

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.

Planning, Construction, and Statistical Analysis of Comparative Experiments

Francis G. Giesbrecht and Marcia L. Gumpertz *Jim Rutherford* 93

The Optimal Design of Blocked and Split-Plot Experiments

Peter Goos *Weiming Ke* 94

Multivariate Statistical Methods in Quality Management

Kai Yang and Jayant Trewn *Andrew M. Kuhn* 94

Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing

Daniel B. Rowe *Michael Conklin* 95

Measurement and Multivariate Analysis

S. Nishisato, Y. Baba, H. Bozdogan, and K. Kanefuji (editors) *Olcay Arslan* 95

Random Graphs for Statistical Pattern Recognition

David J. Marchette *William L. Seaver* 96

Statistical Curves and Parameters: Choosing an Appropriate Approach

Michael E. Tarter *Lianfen Qian* 96

Statistical Models

A. C. Davison *Lorrie L. Hoffman and Elayne Reiss* 97

Logit Models From Economics and Other Fields

J. S. Cramer *Lawrence E. Barker* 97

A First Course in Statistical Methods

R. Lyman Ott and Michael T. Longnecker *Hoon Kim* 98

Basic Statistics and Data Analysis

Larry J. Kitchens *Robert V. Brill* 99

Statistics for Research (3rd ed.)

Shirley Dowdy, Stanley Wearden, and Daniel Chilko *William A. Brenneman* 100

Diagnostic Checks in Time Series

Wai Kung Li *B. D. McCullough* 100

Ranked Set Sampling: Theory and Applications

Zehua Chen, Zhidong Bai, and Bimal K. Sinha *Chunming Zhang* 100

Highly Structured Stochastic Systems

Peter J. Green, Nils Lid Hjort, and Sylvia Richardson (editors) *Nicholas Rose* 101

Bayesian Artificial Intelligence

Kevin B. Korb and Ann E. Nicholson *Daniel Zelterman* 101

Applied Data Mining

Paolo Giudici *Thomas D. Sandry* 102

Reliability and Statistics in Geotechnical Engineering

Gregory B. Baecher and John T. Christian *Donald E. Myers* 103

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

Intelligent Data Analysis: An Introduction (2nd ed.)

Michael Berthold and David J. Hand (editors) 104

Mind on Statistics (2nd ed.)

Jessica M. Utts and Robert F. Heckard 104

Introduction to Spectral Analysis

Petre Stoica and Randolph Moses 104

Design for Six Sigma

C. M. Creveling, J. L. Slutsky, and D. Antis, Jr. 105

Quality Assurance in Analytical Chemistry

B. W. Wenclawiak, M. Koch, and E. Hadjicostas (editors) 105

Nature-Inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks

Riccardo Leardi (editor) 106

Quantitative Methods for Current Environmental Issues

Clive W. Anderson, Vic Barnett, Philip C. Chatwin,
and Abdel H. El-Shaarawi (editors) 106

Practical Statistics for Environmental and Biological Scientists

John Townend 106

Internet Companion for Statistics

Michael D. Larsen 106

Handbook of Statistics 23: Advances in Survival Analysis

N. Balakrishnan and C. R. Rao (editors) 107

Statistics for Epidemiology

Nicholas P. Jewell 107

Multivariate Methods in Epidemiology

Theodore R. Holford 107

Statistical Estimation of Epidemiological Risk

Kung-Jong Lui 107

Disease Mapping With WinBUGS and MLwiN

Andrew B. Lawson, William J. Browne,
and Carmen L. Vidal Rodeiro 108

Bayesian Approaches to Clinical Trials and Health Care Evaluation

David J. Spiegelhalter, Keith R. Abrams, and Jonathan P. Myles 108

Design and Analysis of Clinical Trials (2nd ed.)

Shein-Chung Chow and Jen-Pei Liu 108

Design and Analysis of DNA Microarray Investigations

Richard M. Simon, Edward L. Korn, Lisa M. McShane,
Michael D. Radmacher, George W. Wright, and Yingdong Zhao 108

Data Analysis Tools for DNA Microarrays

Sorin Drăghici 109

Analyzing Microarray Gene Expression Data

Geoffrey J. McLachlan, Kim-Anh Do, and Christophe Ambroise 109

Forthcoming Reviews

The range and detail of the topics covered is best conveyed by a listing of the chapter titles:

1. Introduction
2. Completely Randomized Design
3. Linear Models for Designed Experiments
4. Testing Hypotheses and Determining Sample Size
5. Methods of Reducing Unexplained Variation
6. Latin Squares
7. Split-Plot and Related Designs
8. Incomplete Block Designs
9. Repeated Treatment Designs
10. Factorial Experiments: The 2^N System
11. Factorial Experiments: The 3^N System
12. Analysis of Experiments Without Designed Error Terms
13. Confounding Effects With Blocks
14. Fractional Factorial Experiments
15. Response Surface Designs
16. Plackett–Burman Hadamard Plans
17. General p^N and Nonstandard Factorials
18. Plans for Which Run Order Is Important
19. Sequences of Fractions of Factorials
20. Factorial Experiments With Quantitative Factors: Blocking and Fractions
21. Supersaturated Plans
22. Multistage Experiments
23. Orthogonal Arrays and Related Structures
24. Factorial Plans Derived via Orthogonal Arrays
25. Experiments on the Computer.

This is not a cookbook. For each topic, the authors do a thorough job covering concepts and background and giving practical advice and examples. Exercises and examples are interspersed throughout the narrative. Numerous review exercises are included at the end of most chapters. Catalogues and listings of specific designs appear in appendixes to some chapters. The associated website has data and SAS code for examples and exercises.

As an example of the authors' treatment, in a very coherent and understandable flow, in Chapter 4, in a section titled "Power and Sample Size" (pp. 58–68) is followed by "Sample Size for Binomial Proportions" (pp. 68–72), "Confidence Interval Width and Sample Size" (pp. 72–73), and "Alternative Analysis: Selecting and Screening" (pp. 74–78). Several examples and exercises are discussed and presented in detail, including calculations and four snippets of SAS code. At the end of the chapter, six review exercises are given relating to these topics.

Another good example of the completeness of the authors' treatment is in the introduction for the chapter on incomplete block designs, where a bit more than three pages is devoted to interblock information. Nearly a page of narrative gives the background, meaning, and implications. Traditional hand calculations are then described, followed by matrix notation and various techniques of estimation. Interblock information is revisited at appropriate points throughout the chapter with illustrations of SAS code and use of the traditional Yates method.

The expertise imparted by this book is extremely practical, but time-consuming to read. I was unable to skim this book for the review. I found myself wanting to read every word for the wisdom contained and agreeing with advice given.

I found only a couple apparent typographical errors (p. 128, second paragraph, third sentence, "square x MS row" should be "MS square x row," and p. 378, fourth paragraph, ninth sentence, "actors" should be "factors"). The associated website gives three slightly more substantial errata.

I will instruct statisticians reporting to me to get a copy of this book, and will keep the review copy readily available on my shelf when I have finished rereading a few of the chapters. I recommend that consulting statisticians or teachers who use or teach experimental design get a copy for their shelves. Sections of the book could be read or reread to advantage when working in a less familiar area. A course in experimental design could definitely benefit from inclusion of appropriate parts of this well-written book.

Jim RUTHERFORD
ChevronTexaco

Planning, Construction, and Statistical Analysis of Comparative Experiments, by Francis G. GIESBRECHT and Marcia L. GUMPERTZ, Hoboken, NJ: Wiley, 2004, ISBN 0-471-21395-0, xiv + 693 pp., \$115.00.

The authors present this book as a textbook, handbook, and reference book based on their experiences from teaching graduate-level statistics and consulting. They assume a general background in statistics. Several subsets of the book are suggested as course outlines depending on the intention of the course and background of the students.

The Optimal Design of Blocked and Split-Plot Experiments, by Peter GOOS, Secaucus, NJ: Springer-Verlag, 2002, ISBN 0-387-95515-1, xiii + 244 pp., \$67.95.

This book presents a high-level introduction to optimal design approaches and their applications to the design of blocked and split-plot experiments commonly used in practice. The book provides a theoretical background and a thorough review of recent work in the area of blocked and split-plot experiments, along with some interesting results. Examples illustrate the optimal designs for the blocked and split-plot experiments in various experimental situations. Computer programs for finding the best possible designs are also provided. This book can be used by applied statisticians in physics, chemistry, engineering, and many other fields. It can also be used as a graduate-level textbook as a supplement to a text such as the one by Wu and Hamada (2000) for an experimental design course. The book is well organized and will be easy to read for *Technometrics* readers.

The book comprises 10 chapters:

1. Introduction
2. Advanced Topics in Optimal Design
3. Compound Symmetric Error Structure
4. Optimal Designs in the Presence of Random Block Effects
5. Optimal Designs for Quadratic Regression on One Variable and Blocks of Size Two
6. Constrained Split-Plot Designs
7. Optimal Split-Plot Designs in the Presence of Hard-to-Change Factors
8. Optimal Split-Plot Designs
9. Two-Level Factorial and Fractional Factorial Designs
10. Summary and Future Research.

Chapter 1 provides a broad overview of the experimental design literature. Special attention is given to the D-optimality criterion used throughout the book. The basics of the optimal design approach, the standard response surface designs, and the categorical designs are also introduced. Chapter 2 is mainly concerned with the design of experiments with nonhomogeneous error variance and correlated observations. The topic of blocking response surface designs receives elaborate attention as well. It is shown that orthogonal blocking is an optimal blocking strategy when the block effects are assumed to be fixed. Chapter 3 describes the design problems in detail and introduces an appropriate statistical model. It shows that the variance-covariance structure of the statistical model is compound symmetric. The chapter also contains sections on the analysis of blocked and split-plot designs and on the equivalence of ordinary least squares (OLS) and generalized least squares (GLS). Finally, it shows that the asymptotic variance-covariance matrix is a good approximation to the finite-sample variance-covariance matrix. Chapter 4 studies D-optimal designs for randomly blocked experiments. The first part of the chapter is devoted to optimal designs that do not depend on the degree of correlation. The second part presents the general case, in which the optimal designs depend on the degree of correlation. A design construction algorithm and an adjustment algorithm to improve the resulting designs are developed. It is also shown that orthogonal blocking is an optimal blocking strategy when the block effects are assumed to be random and the block sizes are homogeneous. Chapter 5 derives exact D-optimal designs for an optometry experiment. This chapter provides useful insights in the relationship between discrete designs and continuous designs and in the relationship between random and fixed blocks. Chapter 6 computes D-optimal split-plot designs for experiments with a prespecified number of whole plots and prespecified whole plot sizes. This chapter shows that an optimal design strategy is to arrange the observations such that the resulting design is crossed. It also shows that two-level factorial and fractional factorial designs are D-optimal in the class of crossed split-plot designs. Chapter 7 derives D-optimal designs under the assumption that the levels of the hard-to-change or whole-plot factors are changed as little as possible. As a result, the designs computed in this chapter have only one whole plot per whole plot factor level. It is shown that split-plot designs sometimes provide better estimates than completely randomized designs. Chapter 8 shows that split-plot experiments become more efficient when their number of whole plots is increased. The resulting split-plot designs are far more efficient than completely randomized designs. As a result, split-plot designs are not only easier to carry out, but they are also statistically more efficient than a completely randomized design. Chapter 9 shows how two-level factorial and fractional factorial designs can be used in blocked and split-plot experiments. It also reviews the recent literature

on this topic. For that purpose, the concept of minimum aberration is introduced. A summary of the book is given in Chapter 10, provides a summary and suggests some directions for future research.

In addition to the theoretical analyses and practical examples, the author also provides computer programs for finding the best possible designs for blocked experiments and split-plot experiments. The programs are written in FORTRAN 77 language, which is easy to use and powerful, and can be found at <http://www.econ.kuleuven.ac.be/public/NDBAC47/springer.htm>.

Overall, *The Optimal Design of Blocked and Split-Plot Experiments* is a good overview of the techniques available in the optimal design of blocked and split-plot experiments, including the author's own great research in this field. The optimal design approach advocated in this book will help practitioners of statistics in setting up tailor-made experiments. It is also a good reference book for researchers and students in applied statistics. The interested readers may refer to the papers by Bingham and Sitter (2003), Goos and Vanderbroek (2003), and McLeod and Brewster (2004) for further understanding of this topic.

Weiming KE
University of Memphis

REFERENCES

- Bingham, D., and Sitter, R. R. (2003), "Fractional Factorial Split-Plot Designs for Robust Parameter Experiments," *Technometrics*, 45, 80–89.
- Goos, P., and Vanderbroek, M. (2003), "D-Optimal Split-Plot Designs With Given Numbers and Sizes of Whole Plots," *Technometrics*, 45, 235–245.
- McLeod, R. G., and Brewster, J. F. (2004), "The Design of Blocked Fractional Factorial Split-Plot Experiments," *Technometrics*, 46, 135–146.
- Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.

Multivariate Statistical Methods in Quality Management, by Kai YANG and Jayant TREWN, New York: McGraw-Hill, 2004, ISBN 0-07-143208-6, xvi + 299 pp., \$150.00.

This book presents multivariate methods for use by quality professionals. The Preface states the authors' intention is to discuss theoretical aspects and provide examples of applying multivariate statistics to solve real-world problems.

The intended audience should have a solid understanding of basic statistical concepts (e.g., mean, variance, distributions) and techniques such as regression and ANOVA. The writing style is easy to read and not overly mathematical, given the intended audience. The authors note in the Preface that multivariate methods have not been used in quality engineering, because of the perception that they are complex and difficult to understand. To help alleviate this problem, the mathematical complexity is kept at a minimum here, and a quality engineer who can understand basic linear algebra should not have problems.

After the introductory chapter come five chapters on classical multivariate analysis: multivariate distributions, MANOVA, principal components analysis, discriminate analysis, and clustering. There are also chapters on displaying multivariate data, the Mahalanobis-Taguchi methods, and multivariate control charts. There is a section of how these methods relate to Six Sigma programs. Aside from the small examples, the data used are not provided. Minitab is mentioned for data analysis, but no computing specifics are provided.

In the interest of specificity, although the book's title contains the phrase "quality management," the authors do not focus in any one particular area of quality. They list areas that are ripe for multivariate analysis in industry, from conceptual design of a product, through research and development to manufacturing. I am not sure that a quality organization in a company would be responsible for the data analysis in many of these areas (e.g., product development). Instead of *Multivariate Statistical Methods in Quality Management*, perhaps *Multivariate Statistical Methods in Industry* would have been a more accurate title.

The authors correctly state that this book is distinguishable from other multivariate analysis texts because it has an "orientation toward readers in business and industry and toward quality professionals." They also note that "the selection of examples and case studies are all based on the needs of readers in business and industry." Although I agree with the first statement, I do not think that the authors can strongly defend the second statement. Although the book

contains a few good case studies directly related to quality engineering, many of the examples have nothing to do with industry (e.g., measurements on the shapes and sizes of painted turtle carapaces, weight of middle-aged men in fitness clubs, diagnosis of liver disease). These examples are used to illustrate the methodologies as they are being developed and make up the bulk of the results given in the book. On the bright side, the case studies are engaging. The data analysis objectives of the case studies are clearly described, and the presentations of results show how multivariate analysis can address the objectives.

