TECHNICAL REPORT
R-399
June 2012

Biometrics 68, 657–660
June 2012

BOOK REVIEWS

WIENKE, A. **Frailty Models in Survival Analysis.** Chapman and Hall/CRC, Boca Raton, Florida, 2011. xxi + 301 pp. $104.95/£66.99, ISBN 9781420073881.

As usually understood in survival analysis, frailty models are extensions of the proportional hazards model that incorporate unobserved random multiplicative components into the hazard function. Frailty models have become quite popular in recent years both as a way of modeling otherwise unexplained heterogeneity in survival data and as an approach to modeling associations between survival times, as may arise for example in the study of lifetimes of twins or the times to loss of visual acuity in the left and right eyes of diabetics. Unlike previous books on this topic—Duchateau and Janssen (2008) and Hougaard (2000) come to mind—this book has a special focus on correlated frailty models for bivariate survival data. These models include separate independent frailty components for the pair and for each member of the pair. A strength of the book is the wide variety of real datasets used to illustrate models and methods. Occasionally some idiosyncratic choices are made—for example the treatment variable was omitted from a cancer dataset because the author believes that conclusions about the effectiveness of a treatment should be based only on randomized studies, and the present data were observational. This leaves the reader wondering whether an important variable might have been omitted from the analysis—thus voiding any substantive conclusions that could be drawn from the data.

This book will be a very useful reference for researchers in the area. The concise summaries of relevant literature that appear at intervals throughout the text are particularly valuable in this regard. Many formulas are presented. As a word of warning, some of these formulas are very complicated. Often their derivation may be simple conceptually but complex and tedious in practice (at least if done by hand, it would be interesting to know if any use has been made of symbolic manipulation software in developing or checking this work). Although the author often makes valiant efforts to explain the meaning of the formulas, they can still be quite forbidding. Needless to say, apart from recognizing some "old friends," this reviewer has not made any systematic attempt to verify the derivations.

It is important to emphasize what this book is not. It does not give a structured "how-to" guide for the practitioner. Useful references are given to packages and programs in SAS, STATA, and R, but no specific examples of code are presented. The relevant chapter in Therneau and Grambsch (2000) is still an excellent source in this regard. Nor is the book a textbook in the conventional sense. There are no exercises for the reader, to develop and test comprehension. My biggest complaint is that access to the raw datasets analyzed in the book is not provided. These datasets will mostly be too large to reproduce in print. However the book would have been very much more useful if some form of web access had been provided to at least some of the datasets. Without access to the raw data, the reader cannot verify the numerical results presented or examine the sensitivity of the analyses to different model specifications. And think how many additional citations the author could accumulate if his work could be referenced as a source for interesting data sets! (The same criticism also applies to the book by Duchateau and Janssen, though these authors do provide access to "pseudo datasets" of the same structure as the real data sets to enable the reader at least to go through the motions of a statistical analysis).

The exposition in Wienke's book is generally clear, although it is obvious that the author is not writing in his native language. The proof reading is good. Overall, I would recommend this book to specialists for the breadth of its coverage of the literature and to other readers seeking to sample the flavor of ongoing methodological research in frailty models. For practitioners wanting a more discursive introduction to the general concepts—as well as to the apparent paradoxes and interpretive issues that permeate this fascinating area—I recommend they first look at Chapters 6–8 of Aalen, Borgan, and Gjessing. (2008).

References

Aalen, O. O., Borgan, O. and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Viewpoint*. New York: Springer.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. New York: Springer.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Therneau, T. and Grambsch, P. (2000) *Modeling Survival Data: Extending the Cox Model.* New York: Springer.

David Oakes
Department of Biostatistics and Computational Biology
University of Rochester
Rochester, New York, USA

HILBE, J. M. **Logistic Regression Models.** Chapman & Hall/CRC, Boca Raton, Florida, 2009. xviii + 637 pp. \$83.95/£53.99. ISBN 9781420075755.

This book really does cover everything you ever wanted to know about logistic regression, at least what was worth knowing until 2008—with updates available on the author's website http://works.bepress.com/joseph_hilbe/. Hilbe, a former national athletics champion, philosopher, and expert in astronomy, is a master at explaining statistical concepts and methods. Readers familiar with his other expository work will know what to expect—great clarity.

The book provides considerable detail about all facets of logistic regression. No step of an argument is omitted so that the book will meet the needs of the reader who likes to see everything spelt out, while a person familiar with some of the topics has the option to skip "obvious" sections. The material has been thoroughly road-tested through classroom and web-based teaching. Suggestions from students are graciously acknowledged and incorporated; for instance, there is an appendix of the Greek alphabet with the names of the letters and how they are used in the statistical literature.

