

Sampling

Design and Analysis

Third Edition

Sharon L. Lohr



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

Preface	xiii
Symbols and Acronyms	xxi
1 Introduction	1
1.1 Guidance from Samples	1
1.2 Populations and Representative Samples	3
1.3 Selection Bias	6
1.3.1 Convenience Samples	6
1.3.2 Purposive or Judgment Samples	6
1.3.3 Self-Selected Samples	6
1.3.4 Undercoverage	8
1.3.5 Overcoverage	8
1.3.6 Nonresponse	9
1.3.7 What Good Are Samples with Selection Bias?	9
1.4 Measurement Error	10
1.5 Questionnaire Design	13
1.6 Sampling and Nonsampling Errors	17
1.7 Why Use Sampling?	18
1.7.1 Advantages of Taking a Census	19
1.7.2 Advantages of Taking a Sample Instead of a Census	19
1.8 Chapter Summary	20
1.9 Exercises	22
2 Simple Probability Samples	31
2.1 Types of Probability Samples	32
2.2 Framework for Probability Sampling	34
2.3 Simple Random Sampling	39
2.4 Sampling Weights	44
2.5 Confidence Intervals	46
2.6 Using Statistical Software to Analyze Survey Data	50
2.7 Determining the Sample Size	50
2.8 Systematic Sampling	55
2.9 Randomization Theory for Simple Random Sampling*	56
2.10 Model-Based Theory for Simple Random Sampling*	58
2.11 When Should a Simple Random Sample Be Used?	62
2.12 Chapter Summary	63
2.13 Exercises	66

3	Stratified Sampling	79
3.1	What Is Stratified Sampling?	79
3.2	Theory of Stratified Sampling	83
3.3	Sampling Weights in Stratified Random Sampling	87
3.4	Allocating Observations to Strata	89
3.4.1	Proportional Allocation	89
3.4.2	Optimal Allocation	91
3.4.3	Allocation for Specified Precision within Strata	93
3.4.4	Which Allocation to Use?	94
3.4.5	Determining the Total Sample Size	96
3.5	Defining Strata	96
3.6	Model-Based Theory for Stratified Sampling*	99
3.7	Chapter Summary	100
3.8	Exercises	101
4	Ratio and Regression Estimation	121
4.1	Ratio Estimation in Simple Random Sampling	121
4.1.1	Why Use Ratio Estimation?	122
4.1.2	Bias and Mean Squared Error of Ratio Estimators	125
4.1.3	Ratio Estimation with Proportions	132
4.1.4	Ratio Estimation Using Weight Adjustments	134
4.1.5	Advantages of Ratio Estimation	135
4.2	Regression Estimation in Simple Random Sampling	135
4.3	Estimation in Domains	139
4.4	Poststratification	142
4.5	Ratio Estimation with Stratified Sampling	145
4.6	Model-Based Theory for Ratio and Regression Estimation*	147
4.6.1	A Model for Ratio Estimation	148
4.6.2	A Model for Regression Estimation	151
4.6.3	Differences between Model-Based and Design-Based Estimators	152
4.7	Chapter Summary	154
4.8	Exercises	155
5	Cluster Sampling with Equal Probabilities	167
5.1	Notation for Cluster Sampling	171
5.2	One-Stage Cluster Sampling	172
5.2.1	Clusters of Equal Sizes: Estimation	172
5.2.2	Clusters of Equal Sizes: Theory	174
5.2.3	Clusters of Unequal Sizes	179
5.3	Two-Stage Cluster Sampling	182
5.4	Designing a Cluster Sample	192
5.4.1	Choosing the psu Size	193
5.4.2	Choosing Subsampling Sizes	194
5.4.3	Choosing the Sample Size (Number of psus)	196
5.5	Systematic Sampling	197
5.6	Model-Based Theory for Cluster Sampling*	200
5.6.1	Estimation Using Models	202
5.6.2	Design Using Models	205
5.7	Chapter Summary	205
5.8	Exercises	207

