This is page
Printer: Opaque this

Log-Linear Models and Logistic Regression

Ronald Christensen
Department of Mathematics and Statistics
University of New Mexico

To Sharon and Fletch

Preface to the Second Edition

As the new title indicates, this second edition of *Log-Linear Models* has been modified to place greater emphasis on logistic regression. In addition to new material, the book has been radically rearranged. The fundamental material is contained in Chapters 1-4. Intermediate topics are presented in Chapters 5 through 8. Generalized linear models are presented in Chapter 9. The matrix approach to log-linear models and logistic regression is presented in Chapters 10-12, with Chapters 10 and 11 at the applied Ph.D. level and Chapter 12 doing theory at the Ph.D. level.

The largest single addition to the book is Chapter 13 on Bayesian binomial regression. This chapter includes not only logistic regression but also probit and complementary log-log regression. With the simplicity of the Bayesian approach and the ability to do (almost) exact small sample statistical inference, I personally find it hard to justify doing traditional large sample inferences. (Another possibility is to do exact conditional inference, but that is another story.)

Naturally, I have cleaned up the minor flaws in the text that I have found. All examples, theorems, proofs, lemmas, etc. are numbered consecutively within each section with no distinctions between them, thus Example 2.3.1 will come before Proposition 2.3.2. Exercises that do not appear in a section at the end have a separate numbering scheme. Within the section in which it appears, an equation is numbered with a single value, e.g., equation (1). When reference is made to an equation that appears in a different section, the reference includes the appropriate chapter and section, e.g., equation (2.1.1).

The primary prerequisite for using this book is knowledge of analysis

of variance and regression at the masters degree level. It would also be advantageous to have some prior familiarity with the analysis of two-way tables of count data. Christensen (1996a) was written with the idea of preparing people for this book and for Christensen (1996b). In addition, familiarity with masters level probability and mathematical statistics would be helpful, especially for the later chapters. Sections 9.3, 10.2, 11.6, and 12.3 use ideas of the convergence of random variables. Chapter 12 was originally the last chapter in my linear models book, so I would recommend a good course in linear models before attempting that. A good course in linear models would also help for Chapters 10 and 11.

The analysis of logistic regression and log-linear models is not possible without modern computing. While it certainly is not the goal of this book to provide training in the use of various software packages, some examples of software commands have been included. These focus primarily on SAS and BMDP, but include some GLIM (of which I am still very fond).

I would particularly like to thank Ed Bedrick for his help in preparing this edition and Ed and Wes Johnson for our collaboration in developing the material in Chapter 13. I would also like to thank Turner Ostler for providing the trauma data and his prior opinions about it.

Most of the data, and all of the larger data sets, are available from STATLIB as well as by anonymous ftp. The web address for the datasets option in STATLIB is **http://www.stat.cmu.edu/datasets/**. The data are identified as "christensen-llm". To use ftp, type **ftp stat.unm.edu** and login as "anonymous", enter **cd /pub/fletcher** and either get **llm.tar.Z** for Unix machines or **llm.zip** for a DOS version. More information is available from the file "readme.llm" or at **http://stat.unm.edu/~fletcher**, my web homepage.

Ronald Christensen
Albuquerque, New Mexico
February, 1997

BMDP Statistical Software is distributed by SPSS Inc., 444 N. Michigan Avenue, Chicago, IL, 60611, telephone: (800) 543-2185.

MINITAB is a registered trademark of Minitab, Inc., 3081 Enterprise Drive, State College, PA 16801, telephone: (814) 238-3280, telex: 881612.

MSUSTAT is marketed by the Research and Development Institute Inc., Montana State University, Bozeman, MT 59717-0002, Attn: R.E. Lund.

Preface to the First Edition

This book examines log-linear models for contingency tables. Logistic regression and logistic discrimination are treated as special cases and generalized linear models (in the GLIM sense) are also discussed. The book is designed to fill a niche between basic introductory books such as Fienberg (1980) and Everitt (1977) and advanced books such as Bishop, Fienberg, and Holland (1975), Haberman (1974a), and Santner and Duffy (1989). It is primarily directed at advanced Masters degree students in Statistics but it can be used at both higher and lower levels. The primary theme of the book is using previous knowledge of analysis of variance and regression to motivate and explicate the use of log-linear models. Of course, both the analogies and the distinctions between the different methods must be kept in mind.

[From the first edition, Chapters I, II, and III are about the same as the new 1, 2, and 3. Chapter IV is now Chapters 5 and 6. Chapter V is now 7, VI is 10, VII is 4 (and the sections are rearranged), VIII is 11, IX is 8, X is 9, and XV is 12.]

The book is written at several levels. A basic introductory course would take material from Chapters I, II (deemphasizing Section II.4), III, Sections IV.1 through IV.5 (eliminating the material on graphical models), Section IV.10, Chapter VII, and Chapter IX. The advanced modeling material at the end of Sections VII.1, VII.2, and possibly the material in Section IX.2 should be deleted in a basic introductory course. For Masters degree students in Statistics, all the material in Chapters I through V, VII, IX, and X should be accessible. For an applied Ph.D. course or for advanced Masters students, the material in Chapters VI and VIII can be incorporated. Chapter VI recapitulates material from the first five chapters using matrix notation. Chapter VIII recapitulates Chapter VII. This material is necessary (a) to get standard errors of estimates in anything other than the saturated model, (b) to explain the Newton-Raphson (iteratively reweighted least squares) algorithm, and (c) to discuss the weighted least

squares approach of Grizzle, Starmer, and Koch (1969). I also think that the more general approach used in these chapters provides a deeper understanding of the subject. Most of the material in Chapters VI and VIII requires no more sophistication than matrix arithmetic and being able to understand the definition of a column space. All of the material should be accessible to people who have had a course in linear models. Throughout the book, Chapter XV of Christensen (1987) is referenced for technical details. For completeness, and to allow the book to be used in nonapplied Ph.D. courses, Chapter XV has been reprinted in this volume under the same title, Chapter XV.

The prerequisites differ for the various courses described above. At a minimum, readers should have had a traditional course in statistical methods. To understand the vast majority of the book, courses in regression, analysis of variance, and basic statistical theory are recommended. To fully appreciate the book, it would help to already know linear model theory.

It is difficult for me to understand but many of my acquaintance view me as quite opinionated. While I admit that I have not tried to keep my opinions to myself, I have tried to clearly acknowledge them as my opinions.

There are many people I would like to thank in connection with this work. My family, Sharon and Fletch, were supportive throughout. Jackie Damrau did an exceptional job of typing the first draft. The folks at BMDP provided me with copies of 4F, LR, and 9R. MINITAB provided me with Versions 6.1 and 6.2. Dick Lund gave me a copy of MSUSTAT. All of the computations were performed with this software or GLIM. Several people made valuable comments on the manuscript; these include Rahman Azari, Larry Blackwood, Ron Schrader, and Elizabeth Slate. Joe Hill introduced me to statistical applications of graph theory and convinced me of their importance and elegance. He also commented on part of the book. My editors, Steve Fienberg and Ingram Olkin, were, as always, very helpful. Like many people, I originally learned about log-linear models from Steve's book. Two people deserve special mention for how much they contributed to this effort. I would not be the author of this book were it not for the amount of support provided in its development by Ed Bedrick and Wes Johnson. Wes provided much of the data used in the examples. I suppose that I should also thank the legislature of the state of Montana. It was their penury, while I worked at Montana State University, that motivated me to begin the project in the spring of 1987. If you don't like the book, blame them!

Ronald Christensen
Albuquerque, New Mexico
April 5, 1990
(Happy Birthday Dad)

Contents