The prior knowledge assumed of readers is a solid background in linear regression, with an additional recommendation of calculus and probability theory (which I think would also be essential, together with introductory level familiarity with the principles of statistical inference). There are exercises at the end of every chapter with a solutions manual available from the author or the publisher.

Stata is used extensively, as might be expected given the author's long involvement with this and other statistical software. There is a 27-page introduction to Stata illustrated with numerical examples of logistic regression. Almost every section has examples using Stata. The corresponding R code is provided at the end of every chapter. SAS and SPSS procedures are mentioned, mainly where it is necessary to point out differences from Stata. More detail on the use of SAS, SPSS, and other software for logistic regression is available on the author's website.

A delightful history of logistic regression is given, together with stories of the early software such as Genstat and GLIM. There is even a section about discriminant analysis and how the various models relate to logistic regression. The history extends to the present time with the last chapter being on exact methods.

There is considerable emphasis on interpretation of results. For example, a whole chapter is devoted to interactions: how to formulate and interpret different parameterizations, calculation of effects, and graphical presentations.

In addition to the expected topics such as assessing model fit and modelling overdispersion, some less common topics are covered in depth. These include a variety of alternative ways of specifying and fitting models for variables with more than two categories, and a chapter on panel/clustered/longitudinal data.

The supporting material on the website covers not just typographical errors (which are extremely rare) and datasets, but revised sections of the book which the author felt needed clarification, as well as updates of newer work and software enhancements.

The focus is on helping the reader to learn and understand logistic regression. The audience is not just students meeting the topic for the first time, but also experienced users. I believe the book really does meet the author's goal "to serve as a handbook that researchers can turn to as a resource for applying this range of models to their data."

Annette J. Dobson
Professor of Biostatistics
School of Population Health
University of Queensland, Australia

MORRIS, M. D. **Design of Experiments: An Introduction Based on Linear Models.** Chapman & Hall/CRC, Boca Raton, FL, 2011. xviii + 355 pp. \$89.95/£59.99, ISBN 9781584889236.

Morris' text targets second or third year graduate students who are getting their first look at experimental design after having had a few core courses in statistical theory and methods, in particular, basic linear models. As such, it is more mathematical than most introductory books on design of experiments, but it is not a "theory of design" text. As might be expected, the book relies heavily on matrix algebra, but there are no advanced methods involved.

There are three distinguishing characteristics that separate this text from many related books. First, this book emphasizes design quality, or more precisely, how features of various designs such as nuisance parameters and random or fixed blocks affect the precision of estimates and the power of tests. Second, this is not a book from which to learn data analysis for designed experiments. Chapter 6 covers residual analysis, multiple comparisons, and transformations in just a dozen pages. There is, however, at least one "real world" experiment in each chapter to motivate the designs being used, and usually one or two designs with data in the chapter problems. Third, this text deals with inference for fixed effects only. There is discussion of random block effects, but nothing about variance components, nested effects, and so on. This is reasonable given Morris' overall approach.

Other than the gap left by random effects, coverage is pretty standard but with two less common additions. The first two chapters introduce the basic elements and ideas of experimentation and linear models. The next three chapters handle completely randomized designs, complete blocks, and Latin squares. As would be expected, the material in these three chapters sets up both the style for the remaining chapters and some content that will be seen again in later chapters. The typical chapter motivates the design, develops the matrix notation for the model and computes the information matrix, then considers the variance of estimable linear functions, and computation of noncentrality for power. One thing I appreciated throughout the text was the concise conclusion sections in each chapter.

Chapter 6 is the outlier chapter looking at data analysis rather than design quality. Chapters 7 and 8 follow with balanced incomplete blocks and BIBDs with random block effects. Chapters 9 and 10 bring us to factorial treatment structure and split plot designs. The development here is for balanced designs; this is cleaner and avoids the hand wringing over what kind of test to use, but I would have preferred at least a warning for the student that the issue is out there.

Chapters 11, 12, and 13 cover the basics of two-series designs including confounding into incomplete blocks and fractional factorials. The chapters on confounding and fractioning break from the pattern established in earlier chapters and deal mostly with construction of the designs and only a little with computing information.

Chapters 15 and 16 deal with first and second order response surface experiments, called regression experiments by Morris. I was unfamiliar with the appellation "functional" to describe continuously variable factors, but I like it and am going to adopt it. These two chapters also reap the rewards of having established the machinery for incomplete blocks, split plots, and so on in earlier chapters.