One could argue that the goal of a book describing application of multivariate methods to specific types of problems should aim to provide reads with a “feel” for how these methods work for *their* problems. For example, I have spent years encouraging microbiologists to use “industrial statistics,” such as response surface methods or control charts. I have found that once they see in-depth case studies within their area of science that address the nuances of their data, they are almost completely on board. In this sense, the authors moderately succeed using their case studies. However, I would suggest that they make the data available to the public, because many people learn by reproducing examples that they study.

To enable quality organizations to better use multivariate methods, this text should be supplemented with others. For example, the chapter on discrimination hits the major points for classical multivariate methods; however, a vast array of tools for discrimination currently exist, and the classical methods may not fit the problem well and may be far from optimal. For instance, many on-line measurement systems produce large amounts of data on each sample, such as optical coordinate measurement machines, so there may be more variables than observations. Classical methods cannot handle this situation well (e.g., LDA with more variables than observations). I suggest the text by Hastie, Tibshirani, and Friedman (2001) to supplement this book if more expertise is required for discrimination, clustering, and principal components analysis, and the text by Mason and Young (2001) for a more in-depth resource for multivariate control charting.

All in all, the authors meet their intended goals somewhat. This text provides the motivation to use multivariate analysis, but could do a better job providing the means.

Andrew M. KUHN
Becton Dickinson

Michael CONKLIN
GfK Custom Research

REFERENCES

- Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning*, New York: Springer-Verlag.
Mason, R., and Young, J. (2001), *Multivariate Statistical Process Control With Industrial Application*, Philadelphia: ASA-SIAM.

Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing, by Daniel B. ROWE, Boca Raton, FL: Chapman & Hall/CRC, 2003, ISBN 1-58488-318-9, x + 329 pp., \$89.95.

This book is a thorough exposition of Bayesian modeling techniques. The motivating example used throughout is the source separation model. The author effectively uses the example of a cocktail party with a variety of conversations around the room captured by a set of microphones. The model is designed to separate the different sources (the party guest’s conversations) as they are heard, mixed, from the various microphones. This example allows the author to step through a variety of basic Bayesian approaches to modeling, from linear regression to multivariate regression to factor analysis and, finally, to the source separation model.

After setting up the basic problem of the source separation problem in Chapter 1, the author starts with the basics. Chapter 2 is a basic description of a number of standard univariate and multivariate distributions. This chapter reads as if it was computer generated—the exact same format is used to describe each distribution. Nonetheless, it is thorough and provides a handy reference for determining the key characteristics of many common distributions.

Chapter 3 gives a short but very understandable tutorial on the basics of Bayesian statistical thought. The author effectively explains Bayes theorem on the basis of two events and then expands the explanation to scalar, vector, and matrix variables. Chapter 4 covers various types of prior distributions, including noninformative, conjugate, and generalized conjugate for each of the types

of distributions covered in Chapter 2. Each of these prior distributions is dependent on some hyperparameters, and assessment of the hyperparameters is discussed in Chapter 5. Here the description of how to think about the hyperparameters in terms of prior expert knowledge is most helpful.

Chapter 6 goes into great detail on the use of Gibbs sampling and iterated conditional modes for estimating marginal posterior means and joint maximum posterior modal estimates. This chapter is very helpful in its step-by-step explanation of how these algorithms can be applied to estimate the quantities of interest in a Bayesian analysis.

The next three chapters build the foundation for the Bayesian source separation model that is the book’s foundation. Chapter 7 gives a refresher on the classical multivariate regression model, focusing on the distributional assumptions that are present. Chapter 8 extends the model to a full Bayesian approach and gives detailed instructions on how to calculate results. Chapter 9 covers the Bayesian approach to factor analysis, again with detailed instructions on how to set up the Gibbs sampling algorithm and develop results.

Chapter 10 pulls all of the previous work together into a detailed description of procedures to fit a Bayesian source separation model. One of the really nice features of both Chapters 9 and 10 is the presence of a full dataset and example results, allowing the reader to check his or her understanding of the procedures and algorithms presented. Chapter 11 extends the source separation model to include both observable and unobservable sources.

The last few chapters of the book deal with a case study of data from functional magnetic resonance imaging (fMRI). This allows the author to give a specific application of the techniques that have been developed so far in the book. Further extensions are explored to allow for dynamic coefficient models and dealing with correlated observations.

Overall, the book is well written and gives a detailed step-by-step approach to some widely applicable model types. Previously, when I needed to build a specific model, I could (if the model was very simple) determine a good approach using the books by Gelman et al. (1995) and Congdon (2001). This book helps me understand how to build some complex models using a Bayesian approach with a much better understanding of what effect my decisions will have on the final model results.

REFERENCES

- Congdon, P. (2001), *Bayesian Statistical Modeling*, West Sussex, U.K.: Wiley.
Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Measurement and Multivariate Analysis, edited by S. NISHISATO, Y. BABA, H. BOZDOGAN, and K. KANEFUJI, Tokyo: Springer-Verlag, 2002, ISBN 4-431-70338-1, xvi + 332 pp., \$109.00.

This book contains a collection of papers from an International Conference on Measurement and Multivariate Analysis held May 12–14, 2000, in Banff, Alberta, Canada. Topics cover a wide range of theoretical work in statistics, related disciplines, and applications. All of the papers were reviewed before their inclusion in the proceedings. The papers are organized into seven sections. The title of each section and the number of papers included in each are as follows:

- Keynote Papers (2)
- Introduction (1)
- Scaling (8)
- Structure Analysis (8)
- Statistical Inference (7)
- Algorithms (5)
- Data Analysis (4).

The first keynote paper, “Categories and Quantities” by J. C. Gower, deals with the rank of a matrix of categorical variables. Gower states in this summary that this rank may be used as a basis of multivariate methods for approximating categorical data similar to how the rank of a quantitative data matrix is used for standard multivariate methods. Gower outlines several problems that depend on the notation of categorical rank. The second keynote paper in is “Questionnaire Construction, Data Collection and Data Analysis: An Approach by the

Idea of Data Science.” Predictably, given its title, this paper derives attention to the importance of the data design, data collection, and data quality evaluation to data analysis. The paper starts with the fundamental idea of data science, then discusses construction of the questionnaire in cross-cultural social surveys. The paper ends with construction of the questionnaire in the survey of nonprofit corporations and evaluation of the data.

The introduction section consists of “Measurement and Multivariate Analysis,” by S. Nishisato. This paper serves the purpose of introducing readers to the main theme of the conference. It emphasizes the desirable properties of measurement to make the outcomes of multivariate analysis meaningful. The paper starts with the univariate case, covering classification of the measurements, metric axioms, and the Minkowski power metric. The author then turns attention to the multivariate case, starting with the Young–Householder theorem and considering rectangular coordinates, multidimensional space and principal hyper-spaces. The paper ends with a discussion of two important topics, measurements and multivariate analysis, the main topics of the conference, and finally a note on the importance of the measurement and multivariate analysis.

The topics in the “Scaling” section include the clustering technique based on dual scaling, homogeneity and smoothness analysis for quantifying a longitudinal categorical variable, correspondence analysis, triadic distance models, generalized modification of Scheffe’s paired comparisons, analysis of dependence in large contingency tables, multidimensional scaling, and complex space models for the analysis of asymmetry.

Topics covered in the “Structural Analysis” section include extended redundancy analysis, redundancy index in canonical correlation analysis, regularized kernel discriminant analysis, dimension reduction in hierarchical linear models, time-dependent principal component analysis, and nonlinear generalized canonical correlation analysis by neural network models.

The “Statistical Inference” section is devoted to the papers that propose new statistical methods. This section starts with a paper on the empirical Bayes method for the simultaneous estimation of mean vectors under the mean conjugate prior. The next paper proposes an improved shrinkage estimator for interclass correlation coefficients when independent samples are drawn from multivariate normal distributions. The other papers include Bayesian analysis of structural equation modeling, estimation of the fifth percentile using a subset of order statistics, sensitivity analysis in semiparametric regression models, and cohort analysis of data obtained using multiple choice questions.

The next section, “Algorithms,” collects papers on the algorithms. The section begins with a paper that gives a new algorithm of the a priori type for mining all generalized subsequences of user navigation paths. The next paper proposes a new criterion called structured ratio for measuring clustering results. In a paper written by Bearse and Bozdogan, the authors develop a computationally feasible intelligent data mining and knowledge discovery technique for selecting the best subset of predictors in multivariate regression models. The section ends with two papers on statistical software for sensitivity analysis in multivariate methods and the use of prolog, a logic programming for symbolic manipulation, as a model search engine for data analysis.

The final section of the proceedings is devoted to papers related to data analysis. The section starts with a paper on gender differences in mathematics achievement. Other topics in this section include analyzing foreign financial statements, a dual scaling approach to the international ratio analysis, a three-step approach to factor analysis on data of multiple testlets, and a technique for setting standards and maintaining them over time.

Olcay ARSLAN
Çukurova University, Turkey

Random Graphs for Statistical Pattern Recognition, by David J. MARCHETTE, Hoboken, NJ: Wiley, 2004, ISBN 0-471-22176-0, xiii + 237 pp., \$79.95.

The author rightly notes early in the Preface that there is much literature solely covering random graphs or statistical pattern recognition, but this book’s uniqueness is the intersection of these two topics. This is both a blessing and a curse! The blessing is the brevity and broad strokes taken in the chapters: Preliminaries, Computational Geometry, Neighborhood Graphs, Class Cover Catch Digraphs, Cluster Catch Digraphs, and Computational Methods. The blessing also entails insight into certain interrelationships without a lot of theory and reading. The curse is the brevity and the heavy reference to other literature, where the reader must self-educate himself from this outside literature

to fill in some knowledge holes. However, it needs to be said upfront that the author had a difficult task to accomplish with the number of pages used and should be commended for the efforts. With some background in random graphs but much more in pattern recognition, I personally found the book enlightening and a good reference source.

The best chapters were the first three (Preliminaries, Computational Geometry, and Neighborhood Graphs), which compose about 60% of the book. Chapter 1 quickly lays out concepts and definitions of graphs, digraphs, random graphs, statistical pattern recognition (inclusive of classification, dimensionality, and clustering), and statistical issues. Chapter 2 spells out Voroni cells, Delaunay triangularization, alpha hulls, minimum spanning trees, and the interrelationship between these various graph methods. Chapter 3 covers nearest-neighborhood, k -nearest-neighborhood, relative neighborhood, Gabriel, sphere-of-influence, and sphere-of-attraction graphs. There are great applications in this chapter, including detecting fracture lines in porous materials and detecting outliers.

Chapter 4 delineates the properties of class cover catch digraphs (CCCD) along with applications to classification, dimension reduction, and association analysis. However, one may have to look to other resources to fully understand the content in this chapter. Chapter 5, on cluster catch digraphs, is short and sweet with basic definitions, dominating sets, connected components, and variable metric clustering. Chapter 6 covers the computational methods and difficulties in visualizing data structures. The author provides caveats in this chapter as to the computational feasibility related to the data size and as to whether or not there is a data stream.

The author contends that the book is written at the level of an advanced undergraduate or first-year graduate course, but I would contend that it is at the second-year graduate student level. It is an excellent resource book that would be a valuable addition to the library of someone who has an interest in random graphs, pattern recognition, or especially the intersection of the two topics.

William L. SEAVER
University of Tennessee

Statistical Curves and Parameters: Choosing an Appropriate Approach, by Michael E. TARTER, Natick, MA: A. K. Peters, 2000, ISBN 1-56881-105-5, xiii + 386 pp., \$64.00.

Although the topics it discusses are elementary, this is in no way an elementary statistics book. It can serve as a reference book for those who are unfamiliar with the recent statistics literature but who have at least a one-semester background in calculus. The book covers historical stories about the development of statistical methods covered.

The book is more conceptual than mathematical. It comprises 20 chapters and 3 appendixes. The focus is on model parameters, from parametric models to nonparametric adjustment. Covered are parameter estimation and statistical inferences for parameters, including ones from distribution, regression, and mixture models. The chapter titles are as follows:

1. Introduction
2. Model and Distribution Terminology
3. Variability and Related Curve Properties
4. Moments and Curve Uncertainty
5. Goodness of Fit
6. Variates, Variables, and Regression
7. Mixing Parameters and Data-Generation Models
8. The Association Parameter ρ
9. Regression and Association Parameters
10. Parameters, Confounding, and Least Squares
11. Nonparametric Adjustment
12. Continuous Variate Adjustment
13. Procedural Road Maps
14. Model-Based and Generalized Representation
15. Parameters, Transformations, and Quantiles
16. Noncentrality Parameters and Degrees of Freedom
17. Parameter-Based Estimation
18. Inference and Composite Variates

19. Parameters and Test Statistics
20. Curve Truncation and the Curve $e(x)$.

Lianfen QIAN
Florida Atlantic University

Statistical Models, by A. C. DAVISON, Cambridge, U.K.:
Cambridge University Press, 2003, ISBN 0-521-77339-3,
x + 726 pp., \$70.00.

We know of no other one book like this one. *Statistical Models* has the flavor (or we should spell that “flavour”) of Ross’s *Probability Models* (2000). Like Ross, Davison provides sound statistical tools to tackle applications in a concise, succinct tutelage, yet without abandoning mathematical sophistication or detail. But Davison’s topics are wide-ranging, nearing the level of exhaustive, covering usual subjects, such as Markov chains and regression methods, as well as more arcane topics, like graphical representations of probability expert systems. This, of course, partially explains the book’s long length.

We hesitate to call this book a text. The author does so in his Preface, suggesting its use for a course up through a masters degree in statistics, where the students have had linear algebra and a theory of statistics and probability course. But senior undergraduates without the entire sequence of mathematical statistics will struggle with the latter sections of most chapters. Davison might disagree. He insists that readers seeking a more thorough mathematical treatment look elsewhere, but we believe this to be a self-effacing statement. Rarely does he use the words “beyond the scope of this book.” We see on occasion a declaration such as (during a discussion of the binomial distribution in Chap. 3) “in order not to obscure the main points, the discussion above has been deliberately oversimplified” (p. 56). We suspect that little would look oversimplified to most students.

A classroom text is expected to contain problems for students. Although exercises are included at the end of each section, most of which are of the show-or-prove variety, we would have liked to see a short set of answers at the back of the book. Solutions would lend more credence to the claim that the book can be used as a text. Additionally, the “practicals” listed in Appendix A, which were to include data and S programs and be available on the internet were not yet complete (e-mail correspondence with the author on September 16, 2004) despite the publishing date of 2003. We are not surprised that these exercises were left as an afterthought, because the book is not a “cookbook” presentation enabling matching of real-world problems to examples therein. The examples in the book do illustrate techniques applied to real-world datasets (such as the Maize data, i.e., the Darwin plant heights) and are reworked with different modeling approaches, thus providing a feeling of continuity to the subject matter. As a text, this book would suit a class of exactly what the title implies—a statistical modeling course, one that would need to be created around the book. Instead, the Preface should have prepared us for a wonderfully rich treatise of cleverly attended-to examples and charming asides. Many of these interjections involve biographical sketches of famous mathematicians and statisticians—although Wald (see Wolfowitz 1952) was overlooked. And if the Preface had stated that seemingly every statistical technique available to the modern quantitative researcher would be considered, then we would have fully concurred that the author had met his goal.

