CONTENTS

Preface xv

Acknowledgements xvii

Part 1 Some essential preliminaries 1

Chapter 1 Introduction 3

1.1 Classifications, categorical data, and a methodological transformation 4

1.2 A framework for discussion 7

1.3 Some important tables 8

1.4 Sampling schemes 13

1.4.1 Poisson 14

1.4.2 Multinomial 14

1.4.3 Product-multinomial 14

1.5 Integration and transition 15

1.6 How to use the book and what it assumes 16

Part 2 The basic family of statistical models 19

Chapter 2 Categorical response variable, continuous explanatory variables 21

2A TWO RESPONSE CATEGORIES: THE DICHOTOMOUS CASE 21

2.1 Introduction 21

2.2 A conventional regression model approach 22

2.3 Alternative solutions 24
2.3.1 The linear probability model 25
2.3.2 The logit model 28
2.3.3 The probit model 29
2.4 Estimating the parameters of a logistic/logit model 30
2.4.1 Grouped data and weighted least squares 30
2.4.2 Individual data and maximum likelihood 35
2.5 Some simple examples 38
EXAMPLE 2.1 *Oil and gas exploration in south-central Kansas* 39
EXAMPLE 2.2 *Work trip mode choice in Sydney* 40
EXAMPLE 2.3 *Shopping trip mode choice in Pittsburgh* 42
2.6 Testing hypotheses about the logistic/logit model parameters 43
2.6.1 Separate tests 43
2.6.2 Joint tests 45
2.6.3 Partial joint tests 45
EXAMPLE 2.4 *Oil and gas exploration in south-central Kansas (continued)* 46
2.7 Goodness-of-fit measures, residuals, and predicted values 49
2.7.1 Goodness-of-fit measures 49
2.7.2 Residuals 52
EXAMPLE 2.5 *Oil and gas exploration in south-central Kansas (continued)* 54
EXAMPLE 2.6 *Resource evaluation in Newfoundland* 58
2B MULTIPLE RESPONSE CATEGORIES: THE POLYTOMOUS CASE 62
2.8 Introduction 62
2.9 The extended linear logit and logistic models 62
2.9.1 The linear logit model 63
2.9.2 The logistic model 65
2.9.3 General forms 66
2.10 Estimating the parameters of the extended models 67
EXAMPLE 2.7 *Oil and gas exploration in south-central Kansas (continued)* 69
2.11 Types of explanatory variables and response categories 72
2.11.1 Types of explanatory variables 72
2.11.2 Types of response categories 76

EXAMPLE 2.8 *Shopping trip destination choice in Pittsburgh* 77

EXAMPLE 2.9 *Shopping trip destination choice in West Yorkshire* 78

EXAMPLE 2.10 *Shopping destination and mode of travel choice in San Francisco* 81

EXAMPLE 2.11 *Shopping destination and mode of travel choice in Eindhoven* 82

Appendix 2.1 Alternative derivations of logistic/logit models 84

Chapter 3 Categorical response variable, mixed explanatory variables 91

3.1 Dummy variables in conventional regression models 91

3.2 Categorical explanatory variables in logistic/logit models 94

3.3 Dummy variables within the general typology of explanatory variables 96

3.4 A range of illustrative examples 97

EXAMPLE 3.1 *Determinants of housing tenure in Sydney* 97

EXAMPLE 3.2 *Occupational attainment in the United States* 99

EXAMPLE 3.3 *Housing choice in Pittsburgh* 103

EXAMPLE 3.4 *Work trip mode choice in Washington D.C. and the spatial transferability of models* 106

3.5 Summary 109

Chapter 4 Categorical response variable, categorical explanatory variables: the linear logit model approach 111

4.1 Linear logit models for cell (f): basic forms 112

EXAMPLE 4.1 *Preference for army camp location among American soldiers* 114

EXAMPLE 4.2 *Evaluation of military policemen by negro soldiers from different regions* 117

