Statistics in the Health Sciences

Theory, Applications, and Computing

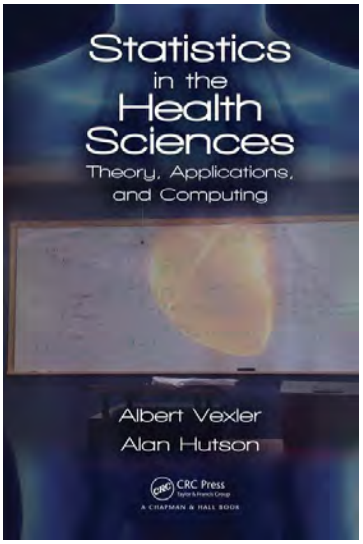
Albert Vexler

Alan Hutson

This book should be appropriate for use both as a text and as a reference. This book delivers a "ready-to-go" well-structured product to be employed in developing advanced courses. In this book the readers can find classical and new theoretical methods, open problems and new procedures.

KEY FEATURES

- A systematic compendium of intermediate and advanced methods in biostatistics
- A good balance of rigor in methodology and applications
- Software code based on both the R and SAS software packages to exemplify the presented topics.
- Detailed worked examples, exercises and exam questions, as well as student projects



SELECTED CONTENTS

Catalog no. K31208
February 2018, 398 pp.
ISBN: 978-1-1381-9689-6
$99.95 / £77.00

Statistics

"This very informative book introduces classical and novel statistical methods that can be used by theoretical and applied biostatisticians to develop efficient solutions for real-world problems encountered in clinical trials and epidemiological studies. The authors provide a detailed discussion of methodological and applied issues in parametric, semiparametric, and nonparametric approaches, including computationally extensive data-driven techniques, such as empirical likelihood, sequential procedures, and bootstrap methods. Many of these techniques are implemented using popular software such as R and SAS."

— Vlad Dragalin, Professor, Johnson & Johnson
Spring House, Pennsylvania

"It is always a pleasure to come across a new book that covers nearly all facets of a branch of science one thought was so broad, so diverse, and so dynamic that no single book could possibly hope to capture all of the fundamentals as well as directions of the field. The topics within the book's purview—fundamentals of measure-theoretic probability; parametric and nonparametric statistical inference; central limit theorems; basics of martingale theory; Monte Carlo methods; sequential analysis; sequential change-point detection—are all covered with inspiring clarity and precision. The authors are also very thorough and avail themselves of the most recent scholarship. They provide a detailed account of the state of the art, and bring together results that were previously scattered across disparate disciplines. This makes the book more than just a textbook: it is a panoramic companion to the field of Biostatistics. The book is self-contained, and the concise but careful exposition of material makes it accessible to a wide audience. This is appealing to graduate students interested in getting into the field, and also to professors looking to design a course on the subject."

— Aleksey S. Polunchenko, Department of Mathematical Sciences,
State University of New York at Binghamton

This book should be appropriate for use both as a text and as a reference. This book delivers a "ready-to-go" well-structured product to be employed in developing advanced courses.

In this book the readers can find classical and new theoretical methods, open problems, and new procedures.

The book presents biostatistical results that are novel to the current set of books on the market and results that are even new with respect to the modern scientific literature. Several of these results can be found only in this book.

- A systematic compendium of intermediate and advanced methods in biostatistics
- A good balance of rigor in methodology and applications
- Software code based on both the R and SAS software packages to exemplify the presented topics
- Detailed worked examples, exercises, and exam questions, as well as student projects

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K31208
ISBN: 978-1-138-19689-6
9 781138 196896 90000
www.crcpress.com

