Sampling Techniques

third edition

WILLIAM G. COCHRAN

Professor of Statistics, Emeritus
Harvard University

John Wiley & Sons
New York · Santa Barbara · London · Sydney · Toronto

Contents

CHAPTER PAGE

1 INTRODUCTION 1

1.1 Advantages of the Sampling Method 1
1.2 Some Uses of Sample Surveys 2
1.3 The Principal Steps in a Sample Survey 4
1.4 The Role of Sampling Theory 8
1.5 Probability Sampling 9
1.6 Alternatives to Probability Sampling 10
1.7 Use of the Normal Distribution 11
1.8 Bias and Its Effects 12
1.9 The Mean Square Error 15

Exercises 16

CHAPTER

2 SIMPLE RANDOM SAMPLING 18

2.1 Simple Random Sampling 18
2.2 Selection of a Simple Random Sample 19
2.3 Definitions and Notation 20
2.4 Properties of the Estimates 21
2.5 Variances of the Estimates 23
2.6 The Finite Population Correction 24
2.7 Estimation of the Standard Error from a Sample 25
2.8 Confidence Limits 27
2.9 An Alternative Method of Proof 28
2.10 Random Sampling with Replacement 29
2.11 Estimation of a Ratio 30
2.12 Estimates of Means Over Subpopulations 34
2.13 Estimates of Totals Over Subpopulations 35
2.14 Comparisons Between Domain Means 39
2.15 Validity of the Normal Approximation 39
2.16 Linear Estimators of the Population Mean 44

Exercises 45

CHAPTER

3 SAMPLING PROPORTIONS AND PERCENTAGES 50

3.1 Qualitative Characteristics 50
3.2 Variances of the Sample Estimates 50
3.3 The Effect of P on the Standard Errors 53
3.4 The Binomial Distribution 55
3.5 The Hypergeometric Distribution 55
3.6 Confidence Limits 57
3.7 Classification into More than Two Classes 60
3.8 Confidence Limits with More than Two Classes 60
3.9 The Conditional Distribution of p 61
3.10 Proportions and Totals Over Subpopulations 63
3.11 Comparisons Between Different Domains 64
3.12 Estimation of Proportions in Cluster Sampling 64

Exercises 68

CHAPTER

4 THE ESTIMATION OF SAMPLE SIZE 72

4.1 A Hypothetical Example 72
4.2 Analysis of the Problem 73
4.3 The Specification of Precision 74
4.4 The Formula for n in Sampling for Proportions 75
4.5 Rare Items—Inverse Sampling 76
4.6 The Formula for n with Continuous Data 77
4.7 Advance Estimates of Population Variances 78
4.8 Sample Size with More than One Item 81
4.9 Sample Size when Estimates Are Wanted for Subdivisions of the Population 82
4.10 Sample Size in Decision Problems 83
4.11 The Design Effect (*Deff*) 85

Exercises 86

CHAPTER

5 STRATIFIED RANDOM SAMPLING 89

5.1 Description 89
5.2 Notation 90
5.3 Properties of the Estimates 91
5.4 The Estimated Variance and Confidence Limits 95
5.5 Optimum Allocation 96

5.6 Relative Precision of Stratified Random and Simple Random Sampling 99
5.7 When Does Stratification Produce Large Gains in Precision? . . . 101
5.8 Allocation Requiring More than 100 Per Cent Sampling 104
5.9 Estimation of Sample Size with Continuous Data 105
5.10 Stratified Sampling for Proportions 107
5.11 Gains in Precision in Stratified Sampling for Proportions 109
5.12 Estimation of Sample Size with Proportions 110

Exercises 111

CHAPTER

5A FURTHER ASPECTS OF STRATIFIED SAMPLING 115

5A.1 Effects of Deviations from the Optimum Allocation 115
5A.2 Effects of Errors in the Stratum Sizes 117
5A.3 The Problem of Allocation with More than One Item 119
5A.4 Other Methods of Allocation with More than One Item 121
5A.5 Two-Way Stratification with Small Samples 124
5A.6 Controlled Selection 126
5A.7 The Construction of Strata 127
5A.8 Number of Strata 132
5A.9 Stratification After Selection of the Sample (Poststratification) . . 134
5A.10 Quota Sampling 135
5A.11 Estimation from a Sample of the Gain Due to Stratification . . . 136
5A.12 Estimation of Variance with One Unit per Stratum 138
5A.13 Strata as Domains of Study 140
5A.14 Estimating Totals and Means Over Subpopulations 142
5A.15 Sampling from Two Frames 144

