

Dr Zlatko Kovacic (1994) *Multivariate analysis*. Faculty of Economics, University of Belgrade (set text for the fourth year course in *Multivariate analysis*).

MULTIVARIATE ANALYSIS

Table of contents

Forward iii

1 INTRODUCTION 1

- 1.1 Definition and classification of the multivariate analysis methods I
- 1.2 Type of data and measurement scales 5
- 1.3 Graphical display multivariate data 6

2 MULTIVARIATE DISTRIBUTIONS 13

- 2.1 Multivariate random variables 13
 - 2.1.1 Joint distribution 14
 - 2.1.2 Marginal distributions 14
 - 2.1.3 Conditional distributions 15
 - 2.1.4 Mean and covariance matrix 16
 - 2.1.5 Correlation matrix 18

3 THE MULTIVARIATE NORMAL DISTRIBUTION 39

- 3.1 Normal density function 39
- 3.2 Marginal and conditional distribution of the multivariate normal variables 44
- 3.3 Samples from multivariate normal distribution 50
- 3.4 Inference about a mean vector 52
 - 3.4.1 Confidence region for a mean vector 52
 - 3.4.2 Testing hypothesis about value of the mean vector 55
 - 3.4.3 Simultaneous confidence regions 57
- 3.5 Inference about covariance and correlation matrix 60

4 CANONICAL CORRELATION ANALYSIS 63

- 4.1 Introduction 63
- 4.2 Canonical variables and canonical correlation 64
 - 4.2.1 Definition and its properties 64
 - 4.2.2 Canonical correlation, correlation coefficients and regression analysis 72
- 4.3 Sample canonical correlation analysis 73
 - 4.3.1 Approximation error matrix 74
 - 4.3.2 Proportion of variance explained 78
 - 4.3.3 Redundancy measure 81
 - 4.3.4 Tests for canonical correlations 83
- 4.4 Interpretation of canonical variables and canonical correlations 85

5 MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA) 89

- 5.1 Introduction 89
- 5.2 Case of two populations 91
 - 5.2.1 Complete random design 91
 - 5.2.2 Paired comparisons 95
- 5.3 Repeated measure design 100
- 5.4 One-way MANOVA 105
 - 5.4.1 One-way MANOVA model 106
 - 5.4.2 Inference for one-way MANOVA 110
- 5.5 Profile analysis 117
- 5.6 Two-way MANOVA analysis 123
 - 5.6.1 Two-way MANOVA model 124
 - 5.6.2 Inference for two-way MANOVA 126

6 DISCRIMINANT ANALYSIS 131

- 6.1 Introduction 131
- 6.2 Descriptive discriminant analysis 134
 - 6.2.1 Fisher's approach - two populations 134
 - 6.2.2 Fisher's approach - more than two populations 138
 - 6.2.3 Inference - canonical discriminant analysis 146
 - 6.2.4 Interpretation of canonical discriminant analysis 147
- 6.3 Discriminant function and classification 152
 - 6.3.1 Classification for two populations 152
 - 6.3.2 Classification for more than two populations 158
 - 6.3.3 Evaluation of classification function 165
- 6.4 Problems in practicing discriminant analysis 169
 - 6.4.1 Variables selection 169
 - 6.4.2 Using qualitative variables 172
 - 6.4.3 Non-normal data 174

7 PRINCIPAL COMPONENTS 177

- 7.1 Introduction 178
- 7.2 Population principal components 180
 - 7.2.1 Definition of principal components 180
 - 7.2.2 Properties of principal components 183
 - 7.2.3 Interpretation of principal components 187
- 7.3 Sample principal components 191
 - 7.3.1 Estimation of principal components 191
 - 7.3.2 Testing significance of principal components 193
 - 7.3.3 Determination of the number of principal components 190
- 7.4 Application of principal components 199
 - 7.4.1 Application in regression analysis 199
 - 7.4.2 Identification of outliers 200
 - 7.4.3 Robust principal components analysis 202

8 FACTOR ANALYSIS 203

- 8.1 Introduction 203
- 8.2 Factor analysis model and estimation 205
 - 8.2.1 Factor analysis model 205
 - 8.2.2 Estimation methods 211
 - 8.2.3 Determination of the number of factors 220
- 8.3 Factor rotation 223
 - 8.3.1 Simple structure 225
 - 8.3.2 Orthogonal rotation methods 226
 - 8.3.3 Non-orthogonal rotation methods 230
 - 8.3.4 Interpretation of factors 230
- 8.4 Factor scores 235
 - 8.4.1 Direct calculation of factor scores 235
 - 8.4.2 Estimation of factor scores using regression method 236
 - 8.4.3 Estimation of factor scores using weighted least squares 238

9 CLUSTER ANALYSIS 241

- 9.1 Introduction 241
- 9.2 Measure of similarity and dissimilarity 245
 - 9.2.1 Measure of similarity and dissimilarity - quantitative variables 245
 - 9.2.2 Measure of similarity and dissimilarity - qualitative variables 251
- 9.3 Measure of similarity and dissimilarity between groups 256
- 9.4 Cluster methods
 - 9.4.1 Hierarchical methods of clustering 260
 - 9.4.2 Dendrogram and derived proximities 265
 - 9.4.3 Choosing the number of clusters 268
 - 9.4.4 Statistical evaluation of cluster quality 271
 - 9.4.5 Non-hierarchical methods of clustering 274
- 9.5 Using cluster analysis with other multivariate methods 276

REFERENCES 279