# Volume 29A

# *Sample Surveys*: *Design, Methods and Applications*

Edited by


## D. Pfeffermann

Department of Statistics, Hebrew University of Jerusalem, Israel

and

Southampton Statistical Sciences Research Institute,
University of Southampton, UK

## C.R. Rao

Director Center for Multivariate Analysis,
Department of Statistics, The Pennsylvania State University,
University Park, PA, USA

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo
North-Holland is an imprint of Elsevier

# Table of Contents

## Volume 29A Sample Surveys: Design, Methods and Applications

Preface to Handbook 29A v

Contributors: Vol. 29A xix

Contributors: Vol. 29B xxiii

Part 1. Sampling and Survey Design 1

Introduction to Part 1 3
*Sharon L. Lohr*

1. Importance of survey design 3
2. Framework and approaches to design and inference 4
3. Challenges in survey design 6

Ch. 1. Introduction to Survey Sampling 9
*Ken Brewer and Timothy G. Gregoire*

1. Two alternative approaches to survey sampling inference 9
2. Historical approaches to survey sampling inference 13
3. Some common sampling strategies 21
4. Conclusion 37

Ch. 2. Sampling with Unequal Probabilities 39
*Yves G. Berger and Yves Tillé*

1. Introduction 39
2. Some methods of unequal probability sampling 40
3. Point estimation in unequal probability sampling without replacement 47
4. Variance estimators free of joint inclusion probabilities 48
5. Variance estimation of a function of means 50
6. Balanced sampling 51

Ch. 3. Two-Phase Sampling 55
*Jason C. Legg and Wayne A. Fuller*

1. Introduction 55
2. Using auxiliary information in estimation 60
3. Three-phase sampling 65
4. Two-phase estimation illustration 66

Ch. 4. Multiple-Frame Surveys 71
*Sharon L. Lohr*

1. What are multiple-frame surveys, and why are they used? 71
2. Point estimation in multiple-frame surveys 76
3. Variance estimation in multiple-frame surveys 83
4. Designing multiple-frame surveys 85
5. New applications and challenges for multiple-frame surveys 85
Acknowledgments 88

Ch. 5. Designs for Surveys over Time 89
*Graham Kalton*

1. Introduction 89
2. Repeated surveys 91
3. Rotating panel surveys 94
4. Panel surveys 97
5. Conclusions 108

Ch. 6. Sampling of Rare Populations 109
*Mary C. Christman*

1. Introduction 109
2. Modifications to classical design-based sampling strategies 110
3. Adaptive sampling designs 115
4. Experimental design 123
5. Confidence interval estimation 123
6. Summary 124

Ch. 7. Design, Conduct, and Analysis of Random-Digit Dialing Surveys 125
*Kirk Wolter, Sadeq Chowdhury and Jenny Kelly*

1. Introduction 125
2. Design of RDD surveys 126
3. Conduct of RDD surveys 134
4. Analysis of RDD surveys 146

Part 2. Survey Processing 155

Introduction to Part 2 157
*Paul Biemer*

1. Overview of data processing steps 157
2. Data quality and data processing 162

Ch. 8. Nonresponse and Weighting 163
*J. Michael Brick and Jill M. Montaquila*

1. Nonresponse in surveys 163
2. Response rates 167
3. The relationship between response rates and nonresponse bias 170
4. Weighting for nonresponse 174
5. Variance and confidence interval estimation 181
6. Discussion 183

Ch. 9. Statistical Data Editing 187
*Ton De Waal*

1. Introduction 187
2. The use of edit rules 188
3. Interactive editing 189
4. Editing during the data collection phase 191
5. Selective editing 192
6. Automatic editing 198
7. Macro-editing 207
8. A strategy for statistical data editing 211
9. Discussion 213

Ch. 10. Imputation and Inference in the Presence of Missing Data 215
*David Haziza*

1. Introduction 215
2. Context and defnitions 216
3. Bias of the imputed estimator 224
4. Variance of the imputed estimator 230
5. Imputation classes 231
6. Variance estimation 235
7. Multiple imputation 243
8. Conclusions 246

Ch. 11. Dealing with Outliers in Survey Data 247
*Jean-François Beaumont and Louis-Paul Rivest*

1. Introduction 247
2. Estimation of the mean of an asymmetric distribution in an infinite population 250
3. The estimation of totals in finite populations containing outliers 256
4. The estimation of totals using auxiliary information in finite populations containing outliers 259
5. Dealing with stratum jumpers 270
6. Practical issues and future work 278