Exercises 146

CHAPTER

6 RATIO ESTIMATORS 150

6.1 Methods of Estimation 150
6.2 The Ratio Estimate 150
6.3 Approximate Variance of the Ratio Estimate 153
6.4 Estimation of the Variance from a Sample 155
6.5 Confidence Limits 156
6.6 Comparison of the Ratio Estimate with Mean per Unit 157
6.7 Conditions Under Which the Ratio Estimate Is a Best Linear Unbiased Estimator 158
6.8 Bias of the Ratio Estimate 160

6.9 Accuracy of the Formulas for the Variance and Estimated Variance 162
6.10 Ratio Estimates in Stratified Random Sampling 164
6.11 The Combined Ratio Estimate 165
6.12 Comparison of the Combined and Separate Estimates 167
6.13 Short-Cut Computation of the Estimated Variance 169
6.14 Optimum Allocation with a Ratio Estimate 172
6.15 Unbiased Ratio-type Estimates 174
6.16 Comparison of the Methods 177
6.17 Improved Estimation of Variance 178
6.18 Comparison of Two Ratios 180
6.19 Ratio of Two Ratios 183
6.20 Multivariate Ratio Estimates 184
6.21 Product Estimators 186

Exercises 186

CHAPTER
7 REGRESSION ESTIMATORS 189

7.1 The Linear Regression Estimate 189
7.2 Regression Estimates with Preassigned b 190
7.3 Regression Estimates when b Is Computed from the Sample . . . 193
7.4 Sample Estimate of Variance 195
7.5 Large-Sample Comparison with the Ratio Estimate and the Mean per Unit 195
7.6 Accuracy of the Large-Sample Formulas for $V(\bar{y}_{lr})$ and $v(\bar{y}_{lr})$ 197
7.7 Bias of the Linear Regression Estimate 198
7.8 The Linear Regression Estimator Under a Linear Regression Model 199
7.9 Regression Estimates in Stratified Sampling 200
7.10 Regression Coefficients Estimated from the Sample 201
7.11 Comparison of the Two Types of Regression Estimate 203

Exercises 203

CHAPTER
8 SYSTEMATIC SAMPLING 205

8.1 Description 205
8.2 Relation to Cluster Sampling 207
8.3 Variance of the Estimated Mean 207
8.4 Comparison of Systematic with Stratified Random Sampling . . . 212
8.5 Populations in "Random" Order 212

8.6 Populations with Linear Trend 214
8.7 Methods for Populations with Linear Trends 216
8.8 Populations with Periodic Variation 217
8.9 Autocorrelated Populations 219
8.10 Natural Populations 221
8.11 Estimation of the Variance from a Single Sample 223
8.12 Stratified Systematic Sampling 226
8.13 Systematic Sampling in Two Dimensions 227
8.14 Summary 229

Exercises 231

CHAPTER 9 SINGLE-STAGE CLUSTER SAMPLING: CLUSTERS OF EQUAL SIZES 233

9.1 Reasons for Cluster Sampling 233
9.2 A Simple Rule 234
9.3 Comparisons of Precision Made from Survey Data 238
9.4 Variance in Terms of Intracluster Correlation 240
9.5 Variance Functions 243
9.6 A Cost Function 244
9.7 Cluster Sampling for Proportions 246

Exercises 247

CHAPTER 9A SINGLE-STAGE CLUSTER SAMPLING: CLUSTERS OF UNEQUAL SIZES 249

9A.1 Cluster Units of Unequal Sizes 249
9A.2 Sampling with Probability Proportional to Size 250
9A.3 Selection with Unequal Probabilities with Replacement 252
9A.4 The Optimum Measure of Size 255
9A.5 Relative Accuracies of Three Techniques 255
9A.6 Sampling with Unequal Probabilities Without Replacement . . . 258
9A.7 The Horvitz-Thompson Estimator 259
9A.8 Brewer's Method 261
9A.9 Murthy's Method 263
9A.10 Methods Related to Systematic Sampling 265
9A.11 The Rao, Hartley, Cochran Method 266
9A.12 Numerical Comparisons 267
9A.13 Stratified and Ratio Estimates 270

