Contents

CHAPTER I  Descriptive Principal Components Analysis and Singular Value Decomposition  1

I.1  Introduction  1
     I.1.1  Scope of the Method  2
     I.1.2  A Geometric Interpretation of Principal Components Analysis  3

I.2  General Analysis  4
     I.2.1  Fitting the Data Points in $\mathbb{R}^p$  4
     I.2.2  Relationship Between Fit in $\mathbb{R}^p$ and $\mathbb{R}^n$  6
     I.2.3  Recreating the Original Data  8

I.3  Applying Principal Components Analysis  10
     I.3.1  Analysis in $\mathbb{R}^p$  10
     I.3.2  Analysis in $\mathbb{R}^n$  12
     I.3.3  Supplementary Variables and Supplementary Individuals  14
     I.3.4  Nonparametric Analysis  16

I.4  Implementation of Principal Components Analysis  18
     I.4.1  Presentation and Interpretation of Results  18
     I.4.2  Example  20

I.5  Mathematical Appendix  26

CHAPTER II  Correspondence Analysis  30

II.1  Geometry of the Configuration of Points and Goodness of Fit Criterion  31
     II.1.1  Construction of the Configuration  31
CONTENTS

II.1.2 Choice of Distances  34
II.1.3 Choice of Goodness of Fit Criterion  36
II.1.4 Summary  36

II.2 Calculation of Principal Axes and Coordinates  37
II.2.1 General Analysis with Any Distance and Any Criteria  38
II.2.2 Analysis in $\mathbb{R}^p$, Calculation of Factors  39
II.2.3 Relationship with Analysis in $\mathbb{R}^n$  40
II.2.4 Supplementary Elements  41
II.2.5 Another Aspect of Correspondence Analysis—Finding the Best Simultaneous Representation  43

II.3 Interpretation of Results  44
II.3.1 Introduction  44
II.3.2 Calculation of Absolute Contributions and Squared Correlations  46

II.4 An Application Example  49
II.4.1 Data  49
II.4.2 Results and Interpretation  51

II.5 Computations  58
II.5.1 Analysis with Respect to the Center of Gravity  58
II.5.2 Symmetrization of $S$  60

CHAPTER III Canonical Analysis and Discriminant Analysis—Theoretical and Technical Considerations: A Brief Review  63

III.1 Canonical Analysis  63
III.1.1 Notation and Formulation of the Problem  64
III.1.2 Calculation of the Canonical Variables  65

III.2 Discriminant Analysis  69
III.2.1 Formulation of Problem  70
III.2.2 Calculation of the Linear Discriminant Functions  73
III.2.3 Relationship with Canonical Analysis  74
CHAPTER IV  Multiple Correspondence Analysis  81

IV.1  Definitions and Notation  81
      IV.1.1  Notation  82
      IV.1.2  Burt's Table Associated with Z  84

IV.2  Two Questions
      (Two-Way Correspondence)  84
      IV.2.1  First Equivalence  84
      IV.2.2  Second Equivalence  86

IV.3  Generalization to More Than Two Questions  88

IV.4  Properties of Multiple Analyses  91
      IV.4.1  $\varphi_q$ Components Are Centered  92
      IV.4.2  Proportion of Variance Due to One Question and to One Category  92
      IV.4.3  Dimensionality of the Configuration of the $p$ Categories in $\mathbb{R}^n$  93
      IV.4.4  Best Simultaneous Representation  94
      IV.4.5  Confidence Interval of a Supplementary Category  94

IV.5  Two Special Cases  95
      IV.5.1  All the Questions Have Two Categories  96
      IV.5.2  The Case Where Multiple Analysis Is Reduced to a Two-Way Analysis  97

IV.6  Example: Survey Processing  99
      IV.6.1  Operational Survey Procedures  99
      IV.6.2  An Example of "Predictive Map"  100

CHAPTER V  Automatic Classification—Clustering Techniques Used with Principal Axes Methods  109

