
PREFACE TO NEW PRINTING (1998)

Rank tests form a class of statistical procedures that have the advantage of great simplicity combined with surprising power. Since their development in the 1940s and 1950s, they have taken their place as strong competitors of the more classical normal theory methods. Rank tests apply only to relatively simple solutions, such as one-, two-, and s -sample problems, and testing for independence and randomness, but for these situations they are often the method of choice.

The current edition of this textbook describes these tests and the estimating procedures derived from them, and gives an account of their properties. The body of work on rank tests was in place when this book was first published in 1975, and it has not changed much since then or lost its usefulness. This book is therefore being reprinted in its original form, with the only additions being this new Preface and an Epilogue that describes some related recent developments.

Even though the field of rank tests has undergone little change, important new methodologies have sprung up which also serve the purpose of freeing statistics from the unrealistic model assumptions that so frequently invalidate its applications. The Epilogue outlines some of these alternative approaches and their connections with the material in this book.

One aspect of statistics that has changed radically during the last 20 years is the computing environment. All the tests discussed here are now available in a variety of statistical packages. Despite the easy availability of such packages, I have retained the tables at the end of the book for students' convenience in working the examples and exercises.

PREFACE

1. History

Methods based on ranks form a substantial body of statistical techniques that provide alternatives to the classical parametric methods. Individual rank tests were proposed much earlier [the earliest use may be that of the sign test by Arbuthnot in 1710; for some additional history see, for example, Kruskal (1957 and 1958)*]; the modern development of the subject may be said to have begun with the papers by Hotelling and Pabst (1936), Friedman (1937), Kendall (1938), Smirnov (1939), and those of Wald and Wolfowitz in the early 1940s. An interesting survey of this work was given by Scheffé (1943).

A full-scale development of rank-based methods seems to have been sparked by the publication in 1945 of a paper by Wilcoxon in which he discussed the two tests, now bearing his name, for comparing two treatments, and by the book of Kendall (1948). Since then there has been a flood of publications that has not yet abated. A bibliography of nonparametric statistics (of which rank-based methods constitute the methodologically most important part) by Savage (1962) lists about 3,000 items. If brought up to date, it probably would contain twice that many entries.

* See *References following Preface*.

The feature of nonparametric methods mainly responsible for their great popularity (and to which they owe their name) is the weak set of assumptions required for their validity. Although it was believed at first that a heavy price in loss of efficiency would have to be paid for this robustness, it turned out, rather surprisingly, that the efficiency of the Wilcoxon tests and other nonparametric procedures holds up quite well under the classical assumption of normality and that these procedures may have considerable advantages in efficiency (as well as validity) when the assumption of normality is not satisfied. These facts were first brought out clearly by Pitman (1948) and were strengthened by results of Hodges and Lehmann (1956) and Chernoff and Savage (1958).

In the early stages, rank-based methods were essentially restricted to testing procedures. They thus did not provide a flexible array of methods, which would include not only tests but point and interval estimates as well as various simultaneous inference procedures. This difficulty is gradually being overcome, although rank-based methods do not yet have the flexibility and the wide applicability to complex linear models that make least squares and normal theory so attractive.

2. The present book

The purpose of this book is to provide an introduction to nonparametric methods for the analysis and planning of comparative studies. Only a relatively small number of basic techniques are presented in detail: These are mainly tests of the Wilcoxon type (which can be obtained from the corresponding classical tests by replacing the observations by their ranks) and the estimation and simultaneous inference procedures based upon these tests. These methods are simple, have good efficiency properties, and most are well tabulated. They are treated rather fully here, with emphasis on the assumptions under which they are appropriate, the accuracy of the various approximations that are required, and the modifications needed for tied observations. For the simplest cases there is also a discussion of power or accuracy and the determination of the sample sizes required to achieve a given accuracy. The use of the methods is exemplified in the text, and numerous problems furnish opportunities for the student to try them for himself. In many cases the data for these illustrations are the results of actual studies reported in the literature.

An indication of some alternatives to and extensions of the above procedures, and of some additional properties, is provided by sections of further developments at the end of each chapter, which give an introduction to the literature on these subjects. Among the topics treated this way are the Normal Scores procedures, permutation tests, sequential methods, and optimum theory. Two topics that are not covered are multivariate techniques (because of lack of space) and goodness-of-fit tests (because both the problem and the data are quite different from those considered here). On the other hand, a discussion of some tests for two-way

contingency tables is included in the text because they can be viewed as special cases of rank tests with tied observations.

As mentioned above, an important advantage of nonparametric tests is the simplicity of the assumptions required for their validity. It is not necessary to postulate a population from which the subjects in a study have been obtained by random sampling, but only that the treatments being compared have been assigned to the subjects at random. All techniques are first discussed in terms of such a *randomization model*. This material (which constitutes Chapters 1 and 3 and some parts of Chapters 5 to 7) requires only the simplest mathematical background for its understanding. All that is needed is an elementary introduction to probability, such as that provided by mathematics courses in many high schools or the first lectures in an introductory course on probability or statistics.

It is possible within this framework to describe the tests, illustrate their use, and discuss the computation of significance or critical values. However, randomization models do not permit an evaluation of power that could be used to plan the size of a study. This is best discussed in terms of a population from which the subjects have been sampled. Unfortunately, an adequate treatment of *population models* (such as those underlying the normal distribution) requires some knowledge of the calculus.

