
Contents

Preface.....	xv
Authors.....	xxi
1. Applied Survey Data Analysis: An Overview	1
1.1 Introduction.....	1
1.2 A Brief History of Applied Survey Data Analysis	2
1.2.1 Key Theoretical Developments.....	2
1.2.2 Key Software Developments.....	4
1.3 Example Data Sets and Exercises	5
1.4 Steps in Applied Survey Data Analysis.....	9
2. Getting to Know the Complex Sample Design	15
2.1 Introduction.....	15
2.1.1 Technical Documentation and Supplemental Literature Review.....	15
2.2 Classification of Sample Designs	16
2.2.1 Sampling Plans.....	17
2.2.2 Other Types of Study Designs Involving Probability Sampling.....	18
2.2.3 Inference from Survey Data	19
2.3 Target Populations and Survey Populations.....	20
2.4 Simple Random Sampling: A Simple Model for Design- Based Inference	21
2.4.1 Relevance of SRS to Complex Sample Survey Data Analysis.....	21
2.4.2 SRS Fundamentals: A Framework for Design-Based Inference.....	22
2.4.3 Example of Design-Based Inference under SRS.....	24
2.5 Complex Sample Design Effects	26
2.5.1 Design Effect Ratio	26
2.5.2 Generalized Design Effects and Effective Sample Sizes....	28
2.6 Complex Samples: Cluster Sampling and Stratification.....	30
2.6.1 Cluster Sampling Plans.....	31
2.6.2 Stratification.....	34
2.6.3 Joint Effects of Sample Stratification and Cluster Sampling	37
2.7 Weighting in Analysis of Survey Data.....	38
2.7.1 Introduction to Weighted Analysis of Survey Data.....	38
2.7.2 Weighting for Probabilities of Selection (w_{sel})	40

2.7.3	Nonresponse Adjustment Weights (w_{nr}).....	42
2.7.3.1	Weighting Class Approach ($w_{nr,wc}$).....	42
2.7.3.2	Propensity Cell Adjustment Approach ($w_{nr,prop}$)....	43
2.7.4	Poststratification Weight Factors (w_{ps}).....	45
2.7.5	Design Effects Due to Weighted Analysis	47
2.8	Multistage Area Probability Sample Designs.....	49
2.8.1	Primary Stage Sampling.....	50
2.8.2	Secondary Stage Sampling.....	51
2.8.3	Third- and Fourth-Stage Sampling of HUs and Eligible Respondents	52
2.9	Special Types of Sampling Plans Encountered in Surveys.....	53
3.	Foundations and Techniques for Design-Based Estimation and Inference.....	55
3.1	Introduction.....	55
3.2	Finite Populations and Superpopulation Models	56
3.3	CI's for Population Parameters	58
3.4	Weighted Estimation of Population Parameters.....	59
3.5	Probability Distributions and Design-Based Inference	62
3.5.1	Sampling Distributions of Survey Estimates.....	62
3.5.2	Degrees of Freedom for t under Complex Sample Designs.....	65
3.6	Variance Estimation.....	67
3.6.1	Simplifying Assumptions Employed in Complex Sample Variance Estimation.....	69
3.6.2	TSL Method	70
3.6.3	Replication Methods for Variance Estimation.....	75
3.6.3.1	Jackknife Repeated Replication.....	76
3.6.3.2	Balanced Repeated Replication	80
3.6.3.3	Fay's BRR Method	83
3.6.3.4	Bootstrap (Rao–Wu Rescaling Bootstrap).....	84
3.6.3.5	Construction of Replicate Weights for Replicated Variance Estimation	85
3.6.4	Example Comparing Results from the TSL, JRR, BRR, and Bootstrap Methods.....	86
3.7	Hypothesis Testing in Survey Data Analysis	89
3.8	TSE and Its Impact on Survey Estimation and Inference	91
3.8.1	Variable Errors.....	91
3.8.2	Biases in Survey Data.....	92
4.	Preparation for Complex Sample Survey Data Analysis	97
4.1	Introduction.....	97
4.2	Final Survey Weights: Review by the Data User.....	98
4.2.1	Identification of the Correct Weight Variable(s) for the Analysis	99

