

9062893

handbook of Statistics 6

Sampling



Edited by

P. R. Krishnaiah[†]

C. R. Rao

0212.1
K92
V.6

- 9062893

Sampling



E9062893

Edited by

P. R. Krishnaiah[†]

C. R. Rao

Department of Mathematics and Statistics
University of Pittsburgh, Pittsburgh, PA, U.S.A.



1988

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

8885208

© ELSEVIER SCIENCE PUBLISHERS B.V., 1988

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 0444 70289 X

Published by:

ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

LIBRARY OF CONGRESS

Library of Congress Cataloging-in-Publication Data

Sampling / edited by P.R. Krishnaiah, C.R. Rao.

p. cm. — (Handbook of statistics : v. 6)

Includes bibliographies and index.

ISBN 0-444-70289-X (U.S.)

I. Sampling (Statistics) I. Krishnaiah. Paruchuri R. II. Rao.

C. Radhakrishna (Calyampudi Radhakrishna). 1920- . III. Series.

QA276.6.S334 1988

88-374

519.5'2--dc 19

CIP

PRINTED IN THE NETHERLANDS

Handbook of Statistics

VOLUME 6

General Editors

P. R. Krishnaiah[†]

C. R. Rao



NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

Preface

The series *Handbook of Statistics* was started by Professor P. R. Krishnaiah, who unfortunately passed away on August 1, 1987 at the prime age of 55 years. The object of bringing out these volumes, as mentioned by the founding editor in the preface to the first volume, is to provide 'comprehensive and self contained reference books to disseminate information on various aspects of statistical methodology and applications'. This is not an easy task and only an erudite scholar like Professor Krishnaiah with a deep knowledge of different fields of statistics and a missionary zeal could achieve it. This is the sixth volume which he edited and shortly before his death he made plans for producing six more volumes devoted to different areas of applications of statistics. These volumes have been well received by the entire statistical community, and scientists in various disciplines who use statistical methodology in their work, which is a great tribute to the imaginative efforts of Professor Krishnaiah.

The present volume, number six in the series, is devoted to the theory and practice of Sample Surveys, which is the most widely used method in statistical practice. The basic ideas of survey sampling methodology were formulated in the twenties, but its firm foundations were laid only in the thirties and forties. A brief historical account of sample surveys is given by D. R. Bellhouse. An overview of the subject and the contents of a course on survey sampling are outlined by T. Dalenius. A. Chaudhuri discusses various sampling strategies and the optimality problems associated with them.

The current developments in sample survey methodology are surveyed by the rest of the authors. P. K. Pathak throws new light on the cost-efficiency of simple random sampling, while V. P. Godambe and M. E. Thompson discuss the role of randomization in inference with special reference to single stage unequal probability sampling. Systematic sampling is the theme of contributions by D. R. Bellhouse, M. N. Murty and T. J. Rao. Repeated sampling over time is considered by D. A. Binder and M. A. Hidirolou. Some theoretical aspects of inference in finite populations are covered in the contributions by W. A. Ericson, Gad Nathan, J. Sedransk and P. J. Smith, P. K. Sen, Ib Thomsen and Dinke Tesfu, and R. M. Royall.

J. C. Koop discusses the concept of interpenetrating subsamples introduced by P. C. Mahalanobis. D. H. Freeman discusses the analysis of contingency tables compiled from survey data. J. N. K. Rao reviews the various methods of variance estimation in sample surveys and P. S. R. S. Rao, the methodology of ratio and regression estimation.

G. P. Patil, G. J. Babu, R. C. Hennemuth, W. L. Meyers, M. B. Rajarshi, C. Tallie, M. T. Boswell, F. L. Ramsey, C. E. Gates and K. P. Burnham review special survey techniques in environmental and ecological studies, while R. Velu and G. M. Naidu review the current sampling methods in marketing research. P. V. Sukhatme discusses methods for controlling and estimating observational errors in sample surveys.