Some level of frustration may surface for the reader early in the book, if he or she left with the feeling that the choice of formulation was achieved through, perhaps, magic. An expansion on the following sentence from Chapter 4 (p. 150): “Model formulation involves judgement, experience, trial and error” inserted as an introductory section would help establish precepts underlying the role of modeling. Chapter 1 presents ample enticing data situations encountered, but a novice to the concept of modeling may need to be more intrepid than intrigued. Chapter 2 broadly discusses items usually seen in the theoretical statistics sequence, including order statistics, moments, and cumulants. Chapter 3 covers topics pertaining to confidence intervals, including pivots and simulations, which are also reexamined later in Chapter 7. Likelihood estimation and model selection are visited in Chapter 4. Davison’s analytical journey includes detailed and important discussions of building block statistical issues, such as describing the mechanics behind the normal distribution (Sec. 3.3) and explaining the workings of likelihood ratio statistics (Sec. 4.5) before moving into the core. We would describe the first four chapters (titled Introduction, Variation, Uncertainty, and Likelihood) as “mathematical statistics revisited,” yet the exploration sometimes features a different perspective on

the topic that is more suited to a book of this style rather than to a plain mathematical statistics text. The book then launches into the meat of the material by describing classes of models in Chapters 5, 6, 8, and 9, including survival models, Markov processes, the general linear model and all of its variants (such as mixed models, with complications such as missing data), and then in Chapter 10 illustrates applications of nonlinear models. Chapter 11 covers the use of Bayesian statistics with accompanying computing techniques (e.g., Gibbs sampler). Altogether these chapters give the reader good grounding in the process of modeling. A synopsis of Chapter 12 could read as “warnings about relying on the old standby likelihood estimation.” A map with plausibly linkable chapters for a feasible subset coverage of the book is offered in Chapter 1. This handy map of chapter dependencies could aid instructors in selecting alternative ways to teach from the book that may be more suitable for their particular classes.

We love owning this book. It gets placed on our shelf among our favorite reference books like the old classic for engineers by Kreyszig (1972), which, incidentally, reviewers seem to either hate or love. We actually learned a lot and deepened our understanding of many topics (e.g., Fisher information, reversible chains, Slutsky’s lemma) while reading Davison’s explanations. Certainly, using a hypercritical eye, one could say that the bibliographies at the chapter ends are a bit skimpy. In our areas of major interest we could reveal holes in the references; for example, for censored data, recent work by Ren (2003) and Zhang, Liu, and Wu (2003) are important citations that are omitted. And a few topics must have been reserved for the second edition, namely circadian rhythm analysis (Qiu 2002). Even so, if asked to summarize *Statistical Models* in a single word, “complete” would serve as the only plausible answer.

Lorrie L. HOFFMAN
Armstrong State University
Elayne REISS
University of Central Florida

REFERENCES

- Kreyszig, E. (1972), *Advanced Engineering Mathematics*, New York: Wiley.
 Qiu, P. (2002), “Fitting a Semi-Parametric Model Based on Two Sources of Information,” *Australian & New Zealand Journal of Statistics*, 44, 87–98.
 Ren, J.-J. (2003), “Goodness-of-Fit Tests With Interval-Censored Data,” *Scandinavian Journal of Statistics*, 30, 211–226.
 Ross, S. M. (2000), *Introduction to Probability Models*, San Diego, CA: Academic Press.
 Wolfowitz, J. (1952), “Abraham Wald, 1902–1950,” *The Annals of Mathematical Statistics*, 23, 1–13.
 Zhang, Y., Liu, W., and Wu, H. (2003), “A Simple Nonparametric Two-Sample Test for the Distribution Function of Event Time With Interval-Censored Data,” *Journal of Nonparametric Statistics*, 15, 643–652.

Logit Models From Economics and Other Fields, by
J. S. CRAMER, Cambridge, U.K.: Cambridge University
Press, 2003, ISBN 0-521-81588-6, x + 173 pp., \$55.00.

Sometimes one must model a categorical variable in terms of independent variables. In the simplest case, this categorical dependent variable takes only the values 0 and 1. Cramer presents more general models; for simplicity, this review considers only the dichotomous case. One model for a dichotomous dependent variable is $P(Y_i = 1) = F(\sum_{j=0}^k \beta_j x_{ij})$, where, for $i = 1, \dots, n$ and $j = 0, \dots, k$, Y_i is the i th realization of the dependent variable, x_{ij} is the i th value of independent variable j , the β_j ’s are unknown parameter (x_{i0} is usually 1), and F is a continuous known cumulative distribution function (cdf). In practice, F is usually the logistic (logit model) or standard normal cdf (probit model).

The logit and probit models are central to respectively epidemiology and toxicology. Both models hold a place of honor in economics. Although many *Technometrics* readers might be more familiar with the latter area than the first two areas, logit and probit models are reasonable candidates whenever dichotomous dependent variables arise; these models belong in every statisticians’ toolbox.

This rather slim text (173 pages, including index and references) provides an overview of logit and probit models, oriented more toward application than

theory. However, this is more than a cookbook. For example, instead of giving a goodness-of-fit test and being done with the matter, Cramer presents three goodness-of-fit tests and compares, in English rather than equations, their interpretation (Table 4.1). Translation of statistics into nontechnical terms is often essential for communication with managers and decision makers. Cramer's text can help with translation.

Cramer also serves the mathematically minded reader. For example, he shows how omitting an independent variable is a more serious error in a logit model than in an ordinary least squares (OLS) model. Cramer also considers the consequences of dependent variable misclassification (1 is recorded as 0, or vice versa). Such misclassification (which might occur more frequently than many logit model users care to admit) can cause serious bias. The definition of an outlier is much less clear-cut in logit/probit models than in OLS models. Cramer presents a good overview of outliers in the logit/probit setting. Further, he discusses state-dependent sampling (i.e., sampling rates depending on value of dependent variable). Given rare outcomes, state-dependent sampling can be the only alternative to very large samples. For a logit model, state-dependent sampling and random sampling differ only in the intercept term. This property makes possible logit models of case-control studies, a workhorse of medical statistics.

Cramer presents a thorough discussion of a real-world dataset (relating auto ownership to demographics in the Netherlands), available for all on the internet. This dataset displays outliers and a lack of theory for determining what dependent variables to include in the model; these are common features of real-world data. The appearance of such messy real-world data, as opposed to "too well-behaved to be true" textbook data, is a strength of the book.

Finally, Cramer presents some fascinating nontechnical aspects of logit/probit models. From the book's beginning, he discusses how multiple disciplines developed such models in parallel, each with its own terminology and way of looking at the data. This "multiple origin" has led to many discoveries of the same methods. Cramer devotes Chapter 9 to a history of the logit/probit model. It is truly fascinating to see how logit and probit models have competed for supremacy over the years.

As with any book, there are a few things I would like to have seen done differently. The absence of problems limit the book's use as a classroom text. Although the consequences of omitting a variable and of outliers (concepts of import for model building) are discussed, Cramer gives no guidance on how one *should* build models. (Economists often advise reliance on theory; the auto ownership data show that even in a discipline as theory-driven as economics, one can not always rely on theoretical direction.)

Despite these minor flaws, *Logit Models From Economics and Other Fields* provides the applied statistician with a solid grounding in logit and probit models. Those who want to quickly teach themselves about these models could do far worse than to begin with this book.

Lawrence E. BARKER
Centers for Disease Control and Prevention

A First Course in Statistical Methods, by R. Lyman OTT and Michael T. LONGNECKER, Belmont, CA: Brooks/Cole, 2004, ISBN 0-534-40806-0, xvii + 741 pp. + CD, \$98.95.

This new edition of *A First Course in Statistical Methods* is tailored to meet the needs of a one-term course in statistical methods for advanced undergraduate and graduate students from a variety of disciplines. It is based on selected topics and material from the book *An Introduction to Statistical Methods and Data Analysis* (5th ed.), by Ott and Longnecker (2001). The authors state that the mathematical requirement for this book is high school algebra.

The authors approach the study of statistics by considering the four-step process in learning from data: (1) defining the problem, (2) collecting data, (3) summarizing data, and (4) analyzing data, interpreting the analyses, and communicating results. Following these four steps in learning from data, the study of statistics closely parallels the scientific method, which relates to the "principles and procedures for the scientific pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses" (p. 2). This book is divided into sections corresponding to the four-step process in learning from data. The relationship among these steps and the book's chap-

ters is shown in the following table.

The four-step process	Chapters
1. Defining the problem	1. Statistics and the Scientific Method
2. Collecting data	2. Collecting Data Using Surveys and Scientific Studies
3. Summarizing data	3. Summarizing Data 4. Probability and Probability Distributions
4. Analyzing data, interpreting the analyses, and communicating results	5. Inferences about Population Central Values 6. Inferences Comparing Two Population Central Values 7. Inferences About Population Variances 8. The Completely Randomized Design 9. More Complicated Experimental Designs 10. Categorical Data 11. Linear Regression and Correlation 12. Multiple Regression 13. Communicating and Documenting the Results of Analysis

Chapter 1 begins with a definition of statistics, based on *Merriam-Webster's Collegiate Dictionary*. The authors relate this definition to their four-step process. They also discuss some current applications of statistics using a variety of real-world examples.

Chapter 2 discusses different survey designs and experimental designs for scientific studies. Intelligent data gathering involves specifying the objectives of the data-gathering exercise, identifying the variables of interest, and choosing an appropriate design for the survey or scientific studies.

Chapter 3 focuses on describing a set of measurements using graphical and numerical techniques. The authors extend the concept of data description to summarize the relations between two variables.

Chapter 4 provides a brief introduction to probability, probability distributions, and sampling distributions. The classical, relative frequency, and subjective interpretations of the probability of an outcome are given. However, only binomial and normal distributions are introduced as the probability distributions. The end of chapter presents Minitab codes for generating random numbers, calculating binomial and normal probabilities, and generating sampling distribution of the sample mean.

Chapter 5 deals with introductory statistical inferences for constructing confidence intervals, performing the testing of hypotheses, and choosing the sample sizes for estimating the population mean. The authors also discuss the inferential procedure about the median, which is more appropriate when the population distribution is highly skewed or very heavy tailed. Chapter 6 extends the ideas of Chapter 5 based on two population/sample situations.

Chapter 7 describes the procedures for making inferences concerning population variances. Estimation and statistical tests concerning a population variance are done with the chi-squared distribution, whereas those concerning the ratio of two population variances are discussed with the *F* distribution. Finally, the authors introduce the Hartley or Levine test statistic for tests concerning differences in more than two population variances.

Chapter 8 introduces a method for analyzing data that extends the results of Chapter 6 to include a comparison among several population means from a completely randomized design. A brief introduction of the notations and the models provides the basis for discussing the analysis of variance for data from other experimental designs in the subsequent chapter.

Chapter 9 expands the discussion of analysis of variance beyond the completely randomized design of Chapter 8 to include several different experimental designs and treatment structures. The designs considered are the randomized complete block design and the Latin square design, which are useful for minimizing the effect of undesirable variability from extraneous variables to obtain more precise comparisons among treatment means.

Chapter 10 examines specific inferences that can be made from experiments involving categorical data. The chapter begins with the basic inferences for one or two population proportions and ends with goodness-of-fit test and tests for independence and homogeneity of distributions.

Chapter 11 is devoted to simple linear regression and correlation between two quantitative variables. Chapter 12 consolidates the methods of Chapter 11 for expressing a dependent variable as a function of two or more independent variables. After presenting various multiple regression models and the interpre-

tations of coefficients in these models, the authors present the normal equations used in obtaining the least squares estimates. They also develop inferential procedures for the coefficients.

Chapter 13 begins with some general pitfalls and difficulties that a researcher may encounter in effectively communicating the results of analysis. This short but comprehensive chapter offers wise and clever comments on how to present the results of a statistical analysis of data and discusses some of the problems in effectively communicating these results to the intended audience.

Overall, *A First Course in Statistical Methods* would be a (not “the”!) good choice for studying the introductory probability and statistics for any college students, although the authors target advanced undergraduate and graduate students from a variety of disciplines. It is not easy to find a book to study with, but this book covers fundamental topics for students who have no prior coursework in statistics. Its statistical contents and mathematical maturity are not overwhelming even for entry-level students. One drawback of this edition is related to computer software. The authors provide outputs from several major statistical packages (Minitab, SAS, and a few others) for examples and exercise questions, but give no information on how to obtain those outputs using statistical packages.

Hoon KIM
California State Polytechnic University, Pomona

REFERENCE

Ott, R. L., and Longnecker, M. (2001), *An Introduction to Statistical Methods and Data Analysis* (5th ed.), Pacific Grove, CA: Duxbury Press.

Basic Statistics and Data Analysis, by Larry J. KITCHENS, Pacific Grove, CA: Brooks/Cole, 2003, ISBN 0-534-38465-X, xiii + 637 pp. + CD, \$91.95.

Professor Kitchens’ book is fairly unique as far as introductory statistics texts go. Besides the usual spate of topics and methods, he applies a heavy dose of nonparametric statistics. Any evaluation of this text’s merit will be strongly determined by the user’s position on the place of nonparametric methods in an introductory course in general and Kitchens’ specific treatment of the topic. Before getting to these issues, let me get some of the basics out of the way.

This book is intended as a one- or two-semester introduction to statistics for a general audience at the undergraduate level. High school algebra is all the math required. There are no derivations and minimal equations. The homework questions at the end of each section are divided into those that require hand/calculator work, interpreting computer output, and those that require a computer to solve. The CD accompanying the text contains all of the data files in the examples and homework in a variety of formats. Answers to selected odd-numbered exercises are provided in the back of the book. All of the computer output examples are from Minitab. There are also instructions on how to use Minitab to perform different statistical tests.

As far as the general topics covered, the book is fairly typical of most introductory texts. The first two chapters deal with summarizing data and looking at data with graphical methods. Throughout the book, Kitchens emphasizes the use of graphical methods to view the data. This is good. Boxplots, stem and leaf plots, and normal probability plots seem to be his favorites. I do think he goes a bit too far by introducing correlation, straight line regression, and contingency tables in the early data exploration sections.

Chapter 3 is a very brief introduction to probability and distribution functions. It covers mutually exclusive and independent events, but does not touch conditional probabilities. As far as distribution functions go, he presents only the binomial and normal. Chapter 4 covers the sampling distributions for \bar{X} and proportions. It introduces the central limit theorem (CLT), but only in the limited context of how averaging tends to produce a statistic that is more normally distributed. The author presents some good simulations that illustrate how the sample size required to get reasonably normally distributed averages depends on the shape of the parent distribution.

