Contents

Preface

Chapter 6 CONTINGENCY TABLES

6.1 Multivariate Data Analysis, Data Matrices and

vii

2

8

10

11

12

17

24

27

Measurement Scales
6.1.1 Data Matrices
6.1.2 Measurement Scales

Quantitative Scales, Qualitative Scales, Measurement
Scales and Analysis

6.1.3 Data Collection and Statistical Inference

Probability Samples and Random Samples, Exploratory
and Confirmatory Analysis

6.1.4 An Outline of the Techniques to be Studied

Topics in Volume II

6.2 Two-Dimensional Contingency Tables

6.2.1 Bivariate Distributions for Categorical Data Joint Density Table, Indepencence, Row and Column Proportions, Row and Column Profiles, Odds Ratios
6.2.2 Statistical Inference in Two-Dimensional Tables The Two-Dimensional Contingency Table, Sampling Models for Contingency Tables, Multinomial, Hypergeometric, Poisson, Product Multinomial, Test of Independence, Sampling Model Assumptions, Poisson Distribution, Product Multinomial Distribution, Standardized Residuals, Correspondence Analysis
6.2.3 Measures of Association

Goodman and Kruskal's Lambda, Inference for Lambda

6.2.4 Models for Two-Dimensional Tables Equal Cell Probability Model, Constant Row or Column Densities, The Independence Model as a Composite of Three Simple Models, The Saturated Model, Loglinear Characterization for Cell Densities, A Loglinear Model for Independence, Parameters for the Loglinear Model, The Loglinear Model with Interaction, Matrix Notation for Loglinear Model

6.2.5 Statistical Inference for Loglinear Models The Loglinear Model Defined in Terms of Cell Frequencies, Multiplicative Form of the Loglinear Model. Estimation for the Loglinear Model, Standardized Estimates of Loglinear Parameters, A Loglinear Representation for Some Simpler Models, Inference Procedures for the Three Simple Models 6.2.6 An Additive Characterization for Cell Densities 6.2.7 Two-Dimensional Contingency Tables in a Multivariate Setting Simpson's Paradox 6.2.8 Other Sources of Information 6.3 Multidimensional Contingency Tables 6.3.1 The Three-Dimensional Contingency Table Models for Three-Way Tables, Inference for the Independence Model, Other Models for Three-Way Tables, Partial Independence, Conditional Independence, No Three-Way Interaction, Saturated Model, Loglinear Models for Three-Way Tables, Definitions of Parameters in Terms of Cell Frequencies, Independence Model, Partial Independence Model, Conditional Independence Model, No Three-Way Interactions Model, Saturated Model, Multiplicative Form of the Loglinear Model, Hierarchical Models, Notation for Loglinear Models, Model Selection, Standardized Estimates and Standardized Residuals, Summary of Loglinear Model Fitting Procedure, Product Multinomial Sampling

43

46

46

47

60

70

77

6.3.2 Some Examples

Three-way Interaction, Goodness of Fit and Model Selection

6.3.3 Four-Dimensional Contingency Tables and Stepwise Fitting Procedures Stepwise Model Selection, Tests of Partial and Marginal Association, Marginal Association
6.3.4 The Effects of Collapsing a Contingency Table and Structural Zeroes Collapsing Contingency Tables, Random Zeroes, Structural Zeroes and Incomplete Tables, Quasi-loglinear Models for Incomplete Tables

Contents xv

82

87

87

87

96

98

102

104

115

116

117

121

131

131

132

6.3.5 Logit Models for Response Variables The Logit Function, Fitting a Logit Model, Relationship to Logistic Regression, Polychotomous Response Variables

6.3.6 Other Sources of Information

6.4 The Weighted Least Squares Approach
6.4.1 The Weighted Least Squares Theory The Product Multinomial Distribution Assumption, Sampling Properties of the Row Proportions, Determining Linear Functions Among the Row Proportions, The Linear Model to be Estimated, Determining the Weighted Least Squares Estimator
6.4.2 Statistical Inference for the Weighted Least Squares Procedure
6.4.3 Some Alternative Analyses Marginal Analysis, Continuation Differences, Averaging or Summing Response Functions, Weighted Sums for Ordinal Responses