V.1  Introduction—Classification and Principal Axes Methods  109

V.2  Clustering Around Moving Centers  111
      V.2.1  Theoretical Basis of the Algorithm  112
V.2.2 Elementary Rationale for the Algorithm 114
V.2.3 Related Techniques 115
V.2.4 Stable Clusters 116

V.3 Hierarchical Classification 117
V.3.1 Introduction 117
V.3.2 Single Linkage Classification and Minimum Spanning Tree 120
V.3.3 Minimum Variance Algorithms and Related Techniques 125

V.4 Reciprocal Neighbors—Chain Search Algorithm 128
V.4.1 Algorithm 128
V.4.2 Remarks 129

V.5 Mixed Methods for Large Data Sets 130
V.5.1 Preliminary Partition 130
V.5.2 Hierarchical Aggregation of the Stable Groups 130
V.5.3 Final Partition, and Description of Classes 130
V.5.4 Comments on Classification Strategy 132

V.6 Classification Example 132
V.6.1 Data 132
V.6.2 Classification Strategy 133
V.6.3 Description of the Groups 137

V.7 Appendix: Equivalence Between Single Linkage and Subdominant Ultrametric 143
V.7.1 Definition of the Distance of the Smallest Maximum Jump 143
V.7.2 Demonstration that \( d^* \) Is an Ultrametric Distance 143
V.7.3 Demonstration that \( d^* \) Is the Subdominant 144
V.7.4 Equivalence of \( d^* \) and \( d_u \) 144

CHAPTER VI Direct Reading Algorithms 146
VI.1 Reduction in the Number of Operations 146
VI.1.1 General Case 146
VI.1.2 The Case of Partitioning Algorithms 147
VI.2 Simultaneous Iterated Power Algorithm 148
    VI.2.1 First Decomposition of Matrix A (Reciprocal Averaging) 149
    VI.2.2 Second Decomposition of Matrix A 149

VI.3 Stochastic Approximation 151
    VI.3.1 Definitions and Notation 151
    VI.3.2 Convergence of the Algorithm 152
    VI.3.3 Comparison with Iterated Power 156

VI.4 Performing the Calculations 157

VI.5 Technical Appendix 158
    VI.5.1 Proof of Lemma 1 158
    VI.5.2 Discussion of Back-and-Forth Readings 159
    VI.5.3 Simultaneous Determination of q Principal Axes 159
    VI.5.4 Numerical Example 161

CHAPTER VII Reliability and Significance of Results 162

VII.1 Which Data Matrices Should We Analyze? How Do We Construct Them? 162

VII.2 What Can We Expect from Multivariate Descriptive Statistical Analysis? 164
    VII.2.1 Technical Advantages 164
    VII.2.2 Fundamental Advantages 165

VII.3 How Do We Evaluate the Quality of the Configurations? 166
    VII.3.1 Hypothesis of Independence 166
    VII.3.2 Percentage of Variance and Information 173
    VII.3.3 Stability of the Patterns 176
    VII.3.4 Confidence Areas for Points on Graphical Displays 182

VII.4 Appendix 1: Eigenvalues and Percentages of Variance in Correspondence Analysis 187
    VII.4.1 An Approximation of the Distribution of the Eigenvalues 187
    VII.4.2 Independence of the Percentages of Variance and of the Trace 189

VII.5 Appendix 2: Information and Eigenvalues 190

CHAPTER VIII A Computer Program: Correspondence Analysis for Large Matrices 193
CONTENTS

VIII.1 Main Features of the Program 193
VIII.2 Parameters 193
  VIII.2.1 Storage Requirements for CORAN 196
VIII.3 Technical Remarks 197
  VIII.3.1 Principal Computational Steps 197
  VIII.3.2 Possible Simplification of the Program 197
VIII.4 Comments on the Output Example 197
VIII.5 Test Data 198
VIII.6 Output Example 199

References 223

Index 229