

Contents

Preface	xv
1. Introduction to Prediction Theory	1
1.1 Sampling Theory and the Rest of Statistics,	1
1.2 Prediction Theory,	2
1.3 Probability Sampling Theory,	7
1.3.1 Techniques Used in Probability Sampling,	9
1.3.2 Some Mathematical Details,	11
1.4 Which Approach to Use?,	14
1.5 Why Use Random Sampling?,	19
Exercises,	22
2. Prediction Theory Under the General Linear Model	25
2.1 Definitions and a Simple Example,	26
2.2 General Prediction Theorem,	29
2.3 <i>BLU</i> Predictor Under Some Simple Models,	31
2.4 Unit Weights,	32
2.5* Asymptotic Normality of the <i>BLU</i> Predictor,	34
2.6 Ignorable and Nonignorable Sample Selection Methods,	35
2.6.1 Examples,	36
2.6.2* Formal Definition of Ignorable Selection,	36
2.7* Comparisons with Design-Based Regression Estimation,	40
Exercises,	44
3. Bias-Robustness	49
3.1 Design and Bias,	50
3.2 Polynomial Framework and Balanced Samples,	53

3.2.1	Expansion Estimator and Balanced Samples,	54
3.2.2*	Order of the Bias of the Expansion Estimator,	55
3.2.3	Ratio Estimator and Balanced Samples,	55
3.2.4	Bias-Robust Strategies,	56
3.2.5	Simulation Study to Illustrate Conditional Biases and Mean Squared Errors,	57
3.2.6	Balance and Multiple Y Variables,	59
3.3	Weighted Balance,	61
3.3.1	Elementary Estimators Unbiased Under Weighted Balance,	62
3.3.2	BLU Estimators and Weighted Balance,	63
3.4	Methods of Selecting Balanced Samples,	65
3.4.1	Simple Random Sampling,	66
3.4.2	Systematic Equal Probability Sampling,	67
3.4.3	Stratification Based on the Auxiliary,	68
3.4.4	Restricted Random Sampling,	71
3.4.5	Sampling for Weighted Balance,	72
3.4.6	Restricted pps Sampling,	74
3.4.7	Partial Balancing,	77
3.5	Simulation Study of Weighted Balance,	77
3.5.1	Results Using the Hospitals Population,	78
3.5.2	Interaction of Model Specification with Sample Configuration,	82
3.6	Summary,	85
3.7	Robustness and Design-Based Inference,	85
	Exercises,	90
4.	Robustness and Efficiency	95
4.1	Introduction,	95
4.2	General Linear Model,	96
4.2.1	BLU Predictor Under the General Linear Model with Diagonal Variance Matrix,	97
4.2.2	Examples of Minimal Models,	101
4.3	Comparisons Using an Artificial Population,	103
4.3.1	Results for Probability Proportional to x Sampling and x -Balance,	105
4.3.2	Results for Probability Proportional to $x^{1/2}$ Sampling and $x^{1/2}$ -Balance,	107
4.3.3	Results for Equal Probability Systematic Sampling and Simple Balance,	110

4.4	Sample Size Determination,	114
4.5	Summary and Perspective,	118
4.6*	Remarks on Design-Based Inference,	119
	Exercises,	122
5.	Variance Estimation	125
5.1	Examples of Robust Variance Estimation,	127
5.1.1	Homoscedastic Through the Origin Model,	127
5.1.2	Variance Estimators for the Ratio Estimator,	130
5.2	Variance Estimation of a Linear Function of the Parameter,	134
5.3	Sandwich Estimator of Variance,	135
5.3.1	Consistency of v_R ,	136
5.3.2	Some Comments on the Requirements for Consistency of the Sandwich Estimator,	138
5.4	Variants on the Basic Robust Variance Estimator,	139
5.4.1	Internal and External Adjustments to the Sandwich Estimator,	139
5.4.2	Jackknife Variance Estimator,	141
5.5	Variance Estimation for Totals,	144
5.5.1	Some Simple Examples,	146
5.5.2	Effect of a Large Sampling Fraction,	149
5.6	Misspecification of the Regression Component,	149
5.7	Hidden Regression Components,	152
5.7.1	Some Artificial Examples,	153
5.7.2	Counties 70 Population,	157
5.7.3	Lurking Discrete Skewed Variables,	161
5.8	Comparisons with Design-Based Variance Estimation,	161
	Exercises,	164
6.	Stratified Populations	167
6.1	Stratification with Homogeneous Subpopulations,	168
6.2	Stratified Linear Model and Weighted Balanced Samples,	173
6.2.1	Optimal Allocation for Stratified Balanced Sampling,	174
6.2.2	Case of a Single Model for the Population,	175
6.2.3	Case of a Single Auxiliary Variable,	177
6.3	Sampling Fractions Greater Than 1,	179
6.4	Allocation to Strata in More Complicated Cases,	181
6.4.1	Contrasts Between Strata,	182
6.4.2	More Than One Target Variable,	184