Chapter 14 on factorial group screening and chapter 16 introducing optimal design are bonus topics not found in many introductory texts. Algorithmically chosen designs are extremely common in some engineering fields, and it is appropriate that the topic is beginning to work its way into more texts.

Overall, this is a book that is easy to like, with good definitions of designs, few typographical errors, and consistent, straightforward explications of the models, but it may be a difficult book to use depending on the curriculum at your school. It works well for its intended niche, but it has too much math for the broad-market introduction to experimental design course, not enough theory for a real theory of design course, and, at least for my taste, not enough data analysis. That said, I can picture a lot of students using a text aimed at a broad-market design course but who need to understand more about what is going on behind the curtain. Morris' text also fills that gap very well.

Gary W. Oehlert
School of Statistics
University of Minnesota
Minneapolis, Minnesota, USA

WEISBERG, H. I. **Bias and Causation, Models and Judgment for Valid Comparisons.** John Wiley & Sons, Inc., Hoboken, New Jersey, 2010. xv + 348 pp. $110.00/€89.90, ISBN 9780470286395.

This is a thoughtful and well written book, covering important issues of causal inference in every field of applied data analysis. However, although the book shines in the motivational and conceptual levels, it fades in the mathematical tools that are harnessed to support the conceptual discussions. My review will focus on this weakness, because it is the one factor that prevented me from enjoying the lucid discussion throughout.

The book is about bias or, more specifically, about "identifying and dealing with bias in statistical research on causal effects" (from the back cover). Naturally, readers expect to find methods, criteria, or algorithms that facilitate the "identification" of bias, its assessment or its control. However, with the exception of the illusive and "catch all" assumption of "ignorability," the book stops short of showing readers what needs to be assumed or what needs to be done to control bias.

This void is a direct consequence of the restricted mathematical language chosen to illuminate examples, concepts and assumptions. This language, which has its roots in Rothman's "sufficient cause" classification (Rothman, 1976) and Rubin's "potential outcome" framework (Rubin, 1974) does not recognize modeling notions such as "processes," "omitted factors," or "causal mechanisms" that guide scientific thoughts, but forces one to articulate knowledge through counterfactual categories such as "doomed," "causal," "preventive," and "immune," and the proportions of individuals in each category. It is an all-or-nothing framework. If one assumes "ignorability," bias disappears; if not, bias persists, and one remains at the mercy of the (wrong) assumption that adjusting for as many covariates as one can measure would reduce bias (Rubin, 2009; Pearl, 2009a, 2009b, 2011a). The question of going from scientific knowledge to bias reduction, as well as the question of defending "ignorability-type" assumptions, remain outside the formal analysis.

The commitment to this mathematical language forces the examples to take the form of numerical tables involving counterfactual variables, rather than depicting the story behind the examples (e.g., through equations or diagrams.) Such tables may convince readers that the phenomenon demonstrated can indeed take place with certain tweaking of parameters, but fail to give readers a sense of the general class of problems where the phenomenon will occur.

Take, for example, Simpson's paradox, which Weisberg (2010, pp. 164–167) describes as "we might find that conditional effects are very similar within each stratum of the third factor (e.g., man and women), but opposite to the direction of the overall effect." (Here, the word "effect" means "association naively presumed to represent effect.") Weisberg rightly continues to the heart of Simpson's paradox and asks:

> "The question then arises of which effect (adjusted or unadjusted) represents a causal effect. Usually, it is assumed that the more "refined" conditional analysis represents the "true" causal effect, reflected in the common effect within strata, whereas the unadjusted effect results from confounding."

At this point the book does not stop to tell us if this "usual" assumption is valid or not (it is not) or how one can go about deciding when it is valid. Instead, we are instructed to construct tables involving doomed/causal/preventive/immune categories of individuals, translate them into spreadsheet, from which we can "calculate the empirical effects under any set of assumptions about the various parameters."

The reader thus gets the impression that, to determine the key question: "which effect (adjusted or unadjusted) represents a causal effect" one needs to guess the relative sizes of the doomed/causal/preventive/immune strata in both the male and female population, for both the exposed and unexposed groups, arrange them in a table like those in Tables 7.12, 7.13, and 7.14, go through the arithmetic, and only then conclude which effect is causal? This is unrealistic, because if we knew the relative sizes of those strata, we would not be facing Simpson's dilemma in the first place, but calculate the causal effect directly.

