Of course, in 1866 Venn developed the frequency notion of probability, which strongly opposed inverse probability calculations.

Another thread is the computation of predictive probabilities first attempted by Price, the publisher of Bayes's *Essay*. Laplace's rule of succession is one example of this type of calculation. Such computations have become so closely associated with the name of Bayes (although his *Essay* contains none) that Pearson (1978) claimed that "Bayes had certainly not reached 'Bayes's Theorem'."

This reader found the repeated presentations of similar types of probability calculations by different workers throughout the eighteenth and nineteenth centuries quite tedious. On the other hand, I found some interesting gems hidden among the integral signs. For example, it appears that one of the very first Bayesian *statistical* calculations was Laplace's computation (1781) of the probability that the chance of birth of a boy in Paris was less than one half in 1778. The numerical value, using data from 1845–1870, differed from one by 1.15×10^{-42}. Subsequent work in this area by Laplace contains tests of statistical hypotheses, such as whether this chance is different between Paris and London.

The author has attempted to express the probability calculations given in the historical works surveyed in modern form. On the other hand, he makes no effort to present the results either in a historical or in a statistical framework. Thus the reader is left with a vast amount of material, but is not carefully guided through it. In particular the author presents no point of view for the reader to take issue or agree with. Thus the work is best suited as a source for further historical research. Perhaps someone else will attempt to paint the picture of Bayesian statistics in a historic perspective.

PETER GUTTORP
University of Washington

REFERENCES

Bayes, T. (1763), "An Essay Towards Solving a Problem in the Doctrine of Chances by the late Rev. Mr. Bayes, F.R.S., Communicated by Mr. Price in a Letter to John Canton, A.M. F.R.S.," *Philosophical Transactions*, 53, 370–418.

Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge, U.K.: Cambridge University Press.

Laplace, P. S. (1781), "Mémoire sur les Probabilités," *Mémoires de l'Academie Royale des Sciences de Paris*, 227–332.

Pearson, K. (ed.) (1978), *The History of Statistics in the 17th and 18th Centuries Against the Changing Background of Intellectual, Scientific, and Religious Thought*, New York: Macmillan.

Stigler, S. (1982), "Thomas Bayes's Bayesian Inference," *Journal of the Royal Statistical Society*, Ser. A, 145, 250–258.

Todhunter, I. (1865), *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, textually unaltered reprint 1965, New York: Chelsea House.

Nonlinear Multivariate Analysis.

Albert Gifi. New York: John Wiley, 1990. xx + 579 pp. $72.95.

It's tempting just to guess what the following acronyms refer to: PRINCALS; OVERALS; HOMALS; CANALS; ANACOR; MORALS; CRIMINALS; MANOVALS; PATHALS; PARTALS. These are the names of the procedures described in "Nonlinear Multivariate Analysis" by Gifi, *nom de plume* for a group of authors who have been at one time or another associated with the Department of Data Theory at the University of Leiden, The Netherlands. The book represents a compilation of the work resulting from intensive efforts of many bright, young scholars for more than 10 years. Behind any such sustained effort must be a master mind who laid out the basic design and kept inspiring the people. Who this is becomes apparent as you read the book.

Nonlinear Multivariate Analysis is about multivariate analysis (MVA) with nonlinear data transformations. The word "nonlinear" in the title thus refers to the type of transformations permitted in the analysis and not to the nonlinearity in fitted multivariate (MV) models. Other books on MVA have emphasized exploratory/descriptive aspects (e.g., Gnanadesikan 1977; Lebart, Morineau, and Warwick 1984; Takeuchi, Yanai, and Mukherjee, 1982). However, Gifi is unique in its emphasis on data transformations. Other important features are the emphasis on graphical displays to present and communicate analysis results and the use of resampling methods (e.g., bootstrap, jackknife) to assess the stabilities of the analysis results. The book covers, with nonlinear twists, such techniques as principal component analysis (PRINCALS), k-set canonical correlation analysis (OVERALS), homogeneity analysis (HOMALS; also known as multiple correspondence analysis), two-set canonical correlation analysis (CANALS), correspondence analysis of simple two-way contingency tables (ANACOR), multiple regression (MORALS), discriminant analysis (CRIMINALS), MANOVA (MANOVALS), path analysis (PATHALS), and partial correlation analysis (PARTALS), with primary emphases attached to the first three methods.

