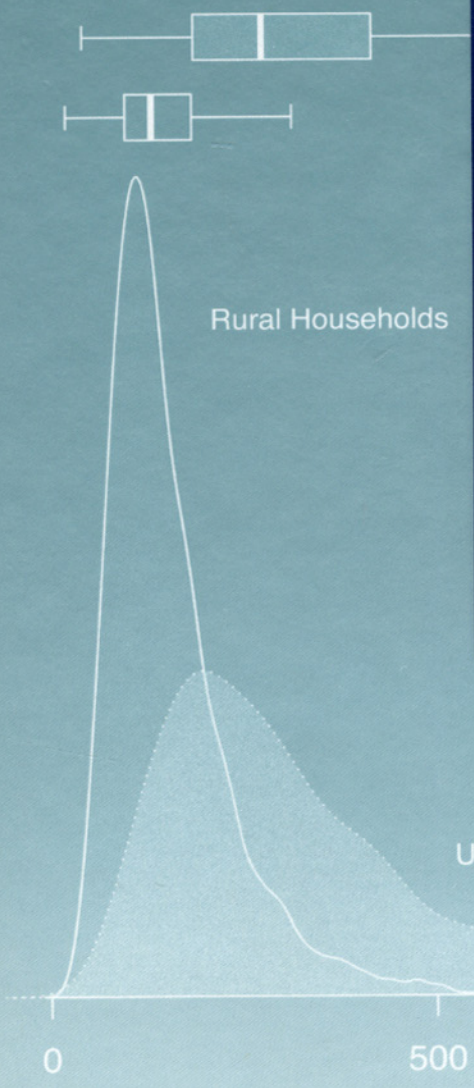

COMPARING GROUPS

Randomization and Bootstrap Methods Using R

Andrew S. Zieffler · Jeffrey R. Harring · Jeffrey D. Long

WILEY

This page intentionally left blank

Comparing Groups

This page intentionally left blank

Comparing Groups

Randomization and Bootstrap Methods Using R

Andrew S. Zieffler
University of Minnesota

Jeffrey R. Harring
University of Maryland

Jeffrey D. Long
University of Iowa

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

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

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Zieffler, Andrew, 1974–
Comparing groups : randomization and bootstrap methods using R / Andrew Zieffler, Jeffrey Harring, Jeffrey D. Long.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-62169-1 (hardback)
1. Bootstrap (Statistics) 2. Random data (Statistics) 3. Psychology—Data processing. 4. R (Computer program language) 5. Distribution (Probability theory) I. Harring, Jeffrey, 1964– II. Long, Jeffrey D., 1964– III. Title.
QA276.8.Z53 2011
519.5'4—dc22 2010054064

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTENTS

LIST OF FIGURES

LIST OF TABLES

FOREWORD

This book is truly different.

It is rare that a textbook with so few technical prerequisites offers so much of value, not just to its intended student audience, but also to all practicing statisticians. The book in your hands is one of those rare ones. It is the product—over years—of careful study and deep thought by its authors, who are practicing statisticians, education researchers, and root-deep innovators.

Statistics, its teaching and its learning, have seen a number of exciting developments over the last quarter century. Among these:

1. Near-universal acceptance of the assertion that applied statistics textbooks should emphasize the priority of real data over mathematically driven abstract exposition.
2. A coming-of-age for computer-intensive methods of data analysis.
3. A parallel recognition by teachers of statistics that, along with an emphasis on real data and reliance on applied context, computer simulation offers an alternative to abstract derivations as an approach to exposition and understanding.
4. A cooperative development and broad embrace of R, the public domain, and open-source software, whose growing community of contributors has made cutting edge, computationally intensive methods freely available around the world.
5. A coming-of-age of statistics education as a research area in its own right, and, as a result, a growing clarity about what is hardest for students to learn, and a sharpened focus on what is most important for students to learn.

6. A change at the foundations of our subject: What we can compute shapes what we can do; what we can do shapes what we actually do in practice; what we do in practice shapes how we think about what we do. (As just one example, when I first was learning statistics in the 1960s, Bayesian analyses were typically dismissed as mere idealism; today, the combination of hierarchical models and Markov chain Monte Carlo has turned Bayes resistors into old fogeys.)

By integrating these developments in a novel way, *Comparing Groups* deserves to be regarded as one of this quarter-century's pioneering textbooks aimed at introducing nonstatisticians to contemporary thought and practice.

