and the neighborhood approaches used widely in application areas. The chapter includes an unusually clear exposition of tree-based classification.

Although error rates and their estimation are discussed throughout the book, this is the topic of Chapter 10, where the common parametric estimators are discussed along with those based on resubstitution, the jackknife, cross-validation, and the bootstrap. Chapter 11 looks more carefully at estimating the posterior probabilities of group membership, primarily in the normal case. This is followed by Chapter 12, covering variable selection and testing for additional discrimination in the normal case.

The final chapter, Chapter 13, is devoted to image analysis. This work is a digression from what one would expect in a text on discriminant analysis and flows from the author's interest in remote sensing using satellite data. Satellite imagery extends conventional classification in two ways: theoretically, in that the training data and the items to be classified are spatially correlated rather than statistically independent, and computationally, in that huge volumes of data are available. Chapter 13 introduces some concepts and procedures in this area and discusses the implications of using conventional independent-sample methodology on spatial image data.

The book includes two indices: an author index and a subject index. The subject index was the only feature of the book with which I was less than delighted. It is only 8 pages long—adequate for most books, but not detailed enough for one with so much distilled information. Ameliorating this brevity a bit is a detailed and accurate table of contents, but even with both sources I found it harder than usual to put my finger on particular topics. Moreover, although dipping into the book to come up to speed on particular topics is generally easy and productive, this would have been facilitated by a glossary of the notation used consistently throughout the book. I noted only a few typographical errors, most of which were of the "multinominal" for "multinomial" sort and should cause no real confusion.

I hope that this broad-brush sketch of the contents makes it sufficiently clear how comprehensive this book is. The author obviously intended to write a state-of-the-art account of classification—and has largely succeeded. The book is clearly not intended to be—nor is it suitable for use as—a textbook (there are no chapter-end exercises, for example), but it is a great backup reference source for advanced students. I believe that it will be used mainly by researchers and advanced users wanting a resource on problems and methods in the broadly construed area of classification. Speaking as someone with a long-standing research interest in applicable multivariate methods, I found *Discriminant Analysis and Statistical Pattern Recognition* exciting and learned much from reading it. It will occupy an easily accessible spot on my bookshelf and within a few years should be as well thumbed as the yellow pages. I believe that most serious workers in multivariate analysis will find it similarly useful.

DOUGLAS M. HAWKINS
University of Minnesota

REFERENCES

Hand, D. J. (1981), *Discrimination and Classification*, New York: John Wiley.
Lachenbruch, P. A. (1975), *Discriminant Analysis*, Hew York: Hafner.
Therrien, C. W. (1989), *Decision Estimation and Classification*, New York: John Wiley.

Exploratory and Multivariate Data Analysis.

Michel Jambu. Boston: Academic Press, 1991. xiii + 474 pp. $79.

This book is another monument to the work of J. P. Benzécri, the most important person in French statistics for about 25 years. It is not the first such monument, not even in the English language: books by Greenacre (1984), Lebart, Morineau, and Warwick (1984), and Jambu and Lebeaux (1983), preceded this book. Gifi (1990) discussed Benzécri's contributions in a somewhat wider context, relating them to psychometrics and to Tukey's form of data analysis. I own 11 French books on *Analyse des Données,* and I am sure there must be many more. Recently (albeit somewhat belatedly), one of the many volumes written by Benzécri himself was translated and published in English (Benzécri 1991).

Jambu's *Exploratory and Multivariate Data Analysis* is the translation of a French book written in 1989. It is not clear whether the French version was ever published or was directly translated into English. Compared to Lebart et al. (1984), it provides more complete coverage of basically the same terrain. Greenacre (1984) focuses on correspondence analysis, which means that it covers less, but in far greater detail and depth and with substantial British influences.

Although *Exploratory and Multivariate Data Analysis* is perhaps the most complete English-language book on French data analysis (FDA), it is becoming more and more clear that FDA comprises only of a very limited class of data analysis techniques. Thus even the best possible book on FDA will not be very exciting. My guess is that this will become increasingly apparent if the influence of Benzécri fades. Jambu's book is certainly not the best possible book on this subject; a translation of Cailliez and Pagès (1976), Bouroche and Saporta (1980), or Fénélon (1981) would have been more interesting.

