

Contents

1	An Overview of Sample Design and Weighting	1
1.1	Background and Terminology	1
1.2	Chapter Guide	7
Part I Designing Single-Stage Sample Surveys		
2	Project 1: Design a Single-Stage Personnel Survey	15
2.1	Specifications for the Study	15
2.2	Questions Posed by the Design Team	17
2.3	Preliminary Analyses	18
2.4	Documentation	21
2.5	Next Steps	23
3	Sample Design and Sample Size for Single-Stage Surveys	25
3.1	Determining a Sample Size for a Single-Stage Design	26
3.1.1	Criteria for Determining Sample Sizes	28
3.1.2	Simple Random Sampling	28
3.1.3	Stratified Simple Random Sampling	42
3.2	Finding Sample Sizes When Sampling with Varying Probabilities	51
3.2.1	Probability Proportional to Size Sampling	51
3.2.2	Regression Estimates of Totals	59
3.3	Other Methods of Sampling	63
3.4	Estimating Population Parameters from a Sample	64
3.5	Special Topics	68
3.5.1	Rare Characteristics	68
3.5.2	Domain Estimates	70
3.6	More Discussion of Design Effects	75
3.7	Software for Sample Selection	76
3.7.1	R Packages	77

3.7.2	SAS PROC SURVEYSELECT	81
3.7.3	Stata Commands	83
	Exercises	84
4	Power Calculations and Sample Size Determination	91
4.1	Terminology and One-Sample Tests	92
4.1.1	Characterizing Hypotheses and Tests	93
4.1.2	One-Sample Test	93
4.1.3	Use of Finite Population Corrections in Variances	95
4.2	Power in a One-Sample Test	97
4.2.1	1-Sided Tests	97
4.2.2	2-Sided Tests	101
4.3	Two-Sample Tests	105
4.3.1	Differences in Means	105
4.3.2	Partially Overlapping Samples	107
4.3.3	Differences in Proportions	108
4.3.4	Arcsine Square Root Transformation	110
4.3.5	Log-Odds Transformation	112
4.3.6	Special Case: Relative Risk	113
4.3.7	Special Case: Effect Sizes	113
4.4	R Power Functions	114
4.5	Power and Sample Size Calculations in SAS and Other Software	123
	Exercises	125
5	Mathematical Programming	129
5.1	Multicriteria Optimization	130
5.2	Microsoft Excel Solver	133
5.3	SAS PROC NLP	144
5.4	SAS PROC OPTMODEL	151
5.5	R Packages	155
5.5.1	R alabama Package	156
5.5.2	R Package nloptr	159
5.6	Allocation for Domain Estimation	162
5.7	Accounting for Problem Variations	165
	Exercises	165
6	Outcome Rates and Effect on Sample Size	169
6.1	Disposition Codes	170
6.2	Definitions of Outcome Rates	172
6.2.1	Location Rate	173
6.2.2	Contact Rate	173
6.2.3	Eligibility Rate	174
6.2.4	Cooperation Rate	175
6.2.5	Response Rate	176

6.3	Rates for Specialized Surveys	177
6.3.1	Probability-Based Panels	177
6.3.2	Nonprobability Surveys	178
6.4	Sample Units with Unknown AAPOR Classification	179
6.5	Weighted Versus Unweighted Rates	181
6.6	Accounting for Sample Losses in Determining Initial Sample Size	182
6.6.1	Sample Size Inflation Rates at Work	182
6.6.2	Sample Replicates	183
	Exercises	185
7	The Personnel Survey Design Project: One Solution	191
7.1	Overview of the Project	191
7.2	Formulate the Optimization Problem	192
7.2.1	Objective Function	192
7.2.2	Decision Variables	193
7.2.3	Optimization Parameters	193
7.2.4	Specified Survey Constraints	194
7.3	One Solution	195
7.3.1	Power Analyses	195
7.3.2	Optimization Results	197
7.4	Additional Sensitivity Analysis	200
7.5	Conclusion	201
	Part II Multistage Designs	
8	Project 2: Designing an Area Sample	205
8.1	Contents of the Sampling Report	206
8.2	Data Files and Other Information	207
9	Designing Multistage Samples	209
9.1	Types of PSUs	210
9.2	Basic Variance Results	211
9.2.1	Two-Stage Sampling	211
9.2.2	Nonlinear Estimators in Two-Stage Sampling	218
9.2.3	More General Two-Stage Designs	221
9.2.4	Three-Stage Sampling	224
9.3	Cost Functions and Optimal Allocations for Multistage Sampling	231
9.3.1	Two-Stage Sampling When Numbers of Sample PSUs and Elements per PSU Are Adjustable	231
9.3.2	Three-Stage Sampling When Sample Sizes Are Adjustable	235
9.3.3	Two- and Three-Stage Sampling with a Fixed Set of PSUs	237
9.3.4	Sample Selection When There Are Sample Losses	241

