



BIostatISTICS

A METHODOLOGY FOR
THE HEALTH SCIENCES

Lloyd D. Fisher
Gerald van Belle

Biostatistics

A Methodology for the Health Sciences

Second Edition

GERALD VAN BELLE

LLOYD D. FISHER

PATRICK J. HEAGERTY

THOMAS LUMLEY

Department of Biostatistics and
Department of Environmental and
Occupational Health Sciences
University of Washington
Seattle, Washington



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2004 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 978-646-8600, 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.

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 please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. 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, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Biostatistics: a methodology for the health sciences / Gerald van Belle . . . [et al.]— 2nd ed.

p. cm. — (Wiley series in probability and statistics)

First ed. published in 1993, entered under Fisher, Lloyd.

Includes bibliographical references and index.

ISBN 0-471-03185-2 (cloth)

I. Biometry. I. Van Belle, Gerald. II. Fisher, Lloyd, 1939— Biostatistics. III. Series.

QH323.5.B562 2004

610'.1'5195—dc22

2004040491

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Ad majorem Dei gloriam

Contents

Preface to the First Edition	ix
Preface to the Second Edition	xi
1. Introduction to Biostatistics	1
2. Biostatistical Design of Medical Studies	10
3. Descriptive Statistics	25
4. Statistical Inference: Populations and Samples	61
5. One- and Two-Sample Inference	117
6. Counting Data	151
7. Categorical Data: Contingency Tables	208
8. Nonparametric, Distribution-Free, and Permutation Models: Robust Procedures	253
9. Association and Prediction: Linear Models with One Predictor Variable	291
10. Analysis of Variance	357
11. Association and Prediction: Multiple Regression Analysis and Linear Models with Multiple Predictor Variables	428
12. Multiple Comparisons	520
13. Discrimination and Classification	550
14. Principal Component Analysis and Factor Analysis	584

15. Rates and Proportions	640
16. Analysis of the Time to an Event: Survival Analysis	661
17. Sample Sizes for Observational Studies	709
18. Longitudinal Data Analysis	728
19. Randomized Clinical Trials	766
20. Personal Postscript	787
Appendix	817
Author Index	841
Subject Index	851
Symbol Index	867

Preface to the First Edition

The purpose of this book is for readers to learn how to apply statistical methods to the biomedical sciences. The book is written so that those with no prior training in statistics and a mathematical knowledge through algebra can follow the text—although the more mathematical training one has, the easier the learning. The book is written for people in a wide variety of biomedical fields, including (alphabetically) biologists, biostatisticians, dentists, epidemiologists, health services researchers, health administrators, nurses, and physicians. The text appears to have a daunting amount of material. Indeed, there is a great deal of material, but most students will not cover it all. Also, over 30% of the text is devoted to notes, problems, and references, so that there is not as much material as there seems to be at first sight. In addition to not covering entire chapters, the following are optional materials: asterisks (*) preceding a section number or problem denote more advanced material that the instructor may want to skip; the notes at the end of each chapter contain material for extending and enriching the primary material of the chapter, but this may be skipped.