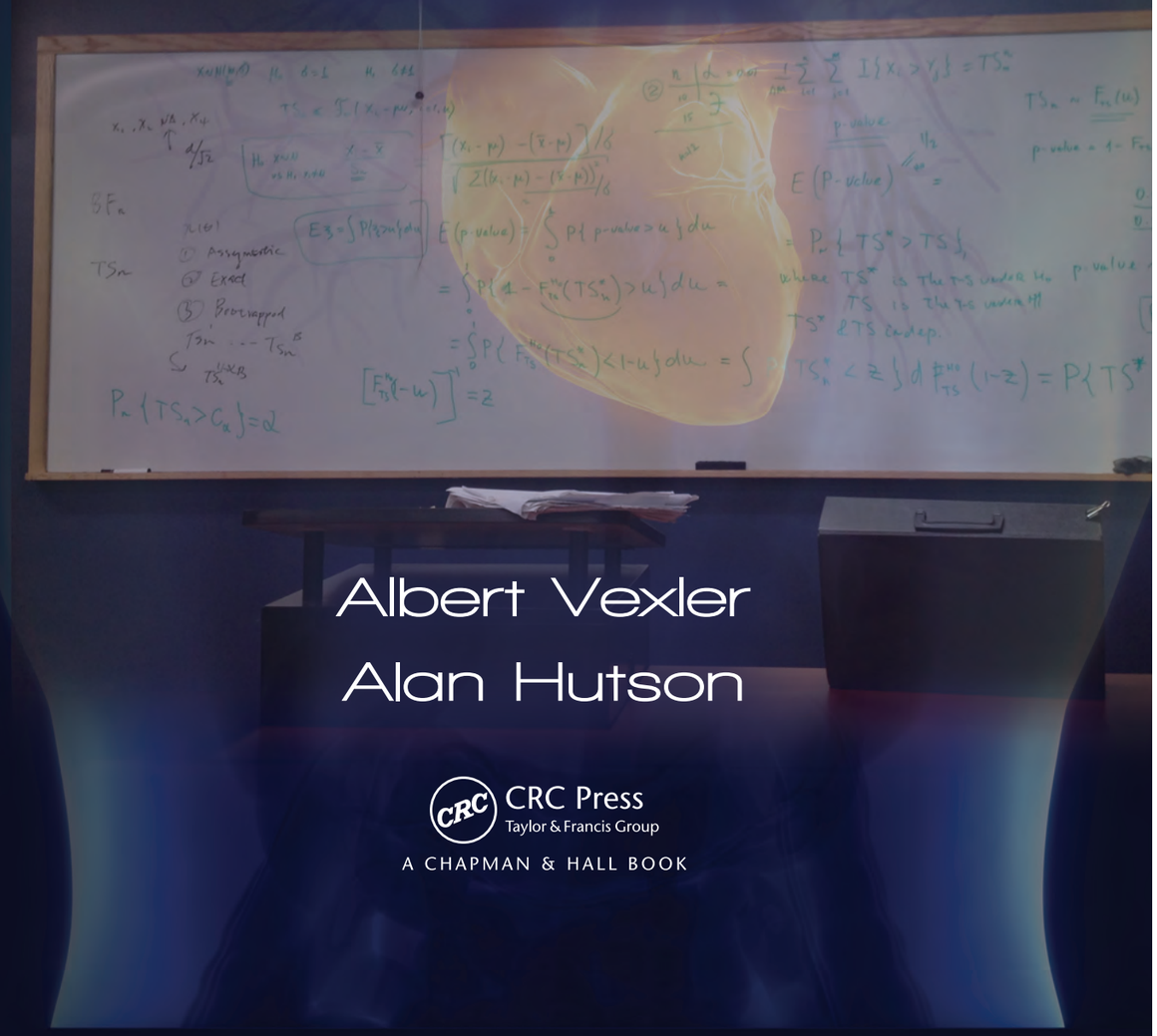
Statistics in the Health Sciences

Vexler
Hutson

CRC

Statistics in the Health Sciences

Theory, Applications and Computing



Albert Vexler
Alan Hutson

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Statistics in the Health Sciences

Theory, Applications and Computing

Statistics in the Health Sciences

Theory, Applications and Computing

By

Albert Vexler

The State University of New York at Buffalo, USA

Alan D. Hutson

Roswell Park Cancer Institute, USA

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-138-19689-6 (Hardback)

Library of Congress Cataloging-in-Publication Data

Names: Vexler, Albert, author. | Hutson, Alan D., author.
Title: Statistics in the health sciences : theory, applications, and computing / Albert Vexler and Alan D. Hutson.
Description: Boca Raton, Florida : CRC Press, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2017031841| ISBN 9781138196896 (hardback) | ISBN 9781315293776 (e-book) | ISBN 9781315293769 (e-book) | ISBN 9781315293752 (e-book) | ISBN 9781315293745 (e-book)
Subjects: LCSH: Medical statistics. | Medicine–Research–Statistical methods. | Epidemiology–Statistical methods. | Statistical hypothesis testing.
Classification: LCC R853.S7 V48 2018 | DDC 610.2/1–dc23
LC record available at https://lccn.loc.gov/2017031841

