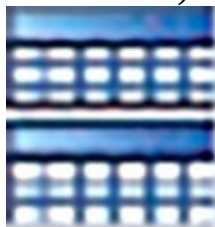
Clicking on this icon is a shortcut to the *Data* ▶ ☰ Split File... dialog box (see Section 6.10.4). In SPSS, we differentiate groups of cases by using a coding variable (see Section 4.6.5), and this function runs any analyses separately for groups coded with such a variable. For example, imagine we test

males and females on their statistical ability. We would code each participant with a number that represents their sex (e.g., 1 = female, 0 = male). If we then want to know the mean statistical ability for males and females separately we ask SPSS to split the file by the variable Sex and then run descriptive statistics.

This icon shortcuts to the *Data* ▶ ⚖ Weight Cases... dialog box. As we shall see, you sometimes need to use the *weight cases* function when you analyse frequency data (see Section 19.7.2). It is also useful for some advanced issues in survey sampling.

This icon is a shortcut to the *Data* ▶ ▦ Select Cases... dialog box, which can be used if you want to analyse only a portion of your data. This function allows you to specify what cases you want to include in the analysis.

Clicking on this icon either displays or hides the value labels of any coding variables in the data editor. We use a coding variable to input information about category or group membership. We discuss this in Section 4.6.5. Briefly, if we wanted to record participant sex, we could create a variable called Sex and assign 1 as female and 0 as male. We do this by assigning *value labels* describing the category (e.g,. 'female') to the number assigned to the category (e.g., 1). In the data editor, we'd enter a number 1 for any females and 0 for any males. Clicking this icon toggles between the numbers you entered (you'd see a column of 0s and 1s) and the value labels you assigned to those numbers (you'd see a column displaying the word 'male' or 'female' in each cell).

4.6 Entering data into IBM SPSS Statistics

4.6.1 Data formats

There are two common data entry formats, which are sometimes referred to as **wide format data** and **long format data**. Most of the time, we enter data into SPSS in wide format, although you can switch between wide and long formats using the *Data* ▶ Restructure... menu. In the wide format *each row represents data from one entity and each column represents a variable.* There is no discrimination between predictor (independent) and outcome (dependent) variables: both appear in a separate column. The key point is that each row represents one entity's data (be that entity a human, mouse, tulip, business, or water sample) and any information about that entity should be entered across the data editor. Contrast this with long format, in which scores on an outcome variable appear in a single column and rows represent a combination of the attributes of those scores. In long format data, scores from a single entity can appear over multiple rows, where each row represents a combination of the attributes of the score (the entity from which the score came, to which level of an independent variable the score belongs, the time point at which the score was recorded, etc.).

We use the long format in , but for everything else in this book we use the wide format, so let's look at an example of how to enter data in this way. Imagine you were interested in how perceptions of pain created by hot and cold stimuli were influenced by whether or not you swore while in contact with the stimulus (Stephens, Atkins, & Kingston, 2009). You could place some people's hands in a bucket of very cold water for a minute and ask them to rate how painful they thought the experience was on a scale of 1 to 10. You could then ask them to hold a hot potato and again measure their perception of pain. Half the participants are encouraged to shout profanities during the experiences. Imagine I was a participant in the swearing group. You would have a single row representing my data, so there would be a different column for my name, the group I was in, my pain perception for cold water and my pain perception for a hot potato: Andy, Swearing Group, 7, 10.

The column with the information about the group to which I was assigned is a grouping variable: I can belong to either the group that could swear or the group that was forbidden, but not both. This variable is a between-group or independent measure (different people belong to different groups). In SPSS we typically represent group membership with numbers, not words, but assign labels to those numbers. As such, group membership is represented by a single column in which the group to which the person belonged is defined using a number (see

Section 4.6.5). For example, we might decide that if a person was in the swearing group we assign them the number 1, and if they were in the non-swearing group we assign them a 0. We then assign a *value label* to each number, which is text that describes what the number represents. To enter group membership, we would input the numbers we have decided to use into the data editor, but the value labels remind us which groups those numbers represent (see Section 6.10.4).

The two pain scores make up a repeated measure because all of the participants produced a score after contact with a hot and cold stimulus. Levels of this variable (see SPSS Tip 4.2) are entered in separate columns (one for pain from a hot stimulus and one for pain from a cold stimulus).

Figure 4.8 The variable view of the SPSS Data Editor
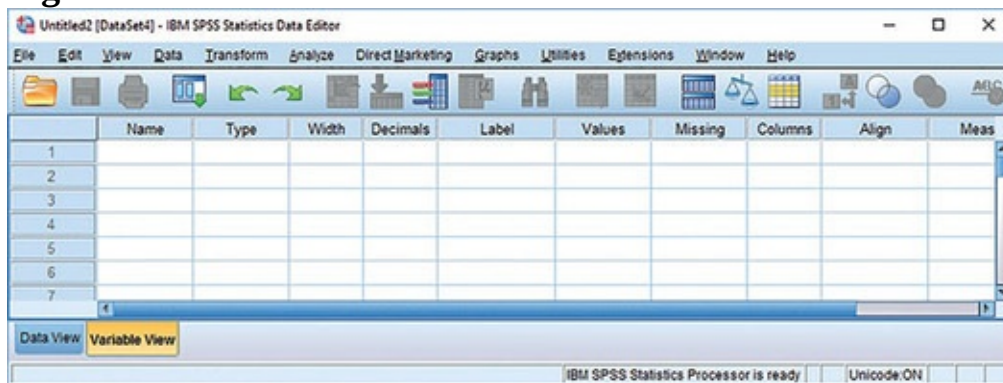

SPSS Tip 4.2 Wide format data entry

When using the wide format, there is a simple rule: data from different things go in different rows of the data editor, whereas data from the same things go in different columns of the data editor. As such, each person (or mollusc, goat, organization, or whatever you

have measured) is represented in a different row. Data within each person (or mollusc, etc.) go in different columns. So, if you've prodded your mollusc, or human, several times with a pencil and measured how much it twitches as an outcome, then each prod will be represented by a column.

In experimental research this means that variables measured with the same participants (a repeated measure) should be represented by several columns (each column representing one level of the repeated-measures variable). However, any variable that defines different groups of things (such as when a between-group design is used and different participants are assigned to different levels of the independent variable) is defined using a single column. This idea will become clearer as you learn about how to carry out specific procedures.
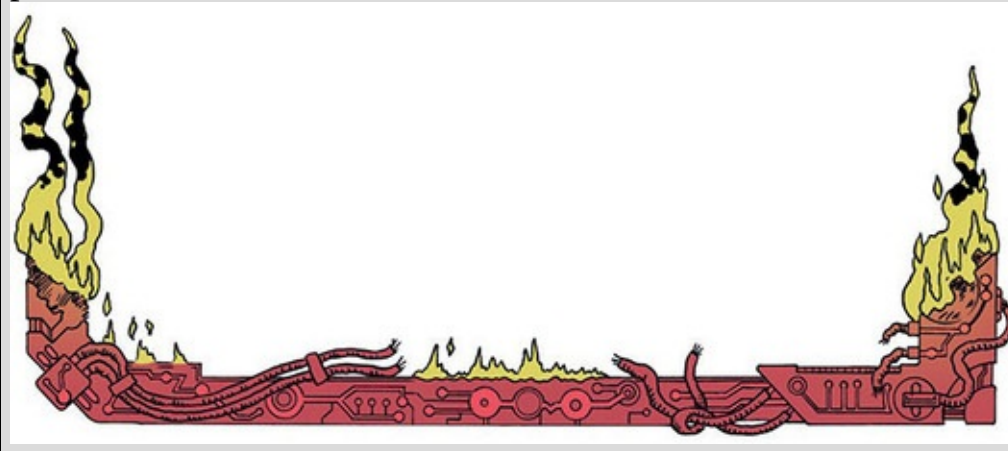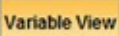
The data editor is made up of lots of *cells*, which are boxes in which data values can be placed. When a cell is active, it becomes highlighted in orange (as in Figure 4.3). You can move around the data editor, from cell to cell, using the arrow keys ← ↑ ↓ → (on the right of the keyboard) or by clicking the mouse on the cell that you wish to activate. To enter a number into the data editor, move to the cell in which you want to place the data value, type the value, then press the appropriate arrow button for the direction in which you wish to move. So, to enter a row of data, move to the far left of the row, type the first value and then press → (this process inputs the value and moves you into the next cell on the right).

4.6.2 The variable view

Before we input data into the data editor, we need to create the variables using

the variable view. To access this view click the 'Variable View' tab at the bottom of the data editor (Data View Variable View); the contents of the window will change (see [Figure 4.8](#)).

Every row of the variable view represents a variable, and you set characteristics of each variable by entering information into the following labelled columns (play around, you'll get the hang of it):

| Name	Enter a name in this column for each variable. This name will appear at the top of the corresponding column in the data view, and helps you to identify variables. You can more or less write what you like, but there are certain symbols you can't use (mainly those that have other uses in SPSS such as +, −, &), and you can't use spaces. (I find 'hard spaces' useful, where you use an underscore instead of a space; for example Andy_Field instead of Andy Field.) If you use a character that SPSS doesn't like, you'll get an error message saying that the variable name is invalid when you try to move out of the cell.
Type	Variables can contain different types of data. Mostly you will use the default **numeric variables** (the variable contains numbers). Other types are **string variables** (strings of letters), **currency variables** (i.e., £s, $s, €s) and **date variables** (e.g., 21-06-1973). So, for example, if you want to type in a person's name, you would need to change the variable type from numeric to string.
Width	By default, SPSS sets up new variables as *numeric* and storing 8 digits/characters. For numeric variables 8 digits is usually fine (unless you have very large numbers), but for string variables you can't write a lot in only 8 characters so you might want to increase the width to accommodate the length of text you intend to use. Width differs from Columns (see below) in that it affects what is stored in the variable rather than what is displayed in the data editor.
Decimals	By default, 2 decimal places are displayed. (If you don't change this option then when you enter whole numbers SPSS adds a decimal place with two zeros after it, which can be disconcerting.) If you want to change the number of decimal places for a given variable, replace the 2 with a new value or increase or decrease the values using ⬍.
Label	The variable Name (see above) has some restrictions on characters, and you also wouldn't want to use huge long names at the top of your columns because they become hard to read. Therefore, you can write a longer variable description in this column. This may seem pointless, but is one of the best habits you can get into (see SPSS Tip 4.3).
Values	This column is for assigning numbers to represent groups of people (see Section 4.6.5).
Missing	This column is for assigning numbers to missing data (see Section 4.6.7).
Columns	Enter a number into this column to determine the width of the column (i.e., how many characters are displayed in the column). The information entered under Columns differs from that for Width (see above) which determines the width of the variable itself. So, you could have a variable with a width of 10 characters but by setting the *column* width to 8 you would see only 8 of the 10 characters of the variable in the data editor. It can be useful to increase the column width if you have a string variable (Section 4.6.3) or a variable with value labels (Section 4.6.5) that exceed 8 characters.
Align	You can use this column to select the alignment of the data in the corresponding column of the data editor. You can choose to align the data to the ≣ Left , ≣ Right or ≣ Center .
Measure	Use this column to define the scale of measurement for the variable as ⬡ Nominal, ⊞ Ordinal or ✏ Scale (Section 1.6.2).
Role	SPSS Statistics has some procedures that attempt to run analyses automatically without you needing to think about what you're doing (one example is *Analyze* ▶ *Regression* ▶ ⬚ Automatic Linear Modeling...). To think on your behalf, SPSS needs to know whether a variable is a predictor (⬊ Input), an outcome (◉ Target), both (◉ Both, although I'm not sure how that works out in practice), a variable that splits the analysis by different groups (⬚ Split), a variable that selects out part of the data (⬚ Partition), or a variable that has no predefined role (⊘ None). It's rarely a good idea to let a computer do your thinking for you, so I'm not a fan of the sorts of automated procedures that take advantage of these role assignments (they have their place, but that place is not in this book). Therefore, I'm not going to mention roles again.

