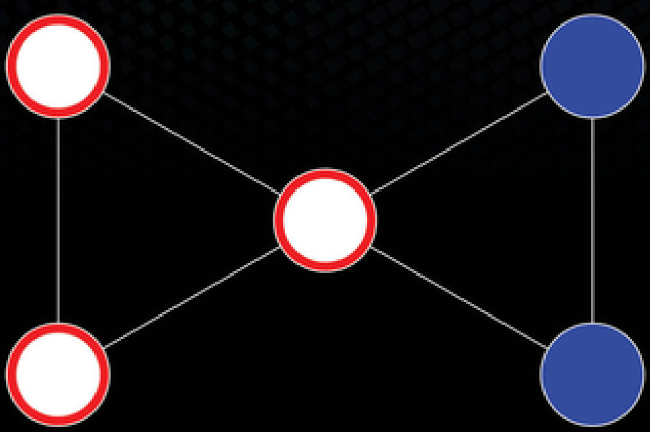
Wiley Series in Probability and Statistics

Multivariate Analysis

Second Edition



Kanti V. Mardia
John T. Kent
Charles C. Taylor

WILEY

Multivariate Analysis

Multivariate Analysis

Second Edition

Kanti V. Mardia, John T. Kent, and Charles C. Taylor

WILEY

This second edition first published 2024

Edition History
© 1st Edition 1979 Academic Press

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Mardia, K. V., author. | Kent, J. T. (John T.), joint author. | Taylor, C. C (Charles C.), joint author.
Title: Multivariate analysis / Kanti V. Mardia, John T. Kent, and Charles C. Taylor.
Description: Second edition. | Hoboken, NJ : Wiley, 2024. | Series: Wiley series in probability and statistics | Includes bibliographical references and index.
Identifiers: LCCN 2023002460 (print) | LCCN 2023002461 (ebook) | ISBN 9781118738023 (cloth) | ISBN 9781118892527 (adobe pdf) | ISBN 9781118892510 (epub)
Subjects: LCSH: Multivariate analysis.
Classification: LCC QA278 .M36 2023 (print) | LCC QA278 (ebook) | DDC 519.5/35–dc23/eng/20230531
LC record available at https://lccn.loc.gov/2023002460
LC ebook record available at https://lccn.loc.gov/2023002461

Cover Design: Wiley
Cover Images: © Liyao Xie/Getty Images, Courtesy of Kanti Mardia

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

*To my daughters **Bela** and **Neeta***

— *with Jainness* (Kanti V. Mardia)

*To my son **Edward** and daughter **Natalie***

(John T. Kent)

*To my wife **Fiona** and my children **Mike, Anna, Ruth, and Kathryn***

(Charles C. Taylor)

Contents

Epigraph *xvii*
Preface to the Second Edition *xix*
Preface to the First Edition *xxi*
Acknowledgments from First Edition *xxv*
Notation, Abbreviations, and Key Ideas *xxvii*

1 Introduction *1*
1.1 Objects and Variables *1*
1.2 Some Multivariate Problems and Techniques *1*
1.2.1 Generalizations of Univariate Techniques *1*
1.2.2 Dependence and Regression *2*
1.2.3 Linear Combinations *2*
1.2.4 Assignment and Dissection *5*
1.2.5 Building Configurations *5*
1.3 The Data Matrix *7*
1.4 Summary Statistics *8*
1.4.1 The Mean Vector and Covariance Matrix *8*
1.4.2 Measures of Multivariate Scatter *11*
1.5 Linear Combinations *12*
1.5.1 The Scaling Transformation *13*
1.5.2 Mahalanobis Transformation *13*
1.5.3 Principal Component Transformation *13*
1.6 Geometrical Ideas *14*
1.7 Graphical Representation *15*
1.7.1 Univariate Scatters *15*
1.7.2 Bivariate Scatters *16*
1.7.3 Harmonic Curves *18*
1.7.4 Parallel Coordinates Plot *19*
1.8 Measures of Multivariate Skewness and Kurtosis *20*
Exercises and Complements *22*

2 Basic Properties of Random Vectors *25*
Introduction *25*
2.1 Cumulative Distribution Functions and Probability Density Functions *25*

2.2 Population Moments *27*
2.2.1 Expectation and Correlation *27*
2.2.2 Population Mean Vector and Covariance Matrix *27*
2.2.3 Mahalanobis Space *29*
2.2.4 Higher Moments *29*
2.2.5 Conditional Moments *30*
2.3 Characteristic Functions *31*
2.4 Transformations *32*
2.5 The Multivariate Normal Distribution *34*
2.5.1 Definition *34*
2.5.2 Geometry *35*
2.5.3 Properties *36*
2.5.4 Singular Multivariate Normal Distribution *39*
2.5.5 The Matrix Normal Distribution *40*
2.6 Random Samples *41*
2.7 Limit Theorems *42*
Exercises and Complements *44*

