Reviews of Books and Teaching Materials

DNA Methylation Microarrays: Experimental Design and Statistical Analysis.

Sun-Chong WANG and Arturas PETRONIS. Boca Raton, FL: Chapman & Hall/CRC, 2008, xxvi + 230 pp., $79.95 (H), ISBN: 978-1-4200-6727-9.

Epigenetics is the study of heritable changes in genome function that occur without a change in DNA sequence. One such change is DNA methylation which involves the addition of a methyl group at a specific location in a DNA sequence, typically reducing gene expression. While the number of *gene expression* microarray technologies have now arguably converged to a few mainstream approaches, that is not yet the case for DNA methylation due to the rapidly changing techniques by which methylation (and other epigenetic modifications) are measured.

This new hardback in the CRC Biostatistics Series is "intended to help researchers and students analyze high throughput epigenomic data, in particular DNA methylayion microarray data, with sound statistics." It is written by a bioinformatics researcher (Wang) and a renowned epigenetics researcher (Petronis). The book gives a good overview of the current state of the art of methylation arrays and other associated techniques, but is not suited as a standalone class text because it does not contain exercises, worked examples, or programming code to carry out the algorithms discussed. The book uses a large number of black-and-white figures and includes a CD-ROM with color versions of the figures in PDF and PostScript.

The text is informally divided into four parts. Part I consists of Chapters 1–4 and is the foundational material for the rest of the book. 'Applied Statistics' (Chapter 1) gives a brief overview of descriptive and inferential statistics and touches on the most basic hypothesis tests (both parametric and nonparametric). 'DNA Methylation Microarrays and Quality Control' (Chapter 2) delves into the array technology, and more importantly, the laboratory techniques by which (un)methylation is measured. This is the most crucial chapter in the book, because it explains how methylation experiments differ from gene expression experiments—differences that can greatly impact the "standard assumptions" upon which many statistical methods for microarray data are based. Bisulfite sequencing (the gold standard for measuring methylation) is mentioned briefly in other chapters but not this one, which is odd. 'Experimental Design' (Chapter 3) follows with a treatment of the topic which is typical for a book about microarrays (outside the excellent section on pooling biological samples). Finally, Part I closes with 'Data Normalization' (Chapter 4). Appropriate normalization methods can differ by methylation technique, but this chapter focuses more on standard approaches that are used for one-color and two-color gene expression arrays.

Part II shifts into hypothesis-based analyses of the normalized data. 'Significant Differential Methylation' (Chapter 5) covers the use of linear models to find differences in methylation level between groups via t and F-tests of contrasts and modified versions of the same. The notation, typesetting, and terminology used in this chapter could be improved. There is also a discussion of p-value adjustments for multiple testing and a permutation test for bisulfite sequencing data which is useful. 'High-Density Genomic Tiling Arrays' (Chapter 7) takes a brief look at the unique problems such arrays induce, and some of the sophisticated solutions for dealing with those. This chapter contains the most specialized material and should be used as a launching point into the literature for necessary details.

Part III consists of Chapters 7–10 which are concerned with exploratory data analysis. Chapters 8 ('Cluster Analysis') and 9 ('Statistical Classification') present standard material on the topics as there is nothing unique about dealing with methylation arrays in those areas. 'Interdependency Network of DNA Methylation' (Chapter 9) is a fascinating but concise chapter given the complexity of the problem, although 'Time Series Experiment' (Chapter 10) can be viewed as an addendum to Chapter 9. It contains an overview of one of the author's published approaches to constructing a regulatory network based on short time series data. Part IV (Chapters 11–13) closes out the text with very useful material for facilitating analysis ('Open Source Software for Microarray Data Analysis' and 'Public Microarray Data Repositories') and interpreting the resultant findings ('Online Annotation').

This text *is not* suited for the biologist turned "do-it-yourself" statistician, as the breadth of tools employed for the analysis of methylation data is both wide and relatively deep (e.g., principal components, wavelets, hidden Markov models, genetic algorithms) or for someone without a generous number of statistics and mathematics courses under their belt. Furthermore, it would not be practical to try to make a specialized book like this self-contained. The book *is* suited for people with some exposure to microarrays, but not necessarily methylation arrays. For microarray novices, start first with Wit and McClure (2004), Drăghici (2003), or Amaratunga and Cabrera (2004) before diving into this book. Even for those statisticians acquainted with arrays, there are still some dense biological passages such as: "DNA methylation can be corroborated or induced by a single proximal methylated CpG dinucleotide via the action of maintenance and de novo DNA methyltransferases." Many readers would find a glossary of terms (e.g., locus, clone, and promoter) beneficial.

My primary criticism of the book is poor copyediting from the publisher. It is a first edition, but there is an unacceptable number of typographical errors, incoherent or ambiguous sentences, and non sequiturs along with some unusual phraseology (e.g., "over than expected," "are comparing to each other," "persuited"). The writing is uneven in places and the segues awkward. Often, a phrase or acronym is used several times before being explained. Finally, it is confusing at times to understand to which technology and/or technique combination (e.g., DMH, mDIP, tiling arrays) a given passage is applicable. For those familiar with methylation studies, these issues pose no problems, but for others it could be frustrating.

In summary, *DNA Methylation Microarrays* shows great promise as the first book on the statistical analysis of DNA methylation data, but it needs some fine tuning. It fills a much needed niche, and I wish the book were available several years ago for my own use.

J. Wade DAVIS
University of Missouri–Columbia

REFERENCES

Amaratunga, D., and Cabrera, J. (2004), *Exploration and Analysis of DNA Microarray and Protein Array Data*, Hoboken, NJ: Wiley.

Drăghici, S. (2003), *Data Analysis Tools for DNA Microarrays*, Boca Raton, FL: Chapman & Hall/CRC.

Wit, E., and McClure, J. (2004), *Statistics for Microarrays: Design, Analysis and Inference*, Hoboken, NJ: Wiley.

Ecological Models and Data in R.

Benjamin M. BOLKER. Princeton, NJ: Princeton University Press, 2008, vii + 396 pp., $55.00 (H), ISBN: 978-0-691-12522-0.

In *Ecological Models and Data in R* Ben Bolker sets out to show ecologists how to use R for the synthesis of data and ecological models. In this sense, the volume is an extension of Hilborn and Mangel (1997) *The Ecological Detective: Confronting Models With Data* paired with the modern statistical computing environment provided by R. I congratulate Bolker for his success in creating a book that scratches so many itches. It provides a solid introduction to R, interleaving high-level descriptions of methods with execution in real working code, and completing the triangle with ample motivation and demonstration on real data. In some sense, it is too bad that the topics in the book are narrowly applied to ecological data, as other disciplines too could benefit from the insights and illustrative use of R provided in this book. Readers from the graduate level onwards will find it useful as both instructional and as a reference text. Its one weakness is that it may not be appropriate as the sole text for a graduate course on ecological (statistical) modeling, since the development is highly nonlinear. However, it would be an essential companion for any course in ecological modeling that has a practical component.

The strengths of this text lie in the R code. From the author's webpage I was able to surmise that the book is authored in Sweave. This allows the R code that is embedded within a LaTeX document to be verified by R at compile time, and to have the resulting figures and output therein generated online. While this may seem like an esoteric point, I assure you that it is not; it is very practical indeed. What this means is that the code that appears in the book is, in some sense, certified by R at the time it went to press. That is, the code works and produces the output that you see on the pages—nothing more, nothing less; no smoke and mirrors. You will not find any bugs or the sloppy syntax errors that too often plague texts providing code. The .Rnw files that the author used to generate (an earlier version of) the book are available on his webpage. Even the R code used to generate figures in the book, which is not provided directly in the text, is available to the reader through these files. I hope to see more of this from authors in the future, and applaud Bolker for going to such lengths. In addition to the .Rnw files the author supports the book further by maintaining a wiki, an elegant errata page, data, and scripts, all in addition to providing an R package called emdbook, available on CRAN, that contains many of the subroutines developed in the text.

The focus of the book is on likelihood based methods. While there are chapters and sections on deterministic modeling, exploratory data analysis and visualization, and stochastic simulation, the majority of the text has to do with the optimization, numerical calculations, and simulation methods needed to extract information from likelihoods that are constructed for the ecological inference problem at hand. The optimization methods provided within R to, say, maximize the likelihood are covered in detail in terms of intuition, methodology, and numerous examples and are provided with code. This alone ought to equip the reader with enough firepower to tackle most inference problems he or she encounters. Bayesian methods are also treated via the R2WinBUGS package. Numerical integration schemes and limited R code for MCMC is also provided. However it would seem that, from their organization in the text and shorter exposition, these topics are somewhat ancillary. This may reasonably be due to the fact that there are relatively fewer R packages supporting Bayesian methods.

Due reverence is given in the text to the need to asses the variability of inferences obtained from the likelihood. Unfortunately, I find that more could have been done here. The bootstrap, which is probably the simplest method available to the computational statistician for assessing the variability of estimators, does not even get a mention. The topic of model uncertainty, and how to analyze samples obtained from the posterior distribution, could have been treated more thoroughly. While AIC and BIC are mentioned, the fully Bayesian treatment via Bayes factors or reversible jump MCMC are not, even though such functionality is available in WinBUGS. Although the CODA package gets a mention, more could be said about how a Bayesian summarizes the inferences obtained from MCMC and WinBUGS in R.

Generally speaking, the book is a pleasure to read. It is peppered with insightful comments about the methods and code it is illustrating. The text is witty, informative, and thoughtful. However, there are some structural aspects that detract somewhat from these strong plusses. Chapters 3 and 4 give excellent descriptions of deterministic modeling and probability, respectively, but they are nearly devoid of R examples. Those that are provided are relegated to an R supplement at the end of the chapters. I would have liked to see more examples here inline with the text, as I believe that this sort of interleaving is what sets books like these apart from their competitors. In contrast, Chapters 9–11 contain examples but read more like appendices, and are almost referred to as such in the main body of the book (Chapters 5–8). Chapter 11 on Dynamic Models grazes the tip of the iceberg of this important statistical toolkit for the ecologist. Without a treatment of particle filtering and associated software it is not clear that the reader would be able to take advantage of the power of state space models. A better use of ink may have been to cover classical time series modeling (e.g., ARMA models), etc. Well-established methods for these are provided in R, while inference for state space models is not. A treatment of spatial statistics (e.g., kriging and Markov random fields), for which many R packages are readily available, would also have been welcome.

I would have liked to see more in the text about good R programming practice (e.g., vectorizing, functions, file structure, and style). The book also lacks comments on how some unavoidably loop-intensive procedures, like (stochastic) dynamic programming (an up-and-coming tool in ecology), are natural candidates for the C-hooks that R provides. A short example outlining how simple it is to construct a "C in R" code, and how favorably it would compare to an R-only implementation, say, for solving a simple dynamic program, would go a long way toward encouraging more advance readers, who may already have C code lying around, to learn more about how it can be ported into R. At it stands, the authors quick pointer to the R *Extensions* manual on p. 243 may be taken as more of a deterrent than an encouragement.