Let's use the variable view to create some variables. Imagine we were interested

in looking at the differences between lecturers and students. We took a random sample of five psychology lecturers from the University of Sussex and five psychology students and then measured how many friends they had, their weekly alcohol consumption (in units), their yearly income and how neurotic they were (higher score is more neurotic). These data are in Table 4.1.

4.6.3 Creating a string variable

The first variable in Table 4.1 is the name of the lecturer/student. This variable is a *string variable* because it consists of names (which are strings of letters). To create this variable in the variable view:

1. Click in the first white cell in the column labelled *Name*.
2. Type the word 'Name'.
3. Move from this cell using the arrow keys on the keyboard (you can also just click in a different cell, but this is a very slow way of doing it).

Well done, you've just created your first variable. Notice that once you've typed a name, SPSS creates default settings for the variable (such as assuming it's numeric and assigning 2 decimal places). However, we don't want a numeric variable (i.e., numbers), we want to enter people's names, so we need a *string* variable, so we have to change the variable type. Move into the column labelled **Type** using the arrow keys on the keyboard. The cell will now look like this **Numeric** ···. Click to activate the *Variable Type* dialog box. By default, the numeric variable type is selected (⊙ **Numeric**) – see the top of Figure 4.9. To change the variable to a string variable, click ⊙ **String** (bottom left of Figure 4.9). Next, if you need to enter text of more than 8 characters (the default width), then change this default value to a number reflecting the maximum number of characters that you will use for a given case of data. Click **OK** to return to the variable view.

SPSS Tip 4.3 Naming variables
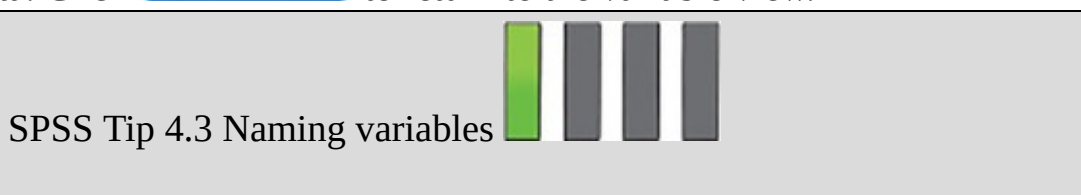
'Surely it's a waste of my time to type in long names for my variables when I've already given them a short one?' I hear you ask. I can understand why it would seem so, but as you go through university or your career accumulating data files, you will be grateful that you did. Imagine you had a variable called 'number of times I wanted to bang the desk with my face during Andy Field's statistics lecture'; then you might have named the column in SPSS 'nob' (short for number of bangs). You thought you were smart coming up with such a succinct label. If you don't add a more detailed label, SPSS uses this variable name in all the output from an analysis. Fast forward a few months when you need to look at your data and output again. You look at the 300 columns all labelled things like 'nob', 'pom', 'p', 'lad', 'sit' and 'ssoass' and think to yourself, 'What does "nob" stand for? Which of these variables relates to face-butting a desk? Imagine the chaos you could get into if you always used acronyms for the variable and had an outcome of 'wait at news kiosk' for a study about queuing. I deal with many data sets with variables called things like 'sftg45c', and if they don't have proper variable labels, then I'm in all sorts of trouble. Get into a good habit and label your variables.

Next, because I want you to get into good habits, move to the cell in the `Label` column and type a description of the variable, such as 'Participant's First Name'. Finally, we can specify the scale of measurement for the variable (see [Section 1.6.2](#)) by going to the column labelled *Measure* and selecting 🔴 Nominal, 📊 Ordinal or 📏 Scale from the drop-down list. In the case of a string variable, it represents a description of the case and provides no information about the order of cases, or the magnitude of one case compared to another. Therefore, select 🔴 Nominal.
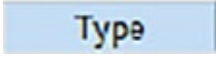
Table 4.1 Some data with which to play

Name	Birth Date	Job	No. of Friends	Alcohol (Units)	Income (p.a.)	Neuroticism
Ben	03-Jul-1977	Lecturer	5	10	20,000	10
Martin	24-May-1969	Lecturer	2	15	40,000	17
Andy	21-Jun-1973	Lecturer	0	20	35,000	14
Paul	16-Jul-1970	Lecturer	4	5	22,000	13
Graham	10-Oct-1949	Lecturer	1	30	50,000	21
Carina	05-Nov-1983	Student	10	25	5,000	7
Karina	08-Oct-1987	Student	12	20	100	13
Doug	16-Sep-1989	Student	15	16	3000	9
Mark	20-May-1973	Student	12	17	10,000	14
Zoë	12-Nov-1984	Student	17	18	10	13

Once the variable has been created, return to the data view by clicking on the 'Data View' tab at the bottom of the data editor (Data View Variable View). The contents of the window will change, and notice that the first column now has the label *Name*. We can enter the data for this variable in the column underneath. Click the white cell at the top of the column labelled *Name* and type the first name, 'Ben'. To register this value in this cell, move to a different cell and because we are entering data down a column, the most sensible way to do this is to press the ↓ key on the keyboard. This action moves you down to the next cell, and the

word 'Ben' should appear in the cell above. Enter the next name, 'Martin', and then press ↓ to move down to the next cell, and so on.
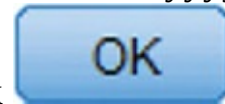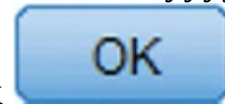
4.6.4 Creating a date variable

The second column in our table contains dates (birth dates to be exact). To create a date variable, we more or less repeat what we've just done. First, move back to the variable view using the tab at the bottom of the data editor (Data View Variable View). Move to the cell in row 2 of the column labelled *Name* (under the previous variable you created). Type the word 'Birth_Date' (note that I have used a hard space to separate the words). Move into the column labelled Type using the → key on the keyboard (doing so creates default settings in the other columns). As before, the cell you have moved into will indicate the default of Numeric , and to change this we click to activate the *Variable Type* dialog box, and click ◉ Date (bottom right of Figure 4.9). On the right of the dialog box is a list of date formats, from which you can choose your preference; being British, I am used to the day coming before the month and have chosen dd-mmm-yyyy (i.e., 21-Jun-1973), but Americans, for example, more often put the month before the date and so might select mm/dd/yyyy (06/21/1973). When you have selected a date format, click OK to return to the variable view. Finally, move to the cell in the column labelled *Label* and type 'Date of Birth'.
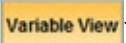
Once the variable has been created, return to the data view by clicking on the 'Data View' tab (Data View Variable View). The second column now has the label *Birth_Date*; click the white cell at the top of this column and type the first value, 03-Jul-1977. To register this value in this cell, move down to the next cell by pressing the ↓ key. Now enter the next date, and so on.

Figure 4.9 Defining numeric, string and date variables

Defining a numeric variable
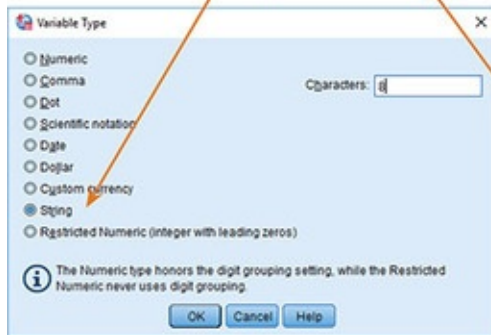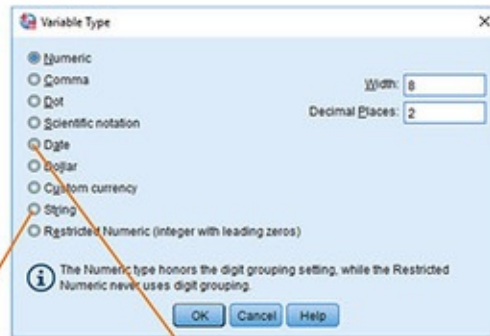
Defining a string variable
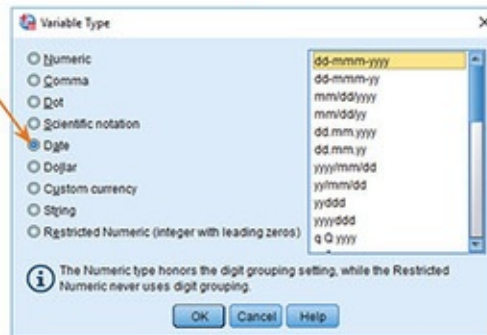
Defining a date variable

4.6.5 Creating coding variables

I've mentioned coding or grouping variables briefly already; they use numbers to represent different groups or categories of data. As such, a coding variable is *numeric*, but because the numbers represent names its variable type is
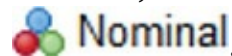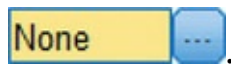
Nominal. The groups of data represented by coding variables could be levels of a treatment variable in an experiment (an experimental group or a control group), different naturally occurring groups (men or women, ethnic groups, marital status, etc.), different geographic locations (countries, states, cities, etc.), or different organizations (different hospitals within a healthcare trust, different schools in a study, different companies).

In experiments that use an independent design, coding variables represent predictor (independent) variables that have been measured between groups (i.e., different entities were assigned to different groups). We do not, generally, use this kind of coding variable for experimental designs where the independent variable was manipulated using repeated measures (i.e., participants take part in all experimental conditions). For repeated-measures designs we typically use different columns to represent different experimental conditions.

Think back to our swearing and pain experiment. This *was* an independent design because we had two groups representing the two levels of our

independent variable: one group could swear during the pain tasks, the other could not. Therefore, we can use a coding variable. We might assign the experimental group (swearing) a code of 1 and the control group (no swearing) a code of 0. To input these data you would create a variable (which you might call **group**) and type the value 1 for any participants in the experimental group, and 0 for any participant in the control group. These codes tell SPSS that the cases that have been assigned the value 1 should be treated as belonging to the same group, and likewise for the cases assigned the value 0. The codes you use are arbitrary because the numbers themselves won't be analysed, so although people typically use 0, 1, 2, 3, etc., if you're a particularly arbitrary person feel free to code one group as 616 and another as 11 and so on.

We have a coding variable in our data that describes whether a person was a lecturer or student. To create this coding variable, we follow the same steps as before, but we will also have to record which numeric codes are assigned to which groups. First, return to the variable view (Data View Variable View) if you're not already in it and move to the cell in the third row under the column labelled *Name.* Type a name (let's call it **Group**). I'm still trying to instil good habits, so move along the third row to the column called *Label* and give the variable a full description such as, 'Is the person a lecturer or a student?' To define the group codes, move along the row to the column labelled Values . The cell will

indicate the default of None . Click --- to access the *Value Labels* dialog box (see Figure 4.10).

The *Value Labels* dialog box is used to specify group codes. First, click in the white space next to where it says *Value* (or press *Alt* and U at the same time) and type in a code (e.g., 1). The second step is to click in the white space below, next to where it says *Label* (or press *Tab,* or *Alt* and L at the same time) and type in an appropriate label for that group. In Figure 4.10 I have already defined a code of 1 for the lecturer group, and then I have typed in 2 as a code and given this a label of *Student.* To add this code to the list click Add . When you have defined all your coding values you might want to check for spelling mistakes in the value labels by clicking Spelling... . To finish, click OK ; if you

do this before you have clicked ![Add] to register your most recent code in the list, SPSS displays a warning that any 'pending changes will be lost'. This message is telling you to go back and click ![Add] before continuing. Finally, coding variables represent categories and so the scale of measurement is nominal (or ordinal if the categories have a meaningful order). To specify this level of measurement, go to the column labelled *Measure* and select ![Nominal] (or ![Ordinal] if the groups have a meaningful order) from the drop-down list.

