CONTENTS

Preface xvii
Acknowledgments xxi

1 Introduction 1
  1.1 WHY MULTIVARIATE ANALYSIS? 1
  1.2 PREREQUISITES 3
  1.3 OBJECTIVES 3
  1.4 BASIC TYPES OF DATA AND ANALYSIS 4

2 Matrix Algebra 7
  2.1 INTRODUCTION 7
  2.2 NOTATION AND BASIC DEFINITIONS 8
    2.2.1 Matrices, Vectors, and Scalars 8
    2.2.2 Equality of Vectors and Matrices 9
    2.2.3 Transpose and Symmetric Matrices 9
    2.2.4 Special Matrices 10
  2.3 OPERATIONS 11
3 Characterizing and Displaying Multivariate Data

3.1 MEAN AND VARIANCE OF A UNIVARIATE RANDOM VARIABLE

3.2 COVARIANCE AND CORRELATION OF BIVARIATE RANDOM VARIABLES

3.3 SCATTERPLOTS OF BIVARIATE SAMPLES

3.4 GRAPHICAL DISPLAYS FOR MULTIVARIATE SAMPLES

3.5 DYNAMIC GRAPHICS

3.6 MEAN VECTORS

3.7 COVARIANCE MATRICES

3.8 CORRELATION MATRICES
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>vii</th>
</tr>
</thead>
</table>

3.9 MEAN VECTORS AND COVARIANCE MATRICES FOR
SUBSETS OF VARIABLES 71
3.9.1 Two Subsets 71
3.9.2 Three or More Subsets 73

3.10 LINEAR COMBINATIONS OF VARIABLES 75
3.10.1 Sample Properties 75
3.10.2 Population Properties 81

3.11 MEASURES OF OVERALL VARIABILITY 81
3.12 ESTIMATION OF MISSING VALUES 82

3.13 DISTANCE BETWEEN VECTORS 84
Problems 85

4 The Multivariate Normal Distribution 91

4.1 MULTIVARIATE NORMAL DENSITY FUNCTION 91
4.1.1 Univariate Normal Density 92
4.1.2 Multivariate Normal Density 92
4.1.3 Generalized Population Variance 93
4.1.4 Diversity of Applications of the Multivariate Normal 93

4.2 PROPERTIES OF MULTIVARIATE NORMAL RANDOM VARIABLES 94

4.3 ESTIMATION IN THE MULTIVARIATE NORMAL 99
4.3.1 Maximum Likelihood Estimation 99
4.3.2 Distribution of $y$ and $S$ 100

4.4 ASSESSING MULTIVARIATE NORMALITY 101
4.4.1 Investigating Univariate Normality 101
4.4.2 Investigating Multivariate Normality 106

4.5 TRANSFORMATIONS TO NORMALITY 108
4.5.1 Univariate Transformations to Normality 109
4.5.2 Multivariate Transformations to Normality 110

4.6 OUTLIERS 111
4.6.1 Outliers in Univariate Samples 112
4.6.2 Outliers in Multivariate Samples 113
Problems 117

5 Tests on One or Two Mean Vectors 125

5.1 MULTIVARIATE VERSUS UNIVARIATE TESTS 125
5.2 TESTS ON $\mu$ WITH $\Sigma$ KNOWN 126
5.2.1 Review of Univariate Test for $H_0: \mu = \mu_0$ with $\sigma$ Known 126
5.2.2 Multivariate Test for $H_0: \mu = \mu_0$ with $\Sigma$ Known 127
5.3 TESTS ON $\mu$ WHEN $\Sigma$ IS UNKNOWN 130
5.3.1 Review of Univariate $t$-Test for $H_0: \mu = \mu_0$ with $\sigma$ Unknown 130
5.3.2 Hotelling's $T^2$-Test for $H_0: \mu = \mu_0$ with $\Sigma$ Unknown 131
5.4 COMPARING TWO MEAN VECTORS 134
5.4.1 Review of Univariate Two-Sample $t$-Test 134
5.4.2 Multivariate Two-Sample $T^2$-Test 135
5.4.3 Likelihood Ratio Tests 139
5.5 TESTS ON INDIVIDUAL VARIABLES CONDITIONAL ON REJECTION OF $H_0$ BY THE $T^2$-TEST 139
5.6 COMPUTATION OF $T^2$ 143
5.6.1 Obtaining $T^2$ from a MANOVA Program 143
5.6.2 Obtaining $T^2$ from Multiple Regression 144
5.7 PAIRED OBSERVATIONS TEST 145
5.7.1 Univariate Case 145
5.7.2 Multivariate Case 147
5.8 TEST FOR ADDITIONAL INFORMATION 149
5.9 PROFILE ANALYSIS 152
5.9.1 One-Sample Profile Analysis 152
5.9.2 Two-Sample Profile Analysis 154
Problems 161

