

Statistics for Biology and Health

Series Editors

K. Dietz, M. Gail, K. Krickeberg, A. Tsiatis, J. Samet

Springer

New York

Berlin

Heidelberg

Hong Kong

London

Milan

Paris

Tokyo

Statistics for Biology and Health

Everitt/Rabe-Hesketh: Analyzing Medical Data Using S-PLUS

Ewens/Grant: Statistical Methods in Bioinformatics: An Introduction.

Hougaard: Analysis of Multivariate Survival Data.

Klein/Moeschberger: Survival Analysis: Techniques for Censored and Truncated Data.

Kleinbaum: Logistic Regression: A Self-Learning Text.

Kleinbaum/Klein: Logistic Regression: A Self-Learning Text, 2nd Ed.

Kleinbaum: Survival Analysis: A Self-Learning Text.

Lange: Mathematical and Statistical Methods for Genetic Analysis.

Manton/Singer/Suzman: Forecasting the Health of Elderly Populations.

Salsburg: The Use of Restricted Significance Tests in Clinical Trials.

Therneau/Grambsch: Modeling Survival Data: Extending the Cox Model.

Zhang/Singer: Recursive Partitioning in the Health Sciences

David G. Kleinbaum Mitchel Klein

Logistic Regression

A Self-Learning Text

Second Edition

With Contributions by
Erica Rihl Pryor



Springer

David G. Kleinbaum
Mitchel Klein
Department of Epidemiology
Emory University
Atlanta, GA 30333
USA

Series Editors

K. Dietz
Institut für Medizinische
Biometrie
Universität Tübingen
Westbahnhofstrasse 55
D-72070 Tübingen
Germany

M. Gail
National Cancer Institute
Rockville, MD 20892
USA

K. Krickeberg
Le Chatelet
F-63270 Manglieu
France

J. Samet
School of Public Health
Department of Epidemiology
Johns Hopkins University
615 Wolfe Street
Baltimore, MD 21205-2103
USA

A. Tsiatis
Department of Statistics
North Carolina State University
Raleigh, NC 27695
USA

Library of Congress Cataloging-in-Publication Data
Kleinbaum, David G.

Logistic regression: a self-learning text / David G. Kleinbaum,
Mitchel Klein.—2nd ed.

p. ; cm.—(Statistics for biology and health)

Includes bibliographical references and index.

ISBN 0-387-95397-3 (hc : alk. paper)

1. Medicine—Research—Statistical methods. 2. Regression
analysis. 3. Logistic distribution. I. Klein, Mitchel. II. Title.
III. Series

R853.S7 K54 2002

610'.7'27—dc21

2002019728

ISBN 0-387-95397-3

Printed on acid-free paper.

© 2002, 1994 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

SPIN 10858950

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

To John R. Boring III

Preface

This is the second edition of this text on logistic regression methods, originally published in 1994.

As in the first edition, each chapter contains a presentation of its topic in “lecture-book” format together with objectives, an outline, key formulae, practice exercises, and a test. The “lecture-book” has a sequence of illustrations and formulae in the left column of each page and a script (i.e., text) in the right column. This format allows you to read the script in conjunction with the illustrations and formulae that highlight the main points, formulae, or examples being presented.

This second edition has expanded the first edition by adding five new chapters and a new appendix. The five new chapters are

- Chapter 9. Polytomous Logistic Regression
- Chapter 10. Ordinal Logistic Regression
- Chapter 11. Logistic Regression for Correlated Data: GEE
- Chapter 12. GEE Examples
- Chapter 13. Other Approaches for Analysis of Correlated Data

Chapters 9 and 10 extend logistic regression to response variables that have more than two categories. Chapters 11–13 extend logistic regression to generalized estimating equations (GEE) and other methods for analyzing correlated response data.

The appendix is titled “Computer Programs for Logistic Regression” and provides descriptions and examples of computer programs for carrying out the variety of logistic regression procedures described in the main text. The software packages considered are SAS Version 8.0, SPSS Version 10.0, and STATA Version 7.0.

Also, Chapter 8 on the Analysis of Matched Data Using Logistic Regression has been expanded to include a discussion of three issues:

- Assessing interaction involving the matching variables
- Pooling exchangeable matched sets
- Analysis of matched follow-up data

Suggestions for Use

This text was originally intended for self-study, but in the eight years since the first edition was published it has also been effectively used as a text in a standard lecture-type classroom format. The text may be used to supplement material covered in a course or to review previously learned material in a self-instructional course or self-planned learning activity. A more individualized learning program may be particularly suitable to a working professional who does not have the time to participate in a regularly scheduled course.

The order of the chapters represents what the authors consider to be the logical order for learning about logistic regression. However, persons with some knowledge of the subject can choose whichever chapter appears appropriate to their learning needs in whatever sequence desired.

The last three chapters on methods for analyzing correlated data are somewhat more mathematically challenging than the earlier chapters, but have been written to logically follow the preceding material and to highlight the principal features of the methods described rather than to give a detailed mathematical formulation.

In working with any chapter, the user is encouraged first to read the abbreviated outline and the objectives and then work through the presentation. After finishing the presentation, the user is encouraged to read the detailed outline for a summary of the presentation, review key formula and other important information, work through the practice exercises, and, finally, complete the test to check what has been learned.