It is easy to criticize a relatively short, 400 page, book on statistical modeling in R for ecology (or any topic) for missing something when the CRAN task view for the "Analysis of Ecological and Environmental Data" in R contains links to thousands of pages of eligible routines and documentation. Surely the point of the book was not to cover everything, but rather to cover a few key aspects well, so as to whet the appetite for the use of R for practical ecological modeling. In this the author hits the nail on the head. So I would like to summarize by reiterating my praise for this book. It is well written and provides an excellent peek into how R can be exploited for applied (statistical) ecology. The value for the money is very high. I would be surprised to find that success in its current form does not bring demand for future editions with many extensions.

Robert B. GRAMACY
University of Cambridge

REFERENCE

Hilborn, R., and Mangel, M. (1997), *The Ecological Detective: Confronting Models With Data*, Princeton, NJ: Princeton University Press.

A First Course in Statistical Programming With R.

W. John BRAUN and Duncan J. MURDOCH. New York: Cambridge University Press, 2007, x + 163 pp., $120.00 (H), ISBN: 978-0-521-87265-2. $50.00 (P), ISBN: 978-0-521-69424-7.

On January 6, 2009 the New York Times published the article "Data Analysts Captivated by R's Power" which described the contemporary and ubiquitous significance of the R language (Vance 2009). R's popularity has grown despite the fact that many practicing statisticians have never taken a course in R (or any other statistical programming language, for that matter). If we were fortunate, perhaps as a junior graduate student, a group of kind-hearted senior students organized seminars for our benefit. Now, however, graduate and undergraduate programs in statistics recognize the significance of statistical programming and are embracing dedicated courses. Correspondingly, many authors are beginning to write books to address this growing audience.

Having just taught a revamped course in statistical computing for undergraduate statistics majors, I found myself searching through these many recent texts. Braun and Murdoch's short (163 pages) book (B&M) stood out from the rest in one important way: instead of focusing on a broad range of common statistical applications [as do Everitt and Hothorn (2006), Maindonald and Braun (2007), and Crawley (2005)], modern statistical applications (e.g., Venables and Ripley 2002), or even introductory statistics (Verzani 2005), it focused on a few simple examples to illustrate and emphasize basic programming concepts and data structures. Students who study from this text should have little difficulty using R for any subsequent applied course because the basic principles of R programming are well illustrated.

B&M assumes that the reader is familiar with very basic statistical concepts and does not consider advanced statistical methods. Rather, the text uses statistical objectives as a method to illustrate programming concepts. The core of the text was motivated by the authors' realization that statistics students need to have a basic understanding of many aspects of programming (numerical computation, data types, style, debugging, and efficiency), but don't have the time or need for a comprehensive computer science curriculum. For the most part, the authors accomplish their goals by providing a simple framework for a student to get a foothold on basic programming as it pertains to our discipline.

The two most important programming chapters are not too dissimilar to those found in other texts (Chapter 2: Introduction to the R Language and Chapter 4: Programming With R). Among other topics, Chapter 2 covers basic data modes and classes, the importance of vectors, and built-in functions. Chapter 4 explains flow control, functions and scope, and provides helpful tips for debugging and efficiency. Since this text is likely to be used by students with little or no programming experience, I would have liked to have seen more on programming style: emphasis on commenting, indentations, choice of variable names, etc. Chapter 3 provides a terse introduction to both high and low-level graphics functions. Because data visualization is one of R's strong suits, I was a bit disappointed at the brevity of this chapter (only 13 pages!). That said, there is enough material to serve as a launching point for an instructor who wishes to take the topic further. (It is entirely possible to write an entire text on graphics in R; e.g., Murrell 2006). The remaining chapters present topics that support statistical computation as well as provide illustrations of other commonly used R functions and programming examples: simulation (including random number generation and Monte Carlo integration), computational linear algebra, numerical optimization, and linear programming.

Exercises appear throughout and range in difficulty from trivial to moderately difficult for beginning programmers. Most of the exercises can be solved with a few lines of code. Missing from the exercises are broader problems which require stitching many functions together to produce a solution… a skill that is oft-used in practice. In my course, I found it necessary to supplement the text exercise with more comprehensive data-driven problems. These problems required that students identify and break down the programming tasks and assemble them together in a meaningful and well-documented manner. Indeed B&M helps students develop skills for the smaller tasks, but does little to help them create more comprehensive code.

Overall, B&M is a useful introductory text that provides enough material to support a 10-week course in programming provided that instructors are willing to supplement the course with additional complex data-driven problems. A longer course would require additional supplementation.

Andrew SCHAFFNER
California Polytechnic State University

REFERENCES

Crawley, M. J. (2005), *Statistics: An Introduction Using R*, Chichester: Wiley.

Everitt, B., and Hothorn, T. (2006), *A Handbook of Statistical Analyses Using R*, Boca Raton: Chapman & Hall.

Maindonald, J., and Braun, J. (2007), *Data Analysis and Graphics Using R: An Example-Based Approach* (2nd ed.), New York: Cambridge University Press.

Murrell, P. (2006), *R Graphics*, Boca Raton: Chapman & Hall/CRC.

Vance, A. (2009), available at *http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html*.

Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics With S* (4th ed.), New York: Springer.

Verzani, J. (2005), *Using R for Introductory Statistics*, Boca Raton: Chapman & Hall/CRC.

Linear Models: The Theory and Application of Analysis of Variance.

Brenton R. CLARKE. Hoboken, NJ: Wiley, 2008, xviii + 241 pp., $99.50 (H), ISBN: 978-0-470-02566-6.

This text provides a terse and easy to read overview of topics pertaining to linear models with emphasis on the analysis of variance (ANOVA). The reader is assumed to have knowledge of matrix theory and a course in probability and statistical inference. The author considers this to be a “graduate text for students at the master’s or Ph.D. level at an American university” (p. xiii). The author later states, “this book is pitched at an advanced level of study” (p. ix). This is due in part to the use of vector space results involving subspaces, projections, and Kronecker products. Despite the level of sophistication, this text would be appropriate for graduate statistics students or practitioners who wish to learn how a matrix based approach can be used to produce the standard formulas used in regression and experimental design.

The first four chapters (76 pages) provide the requisite linear models development. Chapter 1 defines the linear model along with examples illustrating simple linear regression, multiple linear regression, 2-way factorial designs, and the analysis of covariance. Throughout the text, all models assume the parameter vector is estimable. Projection matrices and vector spaces are briefly discussed in Chapter 2. Relevant results are given for orthogonal complements and combinations of projections. It should be noted that generalized inverses are not used. Other important linear algebra properties are scattered throughout the text and in the problem sections. This approach makes it difficult to use this text as a reference, but it does help instruction as it allows the content to motivate the need for these properties. Ordinary least squares, the Gauss–Markov Theorem, and distribution theory are briefly addressed in Chapters 3 and 4. This discussion includes least squares solutions that satisfy constraints. The problem sets for these chapters nicely incorporate results, details, and examples.

Chapters 5 and 6 (82 pages) present the specific orientation of the text. The author states, “In this chapter, I put my own stamp on how to derive distribution theory for ANOVA in higher-order models” (p. 77). The basic technique is to obtain a particular full rank factorization of a projection in terms of Helmert matrices and Kronecker products. This approach, developed in Clarke (2002), requires complete balance in classification models and the specification of the ordering of observations. This text is not orientated towards establishing a general projection based approach for analyzing linear models nor towards statistical computation. Rather, these representations are used to provide insight into the standard ANOVA formulas, noncentrality parameters, and expected mean squares. Two-factor designs, three-factor designs, Latin square designs, and 2 factorial designs are used to demonstrate the application of this method. The problem sections again ask the reader to work through results, details, and examples.

The remaining material in the text, consisting of Chapters 7, 8, 9, 10, 11 (69 pages), contains brief treatments of various topics in linear models. Chapter 7 presents residual diagnostics and robust approaches. More detailed discussion on these topics can be found in Myers (1990) and Clarke (1994, 2000). Chapter 8 discusses variance component models for one factor designs, two factor designs, and split plot designs. The ANOVA approaches discussed in Chapters 5 and 6 are also applied to these models. Chapter 9 introduces maximum likelihood, restricted maximum likelihood, and Bayesian techniques. A good reference for detailed treatment of the material in Chapters 8 and 9 is Searle, Casella, and McCulloch (1992). Chapter 10 presents a nice discussion of uncorrelated residuals borrowing from Theil (1965), Brown, Durbin, and Evans (1975), and Gudolphin and DeTullio (1978). While the above references provide detailed presentations of the topics in these chapters, this text gives a succinct overview along with citations.

This book provides an introduction to a vector space approach for linear models. The niche for this book is as a supplement for statistics students or researchers that would prefer a simple and terse treatment of linear models or a quick overview of the many topics previously mentioned. There are linear model texts, such as Christensen (2002) or Graybill (1976), that contain thorough and rigorous developments of linear models.

Shaun S. WULFF
University of Wyoming

REFERENCES

Brown, R. L., Durbin, J., and Evans, J. M. (1975), “Techniques for Testing the Constancy of Regression Relationships Over Time,” *Journal of the Royal Statistical Society, Ser. B*, 37, 149–192.

Christensen, R. (2002), *Plane Answers to Complex Questions*, New York: Springer-Verlag.

Clarke, B. R. (1994), “Empirical Evidence for Adaptive Confidence Intervals and Identification of Outliers Using Methods of Trimming,” *Australian Journal of Statistics*, 36, 45–58.

——— (2000), “An Adaptive Method of Estimation and Outlier Detection in Regression Applicable for Small to Moderate Sample Sizes,” *Discussiones Mathematicae Probability and Statistics*, 20, 25–50.

——— (2002), “A Representation of Orthogonal Components in Analysis of Variance,” *International Mathematical Journal*, 1, 133–147.

Graybill, F. A. (1976), *Theory and Application of the Linear Model*, Belmont, CA: Wadsworth.

Gudolphin, E. J., and DeTullio, M. (1978), “Invariance Properties of Uncorrelated Residual Transformations,” *Journal of the Royal Statistical Society, Ser. B*, 40, 313–321.

Myers, R. H. (1990), *Classical and Modern Regression With Applications*, Belmont, CA: Duxbury.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

Theil, H. (1965), "The Analysis of Disturbances in Regression Analysis," *Journal of the Americal Statistical Association*, 60, 1067–1079.

Making the Connection: Research and Teaching in Undergraduate Mathematics Education.

Marilyn P. CARLSON and Chris RASMUSSEN. Washington, DC: Mathematical Association of America, 2008, xi + 319 pp., $59.50 (P), ISBN: 978-0-88385-183-8.

Teaching mathematics or statistics is a learning process. After one teaches a particular class, one reflects on the difficulty that students had learning particular concepts. On the basis of this teaching experience, one typically adjusts his or her style of teaching and/or the pace of the class to better communicate the subject matter.

As described in the introduction, this book is intended for mathematics instructors who wish to improve the learning and achievement of students in their undergraduate mathematics courses. Much of the time of the academic mathematician is focused on teaching, preparing to teach, or thinking about ways of improving their teaching. This effort naturally raises questions on the manner of teaching and on how students learn concepts in mathematics. But there is an ongoing effort by the mathematics education community to produce theories, models, and learning materials that address these questions of interest. The intent of this book is to survey mathematics education research for mathematics instructors who may benefit from this knowledge.