Ch. 12. Measurement Errors in Sample Surveys 281
*Paul Biemer*

1. Introduction 281
2. Modeling survey measurement error 282
3. The truth as a latent variable: Latent class models 289
4. Latent class models for three or more polytomous indicators 294
5. Some advanced topics 303
6. Measurement error evaluation with continuous variables 309
7. Discussion 315

Ch. 13. Computer Software for Sample Surveys 317
*Jelke Bethlehem*

1. Survey process 317
2. Data collection 318
3. Statistical data editing 325
4. Imputation 332
5. Weighting adjustment 336
6. Analysis 343
7. Disclosure control 348

Ch. 14. Record Linkage 351
*William E. Winkler*

1. Introduction 351
2. Overview of methods 353
3. Data preparation 366
4. More advanced methods 370
5. Concluding remarks 380

Ch. 15. Statistical Disclosure Control for Survey Data 381
*Chris Skinner*

1. Introduction 381
2. Tabular outputs 384
3. Microdata 388
4. Conclusion 396
Acknowledgments 396

Part 3. Survey Applications 397

Introduction to Part 3 399
*Jack G. Gambino*

1. Frames and designs 399
2. Stratification, allocation and sampling 401
3. Estimation 402

4. Auxiliary information 403
5. Challenges 404

Ch. 16. Sampling and Estimation in Household Surveys 407
*Jack G. Gambino and Pedro Luis do Nascimento Silva*

1. Introduction 407
2. Survey designs 408
3. Repeated household surveys 415
4. Data collection 425
5. Weighting and estimation 427
6. Nonsampling errors in household surveys 435
7. Integration of household surveys 436
8. Survey redesign 438
9. Conclusions 438
Acknowledgments 439

Ch. 17. Sampling and Estimation in Business Surveys 441
*Michael A. Hidiroglou and Pierre Lavallée*

1. Introduction 441
2. Sampling frames for business surveys 442
3. Administrative data 446
4. Sample size determination and allocation 450
5. Sample selection and rotation 457
6. Data editing and imputation 460
7. Estimation 467

Ch. 18. Sampling, Data Collection, and Estimation in Agricultural Surveys 471
*Sarah M. Nusser and Carol C. House*

1. Introduction 471
2. Sampling 473
3. Data collection 480
4. Statistical estimation 482
5. Confidentiality 485
6. Concluding remarks 485
Acknowledgments 486

Ch. 19. Sampling and Inference in Environmental Surveys 487
*David A. Marker and Don L. Stevens Jr.*

1. Introduction 487
2. Sampling populations in space 489
3. Defining sample frames for environmental populations 490
4. Designs for probability-based environmental samples 492
5. Using ancillary information in design 500
6. Inference for probability-based design 502
7. Model-based optimal spatial designs 503

# Volume 29B

## *Sample Surveys: Inference and Analysis*

Edited by


## D. Pfeffermann

Department of Statistics, Hebrew University of Jerusalem, Israel

and

Southampton Statistical Sciences Research Institute,
University of Southampton, UK

## C.R. Rao

Director Center for Multivariate Analysis,
Department of Statistics, The Pennsylvania State University,
University Park, PA, USA

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo
North-Holland is an imprint of Elsevier

8. Plot design issues 506
9. Sources of error in environmental studies 508
10. Conclusions 512
Acknowledgements 512

Ch. 20. Survey Sampling Methods in Marketing Research: A Review of Telephone, Mall Intercept, Panel, and Web Surveys 513
*Raja Velu and Gurramkonda M. Naidu*

1. Introduction 513
2. Telephone surveys 517
3. Fax surveys 527
4. Shopping center sampling and interviewing 527
5. Consumer panels 529
6. Web surveys 534
7. Conclusion 538

Ch. 21. Sample Surveys and Censuses 539
*Ronit Nirel and Hagit Glickman*

1. Introduction 539
2. The use of sample surveys for estimating coverage errors 541
3. The use of sample surveys to evaluate statistical adjustment of census counts 552
4. The use of sample surveys for carrying out a census 558
5. Sample surveys carried out in conjunction with a census 562
6. Concluding remarks 564

Ch. 22. Opinion and Election Polls 567
*Kathleen A. Frankovic, Costas Panagopoulos and Robert Y. Shapiro*

