4 The RESULTS section

Diagnostic questions on statistics for Chapter 4

Before reading this chapter, check that you can answer correctly the following diagnostic questions:

1 What are *inferential* statistics?

2 What is a *statistically significant* difference?

3 What does it mean to say that the *5 percent significance level* was used in an analysis?

4 What is a *type I* error?

You can find the answers at the rear of the book (p. 274). If you had difficulties answering any of these questions, turn to Chapter 11.

Of all of the sections of the report, the RESULTS is probably the one you most frequently mishandle. Yet it really is one of the more straightforward sections of the report. All you need to do in this section is to report the *findings* of your study in the most appropriate manner, resisting in the process any temptation to *interpret* them as you go along. That is, a bit like a journalist of the old school, you must distinguish rigidly between "fact" and "comment". In this section you must not go beyond stating what you *found* ("fact") to discussing what these findings appear to you to *mean* ("comment"). That debate takes place in the DISCUSSION.

With **quantitative** data – data in the form of numbers – there will generally be *two* aspects of your findings to report. First, you must provide a description of the key features of the data you obtained. You use *descriptive* statistics to do this. Second, you must give an account of the type and outcomes of the *inferential* statistical analyses you performed on these data. I will explain the difference between these two forms of statistics next.

Summary

1 Your principal goal in the RESULTS is to report the findings of your study clearly and accurately.

2 Separate fact from comment. In the RESULTS, restrict yourself to stating what you found. Determining how best to interpret these findings takes place in the DISCUSSION.

3 Report both the descriptive statistics and the type and outcomes of the inferential statistics that you used.

4.1 Describing the data: descriptive statistics

If you reported *all* the data from your study – all the numbers you collected – it would be quite difficult for your reader to interpret what the scores of your participants had to say. Very few people are capable of grasping the essential message of a set of data from looking at the raw scores (the actual numbers you gathered from the participants themselves). We therefore need a way of conveniently and simply summarizing the main features of a set of data. **Descriptive statistics** are a way of doing this. These are statistics that describe the key characteristics of a body of data. They enable us to efficiently assess things like whether the scores of our participants tend to be similar to each or whether they vary quite a bit (measures of **variation**, such as the standard deviation or range). Likewise, we can see what score best typifies the data as a whole, or the performance of participants in each condition (measures of **central tendency**, such as the mean, median or mode). Measures like these help us to make sense of our data and of the outcomes of our inferential statistical analyses.

Table 4.1
Scores from an Experiment with One Control and One Experimental Condition

Control	Experimental
17.3	6.8
39.2	14.1
20.7	61.2
79.3	61.7
81.5	14.0
24.7	75.9
55.0	32.3
73.6	22.0
33.0	53.1
5.4	83.2
18.6	7.8
42.7	94.8
56.5	49.7
24.9	37.2
57.9	23.9

For instance, imagine that we ran an experiment and obtained the numbers in Table 4.1. In this form it is difficult to make much sense of the data. It is not easy to tell much about the comparative performance of the participants in the two groups. The scores appear to vary quite a bit from individual to individual, but is the *overall* performance of the participants in the two groups all that different? This is, of course, the question we wish to answer.

In Table 4.2, the same data are expressed in terms of two *descriptive* statistics. There is a measure of *central tendency* (in this case the

Table 4.2
The Same Data as in Table 4.1, Expressed in Terms of Two Descriptive Statistics

	Control	Experimental
M	42.0	42.5
SD	24.2	28.5

Note. M = mean; SD = standard deviation; n = 15 in each condition.

mean, which tells us the average performance in each group) and a measure of *dispersion* (in this case the **standard deviation**, which tells us about the extent to which the scores within each group vary).

Now our task is much easier. From Table 4.2 we can see, perhaps to our surprise, that the typical performance in the two groups is quite similar. In contrast, the scores of the participants *within* the two conditions vary quite a bit (the standard deviations in the two groups are high, given the mean scores). Looking at this table it would be surprising indeed if we were to find a statistically significant difference between our two groups. (If we did, we would probably want to check the analysis.)

The numbers in Table 4.1 are what we call in the trade *raw* scores. As the name implies, **raw scores** are the unprocessed scores that you obtain from your participants. Because of the difficulties inherent in grasping the essential features of raw data, we tend *not* to put such data in the RESULTS. Instead, we present an account of the principal features of the data in the form of descriptive statistics. Such material we usually display in a suitably formatted, appropriately labelled, and informatively titled *table* – like those used in the example RESULTS in this chapter.

Typically, such **tables** should contain an appropriate measure of central tendency, a measure of dispersion, an indication of the numbers of participants per condition and, once you are familiar with them, relevant confidence intervals. Occasionally you may need more than this, but generally you will not. So, just because your statistical software package pumps out every descriptive statistic under the sun, do not dump it all down into the RESULTS. Also, when reporting your descriptive statistics in tables or text, *use two decimal places* at most. Psychological measures are imprecise and do not need the false precision of lots of decimal places, whatever gets splurged out by your statistical package. Where your data are in the form of frequencies and there is only one observation per participant (as for example when you have used chi-square to analyse your data) then your table will consist of the counts in each category (as in Table 1 in Section 4.3).