Exercises 272

CHAPTER

10 SUBSAMPLING WITH UNITS OF EQUAL SIZE 274

10.1 Two-Stage Sampling 274
10.2 Finding Means and Variances in Two-Stage Sampling 275
10.3 Variance of the Estimated Mean in Two-Stage Sampling 276
10.4 Sample Estimation of the Variance 278
10.5 The Estimation of Proportions 279
10.6 Optimum Sampling and Subsampling Fractions 280
10.7 Estimation of m_{opt} from a Pilot Survey 283
10.8 Three-Stage Sampling 285
10.9 Stratified Sampling of the Units 288
10.10 Optimum Allocation with Stratified Sampling 289

Exercises 290

CHAPTER

11 SUBSAMPLING WITH UNITS OF UNEQUAL SIZES 292

11.1 Introduction 292
11.2 Sampling Methods when $n = 1$ 293
11.3 Sampling with Probability Proportional to Estimated Size 297
11.4 Summary of Methods for $n = 1$ 299
11.5 Sampling Methods When $n > 1$ 300
11.6 Two Useful Results 300
11.7 Units Selected with Equal Probabilities: Unbiased Estimator . . . 303
11.8 Units Selected with Equal Probabilities: Ratio to Size Estimate . . 303
11.9 Units Selected with Unequal Probabilities with Replacement: Unbiased Estimator 306
11.10 Units Selected Without Replacement 308
11.11 Comparison of the Methods 310
11.12 Ratios to Another Variable 311
11.13 Choice of Sampling and Subsampling Fractions. Equal Probabilities 313
11.14 Optimum Selection Probabilities and Sampling and Subsampling Rates 314
11.15 Stratified Sampling. Unbiased Estimators 316
11.16 Stratified Sampling. Ratio Estimates 317
11.17 Nonlinear Estimators in Complex Surveys 318
11.18 Taylor Series Expansion 319
11.19 Balanced Repeated Replications 320
11.20 The Jackknife Method 321
11.21 Comparison of the Three Approaches 322

Exercises 324

CHAPTER

12 DOUBLE SAMPLING 327

12.1 Description of the Technique 327
12.2 Double Sampling for Stratification 327
12.3 Optimum Allocation 331
12.4 Estimated Variance in Double Sampling for Stratification 333
12.5 Double Sampling for Analytical Comparisons 335
12.6 Regression Estimators 338
12.7 Optimum Allocation and Comparison with Single Sampling . . . 341
12.8 Estimated Variance in Double Sampling for Regression 343
12.9 Ratio Estimators 343
12.10 Repeated Sampling of the Same Population 344
12.11 Sampling on Two Occasions 346
12.12 Sampling on More than Two Occasions 348
12.13 Simplifications and Further Developments 351

Exercises 355

CHAPTER

13 SOURCES OF ERROR IN SURVEYS 359

13.1 Introduction 359
13.2 Effects of Nonresponse 359
13.3 Types of Nonresponse 364
13.4 Call-backs 365
13.5 A Mathematical Model of the Effects of Call-backs 367
13.6 Optimum Sampling Fraction Among the Nonrespondents 370
13.7 Adjustments for Bias Without Call-backs 374
13.8 A Mathematical Model for Errors of Measurement 377
13.9 Effects of Constant Bias 379
13.10 Effects of Errors that Are Uncorrelated Within the Sample . . . 380
13.11 Effects of Intrasample Correlation Between Errors of Measurement 383
13.12 Summary of the Effects of Errors of Measurement 384
13.13 The Study of Errors of Measurement 384
13.14 Repeated Measurement of Subsamples 386
13.15 Interpenetrating Subsamples 388
13.16 Combination of Interpenetration and Repeated Measurement . . 391
13.17 Sensitive Questions: Randomized Responses 392
13.18 The Unrelated Second Question 393
13.19 Summary 395

Exercises 396

References 400
Answers to Exercises 412
Author Index 419
Subject Index 422