



Journal of Statistical Software

June 2017, Volume 78, Book Review 3.

doi: 10.18637/jss.v078.b03

Reviewer: Andreas Rosenblad
Uppsala University

Multivariate Statistical Methods: A Primer (4th Edition)

Bryan F. J. Manly, Jorge A. Navarro Alberto
Chapman & Hall/CRC, Boca Raton, 2016.
ISBN 978-1-4987-2896-6. 253 pp. USD 59.95 (P).
<https://www.crcpress.com/9781498728966>

Overview

Multivariate Statistical Methods: A Primer has as its stated purpose to introduce multivariate statistical methods to non-mathematicians, intending to keep details to a minimum but still convey a good idea of what can be done in the area of multivariate statistics. The prior three editions appeared in 1986 (159 pages), 1994 (232 pages), and 2004 (224 pages), with Manly as single author. The preface states that the main change since the third edition is the introduction of R code to do all of the analyses in the book. The reader is assumed to have a working knowledge of elementary statistical methods, i.e., significance testing using the normal, t , χ^2 , and F distributions, analysis of variance (ANOVA), and standard linear regression. Additionally, to fully benefit from the text, some facility with algebra is required, as is some knowledge about matrix algebra.

Chapter discussions

The authors' intention has been that each chapter, to some extent, could be read independently of the others. Most chapters are concluded with a discussion of computer programs and further reading, as well as giving one or more end-of-chapter exercises and references. Of the book's 13 chapters, the first five are intended to be preliminary reading, covering general aspects of multivariate data rather than some particular techniques. The first motivational chapter, "The material of multivariate analysis", describes five multivariate data sets from the areas of biology, archaeology, and economics, and provides a brief preview of the multivariate methods discussed in the following chapters and how these may be used to analyze these five data sets. An appendix gives an introduction to R, mainly focusing on subjects useful for performing multivariate analyses in R, such as the handling of vectors and matrices in R, and how to arrange multivariate data in data frames. The second chapter gives a short excursion into matrix algebra, supplemented with an appendix covering the main R functions

for matrix algebra, while the third chapter gives a brief overview of graphical methods for displaying multivariate data, such as scatter plot matrices, Chernoff faces, and profile plots, with the appendix discussing how to produce these plots in R. Chapter 4 is devoted to tests of significance for means and variations in multivariate data, starting with the single-variable case before proceeding to discuss the multivariate case, with the appendix showing how to perform these tests in R. A useful feature is that besides noting the assumptions that the tests rely on, the authors also discuss how robust the tests are to violations of the assumptions. Chapter 5 discusses the measuring and testing of multivariate distances, such as the Penrose and Mahalanobis distances and the Mantel randomization test, with the functions implementing the calculations and tests in R discussed in the appendix.

Chapters 6–12 are devoted to specific multivariate techniques, starting with principal components analysis in Chapter 6. After an easy-to-comprehend definition of the meaning of the term principal components and an overview of the procedure for a principal components analysis (PCA), the book proceeds with a couple of examples applying PCA techniques to data on body measurements of female sparrows and employment in European countries, focusing on the interpretation of the results. The appendix on PCA in R mainly discusses the use of the `prcomp()` function in R's default **stats** package. Chapter 7, although having the general title "Factor analysis", is devoted to exploratory factor analysis (EFA). The confirmatory factor analysis approach is only mentioned at the end of the chapter, with a brief discussion contrasting it to the EFA approach. The authors start by giving a pedagogic overview of the factor analysis model, motivated by Spearman's original example of the correlation between test scores for various school subjects, before describing the general procedure for a factor analysis, followed by an overview of principal components factor analysis. The authors then proceed to discuss how a computer program for factor analysis can be used to do principal components analysis, illustrated with an example using data on employment in European countries, followed by a discussion of available options in computer programs for factor analysis and the general value of using factor analysis methods. The appendix on factor analysis in R mainly discusses resources available in the **psych** package, comparing these to the `factanal()`, `prcomp()`, and `varimax()` functions in R's **stats** package.

