S-Plus for
The Analysis of Biological Data

Jones  ·  Gilliver  ·  Robson  ·  Edwards
S-Plus for the
Analysis of Biological Data

Rhondda Jones
Robin Gilliver
Simon Robson
&
Will Edwards
Contents

Preface xv
  Why S-Plus? xv
How to use the manual xvii
To the instructor xvii
Acknowledgements xviii
Typesetting conventions used in the manual xix

1 Introduction to S-Plus 1
1.1 Starting S-Plus in Windows 2
  Choosing or setting a working directory 2
1.2 The S-Plus main program window 3
  Getting help 3
1.3 Import data to create a new S-Plus data set 4
1.4 Create a new empty data set to enter sample data manually 6
1.5 The Object Explorer 7
  Changing the properties of an object 8
  Examining the details of a Data object 8
1.6 Data types in S-Plus 9
1.7 Data transformation: creating and modifying data 10
  Steps to be taken in S-Plus 11
  • Checking and modifying data types 11
  • Calculating a time interval in days 11
  • Calculating a time interval in weeks 12
  • Calculating a proportional weight gain 12
  • Calculating weekly growth rates 13
  • Creating a logical variable 13
  Single-function transforms 15
  Other options provided by the Data menu 15
  Examining the data set with the Object Explorer 15
  Special values and reserved words 16
1.8 Data Objects in S-Plus 16
1.9 Introduction to the Commands Window  
   Using the commands window as a calculator  
   Using S-Plus standard functions  
   Giving names to objects: assignment commands  
   Creating vectors  
   • Creating vectors with the `c()` function  
   • Creating vectors with the `scan()` function  
   • Naming elements of a vector with the `names()` function  
   • Creating vectors with sequential or repeating data: `seq()` and `rep()`  
   • Creating vectors from probability distributions: `rnorm()` and `runif()`  
   Working with vectors  
   Drawing plots by entering commands  
   Using data frames on the command line  
   Working with parts of a data frame  
   Creating a data frame in the commands window  

1.10 Using the Script Window  
   Opening the script window  
   Entering and running a script  
   Writing functions  

1.11 S-Plus language and functions  

1.12 References and further reading  
   Test Your Skills  

2. Displaying data  

2.1 Displaying frequency distributions  
   Bar graphs for categorical data  
   Frequency tables and histograms for numerical data  

2.2 Quantiles of a frequency distribution  
   Plotting a cumulative frequency distribution  

2.3 Associations between categorical variables  
   Creating a grouped bar plot  
   Creating a stacked bar plot  
   Creating a mosaic plot  

2.4 Comparing numerical variables between groups  
   Using box plots  
   Using trellis graphics to compare histograms  
   Comparing cumulative frequencies for different groups
### Contents

2.5 Displaying relationships between a pair of numerical variables  50  
  Scatter plots  50  
  Line graphs  51  
  Putting several graphs on the same graph sheet  51  
  Varying symbols between groups on the same plot  52  
  Plotting fitted lines to scatter plots  53  
  Test your skills  56

3. Describing data  61  
3.1 Arithmetic mean and standard deviation  61  
  Data as individual values  61  
  Data as a frequency table  62  
3.2 Median and interquartile range  63  
3.3 How measures of location and spread compare  64  
  Descriptive statistics with the GUI  64  
3.4 Proportions  67  
  Calculating proportions using the GUI  67  
  Calculating proportions using the commands window  68  
  Test Your Skills  69

4. Estimating with uncertainty  71  
4.1 The sampling distribution of an estimate  71  
4.2 Measuring the uncertainty of an estimate  74  
  The standard error of the mean  75  
4.3 Standard errors and confidence intervals for the sample mean from the GUI  75  
4.4 Plotting means and their confidence limits  76  
  Test your skills  78

5. Probability distributions  79  
5.1 Some terminology  79  
5.2 Probability  79  
5.3 What is a probability distribution?  80  
5.4 Using S-Plus to calculate probabilities for a binomial distribution  81  
  Calculating binomial probabilities using the GUI  81  
  Calculating binomial probabilities using the commands window  81  
5.5 What other information might you need from a probability distribution?  83  
5.6 Another common discrete probability distribution: the Poisson  84  
5.7 Continuous probability distributions in S-Plus: the normal distribution  85
5.8 Other key continuous probability distributions 88
   The Chi-square distribution 89
   The t-distribution 90
   The F-distribution 91