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To my parents, Octyabrina and Alexander, and my son, David

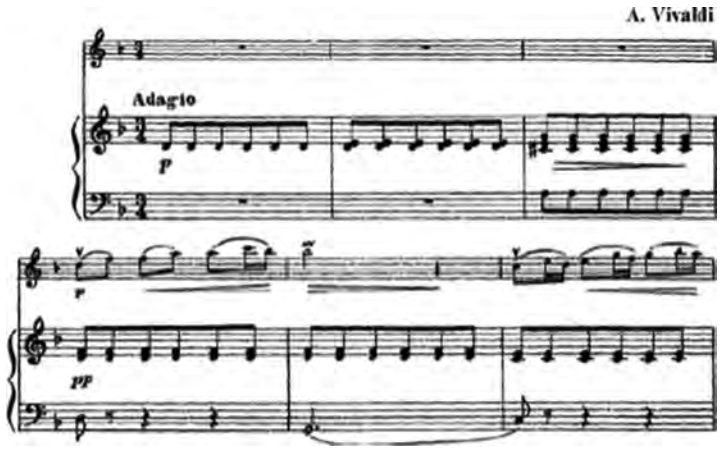
Albert Vexler

To my wife, Brenda, and three kids, Nick, Chance, and Trey

Alan D. Hutson

Table of Contents

Preface—Please Read!



In all likelihood, the Universe of Statistical Science is "relatively" or "privately" infinite and expanding in a similar manner to our Universe, satisfying the property of a science that is alive. Hence, it would be absolute impossibility to include everything in one book written by humans. One of our major goals is to provide the readers a small but efficient "probe" that could assist in Statistical Space discoveries.

The primary objective of the this book is to provide a compendium of statistical techniques ranging from classical methods through bootstrap strategies to modern recently developed statistical techniques. These methodologies may be applied to various problems encountered in health-related studies.

Historically, initial developments in statistical science were induced by real-life problems, when appropriate statistical instruments employed empirical arguments. Perhaps, since the eighteenth century, the heavy convolution between mathematics, probability theory, and statistical methods has provided the fundamental structures for correct statistical and biostatistical techniques. However, we cannot ignore a recent trend toward redundant simplifications of statistical considerations via so called "intuitive" and "applied" claims in complex statistical applications. It is our experience that applied statisticians or users often neglect the underlying postulates when implementing formal statistical procedures and with respect to the interpretation of their results. A very important motivation towards writing this book was to better refocus the scientist towards understanding the underpinnings of appropriate statistical inference in a well-rounded fashion. Maybe now is the time to draw more attention of theoretical and applied researchers to methodological standards in statistics and biostatistics? In contrast with

many biostatistical books, we focus on rigorous formal proof schemes and their extensions regarding different statistical principles. We also show the basic ingredients and methods for constructing and examining correct and powerful statistical processes.

The material in the book should be appropriate for use both as a text and as a reference. In our book readers can find classical and new theoretical methods, open problems, and new procedures across a variety of topics for their scholarly investigations. We present results that are novel to the current set of books on the market and results that are even new with respect to the modern scientific literature. Our aim is to draw the attention of theoretical statisticians and practitioners in epidemiology and/or clinical research to the necessity of new developments, extensions and investigations related to statistical methods and their applications. In this context, for example, we would like to emphasize for whom is interested in advanced topics the following aspects. Chapter 1 lays out a variety of notations, techniques and foundations basic to the material that is treated in this book. Chapter 2 introduces the powerful analytical instruments that, e.g., consist of principles of Tauberian theorems, including new results, with applications to convolution problems, evaluations of sequential procedures, renewal functions, and risk-efficient estimations. In this chapter we also consider problems of reconstructing the general distribution based on the distribution of some observed statistics. Chapter 3 shows certain nontrivial conclusions regarding the parametric likelihood ratios. Chapter 4 is developed to demonstrate a strong theoretical instrument based on martingales and their statistical applications, which include the martingale principle for testing statistical hypotheses and comparisons between the cumulative sum technique and the Shiryayev–Roberts approach employed in change point detection policies. A part of material shown in Chapter 4 can be found only in this book. Chapter 5 can assist the statistician in developing and analyzing various Bayesian procedures. Chapter 8 provides the fundamental components for constructing unconventional statistical decision-making procedures with power one. Chapter 9 proposes novel approaches to examine, compare, and visualize properties of various statistical tests using correct p-value-based mechanisms. Chapter 10 introduces the empirical likelihood methodology. The theoretical propositions shown in Chapter 10 can lead to a quite mechanical and simple way to investigate properties of nonparametric likelihood–based statistical schemes. Several of these results can be found only in this book. In Chapter 14 one can discover interesting open problems. This book also provides software code based on both the R and SAS statistical software packages to exemplify the statistical methodological topics and their applied aspects.