This book rests on the premise that FDA is different from data analysis practiced in other countries. This is what Jambu says in the preface—but is it true? Partly. It is true that from 1970 to 1990, academic statistics in France was very different from academic statistics in England, the United States, or Scandinavia. On the other hand, French academic statistics, under Benzécri, was not unlike psychometrics and not unlike much work done in the Classification Society. This is made quite obvious by the book's references: some psychometricians, some classificators, and only two references that could be called statistics in the sense of classical inferential statistics. The first is Yule and Kendall (1930), and it turns out that this book is only included because of Yule's list of requirements for a coefficient of association. The second reference is Gnanadesikan's (1977) book on multivariate analysis, which is also not really classical statistics and is never actually mentioned in the text. Thus it seems that the work of Benzécri and his school completely ignores statistics. In fact it makes much more sense to say that correspondence analysis and related techniques are in the tradition of exploratory multivariate analysis, and actually they fit into this tradition seamlessly as soon as the appropriate translations are made.

Two interesting questions remain. The first is: How is it possible that canonical analysis of contingency tables and hierarchical clustering completely dominated French academic statistics for so long? It seems that these two topics constitute a fairly narrow menu of techniques from which to choose. The answer must lie in the power structure and the history of French statistics and in its relationship to the probabilists in French universities. Fascinating territory, but I am not going to cover it. The second question is: Why did a brilliant French mathematician use his considerable analytic powers to study these rather pedestrian techniques? Another interesting question, and one that has been studied by the French sociologist Philip Cibois, who related Benzécri's choice of program to his fundamentalist religious convictions. Fascinating, to be sure, but again territory that I will avoid. Historians of statistics should also explore Benzécri's visit to Bell Telephone Labs in Murray Hill in the mid-1960s, which still seems to account for most of the English language references in Jambu's 1991 book! There is some information on this in Benzécri's curious *Histoire and Préhistoire de l'Analyse des Données* (Benzécri 1982).

To be on reasonably safe ground, we give the five data analysis principles of Benzécri (1973, pp. 1–17), using the translation provided by Gifi (1990, pp. 25–26). We give only the catch phrases; there is more in Gifi and obviously much more in Benzécri. Not surprisingly, the principles (without a suitable reference) occur almost literally in Jambu (pp. 4–5):

1. Statistics is not the same thing as probability theory.
2. The model must follow the data, not the other way around.
3. It is convenient to treat simultaneous information on as many variables as possible.
4. For analysis of complex facts, we cannot do without the computer.
5. All techniques designed before the advent of automatic computing must be abandoned.

This will give you an idea about FDA, and it also shows clearly that Tukey's form of data analysis is quite different. Obviously, Tukey would not agree with Principle 3 nor, consequently with Principles 4 and 5. By now, everybody except some radical Bayesians will agree with Principle 1, whereas Principle 3 will appeal to those who admonish us to make our experiments elaborate and will please those whose hardware acquisitions force them to believe Principles 4 and 5. Principle 2 is the pièce de résistance, of course, but in various disguises it has been in that position for more than 2000 years now. A verdict does not seem to be in sight. Jambu typically botches his presentation of the five principles by saying that data analysis is deductive because it deduces only from gathered data. This is usually known as *induction,* and it is at the root of the proud *Hypotheses non fingo* of Newton (another subscriber to Principle 2).

Jambu pays brief lip-service to the important steps of data decision, data conception, data elaboration, data input, data management, data communication, and data presentation. The fervor with which he inserts word "data" in almost every possible place is quite remarkable. But the meat of the book is data analysis, which takes up Chapters 3–11. Finally, in Chapter 12 we have the unavoidable software—in this case the DACL (Data Analysis and Classification Library), which has 38 different flavors.

The classification on which the chapters are based is somewhat shaky. Chapters 3, 4, and 5 are called 1-D, 2-D, and N-D Statistical Data Analysis. This turns out to be descriptive statistics, including graphics, regression, contingency tables, and so on. All of this in less than 100 pages, with no references, and as far as I can see nothing typically French. This part is quite useless and not particularly well done. It is just a long list of things that one can also do with data. Unfortunately, this type of superficial listing of tech-

niques and recipes is typical for this book, although it certainly is not typical for FDA.

