Book Reviews

This section will review those books whose content and level reflect the general editorial policy of *Technometrics*. Publishers should send books for review to Eric R. Ziegel, BP Naperville Complex, Mail Code C-7, 150 West Warrenville Road, Naperville, IL 60563-8460 (*ziegeler@bp.com*).

The opinions expressed in this section are those of the reviewers. These opinions do not represent positions of the reviewer's organization and may not reflect those of the editors or the sponsoring societies. Listed prices reflect information provided by the publisher and may not be current.

The book purchase programs of the American Society for Quality can provide some of these books at reduced prices for members. For information, contact the American Society for Quality, 1-800-248-1946.

Editor Reports on New Editions, Proceedings, Collections, and Other Books

TECHNOMETRICS, FEBRUARY 2004, VOL. 46, NO. 1
DOI 10.1198/004017004000000167

A User-Friendly Guide to Multivariate Calibration and Classification, by Tormod NÆS, Tomas ISAKSSON, Tom FEARN, and Tony DAVIES, Chichester, U.K.: NIR Publications, 2002, ISBN 0-9528666-2-5, viii + 344 pp., $75.00.

Contrary to the perceptions that people may have formed from my extended tenure as *Technometrics* Book Reviews Editor, I have other hobbies besides reading statistics books. For example, I am a serious collector of classical music recordings, especially operas. Publications of reviews for classical music on CDs all publish annual lists of their reviewers' choices for the year's best CDs. Well, if I had a similar feature, this would certainly top my list of favorite statistics books for the recent past. A book with "user-friendly" in its title would automatically grab my attention. This excellent book really is user-friendly, and I have read it with interest and enjoyment. This is a necessary library addition for anyone involved with chemometrics.

This is a privately published book. The first author was previously very familiar to me as an author of the first really comprehensive chemometrics textbook (Martens and Næs 1989), which was reviewed for *Technometrics* by Hagwood (1991). Unfortunately, someone permanently "borrowed" my copy of that book, and I have missed it on many occasions since then. This author and the second author together wrote a column, "The Chemometrics Space," in the *NIR News* from 1991 through 1995. Since then, the third author has written the column. Since 1989, the fourth author has been writing the chemometrics column in *Spectroscopy World*. The Preface, which is also the first chapter, credits this last author with "making the text as reader-friendly as possible" (p. 1). He did his job very well.

In nearly all of the chemometrics books that I have known, the authors have been chemists, and that situation persists here. The first two authors are from Norway, and the last two are from the United Kingdom. In the latter country, and probably in other parts of Europe too, it is possible to obtain a doctorate jointly in chemistry and statistics. As chemists, these authors are certainly very competent in many areas of statistics, especially multivariate statistics. As the publisher's name indicates, the book has emanated from applications in NIR spectroscopy. The authors have tried to generalize their material as much as possible. However there are chapters that relate strictly to NIR, and its applications comprise most of the illustrations. This is also mostly a book about calibration. Only the last of the 16 chapters (40 pages long) is devoted to classification. All of the rest of the book is concerned with applications in multivariate calibration.

Why is this book user-friendly? First, unlike the other books in the chemometrics literature that include significant coverage of multivariate calibration, this book presumes only "simple matrix algebra and basic statistics" (p. 2) as a necessary background. The book does a nice job of providing a lucid and basic introduction to linear calibration and collinearity in multiple regression as the motivation for the multivariate calibration problem. For the primary audience, the persons who actually calibrate NIR instruments, it is totally anchored in their applications. Some chapter topics, including data compression by Fourier analysis and the multiplicative scatter correction (MSC) of spectroscopic data, are very specific to spectroscopic applications. Beer's law is mentioned in several chapters. There is extensive use of visualization. There are even decent margin notes.

The basics of data reduction and calibration by principal components regression (PCR) and partial least squares (PLS) are presented in Chapter 5. The issue of data standardization is discussed, because the audience for this book is not intended to be limited to users of NIR. One does not need to standardize for NIR or similar spectroscopic applications, because all predictors are dimensionally consistent. The chapter ends with a discussion of continuum regression. Interpretation of the loadings and scores from PCR and PLS is discussed in Chapter 6. There is some interesting discussion of nonlinearity. The application in this chapter involves a mixture design. Chapter 7 offers variable selection as an alternative to dimension reduction in multivariate calibration, particularly for the cost-reduction benefit if fewer measurements would be necessary. Statisticians should find the application to wavelength selection for use in PLS models to be an interesting approach. In addition, modern methods, such as genetic algorithms and jackknifes, are discussed for the same situation.

Chapter 8 is for chemists. It focuses primarily on the Fourier transform, although it also includes a section on wavelets as an alternative methodology. Chapter 9 discusses nonlinearity issues in calibration. Several clever methods include using nonlinear effects to get simpler models, adding squares and cross-products in PCR or PLS, and such modern methods as locally weighted regression (LWR) and neural networks. Chapter 10, another chapter for the spectroscopy practitioner, describes MSC. This important data-handling methodology helps improve the predictive capabilities of PCR or PLS. Several other scatter correction methods are also presented. The nonlinearity theme continues in Chapter 11, which offers further consideration of LWR. An algorithmic approach is presented that involves dividing the samples into regions and using PCR within each of the regions. Quadratic functions can be added for flexibility. The same set of principal components is used for all samples. Several nonlinear methodologies are discussed in Chapter 12. First, polynomial PCA creates flexibility by adding squares and cross-predictors to the variables for PCA. Next, cluster analysis is used to create clusters for which linear calibration equations are appropriate. Finally, there is an extended discussion of feed-forward neural networks.

The vast array of modeling techniques clearly requires a full range of diagnostic evaluations. Chapter 13 focuses on the primary validation in calibration, the ability of the model to predict accurate results for future analyses. Generally, some test samples are withheld to evaluate the predictive capability and provide a measure of the root mean squared error. Predictive capability is fundamental to the entire process, because it also helps determine the number of multivariable components needed in the calibration model. The authors recommend setting aside one-third of the samples. Confidence intervals, analysis of variance, and bootstrapping are the statistical methodologies used. A chapter is also devoted to outlier testing. Here the most interesting feature is the evaluation of x-outliers, which are detected graphically in the principal component space. Leverage statistics are calculated versus principal components, not versus the original x-variables. It is noted that the leverage points may cause the component to appear to be necessary, when it really is not. With the many online analyses using multivariable calibration models in place in the chemical industry, x-outlier analysis should be used as a routine screen.

The calibration material concludes with three brief chapters, one on the selection of samples and the other two on the use of the calibration equations. Experimental design for chemical spectral applications offers some unique opportunities, because the predictor variable set can have more than 1000 variables. Because obtaining response properties is often costly, one suggestion is to work with the sample spectra and get response properties on a diverse set of sufficient size. Sample representativeness actually applies in the principal component space. Cluster analysis can be used in this process. In calibration monitoring, analogous to the classical Deming funnel experiment, one does not use daily bias adjustments, so control charts for differences between lab and analyzer samples are recommended. Surprisingly, the authors suggest that subgroups of samples should be accumulated, for example weekly. Another important issue in using calibrations is their transfer between instruments. Several different spectral transfer methods are discussed for use with a set of "standard" samples. There is also discussion of a couple of procedures for making the calibrations more robust.

Finally come the aforementioned 40 pages on clustering and discriminant analysis that represent the "and classification" portion of book. All but the last 10 of these pages are devoted to discriminant analysis. The authors note that discriminant analysis is simply calibration versus a categorical response, then proceed to draw parallels throughout to calibration methods. They use Bayesian methods as the primary motivation for linear discriminant analysis and extend the development to a quadratic discriminant function for the situation where there are not common covariance matrices. Fisher's linear discriminant function is then explained and extended to canonical discrimination for more than two groups. Because of the use or spectral data, the multicollinearity issue arises again. Principal components, SIMCA, and DASCO are presented as solutions to this problem. A section on alternative methods considers K-nearest neighbors and discrimination via classification. Cluster analysis is presented first as methodology for principal components, but hierarchical, partitioning, and fuzzy methods are also presented.

The length of this review certainly indicates that much material has been presented in not many pages. The sparseness of the mathematics is mitigated in part by a 40-page Appendix A, "Technical Details." Although there are illustrations, they are all brief. There is no data analysis. The recent book by Martens and Martens (2001), reviewed by Bayne (2002), is an excellent resource for actual analysis of multivariate chemometrics data. Nevertheless, *A User-Friendly Guide to Multivariate Calibration and Classification* belongs on the bookshelf of any statistician or chemist who is involved with doing multivariate calibration.

Eric R. ZIEGEL
BP

REFERENCES

Bayne, C. (2002), Review of *Multivariate Analysis of Quality: An Introduction*, by H. Martens and M. Martens, *Technometrics*, 44, 186–187.

Hagwood, C. (1991), Review of *Multivariate Calibration*, by H. Martens and T. Næs, *Technometrics*, 33, 366–367.

Martens, H., and Martens, M. (2001), *Multivariate Analysis of Quality: An Introduction*, West Sussex, U.K.: Wiley.

Martens, H., and Næs, T. (1989), *Multivariate Calibration*, New York: Wiley.

Chemometrics: Data Analysis for the Laboratory and Chemical Plant, by Richard G. BRERETON, West Sussex, U.K.: Wiley, 2003, ISBN 0-471-48978-6, xiv + 489 pp., $55.00 (paperback).

My background in chemometrics is somewhat strange. In graduate school, I kept switching between physical chemistry, statistics and genetics. I finally ended with a Ph.D. in physical chemistry, an MS in statistics, and coursework in genetics. I am now a professor of statistics. Given this background, I am in a somewhat unique position to review this book. In short, this is the best book that I have seen covering the entire field of modern chemometrics both for the academic and the industrial user. Most of the techniques used in modern chemometrics are in this book, and the coverage is extremely clear, well done, and very useful.

Some examples include the following:

1. Wavelet methods, which probably are going to be very important in the future, are discussed.
2. Software is provided for some of the methods, references to software for most of the other methods are given, and very clear discussions of the use of software are presented.
3. The discussion of all the methods is clear and complete. The goal is to understand the method and to apply it in a correct manner. The treatment is not "black box," and I think it is first rate.
4. The examples are clear, well chosen, and useful.
5. The discussion of calibration is excellent, but the terminology is a bit nonstandard for statisticians. As long as one knows this, there is no problem. The terminology issue extends to other methods as well. Again, just knowing that the terminology is a bit different is sufficient.
6. The use of graphical methods is, again, very well done. The use of exploratory data analysis is extensive. One also finds such methods as looking for multivariate outliers in the plane of the first two principal components of a multivariate data set.
7. There are good discussions of cross-validation, scaling, and standardization issues for multivariate models.
8. This book contains the best introductory discussion of fractional factorial designs that I am aware of.

Overall, I think this book is superior.

Having said that, I also found some things that I would like to see changed in a second edition. These include the following:

1. The Savitsky–Golay smoothing technique is discussed. However, only Savitsky and Golay's original article is included as a reference. Two modifications that are very important for proper use of the technique were published later; Sebastian, Booth, and Hu (1995) gave those references and discussed their use.
2. Other smoothers, such as loess (Grznar, Booth, and Sebastian 1997), could be discussed.
3. Neural networks are becoming important in chemistry (e.g., Hamburg, Booth, and Weinroth 1996), and could be discussed.
4. The coverage of discriminant analysis is limited to the linear, two-group case. A more general discussion would be of value.
5. This point is a bit longer. The author (as discussed on p. 2) had a bad experience with a statistician. He says, "I was told once by a very advanced mathematical student that it was necessary to understand Gallois (sic) field theory in order to perform multilevel calibration designs, and that everyone in chemometrics should know what a Krilov (sic) space is." I did study Galois theory, but was long gone from graduate school by the time I learned that it has a lot to do with experimental design. I think most people that deal in applications of standard designs do not need to know this. Those that do research in new designs may. I am afraid that I still have not come across Krylov spaces. (A session on the web introduced me to Krylov spaces. If one is researching new methods in statistical computing, they may be helpful. If not, they can be omitted by most of us.) We need to be careful about what we say, because we can turn people off to statistics. I bring this up because of comments the author makes on pages 46 and 47. He is somewhat dismissive of using confidence intervals and hypothesis tests to analyze data and that carries through to the rest of the book, stating that required normality assumptions do not apply, among other reasons. We know that in many cases such problems can be dealt with. The use of confidence intervals and hypothesis tests allow us to quantify uncertainty and deal with variability and thus helps us understand and apply our experimental results. I think the author was turned off to inferential statistics by comments such as the foregoing quote. I hope that he will reconsider inferential statistics and include more of it, as appropriate, in a future edition.
6. More references to the literature would be very helpful.

In summary, this is the best book on chemometrics that I have seen. I highly recommend it.