Indeed, we focused essentially on developing a very informative textbook that introduces classical and novel statistical methods with respect to various biostatistical applications. Towards this end we employ our experience and relevant material obtained via our research and teaching activity across

10–15 years of biostatistical practice and training Master and PhD level students in the department of biostatistics. This book is intended for graduate students majoring in statistics, biostatistics, epidemiology, health-related sciences, and/or in a field where a statistics concentration is desirable, particularly for those who are interested in formal statistical mechanisms and their evaluations. In this context Chapters 1–10 and 12–14 provides teaching sources for a high level statistical theory course that can be taught in a statistical/biostatistical department. The presented material evolved in conjunction with teaching such a one-semester course at The New York State University at Buffalo. This course, entitled "Theory of Statistical Inference," has belonged to a set of the four core courses required for our biostatistics PhD program.

This textbook delivers a "ready-to-go" well-structured product to be employed in developing advanced biostatistical courses. We offer lectures, homework questions, and their solutions, examples of midterm, final, and Ph.D. qualifying exams as well as examples of students' projects.

One of the ideas regarding this book's development is that we combine presentations of traditional applied and theoretical statistical methods with computationally extensive bootstrap type procedures that are relatively novel data-driven statistical tools. Chapter 11 is proposed to help instructors acquire a statistical course that introduces the Jackknife and Bootstrap methods. The focus is on the statistical functional as the key component of the theoretical developments with applied examples provided to illustrate the corresponding theory.

We strongly suggest to begin lectures by asking students to smile!

It is recommended to start each lecture class by answering students' inquiries regarding the previously assigned homework problems. In this course, we assume that students are encouraged to present their work in class regarding individually tailored research projects (e.g., Chapter 14). In this manner, the material of the course can be significantly extended.

Our intent is not that this book competes with classical fundamental guides such as, e.g., Bickel and Doksum (2007), Borovkov (1998), and Serfling (2002). In our course, we encourage scholars to read the essential works of the world-renown authors. We aim to present different theoretical approaches that are commonly used in modern statistics and biostatistics to (1) analyze properties of statistical mechanisms; (2) compare statistical procedures; and (3) develop efficient (optimal) statistical schemes. Our target is to provide scholars research seeds to spark new ideas. Towards this end, we demonstrate open problems, basic ingredients in learning complex statistical notations and tools as well as advanced nonconventional methods, even in simple cases of statistical operations, providing "**Warning**" remarks to show potential difficulties related to the issues that were discussed.

Finally, we would like to note that this book attempts to represent a part of our life that definitely consists of mistakes, stereotypes, puzzles, and so on, that we all love. Thus our book cannot be perfect. We truly thank the reader

for his/her participation in our life! We hope that the presented material can play a role as prior information for various research outputs.

