Wiley Series in Probability and Statistics

THIRD EDITION

Nonparametric Statistical Methods

MYLES HOLLANDER
DOUGLAS A. WOLFE
ERIC CHICKEN

WILEY

Nonparametric
Statistical Methods

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Garrett M. Fitzmaurice, Harvey Goldstein, Iain M. Johnstone, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Ruey S. Tsay, Sanford Weisberg*
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, Joseph B. Kadane, Jozef L. Teugels*

A complete list of the titles in this series appears at the end of this volume.

Nonparametric Statistical Methods

Third Edition

Myles Hollander

Department of Statistics
Florida State University
Tallahassee, Florida

Douglas A. Wolfe

Department of Statistics
Ohio State University
Columbus, Ohio

Eric Chicken

Department of Statistics
Florida State University
Tallahassee, Florida

WILEY

Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Hollander, Myles.
Nonparametric statistical methods. –Third edition / Myles Hollander, Department of Statistics, Florida State University, Tallahassee, Florida, Douglas A. Wolfe, Department of Statistics, Ohio State University, Columbus, Ohio, Eric Chicken, Department of Statistics, Florida State University, Tallahassee, Florida.
pages cm
Includes bibliographical references and indexes.
ISBN 978-0-470-38737-5 (hardback) -
1. Nonparametric statistics. I. Wolfe, Douglas A. II. Chicken, Eric, 1963– III. Title.
QA278.8.H65 2013
519.5–dc23

2013007061

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To our wives,
Glee, Marilyn, and Rebecca.

Contents

Preface

The nonparametric approach is the preferred methodology for statisticians and other scientists. We list some of its advantages in Section 1.1. In the third edition, we retain our emphasis on applications to real-world situations. We want our readers to learn how to apply nonparametric techniques in a variety of settings and to understand the assumptions underlying the methods.

In this third edition, we have improved the 11 chapters of the second edition and added five new chapters. The new chapters cover topics of recent and current interest, namely, density estimation, wavelets, smoothing, ranked set sampling, and Bayesian nonparametrics. `R` programs are now used to perform calculations. See Section 1.5 for a description of `R`.

The second edition was used primarily for a one-semester senior undergraduate/first-year graduate course for students having had a prior course in statistics. With the added coverage here, there is ample material for a two-semester course. Nevertheless, we expect most teachers will still opt for a one-semester course and choose specific chapters in accordance with their interests and those of their students.

Many friends and colleagues have helped us with this project.

Grant Schneider, a graduate student at the Ohio State University, provided invaluable support in the conversion from complete reliance on null distribution tables in our second edition to the exclusive use of `R` programs to obtain appropriate critical values and compute associated *P*-values in this third edition. He wrote new `R` programs for all of the statistical procedures in Chapter 15 and for a majority of the many procedures in Chapters 5–7, and modified existing programs for the other procedures in those three chapters, leading to significantly improved computational speed in most cases. He also organized all of the `R` programs used in this third edition into a documented collection that is formally registered as an `R` package specifically linked to this third edition. We owe Grant a special thanks for his leadership role in this important aspect of our new edition.

Rachel Becvarik wrote new `R` programs for Chapters 11 and 16 and provided a spark.

Jelani Wiltshire and Michael Rosenthal assisted with LaTeX typesetting.

James Stricherz provided computing support and Pamela McGhee and Marylou Tatis provided office support.

Our editors Steve Quigley, Susanne Steitz-Filler, and Sari Friedman were dedicated from the inception to the completion. Our production manager Melissa Yanuzzi carefully guided the manuscript through the production process.

To all these helpmates, we are very grateful.

Tallahassee, Florida
Columbus, Ohio
Tallahassee, Florida
August 2013

MYLES HOLLANDER
DOUGLAS A. WOLFE
ERIC CHICKEN

Chapter 1

Introduction

1.1 ADVANTAGES OF NONPARAMETRIC METHODS

Roughly speaking, a nonparametric procedure is a statistical procedure that has certain desirable properties that hold under relatively mild assumptions regarding the underlying populations from which the data are obtained. The rapid and continuous development of nonparametric statistical procedures over the past $7\frac{1}{2}$ decades is due to the following advantages enjoyed by nonparametric techniques:

1. Nonparametric methods require few assumptions about the underlying populations from which the data are obtained. In particular, nonparametric procedures forgo the traditional assumption that the underlying populations are normal.
2. Nonparametric procedures enable the user to obtain exact *P*-values for tests, exact coverage probabilities for confidence intervals, exact experimentwise error rates for multiple comparison procedures, and exact coverage probabilities for confidence bands without relying on assumptions that the underlying populations are normal.
3. Nonparametric techniques are often (although not always) easier to apply than their normal theory counterparts.
4. Nonparametric procedures are often quite easy to understand.
5. Although at first glance most nonparametric procedures seem to sacrifice too much of the basic information in the samples, theoretical efficiency investigations have shown that this is not the case. Usually, the nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal (the home court of normal theory methods), and they can be mildly or wildly more efficient than these competitors when the underlying populations are not normal.
6. Nonparametric methods are relatively insensitive to outlying observations.
7. Nonparametric procedures are applicable in many situations where normal theory procedures cannot be utilized. Many nonparametric procedures require just the ranks of the observations, rather than the actual magnitude of the observations, whereas the parametric procedures require the magnitudes.
8. The Quenouille–Tukey jackknife (Quenouille (1949), Tukey (1958, 1962)) and Efron's computer-intensive (1979) bootstrap enable nonparametric approaches to be used in many complicated situations where the distribution theory

Nonparametric Statistical Methods, Third Edition. Myles Hollander, Douglas A. Wolfe, Eric Chicken.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.

needed to support parametric methods is intractable. See Efron and Tibshirani (1994).

9. Ferguson's Dirichlet process (1973) paved the way to combine the advantages of nonparametric methods and the use of prior information to form a Bayesian nonparametric approach that does not require distributional assumptions.
10. The development of computer software has facilitated fast computation of exact and approximate P-values for conditional nonparametric tests.

1.2 THE DISTRIBUTION-FREE PROPERTY

The term *nonparametric*, introduced in Section 1.1, is imprecise. The related term *distribution-free* has a precise meaning. The distribution-free property is a key aspect of many nonparametric procedures. In this section, we informally introduce the concept of a distribution-free test statistic. The related notions of a distribution-free confidence interval, distribution-free multiple comparison procedure, distribution-free confidence band, asymptotically distribution-free test statistic, asymptotically distribution-free multiple comparison procedure, and asymptotically distribution-free confidence band are introduced at appropriate points in the text.

Distribution-Free Test Statistic

We introduce the concept of a distribution-free test statistic by referring to the two-sample Wilcoxon rank sum statistic, which you will encounter in Section 4.1.

The data consist of a random sample of m observations from a population with continuous probability distribution F_1 and an independent random sample of n observations from a second population with continuous probability distribution F_2. The null hypothesis to be tested is

$$H_0 : F_1 = F_2 = F, \ F \text{ unspecified.}$$

The null hypothesis asserts that the two random samples can be viewed as a single sample of size $N = m + n$ from a common population with unknown distribution F. The Wilcoxon (1945) statistic W is obtained by ranking the combined sample of N observations jointly from least to greatest. The test statistic is W, the sum of the ranks obtained by the Y's in the joint ranking.

When H_0 is true, the distribution of W does not depend on F; that is, when H_0 is true, for all a-values, the probability that $W \le a$, denoted by $P_0(W \le a)$, does not depend on F.

$$P_0(W \le a) \text{ does not depend on } F. \tag{1.1}$$

The distribution-free property given by (1.1) enables one to obtain the distribution of W under H_0 without specifying the underlying F. It further enables one to exactly specify the type I error probability (the probability of rejecting H_0 when H_0 is true) without making distributional assumptions, such as the assumption that F is a normal distribution; this assumption is required by the parametric t-test.

The details concerning how to perform the Wilcoxon test are given in Section 4.1.

1.3 SOME REAL-WORLD APPLICATIONS

This book stresses the application of nonparametric techniques to real data. The following 10 examples are a sample of the type of problems you will learn to analyze using nonparametric methods.

EXAMPLE 1.1 ***Dose–Response Relationship.***

In many situations, a dose–response relationship may not be monotonic in the dosage. For example, with *in vitro* mutagenicity assays, experimental organisms may not survive the toxic side effects of high doses of the test agent, so there may be a reduction in the number of organisms at risk of mutation. This would lead to a downturn (i.e., an umbrella pattern) in the dose–response curve. The data in Table 6.10 were considered by Simpson and Margolin (1986) in a discussion of the analysis of the Ames test results. Plates containing *Salmonella* bacteria of strain TA98 were exposed to various doses of Acid Red 114. Table 6.10 gives the number of visible revertant colonies on the 18 plates in the study, three plates for each of the six doses (in μg/ml): 0, 100, 333, 1000, 3333, and 10,000. How can we test the hypothesis of equal population median numbers at each dose against the alternative that the peak of the dose–response curve occurs at 1000 μg/ml? How can we determine which particular pairs of doses, if any, significantly differ from one another in the number of revertant colonies? Which particular doses, out of 100, 333, 1000, 3333, and 10,000, differ significantly from the 0 dose in terms of the number of revertant colonies? For doses that significantly differ, how can we estimate the magnitude of the difference? How can we simultaneously estimate all 15 "contrasts," $\tau_1 - \tau_2, \tau_1 - \tau_3, \tau_1 - \tau_4, \tau_1 - \tau_5, \tau_1 - \tau_6, \tau_2 - \tau_3, \tau_2 - \tau_4, \tau_2 - \tau_5, \tau_2 - \tau_6, \tau_3 - \tau_4, \tau_3 - \tau_5, \tau_3 - \tau_6, \tau_4 - \tau_5, \tau_4 - \tau_6, \tau_5 - \tau_6$, where, for example, $\tau_1 - \tau_2$ denotes the difference between the population medians at dose 0 and dose 100. The methods in Chapter 6 can be used to answer these questions.