David E. BOOTH
Kent State University

REFERENCES

Grznar, J., Booth, D. E., and Sebastian, P. R. (1997), "The Use of Robust Smoothers in Nuclear Material Safeguards," *Journal of Chemical Information and Computer Science*, 37, 236–240.

Hamburg, J. H., Booth, D. E., and Weinroth, G. J. (1996), "A Neural Network Approach to the Detection of Nuclear Material Losses," *Journal of Chemical Information and Computer Science*, 36, 544–553.

Sebastian, P. R., Booth, D. E., and Hu, M. Y. (1995), "Using Polynomial Smoothing and Data Bounding for the Detection of Nuclear Material Diversions and Losses," *Journal of Chemical Information and Computer Science*, 35, 442–450.

Statistics With Applications in Biology and Geology, by Preben BLÆSILD and Jørgen GRANFELDT, Boca Raton, FL: Chapman & Hall/CRC, 2003, ISBN 1-58488-309-X, ix + 555 pp., $59.95 (softcover).

If teaching statistics by example is your goal, then this book will provide a wealth of interesting examples that demonstrate the methods of statistics and practical data analysis. The focus of the book is biology and geology, as the title indicates, with a greater emphasis on biology. There are about 70 separate datasets in the examples and exercises, providing a great deal of material for teaching. About one-third of the examples and exercises are based on problems in geology, including sediment composition, sound propagation in rock, metal content of geologic samples, sand sifting rates, pH of core samples, shape and weight of stones, magnetism in lava flows, gas diffusion in stones, mineral content in water, earthquakes, sediment transport, wind direction, and orientation of crystals. The biological exercises and examples encompass problems related to fisheries taxonomy, vertebrate physiology, invertebrate population density and size distribution, ecology, toxicology, both plant and animal genetics, agronomy, microbiology, and human studies involving asthma, twin studies, cancer risks, chronic disease, and physiology.

The authors provide the data on their website along with SAS programs used for analysis. The book devotes more pages to displaying SAS program code than I enjoy reading, and much of this is redundant with the website, but the duplication of material may be helpful for students. The first example of a SAS program includes a SAS macro and programming using PROC IML, a matrix programming language within SAS. This is much too advanced for a student's first encounter with SAS, and is a challenge to even experienced SAS programmers. However, the website is well organized and enables the user to easily locate the programs for the specific examples.

The book comprises 12 chapters, beginning with a discussion of statistical models and inference in Chapter 1. Chapter 2 emphasizes graphics including histograms and probability plots. Chapter 3 discusses normally distributed data procedures, including one- and two-sample tests, one-way analysis of variance, and simple regression. Chapter 4 presents two-way analysis of variance with interaction as an example that extends the simple linear models. Chapter 5 discusses the concepts of power, noncentral distributions, and sample size efficiency through experimental design. Chapter 6 introduces correlation and the bivariate normal distribution. Chapter 7 presents the multinomial distribution and statistical methods for categorical data, and Chapter 8 introduces the Poisson distribution for analysis of rate data. Chapter 9 introduces Poisson and logistic regression as part of a brief discussion of generalized linear models. Chapter 10 presents methods for analyzing directional data in two and three dimensions using circular normal distributions (von Mises and Fisher distributions). Although concepts of maximum likelihood are briefly mentioned in various parts of the book, Chapter 11, "The Likelihood Method," offers more details and a discussion of quadratic approximation. Chapter 12 introduces some nonparametric tests for one-sample, two-sample, and k-sample Kruskal–Wallis tests.

Unfortunately, my copy of the book had too many printing errors, especially in the mathematical notation, which would cause confusion for students seeing this material for the first time. Some of the discussion assumes a background that is too advanced for the intended audience of undergraduate students in biology and geology. Discussions of subspaces, projections about the geometry of linear models, and the distinctions between "affine subspace" and "linear subspace" in the section on generalized linear models seem to be a distraction from the theme of practical data analysis in biology and geology. The book's strength lies in the emphasis on practical examples and realistic research problems that include many interesting datasets. The analyses are presented in a way to facilitate teaching the principles of statistics and the importance of quantitative studies to students in the natural sciences.

James T. WASSELL
National Institute for Occupational Safety and Health
Centers for Disease Control and Prevention

Survival Analysis (2nd ed.), by John P. KLEIN and Melvin L. MOESCHBERGER, New York: Springer-Verlag, 2003, ISBN 0-387-95399-X, xv + 536 pp., $89.95.

Comprising 13 chapters and 5 appendixes, with 97 illustrations and several exercises at the end of each chapter, this book is an excellent graduate-level text for a course in survival analysis. Students will definitely find the authors' systematic treatment of topics, clear discussions and derivations, and numerous detailed examples useful. This book is also a good reference source for practicing statisticians, biostatisticians, and public health professionals with a basic statistics and applied statistics background. Although the examples are biomedical in nature, most methods described in the book for time-to-event data are applicable to other fields, including engineering and economics, and the book should be useful for researchers in these disciplines. The authors use semiparametric and nonparametric methods extensively, and also discuss parametric models. The "Practical Notes" and "Theoretical Notes" provided in many sections are very attractive and give readers information and citations beyond the material in the text.

Chapter 1 introduces 19 datasets that are used to illustrate various aspects of survival analysis throughout the text. These datasets pertain to biomedical or public health examples. As indicated in the Preface, these are available at the authors' website, accessible through the Springer website (*http://www.springer-ny.com*) or the first author's website, (*http://www.biostat.mcw.edu/homepgs/klein/book.html*), which also provides the book's outline, errata, and some SAS macros.

Chapter 2 defines and illustrates basic parameters related to time-to-event data, including the survival function, the hazard function, the density function, cumulative distribution function, and the mean and median residual life at a given time. Table 2.2 summarizes these parameters for 11 common parametric families. A discussion of summary statistics for competing risk probabilities is also included; these quantities are estimated in Chapter 4.

Chapter 3 describes censoring (right, left, and interval) and a common approach for the construction of the corresponding likelihood function. This chapter also defines truncation (left and right) and discusses the use of a conditional distribution for constructing the likelihood function. The counting process approach for construction of the likelihood with truncated or censored survival data is briefly described in the last section.

Chapter 4 describes nonparametric estimation of the cumulative hazard function and survival function for right-censored data. The product-limit estimator and associated confidence intervals are described and illustrated, and a confidence band for the survival function is also discussed. The chapter ends with a section on estimation of the survival function for right-censored and left-truncated survival data. Chapter 5 presents estimation of the survival function under other schemes, including left censoring, double censoring, interval censoring, right-truncation, and grouped data.

Chapter 6 explores a few topics on estimation for univariate survival data. Kernel methods for estimating the hazard function are explained in Section 6.2. The next section summarizes estimation of excess mortality. Section 6.4 discusses Bayesian nonparametric estimation, incorporating MCMC methods.

Chapter 7 takes a detailed look at hypothesis testing for hazards, starting with the simplest single population case and going on to tests for two or more populations, tests for trend (via an ordered alternative), stratified tests (on levels of covariates), and Renyi-type tests (which are powerful in detecting crossing hazards). A new section on testing the equality of survival curves at a fixed time point is an attractive inclusion in the second edition.

Chapter 8 discusses the Cox proportional hazards model with fixed covariates (discrete or continuous). The partial likelihood is clearly explained and is illustrated with examples. Although this chapter generally treats the baseline hazard function as a nuisance parameter, the last section is devoted to estimation of the survival function via a variant of Breslow's estimator of the baseline cumulative hazard. Extensions to the proportional hazards model, such as inclusion of time-dependent covariates, stratified proportional hazards model, left-truncation, and a multistate model for handling time-varying effects, are discussed. Using several examples, Chapter 11 explains model assessment via Cox–Snell residuals, martingale residuals, deviance residuals, and influence diagnostics. Graphical methods for assessing the proportionality assumption are also presented.

Chapter 10 describes two additive hazards regression models. First, Aalen's nonparametric additive model for the conditional hazard given covariates is discussed in detail, allowing the regression coefficients to be functions of time.

Next, the authors discuss an additive model obtained by replacing the time-varying regression coefficients by constants. Chapter 12 presents regression models based on parametric distributions such as the Weibull or log-logistic distributions. Chapter 13 gives a brief summary of multivariate survival analysis, including measures of association and frailty models. This chapter is a good starting point for further study of multivariate survival analysis, for which the book by Hougaard (2000) is an excellent reference.

Nalini RAVISHANKER
University of Connecticut

REFERENCES

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

The Stress-Strength Model and Its Generalizations: Theory and Applications, by Samuel KOTZ, Yan LUMELSKII, and Marianna PENSKY, River Edge, NJ: World Scientific Publishing, 2003, ISBN 981-238-057-4, xvii + 253 pp., $52.00.

This book is all about the probability of one random variable, X (e.g., stress) is less than another random variable, Y (e.g., strength), and in which scenario the unit is considered safe and reliable. Otherwise, the unit is liable for failure and risky. The book's title is hence aptly chosen. In health applications, X and Y could be "susceptibility" and "immunity" to a disease.

The book has seven well-written chapters with an exhaustive bibliography. The exercises in each chapter make this a suitable textbook for a graduate-level course. At least one statistical theory course with sufficient calculus and probability background is necessary to follow the material in this book. The book covers such topics as *definition and history of stress-strength models; motivation for the importance of such concepts; maximum likelihood, Bayes, empirical Bayes, and uniformly minimum variance unbiased estimates of* $P(X < Y)$*; interval estimation of* $P(X < Y)$*; necessary transformations; sampling from Rayleigh, exponential, normal, gamma, Pareto, Weibull,* and *Burr populations; bivariate and multivariate normal, elliptical, and exponential distributions with their properties in relation to* $P(X < Y)$*; hypothesis testing, bootstrap, parametric, and nonparametric methods of making inference of* $P(X < Y)$*; estimation of* $P(X < Y < Z)$*; stochastic processes (including Markov models) for system reliability; linear models; and stochastic time series methods,* among others.

The applied case studies in rocket motor, military, medicine, and psychology applications strengthen the reader's interest, understanding, and appreciation of the concepts. Other interesting and valuable features of this book are similarities versus differences between the stress-strength model and receiver operating characteristic curve and the relationship between process capability index and the stress-strength model.

I enjoyed reading this well-explained book. I strongly recommend it to both theoretical and applied statisticians.

Ramalingam SHANMUGAM
Texas State University

Optimal Reliability Modeling: Principles and Applications, by Way KUO and Ming J. ZUO, Hoboken, NJ: Wiley, 2003, ISBN 0-471-39761-X, xvi + 544 pp., $125.00.

The authors have produced an excellent book on the modeling of system reliability. It is a modern, up-to-date treatment of a number of topics and will be of interest to both the practicing engineer and those interested in reliability as a research topic. It has an extensive set of references and bibliography. An instructor wishing to adopt it as a textbook will need to supply problem sets for homework.

The book is concerned with probabilistic modeling and does not treat matters of statistical inference, which are treated in a number of other books. The book is organized into 12 chapters and covers a number of modern topics.

After a brief introductory chapter, the authors provide some basic probability theory and stochastic processes along with basic reliability concepts and common probability distributions in Chapter 2. Chapter 3 is an introduction to complexity analysis. Chapter 4 provides the basic introduction to the mathematical theory of binary structure functions and coherent systems. Here system logic and minimal path and cut sets are discussed, the notion of a module is introduced, and various performance measures including reliability and availability are discussed. Series and parallel and series-parallel and parallel-series, as well as warm and cold standby systems, are treated in some detail.

Chapter 5 describes the pivotal decomposition and studies methods for generating the minimal path and cut sets from reliability block diagrams. It then describes the inclusion-exclusion method and the sum-of-disjoint-products method for reliability evaluation. The chapter closes with discussions of Markov chain imbeddable structures, delta-star and star-delta transformations, and various methods of producing bounds for system reliability.

Chapter 6 discusses general methodology for system design. Here one finds discussions of redundancy, various measures of component importance, majorization, and optimal arrangements and designs. Chapters 7 and 8 deal with k-out-of-n systems and their designs. This is a popular type of redundancy in fault tolerant systems. Chapters 9 and 10 treat consecutive k-out-of-n systems and their design and multidimensional versions. Chapter 11 treats various extensions of the models in Chapters 7–10, and the final chapter treats the important subject of multistate reliability models.

The authors have provided an up-to-date treatment of these topics. Those interested in the probabilistic end of reliability theory will find this an excellent book.

William S. GRIFFITH
University of Kentucky

Applied Statistics and Probability for Engineers (3rd ed.), by Douglas C. MONTGOMERY and George C. RUNGER, Hoboken, NJ: Wiley, 2003, ISBN 0-471-20545-4, xiv + 706 pp., + CD, $104.95.