6 Multivariate Analysis of Variance 169
6.1 ONE-WAY MODELS 169
6.1.1 Univariate One-Way Analysis of Variance (ANOVA) 169
6.1.2 Multivariate One-Way Analysis of Variance Model (MANOVA) 171
6.1.3 Wilks' Test Statistic 174
6.1.4 Roy's Test 178
6.1.5 Pillai and Lawley--Hotelling Tests 179
6.1.6 Unbalanced One-Way MANOVA 181
6.1.7 Summary of the Four Tests and Relationship to $T^2$ 182
6.1.8 Measures of Multivariate Association 186
6.2 COMPARISON OF THE FOUR MANOVA TEST STATISTICS 189
6.3 CONTRASTS 191
6.3.1 Univariate Contrasts 191
6.3.2 Multivariate Contrasts 192

6.4 TESTS ON INDIVIDUAL VARIABLES FOLLOWING REJECTION OF $H_0$ BY THE OVERALL MANOVA TEST 195

6.5 TWO-WAY CLASSIFICATION 198
6.5.1 Review of Univariate Two-Way ANOVA 198
6.5.2 Multivariate Two-Way MANOVA 201

6.6 OTHER MODELS 207
6.6.1 Higher-Order Fixed Effects 207
6.6.2 Mixed Models 208

6.7 CHECKING ON THE ASSUMPTIONS 210

6.8 PROFILE ANALYSIS 211

6.9 REPEATED MEASURES DESIGNS 215
6.9.1 Multivariate Versus Univariate Approach 215
6.9.2 One-Sample Repeated Measures Model 219
6.9.3 $k$-Sample Repeated Measures Model 222
6.9.4 Computation of Repeated Measures Tests 224
6.9.5 Repeated Measures with Two Within-Subjects Factors and One Between-Subjects Factor 224
6.9.6 Repeated Measures with Two Within-Subjects Factors and Two Between-Subjects Factors 230
6.9.7 Additional Topics 232

6.10 GROWTH CURVES 232
6.10.1 Growth Curve for One Sample 232
6.10.2 Growth Curves for Several Samples 239
6.10.3 Additional Topics 241

6.11 TESTS ON A SUBVECTOR 241
6.11.1 Test for Additional Information 241
6.11.2 Stepwise Selection of Variables 243
Problems 244

7 Tests on Covariance Matrices 259

7.1 INTRODUCTION 259

7.2 TESTING A SPECIFIED PATTERN FOR $\Sigma$ 259
7.2.1 Testing $H_0: \Sigma = \Sigma_0$ 260
7.2.2 Testing Sphericity 261
7.2.3 Testing $H_0: \Sigma = \sigma^2[(1 - \rho)I + \rho J]$ 263

7.3 TESTS COMPARING COVARIANCE MATRICES 265
7.3.1 Univariate Tests of Equality of Variances 265
7.3.2 Multivariate Tests of Equality of Covariance Matrices 266
7.4 TESTS OF INDEPENDENCE 269
7.4.1 Independence of Two Subvectors 269
7.4.2 Independence of Several Subvectors 271
7.4.3 Test for Independence of All Variables 275
Problems 276

8 Discriminant Analysis: Description of Group Separation 281
8.1 INTRODUCTION 281
8.2 THE DISCRIMINANT FUNCTION FOR TWO GROUPS 282
8.3 RELATIONSHIP BETWEEN TWO-GROUP DISCRIMINANT ANALYSIS AND MULTIPLE REGRESSION 286
8.4 DISCRIMINANT ANALYSIS FOR SEVERAL GROUPS 288
8.4.1 Discriminant Functions 288
8.4.2 A Measure of Association for Discriminant Functions 292
8.5 STANDARDIZED DISCRIMINANT FUNCTIONS 292
8.6 TESTS OF SIGNIFICANCE 294
8.6.1 Tests for the Two-Group Case 294
8.6.2 Tests for the Several-Group Case 295
8.7 INTERPRETATION OF DISCRIMINANT FUNCTIONS 298
8.7.1 Standardized Coefficients 298
8.7.2 Partial $F$-Values 299
8.7.3 Correlations Between Variables and Discriminant Functions 300
8.7.4 Rotation 301
8.8 SCATTERPLOTS 301
8.9 STEPWISE SELECTION OF VARIABLES 303
Problems 306