6.4.4 Weighted Least Squares Estimation for Logit Models The Logit Model as a Special Case of a Weighted

Least Squares Model, Continuation Ratios 6.4.5 Two or More Response Variables Defining Response Functions, Repeated Measurement Designs, Adding Interaction Effects 6.4.6 Other Sources of Information Cited Literature and References Exercises for Chapter 6 Questions for Chapter 6



 MULTIVARIATE DISTRIBUTIONS, INFERENCE, REGRESSION and CANONICAL CORRELATION
 7.1 Multivariate Random Variables and Samples
 7.1.1 Multivariate Distributions and Multivariate Random Variables
 Joint Distribution, Partitioning the Random Variable, Conditional Distributions and Independence, Mean Vector and Covariance Matrix, Correlation Matrix

xvi Contents

7.1.2 Multivariate Samples

136

140

144

144

145

146

148

148

150

156

157

157

157

Sample Mean Vector and Covariance Matrix, Sample Correlation Matrix, Sums of Squares and Cross Product Matrices, Multivariate Central Limit Theorem

7.1.3 Geometric Interpretations for Data Matrices

p-Dimensional Space, n-Dimensional Space, Mahalanobis Distance and Generalized Variance, p-Dimensional Ellipsoid, Generalized Variance, Trace Measure of Overall Variance, Generalized Variance for Correlation Matrices, Eigenvalues and Eigenvectors for Sums of Squares and Cross Product Matrices
7.1.4 Other Sources of Information

7.2 The Multivariate Normal Distribution

7.2.1 The Multivariate Normal Density, Constant Probability Density Contour, Linear Transformations, Distribution
7.2.2 Partitioning the Normal Marginal Distributions, Conditional Distributions, Mul-

tivariate Regression Function, Partial Correlation 7.3 Testing for Normality, Outliers and Robust Estimation 7.3.1 Testing for Normality Mahalanobis Distances from the Sample Mean, Multivariate Skewness and Kurtosis, Transforming to Normality

7.3.2 Multivariate Outliers Multivariate Outliers and Mahalanobis Distance, Testing for Multivariate Outliers, Multiple Outliers
7.3.3 Robust Estimation

Obtaining Robust Estimators of Covariance and Cor-

relation Matrices, Multivariate Trimming 7.3.4 Other Sources of Information

7.4 Inference for the Multivariate Normal

7.4.1 Inference Procedures for the Mean Vector Sample Likelihood Function, Hotelling's T^2 , Inference, Simultaneous Confidence Regions, Inferences for Linear Functions

Contents xvii

160

7.4.2 Repeated Measures Comparisons Repeated Measurements on a Single Variable, Profile Characterization, Repeated Measures in a Randomized Block Design, Necessary and Sufficient Conditions for Validity of Univariate F Test

- 7.4.3 Mahalanobis Distance of the Mean Vector from the Origin 163 Mahalanobis Distance of Mean Vector from the Origin, Application to Financial Portfolios
- 7.4.4 Inference for the Covariance and the Correlation Matrices 165

Wishart Distribution, Sphericity Test and Test for Independence, A Test for Zero Correlation, Test Statistics for Repeated Measures Designs, Test for Equal Variance-Equal Covariance Structure, Test for the Hyunh-Feldt Pattern, Equal Correlation Structure, Independent Blocks, Partial and Multiple Correlation

7.4.5 Other Sources of Information

 7.5 Multivariate Regression and Canonical Correlation
 7.5.1 Multivariate Regression
 The Multivariate Regression Function, Estimation of the Multivariate Regression Model, Relationship to 169 169 170

181

190

191

193

Ordinary Least Squares, Residuals, Influence, Outliers and Cross Validation, Estimation of the Error Covariance Matrix, Relationship to Multiple Linear Regression, Testing the Hypothesis that Some Coefficients are Zero, Other Tests, Inferences for Linear Functions, Relationship to Generalized Least Squares, Zellner's Seemingly Unrelated Regression Model

