

Contents

	<i>Preface</i>	page xiii
	<i>Acknowledgements</i>	xvii
1	Basic Statistical Terms, Sample Statistics	1
	1.1 Cases, Variables and Data Types	1
	1.2 Population and Random Sample	3
	1.3 Sample Statistics	4
	1.4 Precision of Mean Estimate, Standard Error of Mean	9
	1.5 Graphical Summary of Individual Variables	10
	1.6 Random Variables, Distribution, Distribution Function, Density Distribution	10
	1.7 Example Data	13
	1.8 How to Proceed in R	13
	1.9 Reporting Analyses	17
	1.10 Recommended Reading	18
2	Testing Hypotheses, Goodness-of-Fit Test	19
	2.1 Principles of Hypothesis Testing	19
	2.2 Possible Errors in Statistical Tests of Hypotheses	21
	2.3 Null Models with Parameters Estimated from the Data: Testing Hardy–Weinberg Equilibrium	26
	2.4 Sample Size	26
	2.5 Critical Values and Significance Level	27
	2.6 Too Good to Be True	29
	2.7 Bayesian Statistics: What is It?	30
	2.8 The Dark Side of Significance Testing	32
	2.9 Example Data	35
	2.10 How to Proceed in R	35
	2.11 Reporting Analyses	37
	2.12 Recommended Reading	37
3	Contingency Tables	39
	3.1 Two-Way Contingency Tables	39
	3.2 Measures of Association Strength	44
	3.3 Multidimensional Contingency Tables	46
	3.4 Statistical and Causal Relationship	47
	3.5 Visualising Contingency Tables	49
	3.6 Example Data	50
	3.7 How to Proceed in R	50

3.8	Reporting Analyses	54
3.9	Recommended Reading	54
4	Normal Distribution	55
4.1	Main Properties of a Normal Distribution	55
4.2	Skewness and Kurtosis	56
4.3	Standardised Normal Distribution	57
4.4	Verifying the Normality of a Data Distribution	58
4.5	Example Data	60
4.6	How to Proceed in R	60
4.7	Reporting Analyses	63
4.8	Recommended Reading	64
5	Student's t Distribution	65
5.1	Use Case Examples	65
5.2	t Distribution and its Relation to the Normal Distribution	66
5.3	Single Sample Test and Paired t Test	67
5.4	One-Sided Tests	70
5.5	Confidence Interval of the Mean	72
5.6	Test Assumptions	73
5.7	Reporting Data Variability and Mean Estimate Precision	74
5.8	How Large Should a Sample Size Be?	77
5.9	Example Data	79
5.10	How to Proceed in R	79
5.11	Reporting Analyses	82
5.12	Recommended Reading	83
6	Comparing Two Samples	84
6.1	Use Case Examples	84
6.2	Testing for Differences in Variance	85
6.3	Comparing Means	87
6.4	Example Data	88
6.5	How to Proceed in R	88
6.6	Reporting Analyses	91
6.7	Recommended Reading	91
7	Non-parametric Methods for Two Samples	92
7.1	Mann–Whitney Test	93
7.2	Wilcoxon Test for Paired Observations	95
7.3	Using Rank-Based Tests	97
7.4	Permutation Tests	97
7.5	Example Data	99
7.6	How to Proceed in R	99
7.7	Reporting Analyses	102
7.8	Recommended Reading	103

8	One-Way Analysis of Variance (ANOVA) and Kruskal–Wallis Test	104
8.1	Use Case Examples	104
8.2	ANOVA: A Method for Comparing More Than Two Means	104
8.3	Test Assumptions	105
8.4	Sum of Squares Decomposition and the F Statistic	106
8.5	ANOVA for Two Groups and the Two-Sample t Test	108
8.6	Fixed and Random Effects	108
8.7	F Test Power	109
8.8	Violating ANOVA Assumptions	110
8.9	Multiple Comparisons	111
8.10	Non-parametric ANOVA: Kruskal–Wallis Test	115
8.11	Example Data	116
8.12	How to Proceed in R	117
8.13	Reporting Analyses	127
8.14	Recommended Reading	128
9	Two-Way Analysis of Variance	129
9.1	Use Case Examples	129
9.2	Factorial Design	130
9.3	Sum of Squares Decomposition and Test Statistics	132
9.4	Two-Way ANOVA with and without Interactions	134
9.5	Two-Way ANOVA with No Replicates	135
9.6	Experimental Design	135
9.7	Multiple Comparisons	137
9.8	Non-parametric Methods	138
9.9	Example Data	139
9.10	How to Proceed in R	139
9.11	Reporting Analyses	149
9.12	Recommended Reading	150
10	Data Transformations for Analysis of Variance	151
10.1	Assumptions of ANOVA and their Possible Violations	151
10.2	Log-transformation	153
10.3	Arcsine Transformation	156
10.4	Square-Root and Box–Cox Transformation	156
10.5	Concluding Remarks	157
10.6	Example Data	158
10.7	How to Proceed in R	158
10.8	Reporting Analyses	163
10.9	Recommended Reading	163
11	Hierarchical ANOVA, Split-Plot ANOVA, Repeated Measurements	164
11.1	Hierarchical ANOVA	164
11.2	Split-Plot ANOVA	167
11.3	ANOVA for Repeated Measurements	169

