Contents

PREFACE xiii

1 Introduction

1.1 Why Categorical Data Analysis? 1
 1.1.1 Defining Categorical Variables 2
 1.1.2 Dependent and Independent Variables 3
 1.1.3 Categorical Dependent Variables 4
 1.1.4 Types of Measurement 5
1.2 Two Philosophies of Categorical Data 7
 1.2.1 The Transformational Approach 8
 1.2.2 The Latent Variable Approach 9
1.3 An Historical Note 11
1.4 Approach of This Book 12
 1.4.1 Organization of the Book 13

2 Review of Linear Regression Models

2.1 Regression Models 15
2.1.1 Three Conceptualizations of Regression 16
2.1.2 Anatomy of Linear Regression 18
2.1.3 Basics of Statistical Inference 20
2.1.4 Tension between Accuracy and Parsimony 22
2.2 Linear Regression Models Revisited 24
2.2.1 Least Squares Estimation 24
2.2.2 Maximum Likelihood Estimation 25
2.2.3 Assumptions for Least Squares Regression 29
2.2.4 Comparisons of Conditional Means 30
2.2.5 Linear Models with Weaker Assumptions 32
2.3 Categorical and Continuous Dependent Variables 37
2.3.1 A Working Typology 38

3 Logit and Probit Models for Binary Data

3.1 Introduction to Binary Data 41
3.2 The Transformational Approach 43
3.2.1 The Linear Probability Model 43
3.2.2 The Logit Model 49
3.2.3 The Probit Model 52
3.2.4 An Application Using Grouped Data 53
3.3 Justification of Logit and Probit Models 55
3.3.1 The Latent Variable Approach 56
3.3.2 Extending the Latent Variable Approach 59
3.3.3 Estimation of Binary Response Models 61
3.3.4 Goodness-of-Fit and Model Selection 63
3.3.5 Hypothesis Testing and Statistical Inference 71
3.4 Interpreting Estimates 75
3.4.1 The Odds-Ratio 75
3.4.2 Marginal Effects 76
3.4.3 An Application Using Individual-Level Data 80
3.5 Alternative Probability Models 83
3.5.1 The Complementary Log–Log Model 83
3.5.2 Programming Binomial Response Models 85
3.6 Summary 85

4 Loglinear Models for Contingency Tables

4.1 Contingency Tables 87
4.1.1 Types of Contingency Tables 88
4.1.2 An Example and Notation 88
4.1.3 Independence and the Pearson χ^2 Statistic 90
4.2 Measures of Association 93
4.2.1 Homogeneous Proportions 93
4.2.2 Relative Risks 94
4.2.3 Odds-Ratios 95
4.2.4 The Invariance Property of Odds-Ratios 97
4.3 Estimation and Goodness-of-Fit 99
4.3.1 Simple Models and the Pearson χ^2 Statistic 100
4.3.2 Sampling Models and Maximum Likelihood Estimation 102
4.3.3 The Likelihood-Ratio Chi-Squared Statistic 104
4.3.4 Bayesian Information Criterion 106
4.4 Models for Two-Way Tables 107
4.4.1 The General Setup 107
4.4.2 Normalization 108
4.4.3 Interpretation of Parameters 110
4.4.4 Topological Model 111
4.4.5 Quasi-independence Model 114
4.4.6 Symmetry and Quasi-symmetry 116
4.4.7 Crossings Model 117
4.5 Models for Ordinal Variables 119
4.5.1 Linear-by-Linear Association 119
4.5.2 Uniform Association 120
4.5.3 Row-Effect and Column-Effect Models 122
4.5.4 Goodman's *RC* Model 124
4.6 Models for Multiway Tables 129
4.6.1 Three-Way Tables 130
4.6.2 The Saturated Model for Three-Way Tables 132
4.6.3 Collapsibility 133
4.6.4 Classes of Models for Three-Way Tables 135
4.6.5 Analysis of Variation in Association 140
4.6.6 Model Selection 145

5 Statistical Models for Rates

5.1 Introduction 147

5.2 Log-Rate Models 148
5.2.1 The Role of Exposure 149
5.2.2 Estimating Log-Rate Models 154
5.2.3 Illustration 156
5.2.4 Interpretation 159
5.3 Discrete-Time Hazard Models 160
5.3.1 Data Structure 161
5.3.2 Estimation 162
5.4 Semiparametric Rate Models 168
5.4.1 The Piecewise Constant Exponential Model 169
5.4.2 The Cox Model 174
5.5 Models for Panel Data 177
5.5.1 Fixed Effects Models for Binary Data 179
5.5.2 Random Effects Models for Binary Data 183
5.6 Unobserved Heterogeneity in Event-History Models 188
5.6.1 The Gamma Mixture Model 190
5.7 Summary 199

6 Models for Ordinal Dependent Variables

6.1 Introduction 201
6.2 Scoring Methods 202
6.2.1 Integer Scoring 202
6.2.2 Midpoint Scoring 203
6.2.3 Normal Score Transformation 204
6.2.4 Scaling with Additional Information 205
6.3 Logit Models for Grouped Data 206
6.3.1 Baseline, Adjacent, and Cumulative Logits 206
6.3.2 Adjacent Category Logit Model 207
6.3.3 Adjacent Category Logit Models and Loglinear Models 209
6.4 Ordered Logit and Probit Models 210
6.4.1 Cumulative Logits and Probits 211
6.4.2 The Ordered Logit Model 212
6.4.3 The Ordered Probit Model 214
6.4.4 The Latent Variable Approach 215
6.4.5 Estimation 217
6.4.6 Marginal Effects 220
6.5 Summary 222

7 Models for Unordered Dependent Variables

7.1 Introduction 223
7.2 Multinomial Logit Models 224
7.2.1 Review of the Binary Logit Model 224
7.2.2 General Setup for the Multinomial Logit Model 225
7.3 The Standard Multinomial Logit Model 227
7.3.1 Estimation 229
7.3.2 Interpreting Results from Multinomial Logit Models 230
7.4 Loglinear Models for Grouped Data 234
7.4.1 Two-Way Tables 234
7.4.2 Three- and Higher-Way Tables 235
7.5 The Latent Variable Approach 238
7.6 The Conditional Logit Model 239
7.6.1 Interpretation 240
7.6.2 The Mixed Model 242
7.7 Specification Issues 245
7.7.1 Independence of Irrelevant Alternatives: The IIA Assumption 245
7.7.2 Sequential Logit Models 249
7.8 Summary 252

A The Matrix Approach to Regression

A.1 Introduction 253
A.2 Matrix Algebra 253
A.2.1 The Matrix Approach to Regression 254
A.2.2 Basic Matrix Operations 255
A.2.3 Numerical Example 259

B Maximum Likelihood Estimation

B.1 Introduction 261
B.2 Basic Principles 261
B.2.1 Example 1: Binomial Proportion 262
B.2.2 Example 2: Normal Mean and Variance 264
B.2.3 Example 3: Binary Logit Model 266
B.2.4 Example 4: Loglinear Model 272
B.2.5 Iteratively Reweighted Least Squares 275
B.2.6 Generalized Linear Models 277
B.2.7 Minimum χ^2 Estimation 281