. Introduction: the reasons for public opinion and election polling 567
:. General methodological issues in public opinion and election polls 575
3. Preelection polling: methods, impact, and current issues 580
4. Exit polling 584
5. Postelection and between-election polls 588
6. Other opinion measurements: focus groups, deliberative polls, and the effect of political events 588
7. Present and future challenges in polling 590
8. Continued interest in public opinion and polling 594
Acknowledgments 595

References 597

Subject Index: Index of Vol. 29A 651

Subject Index: Index of Vol. 29B 661

Handbook of Statistics: Contents of Previous Volumes 671

# Table of Contents

## Volume 29B Sample Surveys: Inference and Analysis

Preface to Handbook 29B v

Contributors: Vol. 29B xix

Contributors: Vol. 29A xxi

Part 4. Alternative Approaches to Inference from Survey Data 1

Introduction to Part 4 3
*Jean D. Opsomer*

1. Introduction 3
2. Modes of inference with survey data 4
3. Overview of Part 4 8

Ch. 23. Model-Based Prediction of Finite Population Totals 11
*Richard Valliant*

1. Superpopulation models and some simple examples 11
2. Prediction under the general linear model 14
3. Estimation weights 17
4. Weighted balance and robustness 18
5. Variance estimation 20
6. Models with qualitative auxiliaries 22
7. Clustered populations 23
8. Estimation under nonlinear models 29

Ch. 24. Design- and Model-Based Inference for Model Parameters 33
*David A. Binder and Georgia Roberts*

1. Introduction and scope 33
2. Survey populations and target populations 34

3. Statistical inferences 38
4. General theory for fitting models 41
5. Cases where design-based methods can be problematic 47
6. Estimation of design-based variances 51
7. Integrating data from more than one survey 52
8. Some final remarks 53

Ch. 25. Calibration Weighting: Combining Probability Samples and Linear Prediction Models 55
*Phillip S. Kott*

1. Introduction 55
2. Randomization consistency and other asymptotic properties 58
3. The GREG estimator 60
4. Redefining calibration weights 65
5. Variance estimation 69
6. Nonlinear calibration 73
7. Calibration and quasi-randomization 76
8. Other approaches, other issues 80
Acknowledgements 82

Ch. 26. Estimating Functions and Survey Sampling 83
*V. P. Godambe and Mary E. Thompson*

1. Introduction 83
2. Defining finite population and superpopulation parameters through estimating functions 84
3. Design-unbiased estimating functions 85
4. Optimality 87
5. Asymptotic properties of sample estimating functions and their roots 89
6. Interval estimation from estimating functions 92
7. Bootstrapping estimating functions 96
8. Multivariate and nuisance parameters 97
9. Estimating functions and imputation 99
Acknowledgment 101

Ch. 27. Nonparametric and Semiparametric Estimation in Complex Surveys 103
*F. Jay Breidt and Jean D. Opsomer*

1. Introduction 103
2. Nonparametric methods in descriptive inference from surveys 107
3. Nonparametric methods in analytic inference from surveys 114
4. Nonparametric methods in nonresponse adjustment 115
5. Nonparametric methods in small area estimation 118

Ch. 28. Resampling Methods in Surveys 121
*Julie Gershunskaya, Jiming Jiang and P. Lahiri*

1. Introduction 121
2. The basic notions of bootstrap and jackknife 123
3. Methods for more complex survey designs and estimators 126

4. Variance estimation in the presence of imputation 134
5. Resampling methods for sampling designs in two phases 137
6. Resampling methods in the prediction approach 138
7. Resampling methods in small area estimation 140
8. Discussion 149
Acknowledgments 151

Ch. 29. Bayesian Developments in Survey Sampling 153
*Malay Ghosh*

1. Introduction 153
2. Notation and preliminaries 154
3. The Bayesian paradigm 156
4. Linear Bayes estimator 161
5. Bayes estimators of the finite population mean under more complex models 164
6. Stratified sampling and domain estimation 174
7. Generalized linear models 179
8. Summary 186
Acknowledgments 187

Ch. 30. Empirical Likelihood Methods 189
*J.N.K. Rao and Changbao Wu*

1. Likelihood-based approaches 189
2. Empirical likelihood method under simple random sampling 191
3. Stratified simple random sampling 193
4. Pseudo empirical likelihood method 194
5. Computational algorithms 205
6. Discussion 206
Acknowledgments 207

Part 5. Special Estimation and Inference Problems 209

Introduction to Part 5 211
*Gad Nathan and Danny Pfeffermann*

1. Preface 211
2. Overview of chapters in Part 5 212

Ch. 31. Design-based Methods of Estimation for Domains and Small Areas 219
*Risto Lehtonen and Ari Veijanen*