7.5.2 Canonical Correlation

Derivation of Canonical Relationships, An Eigenvalue Problem, The Canonical Variables, Sample Canonical Correlation Analysis, Canonical Weights and Canonical Variables, Inference for Canonical Correlation, An Alternative Test Statistic, Structure Correlations or Canonical Loadings, Redundancy Analysis and Proportion of Variance Explained, Redundancy Measure for a Given Canonical Variate, Total Redundancy, Relation to Multiple Regression, Residuals, Influence, Outliers and Cross Validation 7.5.3 Other Sources of Information Cited Literature and References Exercises for Chapter 7

Questions for Chapter 7

195

209

209

209

Chapter 8 MANOVA, DISCRIMINANT ANALYSIS and QUALITATIVE RESPONSE MODELS

8.1 Multivariate Analysis of Variance 8.1.1 One-Way Multivariate Analysis of Variance Comparison to Univariate Analysis of Variance, Notation for Several Multivariate Populations, Mean Vector for Group k and Common Covariance Matrix, Grand Mean Vector, Notation for Samples, Sample Mean Vector and Sample Covariance Matrix for Group k, Sample Grand Mean Vector, The Multivariate Analysis of Variance Model, Within Group Sum of Squares Matrix, Among Group Sum of Squares Matrix, Total Sum of Squares Matrix, Statistical Inference for MANOVA, Wilk's Lambda Likelihood, Ratio F-Statistic, An Alternative Test Statistic, Correlation Ratio, The Special Case of Two Groups, A Bonferroni Approximation, Multiple Comparison Procedures Based on Two Group Comparisons, Testing for the Equality of Covariance Matrices

8.1.2 Indicator Variables, Multivariate Regression and Analysis of Covariance Some Relationships to the Multivariate Regression Test for H_0 : **ABM** = 0, Cell Parameter Coding, The Non-Full Rank Design Matrix, Multivariate Analysis of Covariance

8.1.3 Profile Analysis with Repeated Measurements Comparing Profiles, Parallel Profiles, Equal Profiles Given Parallel Profiles, Horizontal Profiles Given Parallel Profiles, Horizontal Profiles

8.1.4 Balanced Two-Way MANOVA The Model, Sums of Squares Matrices, Inference, The Multivariate Paired Comparison Test
8.1.5 An Unbalanced MANOVA with Covariate
8.1.6 Other Sources of Information

8.2 Discriminant Analysis

8.2.1 Fisher's Discriminant Criterion and Canonical Discriminant Analysis Fisher's Discriminant Criterion, An Eigenvalue Problem, Canonical Discriminant Functions, Inferences for Canonical Discriminant Functions, Bartlett's Test, 239 241 242

244

222

229

234

Contents xix

An Alternative Test Statistic-F, Interpretation of the Discriminant Analysis Solution, Interpretation Using Correlations, Graphical Approach to Group Characterization, Comparison of Correlation Coefficients and Discriminant Function Coefficients, Effect of Correlation Structure on Discriminant Analysis, Discriminant Analysis and Canonical Correlation, Discriminant Analysis and Dimension Reduction

8.2.2 Discriminant Functions and Classification Discrimination Between Two Groups with Param-

eters Known, Classification of an Unknown, Fisher Criterion and Mahalanobis Distance, Maximum Likelihood Criterion, Minimum Total Probability of Misclassification Criterion, Bayes Theorem Criterion, Minimax Criterion, Minimum Cost Criterion, Summary, Quadratic Discriminant Function and Unequal Covariance Matrices, Classification in Practice, Evaluation of a Discriminant Function as a Classification Mechanism, Split Sample, Jackknife Procedure, Multiple Group Classification, Bias When Parameters are Unknown

8.2.3 Tests of Sufficiency and Variable Selection

1. 1. M.