Many features are distinctive:

1. The exposition is visual/intuitive/computational rather than haunted by the old headless horseman of abstract derivations.
2. The writing is strikingly good—the exposition reflects the authors' careful attention to word choice. (In my experience, it is rare in statistical exposition to read with a sense that the writers made thoughtful, deliberate choices about wording after considering a variety of options.)
3. The references are a gold mine.
4. The use of modern graphics is unusual for a book at this level. I refer, in particular, to panel plots and kernel density estimates.
5. Prerequisites are minimal. No calculus is needed. Although it helps to know about vectors and matrices, one does not need a linear algebra course.
6. The emphasis is practical throughout. For example, the exposition refers systematically to APA guidelines and offers sample write-ups.
7. Content reflects the current research literature. For example, the exposition recognizes the importance of effect sizes, and the treatment of multiple testing addresses the recent research on false discovery rates.
8. Overall, the emphasis is on statistics with a purpose, statistics for deep interpretive understanding.

In short, this is a book that reduces prerequisites, avoids technical baggage, focuses on essentials, and teaches via authentic applications. It offers readers our profession's current best sense of how to understand data for comparing groups, taking full advantage of computationally intensive methods such as randomization and the bootstrap.

George W. Cobb
Robert L. Rooke Professor of Statistics, emeritus, Mount Holyoke College
Vernon Wilson Endowed Visiting Professor, Eastern Kentucky University

PREFACE

Computational advances have changed the face of statistical practice by transforming what we do and by challenging how we think about scientific problems.

—R. A. Thisted & P. F. Velleman (1992)

Drawing conclusions and inferences about the differences among groups is an almost daily occurrence in the scholarly life of an educational or behavioral researcher. Group comparisons are at the heart of many research questions addressed by these researchers. These are questions about efficacy, such as "Is a particular curriculum effective in improving students' achievement?" and questions about magnitude such as "How much lower are attendance rates for a particular population of students?"

The content in this book provides the statistical foundation for researchers interested in answering these types of questions through the introduction and application of current statistical methods made possible through computation—including the use of Monte Carlo simulation, bootstrapping, and randomization tests. Rather than focus on mathematical calculations like so many other introductory texts in the behavioral sciences, the approach taken here is to focus on conceptual explanations and the use of statistical computing. We agree with the sentiments of Moore (1990, p. 100), who stated, "calculating sums of squares by hand does not increase understanding; it merely numbs the mind."

At the heart of every chapter there is an emphasis on the direct link between research questions and data analysis. Purposeful attention is paid to the integration of design, statistical methodology, and computation to propose answers to research questions based on appropriate analysis and interpretation of quantitative data. Practical suggestions for analysis and the presentation of results based on suggestions from the *APA Publication Manual* are also included. These suggestions are intended to help researchers clearly communicate the results of a data analysis to their audience.

Computation as a Tool for Research

Computation is ubiquitous in everyday life. This is in large part due to the progress made in technology in the last 20 years. The price of computational power continually becomes less expensive and more powerful. In a recent *New York Times* article Caleb Chung—the inventor of the electronic Furby toy—was quoted as saying, "the price of processing power has dropped to the floor. I can buy the equivalent of an Apple II processor for a dime" (Marriot, 2007, p. 9).

Computing has become an essential part of the day-to-day practice of statistical work. It has not only greatly changed the practice of statistics itself, but has influenced the development of new state-of-the-art statistical methodologies and broadened the types of questions that can now be addressed by research scientists applying these newly derived data analytic techniques. Previous to having the availability of such computational power, the outcome of a research project may have been crucially dependent on the efficiency or numerical accuracy of algorithms employed by a methodologist. Computational advantages have allowed educational and behavioral researchers to capitalize on sheer computing power to solve problems that previously were impractible or intractable.

Although computer-based data analysis probably covers most of the activity that educational and behavioral researchers use in their work, statistical computing is more than just using a software package for the analysis of data. It also encompasses the programming and development of new functionality or software, the analysis of statistical and numerical algorithms, and data management. Furthermore, computing is an important avenue for allowing statisticians and scientists to pursue lines of inquiry that in the past would not have been possible. It is also an essential tool in any modern collaborative effort.

To support and help facilitate the use of scientific computing, examples using the R computer language will be used throughout the book. Rather than relegate examples to the end of chapters, our approach is to interweave the computer examples with the narrative of the monograph. R is free, which means it is available to individuals with various resources. It is also a professional-level software environment that implements many of the modern statistical methods emphasized in the monograph. Furthermore, the architecture of the object orientation in R easily allows data analysis to be performed based on structured inquiry, in which a series of interrelated and increasingly complex questions are posed to guide and inform the process of data analysis. In the classroom, structured inquiry can increase the opportunities for

intellectual engagement (e.g., Badley, 2002; Cooper, 2005; Lee, 2005; Prince & Felder, 2006). Lastly, since R is a functional programming language, it can be easily extended for use in a great many analytic scenarios in the behavioral and educational sciences.