In the Gifi system of nonlinear MVA, observed data are considered as having been generated by nonlinear transformations of underlying linear processes, so that once the data are "transformed" back, they can be fitted by conventional (linear) MV models. Unlike most other procedures that allow data transformations, a specific data transformation to be applied does not have to be specified a priori. Both the optimal data transformation and the best data representation (estimates of parameters in fitted models) are derived from the analysis through some optimization procedure (Young 1981). This approach clearly has its origin in psychometrics, where the numerical quality of data is almost always suspect and data transformations are applied rather routinely to improve the quality of the data. But it was not until Shepard's (1962) and Kruskal's (1964a, b) landmark papers on nonmetric multidimensional scaling (MDS) that the real thrust of the idea came to be fully realized. These papers clearly demonstrated the feasibility of deriving a multidimensional representation of stimuli while at the same time seeking the best monotonic transformation of observed similarity data, thought to convey only ordinal information about the underlying distances. Gifi represents a systematic application of this optimal scaling idea to MVA. The optimal scaling approach has been kept within a limited realm of psychometrics for quite some time, but recently more statisticians are becoming interested, following the lead of psychometricians (Breiman and Friedman 1985; Hastie and Tibshirani 1990).

In the Gifi system the data generally are represented as indicator matrices (matrices of dummy variables), one for each variable with as many columns as there are distinct observations in the variable. Then the data transformation problem can be viewed as a quantification problem of the columns of the indicator matrices subject to various constraints on the quantifications. Particular constraints to be imposed on the quantifications reflect various measurement characteristics of the variables (e.g., scale levels of measurement), and regulate the type of permissible transformations to be applied to the variables (e.g., one to one, monotonic, spline). Note that the case of no data transformations can be thought of as one in which the quantifications are given *a priori*. The quantified data then are approximated by some model.

This approach necessitates a criterion to measure a goodness (or badness) of agreement between the model and the quantified data. In Gifi's MVA, the criterion called "meet" loss plays a central role. (Although the adoption of this criterion is one of the most important contributions of Gifi, the book would have been much more readable if this criterion had been introduced right at the outset and was followed throughout the rest of the book.) This criterion, sometimes called the "homogeneity" criterion, requires the quantified data for k sets of variables to be as homogeneous as possible (Carroll 1968; Meredith 1964). It can be used to fit virtually all the MV models covered in the book. A number of interesting special cases follow from this criterion. For example, ordinary k-set canonical correlation analysis follows when the quantifications for the k sets of variables are assumed given a priori. Homogeneity analysis (multiple correspondence analysis) follows when k sets of variables represent k indicator matrices pertaining to, for example, k multiple-choice items. When in addition $k = 2$, the first case reduces to ordinary canonical correlation analysis between two sets of variables, and the second example reduces to correspondence analysis of simple two-way contingency tables. Finally, principal component analysis results when each of the k variable sets consists of a single prequantified variable. It is an interesting exercise to figure out how these special cases can be derived from the criterion. (A key idea can be found in Sec. 5.2.)

In all these special cases, solutions can be obtained analytically. These are rather exceptional cases, however; in general, the criterion cannot be optimized simultaneously with respect to both model parameters and quantification parameters. Consequently, an iterative procedure, called alternating least squares (ALS), is used to optimize the criterion alternately with respect to the model parameters and the quantification parameters. (The "ALS" suffix for the procedures stands for "by alternating least squares.") More specifically, the criterion is optimized with respect to the model parameters with the quantification parameters fixed and then with respect to the quantification parameters with the model parameters fixed. This process is repeated until convergence is reached. The ALS algorithm has a desirable property of monotonic convergence.