Albert Vexler
Alan D. Hutson

Authors

Albert Vexler, PhD, obtained his PhD degree in Statistics and Probability Theory from the Hebrew University of Jerusalem in 2003. His PhD advisor was Moshe Pollak, a fellow of the *American Statistical Association* and Marcy Bogen Professor of Statistics at Hebrew University. Dr. Vexler was a postdoctoral research fellow in the Biometry and Mathematical Statistics Branch at the National Institute of Child Health and Human Development (National Institutes of Health). Currently, Dr. Vexler is a tenured Full Professor at the State University of New York at Buffalo, Department of Biostatistics. Dr. Vexler has authored and co-authored various publications that contribute to both the theoretical and applied aspects of statistics in medical research. Many of his papers and statistical software developments have appeared in statistical/biostatistical journals, which have top-rated impact factors and are historically recognized as the leading scientific journals, and include: *Biometrics, Biometrika, Journal of Statistical Software, The American Statistician, The Annals of Applied Statistics, Statistical Methods in Medical Research, Biostatistics, Journal of Computational Biology, Statistics in Medicine, Statistics and Computing, Computational Statistics and Data Analysis, Scandinavian Journal of Statistics, Biometrical Journal, Statistics in Biopharmaceutical Research, Stochastic Processes and their Applications, Journal of Statistical Planning and Inference, Annals of the Institute of Statistical Mathematics, The Canadian Journal of Statistics, Metrika, Statistics, Journal of Applied Statistics, Journal of Nonparametric Statistics, Communications in Statistics, Sequential Analysis, The STATA Journal; American Journal of Epidemiology, Epidemiology, Paediatric and Perinatal Epidemiology, Academic Radiology, The Journal of Clinical Endocrinology & Metabolism, Journal of Addiction Medicine, and Reproductive Toxicology and Human Reproduction.*

Dr. Vexler was awarded National Institutes of Health (NIH) grants to develop novel nonparametric data analysis and statistical methodology. His research interests are related to the following subjects: receiver operating characteristic curves analysis; measurement error; optimal designs; regression models; censored data; change point problems; sequential analysis; statistical epidemiology; Bayesian decision-making mechanisms; asymptotic methods of statistics; forecasting; sampling; optimal testing; nonparametric tests; empirical likelihoods, renewal theory; Tauberian theorems; time series; categorical analysis; multivariate analysis; multivariate testing of complex hypotheses; factor and principal component analysis; statistical biomarker evaluations; and best combinations of biomarkers. Dr. Vexler is Associate Editor for *Biometrics* and *Journal of Applied Statistics*. These journals belong to the first cohort of academic literature related to the methodology of biostatistical and epidemiological research and clinical trials.

Alan D. Hutson, PhD, received his BA (1988) and MA (1990) in Statistics from the State University of New York (SUNY) at Buffalo. He then worked for Otsuka America Pharmaceuticals for two years as a biostatistician. Dr. Hutson then received his MA (1993) and PhD (1996) in Statistics from the University of Rochester. His PhD advisor was Professor Govind Mudholkar, a world-renown researcher in Statistics and Biostatistics. He was hired as a biostatistician at the University of Florida in 1996 as a Research Assistant Professor and worked his way to a tenured Associate Professor. He had several roles at the University of Florida including Interim Director of the Division of Biostatistics and Director of the General Clinical Research Informatics Core. Dr. Hutson moved to the University at Buffalo in 2002 as an Associate Professor and Chief of the Division of Biostatistics. He was the founding chair of the new Department of Biostatistics in 2003 and became a full professor in 2007. His accomplishments as Chair included the implementation of several new undergraduate and graduate degree programs and a substantial growth in the size and quality of the department faculty and students. In 2005, Dr. Hutson also became Chair of Biostatistics (now Biostatistics and Bioinformatics) at Roswell Park Cancer Institute (RPCI), was appointed Professor of Oncology, and became the Director of the Core Cancer Center Biostatistics Core. Dr. Hutson helped implement the new Bioinformatics Core at RPCI. He recently became the Deputy Group Statistician for the NCI national NRG cancer cooperative group. Dr. Hutson is Fellow of *the American Statistical Association*. He is Associate Editor of *Communications in Statistics,* Associate Editor of the *Sri Lankan Journal of Applied Statistics*, and is a New York State NYSTAR Distinguished Professor. He has membership on several data safety and monitoring boards and has served on several high level scientific review panels. He has over 200 peer-reviewed publications. In 2013, Dr. Hutson was inducted into the Delta Omega Public Health Honor Society, Gamma Lambda Chapter. His methodological work focuses on nonparametric methods for biostatistical applications as it pertains to statistical functionals. He has several years of experience in the design and analysis of clinical trials.

Subject Index