# EXHIBIT G

## Springer Series in Statistics

*Anderson:* Continuous-Time Markov Chains: An Applications-Oriented Approach.
*Andrews/Herzberg:* Data: A Collection of Problems from Many Fields for the Student and Research Worker.
*Anscombe:* Computing in Statistical Science through APL.
*Berger:* Statistical Decision Theory and Bayesian Analysis, 2nd edition.
*Brémaud:* Point Processes and Queues: Martingale Dynamics.
*Brockwell/Davis:* Time Series: Theory and Methods, 2nd edition.
*Daley/Vere-Jones:* An Introduction to the Theory of Point Processes.
*Dzhaparidze:* Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.
*Farrell:* Multivariate Calculation.
*Fienberg/Hoaglin/Kruskal/Tanur (Eds.):* A Statistical Model: Frederick Mosteller's Contributions to Statistics, Science, and Public Policy.
*Goodman/Kruskal:* Measures of Association for Cross Classifications.
*Grandell:* Aspects of Risk Theory.
*Hall:* The Bootstrap and Edgeworth Expansion.
*Härdle:* Smoothing Techniques: With Implementation in S.
*Hartigan:* Bayes Theory.
*Heyer:* Theory of Statistical Experiments.
*Jolliffe:* Principal Component Analysis.
*Kres:* Statistical Tables for Multivariate Analysis.
*Leadbetter/Lindgren/Rootzén:* Extremes and Related Properties of Random Sequences and Processes.
*Le Cam:* Asymptotic Methods in Statistical Decision Theory.
*Le Cam/Yang:* Asymptotics in Statistics: Some Basic Concepts.
*Manoukian:* Modern Concepts and Theorems of Mathematical Statistics.
*Miller, Jr.:* Simultaneous Statistical Inference, 2nd edition.
*Mosteller/Wallace:* Applied Bayesian and Classical Inference: The Case of *The Federalist Papers*.
*Pollard:* Convergence of Stochastic Processes.
*Pratt/Gibbons:* Concepts of Nonparametric Theory.
*Read/Cressie:* Goodness-of-Fit Statistics for Discrete Multivariate Data.
*Reiss:* Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics.
*Ross:* Nonlinear Estimation.
*Sachs:* Applied Statistics: A Handbook of Techniques, 2nd edition.
*Särndal/Swensson/Wretman:* Model Assisted Survey Sampling.
*Seneta:* Non-Negative Matrices and Markov Chains.
*Siegmund:* Sequential Analysis: Tests and Confidence Intervals.
*Tong:* The Multivariate Normal Distribution.
*Vapnik:* Estimation of Dependences Based on Empirical Data.
*West/Harrison:* Bayesian Forecasting and Dynamic Models.
*Wolter:* Introduction to Variance Estimation.

*(continued after index)*

Carl-Erik Särndal  Bengt Swensson
Jan Wretman

# Model Assisted Survey Sampling

Springer-Verlag
New York Berlin Heidelberg London Paris
Tokyo Hong Kong Barcelona Budapest

Carl-Erik Särndal
Département de mathématiques et de statistique
Université de Montréal
Montréal, Québec H3C 3J7
Canada

Bengt Swensson
Department of Data Analysis
University of Örebro
701 30 Örebro
Sweden

Jan Wretman
Department of Statistics
Stockholm University
106 91 Stockholm
Sweden

The work on this book was supported in part by Statistics Sweden.

With 8 illustrations.

Mathematics Subject Classification: 62D05

Library of Congress Cataloging-in-Publication Data
Särndal, Carl-Erik, 1937–
Model assisted survey sampling / Carl-Erik Särndal, Bengt Swensson, Jan Wretman.
p. cm. — (Springer series in statistics)
Includes bibliographical references and indexes.
ISBN 0-387-97528-4 (alk. paper)
1. Sampling (Statistics) I. Swensson, Bengt. II. Wretman, Jan Håkan, 1939– . III. Title. IV. Series.
QA276.6.S37 1991
001.4'222—dc20 91-7851

Printed on acid-free paper.



Production managed by Bill Imbornoni; manufacturing supervised by Bob Paella.
Typeset by Asco Trade Typesetting Ltd., Hong Kong.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-97528-4 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-97528-4 Springer-Verlag Berlin Heidelberg New York

# Preface

This text on survey sampling contains both basic and advanced material. The main theme is estimation in surveys. Other books of this kind exist, but most were written before the recent rapid advances. This book has four important objectives:

1. To develop the central ideas in survey sampling from the unified perspective of unequal probability sampling. In a majority of surveys, the sampling units have different probabilities of selection, and these probabilities play a crucial role in estimation.
2. To write a basic sampling text that, unlike its predecessors, is guided by statistical modeling in the derivation of estimators. The model assisted approach in this book clarifies and systematizes the use of auxiliary variables, which is an important feature of survey design.
3. To cover significant recent developments in special areas such as analysis of survey data, domain estimation, variance estimation, methods for nonresponse, and measurement error models.
4. To provide opportunity for students to practice sampling techniques on real data. We provide numerous exercises concerning estimation for real (albeit small) populations described in the appendices.