Generally speaking, therefore, there will usually be at least *one* table in your RESULTS; you can find more about how to lay out your tables and how to strike the balance between tables and text in Section 8.4. However, sometimes it is still not possible to fully grasp the main features of a set of data from even a table of descriptive statistics. If you wish to enhance your reader's understanding of the data, therefore, you might go one step further and *graph* it – especially as many of those who go blank at the sight of numbers have little problem in grasping the message of a well-designed graph. Be warned, however; there are some pitfalls. Advice on the use of graphs can be found in Section 8.5.

Do not be afraid, therefore, to deploy the techniques at your disposal to aid your reader's (not to mention your own) understanding of the basic features of your data. However, do not go overboard here. This aspect of your findings is typically less interesting than the outcomes of your *inferential* statistical analyses. Think of the descriptive statistics as preparing the ground for the reporting of the inferential analyses.

Summary of Section 4.1

1 In the opening paragraph of the RESULTS reiterate briefly what data you gathered from your participants (e.g., response times, number of items recalled, number of people reporting nightmares). That is, remind your reader what the DV is (or principal DVs are if you have more than one). This both sets the scene and ensures that the reader does not have to look back at your METHOD to understand your results.

2 In general, you should not include the raw data in this section. Instead, provide a potted account of your data in the form of descriptive statistics. These should generally be presented in tables.

3 Report descriptive statistics to no more than two decimal places.

4 Feel free to graph the data as well if you feel that this will help the reader to understand the findings better. However, read Section 8.5 first.

4.2 Analysing the data: inferential statistics

After drawing attention to the descriptive statistics, tell your reader what analyses you ran on these data and the results of these analyses. These analyses involve the *inferential* statistics, such as *chi-square*, *t tests*, and *analysis of variance* (ANOVA), that you have grown to know and love. This is the part of the RESULTS that gives some of you sleepless nights. Yet, it need not be as daunting as you think. The key point is to tell your reader *clearly* and *accurately* (a) how you analysed the data; (b) what the outcome of the analysis was; and (c) what this result tells us. There are clear rules to guide you in this. These are:

1 State clearly *in what way* you analysed the data, i.e., which inferential statistical test you used. Describe this test precisely. Do not, for example, say that a "*t* test" was used. Instead, state which *type* of *t* test – for example, a "related *t* test". If you are using ANOVA, make sure you state accurately which type you used (see Section 4.6.10)
2 State the significance level that you used and (where appropriate) whether your test was one- or two-tailed.
3 State the precise value that you *obtained* for the statistic (i.e., the value printed in your output or that you calculated by hand). For example, the value of *F* or *t*. However, do this to no more than two decimal places. (Do not slavishly copy out all of the digits to the right of the decimal point on your statistical output, as this looks *very* amateurish.)
4 Provide the additional information your reader needs to look up the relevant *critical* value of your statistic should s/he want. These are usually the *degrees of freedom* or the *numbers of participants* or *observations*. Note that you are expected to provide this information, even though most of the time your statistical software package will give you an exact probability associated with your obtained statistic. (See Appendix 3 for more about critical values.) To ensure you get this right, follow the relevant examples in Section 4.6 of this chapter and on the book's Web site.
5 Wherever you can, report the *exact* probability associated with your obtained statistic regardless of whether the outcome is statistically significant or not. Do this to no more than three decimal places (e.g., p = .002). Where your output prints p = .000 or it is not possible to report the exact probability for other reasons, see Appendix 3 for what to do.
6 State whether the obtained value was statistically significant or not.
7 State explicitly what this result tells us about the data. That is, relate the outcome of your inferential statistic back to the relevant descriptive statistics.
8 Make sure, however, that you distinguish between what you have found and what you believe it means (your inferences and conclusions about your findings). In the RESULTS restrict yourself to describing the findings. Save discussion of how best to interpret these findings until the DISCUSSION.

This sounds like a lot, but in fact the above information can be conveyed surprisingly succinctly. For example: "Participants using the semantically related items were significantly quicker to reach the criterion than those using the semantically unrelated items, $t(42) = 2.23$, $p = .025$ (two-tailed test)", or "The effect of reinforced practice upon the time taken to reach criterion was not statistically significant, $t(40) = 1.30$, $p = .20$ (two-tailed test)." How to do this for some of the statistics that you are likely to use when you first start writing reports can be found in Section 4.6 of this chapter and on the book's Web site.

4.3 An example RESULTS section

Here is an example of how a basic RESULTS governed by these conventions might look. In this case, the data are frequencies and there is only one observation per participant, so the table of data contains counts, rather than measures of central tendency and dispersion.

Results

The number of participants reporting nightmares in each condition of the experiment is shown in Table 1. The data were analysed using chi-square and an alpha level of .05.