This volume consists of a series of articles divided into two parts. In Part 1 titled "Student Thinking," the authors give an overview of the research findings from mathematics education on how students learn particular mathematics concepts. In Part 2 titled "Cross-Cutting Themes," the chapters focus on the implication of this research on the effective teaching of these concepts.

In Part 1, the articles describe the processes in which students learn the concepts of variable and limits and the notion of infinity. One of the hardest things to communicate to a mathematics major is how to formally construct proofs, and several articles describe approaches to help students learn how to construct proofs. This section describes strategies that may be helpful for teaching mathematics induction and concepts in group theory. In Part 2, one article describes achieving a proper balance between "all lecture" and "all student discovery" styles of teaching. Another article describes the use of computer-based tools in helping students in learning mathematics, and a third article describes how problem solving research is useful for facilitating problem solving attributes in students.

This volume provides a useful entry into the large body of research in mathematics education on how students learn mathematical concepts. The articles describe a variety of issues and associated teaching challenges, and I would think all mathematics teachers would benefit from these articles. As a statistician who teaches probability and statistics courses, some of the articles on particular mathematics concepts such as the concept of variable and the notion of a limit didn't seem directly useful in my teaching. But I believe that much of the mathematics education research presented in this book has implications on how we learn statistical concepts. For example, the article "How Do Mathematicians Learn to Teach?" describes the qualities of a good teacher, and many of these attributes would apply to a good statistics teacher. The article "When Students Don't Apply the Knowledge You Think They Have, Rethink Your Assumptions About Transfer" has obvious implications on how students learn to apply statistical methods such as regression, and the article "Computer-Based Technologies and Plausible Reasoning" give conclusions that may help us understand how computer demonstrations are useful in understanding statistical concepts. As summarized on the back cover, I agree that this book opens up a meaningful dialogue between researchers and practitioners with the goal of improving mathematics instruction, and believe a book with a similar focus could greatly benefit statistics instruction.

James H. ALBERT
Bowling Green State University

Medical Biostatistics (2nd ed.).

Abhaya INDRAYAN. Boca Raton, FL: Chapman & Hall/CRC, 2008, xlvi + 771 pp., $99.95 (H), ISBN: 978-1-5848-8887-1.

In the preface to the second edition of *Medical Biostatistics* (page xxix), the author provides the primary motivation for writing his text: "This book arises from the desire to help biostatistics earn its rightful place as a medical, rather than a mathematical, subject." Of course, some might argue that characterizing biostatistics as a medical discipline is overly simplistic, since methodological and theoretical developments in biostatistics are often built upon mathematical as well as computational foundations. Yet few would question the importance of the symbiotic relationship that exists between biostatistics and the medical sciences. Collaborative endeavors greatly benefit not only when biostatisticians possess a basic understanding of the biomedical underpinnings of their projects, but also when clinical researchers possess a basic understanding of biostatistical methods. This book is geared towards facilitating the latter. The author writes (page xxix), "The primary target audiences are students, researchers, and professionals of medicine and health."

The principal strength of this book is its comprehensive coverage, which is almost encyclopedic in breadth. Fundamental topics that would be generally found in introductory biostatistics and/or epidemiologic texts are given a thorough treatment (e.g., interval estimation, hypothesis testing, observational and experimental study designs). More advanced analytical frameworks are also covered, including linear regression, logistic regression, clinical trials, sampling, survival analysis, multivariate analysis, and quality control. In addition, the book opens and closes with intriguing chapters on medical uncertainties and statistical fallacies.

In preparing the second edition, the author has substantially increased the coverage, expanding the existing chapters and adding 7 new chapters. The current version is comprised of 21 chapters, including several that cover topics that would not be found in typical biostatistical texts, such as a chapter on clinimetrics and evidence-based medicine and a chapter on measurement in community health.

In teaching biostatistical concepts and methods to clinical researchers, a primary challenge is to promote an understanding that is accurate, nuanced, and grounded in intuition, and to do so without becoming mired in technical details. The author wisely avoids developments that are overly technical or mathematical. However, the explanations often fail to meet the preceding challenge because they are either unclear or inaccurate. In my opinion, this is the most significant weakness of the book.

In his review of the first edition, Dunstan (2002) criticized the exposition, citing "many cases where sloppiness has led to incorrect statements," and "many instances of confused and contradictory thinking." These criticisms persist with the second edition. For example, the author routinely interprets a computed $100(1-\alpha)\%$ confidence interval as an interval that contains the unknown parameter with probability $100(1-\alpha)\%$. Moreover, a p-value is alternately interpreted as "the probability that a true null hypothesis is wrongly discarded" (page 368) and "the chance that the null hypothesis is true" (page 372). (Even a Bayesian would object to the latter interpretation, since the Jeffreys–Lindley paradox indicates that the p-value and the posterior probability of the null hypothesis may lead to very different conclusions.)

In some instances, the explanations are simply unclear or lack sufficient detail. For instance, in describing the coefficient of determination (page 537), the author writes "The larger the R^2, the better is the fit." The fact that a structurally inappropriate model may lead to a high value of R^2 is not mentioned, although this very issue arises (and is resolved) in a subsequent example (pages 541–543). In describing the t-distribution for the statistic $(\bar{x}-\mu)/(s/\sqrt{n})$, the author writes (page 348), "It is true that the underlying distribution must be Gaussian for t to be valid, but it is robust to mild departures." The relationship between the sample size and the degree to which the t-distribution is robust to normality is not mentioned.

In short, *Medical Biostatistics* is written with an ambitious and laudable objective: to serve as a comprehensive and accessible biostatistics reference for clinical researchers. It achieves this objective in terms of its breadth of coverage, yet falls short in terms of the accuracy and clarity of its explanations.

Joseph CAVANAUGH
The University of Iowa

REFERENCE

Dunstan, F. D. J. (2002), Review of *Medical Biostatistics*, by A. Indrayan and S. Sarmukaddam, *Biometrics*, 58, 475–476.

Numerical Computing With Simulink, Volume I: Creating Simulations.

Richard J. GRAN. Philadelphia, PA: SIAM, 2007, xxi + 306 pp., $81.00 (P), ISBN: 978-0-898716-37-5.

Since its inception in the early 1980s, MATLAB with its following sister product, Simulink, and their accompanied toolboxes have continued a near-exponential growth amongst the scientific community (Moler 2006). The prime factors for such a remarkable success include: (a) powerful built-in computational and graphical operators; (b) open and extensible programming environment with intuitive and concise matrix syntax; and (c) a rich library of ready-to-use functions, bundled as "toolboxes," which cover a wide spectrum of science and technology. These key features have resonated very well with both research and pedagogy, and hence a surge of knowledge dissemination has been produced about or based on the MATLAB family. The developer company, MathWorks, has listed on its website 1006 English and 337 non-English manuscripts related to MATLAB/Simulink (MathWorks 2009). In this list, 124 English and 92 non-English books address some aspects of Simulink in their subject.

Numerical Computing With Simulink, Volume I: Creating Simulations is one in the above-mentioned stream of publications, with a focus on using Simulink as the basis for an integrated design environment (IDE). The content is designed to follow that of a textbook written by the creator of MATLAB, Cleve Moler, namely *Numerical Computing With MATLAB* (Moler 2004). Hence, similar to Moler's book, chapters begin with some mathematics and end with practical applications. The text is also meant to be a supplement for a course on numerical computing through including case studies and exercises in each chapter, which are based on 88 model files that are available online. Indeed, a distinguished value of the book is its realistic case studies, including spacecraft attitude control (Chapter 3), filter design (Chapter 4), and house heating system control (Chapter 7). These examples reflect the author's long engagement with R&D projects, including his five-year involvement in MathWorks as an executive consultant (see MAC 2009 for the author's biography).

The book's 306 pages are divided into 9 chapters followed by a brief discussion of some concluding remarks and a short bibliography and index. Chapters are mostly independent, each discussing a different topic related to systems analysis and design. The text is easy to follow throughout the book, albeit references to underlying theories are succinct and encapsulated. Thus, the reader is expected to have a basic understanding of the subject in each chapter, and use the text only to refresh and/or enrich his or her knowledge through practical examples. To compensate, each chapter ends with a discussion on "further reading" with suggestions for how to learn more about the topic through the literature. While a section on "further reading" is a suitable way of concluding each chapter, some (in Chapter 2, 5, 6, 7, and 9) have room to include more relevant references. With few exceptions (e.g., pp. 111, 166, 179, 185), the graphics are illustrative. The figures, almost all generated in MATLAB, are black and white, which makes those containing a multitude of fluctuating curves difficult to decipher, particularly those in Chapter 5.

Chapter 1 introduces Simulink through three interesting examples. Do not consider yourself as a master after reading this chapter, though! Readers who have worked with Simulink before or studied tutorials, such as Karris (2008), would enjoy more the remaining chapters of the book.

Chapter 2 discusses linear systems, and attempts to review the fundamentals of linear differential equations, matrix algebra, and servo control systems (including Laplace transforms and Bode plots) in 45 pages. The four Simulink case studies in the chapter are definite helpers in this dense learning journey, but the content still feels overwhelming.

Chapter 3 describes how to develop Simulink models of two nonlinear systems, the Lorenz chaotic attractor and rigid body spatial rotation. The chapter also has an instructive section on the use of look up tables and data interpolation and curve fitting in Simulink. The case study of spacecraft attitude servo control, though using a simplified approach (see Paluszek et al. 2008 for an in-depth study of the application), is a good illustration of how systems are modeled in Simulink.

Chapter 4 is aimed at providing insight into how to build Simulink models for digital systems including digital filters. Similar to Chapter 2, it tries to cover an overwhelmingly wide range of fundamentals in the digital world, from difference equations, Fibonacci numbers, and z-transforms to digital filters and signal processing, including fast Fourier transform (FFT) and phase-locked loop (PLL). The reader is cautioned, again, that this chapter cannot be used as a mere means of learning digital systems, and the theoretical content of the chapter can, at best, serve as a refresher on the subject. Aside from the theory, the chapter offers a thorough introduction of a Simulink add-on tool, called Signal Processing Blockset, which is used for design, modeling, and coding analog and digital filters.

Chapter 5 tackles another challenge of real-life system analysis, namely uncertainty in terms of both random functions and stochastic processes. This time, the author has managed to strike a balance between underlying theories and illustrative examples. In particular, the effect of white noise on continuous and discrete time systems is properly exemplified by Simulink models.

Chapter 6 is an illustration of how to convert the partial differential model of a distributed system into a set of coupled ordinary differential equations suitable for Simulink modeling.

Chapters 7 and 8 describe three useful add-on tools of Simulink, namely Stateflow (in Chapter 7) and SimPowerSystems and SimMechanics (in Chapter 8). The Stateflow toolbox, which is remarkably useful for modeling systems with complex logic and state transitions that depend upon events, deserves a book of its own to explain all its functions and features. Nevertheless, Chapter 7 can conveniently walk the reader through basic steps of building state flow model for a house-heating system. Similarly, Chapter 8 gives an illustrative introduction to Simulink object-oriented modeling tools for electric (SimPowerSystem) and mechanical (SimMechanics) systems, albeit it runs short of mentioning how these tools are compared with more sophisticated software packages, such as MSC Adams, LMS Virtual.Lab, and Dymola.

