

Table of Contents

About the Authors..... ix

About the Technical Reviewer xi

Acknowledgments xiii

Introductionxv

Chapter 1: Univariate Data Visualization 1

 1.1 Distribution 2

 Visualizing the Observed Distribution 2

 Stacked Dot Plots and Histograms..... 2

 Density Plots 6

 Comparing the Observed Distribution with Expected Distributions 9

 Q-Q Plots 9

 Density Plots 14

 Fitting More Distributions 15

 1.2 Anomalous Values 21

 1.3 Summary 30

Chapter 2: Multivariate Data Visualization 33

 2.1 Distribution 34

 2.2 Anomalous Values 40

 2.3 Relations Between Variables 44

 Assessing Homogeneity of Variance 53

 2.4 Summary 59

TABLE OF CONTENTS

Chapter 3: GLM 1 61

3.1 Conceptual Background 62

3.2 Categorical Predictors and Dummy Coding..... 65

 Two-Level Categorical Predictors 65

 Three- or More Level Categorical Predictors 66

3.3 Interactions and Moderated Effects 68

3.4 Formula Interface 70

3.5 Analysis of Variance 72

 Conceptual Background..... 72

 ANOVA in R..... 76

3.6 Linear Regression 79

 Conceptual Background..... 80

 Linear Regression in R..... 82

 High-Performance Linear Regression..... 99

3.7 Controlling for Confounds..... 102

3.8 Case Study: Multiple Linear Regression with Interactions 113

3.9 Summary..... 121

Chapter 4: GLM 2 123

4.1 Conceptual Background 124

 Logistic Regression..... 124

 Count Regression..... 126

4.2 R Examples..... 128

 Binary Logistic Regression 129

 Ordered Logistic Regression 136

 Multinomial Logistic Regression 140

 Poisson and Negative Binomial Regression..... 145

4.3 Case Study: Multinomial Logistic Regression 153

4.4 Summary..... 162

Chapter 5: GAMs	165
5.1 Conceptual Overview	166
Smoothing Splines	167
5.2 GAMs in R	173
Gaussian Outcomes	173
Binary Outcomes	202
Unordered Outcomes	209
Count Outcomes	214
5.3 Summary	223
Chapter 6: ML: Introduction	225
6.1 Training and Validation Data	226
6.2 Resampling and Cross-Validation	233
6.3 Bootstrapping	237
6.4 Parallel Processing and Random Numbers	239
foreach	245
6.5 Summary	247
Chapter 7: ML: Unsupervised	251
7.1 Data Background and Exploratory Analysis	252
7.2 kmeans	263
7.3 Hierarchical Clusters	276
7.4 Principal Component Analysis	288
7.5 Non-linear Cluster Analysis	300
7.6 Summary	302
Chapter 8: ML: Supervised	305
8.1 Data Preparation	306
One Hot Encoding	308
Scale and Center	311
Transformations	312
Train vs. Validation Data	318
Principal Component Analysis	319

TABLE OF CONTENTS

8.2 Supervised Learning Models..... 324

 Support Vector Machines 325

 Classification and Regression Trees 335

 Random Forests 341

 Stochastic Gradient Boosting..... 348

 Multilayer Perceptron..... 358

8.3 Summary..... 380

Chapter 9: Missing Data 383

 9.1 Conceptual Background 384

 Multiple Imputation 386

 9.2 R Examples..... 393

 Multiple Imputation with Regression 398

 Multiple Imputation with Parallel Processing 410

 Multiple Imputation Using Random Forests 412

 9.3 Case Study: Multiple Imputation with RFs..... 419

 9.4 Summary..... 432

Chapter 10: GLMMs: Introduction..... 435

 10.1 Multilevel Data 436

 Reshaping Data..... 438

 Daily Dataset..... 440

 10.2 Descriptive Statistics..... 444

 Basic Descriptives 446

 Intraclass Correlation Coefficient (ICC) 454

 10.3 Exploration and Assumptions 457

 Distribution and Outliers 457

 Time Trends 463

 Autocorrelation 465

 Assumptions 470

 10.4 Summary..... 476

Chapter 11: GLMMs: Linear 479

11.1 Theory 480

 Generalized Linear Mixed Models..... 480

 Mixed Effects vs. Multilevel Model Terminology 485

 Statistical Inference..... 485

 Effect Sizes..... 487

 Random Intercept Model 489

 Visualizing Random Effects 489

 Interpreting Random Intercept Models 494

 Random Intercept and Slope Model..... 502

 Intercepts and Slopes as Outcomes 508

11.2 R Examples..... 514

 Linear Mixed Model with Random Intercept..... 514

 Linear Mixed Model with Random Intercept and Slope 528

11.3 Summary..... 549

Chapter 12: GLMMs: Advanced..... 553

12.1 Conceptual Background 554

12.2 Logistic GLMM..... 554

 Random Intercept 554

 Random Intercept and Slope 560

12.3 Poisson and Negative Binomial GLMMs 565

 Random Intercept 565

 Random Intercept and Slope 576

12.4 Summary..... 586