4.2 Weighted least squares estimation of cell (f) linear logit models 121

4.3 Goodness-of-fit and test statistics 125

4.3.1 Joint tests 126

4.3.2 Partial joint tests 127

4.3.3 Separate tests 128

EXAMPLE 4.3 *Automobile accidents and the accident environment in North Carolina* 129

4.4 Coding systems for categorical explanatory variables 132

4.5 Interactions, saturation, hierarchies and parsimony 136

EXAMPLE 4.4 Determinants of homeownership in Boston and Baltimore 140

4.6 Model selection 143

EXAMPLE 4.5 Byssinosis amongst U.S. cotton textile workers 147

4.7 A more general matrix formulation 152

Chapter 5 All variables categorical but no division into response and explanatory 155

5.1 The hypothesis of independence and the chi-square test 156

5.2 Towards a log-linear model of independence 157

5.3 A hierarchical set of log-linear models for two-dimensional contingency tables 161

EXAMPLE 5.1 Some simple two-dimensional contingency tables 163

(a) Pebbles in glacial till 164

(b) Lifetime residential mobility and retirement migration 166

(c) Farm acreage under woodland 168

5.4 Log-linear models for multidimensional contingency tables 169

EXAMPLE 5.2 Age, decay and use of buildings in north-east London 175

EXAMPLE 5.3 Industrial location in Hull 176

5.5 Abbreviated notation systems for log-linear models 179

5.6 Estimation of the parameters and the expected cell frequencies 182

5.6.1 The iterative proportional fitting procedure 184

5.6.2 The iterative weighted least squares procedure 188

5.6.3 The Newton–Raphson procedure 190

5.7 Model selection 190

5.7.1 Strategy 1. Stepwise selection 191

5.7.2 Strategy 2. Abbreviated stepwise selection 194

5.7.3 Strategy 3. Screening 196

5.7.4 Strategy 4. Aitkin's simultaneous test procedure 200

EXAMPLE 5.4 Shopping behaviour in Manchester 204

EXAMPLE 5.5 Non-fatal deliberate self-harm in Bristol 207

5.8 The analysis of residuals 211

EXAMPLE 5.6 Opinions about a television series in urban and rural areas 213

Chapter 6 Categorical response variable, categorical explanatory variables: the log-linear model approach 215

6.1 Fixed marginal totals 215

6.2 Log-linear models for mixed explanatory/response variable situations 216

EXAMPLE 6.1 *A hypothetical three-dimensional table* 217

EXAMPLE 6.2 *Shopping behaviour in Manchester (continued)* 219

EXAMPLE 6.3 *Relationships between ethnic origin, birthplace, age, and occupation in Canada in 1871* 221

6.3 Log-linear models as logit models 223

6.3.1 The dichotomous response variable case 223

6.3.2 The multiple-category response variable case 227

6.3.3 Discussion 230

Chapter 7 Computer programs for categorical data analysis 233

7.1 A classification of available programs 233

7.2 Programs which use function maximization algorithms 233

7.3 Programs based upon weighted least squares algorithms 236

7.3.1 Iterative weighted least squares 236

7.3.2 Non-iterative weighted least squares 237

7.4 Programs which use iterative proportional fitting algorithms 237

Part 3 Extensions of the basic statistical models 239

Chapter 8 Special topics in logistic/logit modelling 241

8.1 Logistic regression diagnostics and resistant fitting 241

8.2 Some logistic/logit model analogues to classical spatial analysis models 246

8.2.1 Trend surface models 246

8.2.2 Space-time models 249

EXAMPLE 8.1 *Aircraft noise disturbance around Manchester Airport* 250

EXAMPLE 8.2 *The space–time pattern of housing deterioration in Indianapolis* 252

