A Chain Ratio-Type Estimator in Finite Population Double Sampling Using Two Auxiliary Variables

By B. Kiregyera, Kampala

Summary: In this paper we construct a chain ratio-type estimator using two auxiliary variables. The performance of the constructed estimator relative to the simple mean, ratio-type estimate based on double sampling and Chand's ratio-type estimator is investigated. A numerical illustration is given.

1. Introduction

Use of auxiliary information is a common phenomenon in sampling theory of surveys. This information is used at planning stage of a survey thereby leading to a better choice of sample design or it is used at estimation stage thereby leading to better choice of estimator, viz. ratio and regression estimators of population characteristics. We concern ourselves with the latter case.

Let \( y \) be a study variable and \( x \), the auxiliary variable. Suppose \( n \) pairs \( (x_i, y_i) \) \((i = 1, 2, \ldots, n)\) of observations are taken on \( n \) units sampled from \( N \) population units using simple random sampling (SRS) design without replacement.

The classical ratio-type estimator of population mean \( \bar{Y} \) is

\[
\hat{Y}_R = \frac{\bar{y}n}{\bar{x}_n} \bar{X}
\]

where

\[
\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{and} \quad \bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i.
\]

It will be taken that \( x_i > 0 \) for \( i = 1, 2, \ldots, N \) and \( \bar{X} \) is assumed known a priori.

It is a well known result [e.g. Cochran] that \( \hat{Y}_R \) will estimate \( \bar{Y} \) to \( O(1/n) \) more precisely than \( \bar{y}_n \) if \( \rho_{y,x} > 1/2 \frac{C_x}{C_y} \); where \( \rho_{y,x} \) is the correlation coefficient between \( x \) and \( y \), and \( C_x, C_y \) are coefficients of variation of \( x, y \) respectively.

Often, however, \( \bar{X} \) is not known before the start of a survey. The usual practice is to estimate it by the simple mean \( \bar{x}_n' = (1/n') \sum_{i=1}^{n'} x_i \), where \( n' \) is the size of a prelimi-

\(^1\) Ben Kiregyera. Institute of Statistics and Applied Economics, Makerere University, P.O. Box 7062, Kampala, Uganda, E. Africa.
nary simple random sample drawn from $N$ without replacement and $n < n'$ is a sub-sample of $n'$. Then the ratio-type estimator in (1) is modified as

$$\hat{Y}_{Rd} = \frac{\bar{y}_n}{\bar{x}_n} \bar{x}_{n'}$$

(2)

and its approximate bias and MSE to 0 ($1/n$) are given [e.g. Sukhatme] as

$$\text{Bias}_1 (\hat{Y}_{Rd}) = K_1 \bar{Y} [C_{20} - C_{11}]$$

(3)

and

$$\text{MSE}_1 (\hat{Y}_{Rd}) = \bar{Y}^2 [K_1 (C_{20} + C_{02} - 2C_{11}) + K_2 C_{02}]$$

(4)

where

$$K_1 = \frac{n' - n}{nn'}, \quad K_2 = \frac{N - n'}{n'N}$$

and

$$C_{rs} = \frac{E(x_i - \bar{X})^r (y_i - \bar{Y})^s}{\bar{X}^r \bar{Y}^s}, \quad r \geq 0, \quad s \geq 0.$$ 

Again subject to the condition that $\rho_{yx} > (1/2) C_x/C_y$, $\hat{Y}_{Rd}$ will be more precise than $\bar{y}_n$ for estimating $\bar{Y}$.

Suppose that $\bar{Z}$, the population mean of another variable $z$ closely related to $x$ but compared to $x$ remotely related to $y$ is available (e.g. $y$ is value of cattle and/or calves sold live in 1964, $x$ is the number of cattle and/or calves sold in 1964 and $z$ is the number of farms reporting sale of cattle and/or calves sold live in 1964, $\rho_{yx} > \rho_{yz}$). Then, as argued before

$$\hat{X}_{Rd} = \frac{\bar{x}_{n'}}{\bar{z}_{n'}} \bar{z}$$

(5)

will estimate $\bar{X}$ to $0(1/n)$ more precisely than $\bar{x}_{n'}$ if $\rho_{xz} > (1/2) C_z/C_x$. Accordingly, Chand [1975] has proposed a chain ratio-type estimator

$$\hat{Y}_{2d} = \frac{\bar{y}_n}{\bar{x}_n} \hat{X}_{Rd}$$

(6)

where subscript 2 on $\hat{Y}_{2d}$ indicates that two auxiliary variables are used. When $\bar{Z}$ is not known a priori, but data on $z$ can be obtained easily and cheaply, then it is estimated from a larger sample $n''$.

The sampling model involving $n$ and $n'$ is as follows: Select $n$ units from the population and observe $y$, $x$ and $z$, from the remaining $(N - n)$ units select $(n' - n)$ units and observe $x$ and $z$. These selections are done by SRS without replacement $| 1 |$.

The approximate expressions for bias and MSE of $\hat{Y}_{2d}$ to $0(1/n)$ are