Chapter 9 is expected to be the fruit of the book, where all the previous discussions on systems modeling and analysis are put into an IDE perspective. However, 19 pages are far less than adequate to fulfill the author's promise in the book Preface "to show how it is possible for Simulink to be the basis of a complete design process." One would expect to see more in this chapter on how to link the previous materials about linear, nonlinear, digital, stochastic, and distributed system modeling (using Simulink) to design for quality, design for optimization, and design for reliability, to name a few. Moreover, the two crucial steps of most design processes, namely hardware-in-the-loop simulation and embedded code generation, deserved more than just few lines in the last two pages of the chapter. Let's hope that "Volume II" will be dedicated to various design approaches as well as methods of testing and implementation.

In summary, Gran's book is a promising portal into the domain of engineering systems, concurrent analysis, and synthesis. It tries to cover a wide range of real-life engineering problems, not only by means of some interesting applications, but also through explaining their computational bases. It may not be a book to start with for learning Simulink or any of the discussed topics, but at a price of $81(U.S.) it is surely recommended to those who want to sharpen their "Simulink skills."

M. Reza EMAMI
University of Toronto

REFERENCES

Karris, S. T. (2008), *Introduction to Simulink With Engineering Applications* (2nd ed.), California: Orchard Publications.

Mathematical Analysis Co. (2009), "Dr. Richard J. Gran, President and CEO," The MAC Web Site, available at *www.math-analysis.com/about.htm*.

Moler, C. (2004), *Numerical Computing With MATLAB*, Philadelphia: SIAM.

——— (2006), "The Growth of MATLAB and MathWorks Over Two Decades," *Cleve's Corner (in The MathWorks Newsletter)*, available at *www.mathworks.com/company/newsletters/news_notes/clevescorner/jan06.pdf*.

Paluszek, M., Thomas, S., Mueller, J., and Bhatta, P. (2008), *Space Attitude and Orbit Control*, New Jersey: Princeton Satellite Systems, Inc.

The MathWorks (2009), "MATLAB and Simulink Based Books," The MathWorks Web Site, available at *www.mathworks.com/support/books/index.html*.

A Primer on Linear Models.

John F. MONAHAN. Boca Raton, FL: Chapman & Hall/CRC, 2008, xiv + 287 pp., $49.95 (P), ISBN: 978-1-4200-6201-4.

Most of the more recent books on linear models have sought to supplement or de-emphasize the traditional "probability model" view of the subject, for example, through a geometric perspective, through visualization, through simulation and computation, or through case studies. *A Primer on Linear Models* by John F. Monahan, takes a more traditional perspective, focusing on the mathematical statistics behind linear models. It aims to be a concise and reasonably comprehensive textbook and reference work on finite sample inference for linear models.

To achieve exact finite sample inference, the discussion is restricted to linear estimators (derived using the least squares principle), and the "errors" (the variation of the response around its conditional mean) are assumed to follow a multivariate normal distribution. Properties of the multivariate normal distribution are used to obtain nearly all the results (a major exception, of course, being the Gauss–Markov theorem). Distribution theory and the language of mathematical statistics play central roles in the presentation, but the focus is exclusively on exact finite sample inference—there is hardly a mention of asymptotics nor a limit taken in the entire book.

The choice of topics seems to have been guided by the feasibility of doing exact, finite sample inference. Classical point and interval estimation, and hypothesis testing fit easily into this framework. But topics like model selection and diagnostics, where the finite sample theory is technically difficult or hard to formulate in precise mathematical terms, are covered more briefly. Some interesting topics like Mallows' C_p, Box–Cox transformations, and partial R^2 are not covered.

There are some innovations in how the material is presented that are thought-provoking even for those with a good background in linear models. Most notable among these is that many of the results are stated for nonfull rank design matrices, simplifying specification of complex ANOVA models. The sections on estimability and identifiability are insightful, and go deeper than the coverage of these topics in other contemporary books on linear models.

A theorem/proof style is followed throughout the text. Monahan's writing style is terse but uniform and clear. As one spends more time with the book, it becomes evident that there is an immense amount of information contained within its pages. However some important topics are mentioned only in passing (R^2 and prediction), only appear in exercises (ridge regression), or are excluded altogether (regression splines). The text was evidently written and edited with great care. No typographical or other errors are apparent.

This book is likely to be used more as a reference book than a textbook, although for specialized courses it could be an ideal text. Deliberately, with a few exceptions, little context or motivation for the statistical methodology is provided before moving into the mathematical results. There is an acknowledged assumption that readers already have a good understanding of the purpose and practical use of the methods being discussed. This works well for the more basic topics, as discussion of practical issues is available elsewhere, and reviewing it here would detract from the elegant and concise mathematical discussion. Mixed and random effects models, however, are conceptually difficult for reasons that are completely distinct from their mathematical challenges, and discussing these issues in isolation feels incomplete.

Consistent with most regression texts published in the last 20 years, Monahan uses linear algebraic concepts and notation to streamline the presentation. "Case space" geometry is used to some extent, but geometry does not play a central role in the presentation. Models expressed algebraically in terms of probability distributions are the dominant way that complex ideas are communicated to the reader.

Monahan's book is unique among the large number of contemporary books on linear models. Historically, it follows in the path of the books by Rao (1973) and Searle (1971) published more than 30 years ago, but with a much more modern approach. Among more recently published books, Stapleton's (1995) *Linear Statistical Models* is perhaps most similar, but Stapleton's book, while mathematically rigorous, is less theoretical and includes substantially more discussion of practical issues. The style and level of Monahan's book is not unlike that of Casella and Berger's (2002) popular text, *Statistical Inference*, and in some ways could be viewed as a massive expansion of their chapter on linear models.

Graduate students in most statistics Ph.D. programs learn about the practice of linear modeling in applied statistics courses, where theoretical justifications for inferential procedures are generally cited without providing completely rigorous proofs. The premise of Monahan's book is that comprehensively covering the theory of linear models is an excellent way to reinforce students' understanding of mathematical statistics, while also deepening their understanding of linear models.

There is merit to integrating graduate training in theoretical and applied statistics. It has been said that modern applied statistics is increasingly driven by heuristics, but linear models can be held up as an unquestionably useful tool that has been deeply characterized in theoretical terms. However linear models may be a unique case in which sampling properties can be described without asymptotic or other approximations. Students spending a lot of time with this book will become skilled technicians in a set of ideas centered on the multivariate normal distribution, but there is no thriving research area for which this specialized knowledge is essential.

An alternative perspective on the role of linear models in graduate education favors an asymptotic emphasis rather than the finite sample approach taken by Monahan. The relative simplicity of linear models makes them a useful framework for exploring some of the most fascinating research themes of modern statistics. Covariate measurement error and multivariate regression, both covered in Monahan's book, are two of these. In addition, functional regression, causal inference, and prediction and variable selection in very high dimensions can also be tackled. Arguably, insight for most of these topics can be gained more quickly through an asymptotic approach, and such an approach is more transferable to other types of regression such as classification.

Regardless of where one stands on making this material the central focus of a course, Monahan's book will be valuable to almost any graduate student in statistics. The book can serve as a supplemental text in a broader course on regression analysis, or as a reference book for researchers in a variety of fields employing linear estimation techniques. The exercises are particularly impressive—they are extensive and varied, and reveal a lot of insight into the subject that would be too cumbersome to cover in the text itself. Monahan and CRC Press should be commended for making such a high quality book available at a very reasonable price.

Kerby SHEDDEN
University of Michigan

REFERENCES

Casella, G., and Berger, R. L. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Thomson Learning.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: Wiley.

Searle, S. R. (1971), *Linear Models*, New York: Wiley.

Stapleton, J. H. (1995), *Linear Statistical Models*, New York: Wiley.

Probability and Statistics by Example: Volume 2, Markov Chains: A Primer in Random Processes and Their Applications.

Yuri SUHOV and Mark KELBERT. New York, NY: Cambridge University Press, 2008, ix + 487 pp., $140.00 (H), ISBN: 978-0-521-84767-4. $60.00 (P), ISBN: 978-0-521-61234-0.

The book consists of three main chapters, one on discrete and one on continuous time Markov chains, and one on statistics for chains in discrete time. The first two have substantial sections of about 30 and 40 pages respectively where examination questions and their solutions are collected. The authors state as one of their important goals "to give the general public a clearer picture of how probability- and statistics-related courses are taught in a place like the University of Cambridge." However, somewhat to the disappointment of the reviewer, much of the material is rather routine, often involving little more than putting numbers into formulas, or working out some matrix multiplications. The "applications" mentioned in the subtitle are mostly little stories that directly lead to a Markov chain problem such as fleas hopping about on the vertices of a triangle, or the seating problems in a rock concert. The book is sprinkled with puns of the type "weapons of math destruction," "Bayesian instincts," "secretaries do it without any problems," etc., but Shakespeare also appears frequently in the citations. There are several enjoyable stories about some of the many outstanding Russian probabilists. In an epilogue, of well-written form and touching content, Markov himself is the main character.

There are some highlights, such as hidden Markov models, Dirichlet forms, large deviations, and the Bayesian analysis of Markov chains, but there is no discussion of martingales and their use in connection with Markov chains. As the authors point out in their preface, the theory of Markov chains has seen a lot of development in the last few decades. Still, the book largely preserves the traditional approach that, in its extreme form, reduces the theory to an application of linear algebra and matrix calculus. In the reviewer's opinion, better books are available, such as Resnick (1992) for undergraduate teaching, or Brémaud (1999) for applications of Markov chains (and Williams 1991 or Körner 1988, for example, if it comes to Cambridge products). Also, as a "primer in random processes and their applications," a book of about 500 pages could or perhaps even should have treated some other stochastic processes too.

The above notwithstanding I did enjoy browsing through this somewhat baroque book.

Rudolf GRÜEBEL
Leibniz Universität Hannover

REFERENCES

Brémaud, P. (1999), *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*, New York, Springer.

Körner, T. W. (1988), *Fourier Analysis*, Cambridge: Cambridge University Press.

Resnick, S. I. (1992), *Adventures in Stochastic Processes*, Boston: Birkhäuser.

Williams, D. (1991), *Probability With Martingales*, Cambridge: Cambridge University Press.

The Role of Statistics in Business and Industry.

Gerald J. HAHN and Necip DOGANAKSOY. Hoboken, NJ: Wiley, 2008, xvii + 344 pp., $84.95 (P), ISBN: 978-0-471-21874-6.

The Role of Statistics in Business and Industry provides an overview of the vital role statistics plays in everyday decisions in business, manufacturing, and industry. The book addresses a wide range of problems starting with product development and manufacturing to quality control and reliability. This book is not just of interest to statisticians, but may be even more valuable to professionals outside of our area, such as managers and engineers, who are frequently involved in data-based decisions. The book can also serve as a resource for aspiring statisticians, providing insight into the daily work of our profession.