6	Sampling with Unequal Probabilities	219
6.1	Sampling One Primary Sampling Unit	221
6.2	One-Stage Sampling with Replacement	224
6.2.1	Selecting Primary Sampling Units	224
6.2.2	Theory of Estimation	226
6.2.3	Designing the Selection Probabilities	229
6.2.4	Weights in Unequal-Probability Sampling with Replacement	230
6.3	Two-Stage Sampling with Replacement	230
6.4	Unequal-Probability Sampling without Replacement	233
6.4.1	The Horvitz–Thompson Estimator for One-Stage Sampling	235
6.4.2	Selecting the psus	239
6.4.3	The Horvitz–Thompson Estimator for Two-Stage Sampling	239
6.4.4	Weights in Unequal-Probability Samples	240
6.5	Examples of Unequal-Probability Samples	243
6.6	Randomization Theory Results and Proofs*	247
6.7	Model-Based Inference with Unequal-Probability Samples*	254
6.8	Chapter Summary	256
6.9	Exercises	258
7	Complex Surveys	273
7.1	Assembling Design Components	273
7.1.1	Building Blocks for Surveys	273
7.1.2	Ratio Estimation in Complex Surveys	275
7.1.3	Simplicity in Survey Design	276
7.2	Sampling Weights	276
7.2.1	Constructing Sampling Weights	276
7.2.2	Self-Weighting and Non-Self-Weighting Samples	279
7.3	Estimating Distribution Functions and Quantiles	280
7.4	Design Effects	286
7.5	The National Health and Nutrition Examination Survey	288
7.6	Graphing Data from a Complex Survey	291
7.6.1	Univariate Plots	292
7.6.2	Bivariate Plots	295
7.7	Chapter Summary	301
7.8	Exercises	303
8	Nonresponse	311
8.1	Effects of Ignoring Nonresponse	312
8.2	Designing Surveys to Reduce Nonresponse	314
8.3	Two-Phase Sampling	319
8.4	Response Propensities and Mechanisms for Nonresponse	320
8.4.1	Auxiliary Information for Treating Nonresponse	320
8.4.2	Methods to Adjust for Nonresponse	320
8.4.3	Response Propensities	321
8.4.4	Types of Missing Data	321
8.5	Adjusting Weights for Nonresponse	323
8.5.1	Weighting Class Adjustments	324
8.5.2	Regression Models for Response Propensities	328
8.6	Poststratification	329
8.6.1	Poststratification Using Weights	330
8.6.2	Raking Adjustments	331

8.6.3	Steps for Constructing Final Survey Weights	333
8.6.4	Advantages and Disadvantages of Weighting Adjustments	334
8.7	Imputation	335
8.7.1	Deductive Imputation	335
8.7.2	Cell Mean Imputation	336
8.7.3	Hot-Deck Imputation	337
8.7.4	Regression Imputation and Chained Equations	338
8.7.5	Imputation from Another Data Source	338
8.7.6	Multiple Imputation	339
8.7.7	Advantages and Disadvantages of Imputation	339
8.8	Response Rates and Nonresponse Bias Assessments	340
8.8.1	Calculating and Reporting Response Rates	340
8.8.2	What Is an Acceptable Response Rate?	342
8.8.3	Nonresponse Bias Assessments	343
8.9	Chapter Summary	346
8.10	Exercises	348
9	Variance Estimation in Complex Surveys	359
9.1	Linearization (Taylor Series) Methods	359
9.2	Random Group Methods	363
9.2.1	Replicating the Survey Design	363
9.2.2	Dividing the Sample into Random Groups	365
9.3	Resampling and Replication Methods	367
9.3.1	Balanced Repeated Replication (BRR)	367
9.3.2	Jackknife	373
9.3.3	Bootstrap	375
9.3.4	Creating and Using Replicate Weights	377
9.4	Generalized Variance Functions	379
9.5	Confidence Intervals	381
9.5.1	Confidence Intervals for Smooth Functions of Population Totals	381
9.5.2	Confidence Intervals for Population Quantiles	382
9.6	Chapter Summary	384
9.7	Exercises	386
10	Categorical Data Analysis in Complex Surveys	395
10.1	Chi-Square Tests with Multinomial Sampling	395
10.1.1	Testing Independence of Factors	397
10.1.2	Testing Homogeneity of Proportions	398
10.1.3	Testing Goodness of Fit	398
10.2	Effects of Survey Design on Chi-Square Tests	399
10.2.1	Contingency Tables for Data from Complex Surveys	400
10.2.2	Effects on Hypothesis Tests and Confidence Intervals	401
10.3	Corrections to Chi-Square Tests	403
10.3.1	Wald Tests	403
10.3.2	Rao–Scott Tests	405
10.3.3	Model-Based Methods for Chi-Square Tests	407
10.4	Loglinear Models	408
10.4.1	Loglinear Models with Multinomial Sampling	409
10.4.2	Loglinear Models in a Complex Survey	410
10.5	Chapter Summary	411
10.6	Exercises	412