8.3 Logistic/logit models for ordered categories and systems of equations 254
8.3.1 Ordered response categories 254
8.3.2 Systems of logistic/logit models 255
8.4 Some extensions of the general matrix formulations of cell (f) linear logit models 259
8.4.1 A multiple response category model 260
8.4.2 A repeated-measurement research design example 261
8.4.3 Paired comparison experiment examples 264
EXAMPLE 8.3 Residential preferences of schoolchildren in Southend-on-Sea 269
8.5 An alternative treatment of error structures in cell (f) linear logit models 271

Chapter 9 Special topics in log-linear modelling 275
9.1 Combining categories and collapsing tables 275
EXAMPLE 9.1 Oak, hickory and maple distributions in Lansing Woods, Michigan 277
9.2 Sampling zeros 279
9.2.1 Sampling zeros and saturated log-linear models 279
9.2.2 Sampling zeros and unsaturated log-linear models 281
EXAMPLE 9.2 Non-fatal deliberate self-harm in Bristol (continued) 282
EXAMPLE 9.3 Relationships between tree species and tree height in the forests of South Island, New Zealand 284
9.3 Structural zeros and incomplete contingency tables 285
EXAMPLE 9.4 Filtering in the housing market of Kingston, Ontario 289
EXAMPLE 9.5 Plant type, soil type and slope aspect 292
9.4 Outliers or rogue cells 293
EXAMPLE 9.6 Opinions about a television series in urban and rural areas (continued) 293
9.5 Square tables, symmetry, and marginal homogeneity 295
9.5.1 Symmetry 295
9.5.2 Marginal homogeneity 296
9.5.3 Quasi-symmetry 297
9.5.4 Symmetry and marginal homogeneity in multidimensional tables 299
9.5.5 Alternative log-linear models for square tables 300

EXAMPLE 9.7 *Filtering in the housing market of Kingston, Ontario (continued)* 301

9.6 Some remaining issues 303

9.6.1 The multiplicative form of the log-linear model 303

9.6.2 Log-linear models for tables with ordered categories 304

9.6.3 Causal analysis with log-linear models 305

9.6.4 Log-linear models and spatially dependent data 307

Part 4 Discrete choice modelling 311

Chapter 10 Statistical models for discrete choice analysis 313

10.1 Random utility maximization, discrete choice theory and multinomial logit models 313

EXAMPLE 10.1 *The collapse and re-opening of the Tasman Bridge* 318

10.2 The IIA property and its implications 324

10.3 The search for less restrictive discrete choice models 326

10.3.1 The multinomial probit model 327

10.3.2 The dogit model 328

10.3.3 The nested logit model 329

10.3.4 Elimination-by-aspects models 332

10.3.5 Weight shifting models 334

EXAMPLE 10.2 *Location decisions of clothing retailers in Boston* 337

EXAMPLE 10.3 *Travel mode choice in Montreal* 340

EXAMPLE 10.4 *Travel mode choice in the Rotterdam/Hague Metropolitan area.* 341

10.4 Assessing and comparing the performance of alternative discrete choice models 343

10.4.1 Tests of the IIA property of the MNL 344

10.4.2 Tests of the MNL against specific alternative discrete choice models 347

10.4.3 A generalized test procedure for comparing the performance of any pair of discrete choice models 349

10.5 A brief guide to some remaining statistical issues 350

10.5.1 Statistical transformations and the search for appropriate functional form 350

10.5.2 Sample design and parameter estimation 351

10.5.3 Panel data and dynamic modelling 353

10.5.4 Specification analysis: improper exclusion or inclusion of explanatory variables 355

10.5.5 Wider themes of empirical application 357

Part 5 Towards integration 359

Chapter 11 An alternative framework 361

11.1 The central classification scheme reconsidered 361

11.2 The GLM framework 362

11.2.1 The linear predictor 363

11.2.2 The link function 364

11.2.3 The error distribution 364

11.2.4 Some examples of GLMs 365

11.3 Conclusion 366

References 367

Author index 383

Example index 387

Subject index 389