In Chapter 11, the authors state that "It is impossible, in a single chapter, to do justice to the many applications of statistics to financial services." Of course, it is equally impossible, in a single book, to do justice to the many applications of statistics in business and industry. But, that being said, the authors have done a remarkable job of taking on the impossible. It is not the book's intention to provide deep technical details about any of the statistical procedures presented, but rather to elucidate their usefulness in informative day-to-day business decision making. Nevertheless, the book does not oversimplify statistical issues, but when appropriate, provides references for further reading. Some general features that enhance the book are its sidebars (gray-shadowed boxes) which concisely summarize important concepts and principles. Every chapter of the book concludes with a summary of "Major Takeaways" and a set of useful and stimulating discussion questions. I appreciate the authors' strategy to motivate statistical tools by first posing the business problem or industrial question of interest.

The book is divided into three parts: "Setting the Stage," "Manufactured Products and Applications," and "Other Applications." Each part consists of several chapters. Part I of the book (Chapters 1–2) gives an overview of the role of statistics and the work of statisticians. Historical notes and facts, such as the beginnings of quality control with W. E. Deming, or W. S. Gossett and the t-distribution, are shared with the reader. The authors illustrate how our profession has evolved over recent decades and challenge statisticians to function beyond being as mere data analysts and taking a proactive role in their professional environment.

Part II of the book (Chapters 3–9) addresses fundamental ideas of product development and manufacturing. General aspects of quality assurance and reliability are outlined. Perhaps not surprisingly, this part begins by describing effective product design (Chapter 3). A discussion of essential ideas, such as desired target performance of a product related to eliminating bias and reduction of product variability follow. The use of simulation and propagation of error in the assessment and verification of desired product characteristics are introduced and well illustrated by examples (e.g., for products consisting of multiple independently manufactured parts). Concepts in Chapter 3 are demonstrated through a case study in Chapter 4. Often textbooks do not have room for such detailed cased studies and I feel that this chapter as well as another case study in Chapter 7 add great value to the book. The reader will also appreciate the presentation on design validation and transition of a product into manufacturing. Chapters 5 and 9 focus on general reliability concepts. For example, the book discusses the different possible failure modes and the different types of reliability estimation, such as accelerated life testing or degradation testing. System reliability and the aspect of in-house versus field reliability are emphasized and illustrated through interesting examples.

The book concludes with four additional chapters in Part III. These chapters are contributed by authors with expertise and years of experience in a specific field. Chapter 10 provides insight into the work of biostatisticians and touches on the general challenges and regulations for product development in pharmaceutical industries. A concise overview of clinical trials and their different phases are presented and supplemented with a case study for a Phase 2 trial. Subsequent chapters (Chapters 11–12) focus on applications in financial services and business processes, where decisions and uncertainty assessments have large-scale monetary consequences. The authors explain how statistical modeling plays a crucial role in financial applications, and they discuss the important roles of variable selection and model validation. A number of interesting examples are provided, and I especially liked one which illustrated credit scoring for customer loan approval. The concluding Chapter 13 of the book gives a brief overview of several additional areas of applications where the expertise of the statistics profession is needed. For example, the authors explain uses of statistics in the food and beverage industry, in medical and security services, and in the communications industry.

My current position involves a lot of teaching (often business students), and I am eager to try this book in my classrooms. The book's diverse applications could potentially make the incorporation of the entire book into a single course challenging; however, Part II of the book can easily be implemented in graduate or higher level undergraduate courses on quality control or reliability, particularly if the audience includes engineering students. Chapter 10 on pharmaceutical products could potentially be used to augment applications in a biostatistics course and to provide an overview of what is expected from a biostatistician in that industry. Chapters 11 and 12 provide an excellent source for a general course in business statistics (both at an introductory or MBA level). Generally speaking, there are parts of the book that would provide a nice supplement to many statistics courses.

Ulrike GENSCHEL
Iowa State University

Sampling Techniques for Forest Inventories.

Daniel MANDALLAZ. Boca Raton, FL: Chapman & Hall/CRC, 2008, xv + 256 pp., $83.95 (H), ISBN: 978-1-5848-8976-2.

In the preface, Mandallaz states, "The objective of this textbook is to provide graduate students and professionals with the up-to-date statistical concepts and tools needed to conduct a modern forest inventory." In his short text, Mandallaz has generally met this objective. As such, the text is an excellent addition to the forest and natural resource inventory literature, and nicely compliments classic and contemporary texts focused broadly on applied concepts and tools but lacking in rigorous statistical treatment. Mandallaz's mathematical approach requires users to have a firm grasp on elementary probability theory and matrix algebra. For graduate students, a two semester sequence in probability theory or mathematical statistics with some forestry or natural resource domain knowledge should suffice. Unfortunately, these requirements exclude many of the students in these more applied fields.

The text presents information in a logical progression beginning with basic sampling theory, and then moves into advanced topics and areas of active research. Relevant exercises close each chapter, and solutions are provided for a subset of these problems. In the later portion of the text, the author devotes two chapters to the analysis of real forest inventory data. The first offers a comparison of various estimators using inventory data from the Zürichberg Forest, and the second details the application of optimal sampling schemes to the Swiss National Forest Inventory.

Chapter 2 addresses fundamental concepts and techniques of design-based survey sampling and inference for finite populations. The emphasis is inclusion probabilities and the Horvitz-Thompson estimator. Rather than exhaustively covering special cases of the various estimators, Mandallaz chooses to detail the fundamental theoretical concepts and offer detailed derivation of the core estimators (a trend that continues in subsequent chapters). This approach equips the reader with the insight necessary to recognize special cases and either construct the appropriate estimator or more fully understand the forms given without derivation in supporting texts.

Chapter 3 details some advanced survey sampling topics including an in-depth discussion on three-stage element sampling which nicely picks up where comparable texts leave off at two-stage or a sketch of multistage sampling (e.g., Särndal, Swensson, and Wretman 1992; Gregoire and Valentine 2007). Here, Mandallaz introduces the super-population model and subsequent distinction between model-assisted versus model-dependent inference offering a prelude to a deeper look into these topics in Chapter 6. Although he provides a very detailed treatment of this and other contemporary topics, Mandallaz misses opportunities to expand the discussion by including key references on this topic (e.g., Gregoire 1998).

Chapters 4 and 5 focus on topics germane to forest inventory, detailing one and two-phase sampling schemes, respectively. These chapters represent a tour of Mandallaz's early contributions to the field. Here, a contemporary perspective is provided by casting the exposition of these methods within the infinite population model (commonly referred to as Monte Carlo approach). Now, within the model-based paradigm, Mandallaz offers a detailed discussion on specifying residual spatial processes, which sets the stage for the subsequent chapter on classic theory of Kriging models. A brief sketch of the *small-area estimation problem* is then presented, but only for settings where ground-based inventory data is available for the small areas of interest. Review of linear models and detailed contingency tables and associated test statistics are presented as tools to more fully explore functions of univariate and multivariate response variables.

The overview of geostatistics in Chapter 7 is terse; however, it is a necessary and important addition typically not found in other forest sampling texts. Following their introduction, ordinary, universal, and double Kriging methods are illustrated in the capstone case study in Chapter 8. There is substantial opportunity here to extend the discussion and self study by including relevant literature (e.g., de Gruijter and ter Braak 1990; Brus and de Gruijter 1993; Diggle, Tawn, and Moyeed 1998).

Continuing to leverage the model-based paradigm, Mandallaz covers anticipated variance for optimal sampling schemes based on a local Poisson model in Chapter 9. This chapter is followed by an illustration of optimal sampling using the Swiss National Forest Inventory data. The final two chapters cover estimation of change and growth and line intersect-sampling, respectively.

In summary, this text is an important contribution to the literature, as it provides a unique, mathematically rigorous, tour of classical and modern topics on forest sampling theory. Of particular importance is the text's emphasis on model-based methods which are becoming increasingly important within forest and natural resources inventory.

Andrew O. FINLEY
Michigan State University

REFERENCES

Brus, D. J., and de Gruijter, J. J. (1993), "Design-Based versus Model-Based Estimates of Spatial Means: Theory and Applications in Environmental Soil Science," *Environmetrics*, 4, 123–152.

de Gruijter, J. J., and ter Braak, C. J. F. (1990), "Model-Free Estimation From Spatial Samples: A Reappraisal of Classical Sampling Theory," *Mathematical Geology*, 22, 407–415.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350.

Gregoire, T. G. (1998), "Design-Based and Model-Based Inference in Survey Sampling: Appreciating the Difference," *Canadian Journal of Forest Research*, 28, 1429–1447.

Gregoire, T., and Valentine, H. (2007), *Sampling Strategies for Natural Resources and the Environment*, New York: Chapman & Hall.

Särndal, C., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer.

Semiparametric Regression for the Social Sciences.

Luke KEELE. Hoboken, NJ: Wiley, 2008, xvi + 213 pp., $80.00 (H), ISBN: 978-0-470-31991-8.

There is a variety of books on nonparametric and semiparametric regression. The level of mathematical difficulty varies immensely from one book to the other: from the measure-theoretic exposition of Kosorok (2008) or Bickel et al. (1998) to the more accessible text of Ruppert, Wand, and Carroll (2003), which keeps the "methodological technicalities to a minimum, so that the book is suitable even for students or working scientists with only a moderate background in regression" (Fahrmeir 2004). But to be completely readable, even Ruppert, Wand, and Carroll (2003) requires a level of comfort with the matrix formulation of regression models. There are few, if any, texts on semiparametric regression that are suitable for a reader who is not interested (or comfortable) with this level of mathematics; that is, until *Semiparametric Regression for the Social Sciences* by Luke Keele. According to the Preface, *Semiparametric Regression for the Social Sciences* is a text for "analysts from the social sciences who would like to extend their statistical toolkit beyond standard parametric function forms." A quick glance through the book immediately reveals the book was not intended for an audience who desires mathematical detail.

Semiparametric Regression for the Social Sciences is an introductory guide to smoothing techniques, semiparametric estimators, and related methods. The book describes the methodology, and illustrates it using carefully explained examples and datasets. It demonstrates the potential of these techniques using detailed empirical examples drawn from the social and political sciences. Each chapter includes exercises and examples and there is a supplementary website containing all the datasets used, as well as computer code in *S-Plus* or *R*, allowing readers to replicate every analysis reported in the book.

Keele's goal in writing *Semiparametric Regression for the Social Sciences* was to "make applied analysts more sensitive to which functional form they adopt for a statistical model." Overall, Keele's approach to semiparametric regression is a gentler approach than Ruppert, Wand, and Carroll (RWC). The notation between the two books is essentially the same. However, Keele's level of mathematical detail is almost nonexistent compared to RWC. I can best explain by describing how I've been using Keele's book in my graduate level course, Semiparametric Regression Models.

In that course, I require the text of Ruppert, Wand, and Carroll (2003). However, the mathematical background of the students in that class is quite varied. I have found this book by Keele to be especially helpful to those less mathematically sophisticated students in their developing a good intuitive understanding of the mathematics and concepts presented in RWC. The read is an easier one for them. The notation in the two books is very similar, and Keele often refers to the more advanced text (RWC).

