

Contents

Preface

Chapter 1 Introduction	1
1.1 Goal of Statistics	1
1.2 Univariate Analysis	3
1.3 Multivariate Analysis	7
1.4 Multivariate Normal Distribution	16
1.5 Unsupervised Learning and Supervised Learning	21
1.6 Data Analysis Strategies and Statistical Thinking	23
1.7 Outline	26
Exercises 1	27
Chapter 2 Principal Components Analysis	29
2.1 The Basic Idea	29
2.2 The Principal Components	30
2.3 Choose Number of Principal Components	34
2.4 Considerations in Data Analysis	35
2.5 Examples in R	37
Exercises 2	43
Chapter 3 Factor Analysis	45
3.1 The Basic Idea	45
3.2 The Factor Analysis Model	46
3.3 Methods for Estimation	47
3.4 Examples in R	50
Exercises 3	54
Chapter 4 Discriminant Analysis and Cluster Analysis	56
4.1 Introduction	56
4.2 Discriminant Analysis	57
4.3 Cluster Analysis	61
4.4 Examples in R	64
Exercises 4	69

Chapter 5 Inference for a Multivariate Normal Population	71
5.1 Introduction	71
5.2 Inference for Multivariate Means	72
5.3 Inference for Covariance Matrices	75
5.4 Large Sample Inferences about a Population Mean Vector	76
5.5 Examples in R	76
Exercises 5	79
Chapter 6 Discrete or Categorical Multivariate Data	80
6.1 Discrete or Categorical Data	80
6.2 The Multinomial Distribution	81
6.3 Contingency Tables	83
6.4 Associations Between Discrete or Categorical Variables	85
6.5 Logit Models for Multinomial Variables	87
6.6 Loglinear Models for Contingency Tables	89
6.7 Example in R	91
Exercises 6	95
Chapter 7 Copula Models	97
7.1 Introduction	97
7.2 Copula Models	99
7.3 Measures of Dependence	102
7.4 Applications in Actuary and Finance	103
7.5 Applications in Longitudinal and Survival Data*	106
7.6 Example in R	107
Exercises 7	110
Chapter 8 Linear and Nonlinear Regression Models	111
8.1 Introduction	111
8.2 Linear Regression Models	112
8.3 Model Selection	114
8.4 Model Diagnostics	116
8.5 Data Analysis Examples with R	117
8.6 Nonlinear Regression Models	122
8.7 More on Model Selection	125
Exercises 8	129
Chapter 9 Generalized Linear Models	131
9.1 Introduction	131

- 9.2 The Exponential Family ····· 132
- 9.3 The General Form of a GLM ····· 133
- 9.4 Inference for GLM ····· 135
- 9.5 Model Selection and Model Diagnostics ····· 137
- 9.6 Logistic Regression Models ····· 140
- 9.7 Poisson Regression Models ····· 146
- Exercises 9 ····· 149
- Chapter 10 Multivariate Regression and MANOVA Models ····· 152**
- 10.1 Introduction ····· 152
- 10.2 Multivariate Regression Models ····· 153
- 10.3 MANOVA Models ····· 156
- 10.4 Examples in R ····· 157
- Exercises 10 ····· 162
- Chapter 11 Longitudinal Data, Panel Data, and Repeated
Measurements ····· 164**
- 11.1 Introduction ····· 164
- 11.2 Methods for Longitudinal Data Analysis ····· 165
- 11.3 Linear Mixed Effects Models ····· 167
- 11.4 GEE Models ····· 171
- Exercises 11 ····· 174
- Chapter 12 Methods for Missing Data ····· 175**
- 12.1 Missing Data Mechanisms ····· 175
- 12.2 Methods for Missing Data ····· 178
- 12.3 Multiple Imputation Methods ····· 181
- 12.4 Multiple Imputation by Chained Equations ····· 183
- 12.5 The EM Algorithm ····· 184
- 12.6 Example in R ····· 187
- Exercises 12 ····· 192
- Chapter 13 Robust Multivariate Analysis ····· 193**
- 13.1 The Need for Robust Methods ····· 193
- 13.2 General Robust Methods ····· 195
- 13.3 Robust Estimates of the Mean and Standard Deviation ····· 199
- 13.4 Robust Estimates of the Covariance Matrix ····· 201
- 13.5 Robust PCA and Regressions ····· 203
- 13.6 Examples in R ····· 205

Exercises 13210

Chapter 14 Selected Topics 211

14.1 Likelihood Methods211

14.2 Bootstrap Methods 214

14.3 MCMC Methods and the Gibbs Sampler215

14.4 Survival Analysis 217

14.5 Data Science, Big Data, and Data Mining220

References224