EXAMPLE 1.2 ***Shelterbelts.***

Shelterbelts are long rows of tree plantings across the direction of prevailing winds. They are used in developed countries to protect crops and livestock from the effects of the wind. A study was performed by Ujah and Adeoye (1984) to see if shelterbelts would limit severe losses from droughts regularly experienced in the arid and semiarid zones of Nigeria. Droughts are considered to be a leading factor in declining food production in Nigeria and in the neighboring countries. Ujah and Adeoye studied the effect of shelterbelts on a number of factors related to drought conditions, including wind velocity, air and soil temperatures, and soil moisture. Their experiment was conducted at two locations about $3\frac{1}{2}$ km apart, near Dambatta. Table 7.7 presents the wind velocity data, averaged over the two locations, at various distances leeward of the shelterbelt. The data are given as percent wind speed reduction relative to the wind velocity on the windward side of the shelterbelt. The data are given for 9 months (data were not available for July, November, and December) and five leeward distances, namely, 20, 40, 100, 150, and 250 m, from the shelterbelt. Does the percent reduction in average wind speed tend to decrease as the leeward distance from a shelterbelt increases? Which particular leeward distances, if any, significantly differ from one another in percent reduction in average wind speed? How can the difference in percent reduction for two leeward distances be estimated? Chapter 7 presents nonparametric methods that will enable you to analyze the data and answer these questions.

EXAMPLE 1.3 ***Nasal Brushing.***

In order to study the effects of pharmaceutical and chemical agents on mucociliary clearance, doctors often use the ciliary beat frequency (CBF) as an index of ciliary activity. One accepted way to measure CBF in a subject is through the collection and analysis of an endobronchial forceps biopsy specimen. This technique is, however, a rather invasive method for measuring CBF. In a study designed to assess the effectiveness of less invasive procedures for measuring CBF, Low et al. (1984) considered the alternative technique of nasal brushing. The data in Table 8.10 are a subset of the data collected by Low et al. during their investigation. The subjects in the study were all men undergoing bronchoscopy for the diagnosis of a variety of pulmonary problems. The CBF values reported in Table 8.10 are averages of 10 consecutive measurements on each subject.

How can we test the hypothesis of independence versus the alternative that the CBF measurements corresponding to nasal brushing and endobronchial forceps are positively associated? If there is evidence that the alternative is true, this would support the notion that nasal brushing is an acceptable substitute to measure CBF for the more invasive endobronchial forceps biopsy technique. How can we obtain an estimate of a measure of the strength of association between the two techniques' CBF values? How can we compute confidence intervals for such a measure? These questions can be answered by the methods described in Chapter 8.

EXAMPLE 1.4 ***Coastal Sediments.***

Coastal sediments are an important reservoir for organic nitrogen (ON). The degradation of ON is bacterially mediated. The mineralization of ON involves several distinct steps, and it is possible to measure the rates of these processes at each step. During the first stage of ON remineralization, ammonium is generated by heterotrophic bacteria during a process called *ammonification*. Ammonium can then be released to the environment or can be microbially transformed to other nitrogenous species. The data in Table 9.4 are from the work by Mortazavi (1997) and are based on four sediment cores that were collected in Apalachicola Bay, Florida, in April 1995 and brought back to the main campus at the Florida State University for analysis. The flux of ammonium to the overlying water was measured in each core during a 6-h incubation period. It is desired to know if there is a significant difference in ammonium flux between the cores. This is a regression problem, and it can be studied using the methods in Chapter 9.

EXAMPLE 1.5 ***Care Patterns for Black and White Patients with Breast Cancer.***

Diehr et al. (1989) point out that it is well known that the survival rate of women with breast cancer tends to be lower in Blacks than Whites. Diehr and her colleagues sought to determine if these survival differences could be accounted for by differences in diagnostic methods and treatments. Diehr et al. reported on various breast cancer patterns; one pattern of interest was *liver scan*. Did patients with local or regional disease have a liver scan or CT scan of the liver? The data are given in Table 10.14. The data are for the 19 hospitals (out of 107 hospitals participating in the study) that had enough Black patients for individual analysis. How can we determine, for a specific hospital, if there was a significant difference between the chance of a White patient receiving a scan and the chance of a Black patient receiving a scan? How can the data from the 19 hospitals be utilized to get an overall assessment? The methods in Chapter 10 provide the means to answer these questions.

