Multivariate Analysis for the Behavioral Sciences

Second Edition

Kimmo Vehkalahti
Brian S. Everitt

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)

Contents

Preface xiii

Preface to *Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences* xv

Authors xix

Acknowledgments xxi

1 Data, Measurement, and Models 1
- 1.1 Introduction . . . 1
- 1.2 Types of Study . . . 2
  - 1.2.1 Surveys . . . 3
  - 1.2.2 Experiments . . . 4
  - 1.2.3 Observational Studies . . . 5
  - 1.2.4 Quasi-Experiments . . . 6
- 1.3 Types of Measurement . . . 7
  - 1.3.1 Nominal or Categorical Measurements . . . 7
  - 1.3.2 Ordinal Scale Measurements . . . 8
  - 1.3.3 Interval Scales . . . 8
  - 1.3.4 Ratio Scales . . . 9
  - 1.3.5 Response and Explanatory Variables . . . 10
- 1.4 Missing Values . . . 10
- 1.5 The Role of Models in the Analysis of Data . . . 11
- 1.6 Determining Sample Size . . . 14
- 1.7 Significance Tests, p-Values, and Confidence Intervals . . . 16
- 1.8 Summary . . . 19
- 1.9 Exercises . . . 20

2 Looking at Data 23
- 2.1 Introduction . . . 23
- 2.2 Simple Graphics—Pie Charts, Bar Charts, Histograms, and Boxplots . . . 24
  - 2.2.1 Categorical Data . . . 24
  - 2.2.2 Interval/Quasi-Interval Data . . . 32
- 2.3 The Scatterplot and beyond . . . 37
  - 2.3.1 The Bubbleplot . . . 40
  - 2.3.2 The Bivariate Boxplot . . . 42
- 2.4 Scatterplot Matrices . . . 45

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)

2.5 Conditioning Plots and Trellis Graphics 48
2.6 Graphical Deception 55
2.7 Summary 59
2.8 Exercises 60

3 Simple Linear and Locally Weighted Regression 63
3.1 Introduction 63
3.2 Simple Linear Regression 64
3.2.1 Fitting the Simple Linear Regression Model to the Pulse Rates and Heights Data 66
3.2.2 An Example from Kinesiology 67
3.3 Regression Diagnostics 69
3.4 Locally Weighted Regression 73
3.4.1 Scatterplot Smoothers 75
3.5 Summary 81
3.6 Exercises 82

4 Multiple Linear Regression 83
4.1 Introduction 83
4.2 An Example of Multiple Linear Regression 85
4.3 Choosing the Most Parsimonious Model When Applying Multiple Linear Regression 90
4.3.1 Automatic Model Selection 95
4.3.2 Example of Application of the Backward Elimination 96
4.4 Regression Diagnostics 98
4.5 Multiple Linear Regression and Analysis of Variance 102
4.5.1 Analyzing the Fecundity of Fruit Flies by Regression 102
4.5.2 Multiple Linear Regression for Experimental Designs 104
4.5.3 Analyzing a Balanced Design 105
4.5.4 Analyzing an Unbalanced Design 106
4.6 Summary 109
4.7 Exercises 110

5 Generalized Linear Models 113
5.1 Introduction 113
5.2 Binary Response Variables 115
5.3 Response Variables That Are Counts 117
5.3.1 Overdispersion and Quasi-Likelihood 119
5.4 Summary 120
5.5 Exercises 121

6 Applying Logistic Regression 123
6.1 Introduction 123
6.2 Odds and Odds Ratios 123
6.3 Applying Logistic Regression to the GHQ Data 125
6.4 Selecting the Most Parsimonious Logistic Regression Model 130

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)

6.5 Driving and Back Pain: A Matched Case–Control Study . . 134
6.6 Summary 136
6.7 Exercises 136

7 Survival Analysis 139
7.1 Introduction 139
7.2 The Survival Function 140
7.2.1 Age at First Sexual Intercourse for Women 142
7.3 The Hazard Function 144
7.4 Cox's Proportional Hazards Model 146
7.4.1 Retention of Heroin Addicts in Methadone Treatment 149
7.5 Summary 152
7.6 Exercises 153