It is a pleasure to review an application-orientated, introductory-level textbook on statistics that does not begin with the author commiserating with the math phobia of students and offering copious assurances that high school algebra is all the mathematics they will need. This book is primarily written as a text for students of engineering or the hard sciences who have had calculus through multiple integration (typically taught in second semester calculus). The only exception is the use of some matrix algebra in the chapter on multiple regression. For this and other reasons I would recommend this book for upper-level undergraduates or lower-level graduate students alike.

This is not a mathematically rigorous book. It does, however, take the attitude that the students are big kids now. They have the math; they are going to use it in other courses, so let's use some of it here. The effect is to turn an otherwise anemic introductory course into a full-course meal (with the last chapter being dessert). In some ways this book is like a standard graduate-level text for a mathematical introduction to probability and statistics where the statements of the theorems are kept and the proofs are replaced with many real world examples. Montgomery and Runger have spent many hours collaborating with industry and have a good idea of what statistical problem-solving skills engineers need to know in the workplace. This is the most complete set of basic statistical methods I have ever seen covered in an introductory text. As a bonus, when the course is over, the text can also serve as an excellent reference.

Chapter 1 introduces the student to the role of statistical thinking in their engineering/scientific understanding of how the world works. Deming's famous funnel experiment is discussed, along with the distinction between retrospective, observational and designed studies. Chapters 2–5 deal with probability theory. This covers a standard set of topics but at a level of detail not seen in most introductory books. For example, the chapters on distribution functions (3 and 4) cover not only the usual suspects (binomial, Poisson, normal, and uniform), but also a host of other useful characters including the geometric, hypergeometric, negative binomial, exponential, Erlang, gamma, Weibull, and lognormal distributions. Examples of when and how to use these are provided. Chapter 5 requires the most knowledge of calculus. It deals with joint, conditional, and marginal distributions.

Chapter 6 provides a needed break from the level of abstraction of the previous three chapters. It deals with samples, some simple summary statistics, and different ways to plot data. Minitab is used exclusively throughout the book for examples of computer outputs.

Chapters 7–10 cover the core of the basic statistical methods taught in an entry-level, one-semester course. Topics covered include point estimates, interval estimates, and hypothesis testing for continuous variables and proportions. Again, what distinguishes this book is the breadth and depth of the coverage. Examples of this include the effective explanation, with examples, of the use of maximum likelihood and method of moments for parameter estimation. Besides discussing type I and II errors, the authors spend a good deal of time explaining statistical power, OC curves, how type II errors are calculated and, most importantly, how to make sample size calculations for a number of tests. Goodness of fit is covered not only for its use with contingency tables, but also for testing the validity of any purposed distribution function. This is all useful stuff.

Chapters 11–16 deal with material usually taught (if at all) in the second semester of an introductory course or in a second-year course on more advanced topics. The subjects covered are simple and multiple linear regression (11 and 12), one-way ANOVA for fixed and random effects (13), design of experiments (14), nonparametric statistics (15), and SQC (16). Chapters 11–15 are a veritable tour de force through their respective subject matters. I think that these chapters border on too much too fast (more on this below). Chapter 16 slows down and ends with some perspectives on the use of statistical methods in corporate America. No reference is made to 6-Sigma or any other corporate-wide program. I think this is wise, because corporate initiatives come and go but basic, sound statistical methods have a much longer life time (I first learned about DOEs and control charts during the total quality heyday of the mid 1980s).

Answers to most of the odd-numbered problems are included in an appendix. Other appendixes contain a good bibliography for further study and a glossary of terms. The book comes with a CD that contains a complete e-copy of the text plus more. The "more" includes complete datasets for some of the examples and problems in the text, completely worked out solutions to a number of the problems, and topics not covered in the paper version. The latter are usually more advanced topics like bootstrapping methods, a proof that the sample standard deviation is biased, lack-of-fit tests, ridge regression, and nonlinear regression models to name a few.

I have only one non "nitpicky" issue with this book. Sometimes the authors get too technical for the intended reader. It is as if they switch and address a more academic, or at least more statistically savvy audience. For example, consider their definition of a random variable: "A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment" (p. 54). True, but unless the student is a math major, more explanation than the authors provided is needed to make the intent of this definition clear.

In the first 10 chapters, this is an episodic problem. In Chapters 11–15 the question of whom they are writing to becomes more of an issue. The authors wrote their book under the assumption that this may be the only course in statistics that the reader will have. Consequently, they filled it with most of what a practicing engineer will need to know. There is indeed a rich array of tools presented in these chapters. For example, the two chapters on linear regression cover variance inflation factors, Mallows's $C\hat{p}$, the Cook distance, the PRESS statistic, regression with dummy variables, and a host of model-building methods, all very useful and most of them typically taught in a subsequent course devoted to linear regression models. The same comments can be made about the other chapters in this section.

The downside is that the background needed to understand the material is presented at a level and a pace that I think will be beyond most of the intended students. The variance estimators, some of the matrix algebra, the partitioning of sums of squares, and degrees of freedom come pretty fast and furious. The presentation is concise and precise to the point at which much of the pedagogical hand-holding appropriate to an introductory text is missing. It is as if the primary intent in these chapters is to write a highly annotated reference book. For a general practitioner of statistics (like this reviewer), the material in these chapters is excellent. For engineers taking their first course in statistics, I fear this too dense. In short, I wish they would slow down in these chapters and spend some more words on explaining "physically, what does this mean and why is it important?" Having said this, I would also add that in the hands of a good teacher, the necessary embellishments could be provided in the classroom that will bridge the "split audience" issue.

This book is as applied and thorough a preparation for the real engineering world as can be expected in a one-year course. The only important material missing is a chapter on reliability modeling. (Reviewers of the previous two editions also made this recommendation.) I think this book will be a long reach for students taking their first course in statistics; but if they can stay the course, they will be far up on the learning curve for the application of statistics to problem solving. This book will also serve as an excellent reference for the practicing industrial statistician.

Robert V. BRILL
Astaris LLC

Probability, Statistical Optics, and Data Testing: A Problem Solving Approach (3rd ed.), by B. R. FRIEDEN, Berlin: Springer-Verlag, 2001, ISBN 3-540-41708-7, xxiv + 493 pp., $79.95.

There is a minor industry devoted to producing books whose titles could be "Probability and Statistics for X," where X stands for a needy group (e.g., industrial engineers, climatologists, "dummies"). For the book under review, X is "students interested in optical applications." A review of a book from this genre should address two questions. First, how closely is the book tied to what members of X are assumed to know? According to the author, the level of presentation requires an understanding of calculus and the basics of matrix manipulations, but "prior formal education in optics is not strictly needed." This latter claim needs to be qualified. The first substantive optical example occurs in Section 2.10, where knowledge of concepts such as "incoherent object" and "intensity profile" is presumed. Readers without a decent background in physics or a related field will have a difficult time tiptoeing through this section and much of the rest of this book.

Second, how well does the book introduce the basic ideas of probability (Chaps. 2–8, roughly half the book) and statistics (Chaps. 9–17)? The general style in each chapter is to introduce the theory and then illustrate the concepts via substantive applications related to optics (although nonoptical examples do crop up involving, e.g., state lotteries, the existence of intelligent life in the universe, and the bull/bear behavior of stock markets). Chapter 1 is a short introduction, with a gem of a mini-essay on chance versus determinism. Chapter 2 is an informal (nonset theoretical) presentation of probability theory. Chapter 3, "Continuous Random Variables," introduces the usual suspects, including probability density functions (PDFs), their moments, and the Poisson, binomial, uniform, exponential, normal and Cauchy distributions, along with some not commonly encountered in an introductory text (a distribution based on the square of the sinc function and a slew of information measures). Chapter 4, "Fourier Methods in Probability," begins with the statement that "there is probably no aspect of probability theory that is easier to learn than its Fourier aspect"—arguably true for students of optics, but questionable for others. The emphasis here is on using the characteristic function to compute moments for a PDF, to determine the PDF for sums of independent random variables (RVs), and to prove a simple version of the central limit theorem. The developments in this chapter are generally good, with some particularly nice examples of sums of RVs in optical applications that do not converge to a normal RV; however, there are two discordant notes. Section 4.15.2 has a bizarre discussion about estimating the unknown mean value, μ, for a normal RV X, in which a nonsensical method to "substantially narrow the range of possibilities" for μ is proposed based on purposely adding random noise to a random sample from X and subtracting a histogram of the noise-contaminated sample from a histogram of the noise-free sample. Chapters 5 and 6 discuss transformations of RVs and the binomial distribution (and its relationship to the Poisson and normal distributions). Chapter 7 is a short, but serviceable, introduction to using uniformly distributed random deviates for generating random deviates from a prescribed PDF; however, the author's opening comment that "most large computers are capable of generating random numbers of only one variety—uniformly random numbers over the interval $(0, 1)$" will strike S-PLUS and Matlab users as rather quaint. Chapter 8 introduces the notion of stochastic processes, but the author's approach is nonstandard and confusing. For example, a stochastic process is defined as a family of curves (the ensemble) that is parameterized by an RV or a vector of RVs, whereas the standard definition is "a family $\{X(t)\}$ of RVs indexed by the symbol t, where t belongs to some given index set" (Priestley 1981). The author's definition leads the reader to puzzle over what parameterization would be appropriate for the ensemble associated with normally distributed stationary processes. The definitions for "autocorrelation" and "autocorrelation coefficient" are also nonstandard, the latter being a mean-corrected version of the former, with both actually being covariances rather than correlations. The power spectrum is defined as a limiting Fourier transform of the "autocorrelation," with no statements about what mathematical conditions are needed for the limit to exist. What is commonly called the gain function (i.e., the absolute value of the transfer function) is here termed the "modulation transfer function."

Chapter 9, the first chapter on statistical theory, discusses the statistical properties of the sample mean, weighted averages, sample variance, and sample median under various distributional assumptions about the underlying independent and identically distributed (iid) data (normal, binomial, and a specialized PDF proportional to the square of a sinc function). The discussions are generally good (particularly for the squared sinc function), although the one on weighted averages might prove to be too abstract for those with little experience in data analysis. Chapter 10 is supposedly a discussion of the estimation of probability laws. Ignoring its confusing introduction, this chapter starts promisingly enough with a section on the orthogonal function approach to estimating PDFs from iid data. A follow-up section discusses a variation involving a Karhunen–Loeve expansion, in which the author fails to point out the practical limitation that knowledge of the unknown PDF is needed to construct the orthogonal functions. After this section, the chapter takes off on a major departure. The data are no longer assumed to be iid realizations of RVs, but rather are taken to be perfectly known functions of the unknown PDF (e.g., our "data" are the mean and variance of the unknown PDF). Under this framework, the author discusses a Bayesian-like approach called the "principle of maximum probability" and relates it to the maximum entropy approach. From a purely mathematical standpoint, this material is fascinating and should be required reading for all advocates of the maximum entropy principle; unfortunately, because of the bizarre notion of what constitutes data, it is only marginally relevant to the statistical problem of PDF estimation.

Chapters 11, 12, and 13 provide fairly standard discussions on the chi-squared test for significance departures from a prespecified distribution, the one-sample t test for a mean and the F test for equality of the variances. Chapters 14 and 15 are bare bones introductions to least squares theory and principal components analysis, whereas Chapter 16 is basically a discussion of Bayesian statistical theory in the context of assessing whether or not coin flips are biased. The book's final chapter, 17, "Introduction to Estimation Methods," starts with a nice overview of maximum likelihood estimators, the Cramer–Rao and Bhattacharyya lower bounds, and Bayesian estimation theory (although purists will object because of the lack of statements about conditions needed for various results to hold). However, as in Chapter 10, the text then takes a major departure away from what most *Technometrics* readers would consider statistical estimation theory. The author discusses a "Fisher information-based approach" that "aims to find the true probability law describing a physical phenomenon by deriving a wave equation that defines the law." This portion of the chapter (which at times has the flavor of a philosophical discussion) is evidently a synopsis of another book by the same author (Frieden 1998).

Missing from this book are many topics that have received considerable attention over the last 20 years in *Technometrics* (bootstrapping and nonparametric regression being two prominent examples). This bias toward older techniques is probably explained by the fact that this book is the third edition of a text that first appeared in 1983. The reference lists for Chapters 1–16 include only a smattering of papers and books that have appeared after 1990. (Indeed, the author has not even bothered to update the references for several books that themselves are now out in newer editions.) Chapter 17 does have numerous references from the last decade and seems to be the main motivation for this new edition, but owners of previous editions might look carefully at the Preface to this edition before purchasing.