274

278

279

258

- Two Groups, More Than Two Groups8.2.4 Discrimination Without Normality
Discrimination Using Ranks, Nearest Neighbor Method2778.2.5 Other Sources of Information278
- 8.3 Qualitative Response Regression Models and Logistic Regression
 - 8.3.1 The Dichotomous Response Model The Point Binomial, Probability as a Function of Other Variables, Alternative Response Functions, Logistic Regression with c Explanatory Variables, Maximum Likelihood Estimation for Dichotomous Logistic Regression, Newton-Raphson Procedure, Infer-

ence for the Dichotomous Logistic Regression Model, Comparing Nested Models and Inference for Coefficients, Goodness of Fit, Hosmer-Lemeshow Goodness of Fit Test, Covariance Matrix for Estimated Coefficients, The Role of the Intercept and Categorical Variables, Testing for Zero Intercept, Dummy Variables as Explanatory Variables – A Caution, The Fitted Model and Classification, The Jackknife Approach, Stepwise Logistic Regression, Influence Di-

agnostics, The Chi Statistic, The Deviance Statistic, Leverage, Influence, The DFBETA Measure 299 8.3.2 The Probit Model 8.3.3 Logistic Regression and Probit Analysis: A Second 303 Example 8.3.4 Multiple Observations and Design Variables 304 The Model and Maximum Likelihood Estimation, The Chi and Deviance Statistics, Weighted Least Squares or Minimum Logit Chi-Square Estimation 306 8.3.5 Other Sources of Information 306 8.3.6 The Multinomial Logit Model Parameterization of the Model, Inference for the Multinomial Logit, Using Multinomial Logit Models, Estimation Using Single Equation Methods, Continuation Ratios, Other Nested Partitions, 321 8.3.7 Other Sources of Information 8.3.8 The Conditional Logit Model and Consumer Choice 321 8.3.9 Multivariate Qualitative Response Models 322Loglinear Models for Dependent Variables, Relation

Between Loglinear Parameters and Logits, A Conditional Probability Approach

Cited Literature and References

Exercises for Chapter 8

Questions for Chapter 8

Chapter 9 PRINCIPAL COMPONENTS, FACTORS and CORRESPONDENCE ANALYSIS 9.1 Principal Components 9.1.1 A Classic Example 329 334

345

040

346

346

346

350

267

201

327

9.1.2 An Ad Hoc Approach

9.1.3 The Principal Components Approach Characterizing the First Principal Component, The Eigenvalue Problem, Generalization to *r* Principal Components, Spectral Decomposition, The Full Rank Case, Alternative Characterizations and Geometry, Principal Components and Multivariate Random Variables, Principal Component Scores

9.1.4 The Various Forms of X'X and Principal Components Interpretations Using Correlations, Standardized Principal Components, Communality or Variance Explained,

Contents XXI

378

How Many Principal Components, Average Criterion, Geometric Mean Criterion, A Test for Equality of Eigenvalues in Covariance Matrices, A Cross Validation Approach, Should all the Variables be Retained

9.1.5 Principal Components, Multiple Regression and Supplementary Points Multiple Regression, Supplementary Dimensions and Points

9.1.6 Outliers and Robust Principal Components Analysis 381 Identification of Outliers, Influence, Robust Principal Components Analysis, Rank Correlation and Robust Principal Components Analysis 9.1.7 Other Sources of Information 388 9.2 The Exploratory Factor Analysis Model 388 9.2.1 The Factor Analysis Model and Estimation 389 The Model, Factor Analysis Using the Correlation Matrix, Indeterminacy, Estimation of the Factor Model Using Principal Components, Estimation of the Common Factor Model, Determination of the Number of Factors, A Useful Preliminary Test, Scree Test,

> The Broken Stick Model, Equal Correlation Structure and the Number of Factors, Principal Factor Approach

9.2.2 Factor Rotation

398

The Theory of Rigid Rotation, Varimax, Other Rotation Methods, Quartimax Criterion, Orthomax, Oblique Rotation, Procrustes Rotation, The Geometry of Factor Analysis

9.2.3 Factor Scores

9.2.4 The Maximum Likelihood Estimation Method The Maximum Likelihood Approach, Goodness of Fit, Cross Validation, Akaike and Schwartz Criteria

