

# Contents

|  |    |
|--|----|
| <b>1 Preliminaries</b> . . . . .   | 1  |
| 1.1 Introduction . . . . .   | 1  |
| 1.2 The Fixed Population Model. . . . .  | 2  |
| 1.3 Different Types of Sampling Designs. . . . .   | 8  |
| 1.4 The Estimators . . . . .   | 11 |
| 1.4.1 A Class of Estimators . . . . .  | 14 |
| 1.5 Model-Based Approach to Sampling . . . . .   | 17 |
| 1.5.1 Uses of Design and Model in Sampling . . . . .                                     | 23 |
| 1.6 Plan of the Book. . . . .  | 24 |
| <b>2 The Design Effects and Misspecification Effects</b> . . . . .                       | 27 |
| 2.1 Introduction . . . . .   | 28 |
| 2.2 Effect of a Complex Design on the Variance of an Estimator . . . . .                 | 30 |
| 2.3 Effect of a Complex Design on Confidence Interval for $\theta$ . . . . .             | 37 |
| 2.4 Multivariate Design Effects. . . . .   | 40 |
| 2.5 Nonparametric Methods of Variance Estimation . . . . .                               | 41 |
| 2.5.1 A Simple Method of Estimation of Variance<br>of a Linear Statistic . . . . .       | 41 |
| 2.5.2 Linearization Method for Variance Estimation<br>of a Nonlinear Estimator . . . . . | 45 |
| 2.5.3 Random Group Method . . . . .  | 48 |
| 2.5.4 Balanced Repeated Replications . . . . .   | 52 |
| 2.5.5 The Jackknife Procedures. . . . .  | 56 |
| 2.5.6 The Jackknife Repeated Replication (JRR) . . . . .                                 | 58 |
| 2.5.7 The Bootstrap. . . . .   | 60 |
| 2.6 Effect of Survey Design on Inference About<br>Covariance Matrix. . . . .             | 64 |
| 2.7 Exercises and Complements . . . . .  | 65 |

|          |  |     |
|----------|--|-----|
| <b>3</b> | <b>Some Classical Models in Categorical Data Analysis</b>                  | 67  |
| 3.1      | Introduction   | 67  |
| 3.2      | Statistical Models   | 68  |
| 3.2.1    | Fitting Statistical Models   | 69  |
| 3.2.2    | Large Sample Estimation Theory   | 69  |
| 3.2.3    | Asymptotic Properties of ML Estimates                                      | 70  |
| 3.2.4    | Testing of Parameters  | 71  |
| 3.2.5    | Transformation of the Central Limit Theorem                                | 73  |
| 3.3      | Distribution Theory for Count Data   | 73  |
| 3.3.1    | Multinomial Models   | 73  |
| 3.3.2    | Poisson Models   | 76  |
| 3.3.3    | The Multinomial-Poisson Connection   | 78  |
| 3.4      | Goodness-of-Fit  | 78  |
| 3.4.1    | Likelihood Ratio Statistic   | 78  |
| 3.4.2    | Pearson's Goodness-of-Fit Statistic  | 79  |
| 3.5      | Binomial Data  | 81  |
| 3.5.1    | Binomial Data and the Log-Likelihood Ratio                                 | 81  |
| 3.6      | Log-Linear Models  | 82  |
| 3.6.1    | Log-Linear Models for Two-Way Tables                                       | 82  |
| 3.6.2    | Log-Linear Models for Three-Way Tables                                     | 84  |
| 3.7      | Logistic Regression Analysis   | 89  |
| 3.7.1    | The Logistic Regression Model  | 89  |
| 3.7.2    | Fitting the Logistic Regression Model                                      | 90  |
| 3.7.3    | The Multiple Logistic Regression Model                                     | 91  |
| 3.7.4    | Fitting the Multiple Logistic Regression Model                             | 92  |
| 3.7.5    | Polytomous Logistic Regression   | 94  |
| 3.8      | Fitting the Logistic Regression Models to Data<br>from Complex Surveys     | 95  |
| <b>4</b> | <b>Analysis of Categorical Data Under a Full Model</b>                     | 97  |
| 4.1      | Introduction   | 97  |
| 4.2      | Tests of Goodness-of-Fit   | 98  |
| 4.2.1    | Pearsonian Chi-Square Statistic  | 98  |
| 4.2.2    | Design-Based Wald Statistic  | 99  |
| 4.2.3    | Neyman's (Multinomial Wald) Statistic                                      | 101 |
| 4.2.4    | Log-Likelihood Ratio Statistic   | 101 |
| 4.2.5    | Asymptotic Distribution of $X_W^2$ and $X_P^2$                             | 101 |
| 4.2.6    | Generalized Design Effect of $\pi$   | 107 |
| 4.2.7    | Modification to $X_P^2$  | 109 |
| 4.2.8    | Fay's Jackknifed Statistic   | 112 |
| 4.3      | Testing for Homogeneity  | 115 |
| 4.3.1    | A Simple Method for Binary Data<br>from Cluster Sampling                   | 120 |
| 4.4      | Effect of Survey Design on Classical Tests<br>of General Linear Hypotheses | 121 |