A Hedayat, C. R. Rao and H. Stufken present some new problems in the design of sample surveys. They provide sampling designs to avoid contiguous units occurring in samples in order to increase the efficiency of estimates.

This volume provides an unusual and useful collection of articles covering many theoretical and practical aspects of sample surveys in social and biological investigations. Written by experts and actual practioners of sample surveys, it would be a valuable guide to those involved in designing sample surveys for collection of data and estimation of unknown population parameters.

I would like to thank North-Holland Publishing Company for their patience and excellent cooperation in bringing out this Volume.

C. R. Rao

Contributors

- G. J. Babu, *Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, Dept. of Statistics, 303 Pond Laboratory, University Park, PA 16802, USA* (Ch. 20)
- D. R. Bellhouse, *Dept. of Statistical & Actuarial Sciences, The University of Western Ontario, Faculty of Science, Room 3005 EMSc, London, Canada N6A 6B9* (Ch. 1, 6)
- D. A. Binder, *Social Survey Methods Division, Statistics Canada, Floor 4, Jean-Talon, Tunney's Pasture, Ottawa, Ontario, K1A 0T6 Canada* (Ch. 8)
- M. T. Boswell, *Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, Dept. of Statistics, 303 Pond Laboratory, University Park, PA 16802, USA* (Ch. 19)
- K. P. Burnham, *Dept. of Statistics, North Carolina State University, Raleigh, NC, USA* (Ch. 19)
- A. Chaudhuri, *Computer Science Unit, India Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India* (Ch. 3)
- T. Dalenius, *Dept. of Statistics, Brown University, Providence, RI 02912, USA* (Ch. 2)
- W. A. Ericson, *Dept. of Statistics, University of Michigan, Ann Arbor, MI 48104, USA* (Ch. 9)
- D. H. Freeman, Jr., *Associate Professor of Public Health (Biostatistics), Yale University, P.O. Box 3333, 60 College Street, New Haven, CT 06510, USA* (Ch. 16)
- C. E. Gates, *Dept. of Statistics, Texas A&M University, College Station, TX 77843, USA* (Ch. 21)
- V. P. Godambe, *Dept. of Statistics, University of Waterloo, Waterloo, Ontario, Canada* (Ch. 5)
- S. Hedayat, *Dept. of Math. Stat. and Comp. Science, 322 Science and Engineering Offices, Box 4348, Chicago, IL 60680, USA* (Ch. 24)
- R. C. Hennemuth, *Northeast Fisheries Center, National Marine Fisheries Service, Woods Hole, MA, USA* (Ch. 20)
- M. A. Hidirolou, *Business Survey Methods Division, Statistics Canada, Floor 11, RHC Bldg., Tunney's Pasture, Ottawa, Ontario, Canada K1A 0T6* (Ch. 8)
- J. C. Koop, *3201 Clark Avenue, Raleigh, NC 27606, USA* (Ch. 13)
- M. N. Murthy, *129 Luz Church Road, Mylapore, Madras 600 004, India* (Ch. 7)
- W. L. Meyers, *Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, Dept. of Statistics, 303 Pond Laboratory, University Park, PA 16802, USA* (Ch. 20)

