# A Review of Multivariate Analysis

**Mark J. Schervish**

A survey of topics in multivariate analysis inspired by the publication of T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley & Sons, New York, 1984, xvii + 675 pages, \$47.50, and WILLIAM R. DILLON and MATTHEW GOLDSTEIN, *Multivariate Analysis: Methods and Applications*, John Wiley & Sons, New York, 1984, xii + 587 pages, \$39.95.

This review and discussion are dedicated to the memory of P. R. Krishnaiah, a leader in the area of Multivariate Analysis, who died of cancer on August 1, 1987.

## 1. INTRODUCTION

It has been a long time coming, but it is finally here. The second edition of T. W. Anderson's classic, *An Introduction to Multivariate Statistical Analysis*, will please all of those who have enjoyed the first edition for so many years. It essentially updates the material in the first edition without going far beyond the topics already included there. A reader who had spent the intervening 26 years on another planet might get the impression that work in multivariate analysis has been concentrated on just those topics with the addition of factor analysis. Of course this impression is mistaken, and Anderson himself notes in the Preface (page vii) that "It is impossible to cover all relevant material in this book." So, in the course of reviewing this book, and comparing it to the first edition, I thought it might be interesting to take a thoroughly biased and narrow look at the development of multivariate analysis over the 26 years between the two editions. A reader interested in a more complete and less personalistic review might refer to Subramaniam and Subramaniam (1973) and/or Anderson, Das Gupta and Styan (1972). Recent reviews of some contemporary multivariate texts (less cluttered by reviewer bias) were performed by Wijsman (1984) and Sen (1986).

Suppose we begin at the end. Nearly simultaneous with the publication of the second edition of Anderson's book is the release of *Multivariate Analysis* by Dillon and Goldstein (the Prefaces are dated June and May 1984, respectively). This text, which is subtitled *Methods and Applications*, is different from Anderson's in every respect except the publisher. It even seems to begin where Anderson leaves off with factor analysis and principal components. I believe that the

*Mark J. Schervish is Associate Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.*

differences between the texts reflect two very different directions in which multivariate analysis has progressed. The topics covered by Dillon and Goldstein have, by and large, been developed more recently than those covered by Anderson. As an illustration, fewer than 18% of the references cited by Dillon and Goldstein are pre-1958, whereas almost 42% of Anderson's references are pre-1958. (Of course Anderson had a headstart, but the other authors had access to his 1958 book. In three places, they cite Anderson's 1958 book in lieu of earlier work.) The major difference in emphasis is between theory and methods. To illustrate this distinction, Anderson had twelve examples worked out with data in his first edition and the same examples appear in the second edition, with no new ones (but one correction). This is due, in large part, to the fact that the topics covered in the two editions are nearly identical. (Although factor analysis has been added as a topic, no numerical examples are given, and no numerical exercises are included.) Dillon and Goldstein work out numerous examples, often reanalyzing the same data several times to illustrate the differences between various techniques.

Since 1958, the development of multivariate theory has been concentrated, to a large extent, in the general areas that Anderson covered in his first edition. Multivariate methods, on the other hand, have taken on a life of their own, with or without the theory that mathematical statisticians would like to see developed. This has led to an entire industry of exploratory and ad hoc methods for dealing with multivariate data. Researchers are not about to wait for theoreticians to develop the necessary theory when they perceive the need for methods that they think they understand. The theoretical statisticians' approach to multivariate analysis seems to have been to follow the first principle of classical inference. "If the problem is too hard, test a hypothesis." The development of procedures like cluster analysis, factor analysis, graphical