Contents

Preface **vii**

1 Introduction **1**
1.1 The Nature of Categorical Data 1
1.2 Organization of This Book 3

2 Gaussian-Based Data Analysis **7**
2.1 The Normal (Gaussian) Random Variable 7
2.1.1 The Gaussian Density Function 7
2.1.2 Large-Sample Inference for the Gaussian Random Variable 8
2.1.3 Exact Inference for the Gaussian Random Variable . 11
2.2 Linear Regression and Least Squares 12
2.2.1 The Linear Regression Model 12
2.2.2 Least Squares Estimation 14
2.2.3 Interpreting Regression Coefficients 15
2.2.4 Assessing the Strength of a Regression Relationship 17
2.3 Inference for the Least Squares Regression Model 18
2.3.1 Hypothesis Tests and Confidence Intervals for $\boldsymbol{\beta}$ . . 18
2.3.2 Interval Estimation for Predicted and Fitted Values 19
2.4 Checking Assumptions 20
2.5 An Example 21
2.6 Background Material 25
2.7 Exercises 26

3 Gaussian-Based Model Building **29**
3.1 Linear Contrasts and Hypothesis Tests 29
3.2 Categorical Predictors 31
3.2.1 One Categorical Predictor with Two Levels 31
3.2.2 One Categorical Predictor with More Than Two Levels 32
3.2.3 More Than One Categorical Predictor 33
3.3 Regression Diagnostics 36
3.3.1 Identifying Outliers 36
3.3.2 Identifying Leverage Points 37
3.3.3 Identifying Influential Points 38
3.4 Model Selection 42
3.4.1 Choosing a Set of Candidate Models 44
3.4.2 Choosing the "Best" Model 45
3.4.3 Model Selection Uncertainty 46
3.5 Heteroscedasticity and Weighted Least Squares 49
3.6 Background Material 51
3.7 Exercises 52

4 Categorical Data and Goodness-of-Fit **55**
4.1 The Binomial Random Variable 55
4.1.1 Large-Sample Inference 56
4.1.2 Sample Size and Power Calculations 59
4.1.3 Inference from a Sample of Binomial Random Variables 61
4.1.4 Exact Inference 61
4.2 The Multinomial Random Variable 68
4.2.1 Large-Sample Inference 68
4.3 The Poisson Random Variable 69
4.3.1 Large-Sample Inference 71
4.3.2 Exact Inference 73
4.3.3 The Connection Between the Poisson and the Multinomial 74
4.4 Testing Goodness-of-Fit 75
4.4.1 Chi-Squared Goodness-of-Fit Tests 75
4.4.2 Partitioning Pearson's X^2 Statistic 80
4.4.3 Exact Inference 82
4.5 Overdispersion and Lack of Fit 84
4.5.1 The Zero-Inflated Poisson Model 84
4.5.2 The Negative Binomial Model 88
4.5.3 Overdispersed Binomial Data and the Beta-Binomial Model 93
4.6 Underdispersion 98
4.7 Robust Estimation Using Hellinger Distance 100
4.8 Background Material 102

4.9 Exercises . . . 103

5 Regression Models for Count Data 125
5.1 The Generalized Linear Model . . . 125
5.1.1 The Form of the Generalized Linear Model . . . 125
5.1.2 Estimation in the Generalized Linear Model . . . 127
5.1.3 Hypothesis Tests and Confidence Intervals for $\boldsymbol{\beta}$ . . . 128
5.1.4 The Deviance and Lack of Fit . . . 129
5.1.5 Model Selection . . . 130
5.1.6 Model Checking and Regression Diagnostics . . . 132
5.2 Poisson Regression . . . 133
5.3 Overdispersion . . . 147
5.3.1 The Robust Sandwich Covariance Estimator . . . 149
5.3.2 Quasi-Likelihood Estimation . . . 149
5.4 Non-Poisson Parametric Regression Models . . . 154
5.4.1 Negative Binomial Regression . . . 155
5.4.2 Zero-Inflated Count Regression . . . 162
5.4.3 Zero-Truncated Poisson Regression . . . 168
5.5 Nonparametric Count Regression . . . 171
5.5.1 Local Likelihood Estimation Based on One Predictor 171
5.5.2 Smoothing Multinomial Data . . . 175
5.5.3 Regression Smoothing with More Than One Predictor . . . 178
5.6 Background Material . . . 181
5.7 Exercises . . . 182