11.4	Example Data	171
11.5	How to Proceed in R	171
11.6	Reporting Analyses	181
11.7	Recommended Reading	182
12	Simple Linear Regression: Dependency Between Two Quantitative Variables	183
12.1	Use Case Examples	183
12.2	Regression and Correlation	184
12.3	Simple Linear Regression	184
12.4	Testing Hypotheses	187
12.5	Confidence and Prediction Intervals	190
12.6	Regression Diagnostics and Transforming Data in Regression	190
12.7	Regression Through the Origin	195
12.8	Predictor with Random Variation	197
12.9	Linear Calibration	197
12.10	Example Data	198
12.11	How to Proceed in R	198
12.12	Reporting Analyses	204
12.13	Recommended Reading	205
13	Correlation: Relationship Between Two Quantitative Variables	206
13.1	Use Case Examples	206
13.2	Correlation as a Dependency Statistic for Two Variables on an Equal Footing	206
13.3	Test Power	209
13.4	Non-parametric Methods	212
13.5	Interpreting Correlations	212
13.6	Statistical Dependency and Causality	213
13.7	Example Data	216
13.8	How to Proceed in R	216
13.9	Reporting Analyses	218
13.10	Recommended Reading	218
14	Multiple Regression and General Linear Models	219
14.1	Use Case Examples	219
14.2	Dependency of a Response Variable on Multiple Predictors	219
14.3	Partial Correlation	223
14.4	General Linear Models and Analysis of Covariance	224
14.5	Example Data	225
14.6	How to Proceed in R	226
14.7	Reporting Analyses	237
14.8	Recommended Reading	238

15	Generalised Linear Models	239
15.1	Use Case Examples	239
15.2	Properties of Generalised Linear Models	240
15.3	Analysis of Deviance	242
15.4	Overdispersion	243
15.5	Log-linear Models	243
15.6	Predictor Selection	244
15.7	Example Data	245
15.8	How to Proceed in R	246
15.9	Reporting Analyses	250
15.10	Recommended Reading	251
16	Regression Models for Non-linear Relationships	252
16.1	Use Case Examples	252
16.2	Introduction	253
16.3	Polynomial Regression	253
16.4	Non-linear Regression	255
16.5	Example Data	256
16.6	How to Proceed in R	256
16.7	Reporting Analyses	259
16.8	Recommended Reading	260
17	Structural Equation Models	261
17.1	Use Case Examples	261
17.2	SEMs and Path Analysis	261
17.3	Example Data	265
17.4	How to Proceed in R	265
17.5	Reporting Analyses	272
17.6	Recommended Reading	272
18	Discrete Distributions and Spatial Point Patterns	274
18.1	Use Case Examples	274
18.2	Poisson Distribution	274
18.3	Comparing the Variance with the Mean to Measure Spatial Distribution	276
18.4	Spatial Pattern Analyses Based on the K-function	279
18.5	Binomial Distribution	280
18.6	Example Data	283
18.7	How to Proceed in R	283
18.8	Reporting Analyses	289
18.9	Recommended Reading	289
19	Survival Analysis	290
19.1	Use Case Examples	290
19.2	Survival Function and Hazard Rate	291

19.3	Differences in Survival Among Groups	293
19.4	Cox Proportional Hazard Model	293
19.5	Example Data	295
19.6	How to Proceed in R	295
19.7	Reporting Analyses	302
19.8	Recommended Reading	302
20	Classification and Regression Trees	
20.1	Use Case Examples	303
20.2	Introducing CART	304
20.3	Pruning the Tree and Crossvalidation	306
20.4	Competing and Surrogate Predictors	307
20.5	Example Data	308
20.6	How to Proceed in R	309
20.7	Reporting Analyses	316
20.8	Recommended Reading	316
21	Classification	
21.1	Use Case Examples	317
21.2	Aims and Properties of Classification	317
21.3	Input Data	319
21.4	Similarity and Distance	319
21.5	Clustering Algorithms	320
21.6	Displaying Results	320
21.7	Divisive Methods	321
21.8	Example Data	322
21.9	How to Proceed in R	322
21.10	Other Software	324
21.11	Reporting Analyses	325
21.12	Recommended Reading	325
22	Ordination	
22.1	Use Case Examples	326
22.2	Unconstrained Ordination Methods	327
22.3	Constrained Ordination Methods	330
22.4	Discriminant Analysis	331
22.5	Example Data	333
22.6	How to Proceed in R	333
22.7	Alternative Software	340
22.8	Reporting Analyses	341
22.9	Recommended Reading	341
Appendix A:	First Steps with R Software	
A.1	Starting and Ending R, Command Line, Organising Data	343
A.2	Managing Your Data	349

A.3	Data Types in R	351
A.4	Importing Data into R	357
A.5	Simple Graphics	359
A.6	Frameworks for R	360
A.7	Other Introductions to Work with R	362
	<i>Index</i>	363

Modern biology is a quantitative science. A biologist weighs, measures and counts, whether she works with aphid or fish individuals, with plant communities or with nuclear DNA. Every number obtained in this way, however, is affected by random variation. Aphid counts repeatedly obtained from the same plant individual will differ. The counts of aphids obtained from different plants will differ more, even if those plants belong to the same species, and samples coming from plants of different species are likely to differ even more. Similar differences will be found in the nuclear DNA content of plants from the same population, in nitrogen content of soil samples taken from the same or different sites, or in the population densities of copepods across repeated samplings from the same lake. We say that our data contain a random component: the values we obtain are random quantities, with a part of their variation resulting from randomness.

But what actually is this randomness? In posing such a question, we move into the realm of philosophy or to axioms of probability theory. But what is probability? A biologist is usually happy with a pragmatic concept: we consider an event to be random if we do not have a causal explanation for it. Statistics is a research field which provides recipes for how to work with data containing random components, and how to distinguish deterministic patterns from random variation. Popular wisdom says that statistics is a branch of science where precise work is carried out with imprecise numbers. But the term statistics has multiple meanings. The layman sees it as an assorted collection of values (football league statistics of goals and points, statistics of MP voting, statistics of cars passing along a highway, etc.). Statistics is also a research field (often called mathematical statistics) providing tools for obtaining useful information from such datasets. It is