Chapters 5–7 cover confidence intervals and statistical inference. Continuous variables and proportions are worked in tandem in these chapters. The statistical inference chapters cover the usual one- and two-sample tests. For the two-sample t -test, the author steers the students toward the nonequal variance version. He does not present the formal F -test for equal variances, because he believes that it is too sensitive to outliers. In its place, he offers the rule of thumb

that if the variance ratio is less than 3, then equal variances may be assumed. If in doubt, use the nonequal variance test should be used. Interpreting p values is stressed over critical values of the test statistic. Using confidence intervals as an alternative to formal hypothesis testing is clearly explained.

Chapters 8–10 deal with categorical methods, regression and ANOVA. Categorical analysis introduces the chi-squared distribution and covers goodness of fit, tests for independence, and tests of the homogeneity of two samples. The linear regression chapter covers only the one-predictor variable case. The coverage is adequate, but the reader will need to review the relevant material in Chapter 2. The final chapter covers one-way ANOVA and Tukey’s multiple comparison test.

This book is not appropriate for students of math, science, and engineering. For one, its lack of any mathematical sophistication severely limits the depth of application of statistics into any of these disciplines. Second, many topics relevant to these majors are covered poorly or not at all. Examples of these are the coverage (or lack thereof) of conditional probabilities, power, sample size, and the use of distribution functions other than the most basic and applications to statistical process control.

As an introduction to statistics for a general audience, the book has some merit. The writing style is very readable, and it has many engaging real world examples. Most of the technical content is taught by working through the examples. A key issue for any perspective user of the book is its treatment of nonparametric statistics. If they cover it at all, most introductory texts relegate nonparametric statistics to an optional chapter toward the end of the book. Kitchens includes a nonparametric test for each of the method categories that he covers. These include confidence intervals for the median, nonparametric versions of the t -test (single-sample, two-sample, and paired), regression on medians (qualitative discussion only), and the Kurskal–Wallis alternative to the one-way ANOVA. The nonparametric versions of the t -test are modifications of the Wilcoxon rank tests, where the usual t -test is performed on rank scores instead of the original data.

The onus born by such an approach is the need to make clear when to use a nonparametric method over methods that make assumptions about normality. The author reiterates many times that the major watch outs are for distributions that have long tails or are highly skewed (sometimes he removes the adjective “highly”). Long tails, he argues, run the risk of including outliers that could bias any particular measurement of a mean. The median is more robust to distortion by outliers. For skewed distributions, he argues that the median by nature is a better descriptor for centrality than the mean. He rightly emphasizes the use of graphical analysis to picture how the data are distributed. These include versions of the histogram (stem and leaf and dot plots) and normal probability plots. Unfortunately, his criteria for when to use the nonparametric methods are not always clear and sometimes are arguably misleading.

As an example, consider his comments concerning the proper two-sample test to use. For large sample sizes ($n \geq 30$) he notes that the CLT does a reasonably good job making \bar{X} close enough to normal to justify using the t -test. “Because the Central Limit Theorem applies regardless of the shape of the underlying distribution, there are hardly any restrictions on using the two-sample t -test when the sample sizes are large” (p. 407). But he then goes on to say the following: “If the parent distributions are skewed, it has been suggested that we make inferences about the difference between population medians, $\Theta_1 - \Theta_2$, instead of the difference between population means. If this is the case, you should use the Wilcoxon procedure instead of the t -test” (p. 407). He appears to be saying that even if the sample means are reasonably normal (a la the CLT), if the parent distributions are skewed, then use the nonparametric test. First, this is confusing. What was the point about the CLT as it applies to averaging? Second, he offers bad advice. If the test statistic is reasonably normally distributed, then the parametric test is more powerful than the nonparametric test.

Things are not better for small sample sizes. For small-sample t -tests, the author presents normal probability plots to establish normality. For the nonparametric example, he looks at a dot plot, notes that it does not look very normal, and thus concludes that the nonparametric test is called for. Had he done a normal probability plot of the data, it would look as “normal” as those used in the t -test examples. The problem, of course, is that for small samples it is hard to get a reliable picture of how the parent distribution is shaped. Kitchens’ advice clearly steers the students toward the nonparametric test, as when he states that, “the applications for the Wilcoxon test are numerous, and in fact, many statisticians suggest that it should be used in all cases except when the parent distribution is known to be normally distributed. For small samples this is good advice” (p. 408). His recommendations are at best confusing and at worst wrong.

To summarize, *Basic Statistics and Data Analysis* is not suitable for students majoring in math, science, and engineering, because of the sparse coverage of statistical topics applicable to these fields. As a general audience text, the book is well written in terms of style and readability. However, the instructor would have to be predisposed to include a heavy dose of nonparametric statistics in an introductory course and plan on presenting clearer guidelines on when such tests are appropriate.

Robert V. BRILL
Astaris

Statistics for Research (3rd ed.), by Shirley DOWDY, Stanley WEARDEN, and Daniel CHILKO, Hoboken, NJ: Wiley, 2004, ISBN 0-471-26735-X, xvi + 627 pp., \$94.95.

Statistics for Research is written to accommodate a two-semester introductory statistical methods course for graduate students coming from various research disciplines in the natural and social sciences. A strong mathematical background is not assumed, and throughout the text mathematical rigor is substituted with a conceptual approach to teaching statistical concepts. However, to stay away from what the authors call a “black box” approach, they do include examples in which the details of the calculations are presented.

There are several additions to the third edition that are worth noting. First, besides discussing SAS software, the authors also discuss the use of JMP, a Windows-based statistical software package developed by the SAS Institute. Second, the authors have added a few sections early in the text that discuss probability concepts. In the previous editions, probability was discussed as needed throughout the text. Third, a section on logistic regression has been added to the last chapter that briefly introduces the idea of maximum likelihood estimation and goes through one detailed example. Finally, other additions include discussions of retrospective studies, risks, odds ratios and repeated-measures studies, the Bonferroni correction, and ratio and difference estimation. There are also additions to the exercises that appear at the end of each section.

The authors do a very nice job by introducing topics with many examples and illustrations. Shortly after a definition is given, an example is sure to follow that makes the idea hit home in a very concrete manner. The exercises at the end of each section are very well thought through and provide the student with a chance to test his or her understanding. I especially like the true–false review exercises at the end of each chapter. One excellent use of the true–false questions would be to have the students write an explanation about why the question is true or false.

The text comprises 14 chapters:

1. The Role of Statistics. Provides a motivational introduction and basic ideas of inferential statistics.
2. Populations, Samples, and Probability Distributions. Covers random variables, probability distributions and expected value; provides many examples.
3. Binomial Distributions. Incorporates the ideas of type I and type II error nicely within testing a hypothesis using a binomial distribution.
4. Poisson Distributions. Motivates the Poisson distribution and shows its approximation to the binomial distribution.
5. Chi-Square Distributions. Focuses on goodness-of-fit tests, tests of homogeneity and independence; discusses relative risks and odds ratios and briefly discusses the nonparametric median test.
6. Sampling Distribution of Averages. Provides a nice discussion of population mean and variance versus the estimated mean and variance.
7. Normal Distributions. Covers the central limit theorem, confidence interval for the mean, approximations to binomial and Poisson distributions, and the nonparametric rank test.
8. Student’s t Distribution. Contains a very nice flowchart for different situations depending on whether the interest lies with testing the mean or variance with continuous data or a proportion with discrete data.
9. Distributions of Two Variables. Covers regression and correlation, with a nice section on computer usage using JMP.
10. Techniques for One-Way Analysis of Variance. Provides a very clear discussion on why ANOVA is concerned with variances when testing means and a nice section on multiple comparison techniques.
11. The Analysis-of-Variance Model. Covers the model, assumptions, and transformations.
12. Other Analysis-of-Variance Designs. Discusses nested, randomized complete block, Latin square, factorial, split-plot, and repeated-measures designs.

13. Analysis of Covariance. Provides very nice examples and explanation of the ANCOVA concept, along with an interpretation of SAS output.
14. Multiple Regression and Correlation. Contains a section on logistic regression.

A selected reading section at the end of each chapter contains a fairly good list of reference material. In addition, answers to odd numbered exercises are given in the back of the text, along with many look-up tables (of, e.g., t and F critical values).

In summary, *Statistics for Research* provides material for a solid two-semester introductory statistics course for graduate students in the natural and social sciences. I could see, however, the need for the text material to be supplemented with lab exercises that clearly show how to perform statistical analysis using statistical software beyond what is done in the text with SAS and JMP. The text is easy to read, and students will enjoy the wide range of examples and illustrations. The additions made to the third edition make it a nice improvement over the first and second editions.

William A. BRENNEMAN
Procter & Gamble Company

Diagnostic Checks in Time Series, by Wai Kung LI, Boca Raton, FL: Chapman & Hall/CRC, 2004, ISBN 1-58488-337-5, xiii + 196 pp., \$69.95.

A researcher expecting to learn about diagnostic checks in time series will be sorely disappointed by this book on two counts, for it is much more narrowly focused than the title suggests. First, it refers only to stationary time series, thus omitting diagnostic tests for cointegrating regressions, for example. Second, it concentrates almost exclusively on portmanteau tests, which constitute only a small subset of all diagnostic tests.

Chapter One (4 pp.) is the introduction. Chapter Two (18 pp.), “Diagnostic Checks for Univariate Linear Models,” covers (partial) autocorrelation tests of residuals, with extensions to seasonal autoregressive moving average (ARMA) models and periodic autoregressions. Chapter Three (19 pp.), “The Multivariate Linear Case,” addresses vector ARMA models, Granger causality tests, and transfer function noise modeling. Chapter Four (17 pp.), “Robust Modeling and Diagnostic Checking,” presents robust varieties of portmanteau and cross-correlation tests. Chapter Five (20 pp.), “Nonlinear Models,” considers tests for nonlinear model structure; tests for linear versus specific nonlinear alternatives, goodness-of-fit tests for nonlinear time series, and choosing between different families of nonlinear models. Chapter Six (31 pp.), “Conditional Heteroscedasticity Models,” presents autoregressive conditional heteroscedasticity (ARCH) model, tests for the presence of ARCH errors, diagnostic checks for ARCH models both univariate and multivariate, and a test for causality in the variance. Chapter Seven (16 pp.), “Fractionally Differenced Process,” considers exact and approximate methods for estimating the parameter of interest, gives a portmanteau test statistic and an example based on tree-ring widths. Chapter Eight (15 pp.), “Miscellaneous Models and Topics,” discusses ARMA models with nonnormal errors and other nonnormal time series, the autoregressive conditional duration model, and a power transformation to induce normality.

The text contains several typos, which suggests that the equations are similarly marred by inadequate proofreading. (I did not check the equations for errors.) A researcher who desires a resource for portmanteau tests applied to stationary time series will find the book useful.

B. D. McCULLOUGH
Drexel University

Ranked Set Sampling: Theory and Applications, by Zehua CHEN, Zhidong BAI, and Bimal K. SINHA, New York: Springer-Verlag, 2004, ISBN 0-387-40263-2, xii + 224 pp., \$59.95.

This comprehensive and up-to-date monograph covers, systematically and in simple language, the theory and applications of ranked set sampling (RSS), an improved technique related to traditional simple random sampling (SRS). Strong emphasis is placed on theoretical developments in RSS. In the meanwhile, the practical orientation and broad coverage will appeal to researchers and scientists working in sampling techniques, experimental designs, nonparametric statistics, and related fields.

The book comprises eight chapters of varying length. The inclusion of sections at the end of the main chapters to collect more technical arguments and to give bibliographic notes works well and helps the readers explore in more depth aspects of RSS in which they are interested.

Chapter 1 introduces the notion and general procedure of RSS. This very useful chapter will enable readers to quickly enter into the realm of RSS, learn about its historical developments, and identify applications of particular interest.

Chapters 2 and 3 discuss balanced RSS. In particular, Chapter 2 focuses on nonparametric RSS, in which no assumption on the underlying distribution of the variable of interest is made. This chapter studies in detail the relative efficiency of RSS with respect to SRS in the estimation of a population mean, a smooth function of means, and population quantiles. The authors also consider the inference procedures, such as the construction of confidence intervals and hypothesis testing. To facilitate the inference procedures based on RSS sample quantiles, they also discuss the kernel method of density estimation. This section is quite interesting. The chapter also presents some robust procedure based on M -estimates with RSS data.

Chapter 3 addresses parametric RSS, where the underlying distribution of the variable of interest is assumed to belong to some parametric family (e.g., location-scale family and shape-scale family) of distributions. The authors nicely lay out the theoretical foundation for the parametric RSS via Fisher information. The maximum likelihood estimate (MLE) based on RSS and its relative efficiency with respect to MLE based on SRS are studied, and the best linear unbiased estimate for location family of distributions is dealt with.

Chapter 4 studies unbalanced RSS. This chapter first develops the methodology of analyzing RSS data for the inferences on distribution functions and quantiles, as well as general statistical functionals. The optimal designs for the parametric location-scale family and for nonparametric estimation of quantiles are discussed in detail. This chapter also contains methods of Bayes design and adaptive design.

Chapter 5 explores classical distribution-free tests in the context of RSS. The authors consider the sign test, signed rank test, and Mann–Whitney–Wilcoxon tests and revisit the issue of the optimal design for distribution-free tests. Readers with a prior knowledge of nonparametric tests at the level of Gibbons and Chakraborti (2003) will find this chapter informative and easy to understand. For readers not familiar with these standard topics, some brief additional explanation and references might have been beneficial for the wider accessibility.

Chapter 6 describes RSS with concomitant variables. A multilayer RSS scheme and an adaptive RSS scheme using multiple concomitant variables are developed; the general regression analysis using RSS is discussed; and the design of optimal RSS schemes for regression analysis, on the basis of the concomitant variables, is explored.

Chapter 7 illustrates RSS as a data reduction tool for data mining, whereas Chapter 8 exemplifies the practical features of RSS via case studies. The inclusion of this last chapter on case studies with RSS further enhances the value of this monograph for practitioners and applied statisticians.

In the development of RSS, the choices of ranked set size k and cycle number m are directly pertinent to practical problems. This book would have been more useful had some more detailed discussion on the choices been added. However, overall I would highly recommend this well-written and reasonably priced book to researchers and practitioners, all of whom are likely to use one or more of the methods it discusses.

Chunming ZHANG
University of Wisconsin

REFERENCE

Gibbons, J. D., and Chakraborti, S. (2003), *Nonparametric Statistical Inference* (4th ed.), New York: Marcel Dekker.

Highly Structured Stochastic Systems, edited by Peter J. GREEN, Nils Lid HJORT, and Sylvia RICHARDSON, New York: Oxford University Press, 2003, ISBN 0-19-851055-1, xxii + 509 pp., \$69.50.

This book is a compilation of 10 years study regarding the development of highly structured stochastic systems (HSSS). This topic has undergone many

major developments during this period, and the contributors in this text are many of the leading developers in this area. The text points out that most of the origins of this field were developed in many diverse fields, and it was the editors's intent to develop and enhance understanding by building a common structure. As such, the ESF initiated and funded a program to develop a cross-disciplinary collaboration. This text is a summary of this work and the multiple workshops held over the past 10 years.