3 Nonnormal Distributions *49*
3.1 Introduction *49*
3.2 Some Multivariate Generalizations of Univariate Distributions *49*
3.2.1 Direct Generalizations *49*
3.2.2 Common Components *50*
3.2.3 Stochastic Generalizations *51*
3.3 Families of Distributions *52*
3.3.1 The Exponential Family *52*
3.3.2 The Spherical Family *53*
3.3.3 Elliptical Distributions *56*
3.3.4 Stable Distributions *57*
3.4 Insights into Skewness and Kurtosis *57*
3.5 Copulas *60*
3.5.1 The Gaussian Copula *61*
3.5.2 The Clayton–Mardia Copula *62*
3.5.3 Archimedean Copulas *63*
3.5.4 Fréchet–Höffding Bounds *63*
Exercises and Complements *65*

4 Normal Distribution Theory *71*
4.1 Introduction and Characterization *71*
4.1.1 Introduction *71*
4.1.2 A Definition by Characterization *72*
4.2 Linear Forms *73*
4.3 Transformations of Normal Data Matrices *75*
4.4 The Wishart Distribution *77*
4.4.1 Introduction *77*

4.4.2 Properties of Wishart Matrices *77*
4.4.3 Partitioned Wishart Matrices *80*
4.5 The Hotelling T^2 Distribution *83*
4.6 Mahalanobis Distance *85*
4.6.1 The Two-Sample Hotelling T^2 Statistic *85*
4.6.2 A Decomposition of Mahalanobis Distance *86*
4.7 Statistics Based on the Wishart Distribution *88*
4.8 Other Distributions Related to the Multivariate Normal *92*
Exercises and Complements *93*

5 Estimation *101*
Introduction *101*
5.1 Likelihood and Sufficiency *101*
5.1.1 The Likelihood Function *101*
5.1.2 Efficient Scores and Fisher's Information *102*
5.1.3 The Cramér–Rao Lower Bound *104*
5.1.4 Sufficiency *105*
5.2 Maximum-likelihood Estimation *106*
5.2.1 General Case *106*
5.2.2 Multivariate Normal Case *107*
5.2.2.1 Unconstrained Case *107*
5.2.2.2 Constraints on the Mean Vector μ *109*
5.2.2.3 Constraints on Σ *110*
5.2.2.4 Samples with Linked Parameters *111*
5.2.3 Matrix Normal Distribution *112*
5.3 Robust Estimation of Location and Dispersion for Multivariate Distributions *112*
5.3.1 M-Estimates of Location *113*
5.3.2 Minimum Covariance Determinant *113*
5.3.3 Multivariate Trimming *114*
5.3.4 Stahel–Donoho Estimator *115*
5.3.5 Minimum Volume Estimator *115*
5.3.6 Tyler's Estimate of Scatter *117*
5.4 Bayesian Inference *117*
Exercises and Complements *119*

6 Hypothesis Testing *125*
6.1 Introduction *125*
6.2 The Techniques Introduced *127*
6.2.1 The Likelihood Ratio Test (LRT) *127*
6.2.1.1 The Hypothesis $H_0 : \mu = \mu_0,\ \Sigma$ Known *129*
6.2.1.2 The Hypothesis $H_0 : \mu = \mu_0,\ \Sigma$ Unknown (Hotelling's One-Sample T^2 Test) *129*
6.2.1.3 The Hypothesis $H_0 : \Sigma = \Sigma_0,\ \mu$ Unknown *130*
6.2.2 The Union Intersection Test (UIT) *131*

6.2.2.1 The Hypothesis $H_0 : \mu = \mu_0$, Σ Known (Union Intersection Approach) *133*
6.2.2.2 The Hypothesis $H_0 : \mu = \mu_0$, Σ Unknown (Hotelling One-Sample T^2 Test, Union Intersection Approach) *133*
6.2.2.3 The Hypothesis $H_0 : \Sigma = \Sigma_0$, μ Unknown (Union Intersection Test) *133*
6.3 The Techniques Further Illustrated *134*
6.3.1 One-sample hypotheses on μ *134*
6.3.1.1 The Hypothesis $H_0 : R\mu = r$, Σ Known (Hypothesis of Linear Constraints) *134*
6.3.1.2 The Hypothesis $H_0 : R\mu = r$, Σ Unknown (Hypothesis of Linear Constraints) *135*
6.3.2 One-sample hypotheses on Σ *136*
6.3.2.1 The Hypothesis $H_0 : \Sigma = k\Sigma_0$, μ Unknown *136*
6.3.2.2 The Hypothesis $H_0 : \Sigma_{12} = 0$, μ Unknown *137*
6.3.2.3 The Hypothesis $H_0 : P = I$, μ Unknown *139*
6.3.3 Multisample Hypotheses *139*
6.3.3.1 The Hypothesis $H_b : \mu_1 = \cdots = \mu_k$ Given That $\Sigma_1 = \cdots = \Sigma_k$ (One-Way Multivariate Analysis of Variance) *139*
6.3.3.2 The Hypothesis $H_a : \Sigma_1 = \cdots = \Sigma_k$ (Test of Homogeneity of Covariances) *141*
6.3.3.3 The Hypothesis $H_c : \mu_1 = \cdots = \mu_k$ and $\Sigma_1 = \cdots = \Sigma_k$ (Test of Complete Homogeneity) *142*
6.4 Simultaneous Confidence Intervals *142*
6.4.1 The one-sample Hotelling T^2 case *143*
6.4.2 The two-sample Hotelling T^2 case *144*
6.4.3 Other Examples *144*
6.5 The Behrens–Fisher Problem *144*
6.6 Multivariate Hypothesis Testing: Some General Points *145*
6.7 Nonnormal Data *146*
6.8 Mardia's Nonparametric Test for the Bivariate Two-sample Problem *149*
Exercises and Complements *151*