EXAMPLE 1.6 ***Times to First Review.***

The data in Table 11.18, from Hollander, McKeague, and Yang (1997), relate to 432 manuscripts submitted to the Theory and Methods Section of the *Journal of the American Statistical Association* (*JASA*) in the period January 1, 1994, to December 13, 1994. Of interest is the time (in days) to first review. When the data were studied on December 13, 1994, 158 papers had not yet received the first review. For example, for a paper received by the *JASA* on November 1, 1994, and still awaiting the first review on December 13, 1994, we know on December 13 that its time to review is greater than 33 days, but at that point we do not know the actual time to review. The observation is said to be *censored*. How can we use the censored and uncensored observations (i.e., the ones for which we know the exact times to first review) to estimate the distribution of the time to first review? Chapter 11 shows how to estimate distributions when some of the data are censored.

EXAMPLE 1.7 ***Spatial Ability Scores of Students.***

In a study examining the relation between student mathematical performance and their preference for solving problems, Haciomeroglu and Chicken (2011) gathered data on a student's spatial ability using four tests of visualization. For each student, these four test scores were combined into a single score representing their overall measure of spatial ability. High scores are associated with students with strong spatialization skills, while low scores reflect weak spatialization. The spatial ability scores for 68 female and 82 male high school students enrolled in advanced placement calculus classes in Florida are given in Tables 12.1 and 12.3, respectively. What is the distribution of spatial ability scores for the population represented by this sample of data? Does the distribution for the male students appear to possess different characteristics than that of the female students? These questions are problems in density estimation. Methods on this are given in Chapter 12.

EXAMPLE 1.8 ***Sunspots.***

Andrews and Herzberg (1985) provide data on mean monthly sunspot observations collected at the Swiss Federal Observatory in Zurich and the Tokyo Astronomical Observatory from the years 1749 to 1983. The data display excessive variability over time, obscuring any underlying trend in the cycle of sunspot appearances. The data do not follow any apparent analytical form or simple parametric model so a general nonparametric regression setting is appropriate. A powerful method for obtaining the trend from a noisy set of observations in cases such as this is by the use of wavelet estimation and thresholding. Wavelet analysis will provide a smoothed and accurate estimate of the noise-free trend underlying the observed data. Chapter 13 provides details on using wavelet methods for this type of problem.

EXAMPLE 1.9 ***Effective Auditing to Detect Fraud.***

Account auditing is one of the most important ways to ensure that a company's stated records accurately represent the true financial transactions of the business. Being able to detect fraudulent accounting practices is vital to the integrity of the business and its management. Statistical sampling is a well-established approach for conducting such audits, as in almost all settings, the number of accounts of interest is far too large for a complete census. One major concern with statistical audits is that assessing the true values of the accounts selected to be part of the statistical sample can be quite time-intensive and, hence, expensive. It is therefore of interest to limit the number of accounts sampled for audit, while still providing adequate assurance that we gather enough information to accurately assess the reliability of the company's financial records. A ranked set sampling approach to select representative observations from a population allows an auditor to formally audit fewer accounts while maintaining the desired level of precision in his or her assessment. This leads to time savings and overall cost reduction for the auditing process. Tackett (2012) provided a collection of sales invoice records data for an electrical/plumbing distribution center that contained some fraudulent accounts where the charges (stated book values) for transactions were larger than the audited values for the materials actually delivered in those transactions. These data are given in Table 15.1. The ranked set sampling techniques described in Chapter 15 provide an effective mechanism for minimizing the auditing expense in assessing the fraudulent nature of these sales invoice records.

EXAMPLE 1.10 ***Times to Death with Cardiovascular Disease.***

The Framingham Heart Study is a well-known ongoing longitudinal study of cardiovascular disease. The original study cohort consisted of a random sample of 5209 adults aged 28 through 62 years residing in Framingham, Massachusetts between 1948 and 1951. The data in Table 16.1 were provided by McGee (2010) and consist of an extinct cohort of 12 men who were 67 years and over at the time of the fourth exam. How can we estimate the survival distribution underlying this population? How can we incorporate expert opinion concerning the remaining life for men under those or similar situations? This is a survival problem that incorporates prior information. It can be studied using the methods of Chapter 16.

1.4 FORMAT AND ORGANIZATION

The basic data, assumptions, and procedures are described precisely in each chapter according to the following format. *Data* and *Assumptions* are specified before the group of particular procedures discussed. Then, for each technique, we include (when applicable) the following subsections: *Procedure*, *Large-Sample Approximation*, *Ties*, *Example*, *Comments*, *Properties*, and *Problems*. We now describe the purpose of each subsection.

