Ira H. Bernstein

Applied Multivariate Analysis

With Calvin P. Garbin
and Gary K. Teng

Springer-Verlag

Ira H. Bernstein

Applied Multivariate Analysis

With Calvin P. Garbin
and Gary K. Teng

With 37 Illustrations

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo

Ira H. Bernstein
Department of Psychology
University of Texas at Arlington
Arlington, Texas 76019
USA

Gary K. Teng
Technical Evaluation and Management Systems, Inc. (TEAMS®)
Dallas, Texas 75240
USA

Calvin P. Garbin
Department of Psychology
University of Nebraska at Lincoln
Lincoln, Nebraska 68588
USA

Library of Congress Cataloging-in-Publication Data
Bernstein, Ira H.
Applied multivariate analysis.
Bibliography: p.
Includes index.
1. Multivariate analysis. I. Garbin, Calvin P.
II. Teng, Gary K. III. Title.
QA278.B457 1987 519.5′35 87-15174

Typeset by Asco Trade Typesetting Ltd., Hong Kong.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, Virginia.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-96542-4 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96542-4 Springer-Verlag Berlin Heidelberg New York

To Linda, Cari, Dina
and
the memory of Jum C. Nunnally

Preface

Like most academic authors, my views are a joint product of my teaching and my research. Needless to say, my views reflect the biases that I have acquired. One way to articulate the rationale (and limitations) of my biases is through the preface of a truly great text of a previous era, Cooley and Lohnes (1971, p. v). They draw a distinction between mathematical statisticians whose intellect gave birth to the field of multivariate analysis, such as Hotelling, Bartlett, and Wilks, and those who chose to "concentrate much of their attention on methods of analyzing data in the sciences and of interpreting the results of statistical analysis .... (and) ... who are more interested in the sciences than in mathematics, among other characteristics."

I find the distinction between individuals who are temperamentally "mathematicians" (whom philosophy students might call "Platonists") and "scientists" ("Aristotelians") useful as long as it is not pushed to the point where one assumes "mathematicians" completely disdain data and "scientists" are never interested in contributing to the mathematical foundations of their discipline. I certainly feel more comfortable attempting to contribute in the "scientist" rather than the "mathematician" role.

As a consequence, this book is primarily written for individuals concerned with data analysis. However, as noted in Chapter 1, true expertise demands familiarity with both traditions.

One consequence of my bias is that even though I have a great love for data, I have long since learned not to worship a particular data set, i.e., I believe in sampling error and replication. I especially believe in Henry Kaiser's (1970) aphorism "It don't make no nevermind" about highly elaborate weighting schemes that more often than not prove less useful than simpler ones such as weighting variables equally. (Kaiser, by the way, shows the limitations of the "rule of thumb" distinction between Mathematician/Platonists and Scientist/Aristotelians; bright people do both as his numerous other contributions such as the varimax rotation and alpha factoring attest.) If you are unsure as to what I mean by "magical equations," pick up almost any scholarly journal containing path analyses and confirmatory factor analyses. Similarly, it should not be too difficult to locate a book in which the author attempts to use the force of copyright law to protect the weighting scheme used in an equation.

The major reason that I refer to Cooley and Lohnes (1971) as belonging to a prior generation is that it was written before the ascendancy of the major computer packages. I assume you will be using these packages. At one point in writing this book, I thought of offering a "translation" table to explain the printouts that are provided by the packages. The reason that I did not devote more space to the topic is the rate at which revisions of the packages make such information obsolete.

The one feature that I hope sets this book apart from the many other excellent books in the field is my use of the computer as a device to teach you about data structures. It is all too easy to see only the data analytic features of computer packages, as wonderful as they are. (Consider that in almost no time at all I can enter the coding to do all variants on a particular analysis that would have been prohibitive in terms of time when I was a student.) Much less apparent to the student is the way that a computer can *generate* data to conform to a model or deviate in specified ways. Scholars have long used Monte Carlo and related simulation approaches; I feel that it is sufficiently easy for students at this level, or even at a more introductory level, to perform simulations so that they should be part of all students' training.

The material in this book is designed to be covered in a standard one semester graduate course. It is assumed that a student has had a conventional first semester graduate class in statistics. I have, however, devoted a chapter to material that is largely a summary of this core material (Chapter 2). Of course, I have also tried to include useful information that would make the book valuable beyond the formal confines of the course (*lagniappe*, as one would say in the language of those to my Southeast).