I am reluctant to recommend *Probability, Statistical Optics and Data Testing* as a textbook for beginning students because of its lack of coverage of important new topics in statistics and because of too many quirky nonstandard definitions that will make it difficult for students to then jump into the bulk of the statistical and engineering literature. There are, however, some real gems contained in the optical applications and the numerous exercises that the author provides. I would encourage instructors and students of the physical sciences to seek out this book for some challenging applications and statistical problems. Alas, in keeping with the generally dated flavor of the book, the author states that "answers to selected problems may be obtained by writing to the author directly, and enclosing a stamped, self-addressed envelope (about $8\frac{1}{2} \times 11$ in) for return"—an odd approach in an age of e-mail permitting bulky attachments.

Donald B. PERCIVAL
University of Washington

REFERENCES

Frieden, B. R. (1998), *Physics From Fisher Information: A Unification*, Cambridge, U.K.: Cambridge University Press.
Priestley, M. B. (1981), *Spectral Analysis and Time Series*, London: Academic Press.

Teaching Statistics, by Andrew GELMAN and Deborah NOLAN, New York: Oxford University Press, 2002, ISBN 0-19-857224, xv + 299 pp., $40.00 (paperback).

The subtitle of this book is "A Bag of Tricks," and that is what this book is. The authors attempt to provide learning exercises for some of the more difficult-to-teach parts of statistics courses. The emphasis is on a noncalculus-based introductory course, but attention is also paid to mathematical statistics, decision theory and Bayesian statistics, survey sampling, and probability. The latter sections are much shorter than the extensive material for an introductory course. The basic assumption is that the students in the introductory course are reasonably adept at algebra and that the teaching is done in a traditional classroom with fewer than about 50 students.

The exercises presented here for the introductory course are by and large quite good. They are also by and large hands-on exercises that are good both to increase student interest and to increase the learning of somewhat-difficult topics. I have dog-eared my copy of the book to mark particular exercises that I plan to try in my course next semester. These include exercises on transformations of data, steps in data collection, design and data collection in experiments, conditional probability, and regression with interaction, to name but a few. The book also includes a course plan with times broken up in approximate 10-minute intervals to cover the usual topics using these exercises. Although most experienced instructors will not want to adopt this plan entirely, it does give one a good idea of about how much time these exercises might take in a real class. However, a beginning instructor could do far worse than simply adopt their plan as is.

The coverage in the case of the more advanced courses is a bit uneven. Survey sampling is dealt with quite well. The other included subjects provide good ideas, but most instructors will not want to limit themselves to just the suggestions made in this book. As usual, I have suggestions for the second edition. I would add material on quality control. The red bead experiment is but one example of some things that could be done. Exercises for elementary control charts and similar topics would be useful to many instructors. I would like to see the material for the more advanced courses increased, because those of us that teach those courses are always in need of new ideas as well as datasets. In particular, I would like to see the coverage for advanced topics extended to include advanced (beyond the introductory course level) linear models and experimental design, and if possible even some material on time series and multivariate analysis. I realize that I am asking for a lot, but new ideas are really needed in all of these areas. Finally, I would ask the authors to consider adding material for two situations that are now common at my university, introductory statistics courses taught to sections of 150–200 students and distance learning sections taught using a teleconference mode with no instructors at the remote sites. We could really use help here.

In summary, *Teaching Statistics* is a first-rate book that should be on the shelf of just about every statistics instructor. The exercises in the "bag of tricks" are very good and should be quite helpful. I also look forward to a second edition, which could be of even more help.

David E. BOOTH
Kent State University

Nonlinear Time Series: Nonparametric and Parametric Methods, by Jianqing FAN and Qiwei YAO, New York: Springer-Verlag, 2003, ISBN 0-387-95170-9, xi + 551 pp., $79.95.

This is both a monograph and a textbook on time series analysis. The focus is on nonparametric regression techniques for nonstationary and nonlinear time series. The argument for nonparametric techniques in time series is aptly provided by the authors in the Preface: "Technological inventions have led to the explosion in data collection. … Nonparametric techniques provide useful exploratory tools for this venture, including suggestion of new parametric models and validation of existing ones." The book contains the following chapters:

1. Introduction
2. Characteristics of Time Series
3. ARMA Modeling and Forecasting
4. Parametric Nonlinear Time Series Models
5. Nonparametric Density Estimation
6. Smoothing in Time Series
7. Spectral Density Estimation and Its Applications
8. Nonparametric Models
9. Model Validation
10. Nonlinear Prediction.

The authors have indicated in the Preface (p. ix) the logical structure of the chapters, and I particularly like the path of using chapters 1→2→3→4→7→8→10 as a very workable course on time series analysis. The surprising inclusion of three chapters on linear time series techniques (Chaps. 2, 3, and 7) together with introductory Chapter 1 provides a good coverage of linear time series analysis at the level of the text by Brockwell and Davis (1991). In addition, Chapter 2 contains some useful results on ergodic processes and mixing, which should prove handy for researchers. Chapter 7 contains some recent research results on spectral estimation and model validation. Presentation of nonlinear theory starts in Chapter 4, which gives a brief but essential review of some of the very useful nonlinear parametric time series models, including threshold, bilinear, ARCH, and GARCH models. The coverage of GARCH models in a general textbook is particularly appropriate, because these models have become very popular for analyzing economical and financial data (Zivot and Wang 2002). Because an important property of GARCH and other models is the martingale difference structure in the errors, I suggest that in future editions the authors make more use of the martingale difference model, which I find to be a more natural and realistic framework for many nonlinear models (e.g., Lu 1999a, b).

Chapter 5 covers kernel density estimation, which is an analog of moving average in time series. The kernel (moving average) idea underlies most other nonparametric methods covered in the book. The local polynomial modeling approach popularized by one of the authors, also known as locally weighted regression, can be considered an extension of the kernel/moving average idea. In particular, the local linear and local quadratic fittings have some nice properties, including automatic bias correction in boundary and nonuniform design problems, that make them attractive in a number of applications, including nonlinear prediction and derivative estimation. Naturally, the local polynomial method is heavily featured in this book and is discussed in the context of various time series problems (Chaps. 6, 7, 8, and 10).

Chapter 6 provides a key introduction to the techniques used later on in the book, and focuses on the univariate case only. The multivariate case is discussed in Chapter 8, which includes discussions on multivariate local polynomial regression, functional coefficient autoregression, additive models, and models for violatility function or conditional variance. I really enjoyed reading this chapter, which showcases some of the recent developments in nonparametric time series analysis. Personally, I would like to see more emphasis on the multivariate local polynomial method, which I consider more realistic models for state-space modeling and prediction (Lu 1999a) than the univariate case, which may be of pedagogical interest only. I guess the reason is probably the worry over the issue of "curse of dimensionality" (pp. 19, 317). However, I do not think this is a fair argument against nonparametric methods such as kernel or local polynomial, because in many real problems there is often low-dimensional structure in the high-dimensional state space, and there are ample evidences that this is really an advantage for high-dimensional prediction (Lu 1999b).

There are two remaining chapters on which I now comment. Chapter 9, on model validation, is basically an account of one of the authors' own work on nonparametric hypothesis testing, which I think is very novel. However, in real applications I recommend that users extensively use graphical tools and exploratory techniques to examine the raw data and residuals before making the inferential step as a final confirmation. Chapter 10, on nonlinear prediction, contains some of the very recent research results on nonlinear prediction. I think that some of the discussions on ways of expressing prediction could apply to other nonlinear prediction methods, such as that of Soofi and Cao (2002), which contains a broad range of multivariate nonlinear prediction techniques. Interested readers are also encouraged to consult the book by Chan and Tong (2001) for discussions of the interesting ramifications of nonlinear prediction from the viewpoints of dynamical systems and chaos.

Nonlinear time series certainly has been an exciting area in which to work in the period after publication of the seminal works of Priestley (1988) and Tong (1990). Given the fact that this research area has grown so fast, the authors have done an excellent job in summarizing some of the recent research work. The "Bibliographic Notes" sections at the end of each chapter should give much broader information and be very helpful in locating relevant literature. Because I have been asked to comment on the book's merits for practical workers in physical and engineering sciences, I must state that practical workers may find the level of mathematics a bit intimidating. This is undoubtedly a very theoretical book, even though most of the techniques discussed are so popular that researchers may not realize that they are using them all the time. A case in point is the lowess, or loess, methodology developed by Bill Cleveland and co-workers at Lucent, made available in S-PLUS and SAS. The book also nicely illustrates the connection between the kernel-type method and the widely known exponential smoothing method (Sec. 6.2) and nonparametric spectral estimation (Sec. 7.2). The book's website made available by the authors provides the datasets and S-PLUS programs used in the book's examples. Readers interested in trying out the techniques discussed in the book should find the website very useful.

Overall, I think that the authors should be congratulated for writing a coherent monograph on modern time series analysis with a focus on nonparametric approaches. I believe that this book will become a standard reference in this area and remain so for a long time. Graduate students in statistics, economics, and financial engineering should be happy to have a much-needed textbook on modern time series methods, which covers not only ARIMA models, but also the newer and more flexible nonlinear and nonparametric techniques.

Z.-Q. John LU
National Institute of Standards and Technology

REFERENCES

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), New York: Springer.

Chan, K. S., and Tong, H. (2001), *Chaos: A Statistical Perspective*, New York: Springer-Verlag.

Lu, Z.-Q. (1999a), "Multivariate Local Polynomial Fitting for Martingale Nonlinear Regression Models," *Annals of Instititute of Statistical Mathematics*, 51, 691–706.

——— (1999b), "Nonparametric Regression With Singular Design," *Journal of Multivariate Analysis*, 70, 177–201.

Priestley, M. B. (1988), *Nonlinear and Non-Stationary Time Series Analysis*, New York: Academic Press.

Soofi, A., and Cao, L. (eds.) (2002), *Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Dordrecht: Kluwer Academic Publishers.

Tong, H. (1990), *Nonlinear Time Series: A Dynamical System Approach*, Oxford, U.K.: Oxford University Press.

Zivot, E., and Wang, J. (2002), *Modeling Financial Time Series With S-PLUS*, New York: Springer.

Spatial Statistics and Computational Methods, edited by Jesper MØLLER, New York: Springer-Verlag, 2003, ISBN 0-387-00136-0, xiv + 202 pp., $59.95.

This book is a compilation of lecture notes from a number of distinguished professors from the United States, England, France, and Germany. This type of publication in statistics is specifically useful for postgraduate students and scientists. This book comprises more than 190 pages organized into four chapters. References, software citations, and others are given at the end each chapter. Application of Markov chain Monte Carlo (MCMC) and spatial statistics are introduced and discussed using data and software to accommodate present day needs for many scientists. Many of the examples introduced are based on the European Union's TMR network.

Chapter 1 is an introduction to MCMC and spatial statistics, written by P. Dellaportas and G. Roberts. This chapter was set as a tutorial to prepare the readers for the coming chapters. Discussions of the Gibbs sampler and Metropolis–Hasting algorithm are supported by practical implementation and illustrative example.

Chapter 2 provides an introduction to model-based geostatistics, written by P. Diggle, P. Ribeiro Jr., and O. Christensen. The Swiss rainfall data and residual contamination of Rongelap Island were introduced. Prediction and parameter estimate were introduced using Gaussian model and variogram analyses.

Chapter 3 is a tutorial on image analysis, written by M. Hurn, O. Husby, and H. Rue. An example of ultrasound imaging was introduced. The role of Bayesian statistics in image analysis and the use of the Hammersley–Clifford theorem and the Ising model are discussed in terms of to how to quantify information from images. Different algorithms are given throughout the chapter for simulations.

Chapter 4 is an introduction to simulation-based inference for spatial point processes, written by J. Møller and R. Waagepetersen. This chapter discusses and explains the meaning of spatial point process, Poisson point processes, model validation based on various kind of summary statistics, cluster processes, Cox process, simulation-based inferences for cluster point patterns, and different aspects of model construction. Two examples, weed plants and Norwegian spruces, are used throughout the chapter.

Spatial Statistics and Computational Methods successfully presents and updates the recent theoretical advances accompanied by examples and applications in simulation-based inferences. This book will be of practical use for many readers, particularly graduate students.

Maliha S. NASH
U.S. Environmental Protection Agency

Spatial Cluster Modelling, edited by Andrew B. LAWSON and David G. T. DENISON, Boca Raton, FL: Chapman & Hall/CRC, 2002, ISBN 1-58488-266-2, xiv + 287 pp., $69.95.

This book is a compilation of the recent developments on spatial cluster models. As such, it is derived from the works of many leading authors in this expanding field. The book discusses many applications to the medical and natural sciences where this topic has been very active. The techniques discussed can easily be envisioned in other settings where identification based on position is required, such as production where issues such as intrabatch defect relationships are of interest. The editors introduce the topic of spatial cluster modeling by developing a short history of the topic and identifying three distinct areas of model development: (1) point process cluster modeling, (2) spatial process cluster modeling, (3) spatio-temporal cluster modeling. The book is further divided into three sections to examine these areas in detail. Before detailing the topics within each section, I note that this text is a compilation of topics and is considered a survey of recent developments, not a comprehensive text on the subject. This has both benefits and drawbacks for practicing statisticians who want to update their appropriate skills.