6.5	Two Traditional Topics,	186	
6.5.1	Efficiency of the Separate Ratio Estimator,	186	
6.5.2	Formation of Strata,	189	
6.6	Some Empirical Results on Strata Formation,	192	
6.7	Variance Estimation in Stratified Populations,	197	
6.8	Stratification in Design-Based Theory,	200	
	Exercises,	205	
7.	Models with Qualitative Auxiliaries		211
7.1	Simple Example,	211	
7.2	Factors, Levels, and Effects,	213	
7.3	Generalized Inverses,	214	
7.4	Estimating Linear Combinations of the Y 's,	217	
7.5	One-Way Classification,	220	
7.6	Two-Way Nested Classification,	223	
7.7	Two-Way Classification Without Interaction,	225	
7.8	Two-Way Classification With Interaction,	226	
7.9	Combining Qualitative and Quantitative Auxiliaries,	231	
7.9.1	General Covariance Model,	232	
7.9.2	One-Way Classification with a Single Covariate,	234	
7.9.3	Examples,	235	
7.10	Variance Estimation,	237	
7.10.1	Basic Robust Alternatives,	238	
7.10.2	Jackknife Variance Estimator,	240	
	Exercises,	242	
8.	Clustered Populations		246
8.1	Intracluster Correlation Model for a Clustered Population,	247	
8.1.1	Discussion of the Common Mean Model,	248	
8.1.2	Simple Sample Designs,	249	
8.2	Class of Unbiased Estimators Under the Common Mean Model,	250	
8.2.1	One-Stage Cluster Sampling,	252	
8.2.2	BLU Predictor,	254	
8.2.3	Variance Component Model,	256	
8.3	Estimation of Parameters in the Constant Parameter Model,	257	
8.3.1	ANOVA Estimators,	257	
8.3.2*	Maximum Likelihood Estimators,	258	

8.3.3*	Lower Bound on the Intracluster Correlation,	261
8.4	Simulation Study for the Common Mean Model,	261
8.5	Biases of Common Mean Estimators Under a More General Model,	264
8.6	Estimation Under a More General Regression Model,	266
8.7*	Robustness and Optimality,	271
8.8	Efficient Design for the Common Mean Model,	275
8.8.1	Choosing the Set of Sample Clusters for the <i>BLU</i> Estimator,	275
8.8.2	Choosing the Set of Sample Clusters for the Unbiased Estimators,	276
8.8.3	Optimal Allocation of Second-Stage Units Given a Fixed Set of First-Stage Sample Units,	277
8.8.4	Optimal First and Second-Stage Allocation Considering Costs,	279
8.9*	Estimation When Cluster Sizes Are Unknown,	283
8.10	Two-Stage Sampling in Design-Based Practice, Exercises,	287 290
9.	Robust Variance Estimation in Two-Stage Cluster Sampling	296
9.1	Common Mean Model and a General Class of Variance Estimators,	296
9.2*	Other Variance Estimators,	299
9.2.1*	Non-Robust ANOVA Estimator,	300
9.2.2*	Alternative Robust Variance Estimators,	301
9.3	Examples of the Variance Estimators,	305
9.3.1	Ratio Estimator,	305
9.3.2	Mean of Ratios Estimator,	307
9.3.3	Numerical Illustrations,	307
9.4*	Variance Estimation for an Estimated Total—Unknown Cluster Sizes,	309
9.5	Regression Estimator,	310
9.5.1	Sandwich Variance Estimator,	311
9.5.2	Adjustments to the Sandwich Estimator,	313
9.5.3	Jackknife Estimator,	314
9.6	Comparisons of Variance Estimators in a Simulation Study, Exercises,	318 320
10.	Alternative Variance Estimation Methods	323
10.1	Estimating the Variance of Estimators of Nonlinear	