- G. M. Naidu, *University of Wisconsin—White Water, 800 West Main Street, Whitewater, WI 53190, USA* (Ch. 22)
- G. M. Nathan, *Dept. of Statistics, The Hebrew University of Jerusalem, Jerusalem, Israel* (Ch. 10)
- P. K. Pathak, *Dept. of Mathematics, University of New Mexico, Albuquerque, NM 87131, USA* (Ch. 4)
- G. P. Patil, *Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, Department of Statistics, 303 Pond Laboratory, University Park, PA 16802, USA* (Ch. 19, 20, 21)
- M. B. Rajarshi, *Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, Dept. of Statistics, 303 Pond Laboratory, University Park, PA 16802, USA* (Ch. 20)
- F. L. Ramsey, *Dept. of Statistics, Oregon State University, Corvallis, OR 97331, USA* (Ch. 21)
- C. R. Rao, *Dept. of Mathematics & Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA* (Ch. 24)
- J. N. K. Rao, *Dept. of Mathematics, Carleton University, Colonel by Drive, Ottawa, Ontario, Canada K1S 5B6* (Ch. 17)
- P. S. R. S. Rao, *Dept. of Statistics, University of Rochester, Dewey Hall 232, Rochester, NY 14627, USA* (Ch. 18)
- T. J. Rao, *Math.-Stat. Division, Indian Statistical Institute, 203 B.T. Road, Calcutta 700035, India* (Ch. 7)
- R. M. Royall, *The John Hopkins University, School of Hygiene & Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA* (Ch. 15)
- J. Sedransk, *Dept. of Mathematics, SUNY at Albany, Albany, NY 12222* (Ch. 11)
- P. K. Sen, *Dept. of Biostatistics, University of North Carolina, Chapel Hill, NC 27514, USA* (Ch. 12)
- P. J. Smith, *International Pacific Halibut Commission, P.O. Box 95009, Seattle, WA 98145-2009, USA* (Ch. 11)
- J. Stufken, *University of Georgia, Athens, GA, USA* (Ch. 24).
- P. V. Sukhatme, *Biometry Dept., Maharashtra Assoc. for the Cultivation of Science, Law College Road, Poona 411004, India* (Ch. 23)
- C. Taillie, *Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, Dept. of Statistics, 303 Pond Laboratory, University Park, PA 16802, USA* (Ch. 20, 21)
- D. Tesfu, *Central Statistical Office, P.O. Box 1143, Addis Ababa, Ethiopia* (Ch. 14)
- M. E. Thompson, *Dept. of Statistics, University of Waterloo, Waterloo, Ontario, Canada* (Ch. 5)
- I. Thomsen, *Central Statistical Office of Norway P.B. 8131 Dep., Oslo 1, Norway* (Ch. 14)
- R. Velu, *University of Wisconsin—White Water, 800 West Main Street, Whitewater, WI 53190, USA* (Ch. 22).

Table of Contents

Preface v

Contributors xv

Ch. 1. A Brief History of Random Sampling Methods 1

D. R. Bellhouse

1. Introduction 1
2. Kiaer 2
3. Bowley 4
4. Neyman 7
5. Hansen and Hurwitz 9
6. Other developments 10
7. The paradigm challenged and defended 11
- References 13

Ch. 2. A First Course in Survey Sampling 15

T. Dalenius

- Introduction 15
- I. The notion of survey sampling 16
 - II. A review of the statistics needed 18
 - III. Getting observational access to the population 21
 - IV. Element sampling 24
 - V. Cluster sampling 32
 - VI. Multi-stage sampling 39
 - VII. The problems of non-sampling errors 41
 - VIII. Survey sampling design 44

Ch. 3. Optimality of Sampling Strategies 47

A. Chaudhuri

1. Introduction 47
2. Repeated sampling approach: Fixed population case 49

3. Optimality under super-population modelling	75
4. Likelihood approach	79
5. Prediction approach	83
6. A summing up: Efficacy of an asymptotic theory	84
7. Roles of labels and randomization: Controversies	88
References	90

Ch. 4. Simple Random Sampling 97

P. K. Pathak

0. Summary	97
1. Introduction	97
2. Simple random sampling without replacement (SRSWOR)	99
3. Simple random sampling with replacement (SRSWR)	102
4. Fixed cost simple random sampling (SRSFC)	104
Acknowledgment	108
References	108

Ch. 5. On Single Stage Unequal Probability Sampling 111

V. P. Godambe and M. E. Thompson

1. Introduction	111
2. Use of randomization	111
3. Estimating finite population means and totals	114
4. Stratified random sampling	115
5. Sampling designs for regression models	117
6. Monetary unit sampling	119
7. A historical note	121
References	122