Although the order of authorship may appear alphabetical, in fact it is random (we tossed a fair coin to determine the sequence) and the book is an equal collaborative effort of the authors. We have many people to thank. Our families have been helpful and long-suffering during the writing of the book: for LF, Ginny, Brad, and Laura; for GvB, Johanna, Loeske, William John, Gerard, Christine, Louis, and Bud and Stacy. The many students who were taught with various versions of portions of this material were very helpful. We are also grateful to the many collaborating investigators, who taught us much about science as well as the joys of collaborative research. Among those deserving thanks are for LF: Ed Alderman, Christer Allgulander, Fred Applebaum, Michele Battie, Tom Bigger, Stan Bigos, Jeff Borer, Martial Bourassa, Raleigh Bowden, Bob Bruce, Bernie Chaitman, Reg Clift, Rollie Dickson, Kris Doney, Eric Foster, Bob Frye, Bernard Gersh, Karl Hammermeister, Dave Holmes, Mel Judkins, George Kaiser, Ward Kennedy, Tom Killip, Ray Lipicky, Paul Martin, George McDonald, Joel Meyers, Bill Myers, Michael Mock, Gene Passamani, Don Peterson, Bill Rogers, Tom Ryan, Jean Sanders, Lester Sauvage, Rainer Storb, Keith Sullivan, Bob Temple, Don Thomas, Don Weiner, Bob Witherspoon, and a large number of others. For GvB: Ralph Bradley, Richard Cornell, Polly Feigl, Pat Friel, Al Heyman, Myles Hollander, Jim Hughes, Dave Kalman, Jane Koenig, Tom Koepsell, Bud Kukull, Eric Larson, Will Longstreth, Dave Luthy, Lorene Nelson, Don Martin, Duane Meeter, Gil Omenn, Don Peterson, Gordon Pledger, Richard Savage, Kirk Shy, Nancy Temkin, and many others. In addition, GvB acknowledges the secretarial and moral support of Sue Goleeke. There were many excellent and able typists over the years; special thanks to Myrna Kramer, Pat Coley, and Jan Alcorn. We owe special thanks to Amy Plummer for superb work in tracking down authors and publishers for permission to cite their work. We thank Robert Fisher for help with numerous figures. Rob Christ did an excellent job of using \LaTeX for the final version of the text. Finally, several people assisted with running particular examples and creating the tables; we thank Barry Storer, Margie Jones, and Gary Schoch.

Our initial contact with Wiley was the indefatigable Beatrice Shube. Her enthusiasm for our effort carried over to her successor, Kate Roach. The associate managing editor, Rose Ann Campise, was of great help during the final preparation of this manuscript.

With a work this size there are bound to be some errors, inaccuracies, and ambiguous statements. We would appreciate receiving your comments. We have set up a special electronic-mail account for your feedback:

<http://www.biostat-text.info>

LLOYD D. FISHER
GERALD VAN BELLE

Preface to the Second Edition

Biostatistics did not spring fully formed from the brow of R. A. Fisher, but evolved over many years. This process is continuing, although it may not be obvious from the outside. It has been ten years since the first edition of this book appeared (and rather longer since it was begun). Over this time, new areas of biostatistics have been developed and emphases and interpretations have changed.

The original authors, faced with the daunting task of updating a 1000-page text, decided to invite two colleagues to take the lead in this task. These colleagues, experts in longitudinal data analysis, survival analysis, computing, and all things modern and statistical, have given a twenty-first-century thrust to the book.

The author sequence for the first edition was determined by the toss of a coin (see the Preface to the First Edition). For the second edition it was decided to switch the sequence of the first two authors and add the new authors in alphabetical sequence.

This second edition adds a chapter on randomized trials and another on longitudinal data analysis. Substantial changes have been made in discussing robust statistics, model building, survival analysis, and discrimination. Notes have been added, throughout, and many graphs redrawn. We have tried to eliminate errata found in the first edition, and while more have undoubtedly been added, we hope there has been a net improvement. When you find mistakes we would appreciate hearing about them at <http://www.vanbelle.org/biostatistics/>.

Another major change over the past decade or so has been technological. Statistical software and the computers to run it have become much more widely available—many of the graphs and new analyses in this book were produced on a laptop that weighs only slightly more than a copy of the first edition—and the Internet provides ready access to information that used to be available only in university libraries. In order to accommodate the new sections and to attempt to keep up with future changes, we have shifted some material to a set of Web appendices. These may be found at <http://www.biostat-text.info>. The Web appendices include notes, data sets and sample analyses, links to other online resources, all but a bare minimum of the statistical tables from the first edition, and other material for which ink on paper is a less suitable medium.

These advances in technology have not solved the problem of deadlines, and we would particularly like to thank Steve Quigley at Wiley for his equanimity in the face of schedule slippage.

GERALD VAN BELLE
LLOYD FISHER
PATRICK HEAGERTY
THOMAS LUMLEY

Seattle, June 15, 2003