Discrete multivariate analysis has also witnessed a phenomenal growth of literature in the past two decades and there are by now several texts covering the main developments in this area in an uptodate fashion. However, there appears to be a different picture with the development of mixed multivariate models involving partly continuous and partly discrete variables. These models arise often in applications and it would be a natural expectation to see an adequate treatment of the theory in a text. This may be a glaring omission in Dillon and Goldstein! Another important area with an outstanding growth of literature in the past fifteen years is the so-called variance component models. Frankly, I expected a more detailed treatment of this important topic in Anderson's second edition, and I am to a greater extent disappointed to see an inadequate treatment of this topic in either of the two books reviewed by Mark Schervish. I would like to make a specific reference to the forthcoming book of Rao and Kleffe (1987) for an in-depth coverage of this important area. I expect a significant amount of applications of these models in various applied areas.

As I tend to draw an overview of modern multivariate statistical analysis, more and more, I feel the need for robust (if not nonparametric) methods. Although some of these methods (mostly, in the context of simple MANOVA models) have been treated adequately in some contemporary textbooks, I have no doubt in my mind that in the coming years, there will be a far reaching impact of this vital area in multivariate analysis.

To summarize, let me congratulate Mark Schervish for a job well done. In principle, I would have argued in favor of a modified title "A Review of Two Texts in Multivariate Analysis." The area of multivariate statistical analysis is indeed too vast to be covered entirely by these two (or, as a matter of fact, by any two) texts. However, Anderson's second edition will naturally help us in identifying the other pockets where an equally sound and lucid treatment of the theory (and methodology) should be developed in the form of a text, and once this has been accomplished, we are all set to close the whole area in the form of two texts. Until then, the second edition is a major step in the right direction.

ADDITIONAL REFERENCES

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.

GIRI, N. C. (1977). *Multivariate Statistical Inference*. Academic, New York.

LEBART, L., MORINEAU, A. and WARWICK, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

RAO, C. R. and KLEFFE, J. (1987). *Estimation of Variance Components and Their Applications*. North-Holland, Amsterdam.

TAKEUCHI, K., YANAI, H. and MUKHERJEE, B. N. (1982). *The Foundations of Multivariate Analysis*. Halsted, New York.

Comment

R. Gnanadesikan and J. R. Kettenring

In our experience, most statistical problems that arise in practice are genuinely multivariate in character. This is almost surely as true in other settings as it is in the telecommunications business that we work in. A recent literature search (Gnanadesikan and Kettenring, 1984) covering seven disciplines over the period 1965 to 1982 turned up 15,000 articles that involved multivariate methods.

It is natural, therefore, to expect that new books on the subject, such as those by Anderson and by Dillon and Goldstein, as well as comprehensive reviews, such as that of Schervish, will have a wide audience. However, our intention in this commentary is not so much to critique either the books or the review as it is to bring out some of our own views on multivariate data analysis.

In outlook, if not detail, these overlap with views of Schervish who makes many telling points about the state of multivariate analysis. The best known and most frequently used of the classical methods have not always served well and often leave the user with the question "What have I really learned about my data and how sure can I be about it?" Much of the elegant theory is of little practical value. Standard multivariate hypothesis tests, which have been so extensively developed (see Schervish's comments in

R. Gnanadesikan is Assistant Vice President of the Mathematical, Communications, and Computer Sciences Research Laboratory and J. R. Kettenring is Division Manager of Statistics and Economics Research at Bell Communications Research, 435 South Street, Morristown, New Jersey 07960.

Section 1), are infrequently used. This helps to explain why applied researchers, both in statistics and in other consumer disciplines, have developed such a variety of alternative numerical and graphical methods in the last 25 years.

In an overall sense, multivariate methods are useful to the extent that they can extract and explicate "structure" in high dimensional observations. In the "exploratory" phase of this search, it is seldom possible to be able to pose questions in the precise way that is required for formal inference. Hence, exploratory methods need to operate flexibly on the data.

Even in the "confirmatory phase," the methods should do much more than provide a specific answer to a tightly posed question. They should, in addition, indicate the adequacy of underlying assumptions, expose unanticipated structure (serendipity!) and peculiarities, and not be overly model-dependent (although to go to the opposite extreme of nonparametric inference is not necessarily the answer).

For both exploratory and confirmatory analyses, more methods are needed that can deal with real world departures from ideal conditions. Diagnostics that are as fully developed and effective as those presently available for linear regression would help considerably. For analyzing multiresponse data from designed experiments, graphical diagnostic tools have been available for over two decades (these are summarized in Gnanadesikan, 1977, Section 6.3) and should become routine parts of multivariate pedagogy. Similarly, methods that are robust against data idiosyncrasies, such as outliers, deserve even more attention than they have already been given. Being satisfied with methods that are robust in the sense that their behavior can be justified by the central limit theorem is unacceptable from the practical perspective.

Promisingly, one can point to numerous relatively recent developments in computer-intensive multivariate methods that move in the direction of providing a better selection of practically oriented tools. Examples include general purpose techniques such as the bootstrap, jackknife, m estimation and dynamic graphical display systems and more focused ones such as ACE (Breiman and Friedman, 1985), CART (Breiman, Friedman, Olshen and Stone, 1984) and projection pursuit (Jones and Sibson, 1987, and references therein).