Figure 4.10 Defining coding variables and their values
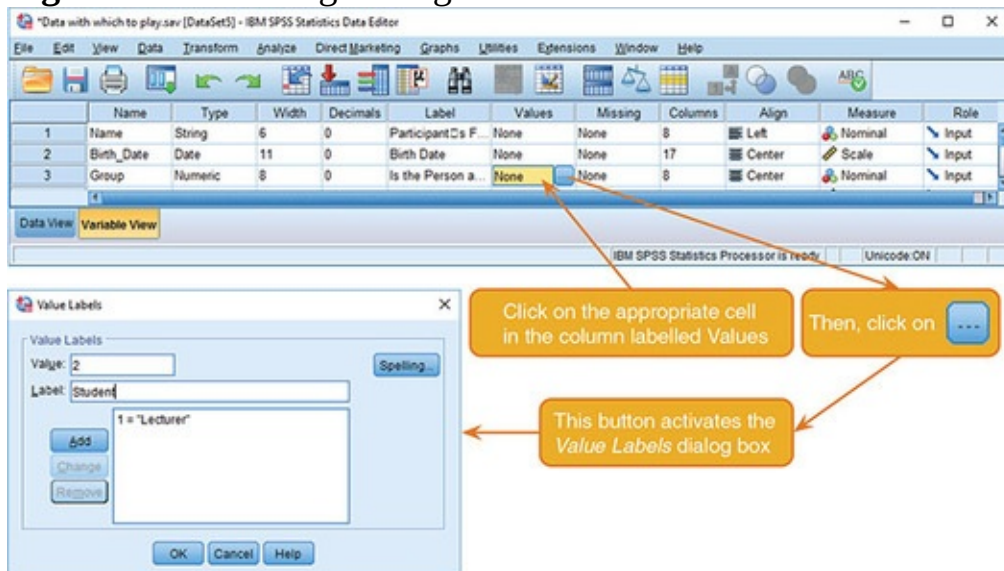
Having defined your codes, switch to the data view and for each participant type the numeric value that represents their group membership into the column labelled *Group*. In our example, if a person was a lecturer, type '1', but if they were a student then type '2' (see SPSS Tip 4.4). SPSS can display either the numeric codes or the value labels that you assigned to them, and you can toggle between the two states by clicking ![A 1] (see Figure 4.11). Figure 4.11 shows how the data should be arranged: remember that each row of the data editor represents data from one entity: the first five participants were lecturers, whereas participants 6–10 were students.
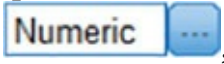
4.6.6 Creating a numeric variable

Our next variable is **Friends**, which is numeric. Numeric variables are the easiest ones to create because they are the default format in SPSS. Move back to the variable view using the tab at the bottom of the data editor (Data View Variable View). Go to the cell in row 4 of the column labelled *Name* (under the previous variable you created). Type the word 'Friends'. Move into the column labelled Type using the → key on the keyboard. As with the previous variables we have created, SPSS has assumed that our new variable is Numeric ..., and because our variable *is* numeric we don't need to change this setting.

The scores for the number of friends have no decimal places (unless you are a very strange person indeed, you can't have 0.23 of a friend). Move to the Decimals column and type '0' (or decrease the value from 2 to 0 using) to tell SPSS that you don't want to display decimal places.

Let's continue our good habit of naming variables and move to the cell in the column labelled *Label* and type 'Number of Friends'. Finally, number of friends is measured on the ratio scale of measurement (see Section 1.6.2) and we can specify this by going to the column labelled *Measure* and selecting Scale from the drop-down list (this will have been done automatically, but it's worth checking).

Figure 4.11 Coding values in the data editor with the value labels switched off and on

Value labels on   Value labels off

SPSS Tip 4.4 Copying and pasting into the data editor and variable viewer

Often (especially with coding variables), you need to enter the same value lots of times into the data editor. Similarly, in the variable view, you might have a series of variables that all have the same value labels (e.g., variables representing questions on a questionnaire might all have value labels of 0 = never, 1 = sometimes, 2 = always to represent responses to those questions). Rather than typing the same number lots of times, or entering the same value labels multiple times, you can use the copy and paste functions to speed things up. All you need to do is to select the cell containing the information that you want to copy (whether that is a number or text in the data view, or a set of value labels or another characteristic within the variable view) and click with the right mouse button to activate a menu within

which you can click (with the left mouse button) on *Copy* (top of Figure 4.12). Next, highlight any cells into which you want to place what you have copied by dragging the mouse over them while holding down the left mouse button. These cells will be highlighted in orange. While the pointer is over the highlighted cells, click with the right mouse button to activate a menu from which you should click *Paste* (bottom left of Figure 4.12). The highlighted cells will be filled with the value that you copied (bottom right of Figure 4.12). Figure 4.12 shows the process of copying the value '1' and pasting it into four blank cells in the same column.

Figure 4.12 Copying and pasting into empty cells
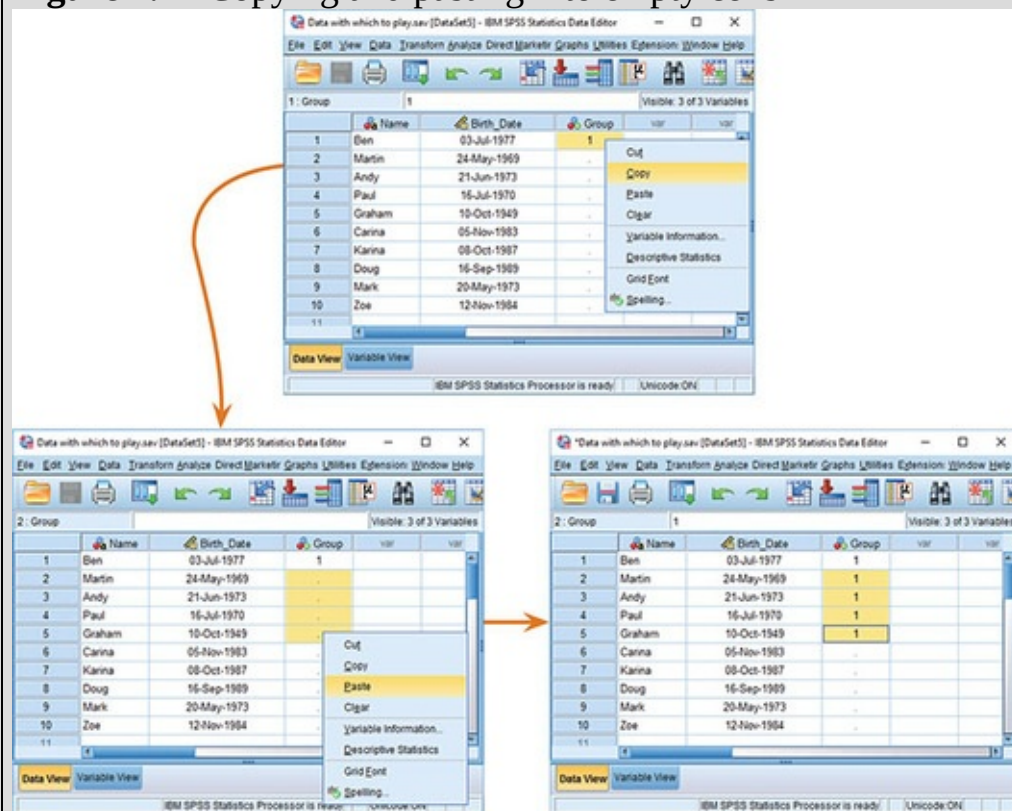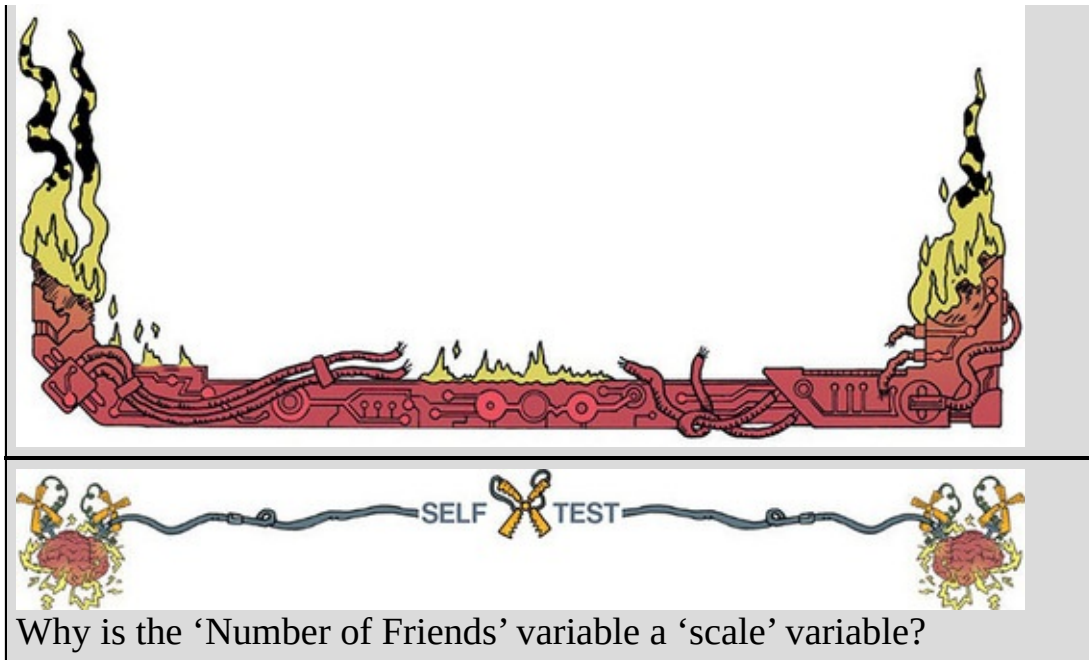
Once the variable has been created, you can return to the data view by clicking on the 'Data View' tab at the bottom of the data editor (Data View Variable View). The contents of the window will change, and you'll notice that the fourth column now has the label *Friends*. To enter the data, click the white cell at the top of the column labelled *Friends* and type the first value, 5. Because we're entering scores down the column the most sensible way to record this value in this cell is to press the ↓ key on the keyboard. This action moves you down to the next cell, and the number 5 is stored in the cell above. Enter the next number, 2, and then press ↓ to move down to the next cell, and so on.
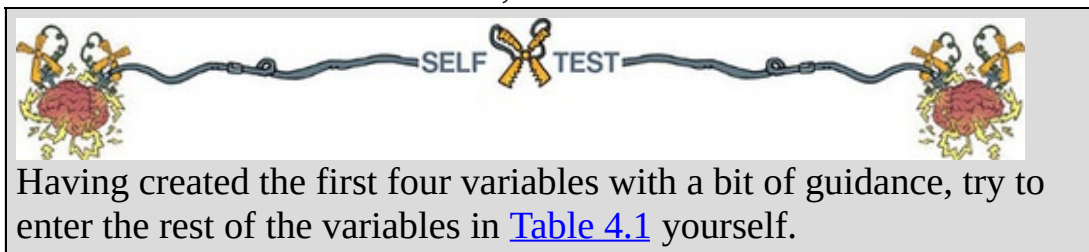
4.6.7 Missing values

Although we strive to collect complete sets of data, often scores are missing. Missing data can occur for a variety of reasons: in long questionnaires participants accidentally (or, depending on how paranoid you're feeling, deliberately to irritate you) miss out questions; in experimental procedures mechanical faults can lead to a score not being recorded; and in research on delicate topics (e.g., sexual behaviour) participants may exert their right not to

answer a question. However, just because we have missed out on some data for a participant, that doesn't mean that we have to ignore the data we do have (although it creates statistical difficulties). The simplest way to record a missing score is to leave the cell in the data editor empty, but it can be helpful to tell SPSS explicitly that a score is missing. We do this, much like a coding variable, by choosing a number to represent the missing data point. You then tell SPSS to treat that number as missing. For obvious reasons, it is important to choose a code that cannot also be a naturally occurring data value. For example, if we use the value 9 to code missing values and several participants genuinely scored 9, then SPSS will wrongly treat those scores as missing. You need an 'impossible' value, so people usually pick a score greater than the maximum possible score on the measure. For example, in an experiment in which attitudes are measured on a 100-point scale (so scores vary from 1 to 100) a good code for missing values might be something like 101, 999 or, my personal favourite, 666 (because missing values *are* the devil's work).