4.2.2	Determining the Distribution and Scaling of the Weight Variable(s)	100
4.2.3	Weighting Applications: Sensitivity of Survey Estimates to the Weights.....	102
4.3	Understanding and Checking the Sampling Error Calculation Model.....	104
4.3.1	Stratum and Cluster Codes in Complex Sample Survey Data Sets	105
4.3.2	Building the NCS-R Sampling Error Calculation Model.....	107
4.3.3	Combining Strata, Randomly Grouping PSUs, and Collapsing Strata	109
4.3.4	Checking the Sampling Error Calculation Model for the Survey Data Set	111
4.4	Addressing Item Missing Data in Analysis Variables	113
4.4.1	Potential Bias due to Ignoring Missing Data	114
4.4.2	Exploring Rates and Patterns of Missing Data Prior to Analysis	114
4.5	Preparing to Analyze Data for Sample Subpopulations	116
4.5.1	Subpopulation Distributions across Sample Design Units	118
4.5.2	Unconditional Approach for Subclass Analysis.....	119
4.5.3	Preparation for Subclass Analyses.....	121
4.6	Final Checklist for Data Users	121
5.	Descriptive Analysis for Continuous Variables	125
5.1	Introduction	125
5.2	Special Considerations in Descriptive Analysis of Complex Sample Survey Data	126
5.2.1	Weighted Estimation	126
5.2.2	Design Effects for Descriptive Statistics.....	127
5.2.3	Matching the Method to the Variable Type	128
5.3	Simple Statistics for Univariate Continuous Distributions.....	128
5.3.1	Graphical Tools for Descriptive Analysis of Survey Data	129
5.3.2	Estimation of Population Totals.....	131
5.3.3	Means of Continuous, Binary, or Interval Scale Data.....	136
5.3.4	Standard Deviations of Continuous Variables	139
5.3.5	Estimation of Percentiles, Medians, and Measures of Inequality in Population Distributions for Continuous Variables	140
5.3.5.1	Estimation of Distribution Quantiles	140
5.3.5.2	Estimation of Measures of Inequality in Population Distributions	142

5.4	Bivariate Relationships between Two Continuous Variables	145
5.4.1	X–Y Scatter Plots.....	145
5.4.2	Product Moment Correlation Statistic (r).....	146
5.4.3	Ratios of Two Continuous Variables	148
5.5	Descriptive Statistics for Subpopulations.....	149
5.6	Linear Functions of Descriptive Estimates and Differences of Means	151
5.6.1	Differences of Means for Two Subpopulations	152
5.6.2	Comparing Means over Time	155
6.	Categorical Data Analysis	159
6.1	Introduction.....	159
6.2	Framework for Analysis of Categorical Survey Data	160
6.2.1	Incorporating the Complex Design and Pseudo Maximum Likelihood	160
6.2.2	Proportions and Percentages.....	160
6.2.3	Crosstabulations, Contingency Tables, and Weighted Frequencies	161
6.3	Univariate Analysis of Categorical Data	162
6.3.1	Estimation of Proportions for Binary Variables	162
6.3.2	Estimation of Category Proportions for Multinomial Variables	166
6.3.3	Testing Hypotheses Concerning a Vector of Population Proportions	169
6.3.4	Graphical Display for a Single Categorical Variable.....	170
6.4	Bivariate Analysis of Categorical Data	172
6.4.1	Response and Factor Variables	172
6.4.2	Estimation of Total, Row, and Column Proportions for Two-Way Tables.....	172
6.4.3	Estimating and Testing Differences in Subpopulation Proportions.....	174
6.4.4	χ^2 Tests of Independence of Rows and Columns.....	175
6.4.5	Odds Ratios and Relative Risks	181
6.4.6	Simple Logistic Regression to Estimate the Odds Ratio	183
6.4.7	Bivariate Graphical Analysis.....	184
6.5	Analysis of Multivariate Categorical Data	185
6.5.1	Cochran–Mantel–Haenszel Test.....	186
6.5.2	Log-Linear Models for Contingency Tables.....	188
6.6	Summary.....	191
7.	Linear Regression Models	195
7.1	Introduction.....	195
7.2	Linear Regression Model	197
7.2.1	Standard Linear Regression Model.....	199
7.2.2	Survey Treatment of the Regression Model.....	200

