Journal of Classification 2:147-150 (1985)

B. S. Everitt and G. Dunn, *Advanced Methods of Data Exploration and Modelling.* London: Heinemann Educational Books Ltd., 1983, pp. 253.

This is a book about multivariate analysis, covering much more material than other multivariate analysis books written at the same level. It is special basically because the authors emphasize *exploratory* methods for multivariate data analysis, and indeed emphasize the exploratory aspects of methods that are usually described as *confirmatory.* A large part of the book discusses principal components analysis, multidimensional scaling, and cluster analysis. This part takes up approximately 40% of its contents. Another 30% deals with (generalized) linear models, 10% with structural covariance modeling, and a final 10% with mathematical preliminaries. The book is not difficult, and is eminently suited for students who have been exposed previously to the rituals of the usual statistics courses. It is a very valuable antidote against the heavy emphasis on significance testing and exact sampling distributions, both of which make elementary statistics courses and advanced multivariate analysis books of very limited practical value. The book is in paperback, and it is not expensive.

A useful way to discuss multivariate analysis books has been applied previously by De Leeuw and Ramsay (1983). They applied correspondence analysis to a content analysis of 20 prominent books, and found that the books could be described well as convex combinations of classical inferential statistics, psychometric correlational analysis, and linear algebra. They can be located in the plane, by using the three components as a triangular coordinate system. If we locate the book by Everitt and Dunn in this triangle, we find that the text is close to Gnanadesikan (1977) and Kendall (1975), not too far from the centroid of the triangle. This representation also suggests that it does not fit very well into the plane of the other books, which is not surprising because structural covariance modeling, generalized linear models, multidimensional scaling, and exploratory graphical techniques are comparatively modern methods that do not get much attention in the older texts. More recent books, which are much closer to Everitt and Dunn, are Mardia, Kent and Bibby (1979), O'Muircheartaigh and Payne (1977), and Barnett (1981).

Readers of the *Journal of Classification* will meet many old acquaintances in this book. Of course, we must realize that it is meant for students and for content-oriented research workers, not for methodological specialists. But nevertheless many specialists could profit from the interesting comparisons between the various classes or techniques, and, above all, from the free and undogmatic spirit of the book. Significance tests are used freely

Reviewer's Address: Jan de Leeuw, Department of Data Theory, University of Leiden, Middelstegracht 4, 2312 TW Leiden, The Netherlands.

in an exploratory way, to aid in the model choice problem, and they are nicely supplemented by modern graphical tools. The decision theoretical orientation of much of modern mathematical statistics is avoided. Everitt and Dunn have included many interesting examples, to illustrate the various steps in the more common multivariate analysis techniques. There are not many exercises, but most of them are eminently useful practical elaborations of points made in the text, with additional interesting examples. The book does not pay any attention to computing and algorithms, taking the very reasonable point of view that users of the book will have a computer with the relevant packages available.

We shall briefly review the contents of the twelve chapters, paying a bit more attention to those topics directly relevant to readers of the *Journal of Classification*. Part I, *Approaches to Analysing Data*, consists of two chapters. The first chapter sets the scene. It discusses briefly multivariate data and similarity data, and the basic types of variables in the case of multivariate data. It then introduces the distinction between exploratory and confirmatory, and subsequently relativizes this distinction again. Models are introduced in the context of 'data = model + residuals'. Thus, models are used to smooth data, and not as a priori's as in decision theory. The second chapter gives some mathematical and statistical background. A few matrices, some statistical concepts, the method of maximum likelihood, and some useful remarks on optimization. Very little mathematics and statistics is required. It is somewhat unfortunate that no reasons are given why the maximum likelihood method has such a prominent place, not only in classical statistics, but actually also in this chapter of the book.

Part II, *Exploring Multivariate Data*, consists of four chapters. The first one, *Preliminary Data Analysis*, treats the arrangement of tables, probability plotting, Andrews Plots, Chernoff Faces, and so on. The chapter gives useful entries into the literature of this somewhat esoterical data analysis discipline. The second chapter of this part discusses *Principal Components Analysis*. Principal components are introduced as linear combinations of the observed variables, that successively account for as much variance as possible. This is a somewhat limited view of the technique, and in this chapter we find perhaps the most serious omissions of the book. Principal components analysis can be described, alternatively, as a technique of fitting a bilinear model to a multivariate table. This brings it much closer to multidimensional scaling, and it also is the natural way to introduce the biplot and correspondence analysis. Everitt and Dunn mention these two types of 'principal component plots', but do not discuss them in any detail. Although we are certainly biased, we think this is unfortunate. Correspondence analysis is possibly the most useful multivariate plotting technique available at the moment, and it has close links with multidimensional scaling and log-linear analysis. Of course, in a book of this size and of this level, some omissions are unavoidable, and perhaps it is a bit early yet to give

