Multivariable Analysis

A Practical Guide for Clinicians and Public Health Researchers

Why do you need this book?

Multivariable analysis is confusing! Whether you are performing your first research project or attempting to interpret the output from a multivariable model, you have undoubtedly found this to be true. Basic biostatistics books are of little or no help to you, since their coverage often stops short of multivariable analysis. However, existing multivariable analysis books are too dense with mathematical formulae and derivations and are not designed to answer your most basic questions. Is there a book that steps aside from the math and simply explains how to understand, perform, and interpret multivariable analyses?

Yes. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers*, as this new edition is titled, is precisely the reference that will lead your way. In fact, Dr. Mitchell Katz has asked and answered all of your questions for you!

Why should I do multivariable analysis?

How do I choose which type of multivariable to use?

How many subjects do I need to do multivariable analysis?

What if I have repeated observations of the same persons?

Answers and detailed explanations to these questions and more are found in this book. Also, it is loaded with useful tips, summary charts, figures, and references.

If you are a medical student, resident, or clinician, *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers* will prove an indispensable guide through the confusing terrain of statistical analysis.

This third edition has been fully revised to build on the enormous success of its predecessors. New features include new sections on Poisson and negative binomial regression, proportional odds analysis, and multinomial logistic regression, and an expanded section on interpretation of residuals.

Praise for first edition

"This is the first nonmathematical book on multivariable analysis addressed to clinicians. Its range, organization, brevity, and clarity make it useful as a reference, a text, and a guide for self-study. This book *is* 'a practical guide for clinicians.'"

Leonard E. Braitman, Ph.D., *Annals of Internal Medicine*

Mitchell H. Katz is Clinical Professor of Medicine, Epidemiology and Biostatistics at the University of California, San Francisco; and Director of the Los Angeles Department of Health Services, Los Angeles, USA.

Multivariable Analysis

A Practical Guide for Clinicians and Public Health Researchers

Third Edition

Mitchell H. Katz

Department of Medicine, Epidemiology and Biostatistics, University of California, USA

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521760980

First published 1999
Second edition published 2006
Third edition published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Katz, Mitchell H., 1959– author.
Multivariable analysis : a practical guide for clinicians and public health researchers /
Mitchell H. Katz, Department of Medicine, Epidemiology, and Biostatistics,
University of California, USA. – 3rd Edition.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-521-76098-0 (hardback) – ISBN 978-0-521-14107-9 (paperback)
1. Medicine–Research–Statistical methods. 2. Multivariate analysis. 3. Biometry.
4. Medical statistics. I. Title.
[DNLM: 1. Multivariate Analysis. 2. Biometry–methods. WA 950]
R853.S7K38 2011
610.72–dc22 2010052187

ISBN 978-0-521-76098-0 Hardback
ISBN 978-0-521-14107-9 Paperback

Every effort has been made in preparing this book to provide accurate and up-to-date information which is in accord with accepted standards and practice at the time of publication. Although case histories are drawn from actual cases, every effort has been made to disguise the identities of the individuals involved. Nevertheless, the authors, editors and publishers can make no warranties that the information contained herein is totally free from error, not least because clinical standards are constantly changing through research and regulation. The authors, editors and publishers therefore disclaim all liability for direct or consequential damages resulting from the use of material contained in this book. Readers are strongly advised to pay careful attention to information provided by the manufacturer of any drugs or equipment that they plan to use.

To my parents, for their unwavering support

Contents

Preface

There has been astounding growth in the use of multivariable analysis in clinical research. When the first edition of this book was published in 1999 logistic regression and proportional hazards models were cutting-edge techniques. Now for many researchers, these are old, staid models and the new edge is mixed-effects models, generalized estimating equations, Poisson regression, and propensity score analysis.

The use of these more sophisticated models is fueled by the development of user-friendly software for constructing multivariable models, increased availability of electronic databases (medical records, disease and procedure registries) that provide longitudinal data on large populations, and increased funding for and interest in clinical effectiveness studies – studies comparing different treatments in use – as a method of improving quality and reducing healthcare costs.

