

Contents

Preface xiii

- The third edition xiv
- How to use this book xiv
 - Packages used xv
 - Example data xv
- Acknowledgements for the first edition xv
- Acknowledgements for the second edition xv
- Acknowledgements for the third edition xvi

1 Eight steps to successful data analysis 1

2 The basics 2

- Observations 2
- Hypothesis testing 2
- P*-values 3
- Sampling 3
- Experiments 4
- Statistics 4
 - Descriptive statistics 5
 - Tests of difference 5
 - Tests of relationships 5
 - Tests for data investigation 6

3 Choosing a test: a key 7

- Remember: eight steps to successful data analysis 7
- The art of choosing a test 7
- A key to assist in your choice of statistical test 8

4 Hypothesis testing, sampling and experimental design 23

- Hypothesis testing 23
- Acceptable errors 23
- P*-values 24
- Sampling 25

- Choice of sample unit 25
- Number of sample units 26
- Positioning of sample units to achieve a random sample 26
- Timing of sampling 27
- Experimental design 27
 - Control 28
 - Procedural controls 28
 - Temporal control 28
 - Experimental control 29
 - Statistical control 29
 - Some standard experimental designs 29

5 Statistics, variables and distributions 32

- What are statistics? 32
- Types of statistics 33
 - Descriptive statistics 33
 - Parametric statistics 33
 - Non-parametric statistics 33
- What is a variable? 33
- Types of variables or scales of measurement 34
 - Measurement variables 34
 - Continuous variables 34
 - Discrete variables 35
 - How accurate do I need to be? 35
 - Ranked variables 35
 - Attributes 35
 - Derived variables 36
- Types of distribution 36
- Discrete distributions 36
 - The Poisson distribution 36
 - The binomial distribution 37
 - The negative binomial distribution 39
 - The hypergeometric distribution 39
- Continuous distributions 40
 - The rectangular distribution 40
 - The normal distribution 40
 - The standardized normal distribution 40
 - Convergence of a Poisson distribution to a normal distribution 41
 - Sampling distributions and the 'central limit theorem' 41
- Describing the normal distribution further 41
 - Skewness 41
 - Kurtosis 43
- Is a distribution normal? 43
- Transformations 43

- An example 44
- The angular transformation 44
- The logit transformation 45
- The t -distribution 46
- Confidence intervals 47
- The chi-square (χ^2) distribution 47
- The exponential distribution 47
- Non-parametric 'distributions' 48
 - Ranking, quartiles and the interquartile range 48
 - Box and whisker plots 48

6 Descriptive and presentational techniques 49

- General advice 49
- Displaying data: summarizing a single variable 49
 - Box and whisker plot (box plot) 49
- Displaying data: showing the distribution of a single variable 50
 - Bar chart: for discrete data 50
 - Histogram: for continuous data 51
 - Pie chart: for categorical data or attribute data 52
- Descriptive statistics 52
 - Statistics of location or position 52
 - Arithmetic mean 53
 - Geometric mean 53
 - Harmonic mean 53
 - Median 53
 - Mode 53
 - Statistics of distribution, dispersion or spread 55
 - Range 55
 - Interquartile range 55
 - Variance 55
 - Standard deviation (SD) 55
 - Standard error (SE) 56
 - Confidence intervals (CI) or confidence limits 56
 - Coefficient of variation 56
 - Other summary statistics 56
 - Skewness 57
 - Kurtosis 57
- Using the computer packages 57
 - General 57
- Displaying data: summarizing two or more variables 62
 - Box and whisker plots (box plots) 62
 - Error bars and confidence intervals 63
- Displaying data: comparing two variables 63
 - Associations 63

- Scatterplots 64
- Multiple scatterplots 64
- Trends, predictions and time series 65
 - Lines 65
 - Fitted lines 67
 - Confidence intervals 67
- Displaying data: comparing more than two variables 68
 - Associations 68
 - Three-dimensional scatterplots 68
 - Multiple trends, time series and predictions 69
 - Multiple fitted lines 69
 - Surfaces 70

