Contents

Preface **ix**

Commonly Used Notation **xiii**

1 Basic Concepts for Multivariate Statistics **1**
1.1 Introduction 1
1.2 Population Versus Sample 2
1.3 Elementary Tools for Understanding Multivariate Data 3
1.4 Data Reduction, Description, and Estimation 6
1.5 Concepts from Matrix Algebra 7
1.6 Multivariate Normal Distribution 21
1.7 Concluding Remarks 23

2 Principal Component Analysis **25**
2.1 Introduction 25
2.2 Population Principal Components 26
2.3 Sample Principal Components 29
2.4 Selection of the Number of Principal Components 40
2.5 Some Applications of Principal Component Analysis 46
2.6 Principal Component Analysis of Compositional Data 57
2.7 Principal Component Regression 60
2.8 Principal Component Residuals and Detection of Outliers 65
2.9 Principal Component Biplot 69
2.10 PCA Using SAS/INSIGHT Software 76
2.11 Concluding Remarks 76

3 Canonical Correlation Analysis **77**
3.1 Introduction 77
3.2 Population Canonical Correlations and Canonical Variables 78
3.3 Sample Canonical Correlations and Canonical Variables 79
3.4 Canonical Analysis of Residuals 91
3.5 Partial Canonical Correlations 92
3.6 Canonical Redundancy Analysis 95
3.7 Canonical Correlation Analysis of Qualitative Data 101

3.8 'Partial Tests' in Multivariate Regression 106
3.9 Concluding Remarks 108

4 Factor Analysis 111
4.1 Introduction 111
4.2 Factor Model 112
4.3 A Difference between PCA and Factor Analysis 116
4.4 Noniterative Methods of Estimation 118
4.5 Iterative Methods of Estimation 139
4.6 Heywood Cases 155
4.7 Comparison of the Methods 156
4.8 Factor Rotation 158
4.9 Estimation of Factor Scores 177
4.10 Factor Analysis Using Residuals 184
4.11 Some Applications 188
4.12 Concluding Remarks 209

5 Discriminant Analysis 211
5.1 Introduction 211
5.2 Multivariate Normality 212
5.3 Statistical Tests for Relevance 231
5.4 Discriminant Analysis: Fisher's Approach 242
5.5 Discriminant Analysis for k Normal Populations 255
5.6 Canonical Discriminant Analysis 282
5.7 Variable Selection in Discriminant Analysis 296
5.8 When Dimensionality Exceeds Sample Size 304
5.9 Logistic Discrimination 314
5.10 Nonparametric Discrimination 333
5.11 Concluding Remarks 344

6 Cluster Analysis 347
6.1 Introduction 347
6.2 Graphical Methods for Clustering 348
6.3 Similarity and Dissimilarity Measures 356
6.4 Hierarchical Clustering Methods 359
6.5 Clustering of Variables 380
6.6 Nonhierarchical Clustering: k-Means Approach 393
6.7 How Many Clusters: Cubic Clustering Criterion 421
6.8 Clustering Using Density Estimation 427
6.9 Clustering with Binary Data 435
6.10 Concluding Remarks 441

7 Correspondence Analysis 443
7.1 Introduction 443
7.2 Correspondence Analysis 444

7.3 Multiple Correspondence Analysis 463
7.4 CA as a Canonical Correlation Analysis 476
7.5 Correspondence Analysis Using Andrews Plots 479
7.6 Correspondence Analysis Using Hellinger Distance 490
7.7 Canonical Correspondence Analysis 498
7.8 Concluding Remarks 509

Appendix: Data Sets **511**

References **535**

Index **543**