Ch. 6. Systematic Sampling 125

D. R. Bellhouse

1. Introduction	125
2. Sampling theory based on the randomization alone	125
3. Trends in the population	128
4. Autocorrelated populations	134
5. Variance estimation	138
6. Spatial sampling	141
References	143

Ch. 7. Systematic Sampling with Illustrative Examples 147

M. N. Murthy and T. J. Rao

1. The basic procedure	147
2. Estimation and sampling variance	149

- 3. Efficiency of systematic sampling 152
- 4. Variance estimation 158
- 5. Illustrative examples 161
- 6. Bias in systematic sampling 164
- 7. Variations in systematic sampling 165
- 8. Systematic sampling with probability proportional to size 173
- 9. Superpopulation models and asymptotic results 175
- 10. Applications and illustrations 180
- 11. Some of the available computer programs 181
- 12. Two-dimensional systematic sampling 182
 - Bibliographical note 183
 - Reference 183

Ch. 8. Sampling in Time 187

D. A. Binder and M. A. Hidirolou

- 1. Introduction 187
- 2. General approaches to designs with overlapping units 188
- 3. The classical approach 190
- 4. Rotation bias 197
- 5. Time series approaches 200
 - Acknowledgements 210
 - References 210

Ch. 9. Bayesian Inference in Finite Populations 213

W. A. Ericson

- 1. Introduction 213
- 2. The basic model 213
- 3. Some basic results 215
- 4. Exchangeability and simple random sampling 217
- 5. Stratified sampling 220
- 6. Two stage sampling 223
- 7. Ratio and regression estimation 227
- 8. Response error and bias 223
 - Discussion 241
 - References 243

Ch. 10. Inference Based on Data from Complex Sample Designs 247

G. Nathan

- 1. Introduction 247
- 2. General methods 251
- 3. Regression and linear models 255
- 4. Categorical data analysis 258
- 5. Other methods of analysis 262
 - References 263

Ch. 11. Inference for Finite Population Quantiles 267

J. Sedransk and P. J. Smith

- 1. Introduction 267
- 2. Confidence interval methods for complex sample designs 268
- 3. Point estimation for finite population quantiles 280
- 4. Bayesian methods 283
- Acknowledgement 288
- References 288

Ch. 12. Asymptotics in Finite Population Sampling 291

P. K. Sen

- 1. Introduction 291
- 2. Asymptotics in SRS 292
- 3. Some probability and moment inequalities for SRS 297
- 4. Jackknifing in finite population sampling 301
- 5. Estimation of population size: Asymptotics 304
- 6. Sampling with varying probabilities: Asymptotics 313
- 7. Successive sub-sampling with varying probabilities: Asymptotics 324
- References 328

Ch. 13. The Technique of Replicated or Interpenetrating Samples 333

J. C. Koop

- 1. Introductory review 333
- 2. Theoretical basis of the technique 337
- 3. Applications 353
- 4. Analysis of variance to study methods of investigation 360
- 5. Summary and comments 364
- Acknowledgements 365
- References 365

Ch. 14. On the Use of Models in Sampling from Finite Populations 369

I. Thomsen and D. Tesfu

- 1. Introduction 369
- 2. On the prediction approach 370
- 3. Application of the prediction approach to two-stage sampling 374
- 4. Estimation in election surveys 378
- 5. Other applications of models 381
- 6. Models for response errors 387
- 7. A probabilistic model for nonresponse 391
- References 396

Ch. 15. The Prediction Approach to Sampling Theory 399

R. M. Royall

1. Introduction 399
2. The linear prediction approach 401
3. The approach based on probability sampling distributions 410
- References 412

Ch. 16. Sample Survey Analysis: Analysis of Variance and Contingency Tables 415

D. H. Freeman, Jr.

1. Introduction 415
2. The KFF methodology for analysis of variance models 417
3. An example: Connecticut Blood Pressure Survey 421
4. Summary 424
- Acknowledgement 425
- References 425