Test your skills 92

6. Hypothesis testing: matching hypotheses and analyses in S-Plus 93
   6.1 Making and using hypotheses 93
      Null hypothesis 93
      Alternative hypothesis 93
      When do we reject the alternative hypothesis? 94
      One-tailed and two-tailed tests 94
   6.2 Hypothesis testing in practice 95
   6.3 Hypotheses about frequencies and proportions 97
      Testing a single proportion or relative frequency 97
      Testing a set of proportions or relative frequencies from a single sample 97
      Testing for independence of two categorical variables, each with two levels (2 x 2 contingency tables) 97
      Testing for independence of two categorical variables with an unspecified number of levels in each (n1 x n2 contingency tables) 98
      Testing for independence of two n categorical variables 98
      Testing whether a proportion is related to (can be predicted from) a numeric variable 98
   6.4 Hypotheses about the shape of distributions 98
      Testing for normality 98
      Testing for Poisson, binomial, chi-square, uniform and a variety of other distributions 98
   6.5 Hypotheses about one, two or n means 99
      Testing a single mean 99
      Comparing two means 99
      Comparing n means 99
   6.6 Hypotheses about one, two or n medians 99
      Testing a single median 99
      Comparing two medians 100
      Comparing n medians 100
   6.7 Hypotheses about variances 100
      Comparing two variances 100
      Comparing n variances 100
6.8 Hypotheses about relationships between two numerical variables $x$ and $y$ 100
   Identifying relationships 101
   Testing the form of relationships 101
   Examining regression assumptions 101

6.9 Hypotheses involving multiple explanatory variables 102
Test your skills 103

7. Analysing proportions 105
7.1 The binomial distribution 105
   Calculating binomial probabilities using the commands window 106
   Properties of the sampling distribution for a proportion 107
7.2 Testing a proportion: the binomial test 108
7.3 Estimating proportions 109
   Estimating the standard error of a proportion 109
   Estimating confidence limits for a proportion 109
Test your skills 111

8. Fitting probability models to frequency data 113
8.1 Example of a random model: the proportional model 113
8.2 $\chi^2$ goodness-of-fit test 114
8.3 Assumptions of the $\chi^2$ goodness-of-fit test 115
8.4 Goodness-of-fit tests when there are only two categories 116
8.5 Fitting the binomial distribution 117
   Testing goodness of fit to a binomial using frequency data 117
   Testing goodness of fit to a binomial distribution with individual values 118
8.6 Random in space or time: the Poisson distribution 119
   Using the S-Plus GUI to test goodness-of-fit to a Poisson distribution 120
   Using the commands window to test goodness-of-fit to a Poisson distribution 124
Test your skills 126

9. Contingency analysis: associations between categorical variables 129
9.1 Associating two categorical variables 129
   Contingency tables, proportional plots, and a $\chi^2$ contingency test on
categorical data for individuals 129
   Creating a mosaic or stacked bar plot 131
9.2 Estimating association in 2 x 2 tables: odds ratio 132
   Using S-Plus to calculate odds and the odds ratio 132
9.3 The \( \chi^2 \) contingency test for \( n \times n \) tables 133
   What if S-Plus warns that some expected values are too low? 135
9.4 Fisher's exact test 137
9.5 Log-linear models and G-tests 138
   Two categorical variables: using log-linear modelling to execute a G-test 138
   A more complex example 140
Test your skills 143

10. The normal distribution 145
10.1 Bell-shaped curves and the normal distribution 145
10.2 Exact probability estimates for normal distributions 146
10.3 Properties of the normal distribution 147
10.4 The standard normal distribution 148
   Using normal distributions to answer questions about populations 149
10.5 The normal distribution of sample means 150
10.6 The Central Limit theorem 151
10.7 The normal approximation for the binomial distribution 151
   Using the normal approximation for the binomial 152
   Using the binomial probabilities 152
Test Your Skills 154

11. Inference for a normal population 155
11.1 The \( t \)-distribution for sample means 155
   Using S-Plus to find values of \( t \) from probability values 157
   Using S-Plus to find probabilities from the values of \( t \) 158
11.2 The confidence interval for the mean of a normal distribution 159
   Calculating confidence limits from the original data in the commands window 160
11.3 The one-sample \( t \)-test 161
   The effects of larger sample size: body temperature revisited 163
11.4 Confidence intervals for the standard deviation and variance of a normal population 164
Test Your Skills 166