Labcoat Leni's Real Research 4.1 Gonna be a rock 'n' roll singer

Oxoby, R. J. (2008). *Economic Enquiry, 47*(3), 598–602.
AC/DC are one one of the best-selling hard rock bands in history, with around 100 million certified sales, and an estimated 200 million actual sales. In 1980 their original singer Bon Scott died of alcohol poisoning and choking on his own vomit. He was replaced by Brian Johnson, who has been their singer ever since.[5] Debate rages with unerring frequency within the rock music press over who is the better frontman. The conventional wisdom is that Bon Scott was better, although personally, and I seem to be somewhat in the minority here, I prefer Brian Johnson. Anyway, Robert Oxoby, in a playful paper, decided to put this argument to bed once and for all (Oxoby, 2008).

Well, until all that weird stuff with W. Axl Rose in 2016, which I'm trying to pretend didn't happen.
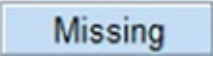
Using a task from experimental economics called the ultimatum game, individuals are assigned the role of either proposer or responder and paired randomly. Proposers are allocated $10 from which they have to make a financial offer to the responder (i.e., $2). The responder can accept or reject this offer. If the offer is rejected neither party gets any money, but if the offer is accepted the responder keeps the offered amount (e.g., $2), and the proposer keeps the original amount minus what they offered (e.g., $8). For half of the participants the song 'It's a long way to the top' sung by Bon Scott was playing in the background, for the remainder 'Shoot to thrill' sung by Brian Johnson was playing. Oxoby measured the offers made by proposers, and the minimum offers that responders accepted (called the minimum acceptable offer). He reasoned that people would accept lower offers and propose higher offers when listening to something they like (because of the 'feel-good factor' the music creates). Therefore, by comparing the value of offers made and the minimum acceptable offers in the two groups, he could see whether people have more of a feel-good factor when listening to Bon or Brian. The offers made (in $) are[6] as follows (there were 18 people per group):
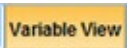
[6] These data are estimated from Figures 1 and 2 in the paper because I couldn't get hold of the author to get the original data files.

- Bon Scott group: 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5
- Brian Johnson group: 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5

Enter these data into the SPSS Data Editor, remembering to include value labels, to set the *measure* property, to give each variable a proper label, and to set the appropriate number of decimal places. Answers are on the companion website, and my version of how this file should look can be found in *Oxoby (2008) Offers.sav*.
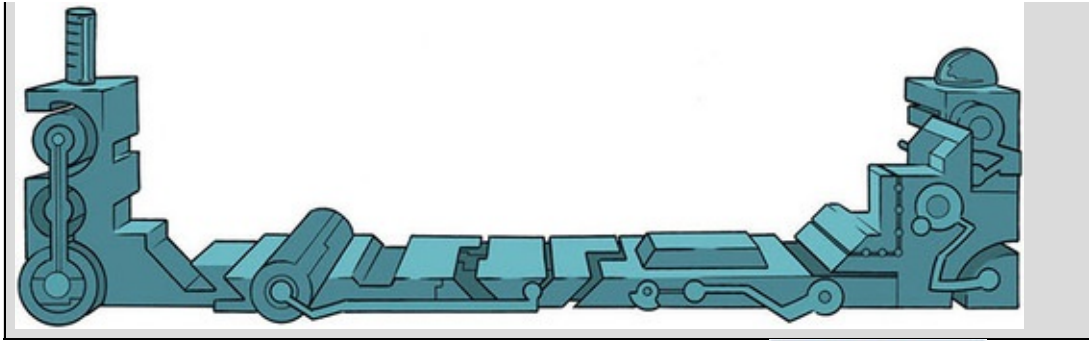
To specify missing values click in the column labelled **Missing** in the variable view (**Data View** **Variable View**) and then click to activate the *Missing Values* dialog box in Figure 4.13. By default, SPSS assumes that no missing values exist, but you can define them in one of two ways. The first is to select discrete values (by clicking on the radio button next to where it says *Discrete missing values*), which are single values that represent missing data. SPSS allows you to specify up to three values to represent missing data. The reason why you might choose to have several numbers to represent missing values is that you can assign a different meaning to each discrete value. For example, you could have the number 8 representing a response of 'not applicable', a code of 9 representing a 'don't know' response, and a code of 99 meaning that the participant failed to give any response. SPSS treats these values in the same way (it ignores them), but different codes can be helpful to remind you of why a particular score is missing. The second option is to select a range of values to represent missing data and this is useful in situations in which it is necessary to exclude data falling between two points. So, we could exclude all scores between 5 and 10. With this last option you can also (but don't have to) specify one discrete value.

4.7 Importing data

We can import data into SPSS from other software packages such as Microsoft Excel, R, SAS, and Systat by using the *File* *Import Data* menu and selecting the corresponding software from the list (Figure 4.14). If you want to

import from a package that isn't listed (e.g., R or Systat), then export the data from these packages as tab-delimited text data (*.txt* or *.dat*) or comma-separated values (*.csv*) and select the *Text Data* or *CSV Data* options in the menu.

Figure 4.13 Defining missing values
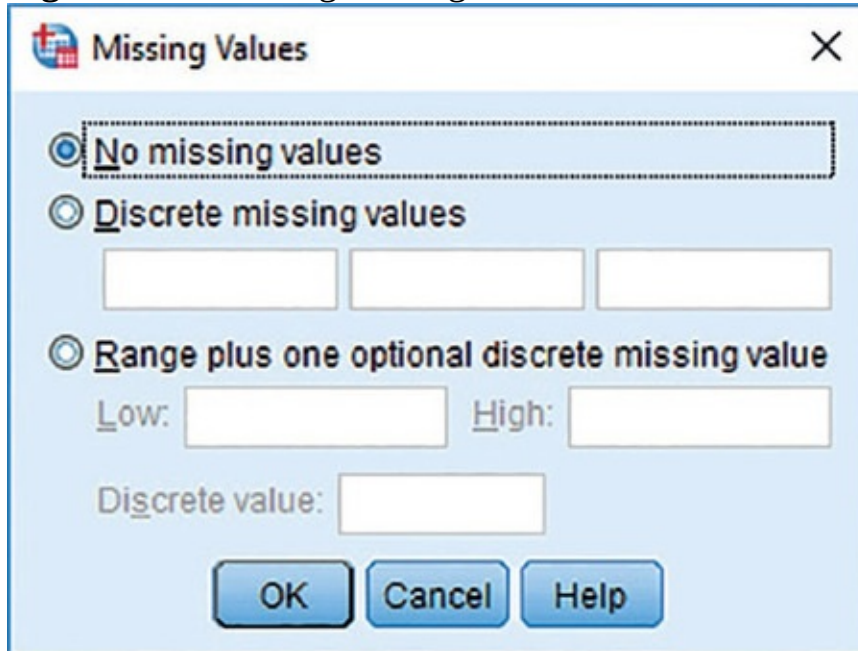
Oditi's Lantern Entering data

'I, Oditi, believe that the secrets of life have been hidden in a complex numeric code. Only by "analysing" these sacred numbers can we reach true enlightenment. To crack the code I must assemble thousands of followers to analyse and interpret these numbers (it's a bit like the chimps and typewriters theory). I need you to follow me. To spread the numbers to other followers you must store them in an easily distributable format called a "data file". You, my follower, are loyal and loved, and to assist you my lantern displays a tutorial on how to do this.'
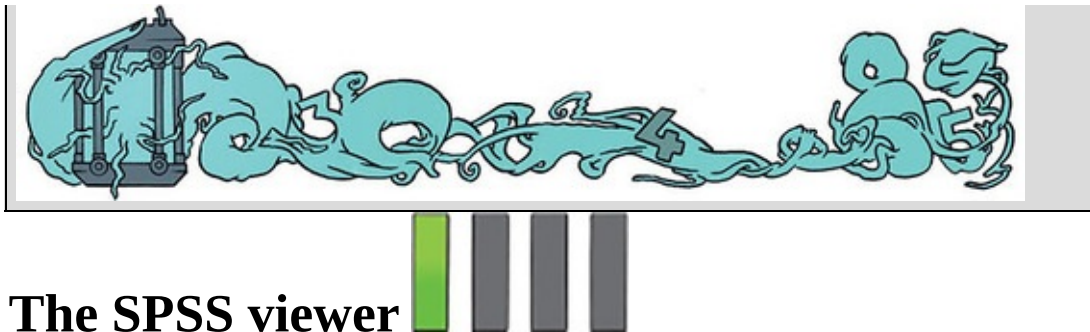
4.8 The SPSS viewer

The *SPSS viewer* appears in a different window than the data editor and displays the output of any procedures in SPSS: tables of results, graphs, error messages and pretty much everything you could want, except for photos of your cat. Although the *SPSS viewer* is all-singing and all-dancing, my prediction in previous editions of this book that it will one day include a tea-making facility have not come to fruition (IBM, take note ☺). Figure 4.15 shows the viewer. On the right there is a large space in which all output is displayed. Graphs (Section 5.9) and tables displayed here can be edited by double-clicking on them. On the left, is a tree diagram of the output. This tree diagram provides an easy way to access parts of the output, which is useful when you have conducted tonnes of analyses. The tree structure is self-explanatory: every time you do something in SPSS (such as drawing a graph or running a statistical procedure), it lists this procedure as a main heading.

In Figure 4.15, I ran a graphing procedure followed by a univariate analysis of variance (ANOVA), and these names appear as main headings in the tree diagram. For each procedure there are subheadings that represent different parts of the analysis. For example, in the ANOVA procedure, which you'll learn more about later in the book, there are sections such as *Tests of Between-Subjects Effects* (this is the table containing the main results). You can skip to any one of these sub-components by clicking on the appropriate branch of the tree diagram. So, if you wanted to skip to the between-groups effects, you would move the on-screen arrow to the left-hand portion of the window and click where it says *Tests of Between-Subjects Effects*. This action will highlight this part of the output in the main part of the viewer (see SPSS Tip 4.5).