The topic of HSSS itself is beginning to emerge with significant practical implications due to the innovations discussed in this text. The book is written by multiple content experts and contains further discussions using addendums written by experts in the respective topics. This allows the reader to delve into other developments on a specific topic with instant access to the major theme. It was apparent that the editors spent a lot of time ensuring that these multiple discussions were complementary to the work being presented.

As a whole, the book is well laid out and the topics are discussed with great detail without being cumbersome. Moreover, the reading is continuous, and great care was taken to keep major themes consistent—an extremely difficult task given the number of contributors to the text. With this in mind, I now provide a detailed discussion of the book.

The first sections of the book are devoted to the introduction of HSSS models and their methodological foundations. The authors assume a basic understanding of the definitions and terms used in these models, but the topics are covered such that an interested reader can pick up the book without being overwhelmed. Although a more formal treatment of DAG models in statistics was given in *Probabilistic Networks and Expert Systems* (Cowell, Dawid, Lauritzen, and Spiegelhalter 1999) two of the authors of that book are contributors to this text. The first chapters discuss graphical models and their interplay with causal inference, specifically DAG models, Bayesian networks, chain graphs and graphical time series models are discussed with applications from areas such as DNA analysis and social science.

The second part of the book deals with algorithms for calculating HSSS models. The major developments regarding Markov chain Monte Carlo (MCMC) theory are presented, and multiple MCMC methodologies are discussed along with their relevance to solving particular problems. These chapters provide a comprehensive discussion of MCMC techniques from theoretical understanding to practical implementation discussions. Although the discussions are well developed, a casual reader can grasp and build from these discussions.

The third set of chapters deals with HSSS applications in epidemiologic models, spatial point processes, spatiotemporal processes, ecologic applications, binary analysis, image analysis, and genetic modeling. The final section discusses HSSS model criticism and Bayesian nonparametric smoothing, model criticism being the HSSS approach to model selection using such methods as sensitivity studies and predictive capability analysis. The nonparametric chapters cover standard material from the Bayesian perspective.

Overall, *Highly Structured Stochastic Systems* provides an extensive summary of the work done over the past 10 years on HSSS models, with numerous examples to demonstrate their strength. In the coming years, the use of HSSS models will become widespread with the rise of data information and the need to specify complexity in applicable terms. This book will be an oft-cited reference when these issues become necessary to analyzing complicated relational data.

Nicholas ROSE
Frequency Marketing, Inc.

REFERENCE

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999), *Probabilistic Networks and Expert Systems*, New York: Springer-Verlag.

Bayesian Artificial Intelligence, by Kevin B. KORB and Ann E. NICHOLSON, Boca Raton, FL: Chapman & Hall/CRC, 2004, ISBN 1-58488-387-1, xxviii + 364 pp., \$79.95.

Bayesian networks (BNs) are located in the confluence of statistics and computer science. They are increasingly being applied in medicine, biology,

engineering, and data mining. This book was written as a graduate or advanced undergraduate text in a computer science curriculum, so it can only hint at some of the greater applications yet to come.

As an example of a BN, the authors propose the problem of a faulty burglar alarm that sometimes goes off when a minor earthquake occurs. Two different neighbors might hear it with different probabilities and independently may or may not call the police, who may or may not respond. The statistical approach to this problem might be to assess the conditional and marginal probabilities of the various events, apply Bayes's theorem as necessary, and be done with it.

The real virtue of the BN methodology is to codify the more general structure of directed graphs and to provide computational algorithms for parsing and updating the links between the nodes. The early chapters make considerable use of fundamental Bayesian ideas, such as incoherence and optimal betting including the Dutch book. These are well explained in simple language that will be easily understood. The mathematics of BN models are expressed in terms of odds and probabilities, as well as systems of Bayesian beliefs. Most of the more abstract concepts of Jeffrey and de Finetti included here will go over the heads of an undergraduate audience. Some of the digressions, including Section 1.8, titled "Are Bayesian Networks Bayesian?," can get really scary, especially when we read that the answer is to this question is "no."

Overall, the mathematical tone is quite tame, and the target audience of computer science or statistics graduate students should be able to appreciate the material. Every chapter is followed by a set of exercises that are also within the target audience's reach. The writing is not primarily intended for statisticians. There are references to such topics as the Turing test and John Searle's Chinese room, for example. These are mentioned without explanation, because it is assumed that intended reader is already very familiar with these ideas. Nevertheless, this statistician reader did follow the thread of the authors' arguments throughout.

There is considerable material early on explaining the causal inference and direction of reasoning that takes place in BNs, reflecting the fundamental work of Judea Pearl, one of the two persons to whom this book is dedicated. The authors explain a number of algorithms for computational work on BNs. There are times when the only solution is through simulation. The simulation algorithms implement Markov chain Monte Carlo and other methods for sampling from the posterior distribution. A nice feature of the book is the extensive survey of the available software, much of it downloadable for free on the web.

Things start getting interesting in Chapters 4 and 5. Decision networks include preferences and utilities that allow us to quantify the consequences of actions taken. Chapter 5 hints at some of the complex applications available, including the problem of medical diagnoses given a multitude of symptoms and laboratory tests that can be performed. This is really the way in which medical decisions are made. Another example appears in weather forecasting. The idea of a BN analysis of a poker game is really delightful. The nice argument generator (NAG) uses natural language to refute your last statement and do it in such a way that your feelings are also taken into consideration. Perhaps a particular family member comes to mind as you read this section. Chapter 9 shows how expert opinion can be used to build the various links of the BN, in a process called knowledge engineering with Bayesian networks (KEBN). Links can be added to and removed from the BN using the Matilda KEBN software as the various experts reply to questions.

Bayesian Artificial Intelligence provides a very solid introduction to BNs for those statisticians who may have heard about BNs but are unfamiliar with their basics. The many examples clearly illustrate the topics, and there are many hints at the broader applications.

Daniel ZELTERMAN
Yale University

Applied Data Mining, by Paolo GIUDICI, West Sussex, U.K.: Wiley, 2003, ISBN 0-470-84678-X, xii + 364 pp., \$99.95.

Subtitled "Statistical Methods for Business and Industry," this book provides a broad description and applications of selected statistical and computational methods for data mining (DM). Aimed at business users, the pedagogy is from art, law, or medicine. Part I devotes six chapters (205 pages) to compact, well-written explanations of the underlying mathematical models. Little space is spent explaining the data analysis problems solved or justifying the methods selected. The six chapters of Part II (143 pages) present business case

studies and apply selected data analysis (DA) methods and interpret their practical results. Multiple models are applied to each problem dataset, and the best model is selected and interpreted. Overall case results are concisely reported in a highly structured summary.

SAS Enterprise Miner (EM) is used exclusively to illustrate and apply methods. The book showcases the capabilities of this software, illustrates the contents of a comprehensive DM tool set, and presents a simplified overview of the collected mathematical models. This approach presumes the reader's access to SAS software and manuals, because no details are provided about code, actual datasets used, or output production.

Chapter 1, "Introduction," establishes the book's purpose. It takes on the daunting tasks of defining DM, systematizing the DM process, and persuading both the computer science and statistical communities that the vast pantheon of DM methods can be comprehended and embraced by collecting both groups' methods in one place and accepting the author's unifying definitions. It addresses the statistician's concerns about "data dredging," defines the DM process, justifies DM to achieve business process insight, and encourages statisticians to lead corporate DM efforts.

Part I, Methodology, comprises the following chapters:

Chapter 2, Organization of the Data. Covers data structures, data classes, the data matrix, frequency distributions and data transformation.

Chapter 3, Exploratory Data Analysis (EDA). Discusses univariate EDA, bivariate EDA, multivariate EDA, and principal components.

Chapter 4, Computational Data Mining. Explores distance measures, cluster analysis, linear regression, logistic regression, discriminant analysis, tree models, neural networks, nearest-neighbor models, local association rules, and retrieval by content.

Chapter 5, Statistical Data Mining. Covers probability, statistical models, statistical inference, nonparametric modeling, normal linear model inferences, generalized linear models, exponential model family, logistic regression, log-linear models, graphical log-linear models, and graphical models compared with neural networks.

Chapter 6, Evaluation of Data Mining Methods. Explores statistical criteria, distance, discrepancy, the Kullback-Leibler discrepancy, scoring function criteria, Bayesian criteria, computational criteria, and (economic) loss criteria.

Part II, Business Cases, comprises the following chapters:

Chapter 7, Market Basket Analysis. Discusses log-linear models and association rules.

Chapter 8, Web Clickstream Analysis. Explores sequence rules, link analysis, probabilistic expert systems, and Markov chains.

Chapter 9, Profiling Web Visitors. Covers cluster analysis and Kohonen maps.

Chapter 10, Customer Relationship Management. Covers logistic regression, radial basis function networks, classification trees, and nearest-neighbor models.

Chapter 11, Credit Scoring. Deals with logistic regression, classification tree models, and multilayer perceptron models.

Chapter 12, Forecasting Television Audience. Topics include linear regression, classification tree models, multilayer perceptron models, radial basis function networks, and nearest-neighbor models.

The author's style is consistently readable. Stripping out all but the barest essential mathematics makes the remaining material very approachable to a model-centric audience. The author defines his audience as "advanced undergraduate and graduate students of business management, computer science, and statistics." However, the presumed breadth of exposure to statistical and mathematical concepts, models, terminology, analysis, and computing strategies requires a profile more typical of experienced professionals.

Eliminating all theory and most computational detail reveals the basic DA strategy common to DM methods. However, deferring nearly all examples to Part II leaves Part I abstracted to a level nearly devoid of motivation. All insights gained from examples are deferred to the case studies in Part II, obliging the reader to jump back and forth.

There is sparse use of or complete omission of many superior textbook features. There are no exercises. Calculations are not annotated or visually linked to the text. No chapter and page reference citations are provided. Equations are not numbered, and many explanations beg for a simple graph. All computer output appears to be default menu generated, including far too many significant figures.

Several of the logic structures and formats used leave much to be desired. Part I leaves the reader with only the barest explanation of which methods apply to which problems. There are rather large leaps in mathematical symbolism that spans algebra, integral calculus, matrices, sets, and directed graphs. Questions regarding the details of computation, code, algorithms, mathematics, and statistics are directed to references. This tactic would be appealing had the citations referred to specific chapters and pages. Another frequent irritation is that much terminology is introduced without definition and many terms are not explained until subsequent chapters. References to case studies in the methods section are few, and the case studies themselves do not obviously point back to the methods section that they support. Consequently, it is easy to get lost in tangential concepts, which could have been avoided with a more sequential presentation.

In Part II, each case is presented according to the following outline: Objectives, description of the data, EDA, model building, model comparison, and summary report. Only the models change between cases. The individual cases are well organized, and the structure of the presentation will appeal to any scientifically inclined reader. My only serious criticism of Part II is that the defense of the "best" method chosen often seems highly subjective, especially when computational methods are being evaluated.

Although the author has achieved the purposes he sets out in Chapter 1, I can recommend this book only to experienced statisticians and analysts who either have access to SAS EM or are trying to justify its purchase to their management. It might also be of value to users trying to assemble a set of DM procedures, because SAS has clearly defined the state of the art in DM packages. Those interested in the do-it-yourself SAS approach may wish to consult the texts by Fernandez (2003) and Rud (2001), reported in *Technometrics* by Caby (2004) and Ziegel (2002).

Thomas D. SANDRY
ENTRIX, Inc.

REFERENCES

- Caby, E. C. (2004), Review of *Data Mining Using SAS® Applications*, by G. Fernandez, *Technometrics*, 46, 260–261.
 Fernandez, G. (2003), *Data Mining Using SAS® Applications*, Boca Raton, FL: Chapman & Hall/CRC.
 Rud, O. P. (2001), *Data Mining Cookbook*, New York: Wiley.
 Ziegel, E. (2002), Editor's Report on *Data Mining Cookbook*, by O. Rud, *Technometrics*, 44, 94.

Reliability and Statistics in Geotechnical Engineering, by Gregory B. BAECHER and John T. CHRISTIAN, West Sussex, U.K.: Wiley, 2003, ISBN 0-471-49833-5, xii + 605 pp., \$160.00.

Geotechnical engineering might be described as the design and construction of structures in or on the earth. Whereas the design of the structures themselves may be well within the basics of physics and mathematics and hence within the control and knowledge of the designer and engineer, earth and soil properties that may impact the structures are likely not well known. The latter properties are not spatially constant, may require far more observation or testing than is feasible, or may be temporally changing. The classic approach would be to "overdesign," but this may not be feasible or practical.

This monograph argues for treating both unknown features and the subsequent reliability of the structures from a probabilistic and statistical perspective. To that end, Chapter 1 begins with both a history of geotechnical engineering, as well as a discussion of some classical examples where there is the potential for significant failure, including offshore platforms, open pit mine slopes, and dams. Besides attempting to convince the reader of both the importance and usefulness of using probability and statistics for geotechnical problems, two other themes appear throughout. A wide variety of actual geotechnical applications are presented, backed up with references to the literature or to reports. There are many drawings, figures, graphs, and tables. Finally, there is at least some emphasis on presenting the probability and statistics that are considered useful. However, this is not a statistics text, not even one intended for students in geotechnical engineering, because there are no exercises or questions. The intended audience is not quite clear; in particular, the prerequisites for reading the book are not clear, but it appears that the reader is assumed to be somewhat fa-

miliar with geotechnical problems. Although many examples of computations, specifically statistical computations, are included there is no discussion of software or reference to the use of any particular software.

It is useful to examine the book's organization. It consists of four parts and an appendix. As noted earlier, Chapter 1 covers motivation and history. The other four chapters in Part I might be considered background material in probability and statistics; the Appendix is also a "primer" on probability theory. Chapter 2 is mostly a philosophical discussion of probability. The part on probability theory that might look most like a text intended for engineers is all contained in Chapter 3, and the material on statistics is all contained in Chapter 4. Chapter 5 explores risk analysis. A significant amount of new material on probability or statistics is scattered throughout subsequent chapters, however. Although there is a discussion of different perspectives on statistical inference in Chapter 4 (i.e., Neyman–Pearson vs. Fisher vs. Jeffreys), the emphasis here and subsequently is on Bayesian methods.

Part II, comprising of seven chapters, focuses on site characterization, including data collection and data analysis. A number of probability and statistics ideas and techniques are introduced in these chapters, including discriminant analysis, trend surfaces, kriging, multivariate Gaussian distributions, maximum likelihood, and (auto) covariance functions.

Part III is concerned with the reliability analysis and comprises seven chapters. Chapter 14 is an extensive discussion of the James Bay dikes and analysis using first-order second moment analysis. This is followed by a discussion in Chapter 15 of Rosenbleuth's point-estimate method. Monte Carlo methods are discussed in Chapter 17, and stochastic finite elements are covered in Chapter 19.

Part IV focuses on the use of expert opinion, the use subjective probabilities and event tree analysis.

