Contents

Preface

1 Introducing nonparametric methods
   1.1 Basic statistics
   1.2 Samples and populations
   1.3 Hypothesis tests
   1.4 Estimation
   1.5 Ethical issues
   1.6 Computers and nonparametric methods
   1.7 Further reading
   Exercises

2 Centrality inference for single samples
   2.1 Using measurement data
   2.2 Inferences about medians based on ranks
   2.3 The sign test
   2.4 Transformation of ranks
   2.5 Asymptotic results
   2.6 Robustness
   2.7 Fields of application
   2.8 Summary
   Exercises

3 Other single-sample inference
   3.1 Inferences for dichotomous data
   3.2 Tests related to the sign test
   3.3 Matching samples to distributions
   3.4 Angular data
   3.5 A runs test for randomness
   3.6 Fields of application
   3.7 Summary
   Exercises

4 Methods for paired samples
   4.1 Comparisons in pairs
   4.2 A less obvious use of the sign test
   4.3 Power and sample size
   4.4 Fields of application
   4.5 Summary
   Exercises

©2001 CRC Press LLC
5 Methods for two independent samples
5.1 Centrality tests and estimates
5.2 Rank based tests
5.3 The median test
5.4 Normal scores
5.5 Tests for survival data
5.6 Asymptotic approximations
5.7 Power and sample size
5.8 Tests for equality of variance
5.9 Tests for a common distribution
5.10 Fields of application
5.11 Summary
      Exercises

6 Three or more samples
6.1 Comparisons with parametric methods
6.2 Centrality tests for independent samples
6.3 Centrality tests for related samples
6.4 More detailed treatment comparisons
6.5 Tests for heterogeneity of variance
6.6 Some miscellaneous considerations
6.7 Fields of application
6.8 Summary
      Exercises

7 Correlation and concordance
7.1 Correlation and bivariate data
7.2 Ranked data for several variables
7.3 Agreement
7.4 Fields of application
7.5 Summary
      Exercises

8 Regression
8.1 Bivariate linear regression
8.2 Multiple regression
8.3 Nonparametric regression models
8.4 Other multivariate data problems
8.5 Fields of application
8.6 Summary
      Exercises

9 Categorical data
9.1 Categories and counts
9.2 Nominal attribute categories

©2001 CRC Press LLC
9.3 Ordered categorical data
9.4 Goodness-of-fit tests for discrete data
9.5 Extension of McNemar's test
9.6 Fields of application
9.7 Summary
  Exercises

10 Association in categorical data
10.1 The analysis of association
10.2 Some models for contingency tables
10.3 Combining and partitioning of tables
10.4 Power
10.5 Fields of application
10.6 Summary
  Exercises

11 Robust estimation
11.1 When assumptions break down
11.2 Outliers and influence
11.3 The bootstrap
11.4 M-estimators and other robust estimators
11.5 Fields of application
11.6 Summary
  Exercises

Appendix

References

Solutions to odd-numbered exercises
Preface

The second edition of this book was written by the first-named author to provide a then (1993) up-to-date introduction to nonparametric and distribution-free methods. It took a midway course between a bare description of techniques and a detailed exposition of the theory. Individual methods and links between them were illustrated mainly by examples, Mathematics was kept to the minimum needed for a clear understanding of scope and limitations. The book was designed to meet the needs both of statistics students making first contact with these methods and of research workers, managers, research and development staff, consultants and others working in various fields who had an understanding of basic statistics and who, although they had little previous knowledge of nonparametric methods, now found or thought they might find them useful in their work.

A positive response from readers and reviewers has encouraged us to retain the basic format while taking the opportunity to introduce new topics as well as changing the emphasis to reflect both developments in computing and new attitudes towards data analysis.

Nonparametric methods are basically analytic tools, but data collection, analyses and their interpretation are interrelated. This is why we have expanded the coverage of topics such as ethical considerations and calculation of power and of sample sizes needed to achieve stated aims. These make their main impact at the planning stage, but also influence the analytic and inferential phases.

There has been widespread criticism in recent years by many statisticians of inappropriate and even improper use of significance tests and the related concept of P-values. However, these tools have a positive role when properly used and understood. To encourage better use the section on hypothesis testing in Chapter I has been rewritten, and throughout the book there is more emphasis on how these concepts should be used and warnings about potential misuse.

The layout of Chapters I to 10 follows the broad pattern of the corresponding chapters in the second edition but there are many changes in order and other aspects of presentation including new and more detailed examples. One or two topics have been dropped or are treated in less detail, and new material has been inserted where appropriate. As well as comments on ethical considerations and discussions on power and sample size, there are new sections on the
analysis of angular data, the use of capture-recapture methods, the measurement of agreement between observers and several lesser additions. Examples have been chosen from a wider range of disciplines. For a few more advanced topics such as regression smoothing techniques and M-estimation we have not given details of specific methods but only a broad overview of each topic to enable readers to judge whether it may be relevant to their particular needs. In such cases references are given to sources that contain the detail needed for implementation.

Chapter 11 has been rewritten to give an elementary introduction to influence functions, the nonparametric bootstrap and robust estimation generally, again with references to source material for those who want to make full use of these ideas. Material that appeared in Chapter 12 of the second edition has been updated and incorporated at relevant points in the text.

We have not included tables for basic nonparametric procedures, mainly because more satisfactory information is provided by modem statistical software, making many standard tables insufficient or superfluous for serious users of the methods. Those who need such tables because they have no access to specialized software are well catered for by standard collections of statistical tables. We give references to these throughout the book and also when relevant to some specialized tables. We have retained the section outlining solutions to odd-numbered exercises.

We are grateful to many readers of the earlier editions who made constructive comments about the content and treatment, or sometimes about the lack of treatment, of particular topics. This input triggered many of the changes made in this edition. Our special thanks go to Jim McGanrick for helpful discussions on physiological measurements and to Professor Richard Hughes for advice on the Guillain-Barré syndrome. We happily renew the thanks recorded in the second edition to Timothy P. Davis and Chris Theobald who supplied us with data sets used initially in that edition for examples that we have retained.

P. Sprent
N. C. Smeeton

July 2000