In Chapter 6 things get more serious. The chapter is supposedly about factor analysis (FA), modern FA even, but actually it turns out that modern FA is just old-fashioned principal components analysis (PCA). The chapter gives basic equations and some geometry, but nothing really beyond Pearson's or Hotelling's basic papers. PCA continues in Chapter 7, where it is applied to various standardizations of the data matrix. By now it is probably well known that in the psychometric tradition, after the data analyst has done the PCA she sits down to *interpret* the solution. This mysterious activity amounts to giving names to the components. It is considered very important and somewhat of an art. It is needed because of the high dimensionality of the usual PCA and FA solutions. In FDA there is a strong emphasis on two-dimensional solutions, one of its more redeeming features. We do not want to *explain* all the variance, nor do we want to extract all *interpretable* variation. We just want to make a two-dimensional picture of the multidimensional variation. Pictures are descriptive statistics, particular ways of looking into the high-dimensional space. The examples in Section 7.8 clearly illustrate this. They also illustrate that the interpretation does not proceed by naming axes, but rather by looking for clusters of points. Also, in FDA we do not emphasize the scalar product, which is difficult to visualize, but emphasize distance wherever possible.

Chapter 7 also illustrates another characteristic of FDA: "The most fruitful part of factor analysis techniques is the possibility of introducing supplementary elements (variables or individuals) into factor graphics" (p. 143). Thus if we have a PCA biplot based on a certain number of individuals (rows) and variables (columns), then we can fit in additional rows and columns just by correlating them with the principal components. Variables that are not in the analysis can be given component loadings and can be used for interpretation. FDA uses this mainly to fit background variables into the plots. These background variables (age, gender, income) and their interrelations are often not the main object of study, but in PCA they sometimes tend to dominate the solution if they are included. This can mean, for instance, that the major effect that we find in our expensive social science study is that older people have a shorter life expectancy and more free time. By not including the demographic variables in the analysis, but using them only in the plots, this does not happen. Systematic use of supplementary elements is a useful tool, no doubt, but nothing really spectacular.

Chapters 8 and 9 discuss 2-D and N-D correspondence analysis (CA), where the D stand for the number of variables. This takes up 140 pages—about a third of the book. Chapter 8 starts with a tantalizing brief history of the Benzécri movement. It then proceeds with a clear, dry, pedestrian account of the equations and the geometry of CA. Next is a murky section on interpretation, which presents rules to translate numbers into statements about *importance* and *explanation.* Practitioners often ask about such rules, but data analysts should not give in so easily. The picture should speak for itself as much as possible. On page 203 we find: "The most fruitful part of correspondence analysis techniques is to introduce supplementary elements onto factor graphics." Again! It is true that in CA plotting in supplementary rows or columns is quite natural because of the centroid interpretation, but to call it "the most fruitful part" is an exaggeration. The technique discussed in Chapter 9 is known as multiple correspondence analysis (MCA), but also as Guttman's PCA of categorical data, and as homogeneity analysis. Jambu does not tell his readers that MCA and CA have been around for many years, long before Benzécri's first publications, and that the techniques were used for actual data analysis by Fisher, Johnson, Guttman, Maung, Lord, Hayashi, and many others. Because of obvious computational problems, their use was limited, and they never got the opportunity to move to center stage as they did in France. Some of the history is in Benzécri (1982); more is in Gifi (1990).

This book has a lot of rules—rules for selecting significant axes, rules for selecting explicative points, and so on. This gives it a "cookbooky" and somewhat authoritarian aftertaste. The sections about the typical shapes of point clouds that can show up in CA are very brief and completely atheoretical. When and why horseshoes (parabolic clouds) can be expected is not explained. The section of acceptable data sets gives a list of types of matrices that can be fit into CA, sometimes by some recoding. An important role in FDA recoding is played by "dédoublement," here translated as doubling. If one has a variable (column) x taking values between 0 and K, then the variable $K - x$ also is included in the analysis. Consequently, all rows of the data matrix add up to the same constant, and the components from CA will be in deviations from the mean. Jambu mentions doubling, he tells the reader how and when to do it, but he does not tell the reader why to do it.

Chapters 10 and 11 provide a short course in cluster analysis. By now the author's pattern is painfully clear. Chapters have three main sections. The first section is a catalogue of the basic formulas, presented in tedious detail. The second is a set of practical rules, telling us when to do it and how to look at it. The third section presents the examples. If you are like me, you can look at such examples for a long time, marvel at them, find interesting nooks and crannies, and feel the temptation to go to the scanner and do some decent analysis. The Appendix gives many of the data sets in a fairly complete form. Although this is certainly a good aspect, it is not enough to make it a good book. If it was, then Andrews and Herzberg (1985) would be the best statistics book ever written.