correspondence analysis a prominent place. The same thing is probably true for the recent generalizations of principal components analysis to multiway matrices. Chapter 5, the third one in this part, gives a highly satisfactory introduction to multidimensional scaling. Classical scaling of the Torgerson-Gower type, nonmetric scaling, three-way scaling, and Gower's techniques for asymmetric matrices all get just the right amount of attention. Choice of the 'number of dimensions' and 'power of the Minkovski metric' are discussed, unfortunately, as if they were serious problems. This means a lapse into the decision-oriented type of data analysis again, at a most unlikely place. Recent statistical and algorithmic developments (associated with programs such as MULTISCALE, ALSCAL, SMACOF) are not discussed yet, but this is perhaps just as well. The final chapter in Part II is on *Cluster Analysis*. It is, not surprisingly, very well done. The emphasis is, quite naturally, on agglomerative hierarchical techniques, but mixture models are also discussed in some detail. Graphical aids are used to illustrate clustering, and the usefulness of cluster analysis is evaluated quite thoughtfully. In summary we can say that, although some of our personal favorites are missing, this part of the book is very useful, and quite unique in a general multivariate analysis text.

Part III has the somewhat unfortunate title *Explanatory and Response Variables*. It contains five chapters, of which the first is called *The Generalised Linear Model*. This would be a much better title for the whole of part III. The chapter is a brief, but clear, introduction to the basic GLIM approach of Nelder and his co-workers. Chapter 8, *Regression and the Analysis of Variance*, treats the linear model properly. It is not very spectacular, but it has an interesting subsection on model selection procedures. In the spirit of the rest of the book, the analysis of residuals gets much attention. In chapter 9, *Regression Models for Categorical Data*, we find the log-linear model and its relatives. There is too much emphasis on maximum likelihood and on estimation in this chapter, at least for our taste, but in the examples Everitt and Dunn show how likelihood ratio statistics can be used nicely in model selection. Chapter 10, *Models for Time-dependent Data*, is a little jewel. It is mainly limited to discrete variables, but it discusses (in 14 pages!) latency models, survival functions, hazard models, and the Cox regression model. Including a chapter on these topics is, indeed, a very good idea. Although much of part III is perhaps not directly relevant for classification, we heartily recommend it.

Part IV, *Latent Variable Models*, has two chapters. The first one is on *Factor Analysis*, the second one on *Structural Equation Models*. Two important omissions here are general latent structure analysis, and latent trait theory. We think the first omission is unfortunate, because the general model makes it easy to introduce factor analysis along with latent class analysis, and logit and probit models such as the one discussed recently by Bartholomew (1980) and Muthén (1979). Of course, factor analysis is a

tricky subject. A number of bad habits from the early days still survive, the claims made for the technique are still quite grotesque, and time and again technical issues are transformed into philosophical or even religious ones. Everitt and Dunn succeed in steering clear of most of these controversies, and at the end of the chapter the authors give a reasonable and moderate evaluation of the technique. Nevertheless some of the rituals they discuss, mainly the ones having to do with rotation, are indeed hard to take. We would have preferred a somewhat more critical discussion. This is even more true for the next chapter. Everyone who is even remotely familiar with the literature in sociology or educational research must be aware of the great dangers and growing misuse of so-called 'causal models'. A balanced and timely evaluation of the situation is Cliff (1983). Everitt and Dunn have a rather weak warning at the end of the chapter, but the 14 pages abound with complicated and impressive looking path diagrams, with or without latent variables. We think that in this part Everitt and Dunn tend to describe and perhaps condone and encourage some very doubtful current practices. In this sense we think that part IV is, by far, the least satisfactory part of the book.

It is probably clear from our review so far, that we like this book a lot. In fact, we think it is perhaps the best book on 'applied multivariate analysis' for students and researchers in the social and behavioral sciences. Specialists will find many nice examples in this book, and perhaps some chapters outside their specialty which are good reading. This text is a very useful, well written book for almost everybody

University of Leiden Jan de Leeuw

References

BARNETT, V.A. (ed.) (1981), *Interpreting Multivariate Data*, New York: Wiley.

BARTHOLOMEW, D.J. (1980), "Factor Analysis for Categorical Data" (with discussion), *Journal of the Royal Statistical Society, B42*, 293-321.

CLIFF, N. (1983), "Some Cautions Concerning the Application of Causal Modeling Methods," *Multivariate Behavioral Research, 18*, 115-126.

DE LEEUW, J., and RAMSAY, J.O. (1983), Review of F. Calliez and J.P. Pagès, "Introduction à l'Analyse des Données," *Psychometrika 48*, 147-152.

GNANADESIKAN, R. (1975), *Statistical Data Analysis of Multivariate Observations*, New York: Wiley.

KENDALL, M.G. (1975), *Multivariate Analysis*, London: Griffin.

MARDIA, K.V., KENT, J.T., and BIBBY, J.M. (1979), *Multivariate Analysis*, New York: Academic Press.

MUTHEN, B. (1979), "A Structural Probit Model with Latent Variables," *Journal of the American Statistical Association, 74*, 807-811.

O'MUIRCHEARTAIGH, C.A., ad PAYNE, C. (1977), *The Analysis of Survey Data*, New York: Wiley.