There was a statistically significant association between the consumption of cheese and the incidence of nightmares, $\chi^2(1, N = 100) = 4.89$, $p = .027$. Participants eating cheese three hours before going to bed reported a higher proportion of nightmares than did participants not eating cheese in that period.

In the post-experimental interviews, after prompting, eight participants were able to describe the experimental hypothesis in full or in part. Of these, five were in the cheese condition (of whom one reported a nightmare). Of the three in the no cheese condition, one reported a nightmare.

Table 1
The Number of Participants Reporting Nightmares and not Reporting Nightmares in the Two Conditions

	Nightmare	
Condition	Yes	No
Cheese	33	17
No cheese	22	28

On the design front, note the use here of post-experimental interviews (Section 13.9.4) to investigate the extent to which participants reported awareness of the experimental hypotheses. These are useful things to include, as you will see when we come to the DISCUSSION of this experiment.

If you have more than one set of data and attendant analyses to report, it may be necessary to include separate tables for the different data. If so, consider the data *and* their attendant analyses as two sides of the same coin. So report them as a unit, by describing the data first and then detailing the outcomes of the analyses *before* moving on to the next data/analysis set. Make sure that you work through these in order of importance, starting with the data and analyses that are central to testing the main predictions of the study and working through to material that is illuminating but essentially ancillary.

3

This, then, is the basic material that you should report in the RESULTS. You should aim at all times for *clarity* and *accuracy* – you must give your reader a clear idea of the type of data you gathered, describe its main features using descriptive statistics, and report appropriately the nature and outcomes of your inferential statistical analyses. Moreover, it should be clear to your reader at any given point *which* data you are talking about and *which* inferential analysis relates to that set of data, especially if you have used more than one DV or wish to test more than one set of predictions. You should also make sure that you provide enough information in this section (e.g., in labelling the conditions) for the reader to be able to make sense of your data *without* having to turn to other parts of the report for clarification.

? SAQ 19 Why should you provide your reader with a clear idea of the type of data you gathered, its main features and the nature and outcomes of the inferential statistics that you used?

? SAQ 20 If you were to come across the following in the RESULTS section of a research report, what criticisms would you have?

> The data were analysed using Analysis of Variance. The results were statistically significant.

4.4 Nine tips to help you avoid common mistakes in your RESULTS section

Here are nine tips to help you avoid some of the commoner mistakes made in this section:

1 The RESULTS should never be just a collection of tables, statistics and figures. This section must *always* have useful and informative text. To learn more about how to strike the balance between tables and text, see Section 8.4.

2 Include enough information in this section for your reader to be able to make sense of the results without having to look elsewhere in the report. For example, take care to label conditions or groups meaningfully in tables, figures and text. Avoid using **ambiguous terms** such as "Condition 1" or "Group A". You should not use meaningless or difficult-to-decipher abbreviations *anywhere* in the report (see Section 8.2).

3 It is important to look at your data and the descriptive statistics before you analyse the data inferentially (Section 13.9.5). However, do not **eyeball the data** by writing that there are "differences" between conditions, or that some means are "higher" or "smaller" than are others, *before* you have reported the outcome of the relevant inferential analysis. This is because these numerical differences may not turn out to be statistically significant and the convention in psychology is to talk about differences between conditions *only* if these have been shown to be statistically significant. For example, it is possible to describe the mean score for the experimental condition in Table 4.2 as higher than that for the control condition. However, this difference is extremely unlikely to be statistically significant. It would be misleading to talk about the data in this way and it is unnecessary to do so. Talk about differences only *after* you have reported the inferential analysis.

4 Once you have reported the outcome of your inferential statistic, you are free to comment on the presence or absence of differences among your conditions. This is no longer eyeballing the data, as now you know whether the sample differences are statistically significant or not. If you only have two conditions you can go even further and describe which condition was higher, better, faster (or whatever the DV measured) than the other. However, be careful here. One of the commonest and *most damaging* failings is to talk

about differences between conditions even though these have failed to achieve statistical significance. The purpose of testing for statistical significance is to help us to decide whether to treat the means in different conditions as equivalent or as different. If your analyses tell you that you should not reject the null hypothesis, then act on that basis and assume that there are no reliable differences between the relevant conditions. If you write about differences when the analysis has not been statistically significant, you are simply ignor-
4 ing the analysis! Your marker may assume that this is because you do not understand what the analysis means.

5 You do not need to give the underlying details of the inferential statistical procedure (e.g., the rationale and workings of the *t* test) or reasoning (e.g., the principles of statistical significance testing) in this section. (Unless your tutor tells you explicitly to include such material.) One of the apparent paradoxes of the practical report is that whereas you must assume that your reader lacks
5 knowledge of the topic that you are investigating, you *may* assume that s/he has a basic grasp of statistics and of the principles of significance testing.