What hasn't changed in the past 11 years is the need for an easy-to-follow guide for nonstatisticians on how to perform and interpret these models. Although the available software (e.g., SPSS, SAS, S-plus, R) doesn't require programming experience or mathematical aptitude to conduct the analyses, if the analysis is not set up correctly, the answer is sure to be wrong! Even when the analysis is performed correctly, researchers may not draw the correct conclusions from the output.

To prevent these problems, throughout the book I have focused on how to set up and interpret multivariable analysis. I use examples from the medical and public health literature because illustrations of how to correctly analyze data and present the results will help you analyze and present your data correctly. Modeling your work based on successful published studies is one of the best and most efficient strategies for correctly analyzing data.

The biggest changes in this edition are that I have written new sections on Poisson and negative binomial regression, proportional odds analysis, and multinomial logistic regression because these models are increasingly in use. I have improved the section on mixed-effects models and generalized

estimating equations, and also expanded the section on checking the underlying assumptions of multivariable models (Chapter 9) using residuals and other techniques.

While taking on new and more complicated material, I have maintained the basic organization of the book. Besides retaining the question-and-answer approach, the order of the book mirrors the process of doing multivariable analysis: deciding whether you need to do multivariable analysis (Chapters 1 and 2), choosing the correct model (Chapter 3), preparing your independent variables (Chapters 4 and 5), setting up the model (Chapter 6), performing the analysis (Chapter 7), interpreting the basic output (Chapter 8), delving deeper into the underlying assumptions of the model (Chapter 9), validating your model (Chapter 12) and publishing your study (Chapter 14). One of the reasons I prefer this approach to the more traditional approach (i.e., having a separate chapter on each type of multivariable model) is that it illustrates the similarities and differences of the different approaches. In my experience, when the results are strong, different (but reasonable) approaches lead to similar answers; conversely, when the results are very different with different techniques be suspicious. Also, I have found that the most efficient way to end an argument over what the best way is to analyze a data set is to analyze it multiple ways and see whether the results differ. If there are few differences then you have strengthened your results. When there are differences, you have probably learned something important about the nature of your data. Also, by structuring the book to parallel the research process, it allows readers to join the book at whatever stage they are at in the research process.

This book assumes that you are familiar with basic biostatistics. If not, I recommend S. Glantz's *Primer of Biostatistics* (sixth edition, McGraw-Hill, 2005). I have also written a basic statistics book using a question-and-answer approach similar to that used in this book called *Study Design and Statistical Analysis: A Practical Guide for Clinicians* (Cambridge University Press, 2006). Some reviewers have suggested that the two books be combined, and while I see the merit in that, I also see a much fatter text that might be more expensive and off-putting to clinical researchers. Please forgive me therefore if I cite that book or my other book on performing interventions (*Evaluating Clinical and Public Health Interventions*, Cambridge University Press, 2010). It is not an exercise of ego, but rather an attempt to keep each book inexpensive and short.

One of the challenges in writing a book for clinical researchers is deciding how much detail to include. One could easily have (and many have) written books larger than this about just one of the procedures described. To keep the presentations short and the material accessible, I direct readers who wish to know more about a particular procedure to more detailed sources in the

footnotes. Since statistical textbooks are expensive, and many journal articles are not easy to find, I have particularly emphasized web resources that I have found useful.

Twenty years of students in the University of California, San Francisco, Clinical Research Program have contributed to this book through their insightful questions and observations. Serving as the Deputy Editor for the *Archives of Internal Medicine* during the past two years has definitely sharpened my eye as to how best to conduct multivariable research. For this opportunity I am grateful to the Editor, Rita Redberg, M.D., our two biostatistical editors who have taught me much, John Neuhaus, Ph.D. and David Glidden, Ph.D., and the other editors, Patrick O'Malley, M.D. and Kirsten Johansen, M.D., who have shared their critical observations with me on hundreds of articles. I greatly appreciate the support of my editor Richard Marley and the staff at Cambridge University Press for encouraging me to do this third edition.

The best part of writing and updating this book is the number of researchers who have emailed me with their comments, compliments, and questions. Writing textbooks is a lonely business and I wouldn't do it unless I had evidence that the books were actually helping people to conduct better research. If you have questions or suggestions for future editions, email me at mhkatz59@yahoo.com