7.3	Four Steps in Linear Regression Analysis.....	202
7.3.1	Step 1: Specifying and Refining the Model.....	202
7.3.2	Step 2: Estimation of Model Parameters.....	203
	7.3.2.1 Estimation for the Standard Linear Regression Model.....	203
	7.3.2.2 Linear Regression Estimation for Complex Sample Survey Data.....	205
7.3.3	Step 3: Model Evaluation.....	210
7.3.4	Step 4: Inference.....	217
	7.3.4.1 Inference Concerning Model Parameters.....	218
	7.3.4.2 Prediction Intervals.....	221
7.4	Some Practical Considerations and Tools.....	223
7.4.1	Distribution of the Dependent Variable.....	223
7.4.2	Parameterization and Scaling for Independent Variables.....	224
7.4.3	Standardization of the Dependent and Independent Variables.....	227
7.4.4	Specification and Interpretation of Interactions and Nonlinear Relationships.....	227
7.4.5	Model-Building Strategies.....	230
7.5	Application: Modeling Diastolic Blood Pressure with the 2011–2012 NHANES Data.....	231
7.5.1	Exploring the Bivariate Relationships.....	232
7.5.2	Naïve Analysis: Ignoring Sample Design Features.....	235
7.5.3	Weighted Regression Analysis.....	236
7.5.4	Appropriate Analysis: Incorporating All Sample Design Features.....	238
8.	Logistic Regression and Generalized Linear Models for Binary Survey Variables.....	257
8.1	Introduction.....	257
8.2	GLMs for Binary Survey Responses.....	258
	8.2.1 Logistic Regression Model.....	260
	8.2.2 Probit Regression Model.....	261
	8.2.3 Complementary-Log-Log Model.....	262
8.3	Building the Logistic Regression Model: Stage 1— Model Specification.....	262
8.4	Building the Logistic Regression Model: Stage 2— Estimation of Model Parameters and Standard Errors.....	263
8.5	Building the Logistic Regression Model: Stage 3— Evaluation of the Fitted Model.....	268
	8.5.1 Wald Tests of Model Parameters.....	268
	8.5.2 GOF and Logistic Regression Diagnostics.....	270
8.6	Building the Logistic Regression Model: Stage 4— Interpretation and Inference.....	272

8.7	Analysis Application	283
8.7.1	Stage 1: Model Specification	283
8.7.2	Stage 2: Model Estimation	285
8.7.3	Stage 3: Model Evaluation.....	286
8.7.4	Stage 4: Model Interpretation/Inference.....	288
8.8	Comparing the Logistic, Probit, and C-L-L GLMs for Binary Dependent Variables.....	293

9.	Generalized Linear Models for Multinomial, Ordinal, and Count Variables.....	299
9.1	Introduction	299
9.2	Analyzing Survey Data Using Multinomial Logit Regression Models.....	299
9.2.1	Multinomial Logit Regression Model.....	299
9.2.2	Multinomial Logit Regression Model: Specification Stage.....	301
9.2.3	Multinomial Logit Regression Model: Estimation Stage.....	302
9.2.4	Multinomial Logit Regression Model: Evaluation Stage.....	304
9.2.5	Multinomial Logit Regression Model: Interpretation Stage.....	304
9.2.6	Example: Fitting a Multinomial Logit Regression Model to Complex Sample Survey Data	305
9.3	Logistic Regression Models for Ordinal Survey Data.....	313
9.3.1	Cumulative Logit Regression Model	314
9.3.2	Cumulative Logit Regression Model: Specification Stage.....	315
9.3.3	Cumulative Logit Regression Model: Estimation Stage.....	315
9.3.4	Cumulative Logit Regression Model: Evaluation Stage.....	316
9.3.5	Cumulative Logit Regression Model: Interpretation Stage.....	317
9.3.6	Example: Fitting a Cumulative Logit Regression Model to Complex Sample Survey Data	319
9.4	Regression Models for Count Outcomes	323
9.4.1	Survey Count Variables and Regression Modeling Alternatives.....	323
9.4.2	Generalized Linear Models for Count Variables.....	326
9.4.2.1	Poisson Regression Model	326
9.4.2.2	Negative Binomial Regression Model.....	327
9.4.2.3	Two-Part Models: Zero-Inflated Poisson and Negative Binomial Regression Models	327

9.4.3	Regression Models for Count Data: Specification Stage.....	329
9.4.4	Regression Models for Count Data: Estimation Stage	329
9.4.5	Regression Models for Count Data: Evaluation Stage.....	330
9.4.6	Regression Models for Count Data: Interpretation Stage.....	330
9.4.7	Example: Fitting Poisson and Negative Binomial Regression Models to Complex Sample Survey Data.....	331
10.	Survival Analysis of Event History Survey Data	339
10.1	Introduction	339
10.2	Basic Theory of Survival Analysis.....	339
10.2.1	Survey Measurement of Event History Data	339
10.2.2	Data for Event History Models	341
10.2.3	Important Notation and Definitions.....	342
10.2.4	Models for Survival Analysis.....	343
10.3	(Nonparametric) K–M Estimation of the Survivor Function.....	344
10.3.1	K–M Model Specification and Estimation.....	345
10.3.2	K–M Estimator: Evaluation and Interpretation	347
10.3.3	K–M Survival Analysis Example.....	347
10.4	The Cox Proportional Hazards (CPH) Model.....	351
10.4.1	CPH Model: Specification.....	351
10.4.2	CPH Model: Estimation Stage.....	352
10.4.3	CPH Model: Evaluation and Diagnostics.....	354
10.4.4	CPH Model: Interpretation and Presentation of Results.....	354
10.4.5	Example: Fitting a CPH Model to Complex Sample Survey Data.....	355
10.5	Discrete Time Survival Models.....	359
10.5.1	Discrete Time Logistic Model.....	359
10.5.2	Data Preparation for Discrete Time Survival Models	360
10.5.3	Discrete Time Models: Estimation Stage.....	363
10.5.4	Discrete Time Models: Evaluation and Interpretation.....	364
10.5.5	Fitting a Discrete Time Model to Complex Sample Survey Data.....	365
11.	Analysis of Longitudinal Complex Sample Survey Data	371
11.1	Introduction	371
11.2	Alternative Analytic Objectives with Longitudinal Survey Data.....	372
11.2.1	Objective 1: Descriptive Estimation at a Single Time Point	372
11.2.2	Objective 2: Estimation of Change across Two Waves	374