1. Introduction 219
2. Theoretical framework, terminology, and notation 221
3. Direct estimators for domain estimation 226
4. Indirect estimators in domain estimation 233
5. Extended GREG family for domain estimation 244
6. Software 248
Acknowledgments 249

Ch. 32. Model-Based Approach to Small Area Estimation 251
*Gauri S. Datta*

1. Introduction 251
2. Model-based frequentist small area estimation 253
3. Bayesian approach to small area estimation 270
4. Concluding remarks 284
Acknowledgements 288

Ch. 33. Design and Analysis of Surveys Repeated over Time 289
*David Steel and Craig McLaren*

1. Overview of issues for repeated surveys 289
2. Basic theory of design and estimation for repeated surveys 293
3. Rotation patterns 296
4. Best linear and composite estimation 297
5. Correlation models for sampling errors 301
6. Rotation patterns and sampling variances 304
7. Time series methods for estimation in repeated surveys 305
8. Seasonal adjustment and trend estimation 309
9. Variance estimation for seasonally adjusted and trend estimates 311
10. Rotation patterns and seasonally adjusted and trend estimates 313

Ch. 34. The Analysis of Longitudinal Surveys 315
*Gad Nathan*

1. Introduction 315
2. Types and problems of longitudinal surveys 316
3. General models for analysis of longitudinal data 318
4. Treatment of nonresponse 322
5. Effects of informative sample design on longitudinal analysis 325

Ch. 35. Categorical Data Analysis for Simple and Complex Surveys 329
*Avinash C. Singh*

1. Introduction 329
2. Likelihood-based methods 332
3. Quasi-likelihood methods 345
4. Weighted quasi-likelihood methods 350
5. Unit-level models 361
6. Conclusions 367
Acknowledgments 369

Ch. 36. Inference on Distribution Functions and Quantiles 371
*Alan H. Dorfman*

. Introduction 371
. Estimating the distribution function with no auxiliary information 374
3. Estimating the distribution function with complete auxiliary information 376
4. Estimating the distribution function using partial auxiliary information 389
5. Quantile estimation 390

6. Variance estimation and confidence intervals for distribution functions 392
7. Confidence intervals and variance estimates for quantiles 393
8. Further results and questions 395

Ch. 37. Scatterplots with Survey Data 397
*Barry I. Graubard and Edward L. Korn*

1. Introduction 397
2. Modifications of scatterplots for survey data 397
3. Discussion 419

Part 6. Informative Sampling and Theoretical Aspects 421

Introduction to Part 6 423
*Danny Pfeffermann*

1. Motivation 423
2. Overview of chapters in Part 6 426

Ch. 38. Population-Based Case–Control Studies 431
*Alastair Scott and Chris Wild*

1. Introduction to case–control sampling 431
2. Basic results 434
3. Two-phase case–control sampling 447
4. Case–control family studies 451
5. Conclusion 453

Ch. 39. Inference under Informative Sampling 455
*Danny Pfeffermann and Michail Sverchkov*

1. Introduction 455
2. Informative and ignorable sampling 459
3. Overview of approaches that account for informative sampling and nonresponse 461
4. Use of the sample distribution for inference 468
5. Prediction under informative sampling 475
6. Other applications of the sample distribution 479
7. Tests of sampling ignorability 484
8. Brief summary 486
Acknowledgements 487

Ch. 40. Asymptotics in Finite Population Sampling 489
*Zuzana Prášková and Pranab Kumar Sen*

1. Introduction 489
2. Asymptotics in SRS 490
3. Resampling in FPS: Asymptotics 495
4. Estimation of population size: Asymptotics 499

5. Sampling with varying probabilities: Asymptotics 502
6. Large entropy and relative samplings: Asymptotic results 510
7. Successive subsampling with varying probabilities: Asymptotics 518
8. Conclusions 521
Acknowledgment 522

Ch. 41. Some Decision-Theoretic Aspects of Finite Population Sampling 523
*Yosef Rinott*

1. Introduction 523
2. Notations and definitions 524
3. Minimax strategies 528
4. UMVU estimators 541
5. Admissibility 542
6. Superpopulation models 546
7. Beyond simple random sampling 552
8. List of main notations 557
Acknowledgements 558

References 559

Subject Index: Index of Vol. 29B 595

Subject Index: Index of Vol. 29A 605

Handbook of Statistics: Contents of Previous Volumes 615