

Contents

PREFACE	xv
1 ASPECTS OF MULTIVARIATE ANALYSIS	1
1.1 Introduction	1
1.2 Applications of Multivariate Techniques	3
1.3 The Organization of Data	5
<i>Arrays,</i>	5
<i>Descriptive Statistics,</i>	6
<i>Graphical Techniques,</i>	11
1.4 Data Displays and Pictorial Representations	19
<i>Linking Multiple Two-Dimensional Scatter Plots,</i>	20
<i>Graphs of Growth Curves,</i>	24
<i>Stars,</i>	26
<i>Chernoff Faces,</i>	27
1.5 Distance	30
1.6 Final Comments	37
Exercises	37
References	47
2 MATRIX ALGEBRA AND RANDOM VECTORS	49
2.1 Introduction	49
2.2 Some Basics of Matrix and Vector Algebra	49
<i>Vectors,</i>	49
<i>Matrices,</i>	54
2.3 Positive Definite Matrices	60
2.4 A Square-Root Matrix	65
2.5 Random Vectors and Matrices	66
2.6 Mean Vectors and Covariance Matrices	68
<i>Partitioning the Covariance Matrix,</i>	73
<i>The Mean Vector and Covariance Matrix</i>	
<i>for Linear Combinations of Random Variables,</i>	75
<i>Partitioning the Sample Mean Vector</i>	
<i>and Covariance Matrix,</i>	77
2.7 Matrix Inequalities and Maximization	78

Supplement 2A: Vectors and Matrices: Basic Concepts	82
<i>Vectors,</i>	82
<i>Matrices,</i>	87
Exercises	103
References	110

3 SAMPLE GEOMETRY AND RANDOM SAMPLING 111

3.1	Introduction	111
3.2	The Geometry of the Sample	111
3.3	Random Samples and the Expected Values of the Sample Mean and Covariance Matrix	119
3.4	Generalized Variance	123
	<i>Situations in which the Generalized Sample Variance Is Zero,</i>	129
	<i>Generalized Variance Determined by R</i>	
	<i>and Its Geometrical Interpretation,</i>	134
	<i>Another Generalization of Variance,</i>	137
3.5	Sample Mean, Covariance, and Correlation As Matrix Operations	137
3.6	Sample Values of Linear Combinations of Variables	140
	Exercises	144
	References	148

4 THE MULTIVARIATE NORMAL DISTRIBUTION 149

4.1	Introduction	149
4.2	The Multivariate Normal Density and Its Properties	149
	<i>Additional Properties of the Multivariate Normal Distribution,</i>	156
4.3	Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation	168
	<i>The Multivariate Normal Likelihood,</i>	168
	<i>Maximum Likelihood Estimation of η and Σ,</i>	170
	<i>Sufficient Statistics,</i>	173
4.4	The Sampling Distribution of X and S	173
	<i>Properties of the Wishart Distribution,</i>	174
4.5	Large-Sample Behavior of X and S	175
4.6	Assessing the Assumption of Normality	111
	<i>Evaluating the Normality of the Univariate Marginal Distributions,</i>	177
	<i>Evaluating Bivariate Normality,</i>	182
4.7	Detecting Outliers and Cleaning Data	187
	<i>Steps for Detecting Outliers,</i>	189
4.8	Transformations to Near Normality	192
	<i>Transforming Multivariate Observations,</i>	195
	Exercises	200
	References	208