Figure 4.14 The *Import Data* menu

File   Edit   View   Data   Transform   Analyze

New ▶
Open ▶
Import Data ▶   Database ▶
Close   Ctrl+F4   Excel...
Save   Ctrl+S   CSV Data...
Save As...   Text Data...
Save All Data   SAS...
Export ▶   Stata...
Mark File Read Only   dBase...
Rename Dataset...   Lotus...
Display Data File Information ▶   SYLK...
Cache Data...   Cognos TM1...
Stop Processor   Ctrl+Period   Cognos Business Intelligence...
Collect Variable Information
Switch Server...
Repository ▶
Print Preview
Print...   Ctrl+P
Welcome Dialog...
Recently Used Data ▶
Recently Used Files ▶
Exit

Figure 4.15 The SPSS viewer
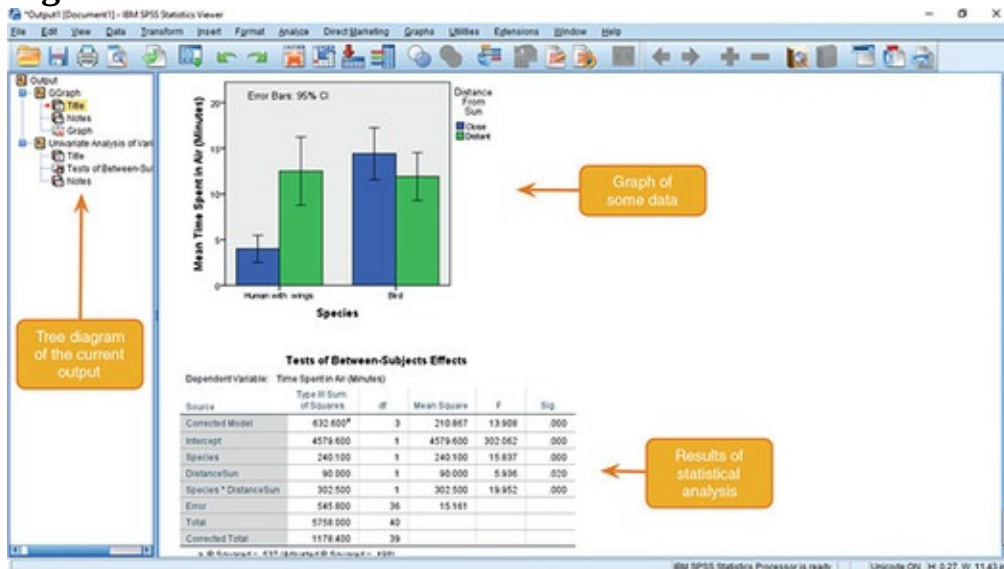
Oditi's Lantern Importing data into SPSS

'I, Oditi, have become aware that some of the sacred numbers that hide the secrets of life are contained within files other than those of my own design. We cannot afford to miss vital clues that lurk among these rogue files. Like all good cults, we must convert all to our cause, even data files. Should you encounter one of these files, you must convert it to the SPSS format. My lantern shows you how.'

Oditi's Lantern Editing tables

'I, Oditi, impart to you, my loyal posse, the knowledge that SPSS will conceal the secrets of life within tables of output. Like the author of this book's personality, these tables appear flat and lifeless; however, if you give them a poke they have hidden depths. Often you will need to seek out the hidden codes within the tables. To do this, double-click on them. This will reveal the "layers" of the table. Stare into my lantern and find out how.'

SPSS Tip 4.5 Printing and saving the planet

Rather than printing all of your SPSS output, you can help the planet by printing only a selection. Do this by using the tree diagram in the SPSS viewer to select parts of the output for printing. For example, if you decided that you wanted to print a particular graph, click on the word *Graph* in the tree structure to highlight the graph in the output. Then, in the *Print* menu you can print just the selected part of the output ([Figure 4.16](#)). Note that if you click a main heading (such as *Univariate Analysis of Variance*) SPSS will highlight all the subheadings under that heading, which is useful for printing all the output from a single statistical procedure.

Figure 4.16 Printing only the selected parts of SPSS output
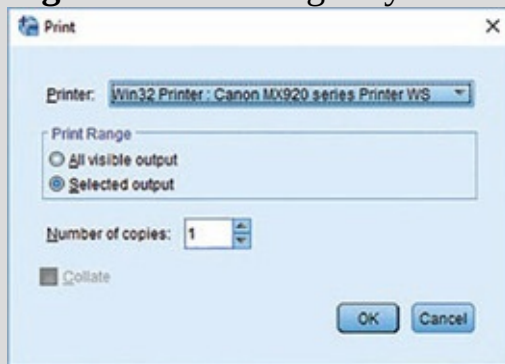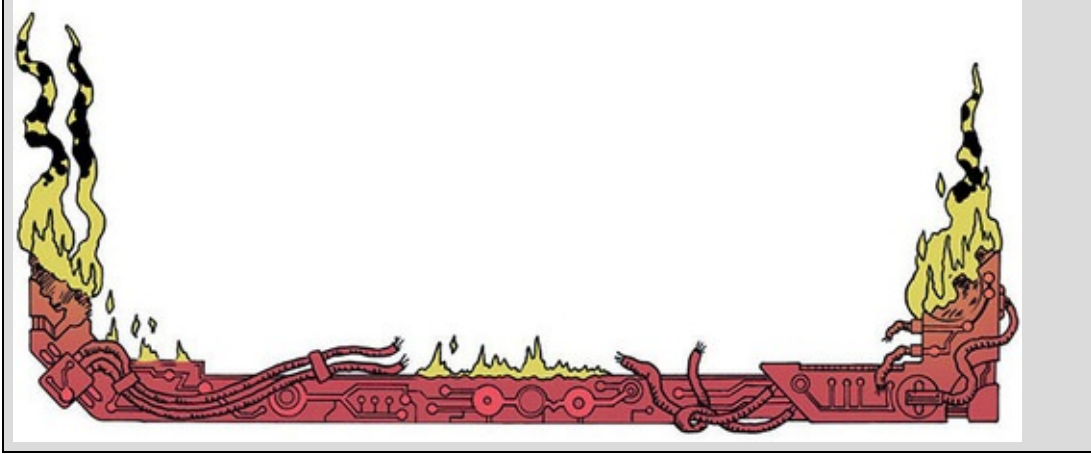
Some of the icons in the viewer are the same as those for the data editor (so refer back to our earlier list), but others are unique.

In the viewer, this icon activates a dialog box for printing SPSS output (see SPSS Tip 4.5).

This icon returns you to the data editor. I'm not sure what the big red star is all about.

This icon takes you to the last output in the viewer (it returns you to the last procedure you conducted).

This icon *promotes* the currently active part of the tree structure to a higher branch of the tree. For example, in Figure 4.15 the *Tests of Between-Subjects Effects* are a sub-component under the heading of *Univariate Analysis of Variance*. If we wanted to promote this part of the output to a higher level (i.e., to make it a main heading) then we'd click this icon.

This icon is the opposite of the above: it *demotes* parts of the tree structure. For example, in Figure 4.15, if we didn't want the *Univariate Analysis of Variance* to be a unique section we could select this heading and click this icon to demote it so that it becomes part of the previous heading (the *Graph* heading). This icon is useful for combining parts of the output relating to a specific research question.

This icon collapses parts of the tree structure: it hides the sub-components under a selected heading. For example, in Figure 4.15, if we selected the heading *Univariate Analysis of Variance* and pressed this icon, all the subheadings would disappear. The sections that disappear from the tree structure don't disappear from the output itself; the tree structure is merely condensed. This option is useful when you have conducted lots of analyses and the tree diagram has become very complex.

This icon expands any collapsed sections. By default, main headings are displayed in the tree diagram in their expanded form. If you have opted to collapse part of the tree diagram (using the icon above) then use this icon to undo your dirty work.

This icon and the following one show or hide parts of the output. If you select part of the output in the tree diagram and click this icon, that part of the output disappears. Some procedures produce copious amounts of output, so this icon can be useful to hide the less relevant parts so you can differentiate the wood from the trees. The output isn't erased, just hidden from view. This icon is like the *collapse* icon above, except that it affects the output rather than the tree structure.

This icon undoes the work of the previous one: if you have hidden part of the output from view, then clicking this icon makes it reappear. By default, this icon is not active because none of the output is hidden, but it becomes active if you use the icon above to hide stuff.

Given this icon looks like it has a slot in which to insert a CD, I had high hopes for it activating some blistering thrash metal, which, as everyone knows, is the best music to have on when doing stats. I was disappointed to discover that rather than blasting out Anthrax at high volume it instead inserts a new heading into the tree diagram. For example, if you had several statistical tests that related to one of many research questions you could insert a main heading and demote the headings of the relevant analyses to fall under this new heading.

Assuming you had done the above, you can use this icon to provide your new heading with a title. The title you type appears in the output. So, you might have a heading like 'Hypothesis 1: thrash metal music helps you to understand stats' which tells you that the analyses under this heading relate to your first hypothesis.

This final icon is used to place a text box in the output window. You can type anything into this box. In the context of the previous two icons, you might use a text box to explain your hypothesis (e.g., 'Hypothesis 1 is that thrash metal music helps you to understand stats. This hypothesis stems from research showing that a 10-minute blast of classical music can enhance learning and memory (Rauscher, Shaw, & Ky, 1993). Thrash metal music has similar rhythmic and melodic complexity to classical music, so why not?').

Oditi's Lantern The SPSS viewer window

'I, Oditi, believe that by "analysing" the sacred numbers we can find the answers to life. I have given you the tools to spread these numbers far and wide, but to interpret these numbers we need "the viewer". The viewer is like an X-ray that reveals what is beneath the raw numbers. Use the viewer wisely, my friends, because if you stare long enough you will see your very soul. Stare into my lantern and see a tutorial on the viewer.'

SPSS Tip 4.6 Funny numbers

SPSS sometimes reports numbers with the letter 'E' placed in the mix just to confuse you. For example, you might see a value such as 9.612 E−02. Many students find this notation confusing. This notation means 9.61×10^{-2} , which might be a more familiar

notation, or could be even more confusing. Think of E−02 as meaning 'move the decimal place 2 places to the left', so 9.612 E−02 becomes 0.09612. If the notation reads 9.612 E−01, then that would be 0.9612, and 9.612 E−03 would be 0.009612. Conversely, E+02 (notice the minus sign has changed) means 'move the decimal place 2 places to the right', so, 9.612 E+02 becomes 961.2.
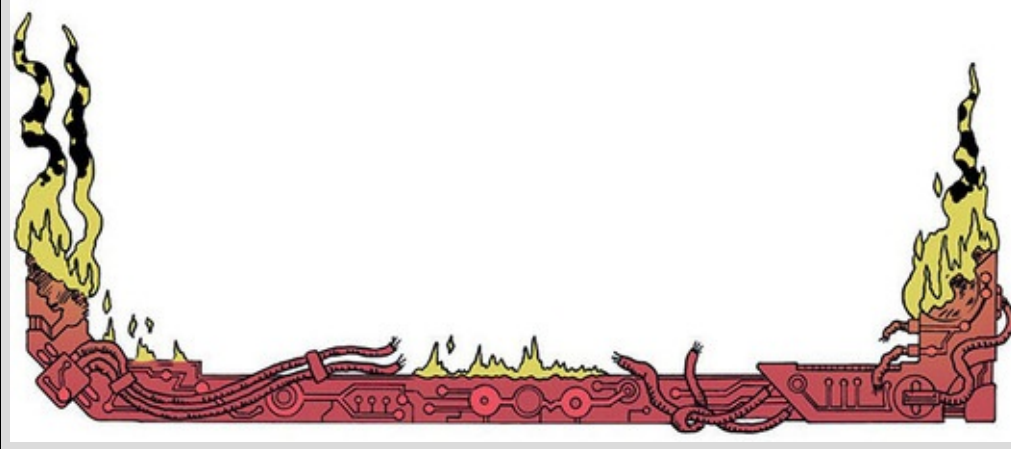
4.9 Exporting SPSS output

If you want to share your SPSS output with other people who don't have access to IBM SPSS Statistics, you have two choices: (1) export the output into a software package that they do have (such as Microsoft Word) or in the portable document format (PDF) that can be read by various free software packages; or (2) get them to install the free IBM SPSS **Smartreader** from the IBM SPSS website. The SPSS Smartreader is basically a free version of the viewer so you can view output but not run new analyses.