This book is quite stimulating from a methodological point of view. It contains many interesting ideas and will be of interest to many methodologically oriented statisticians. However, it apparently is not intended for data analysis practitioners. Little mention is made of practical limitations of the methods described; rather, all methods are described as if they were foolproof. The problems of recovery of underlying structures, local optima, degenerate solutions, and possible nonunique solutions are not discussed in sufficient depth. This is probably because these problems are mathematically less tractable; nonetheless, they may have serious consequences in practical data analytic situations. Recently Buja (1990) systematically investigated recovery issues in the ALS algorithm for two-set nonlinear canonical correlation analysis. More evaluative work of this sort must be conducted before the methods described in the book can be used with ease by data analysis practitioners.

Little is said about convergence properties of the ALS algorithm. The basic logic used in de Leeuw, Young and Takane (1976) for convergence of the ALS algorithm is that the monotonically decreasing bounded sequence must converge. There is nothing incorrect about this argument, but the question is: What's the nature of the convergence point? Aside from convergence to nonglobal optima, the convergence point may be a nonoptimal stationary point (i.e., a saddle point). It may even be a nonstationary point. This may happen because the amount of decrease in the monotonically decreasing sequence generated by the algorithm may get smaller as the iteration proceeds, and the sequence may never get beyond some asymptote. It may be that this is an extremely rare case (or it may never happen), but I have seen no proof that such an event rarely or never occurs.

Having read through the book, I am still not entirely convinced as to why data transformations are useful in MVA. Surely the fit will be improved at least nominally by allowing more flexible data transformations; however, this also means less reliable parameter estimates. Thus there should be a strong reason justifying the data transformations. The book deals primarily with technicalities of how the transformations are implemented algorithmically and not so much with why and when they are needed.

Nonlinear Multivariate Analysis takes conventional MV methods as points of departure and focuses on how various data transformations are incorporated into the conventional MV techniques. Consequently, it never questions adequacies of the conventional techniques, which do have their limitations. For example, multiple correspondence analysis (homogeneity analysis) is known to have many undesirable properties, such as the horseshoe phenomenon, the tendency for less frequent categories to dominate the solution, and so forth. The book is written as if these problems do not exist.

Despite these quibbles, as a developer of new data analysis methods I very much enjoyed reading this book. Above all the book provides an excellent source of new ideas for further research. For me that is the most important reason why a book should be read.

YOSHIO TAKANE
McGill University

REFERENCES

Breiman, L., and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association,* 80, 580–619.

Buja, A. (1990), "Remarks on Functional Canonical Variates, Alternating Least Squares Methods and ACE," *The Annals of Statistics,* 18, 1032–1069.

Carroll, J. D. (1968), "A Generalization of Canonical Correlation Analysis to Three or More Sets of Variables," in *Proceedings of the 76th Annual Convention of the American Statistical Association,* pp. 227–228.

de Leeuw, J., Young, F. W., and Takane, Y. (1976), "Additive Structure in Qualitative Data: An Alternating Least Squares Method With Optimal Scaling Features," *Psychometrika,* 41, 471–503.

Gnanadesikan, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations,* New York: John Wiley.

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models,* London: Chapman & Hall.

Kruskal, J. B. (1964a), "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika,* 29, 1–27.

Kruskal, J. B. (1964b), "Nonmetric Multidimensional Scaling: A Numerical Method," *Psychometrika,* 29, 115–129.

Lebart, L., Morineau, A., and Warwick, K. (1984), "*Multivariate Descriptive Statistical Analysis,* New York: John Wiley.

Meredith, W. (1964), "Rotation to Achieve Factorial Invariance," *Psychometrika,* 29, 187–206.

Shepard, R. N. (1962), "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function, I & II," *Psychometrika,* 27, 125–140, 219–245.

Takeuchi, K., Yanai, H., and Mukherjee, B. (1982), *The Foundations of Multivariate Analysis: A Unified Approach by Means of Projection onto Linear Subspaces,* New Delhi: Wiley Eastern.

Young, F. W. (1981), "Quantitative Analysis of Qualitative Data," *Psychometrika,* 46, 347–388.

Statistical Models in Behavioral Research.

William K. Estes. Hillsdale, NJ: Lawrence Erlbaum, 1991. ix + 159 pp. $36 (cloth); $16.50 (paperback).