One caveat to recommending this book is that some of the the terminology used by Keele may result in some confusion—especially if it is being used as an "easier" supplement to texts with a higher level of mathematical detail. For instance, in Chapters 2 and 4, in his discussion of nonparametric modeling choices, Keele warns that the choice of the "span" is the most critical choice an analyst must make when using nonparametric regression. He goes on to say that the choice of the "span" affects the trade-off between variance and bias. For students who know about the "span" of a set of vectors, but who are yet to learn about the effect of the choice of the smoothing parameter, this unfortunate choice of words might initially lead to some confusion. However, a careful reading of Keele's discussion should prevent it.

For my own purposes, Keele's text provides good examples that compliment those in other texts. With a few exceptions like the one noted above, I liked this book. I would recommend it to students in a class on semiparametric regression as a "gentle" treatment of the subject matter. To teachers of the same subject area, I might recommend it as a supplement or aid that provides additional, easily accessible applications. All in all, for the type of book it is—a low-level treatment of semiparametric regression techniques—it is a useful book.

Jane L. HARVILL
Baylor University

REFERENCES

Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer-Verlag.

Fahrmeir, L. (2004), Review of *Semiparametric Regression*, *Biometrics*, 60 (1), 840.

Jacobson, G. C., and Dimock, M. (1994), "Checking Out: The Effects of Bank Overdrafts on the 1992 House Elections," *American Journal of Political Science*, 38, 601–624.

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, New York: Cambridge University Press.

Software for Data Analysis: Programming With R.

John M. CHAMBERS. New York, NY: Springer, 2008, xiv + 498 pp., $79.95 (H), ISBN: 978-0-387-75935-7.

As a statistician and R enthusiast, already familiar with the previous volume by Chambers (1998) on the S language, it is a pleasure to read the sequel. Chambers' new book explores the popular open source language that is quickly becoming the international standard in statistical computing software. Many readers of the 1998 book on S were using S-Plus version 5.0 and later, and eventually many readers were using R. *Software for Data Analysis: Programming With R* illustrates how to explore and analyze data via trustworthy software, and describes the design and evolution of the R system in the context of the S language. There is throughout a clearly organized approach to presenting concepts at a variety of levels (user, programmer, contributor) and the quality of exposition is truly excellent. Anyone in the R community, whether statistician, researcher, programmer, or developer, will find that this book is a valuable resource. Most who are new to R or S-Plus, however, will probably need a more basic presentation to read first or as a supplement.

New and expanded topics covered in this book include chapters devoted to graphics, text computations, creating packages, and interfaces to other systems. Material on objects, functions, classes, methods and generic functions is reorganized and specific to R. Many examples are included and easily referenced under Examples in the index. There is an appendix on the history of S and a handy index of classes and types. A companion R package SoDA (Chambers 2008) is available that includes the functions and examples from the book.

After an introduction and chapter on the basics of programming with R, Chapter 3 provides an overview of the structure and design of the R language. Functional programming, function calls and function objects, some language details as well as methods for interactive debugging are discussed. There is a blend of introductory notions such as techniques for writing functions, as well as more abstract or technical concepts such as functional programming languages. It is explained why R for the most part could be considered a functional programming language, in contrast to a procedural language like C. The chapter concludes with an accessible but detailed discussion of interactive debugging with browser() and trace() and error handling utilities.

Chapter 4 on R Packages describes the tools available in R to build, check, and install packages. Preparation of documentation in Rd format is carefully explained. This chapter includes many tips that even experienced R package authors may find useful; for example, prompt() and its relatives, including promptAll() and packageAdd() utilities in SoDA to aid in writing new packages. Chapter 11 "Interfaces I: C and Fortran" covers techniques for building functions that may be called from R, and how to use R functions like random number generators and memory management utilities from within the external functions.

Chapters 5 and 6 comprising over 125 pages, present objects and many of the basic concepts of the R language and syntax. The material is presented at several levels. For example, there is a discussion of random number generators, section on efficient programming techniques such as vectorizing, apply() with examples, and data.frame objects in the context of other software such as spreadsheets and SQL. A section on programming simulations includes "Reproducible and repeatable simulations" with an example. Chapter 7 covers data visualization and graphics in R, with sections devoted to the graphics, grid, and lattice packages.

All of Chapter 8 is devoted to text computations in R and other systems. This chapter is a well written and accessible introduction that is largely missing from most other books on R. A section is devoted to writing and using Perl. The topics range from regular expressions and character classes to tasks such as importing text data and parsing dates. Some of the interesting examples given include merging documentation sections in Rd files, removing terms from a model formula, and using Perl to count chunks of text.

Chapters 9 and 10 cover classes, methods, and generic functions. Topics include creating new classes and methods, inheritance, class unions, programming with S3 classes, programming techniques and documentation for methods, generic functions, and much more. Following Chapter 11 on interfaces to C and Fortran is Chapter 12 on interfaces to other systems such as other languages, functional interfaces, databases and spreadsheets. The final chapter "How R Works" discusses topics such as the R program, R evaluator, assignments and replacements, and memory management.

This is a book that will appeal to readers of diverse backgrounds. For R users it has a wealth of information on learning to use R effectively; from efficient and reliable programming to writing packages. It is an authoritative reference for programmers and developers. It is the type of book that will be referenced often, as the reader's experience with R, level of expertise and interest in programming grows.

Maria L. RIZZO
Bowling Green State University

REFERENCES

Chambers, J. M. (1998), *Programming With Data: A Guide to the S Language*, New York: Springer.

——— (2008), "SoDA: Functions and Examples for *Software for Data Analysis*," R Package Version 1.0-3.

Statistical Practice in Business and Industry.

Shirley COLEMAN, Tony GREENFIELD, Dave STEWARDSON, and Douglas C. MONTGOMERY (eds.). Hoboken, NJ: Wiley, 2008, ix + 433 pp., $110.00 (H), ISBN: 978-0-470-01497-4.

Statistical Practice in Business and Industry (SPBI) is a handbook, of sorts, for statistical practitioners. It contains an informative and well-written collection of papers by over 30 experts in statistical methodology from around the world, assembled as a result of the European Commission funded Fifth Framework project entitled European Network for Promotion of Business and Industrial Statistics (pro-ENBIS). The goal of the book is to advance the use and understanding of appropriate statistical methods and to provide a source for the professional development of statisticians working in business and industry.

SBPI begins in Chapter 1 with a brief introduction and a history of statistics in industry, including a discussion of the evolution of mass fabrication, and an overview of the contributions of statistics and statisticians to innovation and industrialization, with special attention to efficient manufacturing, Total Quality Management, and Six Sigma. This chapter builds the case that the key to innovation and competitiveness in the 21st century will be the effective use of statistical methods and tools. The focus of the remaining chapters is either on the interpersonal skills (behaviors, communication styles, personality traits, etc.) that enable the statistical practitioner to communicate with nonstatisticians and successfully participate as a member of a diverse project team, or on the technical statistical skills that most statistical practitioners in business and industry should have in their toolkit.

Chapter 3 delves into statistical consulting. There are several interesting papers on topics that range from the importance of personality and behaviors industrial statisticians can use to work more efficiently and effectively, to how a company can identify Black Belt candidates, to how universities can form commercial or noncommercial statistical consulting units. As anyone who has worked in industry knows, management decisions are ultimately based on the bottom line, and every employee at every level must demonstrate their value and contribution to the company. Chapter 4 offers valuable insight into how statisticians can measure their contribution to projects using *Practical Statistical Efficiency* (PSE). In Chapter 5 the authors discuss management statistics as the foundation for a systematic and informed decision-making process and performance evaluation at all levels of the organization, from the strategic level to the operational level.

Chapters 6–14 focus on the set of specific technical skills and statistical methods that are used to answer a wide variety of business and industrial questions. Chapter 6 illustrates a variety of techniques to improve service quality, including the use of observational studies, surveys, penalty/reward analysis, vignette method, and complaint management, to name a few. Chapter 7 gives a nice discussion on the types of designed experiments most frequently used in industry, including response surface designs, split-plot designs, fractional factorials, and optimal and computer-generated designs. Data mining in business and industry is taken up in Chapter 8 including a brief how-to guide. Chapters 9 and 10 provide a thorough discussion of statistical process control, including process improvement, and the phases of process monitoring in both univariate and multivariate control charts. Chapter 11 contains a complete guide to measurement systems analysis. The reader will glean from this material an understanding of the importance of ideas like repeatability, reproducibility, bias, and uncertainty, and measurement systems analysis tools such as Gauge repeatability and reliability studies. Chapter 12 is comprehensive look at reliability statistics. Chapter 13 is dedicated to the analysis and control of more complex multivariate systems and multiscale data. Statistical methods used for multivariate process monitoring and signal and image processing include principal component analysis (PCA), Partial Least Squares (PLS), wavelets, and multiscale principal component analysis (MSPCA). Chapter 14 presents a nice introduction to the use of computer simulations as a means for improving the performance of a system or product. Coverage includes Markov chain Monte Carlo methods along with several case studies.

Finally, once an experiment or study is carried out and the data analyzed, how does one effectively communicate the results to an audience of lay people, many of whom are the company decision makers? Chapter 15 suggests ways to successfully communicate research/project results, including audience analysis, the different mediums one can use (reports, journal articles, conference presentations, poster presentations, popular media articles, web reports, etc.), and the techniques of good writing, style, etc.

In summary, *Statistical Practice in Business and Industry* meets its goal of providing a guide to a comprehensive set of statistical methods used in business and industry. Written by experts in their respective fields, it is a one-stop reference to almost every statistical tool that one might have occasion to use if working in industry. I wholeheartedly recommend it for statisticians, as well as engineers, managers, basic scientists, and other technical personnel who find themselves serving in the role of statistical "expert" within their company. It is a concise, well-written reference to a multitude of statistical methods and interpersonal skills that every statistical practitioner in business and industry should have sitting on his/her desk.

Karen J. McGaughey
California Polytechnic State University

Statistical Reasoning for Everyday Life (3rd ed.).

Jeffrey O. Bennett, William L. Briggs, and Mario F. Triola. Boston, MA: Pearson Addison Wesley, 2009, xviii + 458 pp., $101.33 (P + CD), ISBN: 978-0-321-28672-7.

Statistical Reasoning for Everyday Life is designed for a first course in statistics. The authors effectively present statistical concepts in an intuitive manner that is both engaging and accessible to undergraduates majoring in "less quantitative" disciplines. The text favors a statistical literacy approach, emphasizing conceptual understanding of methodology instead of development of specific statistical formulas that would necessitate greater mathematical skill. A strength of this text is its consistent use of attention-grabbing issues to introduce and enrich core statistical concepts. Judicious choice and clear presentation of graphs and other figures augment the written text. The importance of being able to reason statistically, to be a critical thinker and decision maker, is clearly the theme carried throughout the text.

