Statistics for Biology and Health

Statistics for Biology and Health

Borchers/Buckland/Zucchini: Estimating Animal Abundance: Closed Populations.
Burzykowski/Molenberghs/Buyse: The Evaluation of Surrogate Endpoints.
Everitt/Rabe-Hesketh: Analyzing Medical Data Using S-PLUS.
Ewens/Grant: Statistical Methods in Bioinformatics: An Introduction, 2nd ed.
Gentleman/Carey/Huber/Irizarry/Dudoit: Bioinformatics and Computational Biology Solutions using R and Bioconductors.
Hougaard: Analysis of Multivariate Survival Data.
Keyfitz/Caswell: Applied Mathematical Demography, 3rd ed.
Klein/Moeschberger: Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed.
Kleinbaum/Klein: Survival Analysis: A Self-Learning Text.
Kleinbaum/Klein: Survival Analysis: A Self-Learning Text, 2nd ed.
Kleinbaum/Klein: Logistic Regression: A Self-Learning Text, 2nd ed.
Lange: Mathematical and Statistical Methods for Genetic Analysis, 2nd ed.
Manton/Singer/Suzman: Forecasting the Health of Elderly Populations.
Martinussen/Scheike: Dynamic Regression Models for Survival Data.
Nielsen: Statistical Methods in Molecular Evolution.
Moyé: Multiple Analyses in Clinical Trials: Fundamentals for Investigators.
Parmigiani/Garrett/Irizarry/Zeger: The Analysis of Gene Expression Data: Methods and Software.
Salsburg: The Use of Restricted Significance Tests in Clinical Trials.
Simon/Korn/McShane/Radmacher/Wright/Zhao: Design and Analysis of DNA Microarray Investigations.
Sorensen/Gianola: Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics.
Stallard/Manton/Cohen: Forecasting Product Liability Claims: Epidemiology and Modeling in the Manville Asbestos Case.
Therneau/Grambsch: Modeling Survival Data: Extending the Cox Model.
Vittinghoff/Glidden/Shiboski/McCulloch: Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.
Zhang/Singer: Recursive Partitioning in the Health Sciences.

Torben Martinussen
Thomas H. Scheike

Dynamic Regression Models for Survival Data

With 75 Illustrations

Springer

Torben Martinussen
Department of Natural Sciences
Royal Veterinary and Agricultural University
1871 Fredriksberg C
Denmark
torbenm@dina.kvl.dk

Thomas H. Scheike
Department of Biostatistics
University of Copenhagen
2200 Copenhagen N
Denmark
ts@biostat.ku.dk

Library of Congress Control Number: 2005930808

ISBN-10: 0-387-20274-9
ISBN-13: 978-0387-20274-7

Printed on acid-free paper.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 2 1

springer.com

To Rasmus, Jeppe, Mie and Ida
To Anders, Liva and Maria

Preface

This book studies and applies flexible models for survival data. Many developments in survival analysis are centered around the important Cox regression model, which we also study. A key issue in this book, however, is extensions of the Cox model and alternative models with most of them having the specific aim of dealing with *time-varying* effects of covariates in regression analysis. One model that receives special attention is the additive hazards model suggested by Aalen that is particularly well suited for dealing with time-varying covariate effects as well as simple to implement and use.

Survival data analysis has been a very active research field for several decades now. An important contribution that stimulated the entire field was the counting process formulation given by Aalen (1975) in his Berkeley Ph.D. thesis. Since then a large number of fine text books have been written on survival analysis and counting processes, with some key references being Andersen et al. (1993), Fleming & Harrington (1991), Kalbfleisch & Prentice (2002), Lawless (1982). Of these classics, Andersen et al. (1993) and Fleming & Harrington (1991) place a strong emphasis on the counting process formulation that is becoming more and more standard and is the one we also use in this monograph. More recently, there have been a large number of other fine text books intended for different audiences, a quick look in a library data base gives around 25 titles published from 1992 to 2002. Our monograph is primarily aimed at the biostatistical community with biomedical application as the motivating factor. Other excellent texts for the same audience are, for example, Klein & Moeschberger (1997) and Therneau & Grambsch (2000). We follow the same direction as Therneau

& Grambsch (2000) and try to combine a rather detailed description of the theory with an applied side that shows the use of the discussed models for practical data. This should make it possible for both theoretical as well as applied statisticians to see how the models we consider can be used and work. The practical use of models is a key issue in biomedical statistics where the data at hand often are motivating the model building and inferential procedures, but the practical use of the models should also help facilitate the basic understanding of the models in the counting process framework.

The practical aspects of survival analysis are illustrated with a set of worked examples where we use the R program. The standard models are implemented in the `survival` package in R written by Terry Therneau that contains a broad range of functions needed for survival analysis. The flexible regression models considered in this monograph have been implemented in an R package `timereg` whose manual is given in Appendix C. Throughout the presentation of the considered models we give worked examples with the R code needed to produce all output and figures shown in the book, and the reader should therefore be able to reproduce all our output and try out essentially all considered models.

The monograph contains 11 chapters, and 10 of these chapters deal with the analysis of counting process data. The last chapter is on longitudinal data and presents a link between the counting process data and longitudinal data that is called marked point process data in the stochastic processes world. It turns out that the models from both fields are strongly related.

We use a special note-environment for additional details and supplementary material. These notes may be skipped without loss of understanding of the key issues. Proofs are also set in a special environment indicating that these may also be skipped. We hope that this will help the less mathematically inclined reader in maneuvering through the book.

We have intended to include many of the mathematical details needed to get a complete understanding of the theory developed. However, after Chapter 5, the level of detail decreases as many of the arguments thereafter will be as in the preceding material. A simple clean presentation has here been our main goal.

We have included a set of exercises at the end of each chapter. Some of these give additional important failure time results. Others are meant to provide the reader with practice and insight into the suggested methods.

Acknowledgments

We are deeply grateful for the support and help of colleagues and friends. Martin Jacobsen introduced us to counting processes and martingale calculus and his teaching has been of great inspiration ever since. Some of

the exercises are taken from teaching material coming from Martin. Our interest for this research field was really boosted with the appearance of the book Andersen et al. (1993). We are grateful to these authors for the effort and interest they have put into this field. We have interacted particularly with the Danes of these authors: Per Kragh Andersen and Niels Keiding. Odd Aalen, Per Kragh Andersen, Ørnulf Borgan, Mette Gerster Harhoff, Kajsa Kvist, Yanqing Sun, Mei-Jie Zhang and some reviewers have read several chapters of earlier drafts. Their comments have been very useful to us and are greatly appreciated. Finally, we would like to thank our co-authors Mei-Jie Zhang, Christian Pipper and Ib Skovgaard of work related to this monograph for sharing their insight with us.

Copenhagen
November 2005

Torben Martinussen
Thomas H. Scheike

Contents