All in all, the range of applications is very broad, as is the range of probabilistic and statistical methods. Many of the applications are likely to be unfamiliar to most statisticians, and for that reason this can be interesting reading. Those two aspects, as well as the extensive documentation from the literature, are real positives.

However, there are some negatives, some merely annoying and others more serious. On page 41, Axiom 3 is the definition of independent events, not an axiom. The equation for differential entropy is given on page 75, but it is claimed to be the same as Shannon's information entropy. Table 4.3 (p. 75) lists "maximum entropy distributions," but the authors never explain what the "maximum" pertains to. In discussing type I and type II errors and error probabilities on page 89, the authors state that "the tool for making this comparison is the power function, which describes the conditional probability of a type II error resulting from the use of that decision rule, given the state of nature." There are a great many errors in Chapter 9 pertaining to the discussion of autocovariance functions, autocorrelation functions, and variograms. Although there are references to and comments about "stationarity," the authors do not mention the distinction between strong stationarity (translation invariance of multivariate distributions), second-order stationarity, and intrinsic stationarity until a later chapter. They do not mention the distinction between trend, a property of the data, and the mean of a random function (field). On page 221 they compare fitted autocovariance functions after fitting with an assumed constant mean, a first-order trend surface, and then a second-order trend surface. They never acknowledge that not all three can be correct; at least two are wrong in that they are based on an incorrect assumption about the mean of the random function. On page 235 (also Fig. 9.25) the authors refer to simulated data, using a given constant mean, given an autocorrelation function and multivariate Gaussian distribution. There is no explanation of the algorithm used. If an L–U decomposition of the covariance/correlation matrix is used, then the autocorrelation function is reproduced only in an average sense, but this is not acknowledged. The explanation pertaining to the sill of a variogram on page 240 is wrong, and that pertaining to the range of a variogram is not quite correct. The authors do not acknowledge the difference between an actual range for spherical model and the "effective" ranges for the exponential and Gaussian models. In discussing kriging (pp. 264–268), the authors make no distinction between simple kriging, ordinary kriging, and universal kriging. Although the Lagrange multiplier is identified as relating to the optimization, there is no mention of the unbiasedness constraint that leads to the use of the Lagrange multiplier. There is no mention of block kriging, which would greatly simplify the discussion of spatially averaged random fields. On page 268, it would appear that the authors are using a zonal anisotropic variogram (Fig. 11.2), but they do not indicate how they ensure that the model is valid (i.e., conditionally negative definite).

On page 461, the authors seem to attribute the turning bands simulation algorithm to Mantoglu and Wilson (1982), whose work is an extension of the work of Journel (1974), which in turn is based on that of Matheron (1973). Although the authors mention the L–U decomposition algorithm, they do not mention other more commonly used algorithms, such as sequential Gaussian and simulated annealing.

The annoying items include the following: on page 44, a sequence is referred to as an “ordered series.” On page 45 (and a number of other places) “cumulative density function” is used in lieu of cumulative distribution function. On page 47, there is failure to distinguish between discrete and continuous random variables when defining the density function. There is a consistent failure to recognize that probability distribution functions may not be differentiable (and, moreover, the derivative may not be the density). There is no recognition that moments may not exist. On page 61 we find separate definitions/explanations for probabilistic correlation and statistical correlation. The latter “means that the two uncertainties are simultaneously estimated from a given set of data and therefore are influenced by a common sampling variability error.” Table 9.1 (p. 229) is titled “One-Dimensional Autocorrelation Models,” but lists several models valid in higher dimensions. Several models claimed to be valid only in one dimension are actually valid in higher dimensions. What is labeled as the linear model is actually the truncated linear model. The power covariance model does not correspond to the power variogram model in Table 9.2. The squared exponential model is incorrectly defined. Table 9.2 (p. 240) is titled “One-Dimensional Variogram Models,” but in fact the models listed are not all “one-dimensional” but are isotropic models. Although the power model is given (albeit incorrectly because of a typo), the linear model with a nugget is given separately and is claimed to be valid only in one dimension. It is possible that the authors intended to give the truncated linear model, which is only valid in one dimension. On page 243 the authors distinguish between a stochastic process (apparently limited to one dimension) and a random field (higher dimensions), but then several lines later define a random field as a joint probability distribution. It is not clear how the number of variables in this joint distribution is supposed to relate to the dimension of the space (or if it does). On page 235, the authors refer to a “spatial field,” although it is clear they mean a random field. On page 264, they call it a “spatial population.” Although γ was used to denote the variogram in the previous chapter, in Chapter 10 it is used to denote the “variance reduction function,” that is, the ratio of the variance of a spatially averaged random process (only one-dimensional examples are considered) to the variable of the process (assuming second-order stationarity). All of the discussion of simple random sampling and systematic sampling (pp. 258–259) is limited to finite populations.

Donald E. MYERS
University of Arizona

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

Intelligent Data Analysis: An Introduction (2nd ed.), edited by Michael BERTHOLD and David J. HAND, Berlin: Springer-Verlag, 2003, ISBN 3-540-43060-1, xi + 514 pp., \$64.95.

This may be a first. Here is a second edition (2E) of a book that is a collection of chapters. All 10 chapters of the first edition (1E), which appeared in 1998 (reviewed in Ziegel 2000), had different authors. Getting all of them interested in doing a revision must have been quite a task.

One excellent feature of the 2E is the addition of two new chapters. First is a chapter on support vector machines (SVMs) and other kernel methods.

SVMs have achieved a certain amount of popularity, especially in the scientific computing community. Second is a new chapter on visualization. This is a particularly nice overview with excellent descriptions and numerous illustrations, most in color, for a wide variety of types of visualizations. The authors should consider expanding this material to a full-length textbook. Besides the two new chapters, the chapter on rule induction was replaced with a new and expanded version. Otherwise, the organization of the nine original chapters remains unchanged.

The editors paid little heed to the comment in my report that “this is not a book for industrial statisticians, unless they are interested in the mathematical background for all these methods.” The graphics chapter is an exception. The other new chapter fits the previous mold, although the highest technical level for the rule induction chapter is a list of classification rules.

REFERENCE

Ziegel, E. (2000), Editor’s Report on *Intelligent Data Analysis*, edited by M. Berthold and D. Hand, *Technometrics*, 42, 442.

Mind on Statistics (2nd ed.), by Jessica M. UTTS and Robert F. HECKARD, Pacific Grove, CA: Brooks/Cole, 2004, ISBN 0-534-39306-3 (Instructor’s Edition), ISBN 0-534-39305-5 (Student Edition), xxi + 646 pp. + CD, \$98.95 (Student Edition).

This title may seem appropriate for a coffee-table statistics book, but it actually is one of many features that the authors have used to try to create an innovative new approach to teaching basic statistics. One presumes that this book has been very successful, because the authors and publisher have pumped out a second edition (2E) just 2 years after the first edition (1E). Seaver (2002), in his review of the 1E, described it as a book that “tries to move in some different direction” in presenting introductory statistics. Books in this category, he noted, will have “some strong advocates and some less than enthusiastic supporters.” Seaver gave both sides, liking “specific chapter features” but expressing reservations about “sequencing issues among chapters and occasionally in chapters.”

The latter perspective certainly was not one that affected the authors. There are only three additions to an otherwise complete repetition of the contents of the 1E. A discussion of “Other Sampling Methods” was added to the chapter on sampling, and two sections were added to the second of two chapters on significance tests. These additions contribute only 20 of the 75 additional pages in the 2E. Some of these additional pages result from new “Turn on Your Computer” sections that extend from the many “Turn on Your Mind” sections appearing throughout the book. These link the user to interactive Java applets that are part of the CD-ROM that comes with the book. Another new feature is the addition of “Mind on the Basics” exercises that “complement the conceptual and data analysis exercises” (Preface, p. xviii).

The authors state that “we wrote this book because we were tired of being told that what statisticians do is boring and difficult” (Preface, p. xvi). The layout is a principal contribution to making the book interesting. In addition to the aforementioned features, it is full of color, beautifully sectioned, effectively highlighted, and imaginatively enhanced with clever pictures. The CD-ROM provides support for the use of Minitab, Excel, SPSS, SAS, and JMP. Despite the lack of examples from science and engineering, I would be happy to hand this book to any of my colleagues who wanted a statistics book to study.

REFERENCE

Seaver, W. (2002), Review of *Mind on Statistics*, by J. Utts and R. Heckard, *Technometrics*, 44, 404–405.

Introduction to Spectral Analysis, by Petre STOICA and Randolph MOSES, Upper Saddle River, NJ: Prentice-Hall, 1997, ISBN 0-13-258419-0, xviii + 319 pp., \$60.00.

This is the second of two books that I requested when I discovered that no books on spectral analysis had been reviewed during my tenure as Editor.

(See Ziegel 2004 for the report on Warner 1998 that appeared in the previous issue.) This modest-sized and reasonably priced volume is considerably more mathematically-oriented than Warner's (1998) book. The size results from the authors' efforts "to keep the book concise without sacrificing rigor of presentation or skipping over essential aspects" (Preface, p. xi). The book is intended for use in a first course. The authors describe such topics as higher-order statistics, time-frequency analysis, and multidimensional spectral analysis as "too advanced for a concise first course in spectral analysis" (Preface, p. xii). However, matrix algebra, discrete-time linear systems, stochastic processes, and the first-course in mathematical statistics are the necessary background for reading this book.

Chapter 1 introduces power spectral density estimation and establishes a generally mathematical orientation for the contents of the book. There follow four chapters on temporal spectral estimation. Chapter 2 discusses nonparametric methods: periodograms, correlograms, and their estimation methods. Chapter 3 presents parametric methods for rational or continuous spectra: autoregressive moving average (ARMA) processes, Yule-Walker methods, and multivariate signals. Chapter 4 continues with spectral line methods for the sinusoidal model: nonlinear regression models, ARMA models, and multiple signal classification methods. Chapter 5 considers a newer nonparametric methodology, filter bank methods. The book's final chapter is devoted to spatial spectral estimation: array models, nonparametric methods, and parametric methods.

This may be the first book I have reviewed that contains not a single dataset or application. The section on "Complements" at the end of each chapter focuses on extensions, not illustrations. Actually, blending this book with the aforementioned book by Warner (1998) might make a reasonably balanced book. She must have had this one in mind when she wrote hers.

REFERENCES

- Warner, R. (1998), *Spectral Analysis of Time-Series Data*, New York: Guilford Press.
 Ziegel, E. (2004), Editor's Report on *Spectral Analysis of Time-Series Data*, by R. Warner, *Technometrics*, 46, 497–498.

Design for Six Sigma, by C. M. CREVELING, J. L. SLUTSKY, and D. ANTIS, Jr., Upper Saddle River, NJ: Prentice-Hall, 2003, ISBN 0-13-009223-1, xxviii + 768 pp., \$89.00.

Subtitled "In Technology and Product Development," this huge book is a comprehensive presentation on tools, best practices, and management methods that does not stint on information about statistics and quality management methods. Previously, a book on Design For Six Sigma (DFSS) by Stamatis (2003), one of a seven-volume series on Six Sigma methodology, was reported by Ziegel (2004). This relatively recent book was discovered during a literature survey for the report on the Stamatis book. Like that author, these authors feel that DFSS is much more involved than the simple adding of a "D" to the conventional measure, analyze, improve, and control (MAIC) cycle found in most references from the Six Sigma literature. This book has an invent, develop, optimize, and verify program for technology development and a concept, design, optimize, and verify program for product design. These processes actually would be precursors to the usual MAIC cycle.

In the Preface, the authors state that "DFSS consists of a set of needs-gathering, engineering, and statistical methods to be used during product development." They present their DFSS program in seven sections:

1. Organizational Leadership and Value Management
2. Product Development Processes
3. Critical Parameter Management (CPM)
4. Tools for Invention and Concept Development
5. Tools for Design Development
6. Tools for Optimization
7. Tools for Verifying Capability.

The first section discusses leadership by management, the measurement of financial results, and the management of value. These are the three designated components for success at the executive level. The second section presents in-

formation about program and project management, technology development, product design, and system design for DFSS. The third part explains CPM, its architecture, its use in product development, its tools and best practices, its project management metrics, and the necessary databases. The three parts that explain what should be accomplished comprise almost half of the book.

The tools for enhancing invention and innovation in Part 4 include a Japanese method (KJ) for assimilating the voice of the customer, quality function deployment, and methods for generating, evaluating, and modeling technology concepts. There is a long list of tools for design development: failure mode and effects analysis, reliability prediction, descriptive statistics, inferential statistics, measurement systems analysis, capability studies, multivar studies, regression, and design of experiments. Chapters on Taguchi techniques, response surface methods, and optimization tools populate the part on optimization. Tolerance design, reliability evaluation, and SPC provide tools for verifying capability. Each of the 21 chapters about the tools has an identical structure that is explicitly built around the answers to six questions:

- Where am I in the process?
- Where do I apply the tool?
- What am I doing in the process?
- What does the tool accomplish?
- What is the end result?
- What are the key deliverables?

Each chapter concludes with an appropriate process flow diagram, a verbal description of each of the steps, and a tools checklist.

The book uses input screens and outputs from Crystal Ball and Minitab. Compared to similar books of this ilk, such as that by Stamatis (2003), there is minimal reliance on lists and a considerably higher prevalence of well-written and understandable text. This book would be an excellent desktop reference for any business that wants to improve its processes for technology and product development.

REFERENCES

- Stamatis, D. (2003), *Six Sigma and Beyond: Design for Six Sigma*, Boca Raton, FL: CRC Press.
 Ziegel, E. (2004), Editor's Report on *Six Sigma and Beyond: Design for Six Sigma*, by D. Stamatis, *Technometrics*, 46, 372.

Quality Assurance in Analytical Chemistry, edited by B. W. WENCLAWIAK, M. KOCH, and E. HADJICOSTAS, Berlin: Springer-Verlag, 2004, ISBN 6-540-40578-X, xvi + 280 pp. + CD, \$69.95.

This is possibly the most unusual book that I have yet encountered as an editor. It consists entirely of small prints of PowerPoint slides on the right side of pages in a two-column format that has brief explanations for the material in the left column. The 16 chapters of PowerPoints are organized around topics including accreditation, good laboratory management, quality management, basic statistics, calibration and detection limits, metrology, measurement uncertainty, control charts, and interlaboratory tests. Altogether, there are about 750 different PowerPoints. As one would expect, there is a CD containing all of them for one's own use or editing. Each of the book's six chapters also includes a bibliography.

The book was created by chemists who were members of Eurachem, a European focus group for analytical chemistry. The editors noted in the Preface, "The aim of this book is to deal with all of these topics in a form that can easily be used for self-training and also for teaching in educational institutions and for in-house training." Their intent was to provide "the basis for presentations by third parties rather than exhaustive and fully comprehensive material." Much of the material has just been assembled in basic slides with titles and bullet points. For the statistical topics, such as calibration, the mathematical detail is such that it is not even readable in the PowerPoints appearing in the book. Some of these slides include results from statistical analyses using Excel.

