Book Reviews

Publishers are invited to send books for review to Dr. Paul I. Feder, Applied Statistics Group, Battelle Memorial Institute, 505 King Avenue, Columbus, OH 43201. Please include the price of the book.

The opinions expressed in this section are those of the reviewers and do not necessarily reflect those of the editorial staff or the sponsoring societies.

Statistical Analysis: A Computer Oriented Approach (2nd Ed.) by A. A. Afifi and S. P. Azen, Academic Press, 1979, xx + 442 pp.

The first edition of this book was published in 1972. This second edition is not greatly different from the first. The two editions are divided into 5 chapters and a 50-page appendix entitled, "Review of Fundamental Concepts." The five chapter titles are the same for the two editions and are 1. "Introduction to Data Analysis," 2. "Elementary Statistical Inference," 3. "Regression and Correlation Analysis," 4. "The Analysis of Variance," and 5. "Multivariate Statistical Methods."

The second edition is 76 pages longer than the first. This is primarily due to the addition of the following sections: in Chapter 1, "Data Screening," in Chapter 2, "Other Measures of Association for Contingency Tables" and "Robust Estimators," and in Chapter 5, "The Multivariate Analysis of Variance;" in Chapter 3 the subsection "Multivariate Missing Observations" is added. A number of new examples and problems have been added. All chapters have a set of problems now whereas Chapter 1 did not in the first edition. No answers are provided for the problems. Most of the examples and problems are from the fields of biology and medicine and unless one is fairly familiar with these subject areas, the examples and problems will not be very meaningful.

In the preface to the first edition the authors indicate that the "book contains both elementary and advanced topics" and has "a wide range of material." In the preface to the second edition they indicate that they have attempted "to broaden the scope of the book." The breadth is perhaps best illustrated by the courses that the authors suggest might be taught from the book; namely, "Elementary Applied Statistical Analysis" (undergraduate), "Applied Statistical Analysis" (first-year graduate), "Applied Multivariate Analysis" (second-year graduate), and "Intensive Course in Data Analysis." I feel that books aimed at a specific well-defined audience are generally more effective than those aimed at a wide general audience, and this book only confirms that opinion.

I tried to put myself in the place of a person who has had only an "elementary course in the fundamentals of statistical inference" and college algebra—the minimum background that the authors feel is required to use the book. I think people with that background would find the book very difficult if they were on their own.

In Chapter 1 the authors say, "... there are two main objectives of this book. The first is to present in a practical manner the fundamental techniques of classical statistical analysis—both univariate and multivariate." They have done a good job of presenting the techniques of classical statistics. However, I do not feel that they have done a good job of interpreting their examples. Most of the interpretations of the analyses consist of little more than saying whether a statistical test is significant—with a P value—or not.

In the analysis of variance (ANOVA) chapter the authors introduce a study (give a literature reference) in which, "The data consisted of measurements Y = the amount of nitrogen expired (in liters) in resting conditions under four dietary regimens." They indicate that, "This example will be analyzed in many different ways throughout the chapter." When they use the study to illustrate the one-way ANOVA, they present some data. This study and these data (or a subset) are then used many times as examples in the chapter. I think real experimental studies and data lose their effectiveness when used this way. One might better, I feel, generate data—and admit it—and do the generation using the model to be illustrated. When a single set of data is used to illustrate successively more complex studies and ends up illustrating the situation for the real experiment that produced the data, it is sometimes effective in that it can show why the real study was set up as it was. However, in the present case it is never indicated what the design was for the study referenced.

There is a brief discussion of power in the appendix. However, I find it strange that there is nothing on power in the body of the text.

At least I do not recall any mention of power and there is no reference to power in the index.

The authors say that "we have updated our first edition by presenting features of the latest versions of the statistical packages BMD-P, SPSS, and SAS. Additional packages such as GLIM and MINITAB are also discussed." In the book there is a two-paragraph (21 lines) discussion of BMD-P, a one-paragraph (12 lines) discussion of SPSS, and a one-paragraph (10 lines) discussion of MINITAB. SAS and GLIM are included in a list of the names of 12 packages, which also lists the names of the developers of each package. Neither SAS, GLIM, MINITAB, nor SPSS are referenced in the index. At a few places in the text there are brief indications of the differences in output among some of the packages. On the basis of what the authors said, I expected to find more than this.

I do not find this book, even though it is "a computer oriented approach," all that different from the statistical analysis books that have been published over the past 10–15 years. The only real difference is that when it comes time to analyze a set of data in this book there is a statement such as, "From a packaged one-way analysis of covariance table, we obtained Table B" (p. 267). Much of the time the package used is not indicated. I feel that the computer has added a dimension to data analysis that this does not reflect.

In summary, it seems to me that the authors promise more than they deliver. By trying to appeal to a wide audience, I think the book is less effective than it might have been if it had been aimed at a more select audience. Being rather familiar with the book now, I expect that I may use it as a reference; however, it is unlikely that I will recommend it to anyone either as a reference or as a textbook.

Phelps P. Crump
USAF School of Aerospace Medicine

The Theory of Linear Models and Multivariate Analysis, by Steven F. Arnold, John Wiley, 1981, xv + 475 pp., $34.95.

This book, which consists of 22 chapters and an appendix, deals with certain aspects of linear models and multivariate statistical analysis. The author uses a coordinate-free approach to the topics covered. The prerequisites are a thorough grounding in vector space algebra, matrix theory, and mathematical statistics at the intermediate level.

The chapter titles are: 1. "Basic Statistical Definitions and Theorems," 2. "Subspace and Projections," 3. "Properties of the Multivariate and Spherical Normal Distributions," 4. "Introduction to Linear Models," 5. "A Sufficient Statistic," 6. "Estimation," 7. "Tests about the Mean," 8. "Simultaneous Confidence Intervals—Scheffe's Type," 9. "Tests about the Variance," 10. "Asymptotic Validity of Procedures under Nonnormal Distributions," 11. "James-Stein and Ridge Estimators," 12. "Inference based on the Studentized Range Distribution and Bonferroni's Inequality," 13. "The Generalized Linear Model," 14. "The Repeated Measures Model," 15. "Random Effects and Mixed Models," 16. "The Correlation Model," 17. "The Distribution Theory for Multivariate Analysis," 18. "The Multivariate One- and Two-Sample Models—Inference about the Mean Vector," 19. "The Multivariate Linear Model," 20. "Discriminant Analysis," 21. "Testing Hypotheses about the Covariance Matrix," 22. "Simplifying the Structure of the Covariance Matrix."