The major benefit of this text is that it compiles multiple discussions of the uses of spatial cluster models and procedures into one source. This allows the reader to compare chapters and develop an understanding based on multiple sources contained within. At first I found the sections very divergent in their presentation, but as I continued to read, I found myself appreciating the fact that I could compare the discussion of a particular topic directly. In some instances the actual model definition was not clear, but by referring to neighboring chapters I could easily establish a consensus opinion. This made reading the material much easier. This brings two drawbacks to light. First, because different authors wrote individual chapters reading was discontinuous in parts and topic flow was hard to comprehend. The editors remedy this problem by dividing the topics into subcategories, but issues still arise within these sections. Second, because each chapter introduces a new topic, there is very little discussion as to why the model, procedure, or topic was being developed. Furthermore, there is little to no discussion of the benefits the presented method has over alternative methodology. This I found to be the hardest issue to overcome for a practitioner. The discussions are held mainly to simple presentations of facts and any person interested in developing the techniques for further use would have to get the original articles to gain a better understanding of why to consider such a model. Having highlighted these basic issues, I now present a detailed discussion of the contents can follow.

The first major topic presented is point process cluster modeling. With the exception of the first and last chapters of this section, the major focus is on Cox processes. Chapter 2 introduces an ad hoc graphical technique to identify intensity regions of a point process. Chapter 3 introduces both Poisson processes and Cox processes and examines the different effects of parameterization on Cox processes. Specifically, Neyman–Scott processes, log-Guassian Cox processes, and Shot noise G Cox processes are introduced, along with definitions of standard summary statistics, intensity, pair correlation, G and L functions, and estimation techniques. In most instances throughout the book calculation is provided by MCMC routines, although alternatives are suggested. Chapter 4 discusses interpolation/extrapolation techniques for different processes using Bayesian cluster model definitions. Chapter 5 introduces perfect sampling algorithms to use instead of standard Markov chain Monte Carlo (MCMC) algorithms for Bayesian cluster models. Examples are showcased but there is little explanation regarding algorithm development. The final chapter in this section examines point clustering using Bayesian models with Voronoi tessellation instead of a parametrically defined Cox process model for point intensity identification.

Chapter 6 provides a bridge to the book's second section, which covers spatial process cluster modeling defined by partition strategies. The first chapters in this section, 7 and 8, deal with different forms of Bayesian cluster models based on partition structures. The models identify how to estimate the posterior distribution of clusters and subsequently discuss such issues as prediction and interpretation. Chapter 9 describes a specialized model, the skew-normal spatial model, developed using a conjugate prior definition, and demonstrates its use by examining geological fluid flow potential. The final chapters, 10 and 11, deal with spatial cluster models involving count data. The examples are derived from astrophysical image analysis and wildlife spread.

Section III develops spatio-temporal cluster models that allow for a time series evolution of clustering. Chapter 12 provides a comprehensive overview of the parameterization techniques, and interpretations of spatio-temporal models based on differing assumptions. This is followed by two chapters applying spatio-temporal models to brain function locations and the clustering of small area health data.

In summary, *Spatial Cluster Modelling* provides a solid overview of the issues associated with spatial cluster methods but does lack a certain degree of practicality because evaluation methods and model building strategies are kept to a minimum.

Nicholas ROSE
Frequency Marketing, Inc.

Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics, edited by Abdol S. SOOFI and Liangyue CAO, Dordrecht, The Netherlands: Kluwer Academic Publishers, 2002, ISBN 0-7923-7680-3, xxviii + 488 pp., $160.00.

This is a collection of invited chapters on nonlinear dynamic techniques oriented toward applications in finance and economics. However, each chapter went through a series of reviews to ensure that the methodology was written for practical workers and contained sufficient details to enable its implementation in application problems. This is the second volume of Kluwer's Studies in Computational Finance series, exploring an area of increasing interest in statistics. The book comprises 22 chapters organized into five parts. Part One (with three chapters) focuses on the techniques of embedding and state-space reconstruction, which deal with issues analogous to variable selection and model selection in time series and regression, but use very different exploratory and data-analytic approaches with such criteria as mutual information and false nearest neighbors. Part Two covers nonlinear modeling and prediction and consists of five chapters on local polynomial prediction and volatility estimation, Kalman filtering, radial basis function, and wavelet networks. Part Three (three chapters) discusses nonlinear prediction of multivariate financial time series, NARMAX polynomial models, and modeling of dynamical systems by error-correction neural networks. Part Four (nine chapters) deals with various practical issues arising in chaotic time series data analysis and modeling, such as surrogate test (bootstrap-type methods) for nonlinearity, model validation, stationarity test, economic delayed-feedback dynamics, global modeling and differential embedding, analyzing noisy datasets, nonlinear noise reduction, model selection criteria, and effects of noise on state-space reconstruction. Part Five concludes with two interesting applications of nonlinear forecasting of noisy financial data and canonical variate (nonlinear Markov) analysis.

This book is truly a multidisciplinary effort, with contributors including economists, electrical engineers, physicists, mathematicians, and statisticians (myself and Jianming Ye). Altrough there are many books on nonlinear dynamic techniques, *Modelling and Forecasting Financial Data* is distinguished by its concerted efforts on practical relevance in financial and economic applications. I am sure that industrial statisticians will find some of the book's novel nonlinear techniques contained equally useful in solving the challenging and dazzling industrial problems.

Z.-Q. Jonh LU
National Institute of Standards and Technology

Combined Survey Sampling Inference, by Ken BREWER, New York: Oxford University Press, 2002, ISBN 0-340-69229-4, xxi + 316 pp., $50.00.

This book is written in a narrative style. In the early chapters the author describes an experience as a consulting statistician charged with coming up with an estimate of the total weight of a small herd of elephants traveling with a particular circus. The story is based on a sampling situation originally described more than 30 years ago (Basu 1971). The elephants are replaced with more typical survey sampling situations in later chapters.

At first I was suspicious of the book's story-telling approach. After all, I'm the kind of person that is more likely to read a textbook or a journal article than to pick up a novel or do any other reading for so-called "enjoyment." However, I was almost immediately interested in the story and the sampling problem that the two consulting statisticians faced, particularly the apparent contradictions that kept arising between sampling theory and practice. I especially liked the drafts of written reports to the client and the descriptions of the interactions of the consultants with their supervisor. I found the examples engaging and the writing entertaining. That is not to say that this is light reading; there is still plenty of rigor in the methods presented. The book should be read with pencil in hand and computer at your side. There are plenty of opportunities to exercise your learning by reproducing the author's results, or by taking on any of the exercises suggested at the end of each chapter.

The author advocates complementary roles of design-based and model-based sampling approaches, a controversy that I appreciate much better after reading this book. The author exhibits a very thorough knowledge of the literature. Sometimes the brief reference to literature is somewhat in passing, perhaps due in part to the author's narrative style, but there are more than sufficient citations to help you find the details if you are interested.

The book comprises 15 chapters, beginning with simple random sampling and stratification. Following this, a couple of chapters cover design-based and prediction-based ratio estimation. Then there are three chapters devoted to various forms of regression estimation. Multistage and multiphase sampling are then covered, and the last couple of chapters cover more specialized topics, including tracking and rotating samples, coping with extreme values, and resampling methods.

Although *Combined Survey Sampling Inference* is not a typical textbook, I can envision using it as a textbook. It is not a typical reference book either, although you can certainly look back and find any of the formulas you need for various methods after having read the book. It would have been nice to have had solutions to the exercises, the data available electronically, and perhaps even sample programs to do computations. However, I really enjoyed working along with the statisticians, and highly recommend this book to anyone that wants to better understand the controversy over, and complementary roles of, design-based and prediction-based sampling.

Steven M. LALONDE
Rochester Institute of Technology

REFERENCE

Basu, D. (1971), "An Essay on the Logical Foundations of Survey Sampling, Part One" (with discussion), in *Foundations of Statistical Inference*, eds. V. P. Godambe and D. A. Sprott, Toronto: Holt, Rinehart and Winston, pp. 203–242.

Making Sense of Data, by Donald J. WHEELER, Knoxville, TN: SPC Press, 2003, ISBN 0-945320-61-2, xii + 395 pp., $59.00.

SPC applies to the service sector as well as the manufacturing, construction, and industrial sectors, but unfortunately there is not much literature on applying SPC tools to the service industries. The reasons for this are apparent. Unless service industry executives and managers conclude that there are economic benefits to the application of TQM and its central set of tools, SPC, its use and application will not grow.

This text is a very good attempt at introducing these tools to service industries. The book begins by explicitly stating the rationale for *continuous improvement* and *service quality*. The author wisely introduces early in the text the notion that understanding one's process is essential in determining the need for and economic benefits from the application of SPC. It is very important that decision makers in industry understand the role of SPC and how it will enable them to become better managers. This book belongs on the bookshelf of not only service managers, but also should be read with by these same people.

I am a teacher, a university professor, and so I wish to review this book also from the standpoint of educating future managers and executives in the service sector. In my classes in SPC, this book could be a very useful text in those environments where the service industries are emphasized. We might note that to truly understand the sense of data, students need to be thoroughly knowledgeable in statistics and probability. I would suspect that instruction in undergraduate and MBA programs and programs in technology require a solid 1-year introduction to these topics taught at the appropriate level. This would be the appropriate prerequisite for courses using this book as a text.

The book's reading level is appropriate for students in the applied programs noted earlier. The writing style is direct, informal, and nonthreatening to students. Stated differently, I endorse the writing level of this text over what I have observed in many texts in the field of SPC, which are written at levels well beyond the reading level of students in academic programs. The examples and illustrations appear to be based on the author's experiences, which is a real positive. Students and faculty desire to learn from the experience of others, and this book lends itself well in this direction. The writing style puts the reader in the center of the learning process, and much of the material can be self-taught. The mathematics is not at a high level and does not require calculus. Students do need to be well grounded in algebra and geometry; the importance of understanding mathematics is illustrated by the kinds of examples and applications throughout the text.

The question remains on what the text is missing. There is no doubt that twenty-first century students require that books be published in multiple colors and that graphic images be drawn in multiple colors as well. Except for the cover, the book's visual style is a throwback to an earlier era when textbooks in advanced methods were not always cosmetically attractive or exciting for students to read. Images must be drawn in a style that induces the reader (students and faculty) to want to read more. The book's style is certainly not a positive and will not induce greater readership. I believe that the cosmetic style of textbooks is very important, because we live in an era when students have many choices. These choices are important because students should want to read books for reasons other than avoiding a below-average grade on an examination.

The book's topical structure is good, but nevertheless a number of items are left out of the discussion. What is essentially missing is the use of computer software for SPC. One need not look too deeply to discover this omission. In my opinion, books of this type should include this material, and I believe Minitab software is the appropriate program for use in the classroom. Minitab is easy to learn, and most college business students have used this software or similar software or can learn it easily. Most important, Minitab contains the various models for applying SPC in the service industries. The modules in Minitab can easily be integrated into the text and easily used to enhance the course in SPC for the service sector. This is not to say that other software cannot be used, but is as the leading software in SPC, Minitab would greatly enhance the book. I would never use the spreadsheet software, Excel as the software application to accompany this text, because it simply does not contain the appropriate modules. Furthermore, it is very difficult software to manipulate in SPC applications. In my experience, students have a great deal of difficulty using Excel for SPC applications.

The book omits many types of QC charts that have wide application in the service sector. For example, exponentially weighted moving average charts (EWMA) are not included in the discussion. I am certain that other reviewers will point out other methods that are omitted.

I am also disturbed that the methods of acceptance sampling are not fully integrated in the text. No doubt there are more applications of acceptance sampling in manufacturing; however, all purchasing decisions do include the notion of incoming as well as outgoing quality. There is also a dearth of end-of-chapter exercises, and capability studies are not included.

I believe that the author has made a very good attempt to introduce a new book in the curriculum of business administration programs. A great deal more work remains to be done to produce a much better text. In particular, decision rules for QC charts include probability-based methods other than the control limits. One can see these methods easily by using such software as Minitab, SAS, and NWA Quality Analyst.

Jeffrey E. JARRETT
University of Rhode Island

Quality From Customer Needs to Customer Satisfaction (2nd ed.), by Bo BERGMAN and Bengt KLEFSJÖ, Lund, Sweden: Studentlitteratur, 2003, ISBN 91-44-04166-7, 606 pp., $68.00.