The three editors are joined by five other chemists as the creators of the PowerPoints. All eight participants are from Germany, Cyprus, Greece, or the U.K. I need to create a basic statistics training program for BP's analytical chemists, and there are certainly some things here that I can use.

Nature-Inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks, edited by Riccardo LEARDI, Amsterdam: Elsevier, 2003, ISBN 0-444-51350-7, xviii + 383 pp., \$240.00.

This book, part of Elsevier's "Data Handling in Science and Technology" series, consists of two main parts and halves, one devoted to each of the two techniques that are the book's topics. Each part begins with a chapter that provides an excellent introduction to the technique. Applications in chemistry for genetic algorithms (GAs) are robust sensor development, molecular modeling, and wavelength selection in spectral datasets. There is also a chapter on hybrid GAs and another that describes MOBYDIGS, a software package for doing regression or classification modeling using GAs. Applications in chemistry for neural networks (NNs) are molecular structure-property studies, calibration for volumetric data in electroanalytical chemistry, prediction of properties and classification of structures in nuclear magnetic resonance, and QSAR modeling for toxicity prediction.

The two introductory chapters begin with the most basic concepts and generally avoid excessive mathematical content, although each has a nonlinear increase in complexity. The book ends with a chapter on actually applying GAs and NNs to chemometric problems. This chapter provides both comparisons between the techniques and guidance on setting the parameters that control each of the techniques. For persons who are involved in chemistry modeling, this would be a good book to own. As always with this publisher, the price tag is high in the U.S., although there is a discount for ASA members. Haupt and Haupt (2004) provided a good resource for an introduction to GAs (see the report in Ziegel 2005). There is no similar book to recommend for NNs.

REFERENCES

- Haupt, R., and Haupt, S. (2004), *Practical Genetic Algorithms* (2nd ed.), Hoboken, NJ: Wiley.
 Ziegel, E. (2005), Editor's Report on *Practical Genetic Algorithms* (2nd ed.), by R. Haupt and S. Haupt, *Technometrics*, forthcoming.

Quantitative Methods for Current Environmental Issues, edited by Clive W. ANDERSON, Vic BARNETT, Philip C. CHATWIN, and Abdel H. EL-SHAARAWI, London: Springer-Verlag, 2002, ISBN 1-85233-294-8, x + 270 pp., \$154.00.

This text results from a joint conference between two environmental organizations; the International Environmetrics Society (TIES) and Statistics in Public Resources, Utilities, and Care of the Environment (SPRUCE). Proceedings of previous SPRUCE conferences have been reported in *Technometrics*, most recently the volume edited by Barnett et al. (1999), which was described by Ziegel (2000). This new volume has papers for 9 of the 11 invited presentations and 3 of more than 70 contributed papers. Section titles for the volume and titles of the papers are as follows:

Spatial and Temporal Models:

1. Modeling Spatio-Temporally Misaligned Areal and Point Process Environmental Data
2. Space and Space-Time Modeling Using Process Convolutions
3. Multivariate Kriging for Interpolating With Data From Different Sources,

Environmental Sampling and Standards:

4. Distance Sampling
5. Setting Standards: A Statistical Approach,

Atmosphere and Ocean:

6. Interpretation and Validation of Measurements in Ocean Wave Directional Spectra
7. Thermal Energy Emission and Propagation From Accidents
8. Methodology to Validate Short-Range Atmosphere Dispersion Models
9. Uncertainty and Sensitivity of Dispersion Model Results to Meteorological Inputs,

Risk and Uncertainty:

10. Approaches to Model Combination

11. Bayesian Analysis of Computer Code Outputs
12. Realities of Decision-Making on Risks.

These presentations mostly do not emphasize any of the mathematical foundations. If any of the titles interest you, track down a library copy or borrow mine for a few days. The book is not priced for individual purchases.

REFERENCES

- Barnett, V., Stein, A., and Turkman, K. (eds.) (1999), *Statistical Aspects of Health and the Environment*, New York: Wiley.
 Ziegel, E. (2000), Editor's Report on *Statistical Aspects of Health and the Environment*, edited by V. Barnett, A. Stein, and K. Turkman, *Technometrics*, 42, 443.

Practical Statistics for Environmental and Biological Scientists, by John TOWNEND, Chichester, U.K.: Wiley, 2002, ISBN 0-471-49664-2, x + 276 pp., \$125.00.

"Practical statistics" for scientists does not necessarily imply simply basic statistics. In fact, this book carries the methodology forward to include principal components analysis and cluster analysis. However, any statistician would certainly characterize the presentation of these topics as very basic. Perhaps that reflects this rather remarkable comment from the author: "There is seldom any necessity to use complex statistical methods to do world-class research in environmental and biological sciences." Curiously, no bibliography or even an affiliation is given for the author. There are very few equations. Most of the equations and all of the computational illustrations are shoved into an appendix. Another appendix tells what one can do with Excel and Minitab, but there are no illustrations.

The book's 15 chapters are organized into two parts, "Statistics Basics" and "Statistical Methods." Highlights for the six chapters on the basics are chapters on designing experiments and surveys and on assumptions for doing statistical tests. Statistical methods start with *t*-tests and *F*-tests. A group of ANOVA chapters includes chapters on MANOVA and repeated measures. There are chapters on correlation and regression, chi-squared tests, and nonparametric tests in addition to the aforementioned chapter on multivariable methods. With a better integration of practical applications and statistical computing, this actually would be a fairly useful book. For a short book that is generally very basic, this book is certainly frightfully expensive.

Internet Companion for Statistics, by Michael D. LARSEN, Belmont, CA: Brooks/Cole, 2004, ISBN 0-534-42356-6, viii + 102 pp., \$19.95.

I will admit to having realized very little value from the statistical content of the Internet. Sure, the websites for the publishers, software vendors, book-sellers, codes for S-PLUS and SAS algorithms, on-line journals, and professional societies are extremely useful to me. However, other than some electronic textbooks that would merit little positive notice in this publication, I do not get information about statistical methods from the Internet. Okay, I have an incredible statistics book library, so do I really need the Internet as a statistics resource?

This book was not written for persons with the particular need that I have described. The author's objective was "to provide educators and students an organized, clear interface to material on the Internet useful for teaching and learning introductory statistics" (Preface, p. v). I create all of BP's statistics training myself, so this particular need is not completely inappropriate for me. The book is "intended as a weekly or more frequent supplement to a longer textbook" (Preface, p. vi). The links in the book mainly take students to sites where there will be problems and data that are appropriate for illustrating and providing practice for the students for a particular topic. The topics are limited to those that appear in the most basic of statistics courses. The material carries the student of basic statistics through hypothesis testing.

There are some references to some of those on-line statistics textbooks that I do not much like. I even found one at a nice website that I had not previously visited. Each of these paperback books comes with a serial number that gives a purchaser access to special material at the publisher's website.

Handbook of Statistics 23: Advances in Survival Analysis, edited by N. BALAKRISHNAN and C. R. RAO, Amsterdam: Elsevier, 2004, ISBN 0-444-50079-0, xxv + 795 pp., \$175.00 (\$145.00 ASA).

Here is my annual report on the latest book in Elsevier's "Handbook of Statistics" series. The second editor here has been the general editor of the series for the past 15 years or so. As with this volume, he has also been a co-editor for many of the handbooks. The first editor has been invited to edit three previous volumes, two volumes on order statistics and most recently a volume on recent advances in reliability (Balakrishnan and Rao 2001). Here they take on the cousin in the world of lifetime statistics, survival analysis, which has come to be associated with living beings.

The editors have organized 40 chapters into 18 categories, which, although perhaps creating my longest list ever, are as follows:

- General Methodology
- Censored Data and Inference
- Truncated Data and Inference
- Hazard Rate Estimation
- Companion of Survival Curves
- Competing Risks and Analysis
- Proportional Hazards Model and Analysis
- Accelerated Models and Analysis
- Frailty Models and Applications
- Models and Applications
- Multivariate Survival Data Analysis
- Recurrent Event Data Analysis
- Current Status Data Analysis
- Disease Progression Analysis
- Gene Expressions and Analysis
- Quality Life Analysis
- Flowgraph Models and Applications
- Repair Models and Analysis.

Eight of the sections, including the last seven, contain just a single chapter.

My ongoing complaint about this series has been that it has been strictly an academic exercise. Perhaps that is actually the design. Here 77 of the 80 contributors come from academia or from government and private research departments. The other three authors work in the pharmaceutical industry in the U.S. or Europe. Two other features of these books always get mentioned. The top-heavy participation from academia results in a book on the advances in the theory and methods for survival analysis. The applications are scarce. On the positive side, these chapters, as always, have extensive reference lists.

If one of the areas in the list interests you, and if you like the academic style of presentation of methodology, then you should probably find a library that has a copy. The publisher has an ASA discount, but the price remains rather steep for individual purchase.

REFERENCE

Balakrishnan, N., and Rao, C. (eds.) (2001), *Handbook of Statistics 20: Advances in Reliability*, Amsterdam: Elsevier.

Statistics for Epidemiology, by Nicholas P. JEWELL, Boca Raton, FL: Chapman & Hall/CRC, 2004, ISBN 1-58488-433-9, xiv + 333 pp., \$69.95.

Multivariate Methods in Epidemiology, by Theodore R. HOLFORD, New York: Oxford University Press, 2002, ISBN 0-19-512440-5, xvii + 408 pp., \$57.95.

Fifteen years ago, when I first had to deal with an epidemiology project, there were no good statistics books for the subject. By now I have reported several that I like a lot, most recently the excellent book by Newman (2001) (reported in Ziegel 2002). In addition, Steven Selvin, author of the first good book on epidemiologic statistics that I encountered (see Ziegel 1995 for a report on the first edition of Selvin 1996) has been listed in the acknowledgments (p. xv) to the Jewell book as a principal colleague and friend.

Jewell's book can certainly be included in any group of useful books on statistics in epidemiology. It actually might be the one with which I would start. It is definitely a first course book for graduate students with some statistics

training. A brief introductory chapter explains the focus on chronic diseases and the analysis of observational data. Chapter 5 focuses on study design (i.e., cohort studies, case-control studies, population-based studies). One of the final two chapters is also concerned with matched studies. Other than a chapter about disease occurrences, this book is about data analysis and specifically methods for binary outcome data, that is, the person survives or does not survive. There also is the presumption that the predictors do not vary over time, so the book does not consider situations that result in longitudinal data. The book then is really a book about binary data analysis. There is a standard book on binary data by Collett (2003), the section edition of which was reported by Ziegel (2004).

Jewell's book begins with introductory data analysis chapters on probability and measures of disease exposure. The next six chapters present the topics related to analysis of data tabulations: 2×2 tables, measures of association, confounding from extraneous factors, control of extraneous factors, interaction, and exposures at several discrete levels. Then come four chapters on logistic regression topics: relating exposure to disease, parameter estimation, confounding and interaction, and goodness-of-fit tests. The last couple of chapters discuss alternatives and extensions to the logistic regression model. The book uses no calculus and presents no proofs. There also is no demonstration of the use of any software.

Holford's book, which seemed to promise an expansion of the tool kit with methods for multivariable data, is unfortunately most noteworthy because of its inappropriate title. By convention, "multivariable" usually refers to situations where a set of measured results occurs for each observation. Then the multivariate normal distribution is the principal reference distribution, and tools such as principal components analysis are used for data analysis. Here the multivariate aspect is actually the multiple number of predictor variables or attributes that might impact the result or response.

Generally, books on statistics in epidemiology are written as much for epidemiologists as they are for statisticians. Any epidemiologists here should have had a good first course in statistical inference. For the statistician, the blend of theory and application plus the excellent selection of methods and problems should provide a good source for learning about the use of statistical tools in epidemiology. The first two chapters in the book's "Concepts and Definitions" part give background for the types of data and models for relating diseases to exposure encountered in epidemiology. The second part, "Non-Regression Methods," contains data analysis chapters for proportions, rates, and time to failures. The remainder of the book is devoted to seven chapters on regression methods. The first two of these chapters deal with models for proportions; the next two consider models for proportional hazard regression. There also are chapters on models for various types of matched studies, sample size requirements to achieve the necessary power for a study, and extensions of the regression modeling approach. A final chapter covers classification and regression trees, splines, generalized additive models, multivariate adaptive recursive splines, and variance component models. There are also discussions on missing data, error in variables, and collinearity.

Holford's book has 60 pages in 8 appendices devoted to additional theory. Because the text already has methodology that uses matrices and calculus, integration of this material would have been appropriate. However, those of us who skip over this material anyway should be happy with the omission.

REFERENCES

- Collett, D. (2003), *Modelling Binary Data* (2nd ed.), Boca Raton, FL: Chapman & Hall/CRC.
- Newman, S. (2001), *Biostatistical Methods in Epidemiology*, New York: Wiley.
- Selvin, S. (1996), *Statistical Analysis of Epidemiologic Data* (2nd ed.), New York: Oxford University Press.
- Ziegel, E. (1995), Editor's Report on *Statistical Analysis of Epidemiologic Data* by S. Selvin, *Technometrics*, 37, 240-241.
- (2002), Editor's Report on *Biostatistical Methods in Epidemiology*, by S. Newman, *Technometrics*, 44, 301-302.
- (2004), Editor's Report on *Modelling Binary Data* (2nd ed.), by D. Collett, *Technometrics*, 46, 119-120.

Statistical Estimation of Epidemiological Risk, by Kung-Jong LUI, Chichester, U.K.: Wiley, 2004, ISBN 0-470-85071-X, xvi + 193 pp., \$95.00.

This is yet another book in Wiley's "Statistics in Practice" series. These modest-sized volumes are intended to be practical and example-oriented intro-

ductions to a particular field of study. The determination of epidemiologic risk occurs within a petrochemical company where concerns about the impact of chemicals on employees have been raised. The government primarily handles the situation where the concern is the public or the environment.

This book deals with estimating risk, a concept by now very familiar to BP management, for whom it is a common feature in decisions concerning most financial matters. The first chapter deals with sampling to determine prevalence, including inverse sampling, which targets a predetermined number of cases. The second chapter provides estimators for the risk differences between normal and affected populations. Relative difference and risk estimators are provided in the next two chapters. There is also a chapter on attributable risk, the proportion of cases that would not occur if the specific risk factor were absent. Linkage to clinical trials is provided from chapters on the odds ratio, the generalized odds ratio, and the number needed to treat.

Despite the practical intent, this book has the feel of a more conventional statistics book. There is a lot of discussion and even a certain amount of derivation built on extensive presentation by algebraic formulas. Examples are generally perfunctory. There is not a lot of actual data from case studies and no effort to use or discuss computer software for problem solutions.

Disease Mapping With WinBUGS and MLwiN, by Andrew B. LAWSON, William J. BROWNE, and Carmen L. VIDAL RODEIRO, West Sussex, U.K.: Wiley, 2003, ISBN 0-470-85604-1, xiii + 277 pp., \$85.00.