Ch. 17. Variance Estimation in Sample Surveys 427

J. N. K. Rao

- Introduction 427
1. Unified approach for linear statistics 428
2. Nonlinear statistics 436
3. Modelling mean square errors 444
- Acknowledgements 446
- References 446

Ch. 18. Ratio and Regression Estimators 449

P. S. R. S. Rao

1. Introduction 449
2. Bias and mean square error of the ratio estimator 450
3. Bias reduction 451
4. Relative merits of the estimators 453
5. Variance estimation and confidence limits 454
6. Regression through the origin and the ratio estimator 455
7. Stratification and the ratio estimators 457
8. Regression estimator 459
9. Multivariate ratio, product and regression estimators 461
10. Two phase sampling for ratio and regression estimators 464
11. Further developments 465
- Acknowledgements 465
- References 465

Ch. 19. Role and Use of Composite Sampling and Capture-Recapture Sampling in Ecological Studies 469

M. T. Boswell, K. P. Burnham and G. P. Patil

Introduction 469

1. Composite sampling 470

2. Capture-recapture sampling 478

Acknowledgements 486

References 486

Ch. 20. Data-based Sampling and Model-based Estimation for Environmental Resources 489

*G. P. Patil, G. J. Babu, R. C. Hennemuth, W. L. Meyers,
M. B. Rajarshi and C. Taillie*

1. Introduction 489

2. Data-based definitions of populations 489

3. Sampling design as design of encounters—The case of living marine resources 493

4. Sampling design as design of encounters—The bias in fisheries harvest data 497

5. Sample-based modeling of populations—An approach with weighted distributions 499

6. Combining recruitment data and kernel approach 502

7. Encountered ecotoxicological data of chronic effects thresholds 507

8. Synthesis 511

Acknowledgements 511

References 512

Ch. 21. On Transect Sampling to Assess Wildlife Populations and Marine Resources 515

F. L. Ramsey, C. E. Gates, G. P. Patil and C. Taillie

1. Introduction 515

2. Transect surveys 517

3. Estimation procedures 518

4. Variable detectability conditions 521

5. Bob-white flushing 524

6. Deep-sea red crab 524

7. Summary 530

Acknowledgements 530

References 530

Ch. 22. A Review of Current Survey Sampling Methods in Marketing Research (Telephone, Mall Intercept and Panel Surveys) 533

R. Velu and G. M. Naidu

1. Introduction 533

2. Telephone Surveys 534

- 3. Shopping center sampling and interviewing 544
- 4. Consumers panels 546
- Acknowledgements 552
- References 552

Ch. 23. Observational Errors in Behavioural Traits of Man and their Implications for Genetics 555

P. V. Sukhatme

- 1. Introduction 555
- 2. Current theory and its limitations: The sample mean and its variance 556
- 3. Estimation of the different components 559
- 4. Application to longitudinal studies 561
- 5. Genetic implications 569
- References 572

Ch. 24. Designs in Survey Sampling Avoiding Contiguous Units 575

A. S. Hedayat, C. R. Rao and J. Stufken

- 1. Introduction 575
- 2. Sample designs excluding contiguous 575
- 3. Results on the existence and construction 579
- 4. Implementation 582
- References 583

Subject Index 585

Contents of Previous Volumes 589



A Brief History of Random Sampling Methods

D. R. Bellhouse

1. Introduction

The field of survey sampling can claim many roots. These roots include a variety of activities over the nineteenth and twentieth centuries in areas such as agriculture, forestry, government administration, and social research. It is the latter two areas, the collection of large sets of data by governments and individual social investigators, that have given the greatest stimulus to the development of the random sampling techniques discussed here. In this chapter the early work (to about 1945) in sampling is reviewed by looking at the work of those who obtained the principal results which form the basis of many of the later survey sampling textbooks. The issues which confronted these early workers are also related to some of the current discussion over model and design-based inference in survey sampling.

