# Contents

**List of Contributors** xi

**Preface** xiii

**CHAPTER 1. GENERAL INTRODUCTION** 1

1.1 Analytic and Descriptive Surveys 1
1.2 Problems with Applying Standard Analytic Procedures to Survey Data 5
1.3 Aggregation and Disaggregation 8
1.4 Survey Design and Population Structure 10
1.5 Objectives and Structure of the Book 12
1.6 Formal Framework, Terminology and Notation 13
1.6.1 Finite Populations, Superpopulation Models and Parameters 13
1.6.2 Auxiliary Information and Population Structure 14
1.6.3 Sampling Designs 15
1.6.4 Inference 17
1.6.5 Additional Notation 19

PART A—AGGREGATED ANALYSIS: STANDARD ERRORS AND SIGNIFICANCE TESTS

**CHAPTER 2. INTRODUCTION TO PART A** *C. J. Skinner* 23

2.1 Introduction 23
2.2 Standard Errors, Design Effects and Misspecification Effects 24
2.3 Confidence Intervals 28
2.4 Significance Tests 30
2.5 Design Factors and Effective Sample Size 31
2.6 Example 2: Cluster Sampling of Unemployment Rates 31
2.7 Design Effects of Without-Replacement Sampling and Finite Population Corrections 35
2.8 Design Effects of Multi-stage Sampling and Intra-Cluster Correlation 36
2.9 Design Effects of Stratified Sampling 39
2.10 Some Actual Design Effects 40
2.11 Multivariate Design Effects 43
2.12 Allowing for Complex Designs 44

2.13 Non-parametric Variance Estimation 46
2.14 Normal Approximations and Confidence Intervals 55
2.15 Introduction to Chapters 3–5 57

**CHAPTER 3. DOMAIN MEANS, REGRESSION AND MULTIVARIATE ANALYSIS** *C. J. Skinner* **59**

3.1 Introduction 59
3.2 Family Expenditure Survey Data 59
3.3 Impact of Design on IID Procedures 60
3.3.1 Introduction 60
3.3.2 Domain Means 60
3.3.3 Differences of Domain Means 64
3.3.4 Linear Regression 65
3.3.5 Other Statistics 70
3.3.6 Hypothesis Testing 71
3.4 Allowing for Complex Designs 75
3.4.1 Domain Means 75
3.4.2 Linear Regression 76
3.4.3 Inference about Covariance Matrices 79
3.4.4 Parametric Modelling 1—Pseudo MLE 80
3.4.5 Parametric Modelling 2—Moment Structures and GLS 84

**CHAPTER 4. CHI-SQUARED TESTS FOR CONTINGENCY TABLE** *J. N. K. Rao and D. R. Thomas* **89**

4.1 Introduction 89
4.2 Impact of Design on Standard Tests 90
4.2.1 Simple Goodness of Fit 90
4.2.2 Multi-way Contingency Tables 92
4.2.3 Nested Models 95
4.3 First-order Corrections to Standard Tests 96
4.3.1 Simple Goodness of Fit 97
4.3.2 Two-way Tables 98
4.3.3 Three-way Tables 99
4.3.4 Nested Models 101
4.4 Second-order Corrections to $X_P^2$ or $X_{LR}^2$ 101
4.5 Wald Tests 102
4.6 Smoothed Estimates 103
4.7 Residual Analysis 103
4.8 Examples 104
4.8.1 Example 1: Goodness of Fit 105
4.8.2 Example 2: Test of Independence 107
4.9 Summary 114

**CHAPTER 5. MEASURES OF ASSOCIATION FOR CONTINGENCY TABLES** *E. A. Molina C.* **115**

5.1 Introduction 115
5.1.1 Measures Based on Pearson's Chi-square Statistic 116

5.1.2 Measures Based on Probabilistic Models of Predictive Activity 117
5.1.3 Measures Based on the Cross-ratios of the Tables 118
5.2 Large-Sample Inference under Multinomial Sampling 122
5.2.1 Inference for Measures with a Ratio Structure 122
5.2.2 Measures Based on Contrasts of Log Ratios 123
5.2.3 Comparison of Two Contingency Tables 124
5.3 Effects of Ignoring the Sampling Scheme 125
5.3.1 Measures with a Ratio Structure 126
5.3.2. Measures Based on Contrasts of Log Ratios 127
5.4 Inference Procedures for Complex Surveys 128
5.4.1 General Large-sample Inference 128
5.4.2 Bounds and Corrections for the Statistics 130

PART B—AGGREGATED ANALYSIS: POINT ESTIMATION AND BIAS

**CHAPTER 6. INTRODUCTION TO PART B** *T. M. F. Smith* **135**

6.1 Introduction 135
6.2 The Bias Effect of Stratfication 136
6.3 The Bias Effect of an Auxiliary Design Variable 139
6.4 The Pearson Adjustment for Multivariate Normal Models 141
6.5 Alternatives to the Pearson Adjustments 143
6.6 Ignorable and Uninformative Sampling Schemes 146