Organization of the Book

The organization of this book has been shaped by the authors' experiences in teaching the statistical and computing content presented in many graduate courses for social science students. The topics introduced represent the authors' beliefs about relevant content for inclusion in an introductory graduate-level statistics course for educational and behavioral science students.

This content covers statistical computing, exploratory data analysis, and statistical inference. The individual chapters in the book endeavor to integrate these ideas to help guide educational and behavioral researchers through the process of data analysis when making group comparisons. While most chapters address multiple strands of the data analytic process, in surveying the computing and statistical content of each chapter, it is clear that each chapter is primarily focused to address a particular strand.

Statistical Computing

The first two chapters are written primarily for educational and behavioral researchers who may not have prior experience with computing environments and, in particular, with the R statistical computing environment. Chapter 1 introduces the fundamentals of R. Basic ideas of computation are presented as the building blocks for the remainder of the monograph. In Chapter 2, these ideas are further developed through the reading in and manipulation of external data (data saved outside of R). This chapter also introduces script files as a way of recording useful syntax.

Exploratory Data Analysis

Chapters 3, 4, and 5 introduce graphical and numerical methods of exploratory analysis for group comparisons. These chapters begin the integration of the statistical and computing content. Chapter 3 focuses on graphical exploration of a marginal distribution to introduce the statistical and computing content without the additional cognitive load of making comparisons. Ideas of kernel density estimation are at the forefront of this chapter. Additionally, the use of R to create publication quality plots is highlighted.

Chapters 4 and 5 focus on the exploration of conditional distributions for two and more than two groups, respectively. Graphical exploration is expanded to facilitate comparisons (e.g., panel plots) and numerical summary measures are introduced to quantify characteristics of the distributions.

Statistical Inference

Chapter 6 is the first of two chapters on the use of Monte Carlo methods for statistical inference. This chapter presents randomization and permutation tests to examine group differences. Chapter 7, the second such chapter, presents parametric and nonparametric bootstrap tests. Both of these chapters highlight the quantification of how likely an observed statistic is given the expected variation in that statistic because of chance. Chapter 6 highlights the expected random variation due to random assignment, whereas Chapter 7 highlights the random variation due to random sampling.

Chapter 8 expands on the differences between the randomization and bootstrap tests by offering philosophical reasons for using each method apart from the design employed by the researcher. This is especially highlighted for observational studies. This chapter also offers an overview of differences between the Fisher and Neyman–Pearson philosophies of statistical testing.

Chapters 9 and 10 expand the realm of statistical inference. Chapter 9 introduces ideas of interval estimation. The bootstrap methodology is developed further in this chapter, but under stratification of groups. Standardized and unstandardized measures of effect size are discussed. In Chapter 10, testing and estimation for dependent samples is introduced. Rather than do this through the use of repeated measures, which is often the case in statistics books in the educational and behavioral sciences, dependence is introduced via the use of blocking in the research design.

Chapters 11 and 12 introduce common methods in the educational and behavioral sciences for comparing more than two groups. Chapter 11 is focused on planned comparisons. Exploration, testing and estimation are revisited in the context of multiple groups. The use of linear contrasts is introduced to facilitate these ideas.

Chapter 12 is focused on unplanned comparisons. Three common methods are presented: (1) adjusted group comparisons without the omnibus test; (2) unadjusted group comparisons following an omnibus test; and (3) adjusted group comparisons following an omnibus test. Strengths, weaknesses, and criticisms of each method are discussed. The use of ensemble-adjusted p-values and adjusted bootstrap intervals are discussed in this chapter.

Extras

This book refers to and uses several data sets throughout the text. Each of these data sets is available online at `http://www.tc.umn.edu/~zief0002/ComparingGroups.html`. The codebook for each data set is also available at the Web site. Many of these data sets were used in actual research studies that have been published and were graciously provided by the authors and researchers involved in those studies.

The R script files for each chapter are also available at the URL mentioned above. The commands are provided so that the reader can reproduce all the output and plots discussed in the monograph.

Problems can be found at the end of each chapter, except Chapter 8. These problems are meant to provide a platform to perform data analyses using real data