6 Analyzing Two-Way Tables 197
6.1 Two-by-Two Tables . . . 197
6.1.1 Two-Sample Tests and Comparisons of Proportions . 197
6.1.2 Two-by-Two Tables and Tests of Independence . . . 199
6.1.3 The Odds Ratio and the Relative Risk . . . 203
6.2 Loglinear Models for Two-Way Tables . . . 208
6.2.1 2×2 Tables . . . 208
6.2.2 $I \times J$ Tables . . . 210
6.3 Conditional Analyses . . . 218
6.3.1 Two-by-Two Tables and Fisher's Exact Test . . . 218
6.3.2 $I \times J$ Tables . . . 221
6.4 Structural Zeroes and Quasi-Independence . . . 225
6.5 Outlier Identification and Robust Estimation . . . 228
6.6 Background Material . . . 234
6.7 Exercises . . . 235

7 Tables with More Structure 247
7.1 Models for Tables with Ordered Categories . . . 247
7.1.1 Linear-by-Linear (Uniform) Association . . . 248

7.1.2 Row, Column, and Row + Column Effects Models . 252
7.1.3 A Log-Multiplicative Row and Column Effects Model 261
7.1.4 Bivariate Discrete Distributions 264
7.2 Square Tables 270
7.2.1 Quasi-Independence 270
7.2.2 Symmetry 270
7.2.3 Quasi-Symmetry 273
7.2.4 Tables with Ordered Categories 276
7.2.5 Matched Pairs Data 278
7.2.6 Rater Agreement Tables 283
7.3 Conditional Analyses 289
7.4 Background Material 291
7.5 Exercises 293

8 Multidimensional Contingency Tables 309
8.1 $2 \times 2 \times K$ Tables 309
8.1.1 Simpson's Paradox 310
8.1.2 Types of Independence 314
8.1.3 Tests of Conditional Association 316
8.2 Loglinear Models for Three-Dimensional Tables 318
8.2.1 Hierarchical Loglinear Models 318
8.2.2 Association Diagrams, Conditional Independence, and Collapsibility 323
8.2.3 Loglinear Model Fitting for Three-Dimensional Tables 324
8.3 Models for Tables with Ordered Categories 330
8.4 Higher-Dimensional Tables 337
8.5 Conditional Analyses 344
8.6 Background Material 347
8.7 Exercises 348

9 Regression Models for Binary Data 365
9.1 The Logistic Regression Model 365
9.1.1 Why Logistic Regression? 365
9.1.2 Inference for the Logistic Regression Model 369
9.1.3 Model Fit and Model Selection 372
9.2 Retrospective (Case-Control) Studies 380
9.3 Categorical Predictors 387
9.4 Other Link Functions 393
9.4.1 Dose Response Modeling and Logistic Regression . . 393
9.4.2 Probit Regression 394
9.4.3 Complementary Log–Log and Log–Log Regression . 396
9.5 Overdispersion 398
9.5.1 Nonparametric Approaches 398

9.5.2 Beta-Binomial Regression 399
9.6 Smoothing Binomial Data 403
9.7 Conditional Analysis 407
9.8 Background Material 411
9.9 Exercises 412

10 Regression Models for Multiple Category Response Data **427**
10.1 Nominal Response Variable 427
10.1.1 Multinomial Logistic Regression 427
10.1.2 Independence of Irrelevant Alternatives 434
10.2 Ordinal Response Variable 435
10.2.1 The Proportional Odds Model 436
10.2.2 Other Link Functions 438
10.2.3 Adjacent-Categories Logit Model 442
10.2.4 Continuation Ratio Models 443
10.3 Background Material 448
10.4 Exercises 449

A Some Basics of Matrix Algebra **455**

References **459**

Index **485**