7 Multivariate Regression Analysis *159*
7.1 Introduction *159*
7.2 Maximum-likelihood Estimation *160*
7.2.1 Maximum-likelihood estimators for B and Σ *160*
7.2.2 The distribution of $\hat{B}$ and $\hat{\Sigma}$ *161*
7.3 The General Linear Hypothesis *162*
7.3.1 The Likelihood Ratio Test (LRT) *163*
7.3.2 The Union Intersection Test (UIT) *164*
7.3.3 Simultaneous Confidence Intervals *165*
7.4 Design Matrices of Degenerate Rank *165*
7.5 Multiple Correlation *167*
7.5.1 The Effect of the Mean *167*
7.5.2 Multiple Correlation Coefficient *168*
7.5.3 Partial Correlation Coefficient *169*
7.5.4 Measures of Correlation Between Vectors *170*

7.6 Least-squares Estimation *171*
7.6.1 Ordinary Least-squares (OLS) Estimation *171*
7.6.2 Generalized Least Squares *172*
7.6.3 Application to Multivariate Regression *173*
7.6.4 Asymptotic Consistency of Least-squares Estimators *173*
7.7 Discarding of Variables *174*
7.7.1 Dependence Analysis *174*
7.7.2 Interdependence Analysis *175*
Exercises and Complements *178*

8 Graphical Models *183*
8.1 Introduction *183*
8.2 Graphs and Conditional Independence *184*
8.3 Gaussian Graphical Models *188*
8.3.1 Estimation *189*
8.3.2 Model Selection *194*
8.4 Log-linear Graphical Models *195*
8.4.1 Notation *196*
8.4.2 Log-linear Models *197*
8.4.3 Log-linear Models with a Graphical Interpretation *199*
8.5 Directed and Mixed Graphs *202*
Exercises and Complements *204*

9 Principal Component Analysis *207*
9.1 Introduction *207*
9.2 Definition and Properties of Principal Components *207*
9.2.1 Population Principal Components *207*
9.2.2 Sample Principal Components *210*
9.2.3 Further Properties of Principal Components *211*
9.2.4 Correlation Structure *214*
9.2.5 The Effect of Ignoring Some Components *215*
9.2.6 Graphical Representation of Principal Components *217*
9.2.7 Biplots *218*
9.3 Sampling Properties of Principal Components *221*
9.3.1 Maximum-likelihood Estimation for Normal Data *221*
9.3.2 Asymptotic Distributions for Normal Data *223*
9.4 Testing Hypotheses About Principal Components *227*
9.4.1 Introduction *227*
9.4.2 The Hypothesis that $(\lambda_1 + \cdots + \lambda_k)/(\lambda_1 + \cdots + \lambda_p) = \psi$ *227*
9.4.3 The Hypothesis that $(p - k)$ Eigenvalues of Σ are Equal *228*
9.4.4 Hypotheses Concerning Correlation Matrices *229*
9.5 Correspondence Analysis *230*
9.5.1 Contingency Tables *230*
9.5.2 Gradient Analysis *235*
9.6 Allometry – Measurement of Size and Shape *237*

9.7 Discarding of Variables *240*
9.8 Principal Component Regression *241*
9.9 Projection Pursuit and Independent Component Analysis *244*
9.9.1 Projection Pursuit *244*
9.9.2 Independent Component Analysis *245*
9.10 PCA in High Dimensions *247*
Exercises and Complements *249*

10 Factor Analysis *259*
10.1 Introduction *259*
10.2 The Factor Model *260*
10.2.1 Definition *260*
10.2.2 Scale Invariance *261*
10.2.3 Nonuniqueness of Factor Loadings *261*
10.2.4 Estimation of the Parameters in Factor Analysis *262*
10.2.5 Use of the Correlation Matrix R in Estimation *263*
10.3 Principal Factor Analysis *264*
10.4 Maximum-likelihood Factor Analysis *266*
10.5 Goodness-of-fit Test *269*
10.6 Rotation of Factors *270*
10.6.1 Interpretation of Factors *270*
10.6.2 Varimax Rotation *271*
10.7 Factor Scores *275*
10.8 Relationships Between Factor Analysis and Principal Component Analysis *276*
10.9 Analysis of Covariance Structures *277*
Exercises and Complements *277*

