Contents

Preface xiii

1 Introduction: Distributions and Inference for Categorical Data 1
   1.1 Categorical Response Data, 1
   1.2 Distributions for Categorical Data, 5
   1.3 Statistical Inference for Categorical Data, 8
   1.4 Statistical Inference for Binomial Parameters, 13
   1.5 Statistical Inference for Multinomial Parameters, 17
   1.6 Bayesian Inference for Binomial and Multinomial Parameters, 22
   Notes, 27
   Exercises, 28

2 Describing Contingency Tables 37
   2.1 Probability Structure for Contingency Tables, 37
   2.2 Comparing Two Proportions, 43
   2.3 Conditional Association in Stratified $2 \times 2$ Tables, 47
   2.4 Measuring Association in $I \times J$ Tables, 54
   Notes, 60
   Exercises, 60

3 Inference for Two-Way Contingency Tables 69
   3.1 Confidence Intervals for Association Parameters, 69
   3.2 Testing Independence in Two-way Contingency Tables, 75
   3.3 Following-up Chi-Squared Tests, 80
   3.4 Two-Way Tables with Ordered Classifications, 86
   3.5 Small-Sample Inference for Contingency Tables, 90
   3.6 Bayesian Inference for Two-way Contingency Tables, 96
   3.7 Extensions for Multiway Tables and Nontabulated Responses, 100
   Notes, 101
   Exercises, 103

vii
# CONTENTS

## 4 Introduction to Generalized Linear Models 113

4.1 The Generalized Linear Model, 113  
4.2 Generalized Linear Models for Binary Data, 117  
4.3 Generalized Linear Models for Counts and Rates, 122  
4.4 Moments and Likelihood for Generalized Linear Models, 130  
4.5 Inference and Model Checking for Generalized Linear Models, 136  
4.6 Fitting Generalized Linear Models, 143  
4.7 Quasi-Likelihood and Generalized Linear Models, 149  
Notes, 152  
Exercises, 153

## 5 Logistic Regression 163

5.1 Interpreting Parameters in Logistic Regression, 163  
5.2 Inference for Logistic Regression, 169  
5.3 Logistic Models with Categorical Predictors, 175  
5.4 Multiple Logistic Regression, 182  
5.5 Fitting Logistic Regression Models, 192  
Notes, 195  
Exercises, 196

## 6 Building, Checking, and Applying Logistic Regression Models 207

6.1 Strategies in Model Selection, 207  
6.2 Logistic Regression Diagnostics, 215  
6.3 Summarizing the Predictive Power of a Model, 221  
6.4 Mantel–Haenszel and Related Methods for Multiple $2 \times 2$ Tables, 225  
6.5 Detecting and Dealing with Infinite Estimates, 233  
6.6 Sample Size and Power Considerations, 237  
Notes, 241  
Exercises, 243

## 7 Alternative Modeling of Binary Response Data 251

7.1 Probit and Complementary Log–log Models, 251  
7.2 Bayesian Inference for Binary Regression, 257  
7.3 Conditional Logistic Regression, 265  
7.4 Smoothing: Kernels, Penalized Likelihood, Generalized Additive Models, 270  
7.5 Issues in Analyzing High-Dimensional Categorical Data, 278  
Notes, 285  
Exercises, 287
CONTENTS

8 Models for Multinomial Responses

8.1 Nominal Responses: Baseline-Category Logit Models, 293
8.2 Ordinal Responses: Cumulative Logit Models, 301
8.3 Ordinal Responses: Alternative Models, 308
8.4 Testing Conditional Independence in $I \times J \times K$ Tables, 314
8.5 Discrete-Choice Models, 320
8.6 Bayesian Modeling of Multinomial Responses, 323
Notes, 326
Exercises, 329

9 Loglinear Models for Contingency Tables

9.1 Loglinear Models for Two-way Tables, 339
9.2 Loglinear Models for Independence and Interaction in Three-way Tables, 342
9.3 Inference for Loglinear Models, 348
9.4 Loglinear Models for Higher Dimensions, 350
9.5 Loglinear—Logistic Model Connection, 353
9.6 Loglinear Model Fitting: Likelihood Equations and Asymptotic Distributions, 356
9.7 Loglinear Model Fitting: Iterative Methods and Their Application, 364
Notes, 368
Exercises, 369