Bayesian Approaches to Clinical Trials and Health Care Evaluation, by David J. SPIEGELHALTER, Keith R. ABRAMS, and Jonathan P. MYLES, West Sussex, U.K.: Wiley, 2004, ISBN 0-471-49975-7, xiv + 391 pp., \$85.00.

These two books are from Wiley's practical applications series, "Statistics in Practice." The books in this series are published in the U.K. and generally have authors from the U.K., although that applies only to the middle author for the book by Lawson et al. Generally the books have focused on health and environmental sciences, which is certainly indicative of where technology, research, and jobs are focused now. Together, the books demonstrate that Bayesian methodology is being used by practitioners in the important area of human health. Both books are written at a level that will be readily accessible to anyone with a modest background in applied statistics.

The book by Lawson et al. is actually specifically focused on modeling for spatial health data and applications concerned with the counted data that results from recorded disease occurrences in small areas. The first chapter gives an introduction to the practice of disease mapping, and the second chapter performs the same function for Bayesian hierarchical modeling. Next multilevel modeling is introduced as a useful tool for accounting for small area differences when sampling occurs from multiple areas, and the appropriate Bayesian methods are presented. The next two chapters deal separately with the use of WinBUGS, a computer program for Bayesian inference using Gibbs sampling, and MLwiN, a computer program for multilevel modeling. Both packages use Markov chain Monte Carlo estimation.

There follow four chapters that apply both WinBUGS and MLwiN to the same data examples. Topical areas for the applications are relative risk estimation, focused clustering applied to the analysis of putative health hazards, ecologic analysis, and spatially correlated survival analysis. The software illustrations are very explicit about both the required input and the display of the output. The book uses a large number of geographic displays to present the results of the data analysis.

The first chapter of the book by Spiegelhalter et al. notes that "advances in health-care typically happen through incremental gains in knowledge, rather than paradigm-shifting breakthroughs" (p. 2). Bayesian methods would seem particularly appropriate for this situation. The book devotes an entire chapter (Chap. 5) to the difficult issue of prior distributions, which have a subjective component that is not necessarily consistent with the need for regulatory standards. The book is specifically about health care interventions, any step that is intended to improve health. A lot of the focus is on conventional randomized controlled trials (RCTs). Chapters 2 and 3 give overviews of classical statistical concepts and the Bayesian approach. Classical and Bayesian approaches are compared in the following chapter. Analyses for RCTs are discussed extensively in Chapter 6. Observational studies and evidence synthesis are discussed in the next two chapters. The final chapter covers cost-effectiveness, policy-making, and regulation. The book is generally discursive

with little direct guidance provided for doing data analysis and modeling. There are no discussions of survival analysis or longitudinal data. An appendix lists websites for software packages.

Design and Analysis of Clinical Trials (2nd ed.), by Shein-Chung CHOW and Jen-Pei LIU, Hoboken, NJ: Wiley, 2004, ISBN 0-471-24985-8, xiii + 729 pp., \$130.00.

Design and Analysis of DNA Microarray Investigations, by Richard M. SIMON, Edward L. KORN, Lisa M. MCSHANE, Michael D. RADMACHER, George W. WRIGHT, and Yingdong ZHAO, New York: Springer-Verlag, 2004, ISBN 0-387-00135-2, x + 199 pp., \$59.95.

Design of experiments has always been a special feature of *Technometrics* and a popular topic with industrial statisticians. These designs, of course, are not the conventional fare of *Technometrics*, but they do represent a significant portion of the types of designs being done in industry as well as the forefront of research in industrial experimental design.

Technometrics has reported occasional books on clinical trials over the last decade, most recently a report by Ziegel (2004) of the book by Moyé (2003). The first edition (1E) of Chow and Liu's book somehow passed under the editorial radar. The only other comprehensive design and analysis book for clinical trials that I have is the classic book by Fleiss (1986), which served me well when Amoco Research ran some clinical trials at Mayo Clinic years ago. Chow and Liu's book, subtitled "Concepts and Methodologies," is twice the size of the Fleiss book and clearly the most comprehensive book of its type. Part of the book's long length comes from its dual intended audience of both clinical scientists and biostatisticians. A 40-page introductory section begins with a chapter titled "What Are Clinical Trials" that explains clinical practices for the statisticians. Two subsequent chapters of similar size present basic statistics and general design concepts for the clinicians. After a chapter on randomization and blinding come three chapters on the actual design of clinical trials, including a new chapter on designs specifically for trials related to cancers.

The focus turns to data analysis in the last half of Chow and Liu's book, including analysis of continuous data, categorical data, and censored data. After a chapter on sample size determination, the book concludes with four chapters on clinical practices: efficacy evaluations, safety assessments, clinical protocols, and data management. The last two topics are covered by relatively short chapters that are new to this edition. The authors provide a nice list of new additions to the original 1E chapters in their Preface. These additions include clinical topics (e.g., clinical significance, target populations, and efficacy evaluations with genomics data) and biostatistics topics (e.g., group-sequential designs, clustered randomized designs, and dose-response trials). The book is certainly comprehensive and should be a standard reference for both clinical scientists and biostatisticians, although the statistical level is more appropriate for the latter group. There is no demonstration of the use of any statistical software.

Unlike clinical trials, microarray studies constitute a very new methodology in life-sciences research. A couple of reports have been published on books on statistical methods for microarray studies; see, for example, the report on Drăghici (2003) and McLachlan et al. (2004), which appears in the issue with this report (Ziegel 2005). That report notes that the book by McLachlan et al. would be a good introduction to the more comprehensive book by Drăghici. The slender volume by Simon et al. would actually be an even better book with which to begin. It has almost no focus on mathematics and equations. Like both of the aforementioned microarray books, it focuses simultaneously on teaching statisticians about microarrays and teaching biologists about statistics. The introductory chapter on microarrays and an appendix on gene expression studies provide the statisticians' background in just 20 pages. The experimental design component of the book is also relatively brief, a chapter giving an overview of different types of study objectives followed by an illustration of a design type for each situation. Three subsequent chapters, all relatively brief, present material devoted to topics concerned with obtaining good data: image analysis, quality control, and array normalization. The brevity might belie the fact that these are critical aspects of creating data worthy of sophisticated analyses.

The remaining three chapters, composing more than half of the book, present data analysis strategies for gene expression data. The first chapter is concerned with comparing different prespecified classes for the genes using various methods which run from *t*-tests to regression models. The second chapter deals with

class prediction, especially prognostic prediction. These data analysis methods apply to gene expression profiles for diseased tissues. The last chapter discusses class discovery, for which graphical displays and clustering algorithms provide the primary tools. Additional features of this fine book are 16 pages of color plates illustrating all aspects of the microarray analysis process, from the robotics that create the arrays to the profile plots that display the cluster results. An appendix lists software packages for the three types of data analysis.

REFERENCES

- Drăghici, S. (2003), *Data Analysis Tools for DNA Microarrays*, Boca Raton, FL: Chapman & Hall/CRC.
- Fleiss, J. (1986), *The Design and Analysis of Clinical Experiments*, New York: Wiley.
- McLachlan, G., Do, K., and Ambrose, C. (2004), *Analyzing Microarray Gene Expression Data*, Hoboken, NJ: Wiley.
- Moyé, L. (2003), *Multiple Analyses in Clinical Trials*, New York: Springer-Verlag.
- Ziegel, E. (2004), Editor's Report on *Multiple Analyses in Clinical Trials*, by L. Moyé, *Technometrics*, 46, 373.
- (2005), Editor's Report on *Data Analysis Tools for DNA Microarrays*, by S. Drăghici, and *Analyzing Microarray Gene Expression Data*, by G. McLachlan, K. Do, and C. Ambrose, *Technometrics*, 47, 109.

Data Analysis Tools for DNA Microarrays, by Sorin DRĂGHICI, Boca Raton, FL: Chapman & Hall/CRC, 2003, ISBN 1-58488-315-4, xxx + 477 pp. + CDs, \$79.95.

Analyzing Microarray Gene Expression Data, by Geoffrey J. McLachlan, Kim-Anh Do, and Christophe Ambrose, Hoboken, NJ: Wiley, 2004, ISBN 0-471-22616-5, xx + 320 pp., \$89.95.

Microarray data analysis is one of the hot new areas in statistics. Last year I gathered up a couple of recent arrivals to the textbook scene (Parmigiani, Garrett, Irizarry, and Zeger 2003; Speed 2003), and cobbled together a brief report (Ziegel 2003). Here, probably not much more than 1 year later, I have already collected two more of these books. One of these, the book by Drăghici, is already the second such book from its publisher. Obviously this is an area of statistical analysis that should be familiar to any statistician working in the sciences.

I really like Drăghici's book. As the author explains in the Preface, the book is intended to serve both the statistician who knows very little about DNA microarrays (which applies to me) and the biologist who has no expertise in data analysis. The author lays out a study plan for the statistician that excludes 5 of the 17 chapters (4–8). These chapters present the basics of statistical distributions, estimation, hypothesis testing, ANOVA, and experimental design. Actually, statisticians should also study the chapter that applies inference to microarrays and likewise the similar aspects of the chapter on experimental design. What that leaves for the statistician is the three-chapter primer on microarrays and image processing, plus all of the data analysis tools specific to the microarray situation. These include corrections for simultaneous inference, catalogues of exploratory and multivariable statistical tools, data preparation issues, selection of differentially regulated genes in comparative experiments, interpretation of data analysis for hypothesizing biological phenomena, and determination of specific microarrays for biological hypotheses. A separate chapter at the end of the book discusses and displays analyses using software available from Insightful and SAS, plus the data capabilities of Spotfire. The biology and statistics are kept at a good level for the intended audience. The softcover book is reasonably priced, and it includes two CDs with trial versions of several specialized software packages. Anyone who uses microarray data should certainly own a copy.

Coincidentally, the book by McLachlan et al. states exactly the same dual objective concerning statisticians and biologists in its Preface, where the statisticians are labeled “biostatisticians.” Here, however, the introduction to gene expression studies with microarrays is considerably briefer—just a single 30-page chapter. There is no introduction to statistical methods. Following a chapter on cleaning and normalizing for the data (a central component of data analysis for microarray data), the book plunges into the first of three sizeable chapters on cluster analysis. Together these chapters encompass about half of the book. The

authors focus their attention primarily on model-based methods. They particularly emphasize a normal mixture-based method called the EMMIX-GENE procedure. They determine the number of components in the mixture model by resampling. Modern technology also prevails for discriminant analysis, where they feature the use of support vector machines. All of these advanced technologies come along with enough methodology to make this book a challenge for the biologists. Application chapters demonstrate the different data analysis methodologies. I liked this book as well. It is the more advanced of the two books, but it also expands on both the tools and the applications of microarray data analysis that are demonstrated in Drăghici's book.

REFERENCES

- Parmigiani, G., Garrett, E., Irizarry, R., and Zeger, S. (eds.) (2003), *The Analysis of Gene Expression Data*, New York: Springer-Verlag.
- Speed, S. (ed.) (2003), *Statistical Analysis of Gene Expression Microarray Data*, Boca Raton, FL: Chapman & Hall/CRC.
- Ziegel, E. (2003), Editor's Report on *The Analysis of Gene Expression Data*, edited by G. Parmigiani, E. Garrett, R. Irizarry, and S. Zeger, and *Statistical Analysis of Gene Expression Microarray Data*, edited by S. Speed, *Technometrics*, 45, 375.

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).

Analysis of Variance for Random Models, Volume II, Unbalanced Data, by Hardeo Sahai and Mario Miguel Ojeda, Springer-Verlag

Applied Bayesian Modeling and Causal Inference From Incomplete-Data Perspectives, edited by Andrew Gelman and Xiao-Li Meng, Wiley

Applied Longitudinal Data Analysis for Epidemiology, by Jos W. R. Twisk, Cambridge University Press

Applied Statistics for Engineers and Scientists (2nd ed.), by Jay Devore and Nicholas Farnum, Thomson Brooks/Cole

Aspects of Multivariate Statistical Analysis in Geology, by Richard A. Reymont and Enrico Savazzi, Elsevier

A Beginner's Guide to Structural Equation Modeling, by Randall E. Schumacker and Richard G. Lomax, Lawrence Erlbaum

Biostatistics: How It Works, by Steve Selvin, Upper Saddle River, NJ

Bootstrap Techniques for Signal Processing, by Abdelhak M. Zoubir and D. Robert Iskander, Cambridge University Press

Chances Are: A Primer on Probability, by Susan F. Wagner, Pearson Prentice Hall

Code Complete (2nd ed.), by Steve McConnell, Microsoft Press

Design and Analysis: A Researcher's Handbook (4th ed.), by Geoffrey Keppel and Thomas D. Wickens, Pearson Prentice Hall

Design of Experiments With Minitab, by Paul Mathews, ASQ Quality Press

Evaluation of Mineral Reserves, by André G. Journel and Phaedon C. Kyriakidis, Oxford University Press

Flowgraph Models for Multistate Time-to-Event Data, by Aparna V. Huzurbazar, Wiley

Generalized Least Squares, by Takeaki Kariya and Hiroshi Kurata, Wiley

Geostatistical Reservoir Modeling, by Clayton V. Deutsch, Oxford University Press

Introduction to Applied Econometrics, by Kenneth G. Stewart, Thomson Brooks/Cole

Introduction to Probability and Statistics for Engineers and Scientists (3rd ed.), by Sheldon M. Ross, Elsevier

Longitudinal and Panel Data, by Edward W. Frees, Cambridge University Press

Nondetects and Data Analysis, by Dennis R. Helsel, Wiley

Nonlinear Signal Processing: A Statistical Approach, by Gonzalo R. Arce, Wiley

Numerical Methods for Nonlinear Estimating Equations, by Christopher G. Small and Jinfang Wang, Oxford University Press

Permutation, Parametric, and Bootstrap Tests of Hypotheses (3rd ed.), by Phillip Good, Springer-Verlag

Practical Business Forecasting, by Michael K. Evans, Blackwell Publishers

Probability and Random Processes, by Scott L. Miller and Donald G. Childers, Elsevier

Procrustes Problems, by John C. Gower and Garmt B. Dijkstrahuis, Oxford University Press

Regression With Social Data: Modeling Continuous and Limited Response Variables, by Alfred DeMaris, Wiley

Statistical Analysis of Stochastic Processes in Time, by J. K. Lindsey, Cambridge University Press

Statistical Methods in Analytical Chemistry, by Peter C. Meier and Richard E. Zünd, Wiley

Statistics of Extremes: Theory and Applications, by Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef Teugels, Wiley

Statistics for Petroleum Engineers and Geoscientists (2nd ed.), by Jerry L. Jensen, Larry W. Lake, Patrick W. M. Corbett, and David J. Goggin, Elsevier

Stereology for Statisticians, by Adrian Baddeley and Eva B. Vedel Jensen, Chapman & Hall/CRC

Structural Equation Modeling, by Structural Equation Modeling, Cambridge University Press

Survival Analysis Using S, by Mara Tableman and Jong Sung Kim, with a contribution from Stephen Portnoy, Chapman & Hall/CRC

Time Series Analysis and Inverse Theory for Geophysicists, by David Gubbins, Cambridge University Press