Basic Biostatistics

Statistics for Public Health Practice

B. Burt Gerstman
Professor
Department of Health Science
San Jose State University
San Jose, California

JONES AND BARTLETT PUBLISHERS
Sudbury, Massachusetts
BOSTON TORONTO LONDON SINGAPORE

World Headquarters
Jones and Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jbpub.com
www.jbpub.com

Jones and Bartlett Publishers
International
Barb House, Barb Mews
London W6 7PA
United Kingdom

Jones and Bartlett Publishers
Canada
6339 Ormindale Way
Mississauga, Ontario L5V 1J2
Canada

Jones and Bartlett's books and products are available through most bookstores and online booksellers. To contact Jones and Bartlett Publishers directly, call 800-832-0034, fax 978-443-8000, or visit our website www.jbpub.com.

Substantial discounts on bulk quantities of Jones and Bartlett's publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones and Bartlett via the above contact information or send an email to specialsales@jbpub.com.

Production Credits
Publisher: Michael Brown
Associate Editor: Katey Birtcher
Production Director: Amy Rose
Production Editor: Tracey Chapman
Associate Production Editor: Rachel Rossi
Marketing Manager: Sophie Fleck
Manufacturing Buyer: Therese Connell

Composition: Graphic World, Inc.
Cover Design: Kristin E. Ohlin
Cover Image: © Sebastian Kaulitzki/
ShutterStock, Inc.; © Li Wa/ShutterStock, Inc.
Printing and Binding: Malloy, Inc
Cover Printing: Malloy, Inc

Library of Congress Cataloging-in-Publication Data
Gerstman, B. Burt.
Basic biostatistics : statistics for public health practice / B. Burt Gerstman.
p. ; cm.
Includes index.
ISBN-13: 978-0-7637-3580-7 (alk. paper)
ISBN-10: 0-7637-3580-9 (alk. paper)
1. Medical statistics. 2. Biometry. 3. Public health—Statistical methods. I. Title.
[DNLM: 1. Biometry—methods. 2. Public Health Practice. WA 950 G383b 2008]
RA409.G47 2008
362.1072'7—dc22

2007003334

6048
Printed in the United States of America
11 10 09 08 07 10 9 8 7 6 5 4 3 2 1

To my mother, Bernadine, and in memory of my father, Joseph.

Table of Contents

Preface

Basic Biostatistics is an introductory text that presents statistical ideas and techniques for students and workers in public health and biomedical practice and research. The book is designed to be accessible to students with modest mathematical backgrounds; no more than high school algebra is needed to understand this book. With this said, I hope to get past the notion that biostatistics is just an extension of math. Biostatistics is much more than that; it is a combination of mathematics and careful reasoning. Do not let the former interfere with the latter.

Biostatistical analysis is more than just number crunching; it considers how research questions are generated, studies are designed, data are collected, and results are interpreted.

> Analysis of data, with a more or less statistical flavor, should play many roles.[a]

Basic Biostatistics pays particular attention to exploratory and descriptive analyses. Whereas many introductory biostatistics texts give this topic intermittent attention, this text gives it ongoing consideration.

> Both exploratory and confirmatory data analysis deserves our attention.[b]

Biostatistics entails formulating research questions and designing processes for exploring and testing theories. I hope students who come to the study of biostatistics asking "What's the right answer?" leave asking questions like "Was that the right question?" and "Has the question been answered adequately?"

> Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise.[c]

---

[a] Tukey, J. W. (1980). We need both exploratory and confirmatory. *American Statistician, 34*(1), 23–25.
[b] Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist, 24*, 83.
[c] Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics, 33*(1), 13–14.

Several additional points bear emphasis:

Point 1: Practice, practice, practice. In studying biostatistics, you are developing a new set of reasoning skills. What is true of developing other skills is true of developing biostatistical skills—the only way to get better is to practice with the proper awareness and attention. To this end, illustrative examples and exercises are incorporated throughout the book. I've tried to make illustrations and exercises relevant. Many are contemporary, and many have historical importance. Carefully following the reasoning of illustrations and exercises is an important opportunity to learn. Answers to odd-numbered exercises are provided toward the back of the book. Qualified instructors may request answers to even-numbered exercises from the publisher.

Point 2: Structure of the book. The structure of this book may differ from that of other texts. Chapters are intentionally brief. They allow for flexibility in the order of coverage. The book is organized into three main parts. Part I (Chapters 1–10) addresses basic concepts and techniques. Students should complete these chapters (or a comparable introductory course) before moving on to Parts II and III.

Part II (Chapters 11–15) covers analytic techniques for quantitative responses. Part III (Chapters 16–19) covers techniques for categorical responses. Chapters in these sections can be covered in many different orders at the discretion of the instructor. One instructor may choose to cover these chapters in sequence, while another may cover Chapter 11 and Chapter 16 simultaneously (as an example), because these chapters both address one-sample problems. (Chapter 11 covers one-sample problems for quantitative responses; Chapter 16 covers one-sample problems for binary responses.) As another example, one could cover the chapters on categorical responses (Chapters 16–19) before covering the chapters on quantitative responses (Chapter 11–15).