- Functions, 323
 - 10.1.1 Variance Estimation for a Ratio of Estimated Totals, 326
 - 10.1.2 Jackknife and Nonlinear Functions, 329
- 10.2 Balanced Half-Sample Variance Estimation, 330
 - 10.2.1 Application to the Stratified Expansion Estimator, 331
 - 10.2.2 Orthogonal Arrays, 333
 - 10.2.3 Extension to Nonlinear Functions, 334
 - 10.2.4 Two-Stage Sampling, 336
 - 10.2.5 Other Forms of the *BHS* Variance Estimator, 340
 - 10.2.6 Partially Balanced Half-Sampling, 341
 - 10.2.7 Design-based Properties, 344
- 10.3 Generalized Variance Functions, 344
 - 10.3.1 Some Theory for *GVF*'s, 345
 - 10.3.2 Estimation of *GVF* Parameters, 347
- Exercises, 348

11. Special Topics and Open Questions

351

- 11.1 Estimation in the Presence of Outliers, 352
 - 11.1.1 Gross Error Model, 354
 - 11.1.2 Simple Regression Model, 361
 - 11.1.3 Areas for Research, 366
- 11.2 Nonlinear Models, 367
 - 11.2.1 Model for Bernoulli Random Variables, 370
 - 11.2.2 Areas for Research, 372
- 11.3 Nonparametric Estimation of Totals, 372
 - 11.3.1 Nonparametric Regression for Totals, 373
 - 11.3.2 Nonparametric Calibration Estimation, 375
- 11.4 Distribution Function and Quantile Estimation, 377
 - 11.4.1 Estimation Under Homogeneous and Stratified Models, 380
 - 11.4.2 Estimation of $F_N(\cdot)$ Under a Regression Model, 382
 - 11.4.3 Large Sample Properties, 387
 - 11.4.4 The Effect of Model Misspecification, 388
 - 11.4.5 Design-Based Approaches, 390
 - 11.4.6 Nonparametric Regression-Based Estimators, 391
 - 11.4.7 Some Open Questions, 393
- 11.5 Small Area Estimation, 394

11.5.1	Estimation When Cell Means Are Unrelated,	395
11.5.2	Cell Means Determined by Class but Uncorrelated,	396
11.5.3	Synthetic and Composite Estimators,	400
11.5.4	Using Auxiliary Data,	402
11.5.5	Auxiliary Data at the Cell Level,	404
11.5.6	Need for a Small Area Estimation Canon,	406
	Exercises,	407
Appendix A. Some Basic Tools		413
A.1	Orders of Magnitude, $O(\cdot)$ and $o(\cdot)$,	413
A.2	Convergence in Probability and in Distribution,	413
A.3	Probabilistic Orders of Magnitude, $O_p(\cdot)$ and $o_p(\cdot)$,	414
A.4	Chebyshev's Inequality,	414
A.5	Cauchy-Schwarz Inequality,	414
A.6	Slutsky's Theorem,	414
A.7	Taylor's Theorem,	415
	A.7.1 Univariate Version,	415
	A.7.2 Multivariate Version,	415
A.8	Central Limit Theorems for Independent, not Identically Distributed Random Variables,	415
A.9	Central Limit Theorem for Simple Random Sampling,	416
A.10	Generalized Inverse of a Matrix,	417
Appendix B. Datasets		420
B.1	Cancer Population,	422
B.2	Hospitals Population,	424
B.3	Counties 60 Population,	428
B.4	Counties 70 Population,	431
B.5	Labor Force Population,	434
B.6	Third Grade Population,	444
Appendix C. S-PLUS Functions		446
Bibliography		460
Answers to Select Exercises		474
Author Index		484
Subject Index		488