The book consists of 10 chapters. The first chapter presents a "big picture" overview of statistics explaining the relationship of samples to populations, various sampling methods, types of statistical studies, and guidelines for critically evaluating a statistical study. Chapter 2 introduces data types, measurement scales, types of errors, and a less-often-seen discussion of the uses of percentages in statistics, including index numbers. In addition to the usual introduction to graphical representations (construction of frequency tables, frequency bar graphs, pie charts, histograms, stem-and-leaf plots, and line graphs), Chapter 3 includes an extensive section of *Graphics in the Media* (variations of the aforementioned graphs in addition to geographical displays of data) along with a cautionary discussion of perceptual distortions. The focus of Chapter 4 is to describe data: shapes of distributions, measures of center and variability, and the somewhat surprisingly grouped topic of statistical paradoxes analyzed through two-way tables. In Chapter 5, basic characteristics of the normal distribution and the 68–95–99.7 rule are thoroughly explained; however, the transformation of a data value to its standard score is presented without mathematical or intuitive development. The authors carefully expound that a z-score is not only a stepping stone to finding accompanying percentiles, but itself a measure of deviation from the mean, a notion often understated in texts and overlooked by students. Unlike many texts where the chapter on probability appears disjointed from the descriptive statistics and inferential statistics chapters that it is attempting to bridge, the authors forefront discussion of quantifying statistical significance clearly and immediately connects Chapter 6 *Probability in Statistics* to statistical inference in the remaining chapters. Comparison of approaches to finding probabilities (theoretical, relative frequency, subjective) is discussed as is finding the probability of the complement and the Law of Large Numbers. The probability of the union and probability of the intersection, topics commonly addressed in other introductory statistics texts, are contained in a supplementary section. The emphasis of Chapter 7 is on making sense of correlation; numerical and graphical comparisons, effect of outliers and inappropriate grouping, and a pointed discussion that correlation does not imply causation. The authors give but a nod to actually calculating the correlation coefficient in an optional subsection. Chapter 8 presents several illustrated examples to develop understanding of sampling distributions of the mean and proportion, leading to discussions of margin of error and 95% confidence intervals. Conceptual foundations of hypothesis testing are carefully built in Chapter 9, including an extended discussion on statistical significance (and p-values) first introduced in the probability chapter. The final chapter extends discussion of hypothesis testing with t-tests, inferences from two-way tables, and ANOVA. True to their conceptual approach, the authors emphasize interpretation and use of statistics: Formulas supporting more sophisticated topics are intuitively justified, or in the case of the standard deviation of the sample proportion, simply stated.

The current edition of *Statistical Reasoning for Everyday Life* includes significant revisions of the last two chapters. The topics of statistical paradoxes and risk and life expectancy, formerly found in the final chapter, have been moved to earlier chapters, allowing the last two chapters to be exclusively devoted to hypothesis testing. To maintain its focus on relevance to everyday life, many of the examples, discussions, and case studies have been updated. The authors report that they have replaced 63% of the exercises and have revised data in most of the others. Each section includes student exercises grouped under three categories: Statistical Literacy and Critical Thinking, Concepts and Applications, Projects for the Internet and Beyond. At the end of each chapter are additional Chapter Review Exercises, a Chapter Quiz, Focus Studies (global warming, reliability of DNA, safety of genetically modified food) and a *Using Technology* feature. This new feature assists the reader in using a wide variety of technology (SPSS, Excel, STATDISK) to calculate and construct graphs relevant to the chapter. Also new to this edition, and expanding upon the *Using Technology* feature, is a companion CD that contains a printable *Technology Manual and Workbook* written by Triola, statistical software, and programs for the TI-83/84 Plus calculator. The chapters of the manual are aligned to those of the text and include instructions for STATDISK (software designed by Triola and downloadable from the CD), Excel, Minitab, SPSS, and TI-83/84 Plus Calculator.

I found the technology manual to be a valuable resource for the software and an informative comparison across software.

In summary, the authors are to be commended for their straightforward approach making statistics relevant to everyday issues and for their clear presentation of concepts. I recommend this text for adoption in introductory courses whose focus is that of statistical reasoning without rigorous mathematical development of statistical formulas.

Linda L. COOPER
Towson University

Statistical Thinking for Non-Statisticians in Drug Regulation.

Richard KAY. Hoboken, NJ: Wiley, 2007, xviii + 276 pp., $80.00 (H), ISBN: 978-0-470-31971-0.

Statistical Thinking for Non-Statisticians in Drug Regulation is a new text from Richard Kay that covers a broad range of topics. There are 16 chapters, some of which are primarily conceptual, with the remainder topically methodological. The book starts off with a well-reasoned, appropriate objective: to foster successful interactions between nonstatisticians and their statistical counterparts, as well as to educate industry researchers to use a critical eye when reading published research. The author does, however, give the inaccurate impression that statistical theory development is confined to universities and other academic settings. Numerous industry statisticians contribute significant amounts to the biostatistics literature and are extremely active in professional circles.

The author provides a nice overview of randomization techniques, and emphasizes pragmatism in two important ways: first, he provides straightforward examples to illustrate the technique; second, he provides quotes from relevant sections of the ICH guidelines to give both practical advice and regulatory perspective. Both of these are extremely important for the target audience. Important topics covered in this text include analysis of datasets (which includes a discussion on missing data), multiple testing, equivalence and noninferiority designs, and meta-analyses. All of these are covered in a skillful manner that is comprehensive from the perspective of the intended audience.

The author describes methodology in a clear, if not dry, manner. The illustration of the interpretation of a 95% confidence interval, as an example, follows a method that isn't necessarily novel, but it is very effective in describing the utility of such a calculation. In some places the logic did not seem to flow as well as it could; for example, the notion of a p-value (with mention of null and alternative hypotheses) is first described in Chapter 3, whereas hypothesis testing along with Type I and Type II errors were left to Chapter 8. This seemed unnecessarily confusing and difficult for someone hoping to learn the basics of hypothesis testing.

It is disappointing that the author does not include a more comprehensive discussion of adaptive trial designs, as this is a very popular area of application in the pharma industry today, and much misunderstanding on the part of nonstatisticians persists (particularly the notion that adaptive trial designs are a panacea for all that ails the industry, and that they come without a cost).

There was, in a number of places, a progression of the described calculations, to a point beyond which, we surrender control to the "computer." The author's attempt to keep material uncluttered by surrendered calculations is noble, but such characterizations might cause some readers to continue to view statistical analyses as a complex web of black boxes and hand waving.

Overall, this is a very well-written text that provides excellent, layman-level information with a proficient use of both textual and graphical descriptions. His use of examples from publications, as well as excerpts from ICH, CPMP, and FDA guidelines, are extremely helpful in characterizing the motivation behind certain techniques and strategies. The flow of ideas can be a bit unorthodox at times, but it is still easy to follow and would not likely lose its audience via overcomplication.

Christopher ASSAID
Merck Research Labs

REFERENCE

ICH E-9 (1998), "Note for Guidance on Statistical Principles for Clinical Trials," guideline, International Conference on Harmonization.

Time Series Analysis: Forecasting and Control (4th ed.).

George E. P. BOX, Gwilym M. JENKINS, and Gregory C. REINSEL. Hoboken, NJ: Wiley, 2008, xxiv + 746 pp., $135.00 (H), ISBN: 978-0-470-27284-8.

Let me begin by saying that I am honored to be asked to review the most recent edition of this classic and truly groundbreaking book on time series analysis. Certainly no other book has had such profound influence on statistical time series analysis. Although I am pleased and excited to have a new edition of this book for my library, my enthusiasm is dampened by the untimely death of Greg Reinsel prior to the printing of this most recent version. A nice tribute to Gwilym Jenkins and Greg Reinsel is included in the preface to the fourth edition by the sole living author, George Box. Professor Box acknowledges that the new material was contributed almost exclusively by Reinsel.

I have taught out of earlier versions of this book (all published by Prentice-Hall) for many years and was planning to use it in the Spring 2008 semester, before I came to the stunning realization that the book was out of print. I was pleased to learn that in fact Wiley was publishing the fourth edition but was disappointed by the fact that it did not come out early enough for my class. I have always found this text to be a very readable book that works well as a textbook for an introductory graduate course in time series analysis. Earlier versions of the book included datasets that have in themselves become classics in the field of time series analysis. These datasets are included (as far as I can tell, without changes) in the new edition.

Because of the popularity and wide distribution of earlier editions of this book, I will restrict my comments to the material that is new to the fourth edition. The fourth edition represents a substantial revision with about 100 new references and over 100 pages of new material. The major addition is the inclusion of two new chapters: Chapter 10—Nonlinear and Long Memory Models and Chapter 14—Multivariate Time Series Models. Additionally, new material has been added throughout the text, some of which involves expanded discussion of state-space models and tests for unit roots that were both new topics in the third edition. The fourth edition also includes new material on regression models with time series error terms, estimation of missing values in ARMA processes, and feedforward control. As with previous editions, the book includes problems which tend to be useful and workable. Except for a few exceptions the problems have been largely unchanged except for the addition of problems for the two new chapters.

The addition of the two new chapters makes this book appropriate as a text for a two-semester graduate course in time series analysis. Chapter 10 provides several useful alternatives to the ARIMA(p, d, q) class of models that are the focus of the earlier chapters. In particular this chapter discusses ARCH/GARCH models, nonlinear models, and long-memory models. ARCH and GARCH models have proven to be very useful in econometrics for purposes of dealing with volatility by considering an error process whose squares are autocorrelated. This chapter contains the necessary background information as well as a discussion on model building and parameter estimation. As an example, they show that the squared errors for the IMA(0, 1, 1) model previously fit to Series B (IBM stock prices) are significantly autocorrelated, and thus that an ARCH/GARCH model is appropriate. The brief section on nonlinear models includes discussion of bilinear and threshold models along with models with time-varying coefficients. The authors discuss several tests for nonlinearity and illustrate the techniques using the logarithm of the Canadian Lynx data. They show that nonlinearity is indicated, and fit a threshold AR model from which they calculate forecasts. The final section in this chapter covers the fractionally integrated ARMA process associated with a factor $(1 - B)^d$ where d can take on fractional values. They discuss model characteristics, parameter estimation, and forecasting. These techniques are illustrated by fitting a fractionally integrated model to Series A. Chapter 14 is a very well-done treatment of multivariate time series analysis. This 45-page chapter covers the basics of stationary multivariate models in the time and frequency domains. The chapter introduces vector ARMA models and their properties, and discusses model identification, estimation, and forecasting in a style consistent with their previous treatments of ARMA models. The authors also discuss nonstationary vector ARMA models (due to unit roots) and tie this very nicely to the testing for cointegration. Discussion is also given to the state-space form of the vector ARMA model.

In summary, this book is a must-have for someone who uses earlier editions of the book. I must say that being someone who was essentially raised on Box and Jenkins, this new edition is valuable in that it uses BJR-type notation andgives a classic BJR-type treatment to the new material. I highly recommend this as a reference book as well as a very teachable textbook.

Wayne WOODWARD
Sonthern Methodist University

Validating Clinical Trial Data Reporting With SAS.

Carol I. MATTHEWS and Brian C. SHILLING. Cary, NC: SAS Institute, Inc., 2008, xii + 206 pp., $36.95 (P), ISBN: 978-1-59994-128-8.

The first time a statistical programmer works with clinical trial data can be a daunting task. Given the highly regulated environment of clinical trials, one might imagine that the data will be orderly and consistent. Unfortunately, clinical trial data reflect the complexity of managing human subjects: schedules are not always kept as planned, responses can be unanticipated, and human error can occur. Fortunately, Matthews and Shilling have written this great introductory book to guide a novice through the unexpected twists and turns of clinical trial data and the difficulties of data validation and reporting.