9 Classification Analysis: Allocation of Observations to Groups 309
9.1 INTRODUCTION 309
9.2 CLASSIFICATION INTO TWO GROUPS 310
9.3 CLASSIFICATION INTO SEVERAL GROUPS 314
9.3.1 Equal Population Covariance Matrices: Linear Classification Functions 315
9.3.2 Unequal Population Covariance Matrices: Quadratic Classification Functions 317
10 Multivariate Regression

10.1 INTRODUCTION

10.2 MULTIPLE REGRESSION: FIXED $x$'s

10.2.1 Model for Fixed $x$'s

10.2.2 Least Squares Estimation in the Fixed-$x$ Model

10.2.3 An Estimator for $\sigma^2$

10.2.4 The Model Corrected for Means

10.2.5 Hypothesis Tests

10.2.6 $R^2$ in Fixed-$x$ Regression

10.2.7 Subset Selection

10.3 MULTIPLE REGRESSION: RANDOM $x$'s

10.4 MULTIVARIATE MULTIPLE REGRESSION: ESTIMATION

10.4.1 The Multivariate Linear Model

10.4.2 Least Squares Estimation in the Multivariate Model

10.4.3 Properties of Least Squares Estimator $\hat{B}$

10.4.4 An Estimator for $\Sigma$

10.4.5 Model Corrected for Means

10.4.6 Estimation in the Seemingly Unrelated Regressions (SUR) Model

10.5 MULTIVARIATE MULTIPLE REGRESSION: HYPOTHESIS TESTS

10.5.1 Test of Overall Regression

10.5.2 Test on a Subset of the $x$'s

10.6 MULTIVARIATE MULTIPLE REGRESSION: PREDICTION

10.6.1 Confidence Interval for $E(y_0)$

10.6.2 Prediction Interval for a Future Observation $y_0$
10.7 MEASURES OF ASSOCIATION BETWEEN THE y’s AND THE x’s 372
10.8 SUBSET SELECTION 374
   10.8.1 Stepwise Procedures 374
   10.8.2 All Possible Subsets 377
10.9 MULTIVARIATE REGRESSION: RANDOM x’s 380
     Problems 381

11 Canonical Correlation 385
   11.1 INTRODUCTION 385
   11.2 CANONICAL CORRELATIONS AND CANONICAL VARIATES 385
   11.3 PROPERTIES OF CANONICAL CORRELATIONS 390
   11.4 TESTS OF SIGNIFICANCE 391
       11.4.1 Tests of No Relationship Between the y’s and the x’s 391
       11.4.2 Test of Significance of Succeeding Canonical Correlations After the First 393
   11.5 INTERPRETATION 395
       11.5.1 Standardized Coefficients 396
       11.5.2 Correlations between Variables and Canonical Variates 397
       11.5.3 Rotation 397
       11.5.4 Redundancy Analysis 398
   11.6 RELATIONSHIPS OF CANONICAL CORRELATION ANALYSIS TO OTHER MULTIVARIATE TECHNIQUES 398
       11.6.1 Regression 398
       11.6.2 MANOVA and Discriminant Analysis 400
     Problems 402

12 Principal Component Analysis 405
   12.1 INTRODUCTION 405
   12.2 GEOMETRIC AND ALGEBRAIC BASES OF PRINCIPAL COMPONENTS 406
       12.2.1 Geometric Approach 406
       12.2.2 Algebraic Approach 410
   12.3 PRINCIPAL COMPONENTS AND PERPENDICULAR REGRESSION 412
   12.4 PLOTTING OF PRINCIPAL COMPONENTS 414
12.5 PRINCIPAL COMPONENTS FROM THE CORRELATION MATRIX 419
12.6 DECIDING HOW MANY COMPONENTS TO RETAIN 423
12.7 INFORMATION IN THE LAST FEW PRINCIPAL COMPONENTS 427
12.8 INTERPRETATION OF PRINCIPAL COMPONENTS 427
12.8.1 Special Patterns in S or R 427
12.8.2 Rotation 429
12.8.3 Correlations Between Variables and Principal Components 429
12.9 SELECTION OF VARIABLES 430
Problems 432