9.4	Estimating Measures of Homogeneity and Variance	242
	Components	242
9.4.1	Two-Stage Sampling	243
9.4.2	Three-Stage Sampling	247
9.4.3	Using Anticipated Variances	251
9.5	Stratification of PSUs	258
9.6	Identifying Certainties	259
	Exercises	260
10	Area Sampling	265
10.1	Census Geographic Units	266
10.2	Census Data and American Community Survey Data	269
10.3	Units at Different Stages of Sampling	270
10.3.1	Primary Sampling Units	271
10.3.2	Secondary Sampling Units	273
10.3.3	Ultimate Sampling Units	274
10.4	Examples of Area Probability Samples	275
10.4.1	Current Population Survey	275
10.4.2	National Survey on Drug Use and Health	278
10.4.3	Panel Arbeitsmarkt und Soziale Sicherung	280
10.5	Composite MOS for Areas	281
10.5.1	Designing the Sample from Scratch	282
10.5.2	Using the Composite MOS with an Existing PSU Sample	287
10.6	Effects of Population Change: The New Construction Issue	292
10.6.1	Option 1: Sample Building Permits	294
10.6.2	Option 2: Two-Phase Sample of Segments	296
10.6.3	Option 3: Half-Open Interval Technique	297
10.7	Special Address Lists	298
10.7.1	Oversampling Demographic Groups Using Enhanced Lists	300
	Exercises	303
11	The Area Sample Design: One Solution	307
11.1	Quality Control Checks on the Frame	308
11.2	Quality Control Checks on the Sample	310
11.3	Additional Considerations	312
Part III Survey Weights and Analyses		
12	Project 3: Weighting a Personnel Survey	317
12.1	Contents of the Weighting Report	319
12.2	Data Files and Other Information	319
12.3	Variable Values and Value Labels for the RESPSTAT Variable	320

13	Basic Steps in Weighting	321
13.1	Overview of Weighting	322
13.2	Theory of Weighting and Estimation	323
13.3	Base Weights	326
13.4	Adjustments for Unknown Eligibility	329
13.5	Adjustments for Nonresponse	331
13.5.1	Weighting Class Adjustments	334
13.5.2	Propensity Score Adjustments	336
13.5.3	Classification Algorithms—CART	354
13.5.4	Classification Algorithms—Random Forests	359
13.6	Collapsing Predefined Classes	361
13.7	Weighting for Multistage Designs	362
13.8	Next Steps in Weighting	364
	Exercises	364
14	Calibration and Other Uses of Auxiliary Data in Weighting	369
14.1	Weight Calibration	371
14.2	Poststratified and Raking Estimators	374
14.3	GREG and Calibration Estimation	382
14.3.1	Links Between Models, Sample Designs, and Estimators: Special Cases	384
14.3.2	More General Examples	386
14.4	Weight Variability	395
14.4.1	Quantifying the Variability	396
14.4.2	Methods to Limit Variability	404
14.5	Survey Weights in Model Fitting	414
	Exercises	415
15	Variance Estimation	421
15.1	Exact Methods	422
15.2	Linear Versus Nonlinear Estimators	425
15.3	Linearization Variance Estimation	426
15.3.1	Estimation Method	426
15.3.2	Confidence Intervals and Degrees of Freedom	430
15.3.3	Accounting for Non-negligible Sampling Fractions	433
15.3.4	Domain Estimation	435
15.3.5	Assumptions and Limitations	436
15.3.6	Special Cases: Multistage Sampling, Poststratification and Quantiles	437
15.3.7	Handling Multiple Weighting Steps with Linearization	443
15.4	Replication	443
15.4.1	Jackknife Replication	444
15.4.2	Balanced Repeated Replication	452

