Contents

1 Introduction 1

1.1 The design-based paradigm 2
1.2 Text content and orientation 4
1.3 What distinguishes this text? 5
1.4 Recommendations for instructors 6
1.5 Sampling theory: A brief history 7

2 Basic concepts 11

2.1 Terminology 12
2.2 Components of a sampling strategy 13
2.3 Selection methods 14
2.4 Properties of estimators 15
2.5 Sampling distribution of an estimator 18
2.6 Judgment sampling versus random sampling 19

3 Equal probability sampling 23

3.1 Without replacement sampling 23
3.1.1 Estimation of the population mean, proportion, and total 25
3.1.2 Sampling variance 27
3.1.3 Estimation of sampling variance 30
3.1.4 Bernoulli sampling 31
3.2 With replacement sampling 32
3.2.1 Estimation of the population mean, proportion, and total 34
3.2.2 Sampling variance and variance estimation 35
3.2.3 Rao–Blackwell theorem 35
3.3 Relative performance of alternative sampling strategies 36
3.3.1 Measures of relative performance 36
3.3.2 An example: SRS/mean-per-unit estimation versus SWR 37
3.4 Sample size to achieve desired level of precision 38
3.4.1 Approximate normality of sampling distributions 38
3.4.2 Confidence interval construction 40
3.4.3 Sample size determination 40
3.5 Nonresponse and oversampling 42

3.6 Sampling in R 44
3.6.1 SRS and SWR 44
3.6.2 Sample spaces 44
3.7 Chapter comments 45
Problems 45

4 Systematic sampling 48
4.1 Linear systematic sampling 50
4.1.1 N/k is integer-valued 50
4.1.2 N/k is not integer-valued 52
4.2 Selection methods that guarantee fixed n 53
4.2.1 Circular systematic sampling 53
4.2.2 Fractional interval random start 54
4.3 Estimation of sampling variance 57
4.3.1 Biased estimation 57
4.3.2 Unbiased estimation 58
4.4 Unpredictable trend in sampling variance with n 62
4.5 Warning: Pathological settings 62
4.6 Nonresponse and oversampling 63
4.7 Chapter comments 64
Problems 65

5 Stratified sampling 68
5.1 Estimation of the population mean 69
5.1.1 Expected value 69
5.1.2 Sampling variance 70
5.1.3 Numerical examples 70
5.2 Estimation of the population proportion 72
5.3 Estimation of the population total 72
5.4 Estimation of sampling variance 73
5.5 Allocation of the sample across strata 74
5.5.1 Optimal allocation: Graphical analysis 75
5.5.2 Optimal allocation: Analytical analysis 75
5.5.3 Comments on optimal allocation 78
5.6 Sample size determination 79
5.7 Relative efficiency 80
5.7.1 Proportional allocation 80
5.7.2 Estimation of finite population variance 81
5.8 Effective degrees of freedom 83
5.9 Post-stratification 84
5.9.1 Unconditional sampling variance 85
5.9.2 Conditional sampling variance 86
5.10 Chapter comments 87
Problems 88

6 Single-stage cluster sampling: Clusters of equal size 92
6.1 Estimation of the population mean 93
6.2 Sampling variance 94
6.2.1 ANOVA/mean squares approach 95
6.2.2 Intracluster correlation approach 96
6.3 Estimation of the population total and proportion 97
6.4 Estimation of sampling variance 98
6.5 Estimation of finite population variance 99
6.6 Sample size determination 100
6.7 Relative efficiency 100
6.8 Chapter comments 101
Problems 102

7 Ratio and regression estimation 104
7.1 Estimation of the mean and total 105
7.1.1 Graphical representation 106
7.1.2 Sample space illustration 107
7.1.3 Bias 109
7.1.4 Sampling variance 110
7.1.5 Estimation of sampling variance 112
7.1.6 Sample size determination 112
7.1.7 Relative efficiency 113
7.2 Ratio estimation of a proportion 115
7.3 Ratio estimation with stratified sampling 117
7.3.1 Combined estimator 117
7.3.2 Separate estimator 117
7.3.3 Choosing between combined and separate estimators 118
7.4 A model-based perspective 118
7.4.1 Estimation of model parameters 119
7.4.2 Prediction of population parameters 124
7.4.3 Prediction error 126
7.4.4 Prediction variance estimators 128
7.5 Monte Carlo performance evaluation 129
7.6 Mark-recapture estimation 132
7.7 Chapter comments 133
Problems 135

8 Unequal probability sampling 140
8.1 Unbiased ratio estimator 141
8.2 Sampling with replacement 143
8.2.1 Hansen–Hurwitz estimator 143
8.2.2 Unbiasedness 144
8.2.3 Sampling variance and variance estimation 144
8.3 Sampling without replacement 145

