

Contents

Preface	vii
1 Basics	1
1.1 First steps	1
1.1.1 An overgrown calculator	3
1.1.2 Assignments	3
1.1.3 Vectorized arithmetic	4
1.1.4 Standard procedures	6
1.1.5 Graphics	7
1.2 R language essentials	10
1.2.1 Expressions and objects	10
1.2.2 Functions and arguments	10
1.2.3 Vectors	12
1.2.4 Missing values	13
1.2.5 Functions that create vectors	13
1.2.6 Matrices and arrays	14
1.2.7 Factors	16
1.2.8 Lists	17
1.2.9 Data frames	18
1.2.10 Indexing	19
1.2.11 Conditional selection	20
1.2.12 Indexing of data frames	21
1.2.13 subset and transform	22
1.2.14 Grouped data and data frames	23

1.2.15	Sorting	24
1.2.16	Implicit loops	26
1.3	The graphics subsystem	27
1.3.1	Plot layout	27
1.3.2	Building a plot from pieces	29
1.3.3	Using <code>par</code>	30
1.3.4	Combining plots	30
1.4	R programming	32
1.4.1	Flow control	32
1.4.2	Classes and generic functions	34
1.5	Session management	34
1.5.1	The workspace	34
1.5.2	Getting help	36
1.5.3	Packages	37
1.5.4	Built-in data	37
1.5.5	<code>attach</code> and <code>detach</code>	38
1.6	Data entry	39
1.6.1	Reading from a text file	39
1.6.2	The data editor	42
1.6.3	Interfacing to other programs	43
1.7	Exercises	44
2	Probability and distributions	45
2.1	Random sampling	45
2.2	Probability calculations and combinatorics	46
2.3	Discrete distributions	47
2.4	Continuous distributions	48
2.5	The built-in distributions in R	49
2.5.1	Densities	49
2.5.2	Cumulative distribution functions	52
2.5.3	Quantiles	53
2.5.4	Random numbers	54
2.6	Exercises	55
3	Descriptive statistics and graphics	57
3.1	Summary statistics for a single group	57
3.2	Graphical display of distributions	61
3.2.1	Histograms	61
3.2.2	Empirical cumulative distribution	63
3.2.3	Q–Q plots	64
3.2.4	Boxplots	65
3.3	Summary statistics by groups	65
3.4	Graphics for grouped data	67
3.4.1	Histograms	67
3.4.2	Parallel boxplots	69

3.4.3	Stripcharts	70
3.5	Tables	72
3.5.1	Generating tables	72
3.5.2	Marginal tables and relative frequency	74
3.6	Graphical display of tables	75
3.6.1	Bar plots	75
3.6.2	Dotcharts	78
3.6.3	Pie charts	78
3.7	Exercises	79
4	One- and two-sample tests	81
4.1	One-sample t test	81
4.2	Wilcoxon signed-rank test	85
4.3	Two-sample t test	86
4.4	Comparison of variances	89
4.5	Two-sample Wilcoxon test	89
4.6	The paired t test	90
4.7	The matched-pairs Wilcoxon test	92
4.8	Exercises	93
5	Regression and correlation	95
5.1	Simple linear regression	95
5.2	Residuals and fitted values	99
5.3	Prediction and confidence bands	103
5.4	Correlation	106
5.4.1	Pearson correlation	106
5.4.2	Spearman's ρ	109
5.4.3	Kendall's τ	109
5.5	Exercises	110
6	ANOVA and Kruskal–Wallis	111
6.1	One-way analysis of variance	111
6.1.1	Pairwise comparisons and multiple testing	115
6.1.2	Relaxing the variance assumption	117
6.1.3	Graphical presentation	118
6.1.4	Bartlett's test	120
6.2	Kruskal–Wallis test	120
6.3	Two-way analysis of variance	121
6.3.1	Graphics for repeated measurements	124
6.4	The Friedman test	124
6.5	The ANOVA table in regression analysis	126
6.6	Exercises	127
7	Tabular data	129
7.1	Single proportions	129

7.2	Two independent proportions	131
7.3	k proportions, test for trend	133
7.4	$r \times c$ tables	135
7.5	Exercises	138
8	Power and the computation of sample size	139
8.1	The principles of power calculations	139
8.1.1	The power of one-sample and paired t tests	140
8.1.2	Power of two-sample t test	142
8.1.3	Approximate methods	142
8.1.4	Power of comparisons of proportions	143
8.2	Two-sample problems	143
8.3	One-sample problems and paired tests	145
8.4	Comparison of proportions	146
8.5	Exercises	146
9	Multiple regression	149
9.1	Plotting multivariate data	149
9.2	Model specification and output	151
9.3	Model search	154
9.4	Exercises	157
10	Linear models	159
10.1	Polynomial regression	160
10.2	Regression through the origin	162
10.3	Design matrices and dummy variables	164
10.4	Linearity over groups	166
10.5	Interactions	170
10.6	Two-way ANOVA with replication	171
10.7	Analysis of covariance	172
10.7.1	Graphical description	173
10.7.2	Comparison of regression lines	177
10.8	Diagnostics	182
10.9	Exercises	188
11	Logistic regression	191
11.1	Generalized linear models	192
11.2	Logistic regression on tabular data	193
11.2.1	The analysis of deviance table	197
11.2.2	Connection to test for trend	199
11.3	Logistic regression using raw data	201
11.4	Prediction	203
11.5	Model checking	204
11.6	Exercises	208

12 Survival analysis	211
12.1 Essential concepts	211
12.2 Survival objects	212
12.3 Kaplan–Meier estimates	213
12.4 The log-rank test	216
12.5 The Cox proportional hazards model	218
12.6 Exercises	220
A Obtaining and installing R	221
B Data sets in the ISwR package	225
C Compendium	247
Index	261