CONTENTS

PREFACE ix

1 INTRODUCTION 1
- 1.1 The Need 1
- 1.2 Why a Book? 1
- 1.3 Different Users 2
- 1.4 Sketch of the Chapters 2
- 1.5 Computer Programs 7
- 1.6 How to Proceed from Here 7

2 STRUCTURAL MODELS FOR COUNTED DATA 9
- 2.1 Introduction 9
- 2.2 Two Dimensions—The Fourfold Table 11
- 2.3 Two Dimensions—The Rectangular Table 24
- 2.4 Models for Three-Dimensional Arrays 31
- 2.5 Models for Four or More Dimensions 42
- 2.6 Exercises 48
- 2.7 Appendix: The Geometry of a 2×2 Table 49

3 MAXIMUM LIKELIHOOD ESTIMATES FOR COMPLETE TABLES 57
- 3.1 Introduction 57
- 3.2 Sampling Distributions 62
- 3.3 Sufficient Statistics 64
- 3.4 Methods of Obtaining Maximum Likelihood Estimates 73
- 3.5 Iterative Proportional Fitting of Log-Linear Models 83
- 3.6 Classical Uses of Iterative Proportional Fitting 97
- 3.7 Rearranging Data for Model Fitting 102
- 3.8 Degrees of Freedom 114

4 FORMAL GOODNESS OF FIT: SUMMARY STATISTICS AND MODEL SELECTION 123
- 4.1 Introduction 123
- 4.2 Summary Measures of Goodness of Fit 124
- 4.3 Standardized Rates 131
- 4.4 Internal Goodness of Fit 136
- 4.5 Choosing a Model 155
- 4.6 Appendix: Goodman's Partitioning Calculus 169

5 MAXIMUM LIKELIHOOD ESTIMATION FOR INCOMPLETE TABLES 177
- 5.1 Introduction 177
- 5.2 Incomplete Two-Way Tables 178
- 5.3 Incomplete Two-Way Tables for Subsets of Complete Arrays 206
- 5.4 Incomplete Multiway Tables 210
- 5.5 Representation of Two-Way Tables as Incomplete Multiway Arrays 225

6 ESTIMATING THE SIZE OF A CLOSED POPULATION 229
6.1 Introduction 229
6.2 The Two-Sample Capture-Recapture Problem 231
6.3 Conditional Maximum Likelihood Estimation of N 236
6.4 The Three-Sample Census 237
6.5 The General Multiple Recapture Problem 246
6.6 Discussion 254

7 MODELS FOR MEASURING CHANGE 257
7.1 Introduction 257
7.2 First-Order Markov Models 261
7.3 Higher-Order Markov Models 267
7.4 Markov Models with a Single Sequence of Transitions 270
7.5 Other Models 273

8 ANALYSIS OF SQUARE TABLES: SYMMETRY AND MARGINAL HOMOGENEITY 281
8.1 Introduction 281
8.2 Two-Dimensional Tables 282
8.3 Three-Dimensional Tables 299
8.4 Summary 309

9 MODEL SELECTION AND ASSESSING CLOSENESS OF FIT: PRACTICAL ASPECTS 311
9.1 Introduction 311
9.2 Simplicity in Model Building 312
9.3 Searching for Sampling Models 315
9.4 Fitting and Testing Using the Same Data 317
9.5 Too Good a Fit 324
9.6 Large Sample Sizes and Chi Square When the Null Model is False 329
9.7 Data Anomalies and Suppressing Parameters 332
9.8 Frequency of Frequencies Distribution 337

10 OTHER METHODS FOR ESTIMATION AND TESTING IN CROSS-CLASSIFICATIONS 343
10.1 Introduction 343
10.2 The Information-Theoretic Approach 344
10.3 Minimizing Chi Square, Modified Chi Square, and Logit Chi Square 348
10.4 The Logistic Model and How to Use It 357
10.5 Testing via Partitioning of Chi Square 361
10.6 Exact Theory for Tests Based on Conditional Distributions 364
10.7 Analyses Based on Transformed Proportions 366
10.8 Necessary Developments 371

11 MEASURES OF ASSOCIATION AND AGREEMENT 373
11.1 Introduction 373
11.2 Measures of Association for 2 × 2 Tables 376
11.3 Measures of Association for $I \times J$ Tables 385
11.4 Agreement as a Special Case of Association 393

12 PSEUDO-BAYES ESTIMATES OF CELL PROBABILITIES 401
12.1 Introduction 401
12.2 Bayes and Pseudo-Bayes Estimators 404
12.3 Asymptotic Results for Pseudo-Bayes Estimators 410
12.4 Small-Sample Results 416
12.5 Data-Dependent λ's 419
12.6 Another Example: Two Social Mobility Tables 426
12.7 Recent Results and Some Advice 429

13 SAMPLING MODELS FOR DISCRETE DATA 435
13.1 Introduction 435
13.2 The Binomial Distribution 435
13.3 The Poisson Distribution 438
13.4 The Multinomial Distribution 441
13.5 The Hypergeometric Distribution 448
13.6 The Multivariate Hypergeometric Distribution 450
13.7 The Negative Binomial Distribution 452
13.8 The Negative Multinomial Distribution 454

14 ASYMPTOTIC METHODS 457
14.1 Introduction 457
14.2 The O, o Notation 458
14.3 Convergence of Stochastic Sequences 463
14.4 The O_p, o_p Notation for Stochastic Sequences 475
14.5 Convergence of Moments 484
14.6 The δ Method for Calculating Asymptotic Distributions 486
14.7 General Framework for Multinomial Estimation and Testing 502
14.8 Asymptotic Behavior of Multinomial Maximum Likelihood Estimators 509
14.9 Asymptotic Distribution of Multinomial Goodness-of-Fit Tests 513

REFERENCES 531

INDEX TO DATA SETS 543

AUTHOR INDEX 547

SUBJECT INDEX 551