Modern treatments of Simpson's paradox can and should tell us how to make this determination directly from the causal story behind the example (See, for example, Pearl, 2009c, p. 383) without guessing relative sizes of strata and without going through the lengthy arithmetic.

Again, it is not a fault of Weisberg, but of the language of tables and strata, which does accept any such notions as "the causal story behind the example." More generally, this language does not allow for causal assumptions to be articulated in a format that matches the way scientific knowledge is stored and communicated. Weisberg has done an incredibly fine job overcoming this basic limitation of the potential outcome language, but there are limits to what good writing can do when mathematical notation is opaque.

Sailing on good writing, Weisberg manages to walk the reader through an impressive array of concepts and topics, including collapsibility, propensity scores, sensitivity analysis, and mediation. Each topic is introduced in a proper technical perspective, starting with its historical roots and ending with its impact on modern causal analysis. It is unfortunate though that the discussion is occasionally marred by myths that once served popular folklore and have since been discarded by analysis.

One such myth is the belief that the use of propensity-score somehow contributes to bias reduction, that it requires no modeling assumptions, and it is "to some degree, capable of providing warnings that the available data may not support a valid causal estimate." (Weisberg, 2010, pp. 141)

Mathematical analysis has overturned these beliefs. The proper choice of covariates into the propensity-score is dependent critically on modeling assumptions (Pearl, 2009a, 2009b, 2011a; Rubin, 2009). The propensity-score method, like any other model-free analysis, cannot give us any warning about the invalidity of the causal estimates. Finally, the propensity-score is merely a powerful estimator, and conditioning on the propensity score would be theoretically equivalent (asymptotically) to controlling on its covariates, regardless of whether strong ignorability holds (Pearl, 2009c, p. 349).

Another myth that finds its way to Weisberg's book concerns causal mediation, sometimes called direct and indirect effects. According to Weisberg (p. 208), "The theory of principal stratification has helped to clarify the essential nature of causal mediation." The hard truth is that principal stratification has helped circumvent, rather than clarify the essential nature of causal mediation. Most participants in a public discussion of the usages of principal strata, including former proponents of this framework now admit that principal strata has nothing to do with causal mediation (Joffe, 2011; Pearl, 2011b; Sjölander, 2011; VanderWeele, 2011).

To summarize, this book would be an excellent companion to standard statistics texts, serving to elucidate the unique problems that data analysts face when challenged to assess causal-effect relationships. When it comes to solving those problems though, the book should be supplemented with one that properly demonstrates the mathematics of modern causal analysis.

References

Joffe, M. (2011). Principle stratification and attribution prohibition: Good ideas taken too far. *The International Journal of Biostatistics* **7,** DOI: 10.2202/1557-4679.1367.

Joffe, M. M., Yang, W. P. and Feldman, H. I. (2010). Selective ignorability assumptions in causal inference. *The International Journal of Biostatistics* **6,** DOI: 10.2202/1557-4679.1199.

Pearl, J. (2009a). Letter to the editor: Remarks on the method of propensity scores. *Statistics in Medicine* **28,** 1415–1416.

Pearl, J. (2009b). Myth, confusion, and science in causal analysis. Tech. Rep. R-348, http://ftp.cs.ucla.edu/pub/stat ser/r348.pdf, University of California, Los Angeles, CA.

Pearl, J. (2009c). *Causality: Models, Reasoning, and Inference*. 2nd edition. New York: Cambridge University Press.

Pearl, J. (2011a). Invited commentary: Understanding bias amplification. *American Journal of Epidemiology* **174,** 1223–1227.

Pearl, J. (2011b). Principal stratification – a goal or a tool? *The International Journal of Biostatistics* **7,** DOI: 10.2202/1557-4679.1322.

Rothman, K. (1976). Causes. *American Journal of Epidemiology* **104,** 587–592.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66,** 688–701.

Rubin, D. (2009). Author's reply: Should observational studies be designed to allow lack of balance in covariate distributions across treatment group? *Statistics in Medicine* **28,** 1420–1423.

Sjölander, A. (2011). Reaction to Pearl's critique of principal stratification. *The International Journal of Biostatistics* **7,** DOI: 10.2202/1557-4679.1324.

VanderWeele, T. (2011). Principal stratification—uses and limitations. *The International Journal of Biostatistics* **7,** DOI: 10.2202/1557-4679.1329.

Weisberg, H. (2010). *Bias and Causation: Models and Judgment for Valid Comparisons*. Hoboken, NJ: John Wiley and Sons, Inc.

Judea Pearl
Computer Science Department
University of California, Los Angeles
Los Angeles, California, USA