# Contents

Preface ..... xiii
Acknowledgments ..... xvii

**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 ..... 9
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 ..... 19

**2. Looking at Data** ..... 21
2.1 Introduction ..... 21
2.2 Simple Graphics—Pie Charts, Bar Charts, Histograms, and Boxplots ..... 22
2.2.1 Categorical Data ..... 22
2.2.2 Interval/Quasi-Interval Data ..... 30
2.3 The Scatterplot and Beyond ..... 35
2.3.1 The Bubbleplot ..... 38
2.3.2 The Bivariate Boxplot ..... 40
2.4 Scatterplot Matrices ..... 44
2.5 Conditioning Plots and Trellis Graphics ..... 45
2.6 Graphical Deception ..... 52
2.7 Summary ..... 58
2.8 Exercises ..... 58

**3. Simple Linear and Locally Weighted Regression** ..... 61
3.1 Introduction ..... 61
3.2 Simple Linear Regression ..... 62
3.2.1 Fitting the Simple Linear Regression Model to the Pulse Rates and Heights Data ..... 64
3.2.2 An Example from Kinesiology ..... 65
3.3 Regression Diagnostics ..... 68
3.4 Locally Weighted Regression ..... 72
3.4.1 Scatterplot Smoothers ..... 73
3.5 Summary ..... 79
3.6 Exercises ..... 80

**4. Multiple Linear Regression** ..... 81
4.1 Introduction ..... 81
4.2 An Example of Multiple Linear Regression ..... 84
4.3 Choosing the Most Parsimonious Model When Applying Multiple Linear Regression ..... 89
4.4 Regression Diagnostics ..... 96
4.5 Summary ..... 100
4.6 Exercises ..... 100

**5. The Equivalence of Analysis of Variance and Multiple Linear Regression, and an Introduction to the Generalized Linear Model** ..... 103
5.1 Introduction ..... 103
5.2 The Equivalence of Multiple Regression and ANOVA ..... 103
5.3 The Generalized Linear Model ..... 110
5.4 Summary ..... 112
5.5 Exercises ..... 113

**6. Logistic Regression** ..... 115
6.1 Introduction ..... 115
6.2 Odds and Odds Ratios ..... 115
6.3 Logistic Regression ..... 117
6.4 Applying Logistic Regression to the GHQ Data ..... 120
6.5 Selecting the Most Parsimonious Logistic Regression Model ..... 124
6.6 Summary ..... 128
6.7 Exercises ..... 128

**7. Survival Analysis** ..... 131
7.1 Introduction ..... 131
7.2 The Survival Function ..... 132
7.3 The Hazard Function ..... 136
7.4 Cox's Proportional Hazards Model ..... 138

7.5 Summary ..... 143
7.6 Exercises ..... 144

**8. Linear Mixed Models for Longitudinal Data** ..... 145
8.1 Introduction ..... 145
8.2 Linear Mixed Effects Models for Longitudinal Data ..... 146
8.3 How Do Rats Grow? ..... 150
8.3.1 Fitting the Independence Model to the Rat Data ..... 151
8.3.2 Fitting Linear Mixed Models to the Rat Data ..... 153
8.4 Computerized Delivery of Cognitive Behavioral Therapy—Beat the Blues ..... 157
8.5 The Problem of Dropouts in Longitudinal Studies ..... 162
8.6 Summary ..... 165
8.7 Exercises ..... 166

**9. Multivariate Data and Multivariate Analysis** ..... 169
9.1 Introduction ..... 169
9.2 The Initial Analysis of Multivariate Data ..... 170
9.2.1 Summary Statistics for Multivariate Data ..... 170
9.2.2 Graphical Descriptions of the Body Measurement Data ..... 173
9.3 The Multivariate Normal Probability Density Function ..... 174
9.4 Summary ..... 180
9.5 Exercises ..... 181

**10. Principal Components Analysis** ..... 183
10.1 Introduction ..... 183
10.2 Principal Components Analysis (PCA) ..... 183
10.3 Finding the Sample Principal Components ..... 185
10.4 Should Principal Components Be Extracted from the Covariance or the Correlation Matrix? ..... 188
10.5 Principal Components of Bivariate Data with Correlation Coefficient $r$ ..... 190
10.6 Rescaling the Principal Components ..... 192
10.7 How the Principal Components Predict the Observed Covariance Matrix ..... 193
10.8 Choosing the Number of Components ..... 193
10.9 Calculating Principal Component Scores ..... 195
10.10 Some Examples of the Application of PCA ..... 196
10.10.1 Head Size of Brothers ..... 196
10.10.2 Crime Rates in the United States ..... 200
10.10.3 Drug Usage by American College Students ..... 205

10.11 Using PCA to Select a Subset of the Variables ..... 208
10.12 Summary ..... 209
10.13 Exercises ..... 210

**11. Factor Analysis** ..... 211
11.1 Introduction ..... 211
11.2 The Factor Analysis Model ..... 212
11.3 Estimating the Parameters in the Factor Analysis Model ..... 215
11.4 Estimating the Numbers of Factors ..... 217
11.5 Fitting the Factor Analysis Model: An Example ..... 218
11.6 Rotation of Factors ..... 220
11.6.1 A Simple Example of Graphical Rotation ..... 222
11.6.2 Numerical Rotation Methods ..... 223
11.6.3 Rotating the Crime Rate Factors ..... 226
11.7 Estimating Factor Scores ..... 227
11.8 Exploratory Factor Analysis and Principal Component Analysis Compared ..... 228
11.9 Confirmatory Factor Analysis ..... 229
11.9.1 Ability and Aspiration ..... 230
11.9.2 A Confirmatory Factor Analysis Model for Drug Usage ..... 233
11.10 Summary ..... 235
11.11 Exercises ..... 236

**12. Cluster Analysis** ..... 239
12.1 Introduction ..... 239
12.2 Cluster Analysis ..... 241
12.3 Agglomerative Hierarchical Clustering ..... 241
12.3.1 Clustering Individuals Based on Body Measurements ..... 243
12.3.2 Clustering Countries on the Basis of Life Expectancy Data ..... 246
12.4 *k*-Means Clustering ..... 250
12.5 Model-Based Clustering ..... 253
12.6 Summary ..... 258
12.7 Exercises ..... 259

**13. Grouped Multivariate Data** ..... 261
13.1 Introduction ..... 261
13.2 Two-Group Multivariate Data ..... 262
13.2.1 Hotelling's $T^2$ Test ..... 262
13.2.2 Fisher's Linear Discriminant Function ..... 265
13.3 More Than Two Groups ..... 270
13.3.1 Multivariate Analysis of Variance (MANOVA) ..... 270
13.3.2 Classification Functions ..... 273

13.4 Summary ..... 277
13.5 Exercises ..... 277

**References** ..... 279
**Appendix: Solutions to Selected Exercises** ..... 285
**Index** ..... 299