Methods for discriminant function analysis is the topic of Chapter 8. After briefly discussing the general problem of separating groups, discrimination using Mahalanobis distances, canonical discriminant functions, tests of significance in the discriminant function analysis context, and the assumptions that the methods are based on, the authors proceed to apply discriminant function analysis methods to examples about the comparison of values for four measurements on five samples of male Egyptian skulls and discriminating between groups of European countries on the basis of employment patterns. This is followed by a few short sections that briefly discuss the possibilities of allowing for prior probabilities of group membership, stepwise discriminant function analysis, jackknife classification of individuals, and the assignment of ungrouped individuals to groups, before proceeding with a discussion of how to make use of logistic regression for discrimination between two groups, illustrated with examples about storm survival of female sparrows and comparison of two of the samples of Egyptian skulls. The appendix covering discriminant function analysis in R focuses on canonical discriminant analysis using the `lda()` function in the R package **MASS** and discriminant analysis based on logistic regression using the standard `glm()` function.

Chapter 9, covering cluster analysis, starts with giving an overview of the uses and usefulness of cluster analysis, the different types of cluster analysis, with special attention to hierarchic

methods, the inherent problems with cluster analysis, and distance measures used for hierarchical clustering algorithms. This is followed by the chapter's main section, devoted to using principal components methods with cluster analysis, applied to examples of the clustering of European countries on the basis of employment patterns and analyzing the relationships between canine species for samples of six living species of dogs and the remains of prehistoric dogs. The appendix on cluster analysis in R mainly discusses the `hclust()` function.

Chapter 10 introduces canonical correlation analysis from the starting point of generalizing a multiple regression analysis, describes the basic procedure for a canonical correlation analysis as well as tests of significance, and then proceeds with the main section on interpreting canonical variates, applied to examples on environmental and genetic correlations for colonies of a butterfly, and soil and vegetation variables in Belize, respectively. The appendix on canonical correlation in R focuses on the wrapper of the `cancor()` function implemented in the `candisc` package and the `cca()` function in the `yacca` package, which the authors recommend as being the functions that are most suitable for performing canonical correlation analyses in R.

Multidimensional scaling, the topic of Chapter 11, is motivated from the perspective of constructing a map from a distance matrix. The main part of the chapter is then devoted to procedures for multidimensional scaling applied to road distances between New Zealand towns, using classical nonmetric multidimensional scaling, and the voting behavior of New Jersey congressmen in the United States House of Representatives, using classical metric multidimensional scaling as well as classical nonmetric scaling. The appendix on multidimensional scaling in R mainly discusses the `isomds()` function from the `MASS` package.

Chapter 12 returns to the topics of principal components analysis and multidimensional scaling in the context of ordination, applied to examples of the abundance of plant species in the Steneryd Nature Reserve in Sweden and types of grave goods found in burials in the Ban-nadi Cemetery in northeast Thailand. The authors also discuss principal coordinates analysis applied to the same examples as well as correspondence analysis as a method of ordination using the Steneryd Nature Reserve data. The overview of ordination methods in R given in the appendix to Chapter 12 is mainly devoted to principal coordinates analysis using the `cmdscale()` function and correspondence analysis using the `ca()` function in the `ca` package. The book's last chapter is an epilogue giving some general advice and comments regarding applying multivariate statistical methods.

Conclusion

Multivariate Statistical Methods: A Primer has a fairly standard coverage of available multivariate statistical methods, but stands out in its presentation of these, which is concise, pedagogic, and easy to follow. Each chapter is fairly short, covering only the most essential details, using mathematical formulas only when it is necessary. The stated purpose of the book, to introduce multivariate statistical methods to non-mathematicians while keeping details to a minimum, but still conveying a good idea of what can be done in the area of multivariate statistics, is thus well fulfilled.

The book takes a practical approach to multivariate statistical methods, with illustrations utilizing real, varying data sets from different disciplines, thus making it useful for the applied statistician. It should be noted that the use of R is mainly limited to the appendices, and

that there is no specific R package accompanying the book. Instead, the authors make use of functions in R's default **stats** package when possible, or otherwise utilize packages developed by other authors. This works well. The R scripts for most of the examples in the text are available at the website accompanying the book, but it would have been useful if these had also been given in the text, together with the examples.

To summarize, this is a very nice book giving a concise not overly technical treatment of multivariate statistical methods that is highly recommended for anyone wanting to have an easy-to-understand overview of this important subject.

Reviewer:

Andreas Rosenblad
Uppsala University
Center for Clinical Research Västerås
Västmanland County Hospital
S-721 89 Västerås, Sweden
E-mail: andreas.rosenblad@regionvastmanland.se
URL: <http://orcid.org/0000-0003-3691-8326>