4.10 The syntax editor

I mentioned earlier that sometimes it's useful to use SPSS syntax. Syntax is a language of commands for carrying out statistical analyses and data manipulations. Most people prefer to do the things they need to do using dialog boxes, but SPSS syntax can be useful. No, really, it can. For one thing, there are things you can do with syntax that you can't do through dialog boxes (admittedly, most of these things are advanced, but I will periodically show you some nice tricks using syntax). The second benefit to syntax is if you carry out very similar analyses on data sets. In these situations, it is often quicker to do the analysis and save the syntax as you go along. Then you can adapt it to new data

sets (which is frequently quicker than going through dialog boxes. Finally, using syntax creates a record of your analysis, and makes it reproducible, which is an important part of engaging in open science practices ([Section 3.6](#)).

Oditi's Lantern Exporting SPSS output

'That I, the almighty Oditi, can discover the secrets within the numbers, they must spread around the world. But non-believers do not have SPSS, so we must send them a link to the IBM SPSS Smartreader. I have also given to you, my subservient brethren, a tutorial on how to export SPSS output into Word. These are the tools you need to spread the numbers. Go forth and stare into my lantern.'

To open a syntax editor window, like the one in [Figure 4.17](#), use *File* ▸ *New* ▸ 🖫 Syntax. The area on the right (the *command area*) is where you type syntax commands, and on the left is a navigation area (like the viewer window). When you have a large file of syntax commands the navigation area helps you find the bit of syntax that you need.

Like grammatical rules when we write, there are rules that ensure that SPSS 'understands' the syntax. For example, each line must end with a full stop. If you make a syntax error (i.e., break one of the rules), SPSS produces an error message in the viewer window. The messages can be indecipherable until you gain experience of translating them, but they helpfully identify the line in the syntax window in which the error occurred. Each line in the syntax window is numbered so you can easily find the line in which the error occurred, even if you don't understand what the error is! Learning SPSS syntax is time-consuming, so

in the beginning the easiest way to generate syntax is to use dialog boxes to specify the analysis you want to do and then click Paste (many dialog boxes have this button). Doing so pastes the syntax to do the analysis you specified in the dialog box. Using dialog boxes in this way is a good way to get a feel for syntax.

Once you've typed in your syntax you run it using the *Run* menu. *Run* will run all the syntax in the window, or you can highlight a selection of syntax using the mouse and select *Run* Selection (or click

in the syntax window) to process the selected syntax. You can also run the syntax a command at a time from either the current command (*Run* Step Through From Current*), or the beginning (*Run* Step Through From Start*). You can also process the syntax from the cursor to the end of the syntax window by selecting *Run* To End.

A final note. You can have multiple data files open in SPSS simultaneously. Rather than having a syntax window for each data file, which could get confusing, you can use one syntax window, but select the data file that you want to run the syntax commands on before you run them using the drop-down list DataSet1 .
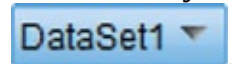
Figure 4.17 A syntax window with some syntax in it
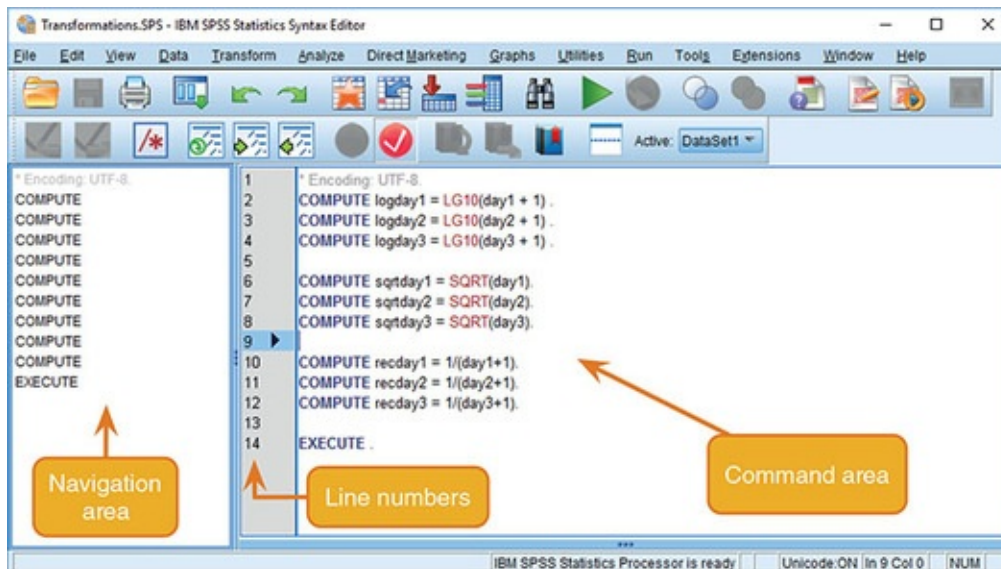
Oditi's Lantern Sin-tax

'I, Oditi, leader of the cult of undiscovered numerical truths, require my brethren to focus only on the discovery of those truths. To focus their minds I shall impose a tax on sinful acts. Sinful acts (such as dichotomizing a continuous variable) can distract from the pursuit of truth. To implement this tax, followers will need to use the sin-tax window. Stare into my lantern to see a tutorial on how to use it.'

4.11 Saving files

Most of you should be familiar with how to save files. Like most software, SPSS

has a save icon and you can use *File* or *File*

 Save as … or *Ctrl* + S (+ S on Mac OS). If the file hasn't been saved previously then initiating a save will open the *Save As* dialog box (see Figure 4.18). SPSS will save whatever is in the window that was active when you initiated the save; for example, if you are in the data editor when you initiate the save, then SPSS will save the data file (not the output or syntax). You use this dialog box as you would in any other software: type a name in the space next to where it says *File name*. If you have sensitive data, you can password encrypt it by selecting . By default, the file will be saved in an SPSS format, which has a *.sav* file extension for data files, *.spv* for viewer documents, and *.sps* for syntax files. Once a file has previously been saved, it can be saved again

(updated) by clicking on .

Figure 4.18 The *Save Data As* dialog box
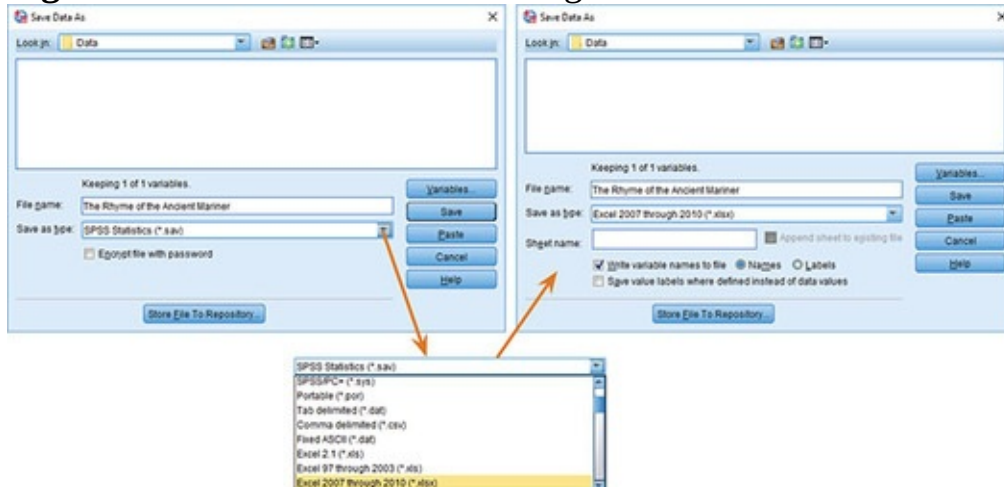
You can save data in formats other than SPSS. Three of the most useful are Microsoft Excel files (*.xls*, *.xlsx*), comma-separated values (*.csv*) and tab-delimited text (*.dat*). The latter two file types are plain text, which means that they can be opened by virtually any spreadsheet software you can think of

(including Excel, OpenOffice, Numbers, R, SAS, and Systat). To save your data file in of these formats (and others), click SPSS Statistics (*.sav) and select a format from the drop-down list (Figure 4.18). If you select a format other than SPSS, the ☐ Save value labels where defined instead of data values option becomes active. If you leave this option unchecked, coding variables (Section 4.6.5) will be exported as numeric values in the data editor; if you select it then coding variables will be exported as string variables containing the value labels. You can also choose to include the variable names in the exported file (usually a good idea) as either the *Names* at the top of the data editor columns, or the full *Labels* that you gave to the variables.
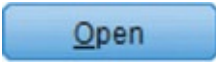
4.12 Opening files

This book relies on you working with data files that you can download from the companion website. You probably don't need me to tell you how to open these file, but just in case … To load a file into SPSS use the icon or select *File*   *Open*   and then Data to open a data file, Output to open a viewer file, or Syntax to open a syntax file. This process opens a dialog box (Figure 4.19), with which I'm sure you're familiar. Navigate to wherever you saved the file that you need. SPSS will list the files of the type you asked to open (so, data files if you selected Data). Open the file you want by either selecting it and clicking on Open, or double-clicking on the icon next to the file you want (e.g., double-clicking on

). If you want to open data in a format other than SPSS (*.sav*), then click SPSS Statistics (*.sav, *.zsav) to display a list of alternative file formats. Click the appropriate file type – Microsoft Excel file (*.xls*), text file (*.dat, *.txt,*) etc.), to list files of that type in the dialog box.

Figure 4.19 Dialog box to open a file
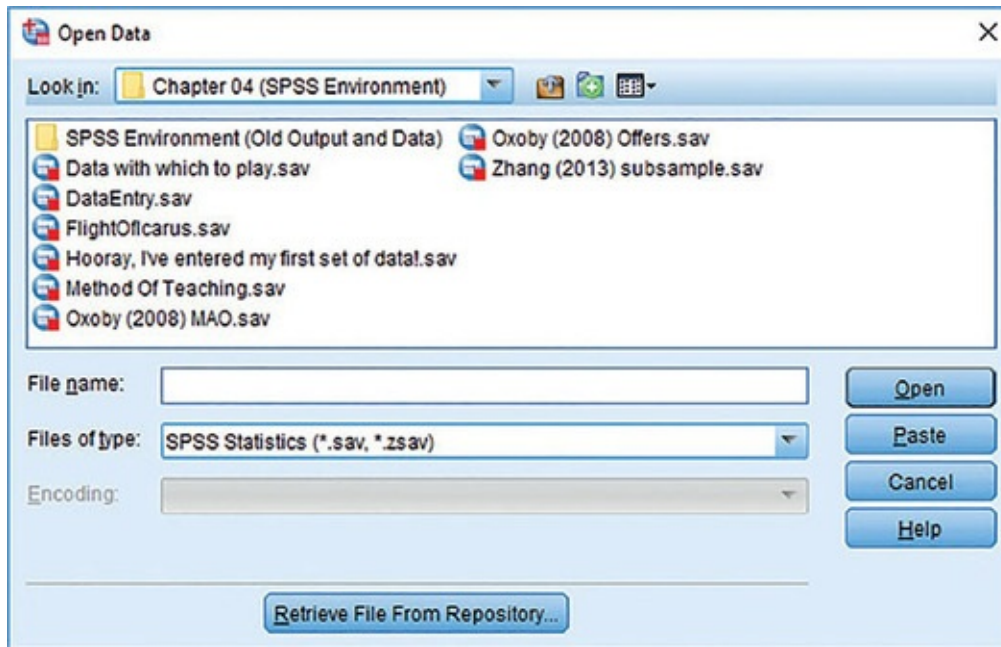
4.13 Extending IBM SPSS Statistics

IBM SPSS Statistics has some powerful tools for users to build their own functionality. For example, you can create your own dialog boxes and menus to run syntax that you may have written. SPSS Statistics also interfaces with a powerful open source statistical computing language called R (R Core Team, 2016). There are two extensions to SPSS that we use in this book. One is a tool called *PROCESS* and the other is the *Essentials for R for Statistics* plugin, which will give us access to R so that we can implement robust models using the *WRS2* package (Mair, Schoenbrodt, & Wilcox, 2015).