In fact, multivariate analysis appears to be entering a revolutionary stage where the limitations associated with classical procedures and their offspring are no longer necessary ones. These developments are being driven by the tremendous computational and graphical power that is here today—and more is coming! Time will be needed not only to complete this revolution but also to filter out the more effective of these new tools and provide them with firmer foundations and levels of understanding. All of this suggests that the interesting theory and relevant practice of multivariate analysis will be (or at least badly needs to be) quite different in a few years than it appears in today's books.

Still, there is no need to await the denouement of this revolution. Already key tools are in hand that move us well beyond the limited capabilities of numerically oriented batch computing and "canned" analyses. The critical "technologies" are: highly interactive computing, with each step of the data analysis indicating what the next step should be, and rudimentary statistical graphics. Most of the graphical displays one needs are in the form of two-dimensional plots of some sort. The plotted quantities and coordinate systems may themselves be the product of standard numerical machinery, for example, a scatter plot of the data in the space of the first two discriminant variables in a discriminant analysis. Indeed, such primitive plotting capabilities are so powerful, basic and well established for multivariate data analysis that they overshadow in importance many other topics on the subject.

Even in a completely conventional presentation of multivariate analysis, we should be able to do more in the way of capturing the underlying principles and mathematical concepts. For example, the fundamental role of the singular value decomposition is seldom exploited. Another example is the interesting alternative motivation of principal components mentioned by Schervish in Section 8. He points out that the first principal component for a set of standardized variables is the linear function of them with the highest sum of squared correlations with the individual members of the set. In fact, there is no need to limit oneself to linear functions of the variables: the first principal component is the best choice among all possible variables according to this criterion. Moreover, this formulation leads to an underlying statistical model that is not only useful for explaining principal components, but also motivates the maximum variance method of canonical correlation analysis for several sets of variables (Kettenring, 1971).

In conclusion, both the "science" and the "technology" of multivariate data analysis have been and are evolving rapidly, and it is time that pedagogical concerns and tools (such as books) reflect these developments adequately if they aim to be relevant and exciting!

ADDITIONAL REFERENCES

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple correlation and regression (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, Calif.

GNANADESIKAN, R. and KETTENRING, J. R. (1984). A pragmatic review of multivariate methods in applications. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds). Iowa State Univ. Press, Ames.

JONES, M. C. and SIBSON, R. (1987). What is projection pursuit (with discussion)? *J. Roy. Statist. Soc. Ser. A* **150** 1–36.

KETTENRING, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika* **58** 433–451.

Comment

S. James Press

I thoroughly enjoyed Mark Schervish's review of multivariate analysis, a subject that has been near and dear to me for many years. The review was written in a very light, free-flowing format that made it interesting and pleasant reading, while at the same time the points made were usually deep and insightful. I will comment generally on the Schervish review by offering my own perspectives on multivariate analysis, and then I will give a few brief specifics on his review. All comments will necessarily be brief but indicative of directions in which the field is moving.

A COMPARISON OF CLASSICAL AND MODERN MULTIVARIATE ANALYSIS

I would like to distinguish "classical" multivariate analysis (CMA) from "modern" multivariate analysis (MMA). I will do so on the basis of how they compare on various (randomly ordered) characteristics.

1. *Distribution theory.* In CMA, the theory derives largely from the multivariate normal and Wishart distributions. It also is concerned with the study of the distribution of latent roots of random matrices.

In MMA there is increasing focus on non-normal inference and distribution theory. It is based upon nonabsolutely continuous distributions, such as the mixed discrete and continuous distributions, or the mixed singular and absolutely continuous distributions, exemplified by the multivariate exponential distribution. Focus has shifted away from the latent root distributions because the models that require them have languished for lack of use.

2. *Estimation.* In CMA, the emphasis was on MLE and moment estimation. In MMA there has been a substantial shift in emphasis to Stein-type estimation, empirical Bayes estimation and Bayes estimation. This shift is natural with the improvements in multidimensional estimation achievable by using higher dimensional shrinkage estimators (for dimension greater than two) and by introducing subjective prior information into a problem in a formal way.

3. *Noncentral distributions.* In CMA, power calculations demanded the development of various noncentral distributions, such as the noncentral Student t and noncentral F distributions, the Hotelling T_0^2 distribution and the noncentral Wishart distribution, which arose in coefficient estimation for simultaneous equation systems.

In MMA a unified theory of noncentral distributions has developed around the theory of hypergeometric functions of matrix arguments, zonal polynomials and generalized distributions.

4. *Distribution theory of sample estimators.* CMA was deeply concerned with the distribution theory of sample estimators, although the introduction of the "bootstrapping" technique (Efron) and the technique of simulating complicated multivariate distributions by simulating functions of known distributions (Kass) have liberated modern multivariate analysts from their former distributional burdens of having to develop the distributional theory of complicated multivariate distributions.

5. *Discrete multivariate analysis.* CMA dealt with discrete data by means of traditional contingency table analysis, i.e., estimating cell probabilities by MLE.

MMA is more concerned with analyzing discrete data by using multivariate log-linear and logistic models; by using models involving ordered categories and by using both dimensions of a contingency table simultaneously to study categorical data, by means of "correspondence analysis."

6. *Factor analysis.* CMA was wary of the factor analysis approach and was concerned with centroid solutions, rotations, maximum likelihood factor analysis and exploratory factor analysis (rather than confirmatory).

MMA has become more accepting of the factor analysis approach. Today the emphasis has shifted to confirmatory factor analysis, Bayesian factor analysis methods and to nonparametric factor analysis

S. James Press is Professor, Department of Statistics, University of California, Riverside, California 92521.