5	INFERENCES ABOUT A MEAN VECTOR	210
5.1	Introduction	210
5.2	The Plausibility of μ_0 as a Value for a Normal Population Mean	210
5.3	Hotelling's T^2 and Likelihood Ratio Tests	216
	<i>General Likelihood Ratio Method,</i>	219
5.4	Confidence Regions and Simultaneous Comparisons of Component Means	220
	<i>Simultaneous Confidence Statements,</i>	223
	<i>A Comparison of Simultaneous Confidence Intervals with One-at-a-Time Intervals,</i>	229
	<i>The Bonferroni Method of Multiple Comparisons,</i>	232
5.5	Large Sample Inferences about a Population Mean Vector	234
5.6	Multivariate Quality Control Charts	239
	<i>Charts for Monitoring a Sample of Individual Multivariate Observations for Stability,</i>	241
	<i>Control Regions for Future Individual Observations,</i>	247
	<i>Control Ellipse for Future Observations,</i>	248
	<i>T^2-Chart for Future Observations,</i>	248
	<i>Control Charts Based on Subsample Means,</i>	249
	<i>Control Regions for Future Subsample Observations,</i>	251
5.7	Inferences about Mean Vectors when Some Observations Are Missing	251
5.8	Difficulties Due to Time Dependence in Multivariate Observations	256
	Supplement 5A: Simultaneous Confidence Intervals and Ellipses as Shadows of the p-Dimensional Ellipsoids	258
	Exercises	261
	References	272
6	COMPARISONS OF SEVERAL MULTIVARIATE MEANS	273
6.1	Introduction	273
6.2	Paired Comparisons and a Repeated Measures Design	273
	<i>Paired Comparisons,</i>	273
	<i>A Repeated Measures Design for Comparing Treatments,</i>	279
6.3	Comparing Mean Vectors from Two Populations	284
	<i>Assumptions Concerning the Structure of the Data,</i>	284
	<i>Further Assumptions When n_1 and n_2 Are Small,</i>	285
	<i>Simultaneous Confidence Intervals,</i>	288
	<i>The Two-Sample Situation When $\Sigma_1 \neq \Sigma_2$,</i>	291
	<i>An Approximation to the Distribution of T^2 for Normal Populations When Sample Sizes Are Not Large,</i>	294
6.4	Comparing Several Multivariate Population Means (One-Way Manova)	296
	<i>Assumptions about the Structure of the Data for One-Way MANOVA,</i>	296

	<i>A Summary of Univariate ANOVA</i>	297
	<i>Multivariate Analysis of Variance (MANOVA)</i>	301
6.5	Simultaneous Confidence Intervals for Treatment Effects	308
6.6	Testing for Equality of Covariance Matrices	310
6.7	Two-Way Multivariate Analysis of Variance	312
	<i>Univariate Two-Way Fixed-Effects Model with Interaction</i>	312
	<i>Multivariate Two-Way Fixed-Effects Model with Interaction</i>	315
6.8	Profile Analysis	323
6.9	Repeated Measures Designs and Growth Curves	328
6.10	Perspectives and a Strategy for Analyzing Multivariate Models	332
	Exercises	337
	References	358

7 MULTIVARIATE LINEAR REGRESSION MODELS

360

7.1	Introduction	360
7.2	The Classical Linear Regression Model	360
7.3	Least Squares Estimation	364
	<i>Sum-of-Squares Decomposition</i>	366
	<i>Geometry of Least Squares</i>	367
	<i>Sampling Properties of Classical Least Squares Estimators</i>	369
7.4	Inferences About the Regression Model	370
	<i>Inferences Concerning the Regression Parameters</i>	370
	<i>Likelihood Ratio Tests for the Regression Parameters</i>	374
7.5	Inferences from the Estimated Regression Function	378
	<i>Estimating the Regression Function at z_0</i>	378
	<i>Forecasting a New Observation at z_0</i>	379
7.6	Model Checking and Other Aspects of Regression	381
	<i>Does the Model Fit?</i>	381
	<i>Leverage and Influence</i>	384
	<i>Additional Problems in Linear Regression</i>	384
7.7	Multivariate Multiple Regression	387
	<i>Likelihood Ratio Tests for Regression Parameters</i>	395
	<i>Other Multivariate Test Statistics</i>	398
	<i>Predictions from Multivariate Multiple Regressions</i>	399
7.8	The Concept of Linear Regression	401
	<i>Prediction of Several Variables</i>	406
	<i>Partial Correlation Coefficient</i>	409
7.9	Comparing the Two Formulations of the Regression Model	410
	<i>Mean Corrected Form of the Regression Model</i>	410
	<i>Relating the Formulations</i>	412
7.10	Multiple Regression Models with Time Dependent Errors	413
	Supplement 7A: The Distribution of the Likelihood Ratio	
	for the Multivariate Multiple Regression Model	418
	Exercises	420
	References	428