4.13.1 The PROCESS tool

The *PROCESS* tool (Hayes, 2018) wraps up a range of functions written by Andrew Hayes and Kristopher Preacher (e.g., Hayes & Matthes, 2009; Preacher & Hayes, 2004, 2008a) to do moderation and mediation analyses, which we look at in Chapter 11. While using these tools, spare a thought of gratitude to Hayes and Preacher for using their spare time to do cool stuff like this that makes it possible for you to analyse your data without having a nervous breakdown. Even if you think you are having a nervous breakdown, trust me, it's not as big as the one you'd be having if *PROCESS* didn't exist. The *PROCESS* tool is what's known as a custom dialog box and it can be installed in three steps (Mac OS users ignore step 2):
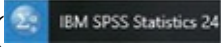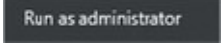
1. *Download the install file*. Download the file **process.spd** from Andrew Hayes's website: http://www.processmacro.org/download.html. Save this file onto your computer.
Figure 4.20 Installing the *PROCESS* menu
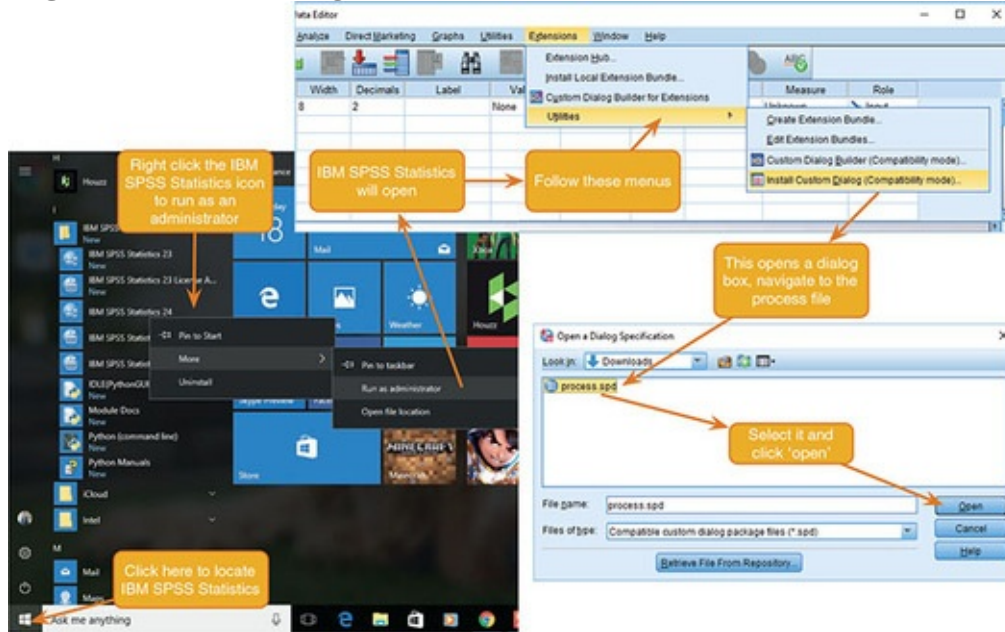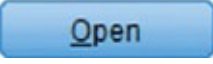
2. *Start up IBM SPSS Statistics as an administrator*. To install the tool in Windows, you need to start IBM SPSS Statistics as an administrator. To do this, make sure that SPSS isn't already running, and click the *Start* menu (

). Locate the icon for SPSS ( ), which, if it's not in your most used list, will be listed under 'I' for IBM SPSS Statistics. The text next to the icon will refer to the version of SPSS Statistics that you have installed (if you have a subscription it will say 'Subscription' rather than a version number). Click on this icon with the *right mouse button* to activate the menu in Figure 4.20. Within this menu select (you're back to using the left mouse button now) . This action opens SPSS Statistics but allows it to make changes to your computer. A dialog box will appear that asks you whether you want to let SPSS make changes to your computer and you should reply 'yes'.

3. Once SPSS has loaded select *Extensions* *Utilities*   , which activates a dialog box for opening files (Figure 4.20).[7] Locate the file

process.spd, select it, and click [Open]. This installs the *PROCESS* menu and dialog boxes into SPSS. If you get an error message, the most likely explanation is that you haven't opened SPSS as an administrator (see step 2).

[7](#) If you're using a version of SPSS earlier than 24, you need to select Utilities Custom Dialogs [Install Custom Dialog (Compatibility mode)...].
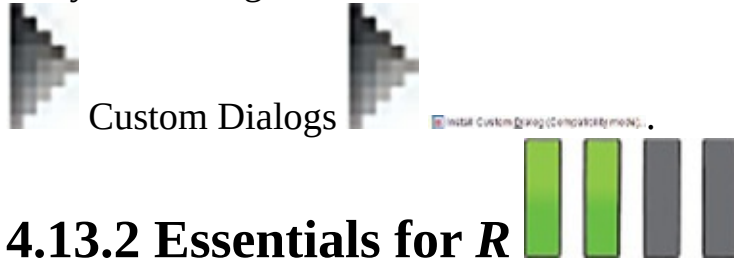
4.13.2 Essentials for *R*

At various points in the book we're going to use robust tests that use R. To get SPSS Statistics to interface with R, we need to install: (1) the version of R that is compatible with our version of SPSS Statistics; and (2) the *Essentials for R for Statistics* plugin from IBM. At the time of writing, the R plugin isn't available for SPSS Statistics version 25, but by the time the book is published it may well be. These instructions are for SPSS Statistics version 24 but you can hopefully extrapolate to other versions. First, let's get the plugin and installation documentation from IBM:

1. Create an account on IBM.com (www-01.ibm.com).
2. Go to https://www-01.ibm.com/marketing/iwm/iwm/web/preLogin.do?source=swg-tspssp
3. There will be a long list of stuff you can download. Select *IBM SPSS Statistics Version 24 – Essentials for R* (or whatever version of SPSS Statistics you're using) and click continue.
4. Complete the privacy information, and read and agree (or not) to IBM's terms and conditions.
5. Download the version of *IBM SPSS Statistics Version 24 – Essentials for R* for your operating system (Windows, Mac OS, Linux, etc.) and the corresponding installation instructions (labelled *Installation Documentation 24.0 Multilingual for xxx*, where *xxx* is the operating system you use). By default the website uses an app called the *Download Director* to manage the download. This app never works for me (on a Mac) and if you have the same problem, switch the tab at the top of the list of downloads to 'Download using http' ([Download using Download Director | Download using http]) and download the files directly through your browser.
6. Open the installation documentation (it should be a PDF file) and check which version of R you need to install.[8]

Having got the *Essentials for R* plugin, *don't install it yet.* You need to check which version of R you need, and download it. SPSS Statistics typically uses an old version of R (because IBM needs to check that the *Essentials for R* plugin is stable before releasing it and by the time they have done that R has updated). Finding old versions of R is tediously overcomplicated; I've tried to illustrate the process in [Figure 4.21](#).

7. Go to https://www.r-project.org/
8. Click the link labelled *CRAN* (under the *Download* heading) to go to a page to select a CRAN mirror. A CRAN mirror is a location from which to download R. It doesn't matter which you choose; because I'm based in the UK, I picked one of the UK links in [Figure 4.21](#).
9. On the next page, click the link for the operating system you use (Windows, Mac, or Linux).
10. You will already know what version of R you're looking for because I told you to check before getting to this point (e.g., SPSS Statistics version 24 uses R version 3.2).[9] What happens next differs for Windows and Mac OS:
    - Windows: If you selected the link to the Windows version you'll be directed to a page for R for Windows. Click the link labelled *Install R for the first time* to go a page to download R for Windows. *Do not* click the link at the top of the page, but scroll down to the section labelled *Other builds,* and click the link to *Previous releases.* The resulting page lists previous versions of R. Select the version you want (for SPSS Statistics 24, select R 3.2.5, for other versions of SPSS consult the documentation).
    - Mac OS: If you selected the link to the OS X version you'll be directed to a page for R for Mac OS X. On this page click the link to the *old* directory. This takes you to a directory listing. You need to scroll down a bit until you find the *.pkg* files. Click the link to the *.pkg* file of the version of R that you want (for SPSS Statistics 24, click R 3.2.4, for other versions consult the documentation).

[8] At the time of writing, the installation documentation for SPSS Statistics 24 links to a PDF file for version 23, which says that you need R 3.1. This is true for version 23 of SPSS Statistics, but version 24 requires R 3.2 onwards.
[9] There will be several versions of R 3.2 which are denoted as 3.2.x, where x is a minor update. It shouldn't matter whether you install version 3.2.1 or 3.2.5, but you may as well go for the last of the releases. In the case of R 3.2, the last update before release 3.3 was 3.2.5.

Figure 4.21 Finding an old version of R is overly complicated …

You should now have the install files for R and for the *Essentials for R* plugin in your download folder. Find them. First, install R by double-clicking the install file and going through the usual install process for your operating system. Having installed R, install the *Essentials for R* plugin by double-clicking the install file to initiate a standard install. If all that fails, there is a guide (at the time of writing) to installing the R plugin via GitHub at https://developer.ibm.com/predictiveanalytics/2016/03/21/r-spss-installing-r-essentials-from-github/ or see Oditi's Lantern.

4.13.3 The WRS2 package

Once the *Essentials for R* plugin is installed (see above) we can access the WRS2 package for R (Mair et al., 2015) by opening a syntax window and typing and executing the following syntax:

```
BEGIN PROGRAM R.
install.packages("WRS2")
END PROGRAM.
```

The first and last lines (remember the full stops) tell SPSS to talk to R and then to stop. All the stuff in between is language that tells R what to do. In this case it tells R to install the package *WRS2*. When you run this program a window will appear asking you to select a CRAN mirror. Select any in the list (it determines from where R downloads the package, so it's not an important decision).

I supply various syntax files for robust analyses in R, and at the top of each one I include this program (for those who skipped this section). However, you only need to execute this program once, not every time you run an analysis. The only times you'd need to re-execute this program would be: (1) if you change computers; (2) if you upgrade SPSS Statistics or need to reinstall the *Essentials for R* plugin, or R itself, for some reason; (3) something goes wrong and you think it might help to reinstall *WRS2*.

Oditi's Lantern SPSS extensions

'I, Oditi, am bearded like a great pirate sailing my ship of idiocy across the vacant seas of your mind. To join my cult you must become pirate-like in my image and speak the pirate language. You must punctuate your speech with the exclamation 'Rrrrrrrrrr'. It will help you uncover the unknown numerical truths embedded in the treasure maps of data. The Rrrrrr plugin for SPSS Statistics will help, and my lantern is primed with a visual cannon-ball of an installation guide that will blow your mind.'

4.13.4 Accessing the extensions

Once the *PROCESS* tool has been added to SPSS Statistics it appears in the *Analyze* *Regression* menu. If you can't see it then the install hasn't worked and you'll need to work through this section again. At the time of writing *WRS2* can be accessed only using syntax.

4.14 Brian's attempt to woo Jane

Brian had been stung by Jane's comment. He was many things, but he didn't think he had his head up his own backside. He retreated from Jane to get on with his single life. He listened to music, met his friends, and played *Uncharted 4*. Truthfully, he mainly played *Uncharted 4*. The more he played, the more he thought of Jane, and the more he thought of Jane, the more convinced he became that she'd be the sort of person who was into video games. When he next saw her he tried to start a conversation about games, but it went nowhere. She said computers were good only for analysing data. The seed was sown, and Brian went about researching statistics packages. There were a lot of them. Too many. After hours on Google, he decided that one called SPSS looked the easiest to learn. He would learn it, and it would give him something to talk about with Jane. Over the following week he read books, blogs, watched tutorials on YouTube, bugged his lecturers, and practised his new skills. He was ready to chew the statistical software fat with Jane.