Procedure

This subsection contains a description of how to apply the procedure under discussion.

Large-Sample Approximation

This subsection contains an approximation to the method described in *Procedure*. The approximation is intended for use when the sample size (or sample sizes, as the case may be) is large. Our R programs enable small-sample and large-sample applications.

Ties

A common assumption in the development of nonparametric procedures is that the underlying population(s) is (are) continuous. This assumption implies that the probability of obtaining tied observations is zero. Nevertheless, tied observations do occur in practice. These ties may arise when the underlying population is not continuous. They may even arise if the continuity assumption is valid. We simply may be unable, owing to inaccuracies in measurement, to distinguish between two very close observations (temperatures, lengths, etc.) that emanate from a continuous population. The *Ties* subsection contains a prescription to adjust the necessary steps in the *Procedure* in order that we may treat tied observations. The adjusted procedure should then be viewed as an approximation.

Example

This subsection is basic to our text. We present a problem in which the procedure under discussion is applicable. The reader has a set of data he or she may use to apply each step of the *Procedure*, to become familiar with our notation, and to gain familiarity in performing the method. In many examples, computations are done directly and using R commands. In addition to practice, the example provides the first step toward developing an appreciation for the simplicity (difficulty) of the procedure and toward developing an intuitive feeling of how the procedure summarizes the data. The enthusiastic reader can seek out the journal article on which the example is based to obtain a more detailed specification of the experiment (in some cases our descriptions of the experiments are simplified so that the examples can be easily explained) and to question whether the *Assumptions* underlying the nonparametric method are indeed satisfied.

Comments

The comments supplement the text. In the comments, we may discuss the underlying assumptions, give an intuitive motivation for the method being considered, relate the method to other procedures in different parts of the book, provide helpful computational hints, or single out certain references including historical references.

Properties

This subsection is primarily intended as a set of directions for the reader who wishes to probe the theoretical aspects of the subject and, in particular, the theory of the procedure under discussion. No theory is presented, but the citations guide the reader to sources furnishing the basic properties and their derivations.

Problems

Typically, the first problem of each *Problems* subsection provides practice in applying the procedure just introduced. Some problems require a comparison of an exact procedure with its large-sample approximation. Other problems are more thought provoking. We sometimes ask the reader to find or create an example that illustrates a desirable or undesirable property of the procedure under discussion.

There are occasional deviations from the format. For example, in many of the sections devoted to estimators and confidence intervals, there is no need for a *Ties* subsection, because the procedures described are well defined even when ties observations occur. In some chapters, the *Assumptions* are given before the particular (group of) sections that contain procedures based on those *Assumptions.*

Efficiency

How do the nonparametric procedures we present compare with their classical competitors, which are based on parametric assumptions such as the assumption of normality for the underlying populations? The answer depends on the particular problem and procedures under consideration. When possible, we indicate a partial answer in an efficiency section at the end of each chapter.

1.5 COMPUTING WITH R

In many of our *Example* subsections, we not only illustrate the direct computation of the procedure but also provide the output obtained using various commands in the statistical computing package R. R is a general-purpose statistical package that provides a wide range of data analysis capabilities. It is an open source program that is available for a variety of computing platforms. Users may obtain the software free of charge through the Comprehensive R Archive Network (CRAN). CRAN is a network of ftp and web servers that provide all the necessary files and instructions for downloading and installing R. It also contains numerous manuals and FAQs to assist users.

One of the strengths of R is its openness. Individuals around the world may create packages of statistical commands and routines to be distributed to any other interested users through CRAN. The standard distribution of R contains the resources to perform many of the nonparametric methods described in this book. Additional packages are readily available that perform more specialized analyses such as the density estimation procedures and wavelet analyses in the book's later chapters. Whenever a command is referenced that is not a part of the standard installation of R but instead comes from an add-on package, we make a note of this and specify which package is needed to perform the analysis.

R is also a programming language. If one cannot find an existing statistical methodology within R that will perform a suitable analysis, it is possible to program unique commands to fill this void. This falls under the topic of programming, rather than statistical analysis. As such, programming within R is not covered. The main procedures discussed in this book have specific sets of existing commands that will perform the appropriate actions.

Many analyses include graphical as well as numeric output. `R` has a significant number of built-in graphing functions and is very flexible in that it allows users to create unique and detailed graphs to suit their specific needs.

The results of statistical analyses performed using R may vary slightly from those presented in the text. When they exist, these differences will be minor and will depend on the hardware configuration of the machine used to run the analyses. We also note that, for large sample sizes, many of the programs will use Monte Carlo approximations by default. Specifying methods=`"Exact"`, while more computationally intensive, will ensure that the user's output matches the text.