Acknowledgments

I would like to thank Springer-Verlag for their faith in this project. Robert Widner's comments on various points of the text were most highly appreciated. Various ideas were outgrowths of conversations I had with various departmental colleagues, especially Jim Nairne, Jim Erickson, Jim Baerwaldt, Bill Ickes, Paul Paulus, Duane Martin, and the ophthalmologists at the Glaucoma Associates of Texas. I also am grateful to Melvin Pierce, Tom Kennedy, Ed Homko, Michael Griffin, and others at the University of Texas Academic Computer center, Rory Gresham of Seabrook Computers, and Sara Kapka of IBM for their technical assistance. The oblique, but nonetheless vital contributions of Emma Peel, Ferdinand LaMenthe, John B. Gillespie, John Coltrane, Charles C. Parker, Jr., Larson E. Whipsnade, Jonathan Steed, John Sheridan, and James Cullum and associates are also noted. Finally, the wisdom of a comment by my long-term friend and colleague, Professor Stanley

Coren of the University of British Columbia that "you never finish writing a book, they just take it away from you" is hereby duly noted.

Arlington, Texas
May 1987

IRA H. BERNSTEIN

Contents

1
Introduction and Preview

Chapter Overview

Chapter 1 contains six major topics:

I. MULTIVARIATE ANALYSIS: A BROAD DEFINITION—I begin by offering a very broad definition of multivariate analysis. The definition is: *an inquiry into the structure of interrelationships among multiple measures.*

II. MULTIVARIATE ANALYSIS: A NARROW DEFINITION—The broad definition is important because virtually all behavioral research deals with questions about structure. Unfortunately, it would subsume all of statistical analysis, making for quite a large textbook! Consequently, it is customary to define multivariate analysis in more limited terms: *the study of linear representations of relations among variables.* Contained within this narrow definition are several interrelated models that are discussed in later chapters: (1) *multiple regression* (Chapters 4 and 5), (2) *factor analysis* (Chapters 6 and 7), (3) *discriminant analysis, related classification techniques, and the multivariate analysis of variance and covariance* (Chapters 8–10), (4) *profile and canonical analysis* (Chapter 11), and (5) *analysis of scales* (Chapter 12).

III. SOME IMPORTANT THEMES—Models like factor analysis and multiple regression used to infer structure are all too often looked upon as if they were separate, unrelated procedures. In reality, certain themes recur regardless of the specific analytic procedure: (1) *equations that are the by-product of a particular analysis should be meaningful*; (2) *cutoffs need to be established defining "how high is high" and "how low is low" when one develops a prediction equation*; (3) *questions of statistical significance often arise*; (4) *outliers can render otherwise elegant analyses meaningless or, worse, highly misleading*; (5) *multivariate investigation should be guided by theory*; and (6) *particular problems are likely to arise when relations among individual items, as opposed to multi-item scales, are the unit of analysis.*

IV. THE ROLE OF COMPUTERS IN MULTIVARIATE ANALYSIS—A generation ago, many analyses that were conceived of were not practicable. The availability of computers and sophisticated computer packages now make such analyses routine. The increasing role of the personal computer is also noted.

V. CHOOSING A COMPUTER PACKAGE—There are several important computer packages now on the market. Some of the pros and cons of the three most popular ones—SAS, SPSSX, and BMDP—are discussed.

VI. PROBLEMS IN THE USE OF COMPUTER PACKAGES—Knowing how to use a computer package is important to multivariate analysis, but it is not sufficient. Four necessary ingredients to a successful analysis are considered: (1) *substantive knowledge of the research topic*; (2) *computer knowledge, including knowledge of the package*; (3) *empirical experience with various kinds of data*; and (4) *formal knowledge of the analytic procedures.*

This textbook is written for those who need to analyze complex behavioral data, whether in field settings such as clinical and social psychology, in applied settings like nursing and marketing research, or in experimental settings like learning and perception. Although I cannot escape some degree of formality, I am writing for students whose primary interests lie more in empirical phenomena as opposed to rigorous mathematical statistics.

Although I very much appreciate why examples related to your specific interests are important, I hope I can "wean" you from examples limited to your own content area as the textbook topics progress. One of the things you should note is that if you are a clinical psychology student working with personality test data, you will have problems that are abstractly *identical* to a market researcher working on a consumer survey. One of the reasons that I find quantitative applications so rewarding is that I can work in a variety of areas, as long as I have someone to provide me with a background of the underlying empirical issue or have that background myself. At any one time, I can and have been working on the issue of selecting police officers, studying patients attitudes toward health care, looking at choices of financial institutions, and evaluating different forms of eye surgery. What I learned in one setting applied to all. I hope you can share the enjoyment that I have been experiencing.

One of the decisions that any author of a quantitative textbook needs to confront is how to choose examples. Conceptually, it should not matter if examples come from clinical psychology, marketing, or visual psychophysics; yet it does matter to most students. What I will do, therefore, is to pick my examples to illustrate the relevance of the various multivariate models in different settings.

Multivariate Analysis: A Broad Definition

As with most topics, it is useful to try to define what is meant by the term "multivariate analysis" before proceeding too far into its details. In the broadest and most literal sense, it means *an inquiry into the structure of*