12. Comparing two means 169
12.1 Paired samples versus independent samples 169
12.2 Paired comparison of means 170
12.3 Two-sample comparison of means 172
   A two-sample t-test where variances can be assumed to be equal 172
   A two-sample t-test where variances cannot be assumed equal 174
12.4 Using the correct sampling units 175
12.5 Avoid indirect comparisons 176
12.6 Interpreting overlap of confidence intervals 177
12.7 Comparing variances 177
Test your skills 179

13. Handling violations of assumptions 181
13.1 Detecting deviations from normality 181
   Graphical methods 181
   Formal tests of normality 184
13.2 When to ignore violations of assumptions 184
13.3 Data transformations 184
13.4 Non-parametric alternatives to one-sample and paired t-tests 186
   The sign test 187
13.5 Non-parametric comparisons of two groups 188
   The Wilcoxon rank-sum test (Mann-Whitney U-test) 188
   The Kolmogorov-Smirnov test 190
Test your skills 192

14. Experimental design 193
14.1 Why (and why not) do experiments? 193
   Confounding variables and experimental artifacts 193
   When an observational study is required 194
14.2 How to reduce bias 195
14.3 How to reduce the impact of sampling error 197
14.4 Experiments with more than one factor 199
14.5 Choosing a sample size 201
   Designing for precision 202
   Designing for power 202
   • Power and sample size calculations for testing a single proportion 203
   • Power and sample size calculations for comparing two means 205
   • Other options 206
   Designing for data loss 206
Test your skills 207
15. Comparing means of more than two groups  209

15.1 The analysis of variance  210

- Executing a one-way ANOVA in S-Plus  211
- Interpreting and reporting the results: the formula  212
- Interpreting and reporting the results: sums of squares, degrees of freedom, and mean squares  213
- Interpreting and reporting the results: the ANOVA table  214
- Interpreting and reporting the results: the R² value  214
- Interpreting and reporting the results: summarising the data values  215
- What you should report  215
- How sums of squares are calculated  215

15.2 Assumptions and alternatives  217

15.3 Planned comparisons  218

- Planned comparisons between two means  219

15.4 Unplanned comparisons  220

- Testing all pairs of means  220

15.5 Fixed and random effects  222

15.6 ANOVA with randomly chosen groups  222

- Variance components and repeatability  224

Test your skills  225

16. Correlation between numerical variables  227

16.1 Estimating a linear correlation coefficient  227

- The correlation coefficient  228
- Standard error and confidence interval  229

16.2 Testing the null hypothesis of zero correlation  229

- Reporting the results of Pearson’s correlation  230

16.3 Assumptions  231

16.4 The correlation coefficient depends on the range  231

16.5 Spearman’s rank correlation  233

- Reporting the results of Spearman’s rank correlation  234
- Assumptions of Spearman’s rank correlation  235

16.6 The effects of measurement error on correlation  235

Test your skills  236
17. Regression 239

17.1 Linear regression 240
   The method of least squares 241
   Executing a regression analysis through the Statistics menu 243
   Interpreting and reporting the output: the function call 243
   Interpreting and reporting the output: the residuals summary 244
   Interpreting and reporting the output: coefficients and their standard errors 244

17.2 Confidence in predictions 244
   Interpreting and reporting the output: the prediction interval 245
   Interpreting and reporting the output: confidence bands 246
   Interpreting and reporting the output: the R² value 247
   Interpreting and reporting the output: testing hypotheses about the regression line 247

17.3 Doing the analysis in the commands window 248

17.4 What you should report 248

17.5 Assumptions of regression 249
   Outliers 249
   Detecting deviations from the assumptions: linearity 251
   Detecting variations from the assumptions: non-normality and unequal variance 253

17.6 Transformations 254

17.7 The effects of measurement error 255

17.8 Non-linear regression 255
   A curve with an asymptote 255
   Quadratic and polynomial curves 257
   Formula-free curve fitting 258
   Logistic regression: fitting a binary response variable 259