1.6 HISTORICAL BACKGROUND

Binomial probability calculations were used early in the eighteenth century by the British physician Arbuthnott (1710) (see Comment 2.13). Nevertheless, Savage (1953) (also see Savage (1962)) designated 1936 as the true beginning of the subject of nonparametric statistics, marked by the publication of the Hotelling and Pabst (1936) paper on rank correlation. Scheffé (1943), in a survey paper, pointed to (among others) the articles by Pearson (1900, 1911) and the presence of the sign test in Fisher's first edition of "Statistical Methods for Research Workers" Fisher (1925). Other important papers, in the late 1930s, include those by Friedman (1937), Kendall (1938), and Smirnov (1939). Wilcoxon (1945), in a paper that is brief, yet elegant in its simplicity and usefulness, introduced his now-famous two-sample rank sum test and paired-sample signed rank test. The rank sum test was given by Wilcoxon only for equal sample sizes, but Mann and Whitney (1947) treated the general case. Wilcoxon's procedures played a major role in stimulating the development of rank-based procedures in the 1950s and 1960s, including rank procedures for multivariate situations. Further momentum was provided by Pitman (1948), Hodges and Lehmann (1956), and Chernoff and Savage (1958), who showed that nonparametric rank tests have desirable efficiency properties relative to parametric competitors. An important advance that enabled nonparametric methods to be used in a variety of situations was the jackknife, introduced by Quenouille (1949) as a bias-reduction technique and extended by Tukey (1958, 1962) to provide approximate significance tests and confidence intervals.

There was major nonparametric research in the 1960s, and the most important contribution was that of Hodges and Lehmann (1963). They showed how to derive estimators from rank tests and established that these estimators have desirable properties. Their work paved the way for the nonparametric approach to be used to derive estimators in experimental design settings and for nonparametric testing and estimation in regression. Two seminal papers in the 1970s are those by Cox (1972) and Ferguson (1973). Cox's paper sparked research on nonparametric models and methods for survival analysis. Ferguson (1973) presented an approach (based on his Dirichlet process prior) to nonparametric Bayesian methods that combines the advantages of the nonparametric approach and the use of prior information incorporated in Bayesian procedures. Susarla and van Ryzin (1976) used Ferguson's approach to derive nonparametric Bayesian estimators of survival curves. Dykstra and Laud (1981) used a different prior, the gamma process, to develop a Bayesian nonparametric approach to reliability. Hjort (1990b) proposed nonparametric Bayesian estimators based on using beta processes to model the cumulative hazard function. In the late 1980s and the 1990s, there was a surge of activity in Bayesian methods due to the Markov chain Monte Carlo (MCMC) methods (see, for example, Gelfand and Smith (1990), Gamerman (1991), West (1992), Smith and Roberts (1993), and Arjas and Gasbarra (1994)). Gilks, Richardson, and Spiegelhalter (1996) give a practical review. Key algorithms for developing and implementing modern Bayesian methods include the Metropolis–Hastings–Green algorithm (see Metropolis et al. (1953), Hastings (1970), and Green (1995)) and the Tanner–Wong (1987) data augmentation algorithm.

One of the important advances in nonparametric statistics in the past $3\frac{1}{2}$ decades is Efron's (1979) bootstrap. Efron's computer-intensive method makes use of the (ever-increasing) computational power of computers to provide standard errors and confidence intervals in many settings, including complicated situations where it is difficult, if not impossible, to use a parametric approach (see Efron and Tibshirani (1994)).

In the new millennium, the development of nonparametric techniques continues at a vigorous pace. The *Journal of Nonparametric Statistics* is solely devoted to nonparametric methods and nonparametric articles are prevalent in most statistical journals. A special issue of *Statistical Science* (Randles, Hettmansperger, and Casella, 2004) contains papers written by nonparametric experts on a wide variety of topics. These include articles on robust analysis of linear models (McKean, 2004), comparing variances and other dispersion measures (Boos and Brownie, 2004), use of sign statistics in one-way layouts (Elmore, Hettmansperger, and Xuan, 2004), density estimation (Sheather, 2004), multivariate nonparametric tests (Oja and Randles, 2004), quantile–quantile (QQ) plots (Marden, 2004), survival analysis (Akritas, 2004), spatial statistics (Chang, 2004), ranked set sampling (Wolfe, 2004), reliability (Hollander and Peña, 2004), data modeling via quantile methods (Parzen, 2004), kernel smoothers (Schucany, 2004), permutation-based inference (Ernst, 2004), data depth tests for location and scale differences for multivariate distributions (Li and Liu, 2004), multivariate signed rank tests in time series problems (Hallin and Paindaveine, 2004), and rank-based analyses of crossover studies (Putt and Chinchilli, 2004).