13 Exploratory Factor Analysis 435
13.1 INTRODUCTION 435
13.2 ORTHOGONAL FACTOR MODEL 437
13.2.1 Model Definition and Assumptions 437
13.2.2 Nonuniqueness of Factor Loadings 441
13.3 ESTIMATION OF LOADINGS AND COMMUNALITIES 442
13.3.1 Principal Component Method 443
13.3.2 Principal Factor Method 448
13.3.3 Iterated Principal Factor Method 450
13.3.4 Maximum Likelihood Method 452
13.4 CHOOSING THE NUMBER OF FACTORS, m 453
13.5 ROTATION 457
13.5.1 Introduction 457
13.5.2 Orthogonal Rotation 458
13.5.3 Oblique Rotation 462
13.5.4 Interpretation 465
13.6 FACTOR SCORES 466
13.7 VALIDITY OF THE FACTOR ANALYSIS MODEL 470
13.8 RELATIONSHIP OF FACTOR ANALYSIS TO PRINCIPAL COMPONENT ANALYSIS 475
Problems 476

14 Confirmatory Factor Analysis 479
14.1 INTRODUCTION 479
14.2 MODEL SPECIFICATION AND IDENTIFICATION 480
14.2.1 Confirmatory Factor Analysis Model 480
14.2.2 Identified Models 482

14.3 PARAMETER ESTIMATION AND MODEL ASSESSMENT 487
  14.3.1 Maximum Likelihood Estimation 487
  14.3.2 Least Squares Estimation 488
  14.3.3 Model Assessment 489

14.4 INFERENCE FOR MODEL PARAMETERS 492
14.5 FACTOR SCORES 495
  Problems 496

15 Cluster Analysis 501
  15.1 INTRODUCTION 501
  15.2 MEASURES OF SIMILARITY OR DISSIMILARITY 502
  15.3 HIERARCHICAL CLUSTERING 505
    15.3.1 Introduction 505
    15.3.2 Single Linkage (Nearest Neighbor) 506
    15.3.3 Complete Linkage (Farthest Neighbor) 508
    15.3.4 Average Linkage 511
    15.3.5 Centroid 514
    15.3.6 Median 514
    15.3.7 Ward's Method 517
    15.3.8 Flexible Beta Method 520
    15.3.9 Properties of Hierarchical Methods 521
    15.3.10 Divisive Methods 529
  15.4 NONHIERARCHICAL METHODS 531
    15.4.1 Partitioning 532
    15.4.2 Other Methods 540
  15.5 CHOOSING THE NUMBER OF CLUSTERS 544
  15.6 CLUSTER VALIDITY 546
  15.7 CLUSTERING VARIABLES 547
  Problems 548

16 Graphical Procedures 555
  16.1 MULTIDIMENSIONAL SCALING 555
    16.1.1 Introduction 555
    16.1.2 Metric Multidimensional Scaling 556
    16.1.3 Nonmetric Multidimensional Scaling 560
  16.2 CORRESPONDENCE ANALYSIS 565
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.2.1 Introduction</td>
<td>565</td>
</tr>
<tr>
<td>16.2.2 Row and Column Profiles</td>
<td>566</td>
</tr>
<tr>
<td>16.2.3 Testing Independence</td>
<td>570</td>
</tr>
<tr>
<td>16.2.4 Coordinates for Plotting Row and Column Profiles</td>
<td>572</td>
</tr>
<tr>
<td>16.2.5 Multiple Correspondence Analysis</td>
<td>576</td>
</tr>
<tr>
<td>16.3 BIPLOTS</td>
<td>580</td>
</tr>
<tr>
<td>16.3.1 Introduction</td>
<td>580</td>
</tr>
<tr>
<td>16.3.2 Principal Component Plots</td>
<td>581</td>
</tr>
<tr>
<td>16.3.3 Singular Value Decomposition Plots</td>
<td>583</td>
</tr>
<tr>
<td>16.3.4 Coordinates</td>
<td>583</td>
</tr>
<tr>
<td>16.3.5 Other Methods</td>
<td>585</td>
</tr>
<tr>
<td>Problems</td>
<td>588</td>
</tr>
<tr>
<td>Appendix A: Tables</td>
<td>597</td>
</tr>
<tr>
<td>Appendix B: Answers and Hints to Problems</td>
<td>637</td>
</tr>
<tr>
<td>Appendix C: Data Sets and SAS Files</td>
<td>727</td>
</tr>
<tr>
<td>References</td>
<td>728</td>
</tr>
<tr>
<td>Index</td>
<td>745</td>
</tr>
</tbody>
</table>