17.9 Multiple linear regression 263
   A warning 268
   A further warning 268

17.10 References 269
   Test your skills 270

18. Multiple explanatory variables 273

18.1 Defining a model in S-Plus 274

18.2 Analysing experiments with blocking: the randomised block design 275
   Analysing data from a randomised block design 275
Chapter 2: Functions for plotting data  328
  2.1 Displaying frequency distributions  328  
    Bar graphs and dot plots for categorical data  328  
    Frequency histogram  329  
  2.2 Cumulative frequency distribution  330  
  2.3 Associations between categorical variables  330  
    Mosaic plot  330  
    Grouped bar plots  330  
    Stacked bar plot  330  
  2.4 Comparing numerical variables between groups  331  
    Trellis graphics  331  
    Comparing frequency histograms  332  
    Box plots  333  
  2.5 Displaying relationships between a pair of numerical variables  333  
    Scatter plots  333  
    Line graphs  333  
    Varying symbols between groups on the same plot  334  
    Plotting fitted lines to scatter plots  334  

Chapter 3: Functions for describing data  336  
  3.1 Examining the whole data frame  336  
  3.2 Descriptive statistics for individual vectors  337  
    Measures of location  338  
    * Measures of dispersion  338  
    Measures of distribution shape  339  
    To calculate descriptive statistics for subsets of an individual vector  340  

Chapter 4: Estimating with uncertainty – functions for calculating standard errors and confidence limits  340  
  4.1 Standard error of the mean  340  
  4.2 Confidence limits of the mean for normally distributed data  341  
  4.3 Confidence limits for the variance and standard deviation for normally distributed data  341  
  4.4 Confidence limits for descriptive statistics which do not require the assumption of normality  341
Appendix 2: Scripts used in each chapter  344

Chapter 1:

plot.summarize()
Generates a set of descriptive plots and returns summary statistics for a numeric vector  344

growth.rate()
Calculates and returns growth rate per time unit given start and end sizes, and times  344

Chapter 4:

sample.means()
Calculates and returns the means of a set of random samples taken from a numerical vector x  345

summaries()
Calculates and returns in a form suitable for plotting, the means, standard deviations, sample sizes, and half-confidence intervals for a numerical vector subdivided according to the levels of one or two categorical variables  345

Chapter 7:

Cl.p.agresti()
Calculates and returns a proportion and its confidence limits using the Agresti-Coull approximation for confidence limits  346

Cl.p.exact()
Calculates proportion and its confidence interval with a specified tolerance  347

Chapter 8:

chisquare.gof()
Executes a chi-square goodness of fit for any specified set of observed counts, expected proportions, and degrees of freedom  348

Chapter 9:

contingency.expected()
Calculates expected values for a contingency table provided as an array or data frame  348

oddsratio()
Calculates an odds ratio and its confidence interval given the numbers of 'successes' and 'failures' in two samples  349

Chapter 10:

p.outside()
Calculates the area of a normal curve outside the interval upper - lower  349

Chapter 11:

Cl.mean()
Calculates the mean and confidence limits for a numerical vector  350

Cl.var()
Calculates sample variances & standard deviation, and a confidence interval for each, from a numeric vector  351
onesample.t()
Executes a one-sample t-test from previously-calculated descriptive
statistics: arguments are a hypothesized mean, a sample mean, a
sample standard deviation, and a sample size 352

Chapter 12:
twosample.t()
Executes a two-sample t-test (assuming homogeneous variances)
from previously-calculated descriptive statistics: arguments are the
mean, standard deviation, and sample size from two samples 352

levene.test()
Executes a Levene test for homogeneity of variances given a numerical
vector and a grouping variable of the same length 353

Chapter 13:
sign.test()
Executes a sign test to test whether the median of x could equal some
specified value 353

Chapter 16:
Cl.r()  
Calculates confidence limits for a previously-calculated Pearson
correlation coefficient, given values for r and the sample size.
Uses the Fisher approximation 354

Index 355
Preface

This manual is designed to teach people to use the statistical software S-Plus and to support the process of learning statistical concepts and methods. It is most useful as a workbook to accompany Whitlock and Schluter's *The Analysis of Biological Data*, published by Roberts & Company, Colorado. Although we include enough statistical background to put the procedures being demonstrated in context, we assume that readers will be acquiring most of their understanding of statistical concepts elsewhere.

Several of the authors of this manual have been teaching introductory biostatistics to undergraduate and postgraduate students on two campuses in Australia for more than a decade (in fact one of us, who would prefer not to be identified, taught a biostatistics course for the first time more than three decades ago). In 2008 we discovered the textbook *The Analysis of Biological Data* (referred to in this manual as ABD). We liked everything about the book: its explanations were beautifully clear and aimed at students much like our own; it used a wide variety of real biological examples; it emphasized concepts and procedures important to biologists and explained how they worked; and it introduced some newer computer-intensive techniques that almost all beginning researchers find themselves needing sooner rather than later. We immediately adopted the book as a text for our own introductory biostatistics course. But this adoption acted as a trigger for making some other changes to our teaching—and in particular, to the way we introduced students to statistical software.