11 Canonical Correlation Analysis *281*
11.1 Introduction *281*
11.2 Mathematical Development *282*
11.2.1 Population Canonical Correlation Analysis *282*
11.2.2 Sample Canonical Correlation Analysis *286*
11.2.3 Sampling Properties and Tests *287*
11.2.4 Scoring and Prediction *287*
11.3 Qualitative Data and Dummy Variables *288*
11.4 Qualitative and Quantitative Data *290*
Exercises and Complements *293*

12 Discriminant Analysis and Statistical Learning *297*
12.1 Introduction *297*
12.2 Bayes' Discriminant Rule *299*
12.3 The Error Rate *300*
12.3.1 Probabilities of Misclassification *300*
12.3.2 Estimation of Error Rate *303*

12.3.3 Confusion Matrix 303
12.4 Discrimination Using the Normal Distribution 304
12.4.1 Population Discriminant Rules 304
12.4.1.1 Linear Discriminant Analysis 304
12.4.1.2 Quadratic Discriminant Analysis 305
12.4.2 The Sample Discriminant Rules 306
12.4.3 Is Discrimination Worthwhile? 311
12.5 Discarding of Variables 312
12.6 Fisher's Linear Discriminant Function 314
12.7 Nonparametric Distance-based Methods 319
12.7.1 Nearest-neighbor Classifier 319
12.7.2 Large Sample Behavior of the Nearest-neighbor Classifier 319
12.7.3 Kernel Classifiers 321
12.8 Classification Trees 323
12.8.1 Splitting Criteria 325
12.8.2 Pruning 328
12.8.2.1 Pessimistic Pruning 329
12.8.2.2 Cost Complexity Pruning 330
12.9 Logistic Discrimination 332
12.9.1 Logistic Regression Model 332
12.9.2 Estimation and Inference 333
12.9.3 Interpretation of the Parameter Estimates 333
12.9.4 Extensions 336
12.10 Neural Networks 336
12.10.1 Motivation 336
12.10.2 Multilayer Perceptron 337
12.10.3 Radial Basis Functions 339
12.10.4 Support Vector Machines 341
Exercises and Complements 342

13 Multivariate Analysis of Variance 355
13.1 Introduction 355
13.2 Formulation of Multivariate One-way Classification 355
13.3 The Likelihood Ratio Principle 356
13.4 Testing Fixed Contrasts 358
13.5 Canonical Variables and A Test of Dimensionality 359
13.5.1 The Problem 359
13.5.2 The LR Test (Σ Known) 359
13.5.3 Asymptotic Distribution of the Likelihood Ratio Criterion 361
13.5.4 The Estimated Plane 361
13.5.5 The LR Test (Unknown Σ) 362
13.5.6 The Estimated Plane (Unknown Σ) 363
13.5.7 Profile Analysis 368
13.6 The Union Intersection Approach 369
13.7 Two-way Classification 370

13.7.1 Tests for Interactions *372*
13.7.2 Tests for Main Effects *372*
13.7.3 Simultaneous Confidence Regions *374*
13.7.4 Extension to Higher Designs *375*
Exercises and Complements *375*

14 Cluster Analysis and Unsupervised Learning *379*
14.1 Introduction *379*
14.2 Probabilistic Membership Models *380*
14.3 Parametric Mixture Models *384*
14.4 Partitioning Methods *386*
14.5 Hierarchical Methods *391*
14.5.1 Agglomerative Algorithms *391*
14.5.1.1 Dendrogram *393*
14.5.2 Minimum Spanning Tree and Single Linkage *395*
14.5.3 Properties of Different Agglomerative Algorithms *396*
14.6 Distances and Similarities *397*
14.6.1 Distances *397*
14.6.2 Similarity Coefficients *402*
14.7 Grouped Data *404*
14.8 Mode Seeking *406*
14.9 Measures of Agreement *408*
Exercises and Complements *412*

15 Multidimensional Scaling *419*
15.1 Introduction *419*
15.2 Classical Solution *421*
15.2.1 Some Theoretical Results *421*
15.2.2 An Algorithm for the Classical MDS Solution *424*
15.2.3 Similarities *426*
15.3 Duality Between Principal Coordinate Analysis and Principal Component Analysis *428*
15.4 Optimal Properties of the Classical Solution and Goodness of Fit *429*
15.5 Seriation *436*
15.5.1 Description *436*
15.5.2 Horseshoe Effect *437*
15.6 Nonmetric Methods *438*
15.7 Goodness of Fit Measure: Procrustes Rotation *440*
15.8 Multisample Problem and Canonical Variates *443*
Exercises and Complements *444*

16 High-dimensional Data *449*
16.1 Introduction *449*
16.2 Shrinkage Methods in Regression *451*
16.2.1 The Multiple Linear Regression Model *451*