Several histories of sampling have been written. Among them are Chang (1976), Duncan and Shelton (1978), Hansen, Dalenius and Tepping (1985), Kruskal and Mosteller (1980), Seng (1951), Stephan (1948), Sukhatme (1966), and Yates (1948). In many ways this article is a review of the previous reviews. No new historical material is examined. What I hope is new is that in reviewing the reviews some of the motivations behind the early work in sampling are discussed along with the tension between the users of models and randomization that has been present over the whole history of the subject.

An examination of this tension is carried out with Kuhn's (1970) theory of paradigms in the history of science in mind. A paradigm or exemplar is followed by later workers because it provides a framework in which adequate answers are given to the questions being asked. New paradigms are put forward when the old framework no longer provides adequate answers. Kuhn (1970, p. 10) has described two general characteristics of paradigms that, in particular, appear in the history of random sampling. The first characteristic is that the paradigm attracts around it a loyal group of followers away from other competing modes of activity; the other is that the paradigm is open-ended enough to provide a number of unresolved problems on which adherents to the paradigm may work. With regard to the first characteristic, the attraction of a loyal group of followers does not

occur in a vacuum. We shall see that in every 'sampling paradigm', either the proponent or a disciple has actively promoted his ideas.

The initial paradigm in survey sampling is that of the desire for a representative sample as first propounded by A. N. Kiaer in the 1890's. Prior to Kiaer there are several examples of sampling procedures; see, for example, Stephan (1948), Kent (1981, Ch. 1, 3), Chang (1976), and Godambe (1976). These examples illustrate the randomness exhibited in research which, according to Kuhn, is typical of preparadigmatic times. Kiaer's contribution was to provide a framework under which sampling became a reasonable activity; in particular, sampling is useful when it provides a miniature of the population, i.e., it is a representative sample. After Kiaer there emerged two competing methods to attempt to achieve representativeness; randomization and purposive selection. For the most part, in large-scale surveys the paradigm of randomization has become dominant. The reasons for this will be examined as the history of sampling is traced.

2. Kiaer

During the nineteenth century in government statistical agencies and among the social reformers of the so-called statistical movement (see for example, Cullen, 1975 and Kent, 1981, for a discussion of this movement), the generally accepted method of coverage was a complete enumeration. Some sampling was done, but it was the exception rather than the rule. The desire for a complete enumeration in social surveys may be traced back to at least Quetelet; see Stigler (1986, pp. 164–165). The position of complete enumeration was challenged by the Norwegian statistician A. N. Kiaer. Kiaer was the first director of the Norwegian Central Bureau of Statistics, a position he held from the Bureau's inception in 1876 until 1913. His responsibilities included the decennial censuses of population and agriculture and many large-scale statistical investigations. Detailed discussions of Kiaer's work and its impact on sampling methodology may be found in Seng (1951) and Kruskal and Mosteller (1980). What follows is a synopsis of the material found in these papers.

At the Berne meeting of the International Statistical Institute (ISI) in 1895 Kiaer (1895/1896) put forward the idea that a partial investigation (i.e. a sample) based on what he called the 'representative method' could provide useful information. The aim of his representative method, the new paradigm in statistical investigations, was that the sample should be an approximate miniature of the population. There had been earlier anticipations of this idea, e.g., Laplace's estimate of the population of France in 1802 (see Cochran, 1978, for a description of Laplace's methodology). However, it was Kiaer's suggestion of the idea and his subsequent campaign for its acceptance that brought about the revolution in data collection. What Kiaer meant by a 'representative sample' is best described by Kruskal and Mosteller (1980, p. 175):

'First, he thought of social and economic surveys in which one could begin by choosing districts, towns, parts of cities, streets, etc., to be followed by systematic,