Contents

Preface *page* **xix**

Content – how the chapters fit together **xxv**

1 A brief introduction to R 1

1.1 An overview of R 1

1.1.1 A short R session 1

1.1.2 The uses of R 6

1.1.3 Online help 7

1.1.4 Input of data from a file 8

1.1.5 R packages 9

1.1.6 Further steps in learning R 9

1.2 Vectors, factors, and univariate time series 10

1.2.1 Vectors 10

1.2.2 Concatenation – joining vector objects 10

1.2.3 The use of relational operators to compare vector elements 11

1.2.4 The use of square brackets to extract subsets of vectors 11

1.2.5 Patterned data 11

1.2.6 Missing values 12

1.2.7 Factors 13

1.2.8 Time series 14

1.3 Data frames and matrices 14

1.3.1 Accessing the columns of data frames – `with()` and `attach()` 17

1.3.2 Aggregation, stacking, and unstacking 17

1.3.3* Data frames and matrices 18

1.4 Functions, operators, and loops 19

1.4.1 Common useful built-in functions 19

1.4.2 Generic functions, and the class of an object 21

1.4.3 User-written functions 22

1.4.4 `if` Statements 23

1.4.5 Selection and matching 23

1.4.6 Functions for working with missing values 24

1.4.7* Looping 24

1.5 Graphics in R 25
1.5.1 The function `plot( )` and allied functions 25
1.5.2 The use of color 27
1.5.3 The importance of aspect ratio 28
1.5.4 Dimensions and other settings for graphics devices 28
1.5.5 The plotting of expressions and mathematical symbols 29
1.5.6 Identification and location on the figure region 29
1.5.7 Plot methods for objects other than vectors 30
1.5.8 Lattice (trellis) graphics 30
1.5.9 Good and bad graphs 32
1.5.10 Further information on graphics 33
1.6 Additional points on the use of R 33
1.7 Recap 35
1.8 Further reading 36
1.9 Exercises 37

2 Styles of data analysis 43
2.1 Revealing views of the data 43
2.1.1 Views of a single sample 44
2.1.2 Patterns in univariate time series 47
2.1.3 Patterns in bivariate data 49
2.1.4 Patterns in grouped data – lengths of cuckoo eggs 52
2.1.5* Multiple variables and times 53
2.1.6 Scatterplots, broken down by multiple factors 56
2.1.7 What to look for in plots 58
2.2 Data summary 59
2.2.1 Counts 59
2.2.2 Summaries of information from data frames 63
2.2.3 Standard deviation and inter-quartile range 65
2.2.4 Correlation 67
2.3 Statistical analysis questions, aims, and strategies 69
2.3.1 How relevant and how reliable are the data? 70
2.3.2 How will results be used? 70
2.3.3 Formal and informal assessments 71
2.3.4 Statistical analysis strategies 72
2.3.5 Planning the formal analysis 72
2.3.6 Changes to the intended plan of analysis 73
2.4 Recap 73
2.5 Further reading 74
2.6 Exercises 74

3 Statistical models 77
3.1 Statistical models 77
3.1.1 Incorporation of an error or noise component 78
3.1.2 Fitting models – the model formula 80

3.2 *Distributions: models for the random component* 81
3.2.1 Discrete distributions – models for counts 82
3.2.2 Continuous distributions 84
3.3 *Simulation of random numbers and random samples* 86
3.3.1 Sampling from the normal and other continuous distributions 87
3.3.2 Simulation of regression data 88
3.3.3 Simulation of the sampling distribution of the mean 88
3.3.4 Sampling from finite populations 90
3.4 *Model assumptions* 91
3.4.1 Random sampling assumptions – independence 91
3.4.2 Checks for normality 92
3.4.3 Checking other model assumptions 95
3.4.4 Are non-parametric methods the answer? 95
3.4.5 Why models matter – adding across contingency tables 96
3.5 *Recap* 97
3.6 *Further reading* 98
3.7 *Exercises* 98

4 A review of inference concepts 102
4.1 *Basic concepts of estimation* 102
4.1.1 Population parameters and sample statistics 102
4.1.2 Sampling distributions 102
4.1.3 Assessing accuracy – the standard error 103
4.1.4 The standard error for the difference of means 103
4.1.5* The standard error of the median 104
4.1.6 The sampling distribution of the t-statistic 105
4.2 *Confidence intervals and tests of hypotheses* 106
4.2.1 A summary of one- and two-sample calculations 109
4.2.2 Confidence intervals and tests for proportions 112
4.2.3 Confidence intervals for the correlation 113
4.2.4 Confidence intervals versus hypothesis tests 113
4.3 *Contingency tables* 114
4.3.1 Rare and endangered plant species 116
4.3.2 Additional notes 119
4.4 *One-way unstructured comparisons* 119
4.4.1 Multiple comparisons 122
4.4.2 Data with a two-way structure, i.e., two factors 123
4.4.3 Presentation issues 124
4.5 *Response curves* 125
4.6 *Data with a nested variation structure* 126
4.6.1 Degrees of freedom considerations 127
4.6.2 General multi-way analysis of variance designs 127
4.7 *Resampling methods for standard errors, tests, and confidence intervals* 128
4.7.1 The one-sample permutation test 128
4.7.2 The two-sample permutation test 129