16.2.2 Ridge Regression *452*
16.2.3 Least Absolute Selection and Shrinkage Operator (LASSO) *453*
16.3 Principal Component Regression *455*
16.4 Partial Least Squares Regression *457*
16.4.1 Overview *457*
16.4.2 The PLS1 Algorithm to Construct the PLS Loading Matrix for $p = 1$ Response Variable *458*
16.4.3 The PLS2 Algorithm to Construct the PLS Loading Matrix for $p > 1$ Response Variables *460*
16.4.4 The Predictor Envelope Model *461*
16.4.5 PLS Regression *462*
16.4.6 Joint Envelope Models *464*
16.5 Functional Data *465*
16.5.1 Functional Principal Component Analysis *466*
16.5.2 Functional Linear Regression Models *470*
Exercises and Complements *473*

A Matrix Algebra *475*
A.1 Introduction *475*
A.2 Matrix Operations *478*
A.2.1 Transpose *478*
A.2.2 Trace *479*
A.2.3 Determinants and Cofactors *479*
A.2.4 Inverse *481*
A.2.5 Kronecker Products *482*
A.3 Further Particular Matrices and Types of Matrices *483*
A.3.1 Orthogonal Matrices *483*
A.3.2 Equicorrelation Matrix *484*
A.3.3 Centering Matrix *484*
A.4 Vector Spaces, Rank, and Linear Equations *485*
A.4.1 Vector Spaces *485*
A.4.2 Rank *486*
A.4.3 Linear Equations *487*
A.5 Linear Transformations *488*
A.6 Eigenvalues and Eigenvectors *488*
A.6.1 General Results *488*
A.6.2 Symmetric Matrices *490*
A.7 Quadratic Forms and Definiteness *495*
A.8 Generalized Inverse *497*
A.9 Matrix Differentiation and Maximization Problems *499*
A.10 Geometrical Ideas *501*
A.10.1 n-dimensional Geometry *501*
A.10.2 Orthogonal Transformations *502*
A.10.3 Projections *502*
A.10.4 Ellipsoids *503*

B **Univariate Statistics** *505*
B.1 Introduction *505*
B.2 Normal Distribution *505*
B.3 Chi-squared Distribution *506*
B.4 F and Beta Variables *506*
B.5 t Distribution *507*
B.6 Poisson Distribution *507*

C **R Commands and Data** *509*
C.1 Basic R Commands Related to Matrices *509*
C.2 R Libraries and Commands Used in Exercises and Figures *510*
C.3 Data Availability *511*

D **Tables** *513*

References and Author Index *523*
Index *543*

Epigraph

> *Everything is related with every other thing, and this relation involves the emergence of a relational quality. The qualities cannot be known* a priori*, though a good number of them can be deduced from certain fundamental characteristics.*
>
> — Jaina philosophy

The *Jaina Philosophy of Non-Absolutism* by S. Mookerjee, q.v.
Mahalanobis (1957).

Preface to the Second Edition

For over 40 years the first edition of this book (which was also translated into Persian) has been used by students to acquire a basic knowledge of the theory and methods of multivariate statistical analysis. The book has also served the wider statistical community to further their understanding of this field. Plans for the second edition started almost 20 years ago, and we have struggled with questions about which topics to add – something of a moving target in a field that has continued to evolve in this new era of artificial intelligence (AI) and "big data". Since the first edition was published, multivariate analysis has been developed and extended in many directions. This new edition aims to bring the first edition up to date by substantial revision, rewriting, and additions, while seeking to maintain the overall length of the book. The basic approach has been maintained, namely a mathematical treatment of statistical methods for observations consisting of several measurements or characteristics of each subject and a study of their properties. The core topics, and the structure many of the chapters, have been retained.

Briefly, for those familiar with the first edition, the main changes (in addition to updating material in several places) are:

- a new section giving Notation, Abbreviations, and Key Ideas used through the book;
- a new chapter introducing some nonnormal distributions. This includes new sections on elliptical distributions and copulas;
- a new chapter covering an introduction to *graphical models*;
- a completely rewritten chapter that begins from discriminant analysis and extends to nonparametric methods, classification and regression trees, logistic discrimination, and multilayer perceptrons. These topics are commonly grouped into the heading of *supervised learning*;
- the above chapter focuses on data in which group memberships are known, whereas "unsupervised learning" has more traditionally been known as cluster analysis, for which the current Chapter 14 has also been substantially updated to reflect recent developments;
- a new (final) chapter introduces some approaches to *high-dimensional data* in which the number of variables may exceed the number of observations. This includes shrinkage methods in regression, principal components regression, partial least squares regression, and functional data analysis;
- further development of discrete aspects, including log-linear models, the EM algorithm for mixture models, and correspondence analysis for contingency tables.

As a consequence of the above new and extended/revised chapters and in order to save space, we have omitted some material from this edition:

- the chapter on econometrics, since there are now dedicated books with an emphasis on statistical aspects (Maddala and Lahiri, 2009); (Wooldridge, 2019);
- the chapter on directional statistics, since there are now related dedicated books by one of the authors (Dryden and Mardia, 2016); (Mardia and Jupp, 2000).