8 Analysis of Longitudinal Data I: Graphical Displays and Summary Measure Approach 155
8.1 Introduction 155
8.2 Graphical Displays of Longitudinal Data 157
8.3 Summary Measure Analysis of Longitudinal Data 159
8.3.1 Choosing Summary Measures 160
8.3.2 Applying the Summary Measure Approach 162
8.3.3 Incorporating Pre-Treatment Outcome Values into the Summary Measure Approach 164
8.3.4 Dealing with Missing Values When Using the Summary Measure Approach 164
8.4 Summary 166
8.5 Exercises 167

9 Analysis of Longitudinal Data II: Linear Mixed Effects Models for Normal Response Variables 169
9.1 Introduction 169
9.2 Linear Mixed Effects Models for Repeated Measures Data . 170
9.3 How Do Rats Grow? 174
9.3.1 Fitting the Independence Model to the Rat Data . . 174
9.3.2 Fitting Linear Mixed Models to the Rat Data 176
9.4 Computerized Delivery of Cognitive Behavioral Therapy—Beat the Blues 181
9.5 Summary 186
9.6 Exercises 187

10 Analysis of Longitudinal Data III: Non-Normal Responses 189
10.1 Introduction 189
10.2 Marginal Models and Conditional Models 190
10.2.1 Marginal Models 190
10.2.2 Conditional Models 194

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)

10.3 Using Generalized Estimating Equations to Fit Marginal Models . . . 196
10.3.1 Beat the Blues Revisited . . . 196
10.3.2 Respiratory Illness . . . 197
10.3.3 Epilepsy . . . 201
10.4 Using Generalized Linear Mixed Effects Models to Fit Conditional Models . . . 203
10.4.1 Respiratory Illness . . . 203
10.4.2 Epilepsy . . . 204
10.5 Summary . . . 206
10.6 Exercises . . . 207

11 Missing Values **209**
11.1 Introduction . . . 209
11.2 Missing Data Mechanisms . . . 210
11.3 Dealing with Missing Values . . . 212
11.4 Imputing Missing Values . . . 213
11.5 Analyzing Multiply Imputed Data . . . 215
11.6 Example of the Application of Multiple Imputation . . . 216
11.6.1 Complete-Case Analysis . . . 217
11.6.2 Mean Imputation . . . 218
11.6.3 Multiple Imputation . . . 219
11.7 Beat the Blues Revisited (Again) . . . 219
11.8 Summary . . . 222
11.9 Exercises . . . 222

12 Multivariate Data and Multivariate Analysis **225**
12.1 Introduction . . . 225
12.2 The Initial Analysis of Multivariate Data . . . 226
12.2.1 Summary Statistics for Multivariate Data . . . 226
12.2.2 Graphical Descriptions of the Body Measurement Data . . . 229
12.3 The Multivariate Normal Probability Density Function . . . 230
12.3.1 Assessing Multivariate Data for Normality . . . 233
12.4 Summary . . . 237
12.5 Exercises . . . 237

13 Principal Components Analysis **239**
13.1 Introduction . . . 239
13.2 Principal Components Analysis (PCA) . . . 239
13.3 Finding the Sample Principal Components . . . 241
13.4 Should Principal Components be Extracted from the Covariance or the Correlation Matrix? . . . 244
13.5 Principal Components of Bivariate Data with Correlation Coefficient r . . . 246

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)

13.6 Rescaling the Principal Components . . . 248
13.7 How the Principal Components Predict the Observed Covariance Matrix . . . 248
13.8 Choosing the Number of Components . . . 249
13.9 Calculating Principal Component Scores . . . 251
13.10 Some Examples of the Application of PCA . . . 252
13.10.1 Head Size of Brothers . . . 252
13.10.2 Crime Rates in the United States . . . 255
13.10.3 Drug Usage by American College Students . . . 260
13.11 Using PCA to Select a Subset of the Variables . . . 264
13.12 Summary . . . 265
13.13 Exercises . . . 266