7 The tests 1: tests to look at differences 72

- Do frequency distributions differ? 72
 - Questions 72
 - G-test 72
 - An example 73
 - Chi-square test (χ^2) 75
 - An example 76
 - Kolmogorov–Smirnov test 86
 - An example 87
 - Anderson–Darling test 89
 - Shapiro–Wilk test 90
 - Graphical tests for normality 90
- Do the observations from two groups differ? 92
 - Paired data 92
 - Paired *t*-test 92
 - Wilcoxon signed ranks test 96
 - Sign test 99
 - Unpaired data 103
 - t*-test 103
 - One-way ANOVA 111
 - Mann–Whitney U 119
- Do the observations from more than two groups differ? 123
 - Repeated measures 123
 - Friedman test (for repeated measures) 123
 - Repeated-measures ANOVA 127
 - Independent samples 128
 - One-way ANOVA 129
 - Post hoc* testing: after one-way ANOVA 138
 - Kruskal–Wallis test 142
 - Post hoc* testing: after the Kruskal–Wallis test 145
- There are two independent ways of classifying the data 145

- One observation for each factor combination (no replication) 146
 - Friedman test 146
 - Two-way ANOVA (without replication) 152
- More than one observation for each factor combination (with replication) 160
 - Interaction 160
 - Two-way ANOVA (with replication) 163
 - An example 164
 - Scheirer-Ray-Hare test 175
 - An example 175
- There are more than two independent ways to classify the data 182
 - Multifactorial testing 182
 - Three-way ANOVA (without replication) 183
 - Three-way ANOVA (with replication) 184
 - An example 184
 - Multiway ANOVA 191
- Not all classifications are independent 192
 - Non-independent factors 192
 - Nested factors 192
 - Random or fixed factors 193
- Nested or hierarchical designs 193
 - Two-level nested-design ANOVA 193
 - An example 193

8 The tests 2: tests to look at relationships 199

- Is there a correlation or association between two variables? 199
 - Observations assigned to categories 199
 - Chi-square test of association 199
 - An example 200
 - Cramér coefficient of association 208
 - Phi coefficient of association 209
 - Observations assigned a value 209
 - 'Standard' correlation (Pearson's product-moment correlation) 209
 - An example 210
 - Spearman's rank-order correlation 214
 - An example 215
 - Kendall rank-order correlation 218
 - An example 218
 - Regression 219
 - An example 220
- Is there a cause-and-effect relationship between two variables? 220
 - Questions 220
 - 'Standard' linear regression 221
 - Prediction 221

- Interpreting r^2 222
- Comparison of regression and correlation 222
- Residuals 222
- Confidence intervals 222
- Prediction interval 223
- An example 223
- Kendall robust line-fit method 230
- Logistic regression 230
 - An example 231
- Model II regression 235
- Polynomial, cubic and quadratic regression 235
- Tests for more than two variables 236
 - Tests of association 236
 - Questions 236
 - Correlation 236
 - Partial correlation 237
 - Kendall partial rank-order correlation 237
- Cause(s) and effect(s) 237
 - Questions 237
 - Regression 237
 - Analysis of covariance (ANCOVA) 238
 - Multiple regression 242
 - Stepwise regression 242
 - Path analysis 243

9 The tests 3: tests for data exploration 244

- Types of data 244
- Observation, inspection and plotting 244
 - Principal component analysis (PCA) and factor analysis 244
 - An example 245
 - Canonical variate analysis 251
 - Discriminant function analysis 251
 - An example 251
 - Multivariate analysis of variance (MANOVA) 256
 - An example 256
 - Multivariate analysis of covariance (MANCOVA) 259
 - Cluster analysis 259
 - DECORANA and TWINSpan 263

Symbols and letters used in statistics 264

- Greek letters 264
- Symbols 264
- Upper-case letters 265
- Lower-case letters 266

Glossary 267**Assumptions of the tests 282**

What if the assumptions are violated? 284

Hints and tips 285

Using a computer 285

Sampling 286

Statistics 286

Displaying the data 287

A table of statistical tests 289**Index 291**