Further changes to this Edition, bringing many subjects up to date, include new graphical representations (Chapter 1), an introduction to the matrix normal distribution (Chapters 2 and 5), elliptical distributions and copulas (Chapter 3), robust estimators for location and dispersion (Chapter 5), a revision of correspondence analysis and biplots (Chapter 9), and projection pursuit and independent component analysis (Chapter 9).

The figures in the first edition have been redrawn in their original style, using the statistical package R. A new Appendix C contains some specific R (R Core Team, 2020) commands applicable to most of the matrix algebra used in the book. In addition, an online addendum to Appendix C contains the data files used in this book as well as the R commands used to obtain the calculations for the examples and figures. This public repository is at github.com/charlesctaylor/MVAdata-rcode. In many cases, we have chosen to use base R functions to mimic the equations used in the text in preference to more "black-box" R functions. Note that intermediate steps in the calculation are generally rounded *only* for display purposes.

Multivariate analysis continues to be a research area of active development. We note that the Journal of Multivariate Analysis, in its 50th Anniversary Jubilee Edition (von Rosen and Kollo, 2022), has published a volume that describes the current state of the art and contains review papers. Beyond mainstream multivariate statistics, there have been developments in the applied sciences; one example in morphometrics is Bookstein (2018).

The first edition was published by Academic Press, and we are grateful to John Bibby for his contributions to that edition. For this edition, we thank the many readers who have offered their advice and suggestions. In particular, we would like to acknowledge the help of Susan Holmes for extensive discussions about a new structure as well as a draft of correspondence analysis material for Chapter 9.

We are extremely grateful to Wiley for their patience and help during the writing of the book, especially Helen Ramsey, Sharon Clutton, Richard Davies, Kathryn Sharples, Liz Wingett, Kelvin Matthews, Alison Oliver, Viktoria Hartl-Vida, Ashley Alliano, Kimberly Monroe-Hill, and Paul Sayer. Secretarial help at Leeds during the initial development was given by Christine Rutherford and Catherine Dobson.

Kanti would like to thank the Leverhulme Trust for an Emeritus Fellowship and Anna Grundy of the Trust for simplifying the administration process. Finally, he would like to express his sincere gratitude to his family for their continuous love, support, and tolerance.

We would be pleased to hear about any typographical or other errors in the text.

May 2022

Kanti V. Mardia
University of Leeds, Leeds, UK and
University of Oxford, Oxford, UK

John T. Kent
University of Leeds, Leeds, UK

Charles C. Taylor
University of Leeds, Leeds, UK

Preface to the First Edition

Multivariate Analysis deals with observations on more than one variable where there is some inherent interdependence between the variables. With several texts already available in this area, one may very well enquire of the authors as to the need for yet another book. Most of the available books fall into two categories, either theoretical or data analytic. The present book not only combines the two approaches but also emphasizes modern developments. The choice of material for the book has been guided by the need to give suitable matter for the beginner as well as illustrating some deeper aspects of the subject for the research worker. Practical examples are kept to the forefront, and, wherever feasible, each technique is motivated by such an example.

The book is aimed at final year undergraduates and postgraduate students in Mathematics/Statistics with sections suitable for practitioners and research workers. The book assumes a basic knowledge of Mathematical Statistics at undergraduate level. An elementary course on Linear Algebra is also assumed. In particular, we assume an exposure to Matrix Algebra to the level required to read Appendix A.

Broadly speaking, Chapters 1–6 and 12 can be described as containing direct extensions of univariate ideas and techniques. The remaining chapters concentrate on specifically multivariate problems that have no meaningful analogs in the univariate case. Chapter 1 is primarily concerned with giving exploratory analyses for multivariate data and briefly introduces some of the important techniques, tools, and diagrammatic representations. Chapter 2 introduces various distributions together with some fundamental results, whereas Chapter 3 concentrates exclusively on normal distribution theory. Chapters 4–6 deal with problems in inference. Chapter 7 [no longer included] gives an overview of Econometrics, while Principal Component Analysis, Factor Analysis, Canonical Correlation Analysis, and Discriminant Analysis are discussed from both theoretical and practical points of view in Chapters 8–11. Chapter 12 is on Multivariate Analysis of Variance, which can be better understood in terms of the techniques of previous chapters. The later chapters look into the presently developing techniques of Cluster Analysis, Multidimensional Scaling, and Directional Data [no longer included].

Each chapter concludes with a set of exercises. Solving these will not only enable the reader to understand the material better but will also serve to complement the chapter itself. In general, the questions have in-built answers, but, where desirable, hints for the solution of theoretical problems are provided. Some of the numerical exercises are designed to be run on a computer, but as the main aim is on interpretation, the answers are provided. We found NAG routines and GLIM most useful, but nowadays any computer center will have some suitable statistics and matrix algebra routines.