The past 15 years have seen the promulgation of many "quality" books. Most feature a total quality management (TQM) focus, whereas others are more specialized, some primarily quantitative. Although the ubiquity of "quality" as a buzzword has waned, there is no dearth of textbooks on the subject. When reviewing a book in this genre, consideration must be given to the potential utility to both the academic and practitioner audiences. *Quality From Customer Needs to Customer Satisfaction* (hereafter, QCNCF) is certainly applicable for both strata.

Bergman and Klefsjö offer a fairly comprehensive volume on quality management from a "soft-side" management and production perspective. There are the obligatory sections on statistical process control (SPC) and experimental design, but the target audience here is clearly the current or aspiring manager. The reader need not have a background in management, because the book is quite readable without a heavy dose of jargon. The required math for the "quantitative" sections is basic algebra. On the academic side, this textbook would be appropriate for an undergraduate course in TQM.

The book begins with an introduction to quality and a historical recap of the gurus and their primary contributions. The topics covered include the quality/productivity relationship, design issues, quality function deployment, reliability, capability, internal/external customers, leadership, processes, quality systems (and applicable standards), and the aforementioned quantitative methods. Concluding with a set of tables and references, the book follows a setup similar to most quality handbooks, as opposed to that of college textbooks. Overall, the coverage is reasonable for the stated audience. A professor using the book would most likely be unable (or unwilling) to cover every chapter, but all of the usual suspects for a general TQM course are more or less here.

A major strength of the book is the interface between traditional management topics and quality-specific management tools. Discussions of the Hawthorne effect and Maslow's hierarchy are juxtaposed with internal customer satisfaction and participation and coworkership. Failure mode and effect analysis (FMEA) and fault tree analysis are discussed in terms of top-down and bottom-up management. Such a presentation is especially helpful for the reader with little or no background in basic management.

The reference section is excellent, providing sources for the interested reader in all subject areas covered. Each chapter also offers a list of references. Again this is crucial, given the book's lack of detailed coverage of certain issues. The table section, is not as strong, especially with respect to the constants for control charts. I suggest that if the authors wish to discuss acceptance sampling in the future, they include some table at the end of each chapter for clarification. I do *not* recommend, however, that the section on acceptance sampling be retained in future editions; it could be eliminated without any substantive loss.

A major drawback of this book is the fact that there are no chapter exercises. In quality-speak, this may offer opportunities for improvement to those willing to write a companion exercise manual for the book. However, it presents a major obstacle for the academician considering this book as a potential undergraduate textbook.

Perhaps the biggest disappointment is the placement of the discussion of processes in Chapter 19. One of the basic tenets of the quality movement and departure from past business practices is the focus on process rather than on product. As a management-oriented work, the book should address this fundamental issue in a much earlier chapter.

As with most general works, the book sacrifices detail for breadth. The authors attempt to walk the tightrope of offering a resource lexicon that can nevertheless be used in a single college course. The usual result of such an effort better available stand-alone resources (e.g., Juran 1999) and textbooks (e.g., Gitlow, Gitlow, Oppenheim, and Oppenheim 1989). One should not infer that this is an indictment of the current work per se, but rather should view it as an expectations caveat for potential consumers.

Overall, I recommend QCNCF as a reference manual for the quality professional or for the practitioner with a primarily quantitative background. I would be somewhat remiss if I did not emphasize that it is *a* reference manual, and not the most highly recommended one. *Juran's Quality Control Handbook* (1999) remains the standard in that niche (though it weighs more than the Tampa Bay Buccaneer's defensive line). I would not recommend QCNCF as a textbook for a TQM course at the undergraduate level; the text by Gitlow et al. (1989) is better. The absence of exercises is a serious omission in any technical academic tome. On a final note, if one needs a quantitative quality resource, I recommend the book by Montgomery (2001).

John D. BARRETT
University of North Alabama

REFERENCES

Gitlow, H. S., Gitlow, S. J., Oppenheim, A., and Oppenheim, R. (1989), *Tools and Methods for Quality Improvement*, Boston: Irwin.
Juran, J. M. (ed.) (1999), *Juran's Quality Handbook*, New York: McGraw-Hill.
Montgomery, D. C. (2001), *Introduction to Statistical Quality Control* (4th ed.), New York: Wiley.

Editor Reports on New Editions, Proceedings, Collections, and Other Books

This section reports on new editions of books previously reviewed in *Technometrics*, collections of papers and conference proceedings, and other statistics books that should have some interest for the readership. Selections and comments do not represent any perspective of the editor's employer or of the sponsoring societies.

Eric R. ZIEGEL
BP

Introduction to the Practice of Statistics (4th ed.), by David S. MOORE and George P. MCCABE, New York: W. H. Freeman, 2003, ISBN 0-7167-9657-0, xxv + 909 pp. + CD, $97.95.

This book, now out in its fourth edition (4E), originated with its first publication in 1989. The *Technometrics* report on the third edition (3E) (Ziegel 2001) gave references for the reviews of the earlier versions. The statistical content of the book has not been changed appreciably. This revision features a considerable upgrading of the examples and the exercises. Also, the internet and media content, which already provided additional chapters on CD-ROM for the 3E, has continued to expand and now includes online applets to provide interactive graphics for visualizing various concepts. There is even a chapter on quality control that has been taken from a new book by these same authors (Moore, McCabe, Duckworth, and Sclove 2003).

The authors contend that this book "was the first book to successfully combine attention to broader content and reasoning with comprehensive presentation of the most-used statistical methods" (To the Teachers, p. ix). This aspect of the book is both a strength and a weakness, at least for the practicing statistician and coworkers. The first 275 pages are devoted to three chapters on collecting and examining data. These pages are filled with applications and plots. The remaining 625 pages compose Part II, "Probability and Inference," and Part III, "Topics in Inference." This material has the full complement of equations. The primary computing support comes from the use of a hand calculator.

The book's size, and presumably its success, can be attributed partly to its efforts to be appropriate to a wide variety of students. The authors (p. xii) indicate "general undergraduates from mixed disciplines," "a quantitatively strong audience—sophomores planning to major in statistics," and "beginning graduate students in such fields as education, family studies, and retailing" as appropriate for their book. One group that is not specified is engineers and scientists in industry. A lack of emphasis on using statistics software is a minus, but certainly those from this group who have had a good college course in statistics would find this book a very useful desktop reference. I would not use it for an industrial short course.

Pedagogy prevails here. Professor Moore was relatively new at Purdue when I was in graduate school there. He was extremely popular in the classroom. My recent Purdue alumni magazine indicated that Professor Moore had been honored by the students (apparently a frequent occurrence) as one of the top instructors at Purdue. These awards must in part reflect the pleasures that the students derive from their time with this book.

REFERENCES

Moore, D., McCabe, G., Duckworth, W., and Sclove, S. (2003), *The Practice of Business Statistics*, New York: W. H. Freeman.

Ziegel, E. (2001), Editor's Report on *Introduction to the Practice of Statistics* (3rd ed.), by D. Moore and G. McCabe, *Technometrics*, 43, 245.

An Introduction to Multivariate Statistical Analysis (3rd ed.), by T. W. ANDERSON, Hoboken, NJ: Wiley, 2003, ISBN 0-471-36091-0, xx + 721 pp., $84.96.

I have noted in previous reports on books about multivariate analysis that I took my first course in the Psychology Department so that I could use an early edition of the text by Morrison (1990), in lieu of the statistics department course that used the first edition (1E) of the classic multivariate statistics book by T. W. Anderson. It appears here in its third edition (3E). That 1E, which seemed very formidable to me, was published in 1958. A second edition (2E) appeared in 1984. The 3E, considerably larger now than the 1E, remains a very challenging book to pick up and use.

In the Preface to the 3E, the author notes that "earlier editions included some methods that could be carried out on an adding machine." This book, however, makes little effort to be anything particularly different than the 1E or the 2E where statistical computing is concerned. This is still a very serious and comprehensive book on the statistical theory of multivariate analysis. The book by Fleury (1997), reported for *Technometrics* by LaLonde (1999), is the only other recent book on the theory of multivariate analysis that I can recollect.

The primary change from the 2E has been the addition of a final chapter, "Patterns of Dependence and Graphical Models." This chapter includes sections on undirected graphs, directed graphs, and chain graphs. There is also some concluding material on statistical inference relating to these graphs. A second feature of the 3E is the addition of coverage of elliptically contoured distributions. These relax the restriction from multivariate normal distributions that the mean vector and covariance matrix are constant on concentric ellipsoids. The Preface (p. xvi) also notes the inclusion of discussion of reduced-rank regression in the chapters on canonical correlations and the distribution of characteristics roots and vectors.

The first chapter and the introductions to the other chapters provide useful and informative reading. Otherwise, this is a book for doctoral students, not for industrial statisticians. The recent book by Rencher (2002), reported for *Technometrics* by Ziegel (2003), is a good current practical book on multivariate statistical methods.

REFERENCES

Flury, B. (1997), *A First Course in Multivariate Statistics*, New York: Springer-Verlag.

LaLonde, S. (1999), Review of *A First Course in Multivariate Statistics*, by B. Flury, *Technometrics*, 40, 262.

Morrison, D. F. (1990), *Multivariate Statistical Methods* (3rd ed.), New York: McGraw-Hill.

Rencher, A. (2002), *Methods of Multivariate Analysis* (2nd ed.), New York: Wiley.

Ziegel, E. (2003), Editor's Report on *Methods of Multivariate Analysis* (2nd ed.), by A. Rencher, *Technometrics*, 45, 109.

The Analysis of Time Series (6th ed.), by Chris CHATFIELD, Boca Raton, FL: Chapman & Hall/CRC, 2004, ISBN 1-58488-317-0, xiii + 333 pp., $49.95 (softcover).

Ziegel (1997) reported on the fifth edition (5E) of this text. Somewhere deep in the archives one perhaps could find a review of the first edition, published almost 30 years ago. Disdainful as I tend to be when the edition number exceeds 2 or 3, I complimented the author in my report on the 5E for the extensive amount of new material, particularly because it accommodated all of the criticisms about missing content that were made by Dickey (1991) in the review of the fourth edition. I recommended the 5E, like the sixth edition (6E) an inexpensive paperback, as a good choice for an addition to one's personal library. That perspective prevails for the 6E as well, but it is not a necessary replacement for one's 5E.

The book comprises 14 chapters. The first 12 are largely the same in the 6E as in the 5E. Each has grown by a couple of pages, but only a few have substantive additions. The chapter on forecasting has a new section on prediction intervals. The penultimate chapter on advanced topics includes three new topics: fractional differencing, long-memory models, and tests for unit roots. A couple of the more advanced chapters have been appended with bibliographies.

Perhaps the highlights for the new material in the 6E come in the section "Handling Real Data," appended to the end of the second chapter, "Simple Descriptive Techniques," and the new final chapter, "Examples and Practical Advice," an upgrade of a 5E appendix on worked examples. This author has been analyzing time series data for more than 30 years. He also has written several other books including one (Chatfield 1995) that is a problem-solving guidebook for statisticians. This is not just advice, this is wisdom. He talks about software, analyzes several sets of data, and especially exudes enthusiasm for analyzing time series data.

If you do not own this book and have any interest in time series data, do buy a copy. Happily the new publisher, CRC Press, has continued Chapman & Hall's fine policy of publishing inexpensive soft-cover editions. After all, the author notes (Preface, p. xi) that "healthy sales figures" were a primary factor in his continuing to pump out revisions regularly. He also states (ibid.) that "I do plan that this should be the sixth and final edition."

REFERENCES

Chatfield, C. (1995), *Problem Solving: A Statistician's Guide* (2nd ed.), London: Chapman & Hall.

Dickey, D. (1991), Review of *The Analysis of Time Series: An Introduction* (4th ed.), by C. Chatfield, *Technometrics*, 33, 364–364.

Ziegel, E. (1997), Editor's Report on *The Analysis of Time Series* (5th ed.), by C. Chatfield, *Technometrics*, 39, 342–343.

Modelling Binary Data (2nd ed.), by David COLLETT, Boca Raton, FL: Chapman & Hall/CRC, 2003, ISBN 1-58488-324-3, xviii + 387 pp., $59.95.

This second edition (2E) follows 12 years after the publication of the first edition (1E). It was completely described and also very well received by Thomas (1993), who described the book (p. 224) as "application-oriented," "very well written," and "an excellent resource." He stated (ibid.) that "the illustration of the methods through the examples is very effective." As the revisions have only enhanced all these perspectives, my judgment remains the same as Thomas's. This is a good and inexpensive resource and an excellent addition to one's desktop collection.