Recommended Preparation

The ideal preparation for this text is a course on quantitative methods in epidemiology and a course in applied multiple regression. The following are recommended references on these subjects, with suggested chapter readings:

Kleinbaum, D., Kupper, L., and Morgenstern, H., *Epidemiologic Research: Principles and Quantitative Methods*, John Wiley and Sons Publishers, New York, 1982, Chaps. 1–19.

Kleinbaum, D., Kupper, L., Muller, K., and Nizam, A., *Applied Regression Analysis and Other Multivariable Methods, Third Edition*, Duxbury Press, Pacific Grove, 1998, Chaps. 1–16.

A first course on the principles of epidemiologic research would be helpful as all modules in this series are written from the perspective of epidemiologic research. In particular, the learner should be familiar with the basic characteristics of epidemiologic study designs (follow-up, case control, and cross sectional) and should have some idea of the frequently encountered problem of controlling or adjusting for variables.

As for mathematics prerequisites, the reader should be familiar with natural logarithms and their relationship to exponentials (powers of e) and, more generally, should be able to read mathematical notation and formulae.

Atlanta, Georgia

David G. Kleinbaum
Mitchel Klein

Acknowledgments

David Kleinbaum and Mitch Klein wish to thank Erica Pryor at the School of Nursing, University of Alabama-Birmingham, for her many important contributions to this second edition. This includes fine-tuning the content of the five new chapters and appendix that have been added to the previous edition, taking primary responsibility for the pictures, formulae, symbols and summary information presented on the left side of the pages in each new chapter, performing a computer analysis of datasets described in the text, and carefully editing and correcting errata in the first eight chapters as well as the new appendix on computer software procedures.

All three of us (David, Mitch, and Erica) wish to thank John Boring, Chair of the Department of Epidemiology at the Rollins School of Public Health at Emory University for his leadership, inspiration, support, guidance, and friendship over the past several years. In appreciation, we are dedicating this edition to him.

Atlanta, GA

Birmingham, AL

David G. Kleinbaum

Mitchel Klein

Erica Rihl Pryor

Contents

Preface vii

Acknowledgments ix

Chapter 1

Introduction to Logistic Regression 1

Introduction 2

Abbreviated Outline 2

Objectives 3

Presentation 4

Detailed Outline 29

Key Formulae 31

Practice Exercises 32

Test 34

Answers to Practice Exercises 37

Chapter 2

Important Special Cases of the Logistic Model 39

Introduction 40

Abbreviated Outline 40

Objectives 40

Presentation 42

Detailed Outline 65

Practice Exercises 67

Test 69

Answers to Practice Exercises 71

Chapter 3

Computing the Odds Ratio in Logistic Regression 73

Introduction 74

Abbreviated Outline 74

Objectives 75

Presentation 76

Detailed Outline 92

Practice Exercises 95

Test 97

Answers to Practice Exercises 99

Chapter 4

Maximum Likelihood Techniques: An Overview 101

Introduction 102

Abbreviated Outline 102

Objectives 103

Presentation	104
Detailed Outline	120
Practice Exercises	121
Test	122
Answers to Practice Exercises	124

Chapter 5 **Statistical Inferences Using Maximum Likelihood Techniques 125**

Introduction	126
Abbreviated Outline	126
Objectives	127
Presentation	128
Detailed Outline	150
Practice Exercises	152
Test	156
Answers to Practice Exercises	158

Chapter 6 **Modeling Strategy Guidelines 161**

Introduction	162
Abbreviated Outline	162
Objectives	163
Presentation	164
Detailed Outline	183
Practice Exercises	184
Test	186
Answers to Practice Exercises	188

Chapter 7 **Modeling Strategy for Assessing Interaction and Confounding 191**

Introduction	192
Abbreviated Outline	192
Objectives	193
Presentation	194
Detailed Outline	221
Practice Exercises	222
Test	224
Answers to Practice Exercises	225

Chapter 8 Analysis of Matched Data Using Logistic Regression 227

Introduction 228
Abbreviated Outline 228
Objectives 229
Presentation 230
Detailed Outline 243
Practice Exercises 245
Test 247
Answers to Practice Exercises 249

Chapter 9 Polytomous Logistic Regression 267

Introduction 268
Abbreviated Outline 268
Objectives 269
Presentation 270
Detailed Outline 293
Practice Exercises 295
Test 297
Answers to Practice Exercises 298

Chapter 10 Ordinal Logistic Regression 301

Introduction 302
Abbreviated Outline 302
Objectives 303
Presentation 304
Detailed Outline 320
Practice Exercises 322
Test 324
Answers to Practice Exercises 325

Chapter 11 Logistic Regression for Correlated Data: GEE 327

Introduction 328
Abbreviated Outline 328
Objectives 329
Presentation 330
Detailed Outline 367
Practice Exercises 373
Test 374
Answers to Practice Exercises 375

Chapter 12

GEE Examples 377

Introduction	378
Abbreviated Outline	378
Objectives	379
Presentation	380
Detailed Outline	396
Practice Exercises	397
Test	400
Answers to Practice Exercises	402

Chapter 13

Other Approaches for Analysis of Correlated Data 405

Introduction	406
Abbreviated Outline	406
Objectives	407
Presentation	408
Detailed Outline	427
Practice Exercises	429
Test	433
Answers to Practice Exercises	435

Appendix: Computer Programs for Logistic Regression 437

Datasets	438
SAS	441
SPSS	464
Stata	474

Test Answers 487

Bibliography 503

Index 507