There are three Appendices A, B, and C, which, respectively, provide a sufficient background of matrix algebra, a summary of univariate statistics, and some tables of critical values. The aim of Appendix A on Matrix Algebra is not only to provide a summary of results but also to give sufficient guidance to master these for students having little previous knowledge. Equations from Appendix A are referred to as (A.x.x) to distinguish them from (l.x.x), etc. Appendix A also includes a summary of results in n-dimensional geometry that are used liberally in the book. Appendix B gives a summary of important univariate distributions.

The reference list is by no means exhaustive. Only directly relevant articles are quoted, and for a fuller bibliography, we refer the reader to Anderson et al. (1972) and Subrahmaniam and Subrahmaniam (1973). The reference list also serves as an author index. A subject index is provided.

The material in the book can be used in several different ways. For example, a one-semester elementary course of 40 lectures could cover the following topics. Appendix A; Chapter 1 (Sections 1.1–1.7); Chapter 2 (Sections 2.1–2.5); Chapter 3 (Sections 3.4.1, 3.5, and 3.6.1, assuming results from previous sections, Definitions 3.7.1 and 3.7.2); Chapter 4 (Section 4.2.2); Chapter 5 (Sections 5.1, 5.2.1a, 5.2.1b, 5.2.2a, 5.2.2b, 5.3.2b, and 5.5); Chapter 8 (Sections 8.1, 8.2.1, 8.2.2, 8.2.5, 8.2.6, 8.4.3, and 8.7); Chapter 9 (Sections 9.1–9.3, 9.4 (without details), 9.5, 9.6, and 9.8); Chapter 10 (Sections 10.1 and 10.2); Chapter 11 (Sections 11.1, 11.2.1–11.2.3, 11.3.1, and 11.6.1). Further material that can be introduced is Chapter 12 (Sections 12.1–12.3 and 12.6); Chapter 13 (Sections 13.1 and 13.3.1); Chapter 14 (Sections 14.1 and 14.2). This material has been covered in 40 lectures spread over two terms in different British universities. Alternatively, a one-semester course with more emphasis on foundation rather than applications could be based on Appendix A and Chapters 1–5. Two-semester courses could include all the chapters, excluding Chapters 7 and 15 on Econometrics and Directional Data, as well as the sections with asterisks. Mathematically orientated students may like to proceed to Chapter 2, omitting the data analytic ideas of Chapter 1.

Various new methods of presentation are utilized in the book. For instance, the data matrix is emphasized throughout, a density-free approach is given for normal theory, the union intersection principle is used in testing as well as the likelihood ratio principle, and graphical methods are used in explanation. In view of the computer packages generally available, most of the numerical work is taken for granted, and therefore, except for a few particular cases, emphasis is not placed on numerical calculations. The style of presentation is generally kept descriptive except where rigor is found to be necessary for theoretical results, which are then put in the form of theorems. If any details of the proof of a theorem are felt tedious but simple, they are then relegated to the exercises.

Several important topics not usually found in multivariate texts are discussed in detail. Examples of such material include the complete chapters on Econometrics, Cluster Analysis, Multidimensional Scaling, and Directional Data. Further material is also included in parts of other chapters: methods of graphical presentation, measures of multivariate skewness and kurtosis, the singular multinormal distribution, various nonnormal distributions and families of distributions, a density-free approach to normal distribution theory, Bayesian and robust estimators, a recent solution to the Fisher–Behrens problem, a test of multinormality, a nonparametric test, discarding of variables in regression, principal component analysis and discrimination analysis, correspondence analysis, allometry, the jack-knifing method in discrimination, canonical analysis of qualitative and quantitative variables, and a test of dimensionality in MANOVA. It is hoped that coverage of these developments will be helpful for students as well as research workers.

There are various other topics that have not been touched upon partly because of lack of space as well as our own preferences, such as Control Theory, Multivariate Time Series, Latent Variable Models, Path Analysis, Growth Curves, Portfolio Analysis, and various Multivariate Designs.

In addition to various research papers, we have been influenced by particular texts in this area, especially Anderson (1958), Kendall (1975), Kshirsagar (1972), Morrison (1976), Press (1972), and Rao (1973). All these are recommended to the reader.

The authors would be most grateful to readers who draw their attention to any errors or obscurities in the book, or suggest other improvements.

January 1979

Kanti V. Mardia
John T. Kent
John C. Bibby

Acknowledgments from First Edition

First of all, we wish to express our gratitude to the pioneers in this field. In particular, we should mention M. S. Bartlett, R. A. Fisher, H. Hotelling, D. G. Kendall, M. G. Kendall, P. C. Mahalanobis, C. R. Rao, S. N. Roy, W. S. Torgeson, and S. S. Wilks.

We are grateful to the authors and editors who have generously granted us permission to reproduce figures and tables.

We are also grateful to many of our colleagues for their valuable help and comments, in particular Martin Beale, Christopher Bingham, Lesley Butler, Richard Cormack, David Cox, Ian Curry, Peter Fisk, Allan Gordon, John Gower, Peter Harris, Chunni Khatri, Conrad Leser, Eric Okell, Ross Renner, David Salmond, Cyril Smith, and Peter Zemroch. We are also indebted to Joyce Snell for making various comments on an earlier draft of the book, which have led to considerable improvements. We should also express our gratitude to Rob Edwards for his help in various facets of the book, for calculations, proofreading, diagrams, etc.