6 Include sufficient information in this section to enable your reader to reach his or her own conclusions about the implications of your data. Your reader should be able to disagree with your interpretation of the data purely on the basis of the information you provide here.

7 Do not *duplicate* analyses. For example, do not report *both* a parametric test *and* its nonparametric equivalent (Section 11.4). Decide
6 which is more appropriate for your data and report the outcomes of the one you choose.

8 Include *all* of the data in this section that you wish to comment upon in the DISCUSSION, however impressionistic, **qualitative** (i.e., not involving numbers) and unamenable to statistical analysis. The inclusion of qualitative data, such as a selection of comments from your participants, can be very useful. However, in *experiments* such data should always be used as a *supplement* to the quantitative, numerical data.

9 Do not necessarily restrict yourself to reporting the obvious analyses – do not be afraid to squeeze all the relevant information from your data that you wish. Bear in mind, however, that your aim is to *communicate* your findings. So avoid overloading this section with irrelevant and unnecessary analyses (see Section 4.2). Nevertheless, there may be some way in which additional analyses of your data

7 can help you to clarify or choose between alternative explanations or to resolve other issues of interpretation. So, think about ways in which further analyses might be useful.

4.5 Rejecting or not rejecting the null hypothesis

Rejecting the null hypothesis entails that you accept the alternative hypothesis; this does not mean, however, that you can conclude that the psychological hypothesis underlying your study has been supported. After rejecting the null hypothesis you have, in fact, to search for the most reasonable explanation of your findings. This may well turn out to be the arguments you summarized in the INTRODUCTION, but need not be. You can only decide this after a suitable discussion in which you consider and examine closely *all* the plausible explanations of your findings. In an experiment, high on the list of issues to examine will be whether you exercised sufficient *control* to enable you to make unequivocal inferences. This is why we require the next section – the DISCUSSION.

Of course, discussion is also needed for studies in which you have *failed* to reject the null hypothesis. Under these circumstances you also have to work out what this means psychologically, as opposed to statistically. As well as examining the adequacy of the control that you exercised in an experiment another thing to consider when you have *failed* to reject the null hypothesis is whether you had enough participants (see Chapter 12).

Where your study is not experimental, you will still need to do these things – to assess rival explanations for your findings and search for the most plausible interpretation. Moreover, given that the level of control you will have been able to exercise is lower than in an experiment, it will be harder to rule out alternative explanations (see Section 5.8).

Summary of Section 4.5

1 Once you have determined whether you have to reject or not reject the null hypothesis, you begin the search for the most reasonable explanation of the findings.

2 This process takes place in the DISCUSSION.

4.6 Reporting specific statistics

Knowing what to report can be bewildering. Even when you have chosen the correct analysis and run it properly, it can be hard to detect the bits that you need to extract from the output to report in the RESULTS. This is especially so when you are a novice, but it can be bewildering even when you are more experienced. So, do not feel alone with this problem. Do *not* under any circumstances, however, cope with it simply by copying over the output from a statistical analysis package, perhaps annotating it by hand, and leaving it at that. I am amazed how often I get grubby bits of second-rate printout stapled into the report, sometimes with no supportive text, often with bits of scribble by way of annotation of the output, and am supposed to be prepared to treat that as if it were a RESULTS! In such circumstances I simply assume that the student is unable to detect the bits of the output s/he needs, cannot be bothered to do the hard work of locating the correct material and laying it out properly and neatly, and assumes that I am too stupid to realize this. As you can imagine, this does not go down well and the mark is not impressive.

The main thing, especially early on in your career, is to demonstrate to your marker that you realize what needs to be reported. The basic rule when reporting inferential statistics is that you need to provide the information that will enable someone else to check whether the outcome of a test is or is not statistically significant.

8 The rest of this chapter contains examples of how to report those specific statistics that you are most likely to use in the first year or so of report writing. More information about each of these statistics, together with some of the issues to watch out for when using them, can be found in Section B of the book's Web site at **http://mcgraw-hill.co.uk/openup/harris/**. The Web site also provides examples of how to report other statistics that you may meet later on in your career as a student of psychology. Before using this section or the Web site you should also be familiar with the issues involved in choosing tests. (These are discussed in Section 11.4 and Section A of the book's Web site.) Please note that, with the exception of the example results for the mnemonic experiment, the tables and figures referred to in these examples do not actually exist!

When reporting, punctuation and the use of italic are important, so pay attention to this when preparing your text. The sequencing is usually also important, as are such seemingly trivial details as whether there is a space between p and = and even whether you put $p = .025$ or $p = 0.025$. This section is a joy for the anally retentive among us; for the rest, it can be a trial.

4.6.1 Chi-square, χ^2

The information you need to provide to enable someone to check the significance of your obtained value of χ^2 is the *degrees of freedom*. You should also report the total number of observations. This example has 1 degree of freedom and is based on 100 observations.

> An alpha level of .05 was used. Analysis of the data in Table 1 using chi-square revealed that breaking the speed limit was significantly associated with gender, $\chi^2(1, N = 100) = 10.83$, $p = .001$. Males tended to break the speed limit whereas females tended not to break the speed limit.

