

The background of the book cover is a photograph of a field of flowers. The top half is dominated by a dense field of red poppies. The bottom half features a mix of red poppies and blue cornflowers. The overall scene is a lush, natural landscape.

CALVIN DYTHAM

Choosing and Using Statistics

A Biologist's Guide

Second Edition



Blackwell
Publishing

Contents

Preface, x

- The second edition, xi
- How to use this book, xii
 - Packages used, xii
 - Example data, xii
- Acknowledgements for the first edition, xii
- Acknowledgements for the second edition, xii

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, 4
 - 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, 22

- Hypothesis testing, 22
- Acceptable errors, 22
- P*-values, 23
- Sampling, 24
 - Choice of sample unit, 24
 - Number of sample units, 25

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

5 Statistics, variables and distributions, 30

- What are statistics? 30
- Types of statistics, 31
 - Descriptive statistics, 31
 - Parametric statistics, 31
 - Non-parametric statistics, 31
- What is a variable?, 31
- Types of variables or scales of measurement, 32
 - Measurement variables, 32
 - Ranked variables, 33
- Types of distributions, 34
 - Why do you need to know about distributions? 34
 - Discrete distributions, 34
 - Poisson distribution, 34
 - Binomial distribution, 35
 - Negative binomial distribution, 37
 - Hypergeometric distribution, 37
 - Continuous distributions, 38
 - Rectangular distribution, 38
 - Normal distribution, 38
 - Describing normal distribution further, 39
 - Transformations, 41
 - t*-distribution, 43
 - Confidence intervals, 43
 - Chi-square distribution, 44
 - Exponential distribution, 44
 - Non-parametric 'distributions', 44
 - Ranking, quartiles and the interquartile range, 45
 - Box and whisker plots, 45

6 Descriptive and presentational techniques, 46

- Displaying data: summarizing a single variable, 46
- Displaying data: showing the distribution of a single variable, 46

- Descriptive statistics, 49
 - Statistics of location, 49
 - Statistics of distribution, dispersion or spread, 50
 - Other summary statistics, 53
- Using the computer packages, 53
- Displaying data: summarizing two or more variables, 57
- Displaying data: comparing two variables, 58
 - Associations, 58
 - Trends, predictions and time series, 59
- Displaying data: comparing more than two variables, 63
 - Associations, 63
 - Multiple trends, time series and predictions, 64

7 The tests 1: tests to look at differences, 66

- Do frequency distributions differ? 66
 - G-test, 66
 - Chi-square test (χ^2), 68
 - Kolmogorov–Smirnov test, 77
 - Anderson–Darling test, 79
- Do the observations from two groups differ? 80
 - Paired data, 80
 - Paired *t*-test, 80
 - Wilcoxon signed ranks test, 83
 - Sign test, 85
 - Unpaired data, 89
 - t*-test, 89
 - One-way ANOVA, 94
 - Mann–Whitney *U* test, 101
- Do the observations from more than two groups differ? 104
 - Repeated measures, 104
 - Friedman test (for repeated measures), 104
 - Repeated measures ANOVA, 107
 - Independent samples, 108
 - One-way ANOVA, 109
 - Post hoc* testing: after one-way ANOVA, 116
 - Kruskal–Wallis test, 119
 - Post hoc* testing: after the Kruskal–Wallis test, 121
- There are two independent ways of classifying the data, 122
 - One observation for each factor combination (no replication), 122
 - Friedman test, 122
 - Two-way ANOVA (without replication), 126
 - More than one observation for each factor combination (with replication), 133

- Interaction, 133
 - Two-way ANOVA (with replication), 137
 - Scheirer–Ray–Hare test, 145
- There are more than two independent ways to classify the data, 151
 - Multifactorial testing, 151
 - Three-way ANOVA (without replication), 151
 - Three-way ANOVA (with replication), 152
 - Multiway ANOVA, 157
- Not all classifications are independent, 158
 - Nested or hierarchical designs, 159
 - Two-level nested design ANOVA, 160

8 The tests 2: tests to look at relationships, 165

- Is there a correlation or association between two variables? 165
 - Observations assigned to categories, 165
 - Chi-square test of association, 165
 - Cramér coefficient of association, 172
 - Phi coefficient of association, 172
 - Observations assigned a value, 173
 - ‘Standard’ correlation (Pearson product–moment correlation), 173
 - Spearman rank-order correlation, 177
 - Kendall rank-order correlation, 179
 - Regression, 181
- Is there a ‘cause and effect’ relationship between two variables? 181
 - ‘Standard’ linear regression, 182
 - Kendall robust line-fit method, 189
 - Logistic regression, 189
 - Model II regression, 192
 - Polynomial, cubic and quadratic regression, 192
- Tests for more than two variables, 193
 - Tests of association, 193
 - Correlation, 193
 - Partial correlation, 194
 - Kendall partial rank-order correlation, 194
 - Cause(s) and effect(s), 194
 - Regression, 194
 - Analysis of covariance (ANCOVA), 195
 - Multiple regression, 198
 - Stepwise regression, 198
 - Path analysis, 199

9 The tests 3: tests for data exploration, 200

- Observation, inspection and plotting, 200
- Principal component analysis (PCA) and factor analysis, 200

- Canonical variate analysis (CVA), 205
- Discriminant function analysis, 205
- Multivariate analysis of variance (MANOVA), 210
- Multivariate analysis of covariance (MANCOVA), 212
- Cluster analysis, 212
- DECORANA and TWINSpan, 214

10 Symbols and letters used in statistics, 216

- Greek letters, 216
- Symbols, 216
- Upper case letters, 217
- Lower case letters, 218

11 Glossary, 219

12 Assumptions of the tests, 234

- What if the assumptions are violated?, 236

13 Hints and tips, 237

- Using a computer, 237
- Sampling, 238
- Statistics, 238
- Displaying the data, 239

14 A table of statistical tests, 241

Index, 243