The bible of FDA is the two-volume set published by Benzécri (1973a, b). One volume is about taxonomy and classification, the other is about correspondence analysis. Again, this is reflected in *Exploratory and Multivariate Data Analysis.* The two techniques seem to live together quite independently, without any interaction and without much of an idea why they happen to be the chosen ones. The book provides recipes for writing down lists of formulas, recipes for writing algorithms, and recipes for interpreting the results. Chapter 12, the ultimate synthesis, translates the recipes into corresponding computer programs in the DACL. This library was put together in 1976, the heyday of FDA, and nothing much seems to have been added to it in last 15 years.

In summary, FDA consists of CA and a series of cluster analysis techniques. It has been applied to numerous interesting examples, and using CA as a basic technique has yielded many very interesting extensions and adaptations. Gifi (1990) showed that classical multivariate descriptive analysis can be fit quite painlessly into multiple correspondence analysis. The philosophy of FDA, as expounded mainly by Benzécri himself, is interesting but very controversial. The class of techniques that have been studied systematically and developed in great detail is quite limited. FDA has been very clever in fitting time series data, survival analysis data, and many other forms of data into the CA framework in the same way that some like to pound on data until they can be used as input for LISREL or GLIM. In many cases these efforts are too Procrustean; we really need tailor-made techniques that adapt much better to the properties of the data and the design. FDA has produced a huge amount of very valuable research, most of it untranslated, and the current trend is to integrate FDA with other areas of statistics.

Jambu's book is disappointing. In introductory statistics teaching we feel that the computer has liberated us from the necessity to teach uninteresting formulas and uninspiring recipes. We certainly don't want the formulas and recipes that we just got rid of to show up in our graduate courses. FDA, if anything, is a class of methods that requires expertise and creativity in the handling of real data. Without the creativity, the discussion, and the controversy, it degenerates to a surprisingly dreary and uninteresting bunch of formulas, rules, and plots.

JAN DE LEEUW
University of California, Los Angeles

REFERENCES

Andrews, D. F., and Herzberg, A. M. (1985), *Data. A Collection of Problems from Many Fields for the Student and Research Worker,* New York: Springer-Verlag.
Benzécri, J.-P. (1973a), *L'Analyse des Données. Volume 1: La Taxinomie,* Paris: Dunod. France.
——— (1973b), *L'Analyse des Données. Volume II: L'Analyse des Correspondence,* Paris: Dunod.
——— (1982), *Histoire et Préhistoire de l'Analyse des Données,* Paris: Dunod.
——— (1991), *Correspondence Analysis Handbook,* New York: Marcel Dekker.
Bouroche, J.-M., and Saporta, G. (1980), *L'Analyse des Données. Que sais je,* Paris: Press Universitaire de France.
Cailliez, F., and Pagès, J.-P. (1976), *Introduction à l'Analyse des Données,* Paris: SMASH.
Fénélon, J.-P. (1981), *Qu'est-ce que l'Analyse des Données?* Paris: LEFONEN.
Gifi, A. (1990), *Nonlinear Multivariate Analysis,* Chichester, U.K.: John Wiley.
Gnanadesikan, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations,* New York: John Wiley.
Greenacre, M. (1984), *Theory and Applications of Correspondence Analysis,* London: Academic Press.
Jambu, M., and Lebeaux, M. O. (1983), *Cluster Analysis for Data Analysis,* Amsterdam: North-Holland.
Lebart, L., Morineau, A., and Warwick, K. M. (1984), *Multivariater Descriptive Statistical Analysis,* New York: John Wiley.
Yule, G. U., and Uendall, M. G. (1930), *Introduction to the Theory of Statistics,* London: Griffin.

Nonparametric Function Estimation, Modeling, and Simulation.

James R. Thompson and Richard A. Tapia. Philadelphia: Society for Industrial and Applied Mathematics, 1990. xvi + 304 pp. $32.50.

This book provides a view of nonparametric function estimation through a maximum likelihood window. The first part of the monograph represents a return to print of the classic Tapia and Thompson (1978) treatise on density estimation. New material by Thompson has been added on other function estimation problems and modeling, which prompted the change in title and priority of authorship.