Books dealing with certain topics in nonparametrics include those on survival analysis (Kalbfleisch and Prentice, 2002 and Klein and Moeschberger, 2003), density estimation, smoothers and wavelets (Wasserman, 2006), rank-based methods (Lehmann and D'Abrera, 2006), reliability (Gámiz, Kulasekera, Limnios, and Lindquist, 2011), and categorical data analysis (Agresti, 2013).

We delineated advantages of the nonparametric approach in Section 1.1. In addition to those practical advantages, the theory supporting nonparametric methods is elegant, and researchers find it challenging to advance the theory. The primary reasons for the success and use of nonparametric methods are the wide applicability and desirable efficiency properties of the procedures and the realization that it is sound statistical practice to use methods that do not depend on restrictive parametric assumptions because such assumptions often fail to be valid.

Chapter 2

The Dichotomous Data Problem

INTRODUCTION

In this chapter the primary focus is on the dichotomous data problem. The data consists of n independent repeated Bernoulli trials having constant probability of success p. On the basis of these outcomes, we wish to make inferences about p. Section 2.1 introduces the binomial distribution and presents a binomial test for the hypothesis $p = p_0$, where p_0 is a specified success probability. Section 2.2 gives a point estimator $\widehat{p}$ for p. Section 2.3 presents confidence intervals for p. Section 2.3 also contains the generalization of the binomial distribution to the multinomial distribution, confidence intervals for multinomial probabilities and a test that the multinomial probabilities are equal to specified values. Section 2.4 presents Bayesian competitors to the frequentist estimator $\widehat{p}$ of Section 2.2. The Bayesian estimators incorporate prior information.

Data. We observe the outcomes of n independent repeated Bernoulli trials.

Assumptions

A1. The outcome of each trial can be classified as a success or a failure.

A2. The probability of a success, denoted by p, remains constant from trial to trial.

A3. The n trials are independent.

2.1 A BINOMIAL TEST

Procedure

To test

$$H_0 : p = p_0, \tag{2.1}$$

where p_0 is some specified number, $0 < p_0 < 1$, set

$$B = \text{number of successes}. \tag{2.2}$$

a. *One-Sided Upper-Tail Test.* To test

$$H_0 : p = p_0$$

Nonparametric Statistical Methods, Third Edition. Myles Hollander, Douglas A. Wolfe, Eric Chicken.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.

versus

$$H_1 : p > p_0$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } B \geq b_\alpha; \text{ otherwise do not reject,} \tag{2.3}$$

where the constant b_α is chosen to make the type I error probability equal to α. The number b_α is the upper α percentile point of the binomial distribution with sample size n and success probability p_0. Due to the discreteness of the binomial distribution, not all values of α are available (unless one resorts to randomization). Comment 3 explains how to obtain the b_α values. See also Example 2.1.

b. *One-Sided Lower-Tail Test.* To test

$$H_0 : p = p_0$$

versus

$$H_2 : p < p_0$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } B \leq c_\alpha; \text{ otherwise do not reject.} \tag{2.4}$$

Values of c_α can be determined as described in Comment 3. Here, c_α is the lower α percentile point of the binomial distribution with sample size n and success probability p_0. For the special case of testing $p = \frac{1}{2}$,

$$c_\alpha = n - b_\alpha. \tag{2.5}$$

Equation 2.5 is explained in Comment 4.

c. *Two-Sided Test.* To test

$$H_0 : p = p_0$$

versus

$$H_3 : p \neq p_0$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } B \geq b_{\alpha_1} \text{ or } B \leq c_{\alpha_2}; \text{ otherwise do not reject,} \tag{2.6}$$

where b_{α_1} is the upper α_1 percentile point, c_{α_2} is the lower α_2 percentile point, and $\alpha_1 + \alpha_2 = \alpha$. See Comment 3.

Large-Sample Approximation

The large-sample approximation is based on the asymptotic normality of B, suitably standardized. To standardize, we need to know the mean and variance of B when the null hypothesis is true. When H_0 is true, the mean and variance of B are, respectively,

$$E_{p_0}(B) = np_0, \tag{2.7}$$

$$\text{var}_{p_0}(B) = np_0(1-p_0). \tag{2.8}$$

Comment 8 gives the derivations for (2.7) and (2.8).

The standardized version of B is

$$B^* = \frac{B - E_{p_0}(B)}{\{\text{var}_{p_0}(B)\}^{1/2}} = \frac{B - np_0}{\{np_0(1-p_0)\}^{1/2}}. \tag{2.9}$$

When H_0 is true, B^* has, as n tends to infinity, an asymptotic $N(0,1)$ distribution. Let z_α denote the upper α percentile point of the $N(0,1)$ distribution. To find z_α, we use the `qnorm`(1-α,0,1). For example, to find $z_{.05}$, we apply qnorm(.95,0,1) and obtain $z_{.05} = 1.645$.