The author introduces the basic statistical concepts in Chapter 1. Chapters 2 and 3 discuss properties of projections and the multivariate normal distribution. In Chapter 4–16 linear models are studied. Chapters 17–22 discuss multivariate analysis. The appendix is devoted to matrix algebra. A good set of exercises is provided at the end of each chapter. The author attempts to unify ideas underlying regression theory and multivariate analysis. The differences between a linear model and a regression model are clearly described. Chapter 10 on the asymptotic validity of procedures under nonnormality and Chapter 11 on James-Stein and ridge estimators form an outstanding feature of the book. In multivariate analysis some important topics like factor analysis and multivariate analysis of covariance are omitted. Perhaps this is because it is very difficult to cover linear models and multivariate analysis in a single book and yet limit it to a reasonable size.

This is a theoretical book with no mention of analysis of data. Printing is far from perfect and some pages are hard to read. There are many typographical errors, for example, page 11 line 8 from the bottom: (θ) should read $\bar{g}(\theta)$; page 37 line 9: $u = \Sigma$ should read $P_v u = \Sigma$; page 66 line 10: $T'T$ should read $\tilde{T}'\tilde{T}$; page 70 line 13: + should be $-$; page 76 exercise #2: $(n-p)\hat{\sigma}^2$ should be $(n-p)\hat{\sigma}^2/\sigma^2$; page 113 line 5: β should be capitalized, and so on.

In spite of these drawbacks, the presentation of the material covered in the book is quite clear. Certainly the book can be recommended to those who share the author's mathematical point of view.

A. K. Gupta
Bowling Green State University

Censoring and Stochastic Integrals, by R. D. Gill, Mathematisch Centrum, 1980, v + 178 pp., price not given.

This tract is a marvelous success in presenting a well-organized, coherent, comprehensive, up-to-date account of the theory of censoring and stochastic integrals.

Censored data, or incomplete observations, arise in various situations such as industrial life-testing, clinical trials, and biological experiments. In industrial life-testing, for example, one is interested in the distribution of the lifetimes of components on test. However, components may suffer removal or withdrawal from further observation owing to an extraneous cause. The author treats the statistical problem of censored data in terms of the theory of stochastic processes.

This tract consists of six chapters and appendices. Chapter 1 is a summary of notation and conventions. Chapter 2 contains results from the theory of stochastic integrals, counting processes, and weak convergence of processes. Chapter 3 begins with nice practical examples that illustrate how censored data from "right censorship" can arise. The product-limit estimator (PLE), an estimator of the cumulative distribution function based on censored observations introduced by Kaplan and Meier (1958), is defined in this chapter in terms of counting processes. The two-sample test statistics of Gehan (1965), Efron (1967), and Cox (1972) are considered as members of a general class of test statistics, and some of the small-sample properties of the PLE and of the test statistics are derived in the latter part of Chapter 3. In Chapter 4 asymptotic results for these statistics are given. In Chapter 5 the author examines efficiencies when testing against specific alternatives. Chapter 6 is a brief description of extensions to the preceding theory. More general types of censorship than the "right censorship" are discussed. The remainder of the tract consists of appendices that provide proofs for results that appear in Chapter 2. This tract is not completely free of misprints, minor errors, and confusing notation. Most sections include helpful intuitive explanations, although the author's approach is mainly theorem-proof oriented. The author's style is terse so that he may present much material. Reading this tract may be a demanding experience for some readers, but those with a good theoretical background will find it abundantly rewarding.

Jee Soo Kim
Florida State University

REFERENCES

COX, D. R. (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society*, Ser. B. 34, 187–202.

EFRON, B. (1967), "The Two Sample Problem With Censored Data," *Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics, and Probability*, 4, 831–853.

GEHAN, E. A. (1965), "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly—Censored Samples," *Biometrika*, 52, 203–223.

KAPLAN, E. L., and MEIER, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.

The Statistical Analysis of Failure Time Data, by J. D. Kalbfleisch and R. L. Prentice, John Wiley, 1980, xi + 321 pp., $24.95.

This text is designed to "collect and unify some statistical models and methods that have been proposed for analyzing failure time data." The authors have done an excellent job of collecting and illustrating many of the recent advances in the area of analysis of survival data. Much of the book deals with the regression problem with failure data, in particular with the proportional hazards rate model. The emphasis is on models and procedures used in the biomedical sciences, though many of these techniques are also useful in the engineering sciences.

The orientation of this book is data-analytic with a rigorous theoretical treatment of the models and methods. Five data sets are described in an appendix and are reanalyzed by the various methods described throughout the text. Another appendix gives sample FORTRAN programs for analyzing data using the proportional hazards rate model with either fixed covariates or time dependent covariates.

Chapter 1 introduces the problem of analyzing failure distributions by introducing several examples. It also introduces the Kaplan-Meier estimator and two-sample log rank test. Chapter 2 discusses the major parametric models in survival analysis. Discussed are the exponential, Weibull, log-normal, gamma, generalized gamma, log-logistic, and generalized F distributions. Regression models, proportional hazards rate models, and accelerated failure-time models are also described.

Chapter 3 deals with making inferences about the parametric models described in Chapter 2. After a brief description of censoring mechanisms, the general form of the likelihood is derived for independent censoring mechanisms. The problem of estimating parameters of the exponential distribution is considered for type-I and type-II censoring. Large-sample likelihood theory is described in detail and illustrated in the case of exponentially distributed failures. The Newton-Raphson method is discussed and is used to illustrate estimation of parameters in log-linear regression models using data from the appendix. The problem of discrimination between parametric models is handled by estimating parameters of the generalized F distribution.