Figure 4.22 What Brian learnt from this chapter
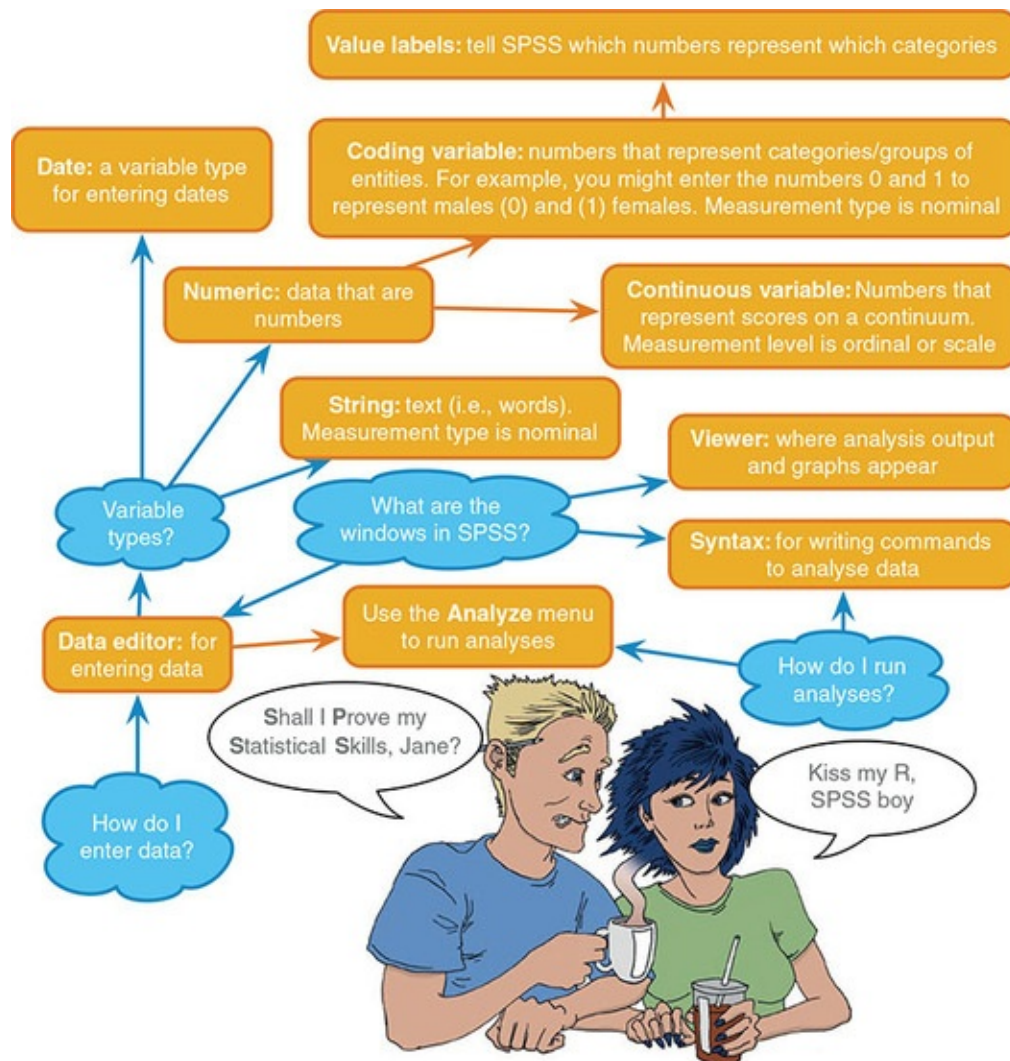
He searched around campus for her: the library, numerous cafés, the quadrangle – she was nowhere. Finally, he found her in the obvious place: one of the computer rooms at the back of campus called the Euphoria cluster. Jane was studying numbers on the screen, but it didn't look like SPSS. 'What the hell …,' Brian thought to himself as he sat next to her and asked …

4.15 What next?

At the start of this chapter we discovered that I feared my new environment of primary school. My fear wasn't as irrational as you might think, because, during the time I was growing up in England, some idiot politician had decided that all school children had to drink a small bottle of milk at the start of the day. The government supplied the milk, I think, for free, but most free things come at some kind of price. The price of free milk turned out to be lifelong trauma. The milk was usually delivered early in the morning and then left in the hottest place

someone could find until we innocent children hopped and skipped into the playground oblivious to the gastric hell that awaited. We were greeted with one of these bottles of warm milk and a very small straw. We were then forced to drink it through grimacing faces. The straw was a blessing because it filtered out the lumps formed in the gently curdling milk. Politicians take note: if you want children to enjoy school, don't force-feed them warm, lumpy milk.

But despite gagging on warm milk every morning, primary school was a very happy time for me. With the help of Jonathan Land, my confidence grew. With this new confidence I began to feel comfortable not just at school but in the world more generally. It was time to explore.

4.16 Key terms that I've discovered

Currency variable
Data editor
Data view
Date variable
Long format data
Numeric variable
Smartreader
String variable
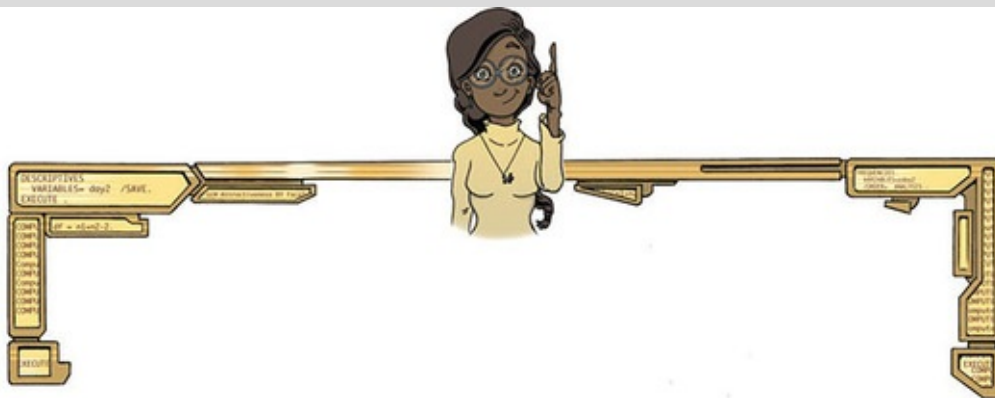Syntax editor
Variable view
Viewer
Wide format data

Smart Alex's tasks

- **Task 1**: Smart Alex's first task for this chapter is to save the data that you've entered in this chapter. Save it somewhere on the hard drive of your computer (or a USB stick if you're not working on your own computer). Give it a sensible title and save it somewhere easy to find (perhaps create a folder called

'My Data Files' where you can save all of your files when

working through this book).
- **Task 2**: What are the following icons shortcuts to?

- **Task 3**: The data below show the score (out of 20) for 20 different students, some of whom are male and some female, and some of whom were taught using positive reinforcement (being nice) and others who were taught using punishment (electric shock). Enter these data into SPSS and save the file as **Method Of Teaching.sav**. (Hint: the data should not be entered

in the same way that they are laid out below.)

Male		Female	
Electric Shock	Being Nice	Electric Shock	Being Nice
15	10	6	12
14	9	7	10
20	8	5	7
13	8	4	8
13	7	8	13

- **Task 4**: Thinking back to [Labcoat Leni's Real Research 4.1](#), Oxoby also measured the minimum acceptable offer; these MAOs (in dollars) are below (again, they are approximations based on the graphs in the paper). Enter these data into the SPSS Data Editor and save this file as **Oxoby (2008) MAO.sav.**

> Bon Scott group: 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5
>
> Brian Johnson group: 0, 1, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 1

- **Task 5**: According to some highly unscientific research done by a UK department store chain and reported in *Marie Claire*

magazine (http://ow.ly/9Dxvy), shopping is good for you. They found that the average woman spends 150 minutes and walks 2.6 miles when she shops, burning off around 385 calories. In contrast, men spend only about 50 minutes shopping, covering 1.5 miles. This was based on strapping a pedometer on a mere 10 participants. Although I don't have the actual data, some simulated data based on these means are below. Enter these data into SPSS and save them as **Shopping Exercise.sav.**

Male		Female	
Distance	Time	Distance	Time
0.16	15	1.40	22
0.40	30	1.81	140
1.36	37	1.96	160
1.99	65	3.02	183
3.61	103	4.82	245

- **Task 6**: This task was inspired by two news stories that I enjoyed. The first was about a Sudanese man who was forced to marry a goat after being caught having sex with it (http://ow.ly/9DyyP). I'm not sure whether he treated the goat to a nice dinner in a posh restaurant beforehand but, either way, you have to feel sorry for the goat. I'd barely had time to recover from that story when another appeared about an Indian man forced to marry a dog to atone for stoning two dogs and stringing them up in a tree 15 years earlier (http://ow.ly/9DyFn). Why anyone would think it's a good idea to enter a dog into matrimony with a man with a history of violent behaviour towards dogs is beyond me. Still, I wondered whether a goat or dog made a better spouse. I found some other people who had been forced to marry goats and dogs and measured their life satisfaction and how much they like animals. Enter these data

  into SPSS and save as **Goat or Dog.sav.**

Goat		Dog	
Animal Liking	Life Satisfaction	Animal Liking	Life Satisfaction
69	47	16	52
25	6	65	66
31	47	39	65
29	33	35	61
12	13	19	60
49	56	53	68
25	42	27	37
35	51	44	72
51	42		
40	46		
23	27		
37	48		

- **Task 7**: One of my favourite activities, especially when trying to do brain-melting things like writing statistics books, is drinking tea. I am English, after all. Fortunately, tea improves your cognitive function – well, it does in old Chinese people, at any rate (Feng, Gwee, Kua, & Ng, 2010). I may not be Chinese and I'm not *that* old, but, I nevertheless, enjoy the idea that tea might help me think. Here are some data based on Feng et al.'s study that measured the number of cups of tea drunk and cognitive functioning in 15 people. Enter these data into SPSS and save the file as **Tea Makes You Brainy 15.sav.**

Cups of Tea	Cognitive Functioning
2	60
4	47
3	31
4	62
2	44
3	41
5	49
5	56
2	45
5	56
1	57
3	40
3	54
4	34
1	46

- **Task 8**: Statistics and maths anxiety are common and affect people's performance on maths and stats assignments; women, in particular, can lack confidence in mathematics (Field, 2010). Zhang, Schmader, & Hall, (2013) did an intriguing study, in which students completed a maths test in which some put their own name on the test booklet, whereas others were given a booklet that already had either a male or female name on it. Participants in the latter two conditions were told that they would use this other person's name for the purpose of the test. Women who completed the test using a different name performed better than those who completed the test using their own name. (There were no such effects for men.) The data below are a random subsample of Zhang et al.'s data. Enter them into SPSS and save the file as **Zhang (2013)**

subsample.sav

	Male			Female		
Female Fake Name	Male Fake Name	Own Name	Female Fake Name	Male Fake Name	Own Name
33	69	75	53	31	70
22	60	33	47	63	57
46	82	83	87	34	33
53	78	42	41	40	83
14	38	10	62	22	86
27	63	44	67	17	65
64	46	27	57	60	64
62	27			47	37
75	61			57	80
50	29				

- **Task 9**: What is a coding variable?
- **Task 10**: What is the difference between wide and long format data?

Answers & additional resources are available on the book's website at **https://edge.sagepub.com/field5e**