4.6.2 Spearman rank correlation coefficient (rho), r_s

The information you need to provide to enable someone to check r_s is the *number of participants*. This example has 40 participants:

> An alpha level of .05 was used. Analysis of the data displayed in the scatter plot in Figure 1 using Spearman's rho (corrected for ties) indicated that ratings of mood were significantly positively correlated with the mean ratings of the attractiveness of the photographs, $r_s(40) = .48$, $p = .002$ (two-tailed test). The ratings were thus moderately correlated, with more positive mood tending to be associated with higher ratings of attractiveness.

4.6.3 Pearson's product moment correlation coefficient, r

The information you need to provide to enable someone to check r is generally the *degrees of freedom*, given by the N of observations minus 2. In this example, there are 40 participants, so there are 38 degrees of freedom:

> An alpha level of .05 was used. Analysis of the data displayed in the scatter plot in Figure 1 using Pearson's r indicated that age was

significantly negatively correlated with the mean ratings of the attractiveness of the photographs, $r(38) = -.37$, $p = .02$ (two-tailed test). The variables were thus moderately correlated, with increases in age tending to be associated with decreases in the ratings of attractiveness.

4.6.4 Mann-Whitney U test, *U*

The information you need to provide to enable someone to check *U* is the *number of participants in group 1* and the *number of participants in group 2*. This example has 14 and 16 respectively in these groups:

> An alpha level of .05 was used. Analysis of the data in Table 1 using the Mann-Whitney U test indicated that ratings of overall satisfaction with the tutor's teaching were significantly higher among those receiving the positive comment than among those not receiving this comment, $U(14, 16) = 56$, $p = .02$ (two-tailed test).

4.6.5 Wilcoxon's Matched-Pairs Signed-Ranks Test, *T*

The information you need to provide to enable someone to check *T* is the *number of participants overall, not counting those with tied ranks* (i.e., not counting those with the same scores in each condition). This example has 18 participants without such tied ranks:

> An alpha level of .05 was used. Analysis of the data in Table 1 using the Wilcoxon test indicated that participants rated their own chances of experiencing the diseases overall as significantly lower than they did the chances of the average student, $T(18) = 27$, $p = .01$ (two-tailed test). (Six participants had tied ranks.)

4.6.6 Kruskal-Wallis one-way analysis of variance, *H*

The information you generally need to provide to enable someone to check *H* is the *degrees of freedom*, given by the *N* of groups minus 1. In this example there are 3 groups and therefore 2 degrees of freedom:

> An alpha level of .05 was used. Analysis of the data in Table 1 using the Kruskal-Wallis one-way analysis of variance indicated that

the ratings of overall satisfaction with the tutor's teaching were significantly different among the three groups, $H(2) = 7.38$, $p. = 025$.

4.6.7 Friedman's ANOVA, χ^2_r

The information you generally need to provide to enable someone to check χ^2_r is the *degrees of freedom* given by the N of groups minus 1. In this example, there are 3 groups and therefore 2 degrees of freedom:

> An alpha level of .05 was used. Analysis of the data in Table 1 using Friedman's ANOVA indicated that ratings of the chances of experiencing the diseases were significantly different for the three targets (self, best friend and average student), $\chi^2_r(2) = 9.21$, $p = .01$.

4.6.8 The independent *t* test, *t*

The information you need to provide to enable someone to check t is the *degrees of freedom*. For the independent t test the degrees of freedom are given by the N of observations minus 2. This example has 15 participants in each condition and therefore 28 degrees of freedom:

> An alpha level of .05 was used. Analysis of the data in Table 1 using the independent t test for equal variances indicated that performance in the end-of-course examination was significantly higher among those receiving the positive comment than among those not receiving this comment, $t(28) = 2.70$, $p = .01$ (two-tailed test).

4.6.9 The related *t* test, *t*

For the related t test, the degrees of freedom are given by the N of observations minus 1. This example has 15 participants and therefore 14 degrees of freedom:

> An alpha level of .05 was used. Analysis of the data in Table 1 using the related t test indicated that performance in the end-of-course examination was significantly higher for the course in which participants received the positive comment than for the course in which they did not receive this comment, $t(14) = 2.49$, $p = .03$ (two-tailed test).

4.6.10 Analysis of variance (ANOVA), *F*

The information you need to provide to enable someone to check *F* is the *two* (NB *two*) *sets of degrees of freedom*: the degrees of freedom for the numerator of the *F* ratio *and* the degrees of freedom for the denominator of the *F* ratio. The numerator is the source under test, such as the main effect of an IV or the interaction between two or more IVs. (See Chapter 12 for a discussion of main effects and interactions.) The denominator is the bit that divides into the numerator and is also called the *error term*.

