AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS

Third Edition

Alan Agresti
University of Florida, Florida, United States

WILEY

CONTENTS

Preface ix

About the Companion Website xiii

1 Introduction 1

1.1 Categorical Response Data 1
1.2 Probability Distributions for Categorical Data 3
1.3 Statistical Inference for a Proportion 5
1.4 Statistical Inference for Discrete Data 10
1.5 Bayesian Inference for Proportions * 13
1.6 Using R Software for Statistical Inference about Proportions * 17
Exercises 21

2 Analyzing Contingency Tables 25

2.1 Probability Structure for Contingency Tables 26
2.2 Comparing Proportions in 2×2 Contingency Tables 29
2.3 The Odds Ratio 31
2.4 Chi-Squared Tests of Independence 36
2.5 Testing Independence for Ordinal Variables 42
2.6 Exact Frequentist and Bayesian Inference * 46
2.7 Association in Three-Way Tables 52
Exercises 56

3 Generalized Linear Models 65

3.1 Components of a Generalized Linear Model 66
3.2 Generalized Linear Models for Binary Data 68
3.3 Generalized Linear Models for Counts and Rates 72
3.4 Statistical Inference and Model Checking 76
3.5 Fitting Generalized Linear Models 82
Exercises 84

4 Logistic Regression 89

4.1 The Logistic Regression Model 89
4.2 Statistical Inference for Logistic Regression 94
4.3 Logistic Regression with Categorical Predictors 98
4.4 Multiple Logistic Regression 102
4.5 Summarizing Effects in Logistic Regression 107
4.6 Summarizing Predictive Power: Classification Tables, ROC Curves, and Multiple Correlation 110
Exercises 113

5 Building and Applying Logistic Regression Models 123

5.1 Strategies in Model Selection 123
5.2 Model Checking 130
5.3 Infinite Estimates in Logistic Regression 136
5.4 Bayesian Inference, Penalized Likelihood, and Conditional Likelihood for Logistic Regression * 140
5.5 Alternative Link Functions: Linear Probability and Probit Models * 145
5.6 Sample Size and Power for Logistic Regression * 150
Exercises 151

6 Multicategory Logit Models 159

6.1 Baseline-Category Logit Models for Nominal Responses 159
6.2 Cumulative Logit Models for Ordinal Responses 167
6.3 Cumulative Link Models: Model Checking and Extensions * 176
6.4 Paired-Category Logit Modeling of Ordinal Responses * 184
Exercises 187

7 Loglinear Models for Contingency Tables and Counts 193

7.1 Loglinear Models for Counts in Contingency Tables 194
7.2 Statistical Inference for Loglinear Models 200
7.3 The Loglinear – Logistic Model Connection 207

7.4 Independence Graphs and Collapsibility 210
7.5 Modeling Ordinal Associations in Contingency Tables 214
7.6 Loglinear Modeling of Count Response Variables * 217
Exercises 221

8 Models for Matched Pairs 227

8.1 Comparing Dependent Proportions for Binary Matched Pairs 228
8.2 Marginal Models and Subject-Specific Models for Matched Pairs 230
8.3 Comparing Proportions for Nominal Matched-Pairs Responses 235
8.4 Comparing Proportions for Ordinal Matched-Pairs Responses 239
8.5 Analyzing Rater Agreement * 243
8.6 Bradley–Terry Model for Paired Preferences * 247
Exercises 249

9 Marginal Modeling of Correlated, Clustered Responses 253

9.1 Marginal Models Versus Subject-Specific Models 254
9.2 Marginal Modeling: The Generalized Estimating Equations (GEE) Approach 255
9.3 Marginal Modeling for Clustered Multinomial Responses 260
9.4 Transitional Modeling, Given the Past 263
9.5 Dealing with Missing Data * 266
Exercises 268

10 Random Effects: Generalized Linear Mixed Models 273

10.1 Random Effects Modeling of Clustered Categorical Data 273
10.2 Examples: Random Effects Models for Binary Data 278
10.3 Extensions to Multinomial Responses and Multiple Random Effect Terms 284
10.4 Multilevel (Hierarchical) Models 288
10.5 Latent Class Models * 291
Exercises 295

11 Classification and Smoothing * 299

11.1 Classification: Linear Discriminant Analysis 300
11.2 Classification: Tree-Based Prediction 302
11.3 Cluster Analysis for Categorical Responses 306
11.4 Smoothing: Generalized Additive Models 310
11.5 Regularization for High-Dimensional Categorical Data (Large p) 313
Exercises 321

12 A Historical Tour of Categorical Data Analysis * 325

Appendix: Software for Categorical Data Analysis 331

A.1 R for Categorical Data Analysis 331
A.2 SAS for Categorical Data Analysis 332
A.3 Stata for Categorical Data Analysis 342
A.4 SPSS for Categorical Data Analysis 346

Brief Solutions to Odd-Numbered Exercises 349

Bibliography 363

Examples Index 365

Subject Index 369