10 Building and Extending Loglinear Models

10.1 Conditional Independence Graphs and Collapsibility, 377
10.2 Model Selection and Comparison, 380
10.3 Residuals for Detecting Cell-Specific Lack of Fit, 385
10.4 Modeling Ordinal Associations, 386
10.5 Generalized Loglinear and Association Models, Correlation Models, and Correspondence Analysis, 393
10.6 Empty Cells and Sparseness in Modeling Contingency Tables, 398
10.7 Bayesian Loglinear Modeling, 401
Notes, 404
Exercises, 407

11 Models for Matched Pairs

11.1 Comparing Dependent Proportions, 414
11.2 Conditional Logistic Regression for Binary Matched Pairs, 418
11.3 Marginal Models for Square Contingency Tables, 424
11.4 Symmetry, Quasi-Symmetry, and Quasi-Independence, 426
CONTENTS

11.5 Measuring Agreement Between Observers, 432
11.6 Bradley–Terry Model for Paired Preferences, 436
11.7 Marginal Models and Quasi-Symmetry Models for Matched Sets, 439
Notes, 443
Exercises, 445

12 Clustered Categorical Data: Marginal and Transitional Models

12.1 Marginal Modeling: Maximum Likelihood Approach, 456
12.2 Marginal Modeling: Generalized Estimating Equations (GEEs) Approach, 462
12.3 Quasi-Likelihood and Its GEE Multivariate Extension: Details, 465
12.4 Transitional Models: Markov Chain and Time Series Models, 473
Notes, 478
Exercises, 479

13 Clustered Categorical Data: Random Effects Models

13.1 Random Effects Modeling of Clustered Categorical Data, 489
13.2 Binary Responses: Logistic-Normal Model, 494
13.3 Examples of Random Effects Models for Binary Data, 498
13.4 Random Effects Models for Multinomial Data, 511
13.5 Multilevel Modeling, 515
13.6 GLMM Fitting, Inference, and Prediction, 519
13.7 Bayesian Multivariate Categorical Modeling, 523
Notes, 525
Exercises, 527

14 Other Mixture Models for Discrete Data

14.1 Latent Class Models, 535
14.2 Nonparametric Random Effects Models, 542
14.3 Beta-Binomial Models, 548
14.4 Negative Binomial Regression, 552
14.5 Poisson Regression with Random Effects, 555
Notes, 557
Exercises, 558

15 Non-Model-Based Classification and Clustering

15.1 Classification: Linear Discriminant Analysis, 565
15.2 Classification: Tree-Structured Prediction, 570
15.3 Cluster Analysis for Categorical Data, 576
Notes, 581
Exercises, 582
## CONTENTS

### 16 Large- and Small-Sample Theory for Multinomial Models

16.1 Delta Method, 587  
16.2 Asymptotic Distributions of Estimators of Model Parameters and Cell Probabilities, 592  
16.3 Asymptotic Distributions of Residuals and Goodness-of-fit Statistics, 594  
16.4 Asymptotic Distributions for Logit/Loglinear Models, 599  
16.5 Small-Sample Significance Tests for Contingency Tables, 601  
16.6 Small-Sample Confidence Intervals for Categorical Data, 603  
16.7 Alternative Estimation Theory for Parametric Models, 610  
Notes, 615  
Exercises, 616

### 17 Historical Tour of Categorical Data Analysis

17.1 Pearson–Yule Association Controversy, 623  
17.2 R. A. Fisher’s Contributions, 625  
17.3 Logistic Regression, 627  
17.4 Multiway Contingency Tables and Loglinear Models, 629  
17.5 Bayesian Methods for Categorical Data, 633  
17.6 A Look Forward, and Backward, 634  

**Appendix A** Statistical Software for Categorical Data Analysis 637  
**Appendix B** Chi-Squared Distribution Values 641  
References 643  
Author Index 689  
Example Index 701  
Subject Index 705  

**Appendix C** Software Details for Text Examples (text website)  
(www.stat.ufl.edu/~aa/cda/cda.html)  
**Appendix D** Solutions to Selected Exercises (text website)  
(www.stat.ufl.edu/~aa/cda/cda.html)