To illustrate how to go about reporting ANOVAs, here is an example RESULTS for the mnemonic experiment. This is a more advanced RESULTS than the earlier example. It is the kind of RESULTS that you will be expected to write as you become more experienced. If you have difficulty understanding any of the terminology in this, then do not hesitate to turn to Chapter 12 and your textbook of statistics for clarification.

Results

9 An alpha level of .05 was used for all statistical tests. The pretest data were analysed using one-way analysis of variance (ANOVA) for unrelated samples with condition (mnemonic or no mnemonic) as the independent variable. This analysis was not statistically significant, $F(1, 38) = 0.33$, $p = .57$, indicating that performance was equivalent in the two conditions (mnemonic group, $M = 10.95$ words recalled, $SD = 1.05$; no-mnemonic group, $M = 11.15$ words recalled, $SD = 1.14$). The groups thus appear to have had equivalent recall abilities for lists of words before the experimental group was taught the mnemonic.

The manipulation check on the imageability of the words used in the experiment was also satisfactory. Analysis of participants' ratings of the imageability of these words used one-way ANOVA for related samples with imageability (easily imaged or hard to image) as the independent variable. This revealed significantly higher ratings for the easily imaged words ($M = 5.21$, $SD = .77$) than for the hard-to-image words ($M = 3.39$, $SD = .95$), $F(1, 38) = 151.72$, $p < .001$.

The mean numbers of words of each type correctly recalled by those in the two conditions (excluding misspellings) are given in Table 2.

The data in Table 2 were analysed using 2 x 2 ANOVA for mixed designs, with imageability (easily imaged or hard to image) as the related samples variable and instruction (mnemonic or no mnemonic) as the unrelated samples variable. There was a statistically significant main effect of instruction, $F(1, 38) = 7.20$, $p = .01$, with those in the

Table 2
The Mean Number of Words from the Easily Imaged and The Hard-to-Image Lists Correctly Recalled in Each Condition

		Imageability	
Instruction	*n*	Easily imaged	Hard to image
Mnemonic	20		
M		18.15	13.15
SD		3.79	4.17
95% confidence interval		16.38–19.92	11.20–15.10
No mnemonic	20		
M		13.80	11.00
SD		3.25	4.62
95% confidence interval		12.28–15.32	8.84–13.16

Note. M = mean; SD = standard deviation.

mnemonic group recalling more items overall than did those in the no-mnemonic group ($M = 15.65$, $SD = 3.97$; $M = 12.40$, $SD = 3.74$, respectively). There was also a statistically significant main effect of imageability, $F(1, 38) = 145.22$, $p < .001$, with more items from the easily imaged list being recalled than from the hard-to-image list ($M = 15.98$, $SD = 4.12$; $M = 12.08$, $SD = 4.48$, respectively). However,

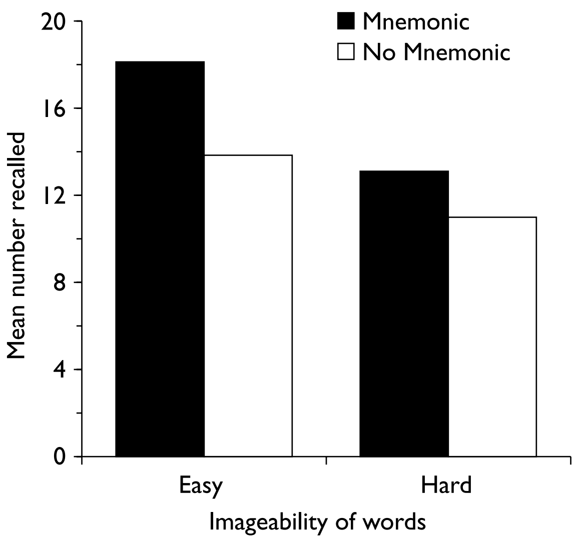

Figure 1. Mean number of easily imaged and hard-to-image words correctly recalled in each condition.

these main effects were qualified by the significant Instruction X Imageability interaction, $F(1, 38) = 11.55$, $p = .002$. Figure 1 shows this interaction.

Inspection of Figure 1 suggests that the significant interaction results from the greater number of easily imaged items recalled by the mnemonic group than by the no mnemonic group (difference = 4.35) relative to the smaller difference between the conditions in recall of the hard-to-image items (difference = 2.15). This is consistent with the experimental hypothesis. However, further analysis is required to confirm this statistically.

10

The majority of participants (18) in the no mnemonic condition claimed to have attempted to remember the items by rote repetition, with the remainder employing simple attempts at organizing the material into semantically related clusters. All those in the mnemonic condition reported using the method of loci, and most of these expressed their surprise at the impact that this appeared to have had on their ability to recall.

Note the use in this example of the abbreviation *M* to stand for *mean* and *SD* for *standard deviation* when reporting these descriptive statistics in text. On the design front, note the use of a pre-test to check that the groups were equivalent before being exposed to the IV and of a manipulation check to confirm that the easily imaged and hard-to-image words did differ in reported imageability. You will see the advantage of these features when we come to the DISCUSSION of this experiment (Section 5.7.2).