Point 3: Hand calculations *and* computational support. While I believe there is still benefit in learning how to calculate statistics by hand, students are encouraged to use statistical software to supplement and check calculations. Use of the proper software tools can free us from some of the tedium of numerical manipulations, leaving more time to step back and think about practical implications of results.

The only way humans can do BETTER than computers is to take a chance of doing WORSE. So we have got to take seriously the need for steady progress toward

teaching routine procedures to computers rather than to people. That will leave the teachers of people with only things hard to teach, but this is our proper fate.[d]

The book is not tied to any particular software package, but does make frequent use of these three programs: *StaTable*, *SPSS,* and *WinPepi*.

- *StaTable*[e] is a freeware program that provides access to 25 commonly used statistical distributions. It is runs on Windows, Palm, and Web-browser (Java) platforms. This utility eliminates the need to look up probabilities in hard-copy tables. It also allows for more exact interpolations for probabilities, especially for continuous random variables. The website for this book includess a link to the *StaTable* website.
- *SPSS*[f] is a commercial software package with versions that run on Windows and MacIntosh computers. A student version of the program can be purchased at most campus bookstores. It can also be purchased online at www.journeyed.com. An economical alternative to purchasing the product is to lease it for short-term use through the Web site www.e-academy.com.
- *WinPepi*[g] stands for **WIN**dows **P**rograms for **EPI**demiologists. This is a series of computer programs written by Joe Abramson of the Hebrew University–Hadassah School of Public Health and Community Medicine, (Jerusalem, Israel) and Paul Gahlinger (University of Utah in Salt Lake City). The programs are designed for use in practice, but are also excellent learning aids. *WinPepi* is free and can be downloaded from the website for this book: http://publichealth.jbpub.com/book/gerstman.

---

[d] Tukey, J. W. (1980). We need both exploratory and confirmatory. *American Statistician, 34,* 23–25.

[e] www.cytel.com/Products/StaTable/, Cytel Inc., 675 Massachusetts Ave., Cambridge, Massachusetts 02139.

[f] SPSS, Inc., Chicago, IL.

[g] Abramson, J. H. (2004). *WINPEPI* (PEPI-for-Windows): Computer programs for epidemiologists. *Epidemiologic Perspectives & Innovations, 1*(1), 6.

Acknowledgments

I wish to express my appreciation to San Jose State University for affording me the leave to work on this book. I would especially like to thank the chair of my department, Kathleen Roe, and dean of my college, Inger Sagatun-Edwards, for administrative support in this regard. I am grateful to the colleagues in my department who taught many of my classes during my absence, especially Jane Pham, Dan Perales, Jenny McNeill, and to those who covered other duties, including Ramani Rangavajhula, Nancy Hikoyeda, Polly Bith-Melander, and Edward Mamary. I greatly appreciate the artistic and technical support of Jean Shiota of the Center for Faculty Development for her work in preparing illustrations for the text. Thanks, Jean. Finally, I wish to express my thanks to those many students in my classes over the years who have provided me with helpful comments, encouragement, and camaraderie.

While writing this book, I had many constructive discussions with Joe Abramson of the Department of Social Medicine, Hebrew University–Hadassah School of Public Health and Community Medicine. I thank Joe for sharing his insights generously. I also greatly appreciate his careful work in developing WINdows Programs for EPIdemiologists.[i] This is really an exceptional set of programs for public health workers. Along these same lines, Paul Gahlinger (University of Utah) deserves credit for conceiving and creating the progenitor of *WinPepi*, *PEPI* (Programs for EPIdemiologists).[j] I also wish to express my thanks to Mads Haahr (University of Dublin, Trinity College, Ireland) for creating his true random number generator at www.random.org and to John C. Pezzullo

[i] Abramson, J. H. (2004). *WINPEPI* (PEPI-for-Windows): Computer programs for epidemiologists. *Epidemiologic Perspectives & Innovations, 1*(1), 6.

[j] Abramson, J. H., & Gahlinger, P. M. (2001). Computer Programs for Epidemiologic Analyses: PEPI v. 4.0. Salt Lake City, UT: Sagebrush Press.

(Georgetown University) for his helpful compilation of web pages that perform statistical calculations at www.statpages.org.

Finally, I would like to acknowledge the contributions of my wife, who has been patient, understanding, supportative, and encouraging throughout the work on this marathon project. As Ralph Kramden (Jackie Gleason) used to tell his wife Alice (Audrey Meadows), "[Honey], you're the greatest!"

About the Author

Dr. Gerstman did his undergraduate work at Harpur College (State University of New York, Binghamton). He later received a doctor of veterinary medicine (Cornell University), a masters of public health (University of California at Berkeley), and a doctor of philosophy degree (University of California, Davis). He has been a U.S. Public Health Service Epidemiology Fellow and epidemiologist at the U.S. Food and Drug Administration and was an instructor at the National Institutes of Health Foundation Graduate School. Since 1990, Dr. Gerstman has been a professor in the Department of Health Science at San Jose State University where he teaches epidemiology, biostatistics, and general education courses. Dr. Gerstman's research interests are in the areas of epidemiologic methods, the history of public health, drug safety, and medical and public health record linkage.