The 1990s brought a lot of development in methods and associated computations for analyzing binary data. It is not surprising that the primary additions to the book are chapters on two of these areas of development, mixed models and exact methods. Mixed models are applied for binary responses in multilevel data, longitudinal data, repeated-measures data, data in meta-analyses, and models for overdispersion. Exact methods are especially focused on logistic regression and result in a variety of exact tests that allow effective analysis for smaller datasets. In addition, the penultimate chapter on "Other Methods" now has been enhanced with a sizeable addition of methods for ordered categorical data. Finally, the book's last chapter, on statistical computing for binary data, has been completely revamped. The 1E was built around applications with GLIM, a software package that has not traveled well outside of the U.K.. Now a variety of software packages are considered useful for the analysis of binary data.

All chapters begin with nice overviews and conclude with sections devoted to "Further Reading." Competitors for this particular book are limited to books on logistic regression or generalized linear modeling, such as that by Myers, Montgomery, and Vining (2002), which was reviewed for *Technometrics* by Ziegel (2002).

REFERENCES

Myers, R., Montgomery, D., and Vining, G. (2002), *Generalized Linear Models*, New York: Wiley.

Thomas, E. (1993), Review of *Modelling Binary Data*, by D. Collett, *Technometrics*, 35, 224.

Ziegel, E. (2002), Review of *Generalized Linear Models*, by R. Myers, D. Montgomery, and G. Vining, *Technometrics*, 44, 287–288.

Sample Survey: Principles & Methods (3rd ed.), by Vic BARNETT, New York: Oxford University Press, 2002, ISBN 0-340-76398-1, xiv + 241 pp., $17.95 (paperback).

Appearing here in its third edition (3E), this book was reviewed in its first edition (1E) by Minton (1993). Although sampling books are generally smaller than books in most other areas of statistics, this one is certainly very modest in size (and in price!). About the 1E, Minton noted (p. 235) that "the book packs a tremendous wallop into only 173 pages." Not its only attribute, as Minton also reported (ibid.) that "a beginner will find this book easily readable, the examples forthrightly illuminating the main idea of each technique." Understandably, Minton concluded that "Barnett's book can be recommended highly."

The first six chapters remain the same as in the first two additions. Following Part 1, which has a single chapter of concepts, Part 2, "Simple Random Sampling," comprises four chapters. These chapters have had additions across the two revisions from the 1E to the 3E, such as multistage procedures and probabilistic sampling. The remainder of the book has been considerably enhanced, however. Part 3, "Practical Aspects of Carrying Out a Survey," which contains a single chapter, "Implementing a Survey," has had updates on a number of topics, including missing data and imputation, panel surveys, telephone surveys, and electronic surveys. Specialized books on these types of features of surveys, usually conference proceedings or invited chapters, are regularly reported in *Technometrics*, most recently the book by Biemer and Lyberg (2003), reported by Ziegel (2003).

Entirely new to the 3E is Part 4, "Environmental Sampling." Here one finds Chapter 7, "Sampling Rare and Sensitive Events," and Chapter 8, "Sampling Natural Phenomena." An example of a rare event is a genetic variant or a search for accounting errors. Issues very personal to individuals, such as sexual behavior, are examples of sensitive events. Sampling methods, usually preceded by preliminary analysis or screening, include disproportionate sampling, multiplicity sampling, snowball sampling, and network sampling. There also are sections on composite sampling and ranked set sampling. For sampling of natural species, the reader can learn about line-intercept sampling, transect sampling, and capture-recapture methods. The type of coverage in Part 4 has always led me to recommend the book by Thompson (2002), most recently reported by Ziegel (2002), as the choice for a desktop reference in sampling. As a low-cost and more concise alternative, this book would be a fine choice as well.

REFERENCES

Biemer, P., and Lyberg, L. (2003), *Introduction to Survey Quality*, Hoboken, NJ: Wiley.

Minton, P. (1993), Review of *Sample Survey Principles and Methods*, by Vic Barnett, *Technometrics*, 35, 235.

Thompson, S. (2002), *Sampling* (2nd ed.), New York: Wiley.

Ziegel, E. (2002), Editor's Report on *Sampling* (2nd ed.), by S. Thompson, *Technometrics*, 44, 407.

——— (2003), Editor's Report on *Introduction to Survey Quality*, by P. Biemer and L. Lyberg, *Technometrics*, 45, 277.

Small Area Estimation, by J. N. K. RAO, Hoboken, NJ: Wiley, 2003, ISBN 0-471-41374-7, xxiii + 313 pp., $99.95.

In the realm of sampling, a small area is a portion of some population for which estimation of adequate precision cannot be done—for example, determining the average age of all illegal immigrants in Chicago. This book describes the methodology for doing this type of estimation from available sample data. As the author notes in his Preface (p. xxi), it is seldom possible to collect sufficient data from sampling to enable estimation for all domains of interest. With linkage to supplementary data through appropriate models, estimation can be accomplished.

The book has 10 chapters. A brief first chapter provides an introduction to the problem, the methods used, and the application for a number of examples. Chapter 2 covers direct domain estimation, the type of estimation that is described in standard sampling textbooks such as that by Levy and Lemeshow (1999), which was reported by Ziegel (2000). It is apparent from this chapter that this book focuses more on methodology than on applications. The examples are brief, and theorems with proofs are included. Chapter 3 presents the traditional demographic methods. These use indirect estimation and implicit linking models, such as regression equations and contingency tables. Indirect domain estimation is studied in Chapter 4. This procedure increases the effective sample and thereby decreases the standard error. The chapter presents synthetic, composite, and shrinkage types of estimators. Chapter 5, "Small Area Models," is concerned with explicit models, mixed models that account for the variation between areas with random area-specific effects.

As the complexity of the methodology increases, the book becomes more mathematical, until with Chapter 6 one finds a whole chapter on the theory of empirical best linear unbiased prediction (EBLUP). Chapter 7 presents the methods drawn from the theory. This chapter includes a number of good examples. Chapter 8 presents a number of models that are extensions of EBLUP. Chapter 9 introduces empirical Bayes models, which are determined to be particularly appropriate for binary and counts data. Finally, hierarchical Bayes methods are featured in Chapter 10, which has a succession of models with nice illustrations.

REFERENCES

Levy, P., and Lemeshow, S. (1999), *Sampling of Populations* (3rd ed.), New York: Wiley.

Ziegel, E. (2000), Editor's Report on *Sampling of Populations* (3rd ed.), by P. Levy and S. Lemeshow, *Technometrics*, 42, 326.

SAS® for Forecasting Time Series (2nd ed.), by John C. BROCKLEBANK and David A. DICKEY, Cary, NC/Hoboken, NJ: SAS Institute/Wiley, 2003, ISBN 1-59047-182-2/0-471-39566-8, x + 398 pp., $64.95.

I have promoted many of these large-format SAS books over the past 10 years, most recently that by Fan, Felsővályi, Sivo, and Keenan (2003); see the editor's report by Ziegel (2003). New SAS books that are mostly redundant with earlier SAS books have lately been going to other reviewers, most recently the book by Fernandez (2003); see the forthcoming review by Caby (2004). The first edition (1E) of this book by Brocklebank and Dickey, published in 1986, has always been a book that I have used a lot. Today I still use SAS for most of the time series modeling that I do. It was nice to see the arrival of this updated edition.

The 1E has been preserved mostly intact within the second edition (2E). The six chapters from the 1E appear in the 2E with mostly identical organization. These chapters cover various aspects of ARIMA modeling, state-space modeling, and spectral analysis. The primary enhancement has been the inclusion of many new examples. Because borrowing the SAS code and imitating the illustrations has been a primary operational mode for me in using these books from SAS Books by Users Press, the new examples are welcome additions. The "Arima Model: Special Applications" chapter from the 1E has been retitled as "ARIMA Model: Introductory Applications." It dwells on basic Box–Jenkins models enhanced with seasonality and transfer function components. A new "Special Applications" chapter has been added. This begins with a brief presentation of time series regression, then presents an extensive section on time series modeling with unequal error variances at various points in time. For this situation, the use of the ARCH and GARCH options within PROC AUTOREG is explained and illustrated. There also is an additional long section with methods and illustrations on cointegration models for multivariate time series. These involve the use of PROC VARMAX in SAS.

The final chapter, "Data Mining and Forecasting," is also new. It is concerned with mostly automatic methods for simultaneously doing forecasting for many different time series. There is an illustration of using SAS Web Analytics to evaluate web traffic data. The SAS Time Series Forecasting System is demonstrated. The use of PROC HPF to forecast many time series at once is also shown. The new material and the update of the excellent 1E, now 17 years in the past, certainly make the 2E a necessary purchase for any user of SAS time series modeling methods.

REFERENCES

Caby, E. (2004), Review of *Data Mining Using SAS Applications*, by G. Fernandez, *Technometrics*, 45, forthcoming.

Fan, X., Felsővályi, Á., Sivo, S., and Keenan, S. (2003), *SAS® for Monte Carlo Studies*, Cary, NC: SAS Institute.

Fernandez, G. (2003), *Data Mining Using SAS Applications*, Boca Raton, FL: Chapman & Hall/CRC.

Ziegel, E. (2003), Editor's Report on *SAS® for Monte Carlo Studies*, by X. Fan, Á. Felsővályi, S. Sivo, and S. Keenan, *Technometrics*, 45, forthcoming.

Extreme Values in Finance, Telecommunications, and the Environment, edited by Bärbel FINKENSTÄDT and Holger ROOTZÉN, Boca Raton, FL: Chapman & Hall/CRC, 2004, ISBN 1-58488-411-8, xix + 405 pp., $89.95.

This sizeable book comprises the seven invited papers from a conference on extreme value theory and its applications. The conference was held in Gothenburg, Sweden in December 2001. These papers vary in length from 16 to 102 pages. (One can only wonder what the session length would be for a 102-page paper.) The initial 70-page overview chapter gives motivating examples, discusses univariate extreme value theory and estimation, and gives several examples each for environmental, insurance, and financial time series extreme values.

Chapter 2, "The Use and Misuse of Extreme Values in Practice," discusses the options for extreme value analysis, which involve the choice of a model, the use of available information, the consideration of estimation uncertainty, and an allowance for nonstationary effects. Chapter 3 discusses the various aspects of the popular value-at-risk (VAR) approach to risk management in finance that relate to extreme value methodology. This 70-page chapter discusses several models for optimizing stock portfolios. Chapter 4 continues the discussion of the use of extremes in economics. Chapter 5, "Modeling Dependence and Tails of Financial Time Series," is the aforementioned 102-page chapter. It discusses the relationship between tail behavior and dependence structure. It teaches the "facts" of financial data, then looks at models, such as GARCH, that try to represent empirical behavior. A considerable dose of theory accounts for much of this chapter's length.

The last two chapters offer some additional applications and insights on extreme values. Chapter 6, "Modeling Data Networks," deals with the heavy tails in quantities such as file sizes, transmission rates, transmission durations, and CPU job completion times, which are familiar to all of us who upload and download large amounts of data. Chapter 7 considers the situation where many of the extreme value occurrences that one might encounter are actually multivariate applications.

Handbook of Statistics 21: Stochastic Processes: Modelling and Simulation, by D. N. SHANBHAG and C. R. RAO, Amsterdam: Elsevier, 2003, ISBN 0-444-50013-8, xvii + 1000 pp., $140.00.

This book is another volume in Elsevier's "Handbook of Statistics" series. By now I have reviewed seven or eight of these handbooks. I may have written the same review seven or eight times, too, because I seem to repeat the same comments. Perhaps if the publisher were to call this series the "Handbook of Research Statistics," I might be more enthusiastic about books in which usually more than 95% of the participating authors are from academia. Here 4 of the 39 authors are actually from outside of the university environment. However, all four of these are employed by government or industry departments where one can do research in statistics, so things are seemingly not much different with this volume.

One thing that is different is that this book, Volume 21 in the series of handbooks, is that it is actually intended as the sequel to Volume 19 (Shanbhag and Rao 2001). That volume was not reviewed by *Technometrics*, because it clearly was focused primarily on the theoretical aspects of stochastic processes. This volume comprises 23 chapters, ordered seemingly haphazardly. Working from the perspective of the title, 14 of the 23 chapters could be categorized as modeling chapters, and 5 of them are concerned with simulation. The other four chapters appear to be chapters that were received too late for inclusion in the earlier volume on theory and methods.

There are some interesting applications for the modeling chapters. These include control for manufacturing systems, models for DNA replication, optimal design and control for telecommunications systems, image analysis problems, reliability, discrete variate time series, astronomy, and ion channels. Simulation applications include self-similar processes, modeling epidemics, and extreme value models. This entire set of chapters has the feel more of a book on mathematical statistics than of a book on applied statistics.

REFERENCE

Shanbhag, D., and Rao, C. (2001), *Handbook of Statistics 19: Stochastic Processes: Theory and Methods*, Amsterdam: Elsevier.