To statistical novices, no statistical software is ‘user-friendly’, and its use needs to be introduced in a structured way which runs in parallel with their acquisition of statistical understanding. At the same time, teaching effort needs to stay focused on statistics rather than software, so that students do not come to see learning to use the software as their primary goal. This manual is intended to allow users to learn to use the software on their own, while keeping a focus on the concepts and procedures which it supports.

We have followed the ABD approach and layout very closely—indeed, we started out with the intention of simply demonstrating in S-Plus every example used in the body of that text. In the end, because everyone has a slightly different view of what should be included in a first statistics course, we added a number of other examples, mostly using our own data, to demonstrate software capabilities that would not otherwise have been covered.

Also we did not include material associated with ABD Chapter 20 (*Likelihood*) or with Chapter 21 (*Meta-analysis*) which are largely conceptual. Most of the computational procedures in Chapter 20 are covered elsewhere in the manual.

**Why S-Plus?**

There are a lot of statistical software options, and most of them will execute all the procedures needed in an introductory course. In choosing a software package, we had four criteria beyond its ability to execute procedures taught in the course:

- It should have little or no cost to students, and should run on operating systems that students are likely to use on their own machines. Some of us (OK, one of us) remembered teaching statistics in the days when the only computing aid available to students was
a hand calculator; the rest of us at least remember being taught that way. While we did not wish to return to those days, they had one huge advantage—students could work on the material anywhere and any time, and not just in computer laboratories provided by the university. Many of our students are part-time, and some are in remote locations. While we can now reasonably expect that students will have access to a computer at home, we cannot reasonably expect them to buy expensive software for themselves. That meant that if we wanted students to work off-campus, we needed to choose software which was either free or very cheap, or which gave students access on their own machines as part of the university’s site licence.

- **It should be useful beyond the course.** We wanted students to use professional-quality software that they would not ‘grow out of’: providing access to all or most of the techniques they were likely to use throughout their careers; and able to import and export data in a wide range of formats (including text files, databases, spreadsheets, and other statistical software).

- **It should have a very strong graphics capability.** We wanted students to realise as quickly as possible that nothing substitutes for an intimate familiarity with the data they are analysing—and easily-usable graphics allow the data to be explored more quickly and thoroughly than anything else. We wanted the graphics capability to cover the whole range from quick-and-dirty exploratory plots to presentation and publication-quality graphs.

- **It should reinforce the statistical concepts we wanted students to grasp, and not get in the way of learning them.** We wanted to avoid both excessive or inappropriate output, and too much difficulty in using the software itself. Excessive output is often a problem with menu-driven software, which may be relatively easy to use, but often provides pages of output that users neither asked for nor know what to do with. Especially for novices, our preference was for software that gives users exactly what they request and offers warnings (or refuses to perform) when what they request is questionable. We believe that someone learning to use statistical procedures should also learn to think about what they are doing and work out exactly what it is they want, rather than making guesses about what button to click in the hope that something useful will happen. On the other hand, if software is too difficult to use, students will inevitably concentrate on learning the mechanics of how to use it rather than developing more fundamental understanding.

In the end we chose S-Plus as the best fit to our needs. That choice committed us to producing this manual: there are some excellent introductory books available for S-Plus, but none that we investigated is targeted at undergraduates who begin as complete statistical novices. S-Plus is very powerful and flexible, has superb editable graphics, and its site licence for universities gives enrolled students permission to use the software on their own computers. While it provides a professional-quality graphical user interface (GUI), it also has an easily-accessible command language. Mostly we use the S-Plus GUI, but we also provide a parallel introduction to the command line and to writing basic scripts.

Since we made this decision, a final issue has become more important. That is, the increasing importance of the open-source statistical software R in the biological research community. Learning the R language as a complete
statistical novice is a hard ask for students, who are often having quite enough difficulty with statistical concepts. But S-Plus shares its command language with R—and our experience has been that students make the transition to R quite easily by the end of an introductory course based on S-Plus.

