STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:2979–2980
Published online in Wiley InterScience (www.interscience.wiley.com).

BOOK REVIEW
Editor: Petra Macaskill

1. *Brady T. West, Kathleen B. Welch, Andrzej T. Galecki*, Linear Mixed Models, A Practical Guide Using Statistical Software.

1. LINEAR MIXED MODELS, A PRACTICAL GUIDE USING STATISTICAL SOFTWARE. Brady T. West, Kathleen B. Welch, Andrzej T. Galecki, Chapman & Hall/CRC, Boca Raton, FL, 2007. No. of pages: xix +353. Price: $71.96. ISBN: 1-58488-480-0

Paralleling the development of computing technology, statistical methodology based on the linear mixed model (LMM) has increasingly become the subject of books and software products over the last 20 years. Statistical literature has actually discussed LMMs for almost 150 years (an interesting piece of trivia I picked up in the book), but computing limitations prevented LMMs from being routinely embraced in applied statistics until the dam broke in the late 1980s with the release of BMDP5V and more importantly SAS Proc MIXED. LMMs are at least an order of magnitude more complex than ordinary linear models, fraught with unique theoretical and practical subtleties. Hence, it is not surprising that software products for LMMs vary tremendously in the way that they specify the models, the user interfaces they provide, and the default and optional outputs that they produce. The current book addresses this variability in modern LMM methodology, with certain software products as the framework for the discussion.

This is a unique book. It is a combination of users' manual, comparative software review, product-oriented guide to a methodology, and case-studies textbook. To lay the foundation for all of this, the book also includes an introductory discussion of history, theory, and analysis strategies for LMMs. I applaud the authors for undertaking—and to a large extent succeeding in—such an ambitious project. For the book to combine all of these elements without exploding into an unreadable tome, the authors have cleverly organized the main chapters (Chapters 3–7) by combining five current LMM programs (SAS Proc MIXED, SPSS MIXED, R lme, STATA xtmixed, and HLM2-3) with five data sets of varying levels of complexity (two-level clustering, three-level clustering, repeated measures, random coefficients, and clustered longitudinal data). All five programs are applied to all five data sets, but as in latin square sampling, one program is highlighted for each of the data sets. This organization allows the authors to concisely but quite naturally demonstrate and compare the programs. It also adds a sense of evenhandedness in evaluating the programs. As a minor point, I wonder why other LMM products such as MLwiN and the REML directive of GENSTAT were left out of the discussion.

The introduction covers the purpose of the book and defines LMM terminology as used in the book. The chapter is a bit hard to follow because of the immediate use of jargon. For example, page 2 seems a bit early for a reader to appreciate a sentence such as

> Although individual cluster specific coefficients are not explicitly estimated, most LMM software produces cluster-specific 'predictions' . . . of the random cluster-specific effects.

The concept is correct, but would be more beneficial in a later chapter. The introductory chapter also includes an enjoyable brief history of LMMs, including key theoretical and software developments. This history would be more useful if references were provided for the theoretical developments.

Chapter 2 is a 40-page overview of the theory and practice of LMMs, interspersed with a few software notes (mostly dealing with SAS). This overview is itself a unique undertaking because of the broad variety of subjects covered. With admirable brevity and logical organization, the chapter addresses terminology, matrix formulation of the model, marginal *versus* subject-specific models, estimation, computational issues, model selection, information criteria, model building strategies,

Copyright © 2008 John Wiley & Sons, Ltd.

diagnostics, and other topics. Certain discussions, such as that on likelihood-ratio tests (LRTs) for covariance parameters when the null-hypothesis includes the boundary of the parameter space, are hard to find in other books. Owing to the scope of the overview, the authors do not bother to spell out their definitions of such terms as covariate, factor, subject, unit of measurement, unit of analysis, and even LMM (as opposed to a marginal model). Further confusion is occasioned by the necessary use of indistinct terms such as clusters, levels, repeated measures, and longitudinal data. Sensing this confusion, the authors provide tables (2.1 and 2.2) to clarify how they use these terms. There is a bit of sloppiness in the chapter, such as in the first paragraph of p. 17, where the authors describe an $\mathbf{X}_i$ matrix that cannot be full-rank, but then say that they will assume it is full-rank. In a minor typo (in equation 2.21), $\mathbf{X}_{i'}$ should be $\mathbf{X}_i'$. The authors vacillate between the terms *variance matrix* and *variance–covariance matrix*. The word *approximate* should be used more frequently to describe the asymptotic distributional results. Relatedly, the authors show a clear preference for LRTs over Wald-type tests for model selection (even for fixed-effect parameters). In this and later chapters, the authors treat LRTs as if they have an almost exact distributional theory.

Chapters 3–7 are well organized and contain thorough case studies of the five data sets. The authors discuss preparing the data, preliminary graphical displays of the data using features other than the LMM programs of the packages, model fitting, model selection (including schematics on relationships among the various models under consideration), diagnostics, and inferences regarding parameters of the final model(s). A central feature of each chapter is a set of tables comparing parameter estimates, standard errors, and hypothesis tests for the five programs. The results are intriguing and instructive as to the capabilities and philosophies of the different programs.

I learned much from this book. As a long-time user of LMMs, I enjoyed the authors' analyses of the data sets, especially when using programs that I was not acquainted with. Although I do not agree with everything in the book, it's an excellent reference, and it delivers what it promises to deliver. The authors clearly have considerable knowledge and experience with the software.

The book is not well suited as a textbook for newcomers to LMMs. It assumes a background in LMMs and in SAS programming. Chapter 2 is best viewed as a reference for Chapters 3–7 rather than as a stand-alone tutorial. I found myself constantly flipping back and forth while reading the book, but the sensible organization of the book allowed me to do this fairly effortlessly. The book is a bit sloppy in defining and using terminology, and it does not include problems. On balance, however, I recommend this book as a very useful reference for applied statisticians.

G. BRUCE SCHAALJE
Department of Statistics
Brigham Young University
UT, U.S.A

(DOI: 10.1002/sim.3167)

Copyright © 2008 John Wiley & Sons, Ltd.