**CHAPTER 7. THE EFFECT OF SELECTION ON REGRESSION ANALYSIS** *G. Nathan and T. M. F. Smith* **149**

7.1 Introduction 149
7.2 Some Examples of Selection Schemes 150
7.3 The Bias of OLS 153
7.4 Alternatives to OLS 154
7.5 Some Empirical Studies of the Effects of Selection 157

**CHAPTER 8. MULTIVARIATE ANALYSIS** *T. M. F. Smith and D. J. Holmes* **165**

8.1 Introduction 165
8.2 Alternative Estimators 166
8.3 Properties of the Estimators under Linearity and Homoscedasticity Conditions 168
8.3.1 Model-based Biases Conditional on $s$ and $\mathbf{z}$ 168
8.3.2 Model-based Biases Conditional on $\mathbf{z}$ only (Unconditional Biases) 169
8.4 Applications in Multivariate Analysis 169
8.5 Study I: Multivariate Normal Data 171
8.5.1 Introduction 171

8.5.2 Principal Component Analysis 172
8.5.3 Regression Analysis 179
8.6 Study II: 'Real' Data 182
8.6.1 Introduction 182
8.6.2 Unconditional Results 182
8.6.3 Conditional Results 189
8.7 Conclusions 187

**CHAPTER 9. SELECTION BASED ON THE RESPONSE VARIABLE IN LOGISTIC REGRESSION** *A. J. Scott and C. J. Wild* **191**
9.1 Introduction 191
9.2 Design-based Estimators 195
9.3 Model-based Estimators 198
9.4 Simulation Results 201
9.5 Stratified Case-control Studies 203
9.6 Conclusion 205

PART C—DISAGGREGATED ANALYSIS: MODELLING STRUCTURED POPULATIONS

**CHAPTER 10. INTRODUCTION TO PART C** *D. Holt* **209**
10.1 Aggregation versus Disaggregation 209
10.2 Modelling the Structure of the Survey Population 215

**CHAPTER 11. MULTILEVEL AND MULTIVARIATE MODELS IN SURVEY ANALYSIS** *H. Goldstein and R. Silver* **221**
11.1 Introduction 221
11.2 The Two-level Model 222
11.3 Estimation in the Multilevel Model 225
11.4 Loglinear Multilevel Models for Proportions 226
11.5 Examples 228
11.6 Discussion 235
11.7 Acknowledgements 235

**CHAPTER 12. REGRESSION MODELS FOR STRATIFIED MULTI-STAGE CLUSTER SAMPLES** *D. Pfeffermann and L. LaVange* **237**
12.1 Introduction 237
12.2 The Model 238
12.3 Estimation of the Fixed Regression Coefficients 240
12.3.1 The Sample 240
12.3.2 Estimation of $\gamma$ 240

12.4 Estimation of the Unknown Variances 242
12.4.1 Preface 242
12.4.2 Maximum Likelihood Estimation 242
12.4.3 Advantages of the Regression Method for ML Estimation 244
12.5 Model Testing 247
12.5.1 Hypotheses of Interest 247
12.5.2 Large-sample Wald Statistics 247
12.6 Empirical Results I: Estimation and Testing of Fixed Regression Coefficients and Variance Components 249
12.6.1 The NHANES Design 249
12.6.2 Statement of Problem and Regression Variables 249
12.6.3 Design Clusters and Design Cluster Characteristics 250
12.6.4 Analysis and Results 250
12.7 Prediction of Design Cluster Coefficients 254
12.7.1 Definitions and Theoretical Formulae 254
12.7.2 Empirical Results II: Prediction of Design Cluster Coefficients 256
12.8 Model Misspecification and $p$-weighted Estimators 258
12.9 Acknowledgements 259
Appendix: Probability Weighted GLS Estimation of $\gamma$ 259

**CHAPTER 13. LOGISTIC MODELS FOR CONTINGENCY TABLES** *D. Holt and P. D. Ewings* **261**
13.1 Introduction 261
13.2 The Standard Logistic Framework 263
13.3 Modification to Allow for Cluster Effects 264
13.4 Estimation and Hypothesis Testing 266
13.5 The Impact of Population Structure on the $\chi^2$ Test 269
13.6 The Interpretation of the Cluster-effect Covariance Matrix 279

**CHAPTER 14. CONCLUSIONS** **285**
14.1 Overview 285
14.2 Impediments to Progress 287
14.3 Areas for Further Work 287

**REFERENCES** **291**

**AUTHOR INDEX** **303**

**SUBJECT INDEX** **307**