Some of the questions are taken from examination papers in British universities, and we are grateful to various unnamed colleagues. Since the original sources of questions are difficult to trace, we apologize to any colleague who recognizes a question of their own.

The authors would like to thank their wives, Pavan Mardia, Susan Kent, and Zorina Bibby.

Finally, our thanks go to Barbara Forsyth and Margaret Richardson for typing a difficult manuscript with great skill.

KVM
JTK
JMB

Notation, Abbreviations, and Key Ideas

Matrices and Vectors

- **Vectors** are viewed as column vectors and are represented using bold lower case letters. Round brackets are generally used when a vector is expressed in terms of its elements. For example,

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

in which the rth element or component is denoted x_r. The transpose of $\boldsymbol{x}$ is denoted $\boldsymbol{x}'$, so $\boldsymbol{x}' = (x_1, \ldots, x_n)$ is a row vector.
- **Matrices** are written using bold upper case letters, e.g. $\boldsymbol{A}$ and $\boldsymbol{\Gamma}$. The matrix $\boldsymbol{A}$ may be written as (a_{ij}) in which a_{ij} is the element of the matrix in row i and column j. If $\boldsymbol{A}$ has n rows and p columns, then the ith row of $\boldsymbol{A}$, written as a column vector, is

$$\boldsymbol{a}_i = \begin{pmatrix} a_{i1} \\ \vdots \\ a_{ip} \end{pmatrix}, \qquad i = 1, \ldots, n,$$

and the jth column is written as

$$\boldsymbol{a}_{(j)} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}, \qquad j = 1, \ldots, p.$$

Hence, A can be expressed in various forms,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_n' \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{(1)} & \cdots & \boldsymbol{a}_{(p)} \end{bmatrix}.$$

We generally use square brackets when a matrix is expanded in terms of its elements. Operations on a matrix $\boldsymbol{A}$ include
  - **transpose:** $\boldsymbol{A}'$
  - **determinant:** $|\boldsymbol{A}|$
  - **inverse:** $\boldsymbol{A}^{-1}$
  - **generalized inverse:** $\boldsymbol{A}^{-}$

where for the final three operations, $\boldsymbol{A}$ is assumed to be square, and for the inverse operation, $\boldsymbol{A}$ is additionally assumed to be nonsingular. Different types of matrices are given in Tables A.1 and A.3. Table A.2 lists some further matrix operations.

Random Variables and Data

- In general, a random vector and a nonrandom vector are both indicated using a bold lower case letter, e.g. $\boldsymbol{x} = (x_1, \dots, x_p)'$. Thus, the distinction between the two must be determined from the context. This convention is in contrast to the standard convention in statistics where upper case letters are used to denote random quantities, and lower case letters their observed values.
- The reason for our convention is that bold upper case letters are generally used for a data matrix $\boldsymbol{X}(n \times p)$, both random and fixed.
- In spite of the above convention, we very occasionally (e.g. parts of Chapters 2 and 10) use bold upper case letters $\boldsymbol{X} = (X_1, \dots, X_p)'$ for a random vector when it is important to distinguish between the random vector $\boldsymbol{X}$ and a possible value $\boldsymbol{x}$.
- The phrase "high-dimensional data" often implies $p > n$, whereas the phrase "big data" often just indicates that n or p is large.

Parameters and Statistics

Elements of an $n \times p$ data matrix $\boldsymbol{X}$ are generally written x_{ri}, where indices $r, s, \dots$, are used to label the *observations*, and indices $i, j, k, \dots$, are used to label the *variables*.

If the rows of a data matrix $\boldsymbol{X}$ are normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$, and $\boldsymbol{\Delta} = \text{Diag}\,(\boldsymbol{\Sigma})^{1/2}$, the following notation is used to distinguish various population and sample quantities:

Parameter	Sample	
$\boldsymbol{\mu}$	$\overline{\boldsymbol{x}}$	Mean vector
$\boldsymbol{\Sigma}$	$\boldsymbol{S} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} - \overline{\boldsymbol{x}}\,\overline{\boldsymbol{x}}'$	Covariance matrix
	$\boldsymbol{S}_u = \frac{n}{n-1}\boldsymbol{S}$	Unbiased covariance matrix
$\boldsymbol{K} = \boldsymbol{\Sigma}^{-1}$	$\boldsymbol{K}^{\text{sample}} = \boldsymbol{S}^{-1}$	Concentration matrix
$\boldsymbol{P} = (\rho_{ij}) = \boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Delta}^{-1}$	$\boldsymbol{R} = (r_{ij})$	Correlation matrix

Distributions

The following notation is used for univariate and multivariate distributions. Appendix B summarizes the univariate distributions used in the Book.