15.4.3	Bootstrap	456
15.4.4	Handling Multiple Weighting Steps with Replication	463
15.5	Combining PSUs or Strata	466
15.5.1	Combining to Reduce the Number of Replicates	466
15.5.2	How Many Groups and Which Strata and PSUs to Combine	469
15.5.3	Combining Strata in One-PSU-per-Stratum Designs	471
15.6	Handling Certainty PSUs	473
	Exercises	476
16	Weighting the Personnel Survey: One Solution	481
16.1	The Data Files	482
16.2	Base Weights	483
16.3	Disposition Codes and Mapping into Weighting Categories	484
16.4	Adjustment for Unknown Eligibility	487
16.5	Variables Available for Nonresponse Adjustment	488
16.6	Nonresponse Adjustments	490
16.6.1	Propensity Models	490
16.6.2	Regression Tree	491
16.7	Calibration to Population Counts	494
16.7.1	Identifying Variables to Use	496
16.7.2	Imputing for Missing Values	499
16.8	Writing Output Files	502
16.9	Example Tabulations	503
Part IV Other Topics		
17	Multiphase Designs	507
17.1	What Is a Multiphase Design?	508
17.2	Examples of Different Multiphase Designs	510
17.2.1	Double Sampling for Stratification	510
17.2.2	Nonrespondent Subsampling	513
17.2.3	Responsive Designs	519
17.2.4	General Multiphase Designs	523
17.3	Survey Weights	524
17.3.1	Base Weights	524
17.3.2	Analysis Weights	527
17.4	Estimation	531
17.4.1	Descriptive Point Estimation	531
17.4.2	Variance Estimation	533
17.4.3	Generalized Regression Estimator (GREG)	539
17.5	Design Choices	543
17.5.1	Multiphase Versus Single Phase	543

17.5.2	Sample Size Calculations	546
17.5.3	Response Rates	555
17.6	R Software	558
	Exercises	561
18	Nonprobability Sampling	565
18.1	Some History of Nonprobability Samples	567
18.2	Types of Nonprobability Samples	568
18.3	Potential Problems	571
18.4	Approaches to Inference	573
18.4.1	Quasi-randomization	573
18.4.2	Superpopulation Models	585
18.5	A Bayesian Approach	594
	Exercises	600
19	Process Control and Quality Measures	605
19.1	Design and Planning	606
19.2	Quality Control in Frame Creation and Sample Selection	609
19.3	Monitoring Data Collection	610
19.4	Performance Rates and Indicators	614
19.5	Data Editing	617
19.5.1	Editing Disposition Codes	618
19.5.2	Editing the Weighting Variables	620
19.6	Quality Control of Weighting Steps	620
19.7	Specification Writing and Programming	624
19.8	Project Documentation and Archiving	626
A	Notation Glossary	629
A.1	Sample Design and Sample Size for Single-Stage Surveys (Chap. 3)	629
A.1.1	General Notation	629
A.1.2	Single-Stage Sampling	630
A.1.3	Stratified Single-Stage Sampling	631
A.2	Designing Multistage Samples (Chap. 9)	632
A.2.1	Two-Stage Sampling	632
A.2.2	Variance Component Estimation in Two-Stage Sampling	634
A.2.3	Three-Stage Sampling	634
A.2.4	Variance Component Estimation in Three-Stage Sampling	636
A.2.5	Cost Functions in Two-Stage and Three-Stage Sampling	638
A.3	Basic Steps in Weighting (Chap. 13)	638
A.4	Calibration (Chap. 14)	639
A.5	Variance Estimation (Chap. 15)	640

A.5.1	Jackknife Variance Estimator	641
A.5.2	Balanced Repeated Replication (BRR) Variance Estimator	642
A.5.3	Bootstrap Variance Estimator	642
A.6	Multiphase Designs (Chap. 17)	643
A.6.1	Variance Estimation in a Two-Phase Design	644
B	Data Sets	647
B.1	Domainly2	647
B.2	Hospital	647
B.3	Labor	648
B.4	MDarea.pop	648
B.5	mibrfss	650
B.6	nhis	651
B.7	nhis.large	653
B.8	smho.N874	654
B.9	smho98	655
C	R Functions Used in This Book	657
C.1	R Overview	657
C.1.1	Documentation and Resources	657
C.1.2	Download a New Version of R	658
C.1.3	R Packages/Libraries	658
C.1.4	Updating R	659
C.1.5	Creating and Executing R Code	660
C.2	Author-Defined R Functions	660
	References	687
	Solutions to Selected Exercises	711
	Author Index	761
	Subject Index	767