11.2.3	Objective 3: Trajectory Estimation Based on Three or More Waves.....	375
11.2.3.1	Approach 1: Weighted Multilevel Modeling....	375
11.2.3.2	Approach 2: Covariance Structure Modeling....	382
11.2.3.3	Approach 3: Weighted GEE Estimation.....	384
11.2.3.4	Approach 4: Multiple Imputation Analysis.....	387
11.2.3.5	Approach 5: Calibration Adjustment for Respondents with Complete Data	388
11.3	Alternative Longitudinal Analyses of the HRS Data	389
11.3.1	Example: Descriptive Estimation at a Single Wave.....	390
11.3.2	Example: Change across Two Waves	395
11.3.2.1	Accounting for Refreshment Samples When Estimating Mean Change.....	401
11.3.3	Example: Weighted Multilevel Modeling.....	402
11.3.3.1	Example: Veiga et al. (2014)	410
11.3.4	Example: Weighted GEE Analysis.....	416
11.4	Concluding Remarks	422
12.	Imputation of Missing Data: Practical Methods and Applications for Survey Analysts	427
12.1	Introduction	427
12.2	Important Missing Data Concepts	429
12.2.1	Sources and Types of Missing Data	429
12.2.2	Patterns of Item Missing Data in Surveys.....	430
12.2.3	Item Missing Data Mechanisms	431
12.2.4	Review of Strategies to Address Item Missing Data in Surveys.....	433
12.3	Factors to Consider in Choosing an Imputation Method	436
12.4	Multiple Imputation.....	439
12.4.1	Overview of MI and MI Phases.....	439
12.4.2	Models for Multiply Imputing Missing Data	440
12.4.2.1	Choosing the Variables to Include in the Imputation Model	441
12.4.2.2	Distributional Assumptions for the Imputation Model	444
12.4.3	Creating the MIs	444
12.4.3.1	Transforming the Imputation Problem to Monotonic Missing Data.....	445
12.4.3.2	Specifying an Explicit Multivariate Model and Applying Exact Bayesian Posterior Simulation Methods.....	445
12.4.3.3	SR or “Chained Regressions”	446
12.4.4	Estimation and Inference for Multiply Imputed Data....	448
12.4.4.1	Estimators for Population Parameters and Associated Variance Estimators.....	448

12.4.4.2	Model Evaluation and Inference	449
12.5	Fractional Imputation	451
12.5.1	Background	451
12.5.2	Creating the FIs	452
12.5.3	Estimation and Inference with Fractionally Imputed Data	454
12.5.4	FI Software	455
12.6	Application of MI and FI Methods to the NHANES 2011–2012 Data	456
12.6.1	Problem Definition	456
12.6.2	Imputation Models for the NHANES DBP Example	457
12.6.3	Imputation of the Item Missing Data	459
12.6.3.1	Multiple Imputation	459
12.6.3.2	FEFI: Hot Deck Method	461
12.6.4	Estimation and Inference	464
12.6.4.1	Multiple Imputation	464
12.6.4.2	FI Estimation and Inference	467
12.6.5	Comparison of Example Results from Complete Case Analysis, MI, and FEFI	468
13.	Advanced Topics in the Analysis of Survey Data	473
13.1	Introduction	473
13.2	Bayesian Analysis of Complex Sample Survey Data	474
13.3	GLMMs in Survey Data Analysis	478
13.3.1	Overview of GLMMs	478
13.3.2	GLMMs and Complex Sample Survey Data	481
13.3.3	Alternative Approaches to Fitting GLMMs to Survey Data: The PISA Example	485
13.4	Fitting Structural Equation Models to Complex Sample Survey Data	489
13.4.1	SEM Example: Analysis of ESS Data from Belgium	491
13.5	Small Area Estimation and Complex Sample Survey Data	498
13.6	Nonparametric Methods for Complex Sample Survey Data	499
References	501
Appendix A: Software Overview	519
Index	549