

# Contents

**Preface**

*page ix*

## PART ONE: INTRODUCTION

<b>1. Introduction</b>	<b>3</b>
1.1 Examples	3
1.2 Notation	5
1.3 Review of objectives and different approaches	6
1.4 Some general comments	9
1.5 Review of books on multivariate analysis	10
1.6 Some matrix algebra revision	11
1.7 The general linear model	15
Exercises	17
<b>2. Multivariate distributions</b>	<b>18</b>
2.1 Multivariate, marginal and conditional distributions	18
2.2 Means, variances, covariances and correlations	23
2.3 The multivariate normal distribution	28
2.4 The bivariate normal distribution	29
2.5 Other multivariate distributions	31
2.5.1 <i>Multivariate discrete distributions</i>	31
2.5.2 <i>Multivariate continuous distributions</i>	31
Exercises	32
<b>3. Preliminary data analysis</b>	<b>34</b>
3.1 Processing the data	34
3.1.1 <i>Data editing</i>	36
3.2 Calculating summary statistics	38
3.2.1 <i>Interpreting the sample correlation matrix</i>	40
3.2.2 <i>The rank of <math>R</math></i>	48
3.3 Plotting the data	49
3.4 The analysis of discrete data	51
Exercises	53

## PART TWO: FINDING NEW UNDERLYING VARIABLES

<b>4. Principal component analysis</b>	<b>57</b>
4.1 Introduction	57
4.2 Derivation of principal components	58
4.2.1 <i>Principal components from the correlation matrix</i>	61
4.2.2 <i>Estimating the principal components</i>	62
4.3 Further results on PCA.	63
4.3.1 <i>Mean-corrected component scores</i>	64
4.3.2 <i>The inverse transformation</i>	64
4.3.3 <i>Zero eigenvalues</i>	64
4.3.4 <i>Small eigenvalues</i>	64
4.3.5 <i>Repeated roots</i>	65
4.3.6 <i>Orthogonality</i>	66
4.3.7 <i>Component loadings/component correlations</i>	66
4.3.8 <i>Off-diagonal structure</i>	67
4.3.9 <i>Uncorrelated variables</i>	68
4.4 The problem of scaling in PCA	68
4.5 Discussion	71
4.5.1 <i>The identification of important components</i>	72
4.5.2 <i>The use of components in subsequent analyses</i>	75
4.6 PCA for multivariate normal data	77
4.7 Summary	78
Exercises	79
<b>5. Factor analysis</b>	<b>82</b>
5.1 Introduction	82
5.2 The factor-analysis model	83
5.3 Estimating the factor loadings	86
5.4 Discussion	87

## PART THREE: PROCEDURES BASED ON THE MULTIVARIATE NORMAL DISTRIBUTION

<b>6. The multivariate normal distribution</b>	<b>93</b>
6.1 Introduction	93
6.2 Definition of the multivariate normal distribution	93
6.3 Properties of the multivariate normal distribution	97
6.4 Linear compounds and linear combinations	101
6.5 Estimation of the parameters of the distribution	103
6.6 The Wishart distribution	104
6.7 The joint distribution of the sample mean vector and the sample covariance matrix	107
6.8 The Hotelling $T^2$ -distribution	108
Exercises	111

<b>7. Procedures based on normal distribution theory</b>	<b>114</b>
7.1 Introduction	114
7.2 One-sample procedures	114
7.3 Confidence intervals and further analysis	117
7.4 Tests of structural relations among the components of the mean	122
7.5 Two-sample procedures	124
7.6 Confidence intervals and further analysis	125
7.7 Tests of structural relations among the components of the means	129
7.8 Discriminant analysis	133
Exercises	138
<b>8. The multivariate analysis of variance</b>	<b>140</b>
8.1 Introduction	140
8.2 MANOVA calculations	140
8.3 Testing hypotheses	144
8.3.1 <i>The special case: The univariate procedure</i>	144
8.3.2 <i>The multivariate model for Example 8.1</i>	145
8.3.3 <i>Multivariate test procedures</i>	146
8.3.4 <i>Distributional approximations</i>	148
8.3.5 <i>Applications of the methodology</i>	149
8.4 Further analysis	151
8.5 The dimensionality of the alternative hypothesis	152
8.6 Canonical variates analysis	153
8.7 Linear functional relationships	159
8.8 Discriminant analysis	160
Exercises	160
<b>9. The multivariate analysis of covariance and related topics</b>	<b>162</b>
9.1 Introduction	162
9.2 Multivariate regression	162
9.2.1 <i>The special case: Univariate multiple regression</i>	163
9.2.2 <i>The general case: Multivariate regression</i>	165
9.3 Canonical correlation.	169
9.4 The multivariate analysis of covariance	173
9.4.1 <i>The special case: Univariate analysis of covariance</i>	173
9.4.2 <i>The multivariate case: An example</i>	176
9.4.3 <i>The multivariate case: General results</i>	178
9.5 The test for additional information	180
9.6 A test of an assigned subset of linear compounds	184
Exercises	186

## PART FOUR: MULTIDIMENSIONAL SCALING AND CLUSTER ANALYSIS

<b>10. Multidimensional scaling.</b>	<b>189</b>
10.1 Introduction	189
10.2 Measures of similarity and dissimilarity	190
10.2.1 <i>Similarity coefficients for binary data</i>	194
10.3 Classical scaling	198
10.3.1 <i>The calculation of co-ordinate values from Euclidean distances</i>	198
10.3.2 <i>The relationship between classical scaling and principal component analysis</i>	200
10.3.3 <i>Classical scaling for a dissimilarity matrix</i>	201
10.3.4 <i>Interpretation of the results</i>	202
10.3.5 <i>Some related methods</i>	204
10.4 Ordinal scaling	204
10.4.1 <i>The iterative procedure</i>	205
10.4.2 <i>Interpreting the results</i>	208
10.5 A comparison	209
10.6 Concluding remarks	210
Exercises	210
<b>11. Cluster analysis</b>	<b>212</b>
11.1 Introduction	212
11.1.1 <i>Objectives</i>	214
11.1.2 <i>Clumping, dissection and clustering variables</i>	215
11.1.3 <i>Some other preliminary points</i>	215
11.2 Visual approaches to finding a partition	216
11.3 Hierarchical trees	219
11.4 Single-link clustering	220
11.5 Some other clustering procedures	224
11.5.1 <i>Method or algorithm?</i>	226
11.6 A comparison of procedures	226
11.6.1 <i>Some desirable conditions for hierarchical clustering methods</i>	226
11.6.2 <i>A comparison</i>	227
Exercises	229
<b>References</b>	<b>231</b>
<b>Answers to exercises</b>	<b>235</b>
<b>Name Index</b>	<b>241</b>
<b>Subject Index</b>	<b>243</b>