410

413

417 9.2.5 Results From a Simulation Study 419 9.2.6 A Second Example 426 9.2.7 Other Sources of Information 9.3 Singular Value Decomposition and Matrix Approximation 426 9.3.1 Singular Value Decomposition and Principal Components 427 428 9.3.2 Biplots and Matrix Approximation Constructing Biplots, The Principal Components Biplot, Covariance Biplot, Symmetric Biplot

9.3.3 Other Sources of Information

9.4 Correspondence Analysis 9.4.1 Correspondence Analysis for Two-Dimensional Tables Some Notation, Correspondence Matrix and Row and Column Masses, Row and Column Profiles, Departure from Independence, Averaging the Profiles, Relationship to Pearson Chi-square Statistic, Total Inertia, Generalized Singular Value Decomposition, Coordinates for Row and Column Profiles, Partial Contributions to Total Inertia, Squared Cosines, Principle of Distributional Equivalence, Generalized Least Squares Approximation, Relationship to Generalized Singular Value Decomposition of O, Row and Column Profile Deviations and Eigenvectors, Correspondence Analysis for Multidimensional Tables 9.4.2 Other Sources of Information 9.4.3 Correspondence Analysis and Frequency Response Tables A Dual Scaling Approach, Review of One-way ANOVA Notation, Scaling the Response Categories, Some Alternative Approaches to Correspondence Analysis, Bivariate Correlation, Simultaneous Linear Regression, Canonical Correlation 9.4.4 Other Sources of Information 9.4.5 Correspondence Analysis in Multidimensional Tables Multiple Correspondence Analysis and Burt Matrices 9.4.6 Other Sources of Information **Cited Literature and References** Exercises for Chapter 9

453

453

461

462

431

431

433

Questions for Chapter 9

Chapter 10
CLUSTER ANALYSIS and MULTIDIMENSIONAL SCALING
10.1 Proximity Matrices Derived from Data Matrices
10.1.1 The Measurement of Proximity Between Objects Similarity, Dissimilarity, Euclidean Distance, Using Mean-Centered Variables, Euclidean Distance in Matrix Form, Standardized Euclidean Distance, Mahalanobis Distance and Multivariate Distance, Euclidean Distance and the Centroid, Manhattan or City Block 469

474

467

465

483484486

Contents xxiii

Metric, Minkowski Metrics, Distance Measures Averaged Over Variables, Correlation Type Measures of Similarity, Similarity Matrices, Double Mean-Centered, Profile Shape, Scatter and Level, Some Relationships Between Similarity and Euclidean Distance, Proximity Measures for Categorical Data, Matching Coefficients for Binary Variables, Mixtures of Categorical and Interval Scaled Variables

10.1.2 The Measurement of Proximity Between Groups Single Linkage or Nearest Neighbor, Complete Link-

age of Furthest Neighbor, Average Linkage, An Algorithm for Updating the Proximity Measures, Distance Between Centroids, Incremental Sums of Squares, Relationship to Analysis of Variance, Algorithms for Determining Proximity, Measures Based on Centroids and Sums of Squares, Ultrametric Inequality, Sums of Squares Derived from MANOVA Matrices, A Multivariate Measure of Proximity

10.2 Cluster Analysis

10.2.1 Hierarchical Methods Agglomerative versus Divisive Processes, Comparison of Group Proximity Measures, Some Multivariate Approaches to Hierarchical Clustering, An Example with Outliers 518 519

10.2.2 Assessing the Hierarchical Solution and Cluster Choice Dendograms and Derived Proximities, Cophenetic Correlation and Cluster Validity, Stress, Alternative Derived Proximities Based on Centroids, Choosing the Number of Clusters, A Binary Data Example, Test Statistics for Number of Clusters, Some ANOVA-Type Statistics, Pseudo-F, Pseudo- t^2 and Beales' F-Ratio, R^2 -Type Measures, Correlation-Type Measures of Cluster Quality, Point-Biserial Correlation,