Chapter 4 is the real meat of the book. This chapter discusses estimation in the proportional hazards rate model in detail. Both the marginal-likelihood and the partial-likelihood approaches to estimation are discussed in detail. Approximations in the event of ties is handled nicely. The data analyzed in the previous chapter using a Weibull log linear model are reanalyzed using Cox's model. Graphical checks on the validity of the proportional hazards rate model are proposed and estimation in discrete time is discussed. The estimation procedure is extended to allow the inclusion of strata. A very complete discussion of the properties of the estimators is given. Included are discussions of the efficiency and asympotic distributional properties of the estimators.

The rest of the book deals with a variety of special topics. In Chapter 5 the idea of noninformative censoring and the resulting likelihood construction is considered. The problem of time-dependent covariates is discussed and the solution is illustrated using one of the data sets in the appendix. The efficiency of the regression estimator is discussed with particular attention to comparing the proportional hazards model to the Weibull.

Chapter 6 deals with rank statistics in censored data experiments. Linear rank tests in the two-sample problem are described and illustrated. The general theory of rank scores in the general regression model is considered in detail.

Chapter 7 deals with competing-risks models. The approach is to model the overall hazard rate as the sum of cause specific hazard functions which are directly observable. Relationships between the competing risks can be studied through the use of time dependent covariates. A brief section on the latent failure time approach is included. The chapter concludes with a discussion of some very sophisticated models for modeling multivariate failure times with competing risks.

Chapter 8 deals with a variety of miscellaneous topics including the analysis of paired failure time data, fixed study and bioassay problems, analysis of retrospective studies, and a Bayesian approach to the analysis of the proportional hazards model. The chapter concludes with a detailed analysis of one of the data sets in the appendix.

In general this book gives a very thorough, data-analytic account of the modeling and analysis of survival data with covariates. Most results are theoretically justified. An appendix of 67 problems is given, though few are data oriented.

Several minor criticisms may be made of the book. First, the orientation of the book is on the proportional-hazards-rate model with noninformative censoring. Other approaches to analyzing failure time data are somewhat neglected. The proportional-hazards-rate model is proposed as the solution to almost all failure time data problems. In practice this may not be the case. The proportional hazards rate assumption of factorability of the hazard as a function of time alone and another function of the covariates alone is a very strong distributional assumption. The user of these techniques needs to be wary of being misled by making such an assumption.

Secondly, throughout the book the assumption of an independent or noninformative censoring mechanism is made. In many applications patients may remove themselves from the test prior to the event of interest for reasons related to the treatment under study. In such cases an investigator may be seriously misled by assuming the censoring mechanism is noninformative.

Thirdly, the notion of latent failure times in competing risks models is given very little discussion. In modeling competing risks in engineering applications and, perhaps, in studying the effects of noninformative censoring, these may be useful tools.

Lastly, the chapter on parametric models seemed too oriented to modeling the use of covariates in the analysis. Little mention of the notions of hazard or probability plotting were given. Estimation focused entirely on the maximum likelihood method and little was done on the small-sample properties of these estimators or other estimation techniques.

On the whole these criticisms are minor in light of the contribution of this book as a reference book on analyzing survival data with covariates. This book should soon become a standard reference book for any statistician who analyzes survival data and a standard text for advanced courses in biostatistics.

John P. Klein
Ohio State University

Multivariate Analysis (2nd Ed.), by M. Kendall, Macmillan Publishing Company, 1980, vii + 210 pp., $29.95.

This second edition is basically identical to the first edition, which was published in 1975. The book is intended mainly for the

practitioner. It covers principal components, classification and clustering, factor analysis, canonical correlations, some distribution theory, regression analysis, functional relationship, tests of hypothesis, discrimination, and categorized multivariate data. An appendix describes briefly multidimensional scaling.

The book is relatively easy to read and includes many numerical examples. Even though some of the recent developments such as growth curves are not included, the book is recommended for people interested in multivariate methods.

Jack C. Lee
Bell Laboratories

Probabilistic Models in Engineering Sciences, Vols. I and II, by H. J. Larson and B. O. Shubert, John Wiley, 1979, Vol. 1, x + 544 pp., $32.95, Vol. 2, x + 737 pp., $36.95.

What should statisticians know about probability? Currently there seems to be a wide diversity of answers to this question among departments of statistics. Statistical inference problems traditionally start with f(sample | parameter), the probability density function or probability mass function for a finite number of random variables (representing the observed sample) as a function of a finite-dimensional parameter to be estimated. Statisticians should be aware of the emerging discipline of "functional inference," in which the sample and/or parameter is a function (such as a function on the unit interval). The theories of time series analysis and stochastic processes study probability densities for samples that are functions. The theories of nonparametric density estimation and nonparametric regression are examples of fields that study parameters that are functions.

Probability theory may be said to have two distinct applied roles, which we call applied probability and mathematical statistics. Applied probability is studied by applied scientists (such as operation researchers and systems electrical engineers) who seek probability models for the process they study. Mathematical statistics uses probability to develop the small- and large-sample properties of the statistical inference and data analysis methods used by statisticians.

How much of the foregoing is learned by the median new statistics Ph.D. (to say nothing of the new statistics M.S. student) I leave to the conscience of my academic colleagues. That much can be learned (and applied to develop important applied theory) without overloading with extraneous mathematics was eloquently pointed out in the preface of J. L. Doob's classic treatise *Stochastic Processes*. There, Doob makes the celebrated remark that the advanced theorems of probability are understood by many without the mathematical background to understand their proofs.

The two-volume introductory probability text by Larson and Shubert has as its aim to provide a text in applied probability that enables applied scientists to understand and apply it using appropriate levels of rigor. It emphasizes electrical-engineering applications. Its distinguishing features are (a) random variables and stochastic processes are introduced simultaneously; (b) it is an extremely carefully written book that discusses basic concepts in great detail. Consequently, it is a commendable book that should be seriously considered for adoption as a text in a course in applied probability for engineers. It requires only a calculus background and is suitable for masters-level graduate students.

Volume I discusses (in 500 pages) basic probability models for random variables and stochastic processes, expectation, central limit theorem, and ergodicity. The concepts of martingale, stationarity, correlation function, and Markov property are introduced.