The authors set the stage for clinical data validation with a brief overview of the regulations, guidelines, and industry standards that govern the execution of clinical trials. Through clear, concise explanations, Matthews and Shilling provide a framework for validation from a general approach, such as the pros and cons of first level versus second-level validation, to more specific details including validation checklists. A thorough discussion of the merits of program documentation with useful practical tips provides good advice to anyone performing SAS programming. The authors continue with a chapter of general techniques to facilitate validation that is filled with SAS programming examples. The authors emphasize the practical with basic SAS procedures such as PROC PRINT and PROC FREQ but leverage their efficiency with SAS options that can facilitate checking. They provide sound advice on the use (and overuse) of SAS macros as part of the data validation process as well as tools to validate and debug macros. A particularly useful section describes methods for ensuring proper merging of datasets and how to track dropped data. Details in a chapter on importing and exporting data delineate processes that should be followed and a discussion of the common issues that can occur and how to prevent or identify them. The authors then move into the reporting process including the development of accurate analysis datasets. Again, the authors emphasize the practical: tips to make sure that subjects are assigned properly to various analysis populations (e.g., safety and intent-to-treat populations) and a logical review of various data types (e.g., how to ensure that units are consistent in vital signs and, if the action taken of an adverse event indicates that study drug was stopped, did the subject discontinue due to an adverse event). The authors' discussion of validation of summary tables focuses on frequent issues such as truncation, and rounding and format problems. Again, useful advice with clear examples point a programmer in the correct direction and help prevent wasted time and ensure correct data summaries.

This book has filled a needed niche as a resource for SAS programmers new to clinical trials. In addition, the authors' clear discussion and practical advice make this an excellent reference for experienced programmers. The task of clinical trial data validation has been made less daunting with this accessible book.

Camille ORMAN
Johnson & Johnson

Brief Reviews of Teaching Materials

Analysis of Multivariate Social Science Data (2nd ed.).

David J. BARTHOLOMEW, Fiona STEELE, Irini MOUSTAKI, and Jane GALBRAITH. Boca Raton, FL: Chapman & Hall/CRC, 2008, xi + 371 pp., $69.95 (P), ISBN: 978-1-584-88960-1.

The second edition of this book continues the authors' effort of exposing multivariate statistics to the nonmathematical community, predominantly to the quantitative social scientists. The title of this version is shorter than its first edition, *The Analysis and Interpretation of Multivariate Data for Social Scientists*, although the content has been expanded with the basics intact. The expansion is based on authors teaching experiences in the last several years and their associated professional activities. These are timely additions. For example, there are additional chapters on regression analysis, structural equation models and on multilevel models, invaluable tools in social research. The website *http://www.cmm.bristol.ac.uk/team/amssd.shtml* contains the datasets and software to analyze them, and a useful tool for instructions. The book could certainly be used as a text book for nonstatistics graduate students mainly interested in social research.

Tapabrata MAITI
Michigan State University

Generalized, Linear, and Mixed Models (2nd ed.).

Charles E. MCCULLOCH, Shayle R. SEARLE, and John M. NEUHAUS. Hoboken, NJ: Wiley, 2008, xxv + 384 pp., $143.50 (H), ISBN: 978-0-470-07371-1.

For the second edition of *Generalized, Linear, and Mixed Models*, Charles McCulloch and Shayle Searle have added a co-author, John Neuhaus, and three new chapters to their book. These changes have strengthened an already valuable reference and teaching resource. The book offers 14 chapters and two useful appendices on the necessary matrix algebra and statistical background. After an introductory chapter that presents several examples and discusses the fixed versus random effect distinction, two chapters are devoted to models with a single categorical (Chapter 2) or single continuous (Chapter 3) predictor. Linear, linear mixed, generalized linear, and generalized linear mixed models are introduced for these simple modeling contexts and many of the methods and results of estimation and inference in the more general model classes are introduced here. Chapters 4–7 are devoted to the general classes of linear models, generalized linear models, linear mixed models, and generalized linear models, respectively. Specialized models are discussed in Chapters 8–11, which deal with models for longitudinal data (Chapter 8), marginal models (Chapter 9), multivariate models (Chapter 10), and, very briefly, nonlinear models (Chapter 11). The last section of the book is devoted to general issues, including chapters on departures from assumptions (Chapter 12), prediction (Chapter 13), and computing (Chapter 14). There has been some helpful reordering, but little revision, of the chapters from the first edition. In addition, Chapters 9, 10, and 12 are new, although some of the material on marginal models was included elsewhere previously. The strongest of the new chapters is Chapter 12, which collects important results concerning the effects of model misspecification from the literature, drawing heavily on the authors' own research.

Generalized, Linear, and Mixed Models has several strengths. It presents theoretical and methodological results concerning prediction, estimation, and inference with greater clarity and detail than in most competing texts. Documenting these results in a single volume is quite valuable, but it has its drawbacks. In several places results and their derivations offer little insight and are too intricate to be of interest to many readers. The authors could have done more to signpost some material as unessential, at least at first reading, or make greater use of appendices. The book contains exercises, which greatly enhance the book's potential for use as a course text; although it would have been nice to see the authors add to the exercises for this revision. Another missed opportunity is the failure to upgrade the graphics in this book, which remain too few and are of mediocre quality.

The book devotes substantial attention to linear and generalized linear models. This is useful for establishing notation, methods, and results that carry over to the broader mixed effect model classes that the book focuses on. However, most readers will need a more thorough understanding of fixed effect models than this book can provide, so the utility of the book's medium-depth coverage of these topics is questionable. If the book were to be used as a course text, for example, such a course would almost certainly have linear and generalized linear model prerequisites. I would have preferred that the authors concentrate on mixed models and complement their excellent presentation of the theory and methodology of these model classes with attention to their application for data analysis.

In my opinion we still do not have a definitive text on mixed models. Students of this area still need to draw from a variety of sources, but the second edition of *Generalized, Linear, and Mixed Models* is an excellent place to start. As a teacher of courses on mixed models and a researcher in this area, I find it to be an essential resource.

Daniel B. HALL
The University of Georgia

Introductory Statistics With R (2nd ed.).

Peter DALGAARD. New York, NY: Springer, 2008, xvi + 363 pp., $59.95 (P), ISBN: 978-0-387-79053-4.

The second edition of this clear and well-written book brings two major improvements with respect to its first edition: two new chapters on statistical methodology, covering Poisson regression (Chapter 15) and nonlinear curve fitting (Chapter 16), and the solution sketches to all the exercises (Appendix D). In addition, the original first chapter was expanded and broken into two chapters: Chapter 1 and Chapter 10. Chapter 1 now covers the very basics of the R language. Now Chapter 10 considers more advanced data handling tasks such as techniques to construct derived variables (e.g., grouping quantitative data and combining factor levels) and techniques for combining and restructuring data frames.

Raquel PRADO
University of California, Santa Cruz

SPSS for Intermediate Statistics: Use and Interpretation (3rd ed.).

Nancy L. LEECH, Karen C. BARRETT, and George A. MORGAN. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2008, xii + 270 pp., $ 34.95 (P + CD), ISBN: 978-0-80586-267-6.

As the title implies, this book is designed for people who know some statistics. The bulk of the material presented comes in the form of examples. These examples allow the reader to try the methods using SPSS, but the authors also break down the output for explanation.

The book has 11 chapters including Data Coding and Exploratory Analysis (EDA), Exploratory Factor Analysis and Principal Components Analysis, Selecting and Interpreting Inferential Statistics, Multiple Regression, Repeated-Measures and Mixed ANOVAs, and Multivariate Analysis of Variance (MANOVA) and Canonical Correlation. The multiple examples in each of these chapters walks the reader through the SPSS process, complete with screenshots, from data to interpretation. The analyses generally show common options used for similar problems, but sometimes are somewhat narrow. The chapter Several Measures of Reliability contains Cronbach's alpha, test-retest reliability, and Cohen's kappa. While these are common reliability measures, this is a narrower view of this subject than offered in other chapters.

The interpretations of the data are perhaps the best part of the book. Too often books focus solely on the analysis and the numerical results of the analysis, but this book remembers that its target audience is perhaps not intimately familiar with the statistical methodology and are unfamiliar with SPSS. The authors show the actual SPSS output and give explanations about each output area and then tie it all together in a brief summary.

Overall, this is a good first book for a person who is familiar with basic statistics and with some intermediate topics. They can teach themselves how to carry out data analysis using SPSS.

Jeffrey R. STANLEY
Texas A&M University

Statistical Rules of Thumb (2nd ed.).

Gerald VAN BELLE. Hoboken, NJ: Wiley, 2008, xxi + 272 pp., $63.95 (P), ISBN: 978-0-470-14448-0.

Most statistics books are texts or monographs concentrating on specific topics within the field or on certain application areas. Such books are essential, of course, but they are typically more work than fun to read. So it is a pleasure to read a book like *Statistical Rules of Thumb*, by Gerald van Belle, which has broad scope and appeal, and which is light enough to be read casually while still delivering useful ideas and insight. Dr. van Belle has assembled 122 "rules of thumb" organized into 10 chapters: Chapter 1, The Basics; Chapter 2, Sample Size; Chapter 3, Observational Studies; Chapter 4, Covariation; Chapter 5, Environmental Studies; Chapter 6, Epidemiology; Chapter 7, Evidence-Based Medicine; Chapter 8, Design, Conduct, and Analysis (of experiments); Chapter 9, Words, Tables, and Graphs; and Chapter 10, Consulting. These topics reflect the author's experience and interests, but are broad enough to appeal to almost all statisticians. The chapters on observational studies and evidence-based medicine are new to the second edition, and the other chapters have been revised and reordered slightly. The revision has involved the removal of a few rules of thumb and the addition of others. Some of the new rules were introduced on a companion website which originally promised to present a "Rule of the Month," a project which, understandably, has since been abandoned as overly ambitious.

Each rule is presented in an identical format: an introduction followed by a statement of the rule, illustration, basis of the rule, and a section on discussion and extensions. This uniform format is sensible and clear, but strained at times. For example, some of the sections "basis of the rule" are little more than restatements of the rule itself. Some rules are what I think of as *true* rules of thumb like, "Obtain at Least Ten Subjects for Every Variable Investigated" and "At Least Twelve Observations for a Confidence Interval." Others are better characterized as general information or principles regarding particular topics which help hold the chapters together. For instance, the first rule of Chapter 3 is "The Model for an Observational Study is the Sample Survey."

In a book like this any experienced statistician is likely to have quibbles about what is included and what is not, but overall, Dr. van Belle has done an admirable job of assembling these rules and pulling them together in a coherent book, rather than as just a list. Students with some coursework and experience under their belt (e.g., statistics or biostatistics graduate students in their second year or beyond) and all practicing applied statisticians will benefit from reading this book. It would be ideal as a secondary text for a course on statistical consulting. The book price is a bit high for a paperback and there are quite a few typographical errors, but despite these minor flaws, I strongly recommend this book to anyone interested in the practical application of statistics.

Daniel B. HALL
The University of Georgia

Material Reviews An excellent way to start professional writing is by submitting reviews. For many novice writers writing reviews of books, Web sites, software and other teaching materials is a comfortable entry threshold into the world of academic publication. Reviews are normally short at about 700 words in length. There are three advantages for novices in writing reviews.