Statistical Data Mining and Knowledge Discovery, edited by Hamparsum BOZDOGAN, Boca Raton, FL: Chapman & Hall/CRC, 2004, ISBN 1-58488-344-8, xxviii + 588 pp., $99.95.

This book emanates from an international conference in Knoxville, Tennessee in June 2002. The book's 34 chapters represent a selection of the 70 papers presented at the conference. Eight interesting keynote lectures are listed in the Preface (p. viii), but unfortunately only three of these landed in the book. The conference apparently was very much an academic affair with 33 of the 36 authors from academia and only 1 from industry.

The book chapters appear in apparent random order. The three keynote papers are titled: "The Role of Bayesian and Frequentist Multivariate Modeling in Statistical Data Mining"; "Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms"; and "Econometric and Statistical Data Mining, Prediction, and Policy-Making." The other interesting statistics chapters are as follows: "Data Mining and Traditional Regression"; "An Extended Slice Inverse Regression"; and "Kernal PCA for Feature Extraction With Information Complexity."

The book's title generally would have been more apt had the word "statistical" been omitted. Some of the remaining papers deal with issues in or aspects of the data mining process for various types of data. The rest apply the tools in different types of applications, including chemical warfare agents, disability survey data, polygraph data, sensor validation in nuclear power plants, time series prediction, plant devices monitoring, customer preferences identification, brand attribute perceptions, steel making, value-added services for scientific libraries, human reliability analysis, scientific and technological innovation, and semantic conference organization. The technical content of these chapters is mostly not very advanced.

Statistical Modeling and Analysis for Database Marketing, by Bruce RATNER, Boca Raton, FL: Chapman & Hall/CRC, 2003, ISBN 1-57444-344-5, xx + 362 pp., $59.95.

As the Preface (p. iii) begins, "This book is a compilation of essays." That is the only negative comment that I have about this book, which is remarkable if one considers the tone of the comments that I have made in the various reports that I have served up in these pages for a host of unexceptional data mining books. For one thing, this book is not at all focused on the myriad of data processing steps that data mining practitioners undertake once their datasets are available, an effort that is often described as more than 80% of the time needed for a data mining project. For another thing, the essays in this book were written by a statistician. Subtitled "Effective Techniques for Mining Big Data," the book focuses strictly on modeling with a designated dependent variable. Also, the author builds a historical basis for data mining that includes EDA as a significant catalyst—a perspective that should interest most statisticians.

The first few chapters (or essays) deal with the methods for developing predictions for dependent responses: correlation and multiple regression for measured responses and logistic regression for categorical responses. A number of chapters discuss a methodology for including interactions in the logistic regression model. This method, called CHAID, is an automatic interaction detection procedure from the machine-learning world that utilizes regression trees. Applications for CHAID include market segment classification and modeling and the finding of replacements for missing-data values and categories.

With the modeling tools and methods in hand, the author next provides several chapters that deal with database marketing, customer identification developed from all of the information in these big databases. Separate chapters cover using bootstrapping and visualization to validate and evaluate the database models. Two chapters explore the use of genetic modeling, which is proposed as a method for variable selection. The book ends with a chapter on coefficient-free models, another product of the machine-learning world.

Note that this book's content is the antithesis for technical difficulty to the fine statistical learning book by Hastie, Tibshirani, and Friedman (2001); see the long-winded *Technometrics* review by Ziegel (2003). Where that book was replete with technical detail, this book is almost devoid of technical content. More than half of these essays previously appeared in the *Journal of Targeting, Measurement and Analysis for Marketing*. However, the author's background and experience make this book a useful addition to the library of any statistician involved with data mining. There is even SAS® code in a few of the chapters.

REFERENCES

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning*, New York: Springer-Verlag.

Ziegel, E. (2003), Review of *The Elements of Statistical Learning*, by T. Hastie, R. Tibshirani, and J. Friedman, *Technometrics*, 45, 267–268.

Quantitative Methods in Population Health, by Mari PALTA, Hoboken, NJ: Wiley, 2003, ISBN 0-471-45505-9, xxvi + 311 pp., $89.95.

It was this book's subtitle, "Extensions of Ordinary Regression," that first caught my attention. In the chemical industry we still use ordinary least squares regression analysis for many applications. This book, which is based on some actual observational studies from the author's research projects, provides the most pages of illustrations relative to pages of text of any book that I can recall. The illustrations include fairly detailed instructions on obtaining the results using SAS. The book is intended as a third course in applied statistics for students familiar with ordinary least squares and logistic regression methods. Some background with SAS is also presumed.

This is a fantastic book for practitioners. The three examples involve health studies of lung development of newborns of low birth weight, diabetes in persons under age 30, and sleep-related problems in individuals over age 65. Normality assumptions in ordinary regression are not met, because the responses are binary or perhaps categorical. Independence assumptions are not satisfied, because longitudinal data are often obtained, which results in correlated residuals. Variability is nonconstant, because it is often a function of the level for the measured responses.

In the early chapters, maximum likelihood estimation using PORC MIXED is compared with ordinary regression using PROC REG. There is a certain amount of methodology (which is done using matrices) in the developmental material, but generally the focus is on the illustrations. PROC MIXED is used to deal with the unequal variance situation. Contrasts are made to response transformations, and weighted regression with PROC REG is used.

Somewhat more complex methodology development accompanies the analyses for correlated response data. There also are chapters on a transformation approach and on a random-effects methodology for handling the same types of data. Again these are supported by analyses done using PROC MIXED. Next, regression methods are generalized to nonnormal responses, which results in the use of PROC GENMOD through applications to the exponential family of distributions. Here a SAS comparison is made versus PROC LOGIST. There are additional separate chapters for binomial and Poisson outcomes. A final chapter combines the correlated and nonnormal considerations to demonstrate the use of generalized estimating equations in PROC GENMOD.

Leading Six Sigma, by Ronald D. SNEE and Roger W. HOERL, Upper Saddle River, NJ: Prentice-Hall, 2003, ISBN 0-13-008457-3, xxi + 279 pp., $29.95.

Although it is certainly not apparent from this book's title or even its subject matter, the two authors were once both very successful industrial statisticians. Their career directions have clearly evolved, as the subtitle, "A Step-by-Step Guide Based on Experience with GE and Other Six Sigma Companies," indicates. This book is both published and priced as a high-volume business press book. The authors claim that all the other Six Sigma books have a "paucity of practical guidance on the deployment of Six Sigma" (Preface, p. xvi). Reports in this section have more than once praised the Six Sigma book by Pande, Neuman, and Cavanagh (2000), reported by Ziegel (2002), for its content on how to implement Six Sigma. Also, the book by Breyfogle, Cupello, and Meadows (2001), reported by Ziegel (2001), is focused on managing Six Sigma initiatives. Leadership, however, connotes a vision and implies a plan that may transcend implementation and management.

This book has a couple of very significant differences compared with all of the other Six Sigma books. First, the authors assume that the readers have already absorbed "a basic knowledge of Six Sigma" (p. 2) from, for example, a book like that of Harry and Schroeder (2000), which was reported here by Alexander (2001). Second, it has been written for use by all participants in Six Sigma programs, from executive-level leadership to the Green Belts. This certainly makes the point that transparency should be a key feature in Six Sigma implementations. One of the book's primary strengths is the effort expended toward that goal.

The book begins with a provocative title for its first chapter, "So You Want to Do Six Sigma?" At BP, that perspective seems to begin and end with me. The authors, who evidently have learned about the "hype" that seems to be part of Six Sigma books, next proceed to report on "Why You Need This Book." There is also a nice overview of Six Sigma, including the roles of the various leaders. Chapter 2 examines "What Works and What Doesn't" through case studies from GE, W. R. Grace, Royal Chemicals, and Diversified Paper, representing two successes and two failures. Chapter 3, "How to Successfully Deploy Six Sigma," looks at why the four companies were or were not successful and then presents the keys to success. There is the inevitable roadmap.

The next three chapters deal with introducing Six Sigma into a business as a new process for bringing improved performance. Chapter 4, "Launching the Initiative," provides guidance on creating a deployment plan. This includes selecting the projects, the people to execute them, and the training that they will need. Chapter 5, "Managing the Effort," deals with project reviews, communications, reward, and recognition. Chapter 6, "Sustaining Momentum and Growth," advises on the difficult tasks of sustaining momentum following the initial Six Sigma deployment and then of growing the process.

Chapter 7, "The Way We Work," ties up some loose ends, including the integrating of Six Sigma with all the other management systems already in a business or with the other initiatives, such as ISO 9000, that are already part of an organization. Chapter 8, "Final Thoughts for Leaders," primarily gives the authors' view on the role of all the available Six Sigma tools. Books such as that by Breyfogle (2003), a vast collection of the statistical methods useful for Six Sigma projects (reported in Ziegel 2003), cause people to equate Six Sigma with statistical methods. The book ends with 31 questions and answers relative to a Six Sigma deployment. This book is a good and inexpensive guide and reference for anyone involved with a Six Sigma implementation.

REFERENCES

Alexander, M. (2001), Review of *Six Sigma*, by M. Harry and R. Schroeder, *Technometrics*, 43, 370.

Breyfogle, F. (2003), *Implementing Six Sigma* (2nd ed.), Hoboken, NJ: Wiley.

Breyfogle, F., Cupello, J., and Meadows, B. (2001), *Managing Six Sigma*, New York: Wiley.

Harry, M., and Schroeder, R. (2000), *Six Sigma*, New York: Currency/Doubleday.

Pande, P., Neuman, R., and Cavanagh, R. (2000), *The Six Sigma Way*, New York: McGraw-Hill.

Ziegel, E. (2001), Editor's Report on *Managing Six Sigma*, by F. Breyfogle, J. Cupello, and B. Meadows, *Technometrics*, 43, 382.

——— (2002), Editor's Report on *The Six Sigma Way*, by P. Pande, R. Neuman, and R. Cavanagh, *Technometrics*, 44, 299–300.

——— (2003), Editor's Report on *Implementing Six Sigma* (2nd ed.), by F. Breyfogle, *Technometrics*, 43, 372.

Forthcoming Reviews

Books listed here have been assigned for review in the past quarter. Publication of their reviews or reports generally would occur within the next four issues of the journal. Persons interested in reviewing specific books must notify the editor soon after the publication date for the book. Persons interested in being reviewers or reviewing specific books should contact the editor by electronic mail (*ziegeler@bp.com*).

Analyzing Categorical Data, by Jeffrey S. Simonoff, Springer-Verlag

Applied Bayesian Modelling, by Peter Congdon, Wiley

Bayesian Data Analysis (2nd ed.), by Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin, Chapman & Hall/CRC

Bayesian Inference: Parameter Estimation and Decisions, by H. L. Harney, Springer-Verlag

Data Analysis and Graphics Using R, by John Maindonald and John Braun, Cambridge University Press

Data Analysis Tools for DNA Microarrays, by Sorin Drăghici, Chapman & Hall/CRC

Data Matters: Conceptual Statistics for a Random World, by Nicholas Maxwell, Key College Publishing

The Design and Analysis of Computer Experiments, by Thomas J. Santner, Brian J. Williams, and William J. Notz, Springer-Verlag

Designing Experiments and Analyzing Data (2nd ed.), by Scott E. Maxwell and Harold D. Delaney, Lawrence Erlbaum Associates

A First Course in Stochastic Models, by Henk C. Tijms, Wiley

Handbook of Statistics 22: Statistics in Industry, edited by R. Khattree and C. R. Rao, Elsevier Science

Highly Structured Stochastic Systems, edited by Peter J. Green, Nils Lid Hjort, and Sylvia Richardson, Oxford University Press

Intro Stats, by Richard D. De Veaux and Paul F. Velleman, with contributions by David E. Bock, Addison-Wesley

Measures of Interobserver Agreement, by Mohamed M. Shoukri, Chapman & Hall/CRC

Modeling the Internet and the Web, by Pierre Baldi, Paolo Frasconi, and Padhraic Smyth, Wiley

Multiple Analyses in Clinical Trials, by Lemuel A. Moyé, Springer-Verlag

Partial Identification of Probability Distributions, Charles F. Manski, Springer-Verlag

Probability and Statistics for Computer Science, by James L. Johnson, Wiley

Random Number Generation and Monte Carlo Methods (2nd ed.), by James E. Gentle, Springer-Verlag

Resampling Methods for Dependent Data, by S. N. Lahiri, Springer-Verlag

Risk Analysis in Engineering and Economics, by Bilal M. Ayyub, Chapman & Hall/CRC

Sample Survey Theory, by Paul Knottnerus, Springer-Verlag

The Short Road to Great Presentations, by Peter Reimold and Cheryl Reimold, Wiley

Statistical Methods for Rates and Proportions (3rd ed.), by Joseph L. Fleiss, Bruce Levin, and Myunghee Cho Paik, Wiley.