Volume II contains 28 sections in 6 chapters, which can be combined in diverse ways to form an introductory course in stochastic processes even for students who have not studied probability theory from Volume I. The main topics discussed are discrete time processes, decomposition, and prediction; Markov chains; continuous time processes and mean square calculus; spectral representation of stationary processes; Markovian diffusion processes; and point processes.

These books are about applied probability rather than mathematical statistics applications of advanced probability. They are a useful addition to the introductory literature in the topics that they cover.

Emanuel Parzen
Texas A & M University

Frontiers in Statistical Quality Control, Edited by H. J. Lenz, G. B. Wetherill, and P.-Th. Wilrich, Birkhaüser Boston, 1981, 294 pp., $60.00 (paperbound).

This volume consists of 22 papers presented to an international workshop in Berlin in June 1980 by researchers making significant contributions to the field of statistical quality control during the last five years or so. The stimulus for such a workshop was that papers in this field are published in a wide variety of journals, and that there is little contact between the researchers and the practitioners.

The papers given were either in the field of sampling inspection or in process control and are presented alphabetically. The titles are:

CH. ASANO and K. JOJIMA, "Some Generalized Sequential Selection Plans, Basing on the Play-the-Winner Sampling and Successive Success Stopping Rule and Their Properties"

R. C. BAKER, "Relationships Between MIL-STD-105D Operating Characteristics and Conditional Sampling Operative Characteristics"

E. V. COLLANI, "On the Choice of Optimal Sampling Intervals to Maintain Current Control of a Process"

R. A. DAYANANDA and P. THYREGOD, "Some Examples of Matching Bayesian Variables and Attribute Acceptance Sampling Plans"

S. S. ERCAN, "Interrelated Minimizing Single Sampling Plans. Dodge-Romig Sampling Plans Revisited"

D. J. G. FARLIE, "Sampling Plans with Property Q"

W. C. GUENTHER, "Sample Size Determination for Tests Involving Means of One or More Normal Distributions"

Y. HOSONO, H. OHTA, and S. KASE, "Design of Single Sampling Plans for Doubly Exponential Characteristics"

T. KOYAMA, "Average Outgoing Quality Through Sampling Systems"

J. KÖLLERSTRÖM and G. B. WETHERILL, "Sampling by Variables with a Normal-Gamma Prior"

M. KÜHLMEYER, "Construction of an Economic and Secure Inspection and Production Strategy"

H.-J. LENZ, D. REETZ, and A. REIMANN, "A Stochastic Model for Optimal Inspection of Lots"

B. S. LIEBESMAN, "The Characterization of MIL-STD-105D Sampling Plans Using Normalized OC Curves"

E. MENIPAZ, "Managing Inspection Sampling Procedure in a Production Context: An Operations Research Methodology"

D. C. MONTGOMERY, R. G. HEIKES, and M. R. SCHEFFLER, "Probability Models for the Occurrence of Defects"

H. MOSKOWITZ and A. RAVINDRAN, "Algorithms and Optimization Schemes for Bayesian Acceptance Quality Control Plans"

H. RINNE, "Cost Minimal Process Control"

H. B. SACKROWITZ, "ARL Comparisons for Multi-level Sampling Plans"

E. M. SANIGA and T. ANTONIUK, "Operating Characteristics

of $\bar{X}/R$ Control Charts Under Non-Normal Assignable Cause Distributions"
K. SARKADI, "Some Remarks on Control Charts"
E. G. SCHILLING, "A Modified General Procedure for Sampling Inspection"
A. K. SHAHANI, "Choice of a Process Inspection Scheme: Some Basis Considerations"

In reading these papers, it is fairly apparent that there is a clustering of papers that can be grouped according to the types of research presented. The papers of Baker, Dayananda et al., Farlie, Hosono, Koyama, Köllerström, Liebesman, Sackrowitz, and Schilling might be grouped as Sampling Plans, covering standard plans (e.g., MIL-STD-105D), attributes or variables, and even non-normality. Another major subdivision might be called Optimality and would include the works of Asano et al., Ercan, Guenther, Kühlmeyer, Moskowitz, and Rinne. A third grouping could be Inspection and would contain the papers of Collani, Lenz et al., Menipaz, Montgomery et al., and Shahani. Finally the last grouping of papers, Control Charts, would contain the works of Saniga and Sarkadi.

No matter what the groupings are called (there is obviously an overlap of some of the categories and the papers themselves), this volume might be more useful if the papers were put together in some classification scheme rather than in the alphabetical ordering chosen by the editors. The papers themselves are of high research quality, with varying levels of mathematical/statistical sophistication. One thing that disturbed this reviewer was the two different spellings of the name Romig (in the table of contents and the title of the paper), something that seems incongruent.

Whether the research the authors presented in these papers becomes a tool of the practitioners, a purpose of this forum, only time will tell.

Robert L. Perry
The Proctor & Gamble Company

Sampling for Health Professionals, by Paul S. Levy and Stanley Lemeshow, Lifetime Learning Publications, a Division of Wadsworth, 1980, xv + 320 pp., $24.95.

This book covers the basic concepts of sampling theory with examples taken from the health field. The outstanding features of the book include

1. Excellent clarity of presentation.
2. Helpful format—the formulas are set out in boxes and there are statements in the margins to indicate major points of interest.
3. Clearly illustrated concepts—examples are given before the conceptual details are presented.
4. Boldface type used to highlight important terms.

The concepts covered in the book include simple random sampling, systematic sampling, stratification and stratified random sampling, network sampling, ratio and synthetic estimation, and randomized response. Cluster sampling is dealt with also, and there are separate chapters on simple one-stage cluster sampling, simple two-stage cluster sampling, and cluster sampling from clusters having unequal numbers of listing units. The final chapters of the book cover missing data, construction of forms and data-collection instruments, the interpretation of data, and writing survey reports.

All of the material is well written and clearly presented for the reader who has an elementary knowledge of statistics and distribution theory. This book, however, will also be useful to researchers with no statistical background who need to know about sampling. Because of the excellent presentation and the breadth and depth of the material covered, I feel that this book would be a very useful addition to the library of any statistician consulting in the health area.