The determination of sample size and the evaluation of the power of a test (which plays an analogous role as the variance of an estimate) seemed too important to omit. So as to include these somewhat more advanced topics, I have allowed the level of the book to vary with the requirements of the material. Despite the obvious disadvantages of such inconsistency, it is my hope that even the reader with little mathematical background will be able to follow the main ideas of the more advanced parts (Chapters 2 and 4 and portions of Chapters 5 to 7) and that the reader whose background is stronger will not be put off by the slow pace of the more elementary sections. I am encouraged in this hope by the fact that courses I have taught along these lines to students with very disparate backgrounds seem to have been reasonably successful.

The main text is followed by an appendix that provides the large-sample theory underlying the many approximations required where tables are not available and exact computations are too laborious. This material is at an intermediate level. It requires substantially more mathematical sophistication than the rest of the book, but it is much less advanced than the books by Hájek and Sidak (1967) and Puri and Sen (1971). A number of standard limit theorems from probability theory are stated and discussed but not proved, and on this basis the needed results are derived with relatively little effort.

3. Acknowledgments

There remains the pleasant task of acknowledging the many debts that I incurred during the writing of the book. I should like to thank Peter Bickel, Kjell Doksum, Gus Haggstrom, Joe Hodges, Bill Kruskal, Vida Lehmann, and Juliet Popper Shaffer who at various stages read portions of the manuscript and gave me their moral support and the benefit of their advice. Of the reviewers who examined the manuscript for the publishers, I am especially grateful to Ralph D'Agostino and Alfred Forsyth for their many valuable criticisms and ideas for improvements, and to Gottfried Noether for his crucial support and suggestions.

To two friends I owe a special debt. Ellen Sherman read critically, checked the computations, and prepared the tables for an early version of the first chapters. Above all, Howard D'Abrera performed the same task for later versions of the whole manuscript and prepared the answers for the problems. He corrected innumerable errors of style, thought, and arithmetic. Without his help I could not have brought this project to its completion.

The aspect of the book that caused me the most difficulty was to find suitable live examples and problems. Authors typically do not publish their data, and when a set of published data is potentially suitable, it usually turns out that the sample size is too large or small, there are too many or too few ties, the results are too obviously significant or too obviously not, or that the design or sampling procedure is not what is required to illustrate the particular point in question. I am grateful to the colleagues who put their unpublished data at my disposal as well as those who published data that I was able to use. To all I would like to extend an apology. When some minor modification of the data made them more suitable for my purpose, I have carried out such modifications (always with an acknowledgment). More seriously, I have used the data to illustrate the point I was trying to make, though this may have borne little relation to the purpose for which they were collected, and I have asked questions of the data that may seem foolish to someone more familiar with the actual situation. I hope that I will be forgiven for these violations of the authors' intentions, and I should like to ask readers who have available more suitable data to illustrate the techniques discussed here to let me know about them for a possible later revision.

Finally, it is a pleasure to thank the Office of Naval Research for their generous support of the research that has gone into this book.

E. L. Lehmann

Berkeley
1974

PREFACE REFERENCES

- Arbuthnot, J. (1710): "An Argument for Divine Providence, Taken from the Constant Regularity Observed in the Births of Both Sexes," *Philos. Trans.* **27**:186–190.
- Chernoff, Herman and Savage, I. Richard (1958): "Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics," *Ann. Math. Statist.* **29**:972–994.
- Friedman, Milton (1937): "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *J. Amer. Statist. Assoc.* **32**:675–701.
- Hájek, J. and Sidak, Z. (1967): *Theory of Rank Tests*, Academic Press.
- Hodges, J. L., Jr. and Lehmann, E. L. (1956): "The Efficiency of Some Nonparametric Competitors of the t -Test," *Ann. Math. Statist.* **27**:324–335.
- Hotelling, Harold and Pabst, Margaret Richards (1936): "Rank Correlation and Tests of Significance Involving No Assumptions of Normality," *Ann. Math. Statist.* **7**:29–43.
- Kendall, Maurice G. (1938): "A New Measure of Rank Correlation," *Biometrika* **30**:81–93.
- (1948): *Rank Correlation Methods*, 4th edition (1970), Griffin, London.
- Kruskal, William H. (1957): "Historical Notes on the Wilcoxon Unpaired Two-sample Test," *J. Amer. Statist. Assoc.* **52**:356–360.
- (1958): "Ordinal Measures of Association," *J. Amer. Statist. Assoc.* **53**:814–861.
- Pitman, E. J. G. (1948): *Lecture Notes on Nonparametric Statistics*, Columbia Univ., New York.
- Puri, M. L. and Sen, P. K. (1971): *Nonparametric Methods in Multivariate Analysis*, John Wiley.
- Savage, I. Richard (1962): *Bibliography of Nonparametric Statistics*, Harvard Univ. Press.
- Scheffé, Henry (1943): "Statistical Inference in the Nonparametric Case," *Ann. Math. Statist.* **14**:305–332.
- Smirnov, N. V. (1939): "On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples," *Bull. Math. Univ. Moscow* **2**, No. 2, 3–14.
- Wilcoxon, Frank (1945): "Individual Comparisons by Ranking Methods," *Biometrics* **1**:80–83.