handbook of Statistics 6

Sampling

Edited by

P. R. Krishnaiah†

C. R. Rao

Sampling

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

ISBN: 0 444 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.)
1. Sampling (Statistics) I. Krishnaiah, Paruchuri R. II. Rao, C. Radhakrishna (Calyampudi Radhakrishna), 1920- . III. Series.
QA276.6.S334 1988 88-374
519.5'2--dc19 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. Hidiroglou. 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 occuring 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. Hidiroglou, *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

P. R. Krishnaiah and C. R. Rao, eds., *Handbook of Statistics, Vol. 6*

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,