Susan D. Horn
Johns Hopkins University

On Reduced Risk Estimation in Linear Models, by Erkki Liski, Tampereen Yliopisto, Tampere, Finland, 1979, 122 pp., dollar price not given.

The problem considered in this work is the estimation of parameters in the general linear model. In particular, let $y = X\beta + \varepsilon$, where y is an $n \times 1$ vector, β is a $q \times 1$ vector, X is a known $n \times q$ matrix of full column rank, and ε is composed of independent, mean zero errors. To estimate β under quadratic loss, the usual least squares estimator $\hat{\beta}$ given by $\hat{\beta} = (X'X) - X'Y$ is often used. (The presentation in the book also includes consideration of arbitrary quadratic and matrix losses. The remarks below carry over to these cases.) This estimator, despite its well-known advantages, suffers from several disadvantages. First, if $X'X$ is ill conditioned, $\hat{\beta}$ has an undesirable covariance matrix. From a decision theoretic point of view, $\hat{\beta}$ is usually inadmissible under the mean squared error (MSE) criterion when $q \geq 3$. Also, usually some (possibly vague) prior information about β is available. Such information is ignored by $\hat{\beta}$.

This book presents a thorough review of the literature concerning one possible approach to the construction of alternatives to $\hat{\beta}$. That approach is the consideration of linear estimators of the form $\tilde{\beta} = A\hat{\beta} + (I - A)d$, where A is a fixed $q \times q$ symmetric matrix with all eigenvalues in the interval $[0, 1]$ and d is any fixed $q \times 1$ vector. It is shown (also see Rao 1976 and Cohen 1966) that $\hat{\beta}$ is admissible *within the class* of linear estimators of β. (Of course, if $A = I$, $q \geq 3$, and the distribution of errors is normal, then $\hat{\beta}$ is overall inadmissible.)

After defining $\tilde{\beta}$, its MSE is computed. Also, various estimators (for example, some ridge estimators) are shown to be of the form $\tilde{\beta}$. Then the author obtains conditions under which $\tilde{\beta}$ is better in MSE than $\hat{\beta}$. However, since $\hat{\beta}$ is admissible in the class of linear estimators, any such conditions must involve β. That is, linear estimators can improve upon $\hat{\beta}$ only in some subset, say Θ_0, of the parameter space. (The region Θ_0 is an ellipsoid (determined by A and X) centered at d.) Hence, the optimal linear estimator, say $\tilde{\beta}_0$, is a function of β. The subscriber to linear estimation is now faced with a dilemma. He/She may replace β by $\hat{\beta}$ in $\tilde{\beta}_0$. The resulting rule is no longer linear and so no optimality properties can be asserted employing linear theory. Instead, β may be replaced by a prior guess at β in $\tilde{\beta}_0$. In any event, the risk of any $\tilde{\beta}$ ($\neq \hat{\beta}$) increases quadratically in the components of β as β moves away from d. Hence, if $\tilde{\beta}$ "shrinks" toward d and β is far from d, the risk of $\tilde{\beta}$ may be huge. To avoid this problem, it is necessary to consider nonlinear estimators.

The final chapter of the book begins with a review of the notion of preliminary test estimators. The author employs reasoning similar to that above to develop the following estimator. First, consider a fixed linear rule $\tilde{\beta}$ and its corresponding region of improvement over $\tilde{\beta}$, Θ_0. A classical (F) test of the hypothesis $\beta \in \Theta_0$ is then presented (assuming the errors are normally distributed). The implied estimator, say δ, is then defined to be $\tilde{\beta}$ if the test rejects and $\hat{\beta}$ if the test accepts. First, note that, as in the case of preliminary test estimators, δ is inadmissible since it is not smooth enough to be a Bayes or generalized Bayes rule. However, this rule may be reasonable in the following sense. When $|\beta - d|^2$ is very large, δ will be $\hat{\beta}$ with high probability (as desired). Hence, I expect that the MSE of δ is close to the MSE of $\hat{\beta}$ as $|\beta - d|^2 \to \infty$. This is clearly preferable to the MSE of $\tilde{\beta}$ as $|\beta - d|^2 \to \infty$.

The main weakness in the analysis given by the author is that it

includes no discussion of what choice of the matrix A is to be used. It seems to me that the most reasonable approach to this problem is Bayesian. In a Bayesian analysis A is determined by the prior covariance matrix for β (and, of course, X). If the user is then asked to specify both a prior mean (d) and a covariance matrix for β, it seems logical (when $q \geq 3$) to recommend the admissible, often minimax (and never "horrible" in MSE), generalized Bayes rules given in Berger (1980).

This book could be read both by theoreticians and by practitioners of linear model analysis. Knowledge of the linear model theory and multivariate analysis, as found in Rao (1965) for example, should be sufficient background for the reader.

Mark Berliner
Ohio State University

REFERENCES

BERGER, J. (1980), "A Robust Generalized Bayes Estimator and Confidence Region for a Multivariate Normal Mean," *Annals of Statistics*, 8, 716–761.

COHEN, A. (1966), "All Admissible Linear Estimates of the Mean Vector," *Annals of Mathematical Statistics*, 37, 458–463.

RAO, C. R. (1965), *Linear Statistical Inference and Its Applications*, New York: John Wiley.

RAO, C. R. (1976), "Estimation of Parameters in a Linear Model," *Annals of Statistics*, 4, 1023–1037.

Approximation Theorems of Mathematical Statistics, by Robert J. Serfling, John Wiley, 1980, xiv + 371 pp., $34.95.

This book presents a systematic treatment of the fundamental and most useful results of asymptotic-distribution theory in statistics. It emphasizes the application and manipulation of probability theorems to obtain limit theorems for statistics. A one-semester course in probability theory and mathematical statistics at the beginning graduate level is presupposed.