This book grew in part out of our work as leaders of survey methodology development projects at Statistics Sweden. In supervising younger colleagues, we repeatedly found it more fruitful to stress a few important general principles than to consider every selection scheme and every estimator formula as a separate estimation problem. We emphasize a general approach.

This book will be useful in teaching basic, as well as more advanced, university courses in survey sampling. Our suggestions for structuring such courses are given below. The material has been tested in our own courses in Montréal, leads to

$$\hat{\bar{y}}_{U_d \text{alt}} = \sum_{s_d} y_k/n_d = \bar{y}_{s_d} \tag{3.3.19}$$

This intuitively sound estimator is simply the mean of the $y$-values observed in the domain. The variance and the variance estimator are derived later with the aid of Result 5.8.1.

**Remark 3.3.3.** We note that several of the parameters encountered in cases 2 and 3 can be expressed as

$$\theta = \sum_U c_k y_k$$

where $c_1, \ldots, c_N$ are constants. For the population mean dealt with in case 2, $c_k$ has the known value $1/N$ for all $k$. In case 3, estimation of the domain total $\sum_{U_d} y_k$, we have $c_k = 1$ for $k \in U_d$ and $c_k = 0$ otherwise. Normally, the domain membership indicator $c_k$ is not known beforehand for all $k \in U$. But for elements $k$ in the sample $s$, it may be possible to determine $c_k$. This is sufficient for the $\pi$-estimation method to function. By contrast, for estimation of the domain mean, $\bar{y}_{U_d}$, in case 3, we have $c_k = 1/N_d$ for $k \in U_d$ and $c_k = 0$ otherwise. If $N_d$ is unknown, the $c_k$ remain unknown, even for elements $k$ appearing in the sample. This is why the $\pi$-estimation method fails when $N_d$ is unknown.

### 3.3.2. Simple Random Sampling with Replacement

Simple random sampling with replacement ($SIR$), considered briefly in Examples 2.9.1 and 2.9.2, is the ordered design that gives the same selection probability $1/N^m$ to every ordered sample

$$os = (k_1, \ldots, k_i, \ldots, k_m)$$

where $k_i$ is the element obtained in the $i$th draw. An element may be drawn more than once. The ordered sample contains at most $m$ distinct elements.

The $SIR$ design can be implemented by the draw-sequential scheme given in Example 2.9.1 with $m$ independent with-replacement draws, such that each draw gives any one element a chance of $1/N$ to be drawn. In $m$ draws, any given element will appear in the ordered sample a certain number of times, say $r$, such that $r$ is a binomially distributed random variable with mean $m/N$ and variance

$$\frac{m}{N}\left(1 - \frac{1}{N}\right) \doteq \frac{m}{N}$$

When $N$ is moderate to large, the Poisson distribution with mean $m/N$ will be an excellent approximation to the distribution of $r$. Under the $SIR$ design, the pwr estimator (see Section 2.9) for the population total $t = \sum_U y_k$ takes the form

$$\hat{t}_{\text{pwr}} = N\bar{y}_{os} \tag{3.3.20}$$



where $\bar{y}_{os}$ is the ordered sample mean, repeated elements included,

$$\bar{y}_{os} = \frac{1}{m}\sum_{i=1}^{m} y_{k_i} \tag{3.3.21}$$

Defining the ordered sample variance as

$$S_{os}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_{k_i} - \bar{y}_{os})^2 \tag{3.3.22}$$

we have the following result, which follows from Result 2.9.1.

**Result 3.3.4.** *Under the $SIR$ design, the pwr estimator of the population total $t = \sum_U y_k$ takes the form of equation* (3.3.20). *The variance is given by*

$$V_{SIR}(\hat{t}_{\text{pwr}}) = N(N-1)S_{yU}^2/m \tag{3.3.23}$$

*An unbiased variance estimator is*

$$\hat{V}_{SIR}(\hat{t}_{\text{pwr}}) = N^2 S_{os}^2/m \tag{3.3.24}$$

Let us compare sampling with and without replacement. If $n$, the sample size in the without-replacement case, equals $m$, the ordered sample size in the with-replacement case, then

$$\frac{V_{SIR}(N\bar{y}_{os})}{V_{SI}(N\bar{y}_s)} = \frac{1 - N^{-1}}{1-f} \doteq \frac{1}{1-f}$$

This shows that the two estimation strategies are roughly of equal efficiency when the sampling fraction $f = n/N$ is very small. On the other hand, if $f$ is substantial, considerable efficiency is lost in with-replacement sampling. For example, if $f = 50\%$, the with-replacement variance is double the without-replacement variance.

**Remark 3.3.4.** For the $SIR$ design, there exist other unbiased estimators than $N\bar{y}_{os}$. These are discussed in Section 3.8.

## 3.4. Systematic Sampling

### 3.4.1. Definitions and Main Result

Systematic sampling refers to a set of procedures that offer several practical advantages, particularly its simplicity of execution. We concentrate on systematic sampling in its basic form. A first element is drawn at random, and with equal probability, among the first $a$ elements in the population list. The positive integer $a$ is fixed in advance and is called the *sampling interval.* No