534

553

Gamma and G(+),

10.2.3 Combining Hierarchical Cluster Analysis With Other Multivariate Methods Interpretation of the Cluster Solution, ANOVA, MANOVA and Discriminant Analysis, Principal Components and Factor Analysis, Principal Components Analysis Prior to Cluster Analysis

xxiv Contents

10.2.4 Other Clustering Methods Partitioning Methods, The k-Means Algorithm, Selecting the Initial Partition, Classification Typologies and Q-Sort Methods, Density Methods, Clumping Techniques or Fuzzy Clustering

10.2.5 Cluster Validity and Cluster Analysis Methodology Cluster Validity, Monte Carlo Studies, The Underlying Cluster Population, Evaluation of Clustering Algorithms, Evaluation of Internal Criterion Mea-

sures, Cluster Choice, Variable Standardization Procedures, On the Measurement of Cluster Recovery and External Measurement Criteria 10.2.6 Other Sources of Information 10.3 Multidimensional Scaling 10.3.1 Metric Multidimensional Scaling Constructing a Positive Semidefinite Matrix Based on D, The Fundamental Theorem of MDS, The MDS Solution, An Approximate Solution, Metric Multidimensional Scaling Beginning with D, Relation to Cluster Analysis, Improving the Solution, Using Similarities, Metric MDS and Principal Coordinates Analysis, Relation to Cluster Analysis, An Alternative Derivation for A, The Additive Constant Problem, **Application of Metric Scaling** 10.3.2 Nonmetric Multidimensional Scaling Ordinal Scaling, Shepard-Kruskal Algorithm, The Nonmetric Phase and Monotone Regression, The Pool Adjacent Violators Algorithm, Ties and Types of Monotonicity, Ties in the Original Dissimilarities, The Metric Phase, The Evaluation Phase, Selection and Interpretation Phase, Monte Carlo Studies of the Stress Function, The ALSCAL Algorithm 10.3.3 Other Scaling Models Individual Difference Models, Preference Models and Multidimensional Unfolding 10.3.4 Other Sources of Information **Cited Literature and References** Exercises for Chapter 10 Questions for Chapter 10

584

601

602

603

606

612

559

563

568

568

570

Contents XXV

APPENDIX

- 1. Matrix Algebra
 - 1.1 Matrices

Matrix, Transpose of a Matrix, Row Vector and Column Vector, Square Matrix, Symmetric Matrix, Diagonal Elements, Trace of a Matrix, Null or Zero Matrix, Identity Matrix, Diagonal Matrix, Submatrix

1.2 Matrix Operations





Equality of Matrices, Addition of Matrices, Additive Inverse, Scalar Multiplication of a Matrix, Product of Two Matrices, Multiplicative Inverse, Idempotent Matrix, Kronecker Product

1.3 Determinants and Rank Determinant, Nonsingular, Relation Between Inverse and Determinant, Rank of a Matrix

1.4 Quadratic Forms and Positive Definite Matrices Quadratic Form, Congruent Matrix, Positive Definite, Positive Semidefinite, Negative Definite, Nonnegative Definite

630

625

1.5 Partitioned Matrices Product of Partitioned Matrices, Inverse of a Partitioned Matrix, Determinant of a Partitioned Matrix **1.6 Expectations of Random Matrices** 1.7 Derivatives of Matrix Expressions 2. Linear Algebra 2.1 Geometric Representation for Vectors n Dimensional Space, Directed Line Segment, Coordinates, Addition of Vectors, Scalar Multiplication, Length of a Vector, Angle Between Vectors, Orthogonal Vectors, Projection

636

639

2.2 Linear Dependence And Linear Transformations Linearly Dependent Vectors, Linearly Independent Vectors, Basis for an *n*-Dimensional Space, Generation of a Vector Space and Rank of a Matrix, Linear Transformation, Orthogonal Transformation, Rotation, Orthogonal Matrix

2.3 Systems of Equations Solution Vector for a System of Equations, Homogeneous Equations – Trivial and Nontrivial Solutions