11 Regression with Complex Survey Data	419
11.1 Model-Based Regression in Simple Random Samples	420
11.2 Regression with Complex Survey Data	423
11.2.1 Point Estimation	424
11.2.2 Standard Errors	427
11.2.3 Multiple Regression	430
11.2.4 Regression Using Weights versus Weighted Least Squares	432
11.3 Using Regression to Compare Domain Means	433
11.4 Interpreting Regression Coefficients from Survey Data	435
11.4.1 Purposes of Regression Analyses	435
11.4.2 Model-Based and Design-Based Inference	436
11.4.3 Survey Weights and Regression	437
11.4.4 Survey Design and Standard Errors	438
11.4.5 Mixed Models for Cluster Samples	439
11.5 Logistic Regression	440
11.6 Calibration to Population Totals	442
11.7 Chapter Summary	446
11.8 Exercises	448
12 Two-Phase Sampling	457
12.1 Theory for Two-Phase Sampling	459
12.2 Two-Phase Sampling with Stratification	461
12.3 Ratio and Regression Estimation in Two-Phase Samples	464
12.3.1 Two-Phase Sampling with Ratio Estimation	464
12.3.2 Generalized Regression Estimation in Two-Phase Sampling	466
12.4 Jackknife Variance Estimation for Two-Phase Sampling	467
12.5 Designing a Two-Phase Sample	469
12.5.1 Two-Phase Sampling with Stratification	469
12.5.2 Optimal Allocation for Ratio Estimation	471
12.6 Chapter Summary	471
12.7 Exercises	472
13 Estimating the Size of a Population	483
13.1 Capture–Recapture Estimation	483
13.1.1 Contingency Tables for Capture–Recapture Experiments	484
13.1.2 Confidence Intervals for N	485
13.1.3 Using Capture–Recapture on Lists	486
13.2 Multiple Recapture Estimation	488
13.3 Chapter Summary	491
13.4 Exercises	492
14 Rare Populations and Small Area Estimation	499
14.1 Sampling Rare Populations	500
14.1.1 Stratified Sampling with Disproportional Allocation	500
14.1.2 Two-Phase Sampling	501
14.1.3 Unequal-Probability Sampling	501
14.1.4 Multiple Frame Surveys	502
14.1.5 Network or Multiplicity Sampling	504
14.1.6 Snowball Sampling	505
14.1.7 Sequential Sampling	506
14.2 Small Area Estimation	506

14.2.1	Direct Estimators	507
14.2.2	Synthetic and Composite Estimators	508
14.2.3	Model-Based Estimators	509
14.3	Chapter Summary	510
14.4	Exercises	512
15	Nonprobability Samples	517
15.1	Types of Nonprobability Samples	518
15.1.1	Administrative Records	518
15.1.2	Quota Samples	519
15.1.3	Judgment Samples	522
15.1.4	Convenience Samples	523
15.2	Selection Bias and Mean Squared Error	524
15.2.1	Random Variables Describing Participation in a Sample	525
15.2.2	Bias and Mean Squared Error of a Sample Mean	528
15.3	Reducing Bias of Estimates from Nonprobability Samples	531
15.3.1	Weighting	531
15.3.2	Estimate the Values of the Missing Units	536
15.3.3	Measures of Uncertainty for Nonprobability Samples	537
15.4	Nonprobability versus Low-Response Probability Samples	539
15.5	Chapter Summary	542
15.6	Exercises	544
16	Survey Quality	557
16.1	Coverage Error	559
16.1.1	Measuring Coverage and Coverage Bias	559
16.1.2	Coverage and Survey Mode	560
16.1.3	Improving Coverage	562
16.2	Nonresponse Error	562
16.3	Measurement Error	564
16.3.1	Measuring and Modeling Measurement Error	565
16.3.2	Reducing Measurement Error	567
16.3.3	Sensitive Questions	568
16.3.4	Randomized Response	568
16.4	Processing Error	570
16.5	Total Survey Quality	571
16.6	Chapter Summary	573
16.7	Exercises	575
A	Probability Concepts Used in Sampling	579
A.1	Probability	579
A.1.1	Simple Random Sampling with Replacement	580
A.1.2	Simple Random Sampling without Replacement	581
A.2	Random Variables and Expected Value	582
A.3	Conditional Probability	585
A.4	Conditional Expectation	587
A.5	Exercises	591
	Bibliography	593
	Index	641