The book begins with a description of the foundational results (without proofs) from probability theory upon which large-sample distribution theory is built. It then deals with basic statistical forms such as sample moments and quantiles and the large-sample approximations to them, for example, the Bahadur representation of a quantile. Statistics created as functions of these more basic forms are then investigated by means of Taylor series approximations and other techniques. These building blocks are then followed by a careful treatment of classes of statistics, including common parametric forms such as maximum likelihood, likelihood ratio, and minimum chi squared statistics. In addition, separate chapters describe results for the popular estimator classes: U statistics (V statistics), von-Mises differentiable statistical functions, M-estimators, L-estimators, and R-estimators. The book concludes with a chapter devoted to the common types of asymptotic efficiencies.

This text provides an important resource for investigating the large-sample behavior of statistics encountered in a wide variety of settings. It includes conditions and careful proofs that are written in a fashion which is self-contained and which challenges, but does not exceed, the students' abilities. By working through the material and exercises, the reader acquires fundamental tools available for studying the large-sample behavior of statistics, an objective that is just as important from the researchers' and teachers' points of view as the theorems themselves.

One of the unique features of this book is its treatment of differentiable statistical functions and the conceptual framework that they provide for the nature of appropriate large-sample approximations. The insight furnished by this framework is very useful. A second key feature is the parallel treatment of popular classes of statistics providing large-sample approximations, limiting distributions, strong and weak convergence theorems, laws of the iterated logarithm, and Berry-Esseen results for members of appropriate U, V, M, L, and R estimator classes. The parametric estimators, such as maximum likelihood ones, are treated in somewhat less detail.

This book has already had a substantial impact on the study and application of large-sample results in statistics. Moreover, the conscientious reader is certain to acquire fundamental skills, valuable in research activities. As a result, the book is already considered by many to be a classic and a most important contribution to the statistical literature.

Ronald H. Randles
University of Florida

Sampling Methods for Censuses and Surveys (4th Ed.) revised and enlarged by Frank Yates, Macmillan, 1981, xvi + 458 pp., $55.00.

This book, first published in 1950, was "written primarily for those who have little or no previous training in mathematical statistics but who have some training or experience in the presentation and handling of statistical data." This fourth edition is a reprint of the third edition with minor revisions to two sections, major revision of the chapter on critical analysis of survey data, and a complete rewrite of the chapter on computer programs.

Chapter 1, entitled "The Place of Sampling in Census Work," describes the structure of the book and provides some key definitions. Chapter 2 ("Requirements of a Good Sample") is mainly a discussion of bias; how it arises, when it is permissible, and how to reduce its effects. Chapter 3 ("The Structure of Various Types of Sample") provides a discussion of the various types of samples from simple random sampling to composite sampling schemes. Examples are presented to illustrate the different sampling schemes. Chapter 4 ("Practical Problems Arising in the Planning of a Survey") and Chapter 5 ("Problems Arising in the Execution and Analysis of a Survey") deal with those items that need to be considered during the planning and execution of a survey. Lists of problems are presented and discussed, along with example surveys that illustrate these problems and their solutions.

Chapters 6 to 9 deal with the more mathematical and statistical aspects of survey work. Chapter 6 ("Estimation of the Population Values") covers estimates for the total and mean for the population for each type of sample discussed in Chapter 3. Included is a fairly clear discussion on the use of supplementary information in estimation. Chapter 7 ("Estimation of the Sampling Error") is a comprehensive discussion of error estimation for each sampling scheme. Chapter 8, entitled "Efficiency," reviews the concepts of relative accuracy, relative precision, and relative efficiency and uses these measures to compare estimates from different sampling schemes. Also included are techniques for determining optimal sampling fractions and the size of sample required for a given accuracy.

The two chapters, 9 and 10, that received major revisions for this edition are a disappointment. Chapter 9 ("Critical Analysis of Survey Data") was reorganized for the purpose of including more detail on analysis of data using linear and generalized linear models. The chapter concentrates on the relationship of these models to analysis of variance and regression methods. The topic is certainly much larger in scope than is possible to discuss in one chapter. Some of the fundamentals of linear model analysis are covered, but unlike other chapters in this book, Chapter 9 presents insufficient information to guide the novice in actually performing an analysis. The basic references to the use of linear models in categorical data analysis, for example, Grizzle, Starmer, and Koch

(1969), and to discrete multivariate analysis for cross-classification tables, for example, Bishop, Fienberg, and Holland (1975), are not given. These omissions limit the usefulness of this chapter for researchers who would be interested in the more complex approaches used in analyzing survey data.

Chapter 10 ("Computer Programs for Survey Analysis") was rewritten to take into account the more recent advances made in computer handling of survey data. The chapter outlines some general programs and examines some of their strengths and weaknesses. Owing to the rapid advances made in computer software for statistical analysis, there is some question as to the usefulness of discussion of software packages, which are constantly being modified. Weaknesses in programs identified in the book have been or will be shortly be corrected, with new problems arising as new revisions are implemented. Further argument on this subject can be found in Hamer (1981). An alternative would have been a more general discussion on this area and a statement encouraging the researcher to contact a statistician or computer scientist as to the limitations of locally available packages.

A major portion of the book remains unchanged from the first edition. This means that almost all of the examples date from the 1950's. Overlooking this weakness, however, most readers would find Dr. Yates's clear exposition on planning, conducting, and analysis of surveys an excellent reference text for applied survey research. I would recommend this book to those who do not already own a previous edition.

Kenneth M. Portier
University of Florida

REFERENCES

BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, Mass.: The MIT Press.

GRIZZLE, J. E., STARMER, C. F. and KOCH, G. G. (1969), "Analysis of Categorized Data by Linear Models," *Biometrics*, 25, 489–504.

HAMER, R. M. (1981), "Papers That Evaluate Computer Programs," *The American Statistician* 35, 264.

PUBLICATIONS RECEIVED

BANCROFT, T. A. and HAN, CHIEN-PAN, *Statistical Theory and Inference in Research*, Marcel Dekker, 1981, xiv + 372 pp., $34.50 (special student price of $24.50 for five or more copies).

Introductory text on the theory of statistics for graduate and advanced undergraduate students or researchers from substantive disciplines in which statistical methodology is applied. The aim of the theory discussion is to educate users of the statistical methods about their assumptions, validity, and limitations. Mathematical development is deemphasized. Topics discussed include basic discrete and continuous univariate and multivariate distribution theory, point and interval estimation, hypothesis testing, sampling, regression analysis, analysis of variance, and nonparametric testing procedures.