**How to use the manual**

If you are a student using ABD as a text, and you have access to S-Plus, you can use S-Plus to work through each chapter of the manual independently. Every example is demonstrated in enough detail for you to carry it out on your own after reading the ABD chapter and/or covering the statistical concepts in class. You should execute every example yourself to make sure that you can carry out the procedures correctly and get to the right result. A set of exercises is provided at the end of each chapter for you to test your skills. You should make sure that you can do all those marked essential—you may require assistance from your instructor to successfully complete some of them, possibly the advanced exercises. All the data and scripts required for each chapter are available in the resource material provided with the manual.

The first chapter of the manual is a basic introduction to S-Plus, and is one of a few chapters whose content is not linked to ABD. The second chapter introduces you to S-Plus graphics. The remaining chapters can be covered in several different orders, but you need to work through these two first. Not all the material in later chapters will necessarily be included in an introductory course.

In most chapters, we show how to execute statistical procedures using both the GUI and the command line. A few procedures require the use of scripts (short programs written in the S language). Where this is the case, we provide the scripts in the resource material, with them being reproduced in Appendix 2—and we show you how to load and use them (but we also encourage you to learn to write your own). In many cases, there are more efficient or elegant ways to write scripts than we have used here—in general, we have tried to produce scripts whose logic can be easily understood by beginners, rather than trying for maximum computational efficiency. Appendix 1 provides a more extensive summary of the S language and S-Plus functions relevant to each chapter of the manual.

**To the instructor**

We believe that learning statistics is like learning to play the piano—there is no substitute for practice. Consequently, in our own teaching, we provide a lot of incentives for students to practice.

In the introductory course that we teach, we expect students to have worked through the appropriate chapter(s) in the manual and attempted the exercises before they arrive at the relevant practical class or tutorial—and the first 20 minutes of each 2-hour practical class includes a simple open-book practical test, marked in class, which requires them to analyse some new data using techniques covered in the chapter. (By the end of the course, most students score full marks on most of these tests.) We also run formal (but also open-book) practical exams twice during the course, where the emphasis is on demonstrating that students can make sensible decisions about what to do as well as demonstrating that they can do it. These are also graded immediately. Because students can take this manual—or anything else—into practical tests and exams, we are explicitly not testing how well they remember what buttons to click.
When we first changed to this very assessment-oriented approach to the acquisition of practical skills, one unexpected result was that the average grade on the theory exam at the end of the course (which was in the same format and covered the same material as previously) was significantly higher than that achieved by any previous class; it has remained at this level in subsequent years. Perhaps the development of practical skills really does improve theoretical understanding.

Acknowledgements
As noted above, the structure and content of this manual owes a huge debt to Whitlock and Schluter's text, which provides the best introduction we know of to statistical methods for biology students. We are also very grateful to the students in our 2009 biometrics class, and especially to the practical class tutors (Clwedd Burns, Gavin Coombes, Rie Hagihara, and Philip Newey) whose combined input and feedback improved the manual immensely. Finally, for all his help our thanks to Kris Angelovski of SolutionMetrics Pty Ltd, the Australian distributor of S-Plus.
Typesetting conventions used in the manual

To make it easier to use the manual, there are several conventions to note.

Navigating – The instructions about how to navigate around S-Plus are always given in a particular typeface. For example, the way to navigate from the drop-down menus looks like:

Statistics > Regression > Log-linear (Poisson)...

Similarly, this typeface is used to indicate the parts of a dialogue box you need to change.

Entering new information – Where you are required to enter a name or value in a dialogue box or change an existing name, such as perhaps the name of a column (vector) in a data set, the instruction might look like:

Right click at the top of the No.deaths column, select Properties, and enter **Number of cases (Frequency)** in the Description box.

This **typeface** represents a name or value that you can either enter or change, whereas the **navigation typeface** cannot be altered.

Coding – You learn in Chapter 1 the significance of single lines of code. When programming, pressing the Enter key always means something. Where we have reproduced lines of code that you will see on your computer screen, it has often been necessary to spread them over more than one line because the printed page is narrower than your screen. Where this has happened, the subsequent lines of that instruction have been indented. The following example shows two separate instructions (in a lighter shade) and the screen output, with the screen prompts [ > ] omitted.

```r
  titanic.array = table(titanicAdults$Gender,
                        titanicAdults$Outcome)
  titanic.array
  Died  Survived
  Female  109  316
  Male    1329 338
```

Quotation marks when used in code should be straight (" " ) and not curly ( " " ), sometimes called “smart” quotes. In S-Plus code, curly quotes will generate an error message.

These conventions will become clear as you work through the chapters.