14 Multidimensional Scaling and Correspondence Analysis 267
14.1 Introduction . . . 267
14.2 Multidimensional Scaling . . . 269
14.2.1 Classical Multidimensional Scaling . . . 270
14.2.2 Connection to Principal Components . . . 273
14.2.3 Road Distances in Finland . . . 274
14.2.4 Mapping Composers of Classical Music . . . 278
14.2.5 Nonmetric Multidimensional Scaling . . . 280
14.2.6 Re-mapping Composers of Classical Music . . . 281
14.3 Correspondence Analysis . . . 284
14.3.1 Simple Example of the Application of Correspondence Analysis . . . 286
14.3.2 Connections of Work Activities and Job Advantages 288
14.4 Summary . . . 291
14.5 Exercises . . . 292

15 Exploratory Factor Analysis 295
15.1 Introduction . . . 295
15.2 The Factor Analysis Model . . . 296
15.3 Estimating the Parameters in the Factor Analysis Model . . 299
15.4 Determining the Number of Factors . . . 301
15.5 Fitting the Factor Analysis Model: An Example . . . 302
15.6 Rotation of Factors . . . 304
15.6.1 A Simple Example of Graphical Rotation . . . 306
15.6.2 Numerical Rotation Methods . . . 309
15.6.3 A Simple Example of Numerical Rotation . . . 311
15.7 Estimating Factor Scores . . . 311
15.7.1 Analyzing the Crime Rates by Factor Analysis . . . 312
15.8 Exploratory Factor Analysis and Principal Component Analysis Compared . . . 315
15.9 Summary . . . 316
15.10 Exercises . . . 317

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)

16 Confirmatory Factor Analysis and Structural Equation Models **319**
16.1 Introduction . . . 319
16.2 Estimation, Identification, and Assessing the Fit for Confirmatory Factor Analysis and Structural Equation Models . . 320
16.2.1 Estimation . . . 320
16.2.2 Identification . . . 321
16.2.3 Assessing the Fit . . . 322
16.3 Examples of Confirmatory Factor Analysis . . . 324
16.3.1 Ability and Aspiration . . . 325
16.3.2 Drug Usage among Students . . . 327
16.4 Eight Factors of Systems Intelligence . . . 331
16.4.1 Testing the Factorial Validity of the SI Inventory . . 333
16.5 Structural Equation Models . . . 335
16.5.1 Example of a Structural Equation Model . . . 335
16.6 Summary . . . 337
16.7 Exercises . . . 337

17 Cluster Analysis **341**
17.1 Introduction . . . 341
17.2 Cluster Analysis . . . 343
17.3 Agglomerative Hierarchical Clustering . . . 344
17.3.1 Clustering Individuals Based on Body Measurements 347
17.3.2 Clustering Countries on the Basis of Life Expectancy 348
17.4 k-Means Clustering . . . 352
17.4.1 Clustering Crime Rates . . . 355
17.5 Model-Based Clustering . . . 356
17.5.1 Clustering European Countries . . . 359
17.6 Summary . . . 362
17.7 Exercises . . . 363

18 Grouped Multivariate Data **365**
18.1 Introduction . . . 365
18.2 Two-Group Multivariate Data . . . 366
18.2.1 Hotelling's T^2 Test . . . 366
18.2.2 Fisher's Linear Discriminant Function . . . 369
18.3 More Than Two Groups . . . 374
18.3.1 Multivariate Analysis of Variance (MANOVA) . . . 374
18.3.2 Classification Functions . . . 378
18.4 Summary . . . 382
18.5 Exercises . . . 382

References **385**

Index **401**

Libros de Estadística-Ciencia de Datos|Statistics-Data Science Books (PDF)