430

- FACTOR ANALYSIS AND INFERENCE
FOR STRUCTURED COVARIANCE MATRICES 481

- 9.1 Introduction 481
- 9.2 The Orthogonal Factor Model 482
- 9.3 Methods of Estimation 488
 - The Principal Component (and Principal Factor) Method, 488*
 - A Modified Approach—the Principal Factor Solution, 494*
 - The Maximum Likelihood Method, 495*
 - A Large Sample Test for the Number of Common Factors, 501*
- 9.4 Factor Rotation 504
 - Oblique Rotations; 512*
- 9.5 Factor Scores 513
 - The Weighted Least Squares Method, 514*
 - The Regression Method, 516*
- 9.6 Perspectives and a Strategy for Factor Analysis 519
 - Supplement 9A: Some Computational Details for Maximum Likelihood Estimation 527
 - Recommended Computational Scheme, 528*
 - Maximum Likelihood Estimators of $\rho = L_Z L'_Z + \psi_Z$ 529*
 - Exercises 530
 - References 538

- 10.1 Introduction 539
- 10.2 Canonical Variates and Canonical Correlations 539
- 10.3 Interpreting the Population Canonical Variables 545
Identifying the Canonical Variables, 545
Canonical Correlations as Generalizations of Other Correlation Coefficients, 547
The First r Canonical Variables as a Summary of Variability, 548
A Geometrical Interpretation of the Population Canonical Correlation Analysis 549
- 10.4 The Sample Canonical Variates and Sample Canonical Correlations 550
- 10.5 Additional Sample Descriptive Measures 558
Matrices of Errors of Approximations, 558
Proportions of Explained Sample Variance, 561
- 10.6 Large Sample Inferences 563
 Exercises 567
 References 574

- 11.1 Introduction 575
- 11.2 Separation and Classification for Two Populations 576
- 11.3 Classification with Two Multivariate Normal Populations 584
Classification of Normal Populations When $\Sigma_1 = \Sigma_2 = \Sigma$, 584
Scaling, 589
Fisher's Approach to Classification with Two Populations, 590
Is Classification a Good Idea?, 592
Classification of Normal Populations When $\Sigma_1 \neq \Sigma_2$, 593
- 11.4 Evaluating Classification Functions 596
- 11.5 Classification with Several Populations 606
The Minimum Expected Cost of Misclassification Method, 606
Classification with Normal Populations, 609
- 11.6 Fisher's Method for Discriminating among Several Populations 621
Using Fisher's Discriminants to Classify Objects, 628
- 11.7 Logistic Regression and Classification 634
Introduction, 634
The Logit Model, 634
Logistic Regression Analysis, 636
Classification, 638
Logistic Regression with Binomial Responses, 640
- 11.8 Final Comments 644
Including Qualitative Variables, 644
Classification Trees, 644
Neural Networks, 647
Selection of Variables, 648

Testing for Group Differences,	648
Graphics,	649
Practical Considerations Regarding Multivariate Normality,	649
Exercises	650
References	669

12 CLUSTERING, DISTANCE METHODS, AND ORDINATION 671

12.1	Introduction	671
12.2	Similarity Measures	673
	<i>Distances and Similarity Coefficients for Pairs of Items,</i>	673
	<i>Similarities and Association Measures for Pairs of Variables,</i>	677
	<i>Concluding Comments on Similarity,</i>	678
12.3	Hierarchical Clustering Methods	680
	<i>Single Linkage,</i>	682
	<i>Complete Linkage,</i>	685
	<i>Average Linkage,</i>	690
	<i>Ward's Hierarchical Clustering Method,</i>	692
	<i>Final Comments—Hierarchical Procedures,</i>	695
12.4	Nonhierarchical Clustering Methods	696
	<i>K-means Method,</i>	696
	<i>Final Comments—Nonhierarchical Procedures,</i>	701
12.5	Clustering Based on Statistical Models	703
12.6	Multidimensional Scaling	706
	<i>The Basic Algorithm,</i>	708
12.1	Correspondence Analysis	716
	<i>Algebraic Development of Correspondence Analysis,</i>	718
	<i>Inertia,</i>	725
	<i>Interpretation in Two Dimensions,</i>	726
	<i>Final Comments,</i>	726
12.8	Biplots for Viewing Sampling Units and Variables	726
	<i>Constructing Biplots,</i>	727
12.9	Procrustes Analysis: A Method for Comparing Configurations	732
	<i>Constructing the Procrustes Measure of Agreement,</i>	733
	Supplement 12A: Data Mining	740
	<i>Introduction,</i>	740
	<i>The Data Mining Process,</i>	741
	<i>Model Assessment,</i>	742
	Exercises	747
	References	755

APPENDIX 757

DATA INDEX 764

SUBJECT INDEX 767