BERNARDO, J. M., DE GROOT, M. H., LINDLEY, D. V., and SMITH, A. F. M. (Editors), *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia (Spain)*, University Press, 1980, 647 pp., $25.00 (paperbound).

Collection of 26 papers presented at an international conference on Bayesian statistical inference. Participants included some of the leading U.S. and European Bayesian statisticians. Topics covered include Bayesian approaches to regression and time series analysis, Bayesian hypothesis testing and nonparametric procedures, Bayesian approaches to robust statistical analysis and outlier detection, Bayesian inference for group judgement formulations, and others. This volume will interest those readers who wish to have a research-level reference on various facets of Bayesian theory.

BRANS, J. P. (Editor), *Operational Research '81*, North-Holland, 1981, xx + 984 pp., $106.50.

Collection of about 70 papers presented at the Ninth IFORS International Conference on Operational Research held at Hamburg, Germany in July, 1981. Topics covered include OR in organizations, in developing countries, in the public sector; multiple objective programming, mathematical programming; location and transport problems; production management; simulation and games; queues and Markov processes, and many other subjects. This advanced reference volume provides an overview of current international research in the theory and applications of operations research.

COX, D. R. and SNELL, E. J., *Applied Statistics: Principles and Examples*, Chapman and Hall, 1981, vii + 189 pp., $32.00 ($14.95 paperbound).

Part I deals with general principles and ideas involved with applying statistical methods in real situations. Part II makes up about two-thirds of the book and consists of a collection of examples illustrating a wide variety of techniques with real data sets. Examples illustrate graphical techniques, descriptive statistics, analysis of variance and regression analysis, survival analysis, analysis of binary data, and other topics. The data are presented so that readers may attempt their own analyses. A number of additional sets of data are included, without analyses, for use by students. This book may be used as a text or supplementary text for statistical methods courses. Experienced data analysts may benefit from the general discussion in the first part.

DAGNELIE, PIERRE, *Principes d'Experimentation*, Les Presses Agronomiques de Gembloux, 1981, 182 pp., price not given (paperbound), in French.

Basic text on experimental design. Discussion of notions of experiments, experimental units, primary and ancillary observations. Designs discussed include completely randomized, randomized complete block, split plot, Latin square, incomplete blocks, factorial experiments, experiments over time. Discussion is illustrated with examples and analyses based on real experiments and real data.

DAGNELIE, PIERRE, *Théorie et Méthods Statistiques Exercises*, Les Presses Agronomiques de Gembloux, 1981, 186 pp., price not given (paperbound), in French.

Collection of exercises on statistical theory and methodology. Solutions are given. Topics covered include descriptive statistics, probability distributions and transformation of variables, estimation, hypothesis testing, statistical methods for means and variances, goodness of fit tests, contingency table tests, analysis of variance, regression analysis, multiple comparisons, analysis of covariance, nonparametric methods.

DEUTLER, T., *Schätz- und Testverfahren Bei Normalverteilung Mit Bekanntem Variationskoeffizienten*, Springer-Verlag, 1981, viii + 195 pp., approximately $33.40 (paperbound), in German.

Discussion of the theoretical properties of various normal theory-based estimation and hypothesis testing procedures. Discussion is intended for people from substantive disciplines who will apply the methods, as well as for statistics students.

FINDLEY, DAVID F. (Editor), *Applied Time Series Analysis, Vol. II*, Academic Press, 1981, xii + 798 pp., $19.50.

Collection of papers that were presented at the Second Applied Time Series Symposium at the University of Tulsa in 1980. Topics discussed include analysis of multidimensional signals, classification of time series by nearest neighbor rules, identification of ARMA models, approaches to time series modeling, time series theory when the data are irregu-

larly spaced in time, forecasting, and many others. This volume will be of interest to researchers in and practitioners of time series analysis as a reference on current developments in applied time series research.

GORDON, A. D., *Classification Methods for the Exploratory Analysis of Multivariate Data*, Chapman and Hall, 1981, xii + 193 pp., $26.00.

Introduction to clustering and geometrical approaches to dividing large multivariate data sets into relatively homogeneous subsets. Book is aimed at researchers from substantive fields who might wish to use those methods, as well as undergraduate and graduate students in statistics. Topics discussed include similarity measures, clustering algorithms, multidimensional scaling, measures of effectiveness of the classification procedures, and others. Algorithms and procedures are illustrated by examples and several case studies are given.

HAND, D. J., *Discrimination and Classification*, John Wiley, 1981, x + 218 pp., $36.95.

Collection of various approaches to discrimination and classification analysis. Methods based on normal theory and nonparametric methods are discussed. Methods for discrete variables, sequential methods, Bayesian methods, and variable selection techniques are discussed. Cluster analysis methods are briefly discussed. Algorithms are described and are illustrated with examples based on real data. Exercises are included at the ends of chapters and are a mixture of theory, numerical examples, and computer projects. A relatively large bibliography is also included.

IKEDA, NOBUYUKI and WATANABE, SHINZO, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, 1981, xiv + 464 pp., $85.25.

Systematic treatment of the theory of stochastic integrals and stochastic differential equations. Highly advanced, theoretical, and mathematical work. Topics discussed include martingales, Brownian motions, stochastic integrals and differentials, stochastic differential equations, diffusion processes on manifolds, and others. This book will interest those readers with strong backgrounds in stochastic processes who wish to study a rigorous probabilistic approach to the theory of stochastic integrals and differential equations.

LAWLESS, J. F., *Statistical Models and Methods for Lifetime Data*, John Wiley, 1982, xi + 580 pp., $36.95.

Reference and graduate-level text on theory and methodology of life data analysis. Both parametric and nonparametric approaches to the analysis of censored data are discussed. Graphical procedures are included. Parametric and nonparametric approaches to regression analysis are discussed, including the proportional hazards regression model. Goodness-of-fit tests are also considered. Procedures are illustrated with examples drawn mainly from engineering and biomedical applications. Problems, both theoretical and numerical, are contained at the ends of chapters. This book will be of interest to those readers who would like a text or reference on the various facets of life data analysis that presupposes no prior background but that provides an overview of the field.