4.6.11 Four tips to help you avoid common mistakes when reporting ANOVA

Here are four tips to help you avoid some of the commoner mistakes students make when reporting ANOVA.

1 *Make sure that you indicate precisely which type of ANOVA you used.* The term ANOVA refers to a family of tests, not a single test. You will use different variants of ANOVA for different designs. For this reason it is important to specify *which* particular version of ANOVA you used. This is not hard: you can simply use the labelling convention that you employed in the DESIGN. For example, you can say that you analysed the data “using ANOVA

for two-way, mixed designs". Alternatively, you could say that you analysed the data "using ANOVA for a 2 × 2 mixed design". (See Section 13.3 for more on how to label designs.)

2 *Make sure that you indicate the number of levels on each IV*. Whichever way you label the design, make sure that you specify which IVs had which number of levels and what these levels were; with mixed designs, do not forget to specify which of the IVs used related samples and which used unrelated samples.

3 *Make sure that you report both sets of degrees of freedom for each F ratio*. This is an extremely common mistake and one that is both very damaging and yet easy to avoid. Remember that *every* value of *F* that you report has *two* sets of degrees of freedom. One set of degrees of freedom belongs to the *numerator* of the *F* ratio, the other set belongs to the *denominator* of the *F* ratio. Learn to identify each from your output and make sure that you include both.

4 *Do not copy over the entire ANOVA summary table and hope for the best!* This is no solution. You have to identify in text the various sources that you are testing with your ANOVA and the precise value of *F* and associated degrees of freedom in each case. In fact, researchers tend not to put ANOVA summary tables in RESULTS sections unless there are a lot of effects to report. When they do include ANOVA summary tables, they report only certain parts of the output. You can find more about this in Section B of the book's Web site at **http://mcgraw-hill.co.uk/openup/harris/**

4.6.12 Linear regression

11 Linear regression techniques are a family of flexible and useful analytical tools based on correlation. You can find more both about regression and how to write up studies using regression in Section B of the Web site that accompanies this book at: **http://mcgraw-hill.co.uk/openup/harris/.**

In linear regression, we extend the analysis of correlation by treating one or more of the variables as predictors and one of the variables as an outcome or predicted variable. This has many advantages, but if you have used regression you should watch out in your DISCUSSION for one major disadvantage – the terminology of prediction makes it very easy to forget that you are usually dealing with correlational data. This can be exacerbated by the tendency in statistics packages and some statistics textbooks to refer to the predictor variables as IVs and the outcome variable as the DV. This is why some researchers have suggested we should use more neutral terms (see Section 13.8 for more on this). Remember, you cannot infer that the predictor *causes* any significant changes in the outcome variable unless the predictor has been experimentally manipulated (and only then after assessing the level of control in the experiment – see Chapter 5). For most of you, most of the time, when you use linear regression this will *not* be the case and you will need to be careful not to infer causality from your data.

In simple linear regression we have one predictor and one outcome variable. In multiple linear regression we have more than one predictor variable. A good set of predictors for multiple linear regression will be strongly correlated with the outcome variable but not with each other.

Linear regression tells us about the extent to which we can describe the relationship between two variables, a predictor variable and an outcome variable, using a straight line. Regression analysis provides information about how much the outcome variable changes with every unit change in the predictor variable (i.e., the *slope* of the line). This can be really useful. Regression provides two statistics to describe this:

B describes this relationship using the original units. It shows how much the outcome variable changes in whatever units it was measured for every 1 unit change in the predictor variable (in whatever units that was measured). So, for example, $B = +0.50$ between a rating of attitude (predictor) and a rating of intention (outcome) shows that for every 1 point increase in the rating of attitude, intention goes up half (.50) a rating on the intentions scale.

β (pronounced "beta") describes exactly the same relationship, but translated into a standard score rather than using the original units of measurement. It tells us how many standard deviations the outcome variable changes for 1 standard deviation change in the predictor variable. This sounds esoteric, but is actually not that hard to understand and it is really useful as it means we can compare the effects of different predictors, even though they may have different units of measurement.

Simple linear regression

Statistical packages will provide you with a range of statistics following a regression analysis. As well as B and β, you should find a t value that tests whether the slope of the line differs significantly from zero and also the correlation coefficient, Pearson's r, which provides information about the strength and direction of the linear relationship between the two variables (see Section 4.6.3). You will also find an effect-size estimate, R^2. As a minimum, you should report these. You can find out more about each of these statistics, plus some of the other statistics you will get in your output, in Section B of the book's Web site and also in the statistical textbooks that are paired with this book.

In the following study, 102 participants completed an individual difference measure of anxiety and also rated how lucky they thought they were. How much does being anxious predict how much they think of themselves as being lucky?

