

Preface

The first edition of *Applied Longitudinal Analysis* was designed to serve as a textbook for a course on modern statistical methods for longitudinal data analysis, and subsequently, as a reference resource for students and researchers. The book was targeted at a broad audience: graduate students in statistics, statisticians working in the health sciences, pharmaceutical industry, and governmental health-related agencies, as well as researchers and graduate students from a variety of substantive fields. In the seven years that have elapsed since publication of the first edition, *Applied Longitudinal Analysis* has been used extensively in university classrooms throughout the United States and abroad. We are grateful to many colleagues, course instructors, students, and readers who have offered constructive suggestions on how the book could be improved. This feedback has been invaluable and helped shape the content of the second edition.

The feedback we received has encouraged us to retain the general structure and format of the first edition while taking the opportunity to introduce a number of new and important topics. Although there is much new material in this second edition, the principles that guided us in writing the first edition have not changed. Our primary goal is to present a rigorous and comprehensive description of modern statistical methods for the analysis of longitudinal data that is accessible to a wide range of readers. A strong emphasis is placed on the application of these methods to longitudinal data and the interpretation of results. Although the methods are presented in the setting of numerous applications to actual data sets drawn from studies in health-related fields, reflecting our own research interests in the health sciences, they apply equally to other areas of application, for example, education, psychology, and other branches of the behavioral and social sciences.

How does this edition differ from its predecessor? The major changes in this edition have resulted from the addition of six new chapters:

1. A chapter (Chapter 9) on “fixed effects models,” in which subject-specific effects are treated as fixed rather than random, has been added. This chapter complements the existing chapter on mixed effects models (Chapter 8) and includes a discussion of the relative advantages of these two classes of models.
2. In the first edition, a single chapter was devoted to marginal models and generalized estimating equations (GEE) that focused exclusively on binary and count data. We now devote two chapters (Chapters 12 and 13) to marginal models and GEE, with new material on models for ordinal responses, residual diagnostics, and issues that arise when modeling time-varying covariates.
3. A chapter (Chapter 15) on approximate methods for generalized linear mixed effects models discusses penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) methods. We highlight settings where these approximations are unlikely to be accurate and can yield biased estimates of effects.
4. A second chapter (Chapter 18) on missing data and dropout, focusing on multiple imputation and inverse probability weighting (IPW) methods, has been added. To give greater prominence to methods for accounting for missing data and dropout in longitudinal analyses, the two companion chapters (Chapters 17 and 18) now appear before the *Advanced Topics* part of the book.
5. A chapter (Chapter 19) on smoothing longitudinal data has been added to the *Advanced Topics*. This chapter focuses on the connection between penalized splines and linear mixed effects models.
6. A chapter on sample size and power (Chapter 20) has been added to the *Advanced Topics*. This chapter considers issues of sample size, power, number of repeated measurements, and study duration for longitudinal study designs.

In addition the chapter on residual analyses and diagnostics (Chapter 10) has been revised to include material on recently developed model-checking techniques based on cumulative sums of residuals and the chapters that review generalized linear models (Chapter 11) and generalized linear mixed effects models (Chapter 14) have been updated to include new material on models for ordinal data and on methods for handling overdispersion. Finally, extra problem sets have been added to many of the chapters.

As in the first edition, the prerequisites for a course based on this book are an introductory course in statistics and a strong background in regression analysis. Some previous exposure to generalized linear models (e.g., logistic regression) would be helpful, although these models are reviewed in detail in the text. An understanding of matrix algebra or calculus is not assumed. Although we do not assume a high level of mathematical preparation, we have written this book for the motivated reader who is willing to consider mathematical ideas. The more technical or mathematical sections of the book are signposted with asterisks and may be omitted at first reading without loss of continuity.

The methods described in this book require the use of appropriate statistical software. As before, we include illustrative *SAS* commands for performing the analyses presented throughout the text at the end of many chapters, with basic descriptions of their usage. Because many of the analyses we discuss can be performed using alternative software packages (e.g., *R*, *S-Plus*, *Stata*, and *SPSS*), this book can be supplemented with any one of them. Readers are encouraged to perform and verify the results of analyses using statistical software of their choice. Programming statements and computer output for selected examples, prepared using *SAS*, *Stata*, and *R*, can be downloaded from the website: www.biostat.harvard.edu/~fitzmaur/ala2e. Because statistical software is constantly evolving, we will endeavor to update the website as new procedures become available in the major statistical software packages. The thirty-two real data sets used throughout the text and problem sets to illustrate the applications of longitudinal methods also can be downloaded from the website.

We hope this second edition of *Applied Longitudinal Analysis* provides a broader foundation in modern methods for the analysis of longitudinal data and will prove a worthy successor to the first edition. The original impetus for writing this book arose from teaching a graduate-level course on “Applied Longitudinal Analysis” at the Harvard School of Public Health. We are especially grateful to the students who have participated in the course since its inception almost twenty years ago; we have learned much from these extraordinary students. The collection of individuals who gave us useful feedback on the first edition is far too long to list. However, we would like to thank the many friends and colleagues who have helped us with this project. A special word of thanks to Amy Herring and Russell Localio. We thank Amy for her many helpful and constructive suggestions on how the book could be improved. We thank Russell for reading a draft of the new chapters and for providing invaluable feedback and suggestions that improved their content. Thanks also to Nick Horton, Stu Lipsitz, and Caitlin Ravichandran for their helpful suggestions and insightful comments on several chapters. Finally, we thank Steve Quigley and Susanne Steitz-Filler of Wiley, for their advice and encouragement during all stages of this project.

GARRETT M. FITZMAURICE
NAN M. LAIRD
JAMES H. WARE

Boston, Massachusetts
May, 2011