The normal approximation to procedure (2.3) is

$$\text{Reject } H_0 \text{ if } B^* \geq z_\alpha; \text{ otherwise do not reject.} \tag{2.10}$$

The normal approximation to procedure (2.4) is

$$\text{Reject } H_0 \text{ if } B^* \leq -z_\alpha; \text{ otherwise do not reject.} \tag{2.11}$$

The normal approximation to procedure (2.6), with $\alpha_1 = \alpha_2 = \alpha/2$, is

$$\text{Reject } H_0 \text{ if } |B^*| \geq z_{\alpha/2}; \text{ otherwise do not reject.} \tag{2.12}$$

EXAMPLE 2.1 ***Canopy Gap Closure.***

Dickinson, Putz, and Canham (1993) investigated canopy gap closure in thickets of the clonal shrub *Cornus racemosa.* Shrubs often form dense clumps where tree abundance has been kept artificially low (e.g., on power-line right of ways). These shrub clumps then retard reinvasion of the sites by trees. Individual clumps may persist for many years. Clumps outlast the lives of the individual stems of which they are formed; stems die and leave temporary holes in the canopies of the clumps. Closure of the hope (gap) left by dead stems occurs in part by the lateral growth of stems that surround the hole. Opening of the gap often occurs when individual branches of hole-edge stems die. Between sample dates, more branches in six out of seven gaps in clumps, at a site with nutrient-poor and dry soil, died than lived. Let us say we have a success if more branches die than live in the gaps in clumps. Let p denote the corresponding probability of success. We suppose that the success probability for sites that are nutrient rich with moist soil has been established by previous studies to be 15%. Do the nutrient-poor and dry soil sites have the same success probability as the nutrient-rich and moist soil sites or is it larger? This reduces to the hypothesis-testing problem

$$H_0 : p = .15$$

versus

$$H_1 : p > .15.$$

Our sample size is $n = 7$ and we observe $B = 6$ successes. From the R command `round(pbinom(0:7,7,.15,lower.tail=F),4)`, we obtain, rounded to four places, the probabilities $P_{.15}(B > x)$ for $x = 0, \dots, 7$. (The notation $P_{.15}(B > x)$ is shorthand for the probability that $B > x$, computed under the assumption that the true success probability is .15.) The $P_{.15}(B > x)$ probabilities are

x	0	1	2	3	4	5	6	7
$P_{.15}(B > x)$	.6794	.2834	.0738	.0121	.0012	.0001	.0000	.0000

To find $P_{.15}(B \geq x)$ note $P_{.15}(B \geq x) = P_{.15}(B > x - 1)$. Reasonable possible choices for α are .0738, .0121, .0012, .0001. Suppose we choose to use $\alpha = .0121$. We note $P(B > 3) = P(B \geq 4) = .0121$ and thus we see $b_{.0121} = 4$. Thus the $\alpha = .0121$ test is

$$\text{Reject } H_0 \text{ if } B \geq 4; \text{ otherwise do not reject.}$$

Our observed value is $B = 6$ and thus we reject H_0 at $\alpha = .0121$. To find the P-value, which is $P_{.15}(B \geq 6)$, we can find $P_{.15}(B > 5)$ using the R command `pbinom(5,7,.15,lower.tail=F)`. Alternatively, we can find the P-value using the R command `binom.test(6,7,.15,"g")`. We find $P = .000069$, or rounded to four places, P is .0001. This is the smallest significance level at which we can reject H_0 (in favor of the alternative $p > .15$) with our observed value of B. We conclude that there is strong evidence against H_0 favoring the alternative. For more on the P-value, see Comment 9.

EXAMPLE 2.2 ***Sensory Difference Tests.***

Sensory difference tests are often used in quality control and quality evaluation. The triangle test (cf. Bradley, 1963) is a sensory difference test that provides a useful application of the binomial model. In its simplest form, the triangle procedure is as follows. To each of n panelists, three test samples are presented in a randomized order. Two of the samples are known to be identical; the third is different. The panelist is then supposed to select the odd sample, perhaps on the basis of a specified sensory attribute. If the panelists are homogeneous trained judges, the experiment can be viewed as n independent repeated Bernoulli trials, where a success corresponds to a correct identification of the odd sample. (If the panelists are not homogeneous trained judges, we may question the validity of Assumption A2.) Under the hypothesis that there is no basis for discrimination, the probability p of success is $\frac{1}{3}$, whereas a basis for discrimination would correspond to values of p that exceed $\frac{1}{3}$.

Byer and Abrams (1953) considered triangular bitterness tests in which each taster received three glasses, two containing the same quinine solution and the third a different