4.5 Tests of Independence in a Two-Way Table . . . . . 123

4.6 Some Evaluation of Tests Under Cluster Sampling . . . . . 128

4.7 Exercises and Complements . . . . . 130

**5 Analysis of Categorical Data Under Log-Linear Models . . . . . 135**

5.1 Introduction . . . . . 135

5.2 Log-Linear Models in Contingency Tables . . . . . 136

5.3 Tests for Goodness of Fit . . . . . 138

5.3.1 Other Standard Tests and Their First- and Second-Order Corrections . . . . . 139

5.3.2 Fay’s Jackknifed Tests . . . . . 143

5.4 Asymptotic Covariance Matrix of the Pseudo-MLE  $\hat{\pi}$  . . . . . 145

5.4.1 Residual Analysis . . . . . 146

5.5 Brier’s Model . . . . . 147

5.6 Nested Models . . . . . 149

5.6.1 Pearsonian Chi-Square and the Likelihood Ratio Statistic . . . . . 150

5.6.2 A Wald Statistic . . . . . 153

5.6.3 Modifications to Test Statistics . . . . . 154

5.6.4 Effects of Survey Design on  $X^2_p(2|1)$  . . . . . 155

**6 Analysis of Categorical Data Under Logistic Regression Model . . . . . 157**

6.1 Introduction . . . . . 157

6.2 Binary Logistic Regression . . . . . 158

6.2.1 Pseudo-MLE of  $\pi$  . . . . . 159

6.2.2 Asymptotic Covariance Matrix of the Estimators . . . . . 160

6.2.3 Goodness-of-Fit Tests . . . . . 163

6.2.4 Modifications of Tests . . . . . 164

6.3 Nested Model . . . . . 164

6.3.1 A Wald Statistic . . . . . 167

6.3.2 Modifications to Tests . . . . . 168

6.4 Choosing Appropriate Cell-Sample Sizes for Running Logistic Regression Program in a Standard Computer Package . . . . . 169

6.5 Model in the Polytomous Case . . . . . 170

6.6 Analysis Under Generalized Least Square Approach . . . . . 172

6.7 Exercises and Complements . . . . . 176

**7 Analysis in the Presence of Classification Errors . . . . . 179**

7.1 Introduction . . . . . 179

7.2 Tests for Goodness-of-Fit Under Misclassification . . . . . 180

7.2.1 Methods for Considering Misclassification Under SRS . . . . . 180

7.2.2 Methods for General Sampling Designs . . . . . 182

7.2.3 A Model-Free Approach . . . . . 183

- 7.3 Tests of Independence Under Misclassification . . . . . 185
  - 7.3.1 Methods for Considering Misclassification Under SRS. . . . . 186
  - 7.3.2 Methods for Arbitrary Survey Designs. . . . . 186
- 7.4 Test of Homogeneity . . . . . 188
- 7.5 Analysis Under Weighted Cluster Sample Design . . . . . 192
- 8 Approximate MLE from Survey Data. . . . . 195**
  - 8.1 Introduction . . . . . 195
  - 8.2 Exact MLE from Survey Data. . . . . 196
    - 8.2.1 Ignorable Sampling Designs . . . . . 196
    - 8.2.2 Exact MLE . . . . . 197
  - 8.3 MLE’s Derived from Weighted Distributions . . . . . 198
  - 8.4 Design-Adjusted Maximum Likelihood Estimation. . . . . 200
    - 8.4.1 Design-Adjusted Regression Estimation  
with Selectivity Bias . . . . . 205
  - 8.5 The Pseudo-Likelihood Approach to MLE  
from Complex Surveys. . . . . 208
    - 8.5.1 Analysis Based on Generalized Linear Model . . . . . 209
    - 8.5.2 Estimation for Linear Models . . . . . 212
  - 8.6 A Mixed (Design-Model) Framework. . . . . 216
  - 8.7 Effect of Survey Design on Multivariate Analysis  
of Principal Components . . . . . 220
    - 8.7.1 Estimation of Principal Components . . . . . 222
- Appendix A: Asymptotic Properties of Multinomial Distribution . . . . . 223**
- References . . . . . 241**



<http://www.springer.com/978-981-10-0870-2>

Complex Surveys

Analysis of Categorical Data

Mukhopadhyay, P.

2016, XV, 248 p., Hardcover

ISBN: 978-981-10-0870-2