8.3.1 Horvitz–Thompson estimator 146
8.3.2 Unbiasedness 147
8.3.3 Sampling variance and variance estimation 148
8.3.4 Alternative selection methods 150
8.3.5 Strategy performance comparisons 152
8.3.6 Survey cost comparisons 155
8.4 Sampling distribution 157
8.5 Systematic sampling 159
8.6 Generality of Horvitz–Thompson estimation 161
8.7 Generalized Horvitz–Thompson estimation 162
8.7.1 Variance, covariance, and correlation estimators 163
8.7.2 Mean-per-unit, ratio, and regression estimators 163
8.7.3 Performance of generalized Horvitz–Thompson estimators 163
8.8 Poisson sampling 164
8.9 Nonresponse and oversampling 165
8.9.1 Hansen–Hurwitz estimator 165
8.9.2 Horvitz–Thompson estimator 166
8.10 Chapter comments 167
Problems 168

9 Multi-stage sampling 173
9.1 Two-stage sampling: Clusters of equal size 174
9.1.1 Estimation of the population mean 175
9.1.2 Expectation 176
9.1.3 Sampling variance and its estimation 179
9.1.4 Optimal allocation 181
9.1.5 Net relative efficiency 184
9.2 Two-stage sampling: Clusters of unequal size 185
9.2.1 Single-stage cluster sampling 185
9.2.2 Two-stage estimation of the population mean and total 187
9.2.3 Sampling variance and its estimation 187
9.2.4 Optimal allocation 194
9.3 Chapter comments 195
9.3.1 Generality of the multi-stage framework 195
9.3.2 Taking advantage of ecological understanding 196
9.3.3 Implications for large-scale natural resource surveys 196
Problems 197

10 Multi-phase sampling 200
10.1 Two-phase estimation of the population mean and total 201
10.1.1 Estimators 202
10.1.2 Sampling variance and its estimation 202
10.1.3 Sample space illustration 203
10.1.4 Optimal allocation 205
10.1.5 Net relative efficiency 207

10.2 Two-phase ratio estimation of a proportion 209
10.3 Two-phase sampling with stratification 212
10.3.1 Estimation of the population mean and total 212
10.3.2 Sampling variance and its estimation 212
10.3.3 Optimal allocation 213
10.3.4 Net relative efficiency 215
10.4 Chapter comments 215
Problems 216

11 Adaptive sampling 219
11.1 Adaptive cluster sampling 220
11.1.1 Basic scheme 220
11.1.2 Definitions 221
11.1.3 Inclusion probabilities and expected sample size 222
11.1.4 Estimators and relative efficiency 224
11.2 Other adaptive sampling designs 231
11.2.1 Single-stage strip and systematic designs 231
11.2.2 Two-stage complete allocation cluster sampling 232
11.3 Chapter comments 236
Problems 238

12 Spatially balanced sampling 240
12.1 Introduction 240
12.2 Finite populations 244
12.2.1 Generalized random tessellation stratified sampling 244
12.2.2 Balanced acceptance sampling 251
12.2.3 Estimation 260
12.3 Infinite populations 263
12.3.1 Generalized random tessellation stratified sampling 263
12.3.2 Balanced acceptance sampling 264
12.3.3 Estimation 265
12.4 Chapter comments 266

13 Sampling through time 269
13.1 Sampling on two occasions 270
13.1.1 Design 1: Full retention of units across occasions 271
13.1.2 Design 2: Independent SRS on each occasion 272
13.1.3 Comparison of full retention and independent SRS designs 272
13.1.4 Design 3: Partial retention/partial replacement 273
13.2 Monitoring design 275
13.2.1 Membership design 276
13.2.2 Revisit design 277
13.3 Estimation of status and trend 282
13.3.1 Design-based estimation 283
13.3.2 Estimators for some specific designs 284

13.4 Sample size determination 286
13.5 Dual frame sampling 288
13.6 Chapter comments 290

A Mathematical foundations 295

A.1 Counting techniques 296
A.1.1 Permutations 296
A.1.2 *k*-Permutations 296
A.1.3 Combinations 298
A.1.4 Partitions 298
A.2 Basic principles of probability theory 299
A.2.1 Random experiment 299
A.2.2 Sample space 300
A.2.3 Outcome probability 300
A.2.4 Event 301
A.2.5 Event relations 301
A.2.6 Union and intersection 302
A.2.7 Conditional probability 302
A.2.8 Other probability relations 302
A.3 Discrete random variables 303
A.3.1 Definition 303
A.3.2 Probability distributions 303
A.3.3 Probability relations 305
A.3.4 Expectation 305
A.3.5 Variance and coefficient of variation 307
A.3.6 Covariance and correlation 309
A.3.7 Total expectation and variance 310
A.3.8 Indicator variables 311
A.4 Key discrete probability distributions 312
A.4.1 Uniform 312
A.4.2 Bernoulli 313
A.4.3 Multinoulli 313
A.4.4 Binomial 314
A.4.5 Multinomial 315
A.4.6 Hypergeometric 316
A.4.7 Multivariate hypergeometric 317
A.5 Population variables 317
A.5.1 Definition 317
A.5.2 Population parameters 318
A.6 Probability sampling 319
A.6.1 Sampling experiment 319
A.6.2 Sampling designs 320
A.6.3 Inclusion probabilities 320
A.6.4 Inclusion indicator variables 320

A.7 Estimators 321
A.7.1 Definition 321
A.7.2 Sampling distribution 322
A.7.3 Statistical properties 323
A.7.4 Confidence intervals 324
A.8 Delta method 325
A.9 Lagrange multipliers 327

References 329
Index 337