Estes's refreshingly slender book distills his long experience in teaching analysis of variance and, to a lesser extent, regression to serious psychology students. He states (in chap. 1) that the book is not intended as a text, but rather as a supplement to any standard "psych stat" text. It provides a coherent discussion of standard ANOVA theory, assuming that calculations are done by computer. The target audience is stated as advanced undergraduate and graduate students in psychology; I would add Ph.D.-level researchers to the list.

There's considerable good sense to be found in this book. The book is effective in its treatment of the one-way layout and noteworthy in its discussion of the two-way layout, but somewhat sketchy in considering repeated-measures designs and the problems of unbalanced design. Estes is particularly effective in ramming home the points that neither a statistical model nor a scientific model is a perfect representation of reality, and that a statistical model is only relevant to the extent that it illuminates a scientific question. Both statisticians and psychologists need to be reminded of this fact periodically. Estes' comments about the two-way layout are particularly welcome to those of us who have grown weary of reading that "the interaction of whosis and whatzat is significant (F = so much, p = such and such), as is the main effect of whosis." He makes very clear that the real scientific issue very often can be better examined by looking at contrasts, rather than at the omnibus F statistics. He is particularly compelling in suggesting simple effects analysis rather than the conventional ANOVA table analysis. Although he does concede that some scientific studies are designed to examine interaction, he fails (to my disappointment) to consider interaction contrasts. A serious student (or researcher) could benefit from this book.

Yet I hesitate to suggest it as supplementary reading for a course. Most crucially, it contains no exercises. Keep in mind Cobb's Dictum: Judge a book by its exercises, and you cannot go far wrong; see Cobb (1987). This applies to supplements almost as fully as to texts. Without wrestling with the application of a book's principles to new situations, most students will not master the principles. Assuredly, developing meaningful and effective exercises is difficult and unpleasant; yet an elementary statistics book without exercises is only half a book.

In addition, the book feels rather old-fashioned. It reminded me of the Hays text (1988) in that it mixes discussions and mathematical derivation willy-nilly in a single paragraph. My experience with the "Haysy" style was not happy; students got so tied up following the derivation that they forgot the point. The book also is innocent of any of the robustness/diagnostics developments of the last decade; there is essentially no discussion of outliers, influence, or transformation. Apart from the reliance on computer packages to do the arithmetic, the book could have been written 20 years ago.

Teachers of "psych stat" could benefit from this book. A thoughtful scholar distilled considerable insight into its pages. But before selecting it as a supplement for a course, however, an instructor should consider carefully exactly how it will be used.

DAVID K. HILDEBRAND
University of Pennsylvania

REFERENCES

Cobb, G. (1987), "Introductory Textbooks: A Framework for Review," *Journal of the American Statistical Association,* 82, 321–339.

Hays, W. L. (1988), *Statistics* (4th ed.), New York: Holt, Rinehart and Winston.

Applied Computational Statistics in Longitudinal Research.

Michael J. Rovine and Alexander von Eye. San Diego, CA: Academic Press, 1991. xiii + 237 pp. $34.50 (paperback).

This book aims to provide a much-requested service: describing and demonstrating methods of analyzing longitudinal data for scientists. It provides concise descriptions of selected methods for analyzing longitudinal data with annotated computer codes and outputs to demonstrate how the methods can be used with real data. Selected methods are categorized into four general groups: repeated-measures ANOVA, methods for measuring structural change, time series analysis, and categorical data analysis. Given the book's lofty goals, it is not surprising that the authors did not fully achieve their aims.

This book should be very useful for scientists who want good, concise descriptions of the selected methods. In particular, the chapters on repeated-measures ANOVA methods and methods for structuring change are very well written.

Two major weaknesses prevent the book from meeting its goals, however. Perhaps the most important weakness is the omission of some important methods—most notably, methods for estimating random effects or mixed effects general linear models in the growth curve section. Although the discussion of Tucker's principal components approach is useful, the omission of the mixed or random effect model is very unfortunate. Further, some other important topics are not covered, including power, nonlinear models, and general estimating equation models.