> An alpha level of .05 was used. Analysis of the data using simple linear regression revealed that anxiety explained a small but statistically significant (5%) proportion of the variance in belief in personal luck, $R^2 = .054$, adjusted $R^2 = .044$, $F(1, 100) = 5.68$, $p = .019$. The relationship between anxiety and belief in luck was negative,
> 12 $\beta = -.23$, $p = .019$, with increases in anxiety being associated with lower level of belief that one was a lucky person.

Multiple linear regression

In multiple linear regression there is more than one predictor variable. The output will give you an estimate of the combined predictive power of all the variables together, R, plus an estimate of the effect size of these combined predictors, R^2, which tells you the percentage variance in the outcome variable predicted (or "explained") by the combined predictors. As well as this, your output will include B and β for each predictor separately, together with a t value that tests whether the slope of the line between *that* predictor and the outcome variable differs significantly from zero. The difference between B and β here and in a simple regression is that in multiple regression these statistics control for the presence of the other predictors in the model. What this means in practice is that β tells you the relative contribution of each individual predictor to the model. So, with multiple linear regression we are able not only to assess how much variance in the outcome variable can be accounted for by the combined predictors but also how much each individual predictor contributes to the model.

In the above study, in fact, the 102 participants had completed a number of other measures, including a measure of self-esteem and a measure of how much they believed that events in the world (good or bad) happened to people at random. For theoretical reasons, the researchers were interested in whether these variables – anxiety, self-esteem and randomness – combined to predict belief in personal luck and which of them would be the most important predictor. All variables were scored so that higher values indicated more anxiety, self-esteem and belief in luck and randomness.

> An alpha level of .05 was used. The means, standard deviations and intercorrelations between the variables are presented in Table 1. As can be seen in Table 1, the outcome variable, belief in personal luck, was significantly correlated with all predictor variables, with greater belief in personal luck being associated with greater belief in randomness, higher self-esteem and lower anxiety. The predictor variables had small but significant correlations with each other, the strongest being that between anxiety and self-esteem, $r(100) = -.37$, $p < .001$.
>
> Analysis of the data using simultaneous multiple linear regression revealed that the combined predictors explained 25% of the variance in belief in personal luck, $R^2 = .25$, adjusted $R^2 = .22$, $F(1, 97) = 10.63$, $p < .001$. Belief in randomness was the only significant individual predictor, $\beta = .43$, $p < .001$; neither self-esteem, $\beta = .07$, $p = .50$, nor anxiety, $\beta = -.15$, $p = .13$, were significant individual predictors in the final model.

Where you have more than three predictors it may make it easier for your reader if you put the results from the analysis in suitably formatted and labelled table rather than in text. You can find advice on what to put in such a table in Section B of the book's Web site at **http://mcgraw-hill.co.uk/openup/harris/**

Like ANOVA, in practice, there are several different variants of multiple regression as well as different labels for the same method; for example, you may use simultaneous, hierarchical or logistic regression, and you may also encounter stepwise regression. You can find out more about these different types in Section B of the book's Web site. As ever, you should be specific and accurate when reporting which the type you have used.

4.6.13 Statistics of effect size

Once you know how to calculate statistics of effect size and understand what they mean, you should include them when reporting your inferential statistics. Effect size statistics, such as *d* or partial eta squared (η^2), estimate the effect your IV has on your DV. Others, such as r^2, estimate the strength of the relationship between two correlated variables. Most of the time all you need to do to report these statistics is to add them to the relevant part of the text or the table in which you report the statistic whose effect they estimate. You can find more about effect size in Chapter 12, where you can also find an example of a RESULTS section that incorporates the reporting of an effect size statistic, partial eta squared (η^2). You should only include such statistics once you have been taught about them. As ever, you should never include statistics that you do not understand or recognize, however tempting it is to try to give the impression that you are on top of things when you are not.

4.7 What you can find on the book's Web site

As you progress you may well find yourself needing to report other statistics than those covered in this chapter or to supplement your reports with additional analyses. You can find in Section B of the Web site material on more advanced reporting, including coverage of issues such as the reporting of planned comparisons or post hoc tests, of tests designed to unpack statistically significant interactions (tests of simple effects), a discussion of when to include ANOVA tables in your report and what to put in them, and more on linear regression. You can find in Section H of the Web site a discussion of what material to put in your later RESULTS and in Section I material on preparing a final year project. You can find a full listing of the contents of the Web site at the front of this book.

4.8 What you can find in the statistics textbooks paired with this book

The books paired with *Designing and Reporting Experiments* have the following coverage:

Learning to use statistical tests in psychology – Greene and D'Oliveira
Greene and D'Oliveira cover all the statistical tests described in Sections 4.6.1–4.6.12 of this book.

SPSS survival manual – Pallant
Pallant discusses descriptive statistics in Chapter 6 and covers all the statistical tests described in Sections 4.6.1–4.6.13 of this book. (Note, however, that she refers to predictors and outcomes in linear regression as independent and dependent variables. See Section 4.6.12 for more on why this may pose problems for you.)

Both books have very good coverage of the issues involved in choosing statistical tests.