MANSKI, CHARLES F. and MCFADDEN, DANIEL (Editors), *Structural Analysis of Discrete Data With Econometric Applications*, The MIT Press, 1981, xxv + 477 pp., $29.95.

Collection of 13 papers dealing with econometric analysis of discrete data. They are concerned with problems of statistical inference on the structure of stimulus-response relationships when the set of possible responses is finite. There are four main categories of papers: statistical analysis of discrete probability models, dynamic discrete probability models, structural discrete probability models derived from theories of choice, and structural equations models with discrete endogenous variables. The papers in this volume are original research contributions. The grouping of papers is intended to give an overview of the field. This collection of papers will be of interest to those readers interested in analysis of structural models with discrete data, involving both econometric and other applications.

MARSHALL, ALBERT W. and OLKIN, INGRAM, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979, xx + 569 pp., price not given.

Text and reference on the theory of majorization and the use of this theory as a general approach to deriving inequalities. Part I discusses the basic theory. Parts II and III discuss mathematical, probabilistic, and statistical applications. Parts IV and V discuss generalizations and complements such as convex functions, stochastic ordering, and matrix factorizations. Book is theoretical and mathematical. This book will appeal to those mathematically well-prepared readers who are interested in the theory of inequalities in general or in particular to the applications of inequalities in probability and statistics.

MEEKER, WILLIAM Q., JR., CORNWELL, LARRY W., and AROIAN, LEO A., *Selected Tables in Mathematical Statistics, Volume VII, The Product of Two Normally Distributed Random Variables*, American Mathematical Society, 1981, vii + 256 pp., price not given.

Tables for the fractiles of the distribution of the product of two normally distributed random variables with means δ_1, δ_2, standard deviations 1, 1, and correlation ρ are presented. Numerical methods underlying the tables and interpolation between table entries are discussed. Applications to sampling problems in accounting and agriculture are briefly mentioned. Other applications to quality control, economic theory, reliability modeling, statistical inference theory, and business are also mentioned.

NALIMOV, V. V., *Faces of Science*, ISI Press, 1981, xv + 297 pp., $22.50.

Philosophical discussion on the nature of scientific thought and development by an eminent Soviet mathematician and statistician. Collection of related essays on the logic of scientific methodology, formulating and testing hypotheses, Bayesian approaches to scientific reasoning, deterministic and probabilistic mathematical modeling and the shift toward probabilistic models, as well as other topics. Nalimov also discusses the relation between scientific and humanistic thinking and describes the problem of their separation in the USSR. This book will interest those readers interested in the foundations of scientific thought and in the role of probabilistic reasoning in scientific development.

PATEL, JAGDISH, K. and READ, CAMPBELL, B., *Handbook of the Normal Distribution*, Marcel Dekker, ix + 337 pp., $35.00.

Compendium of results relating to the normal distribution. Distributional properties of the normal law and of derived statistics are emphasized. Topics discussed include historical development, distributional properties, tables and algorithms, order statistics, Gaussian processes, characterizations, bivariate normal law, and approximations to other distributions. This book will be of interest to those readers who would like a relatively concise reference to the distributional properties of the normal law and an introduction to the literature of that field.

PLACKETT, R. L., *The Analysis of Categorical Data, Second Edition*, Griffin's Statistical Monograph No. 35, MacMillan Publishing Co., Inc., 1981, xi + 207 pp., price not given (paperbound).

Discussion of various methods for analysis of categorical data. Results are illustrated with examples based on real data. Topics discussed include distributional properties of the Poisson, binomial, and multinomial distributions, tests and measures of association for two- and three-way classifications, logit

regression, log-linear models, as well as others. Exercises and a relatively large set of references are also included.

SEN, PRANAB K., *Sequential Nonparametrics*, John Wiley, 1981, xv + 421 pp., $42.50.

Advanced graduate-level text on sequential inference procedures based on nonparametric statistics. Required background includes parametric sequential analysis, nonsequential nonparametric methods, graduate-level probability, and convergence of probability measures on function spaces. Discussion is theoretical and mathematical. The first part of the book discusses weak convergence of various types of statistics used in robust and nonparametric statistical inference. The second part of the book applies this theory to derive asymptotic properties of sequential nonparametric procedures. Includes exercises following each chapter.

TAILLE, CHARLES, PATIL, GANAPTI, P., and BALDESSARI, BRUNO A. (Editors), *Statistical Distributions in Scientific Work*, D. Reidel, 1981.

Volume 4: *Models, Structures, and Characterizations*, xxii + 456 pp., $59.00.

Volume 5: *Inferential Problems and Properties*, xxii + 439 pp., $59.00.

Volume 6: *Applications in Physical, Social, and Life Sciences*, xxii + 445 pp., $59.00.

Special price for the set $156.00.

Three-volume set containing the proceedings of the NATO Advanced Study Institute on Statistical Distributions held at Trieste during July and August 1980. Volume 4 contains papers on continuous models, discrete models, structural properties of distributions, computer generation of random variables, and characterizations. Volume 5 contains papers on goodness-of-fit testing, parameter estimation, hypothesis testing, approximations, and reliability and life testing. Volume 6 discusses applications to the physical, social, and life sciences. These volumes will be useful references to those readers interested in recent developments in statistical distribution theory, inference theory, and scientific applications.

VINOD, HRISHIKESH and ULLAH, AMAN, *Recent Advances in Regression Methods*, Marcel Dekker, 1981, xii + 361 pp., $39.50.

Graduate-level text on biased and robust estimation techniques as alternatives to least squares methods in the linear regression model. Ridge regression and Stein-type shrinkage estimators are discussed. Ridge regression ideas are applied to discriminant analysis and canonical correlation analysis. Robust regression techniques are also discussed. ARMA models, distributed lag models, and simultaneous equation models are discussed. This text will interest those readers who would like an expository account of recent developments in biased estimation in regression, along with examples and exercises.