Optimum Replacement Policy at Current Occassion under Fixed Cost and Fixed Precision Requirements

Sandeep

Centre for Women Development Studies, New Delhi-01

Abstract—*The present paper deals with the estimation of finite population mean using ratio method of estimation when two phase sampling scheme is used. The optimum replacement policy and optimum number of units to be selected a fresh on the current occasion is discussed for regression and ratio estimators in cases when cost of the survey is fixed and when the precision is fixed separately. The adoption of these estimators has also been discussed under different situations.*

Keywords*: Successive sampling, Ratio estimator, Regression estimator, Optimum replacement policy, Cost analysis.*

1. INTRODUCTION

In many situations the study character of a finite population changes over time. A survey over a single occasion does not provide any information on the nature or rate of change of the characteristic or the average value of the characteristic at the current occasion. To meet this requirement, a portion of the sample taken on the previous occasion is retained and a few new units of the population are selected from the population to complete the sample. This sampling is known as successive sampling and it provides with a strong tool to obtain reliable estimates of population parameter at the current occasion.

Theory of successive sampling started with the work of Jessen (1942) where the he collected information on the previous occasions and later this theory was extended by the Patterson (1950), Rao and Graham (1964), Gupta (1979), Das (1982), among others. Sen (1971, 1972,1973), Singh et. al. (1991), Singh and Singh (2001), Singh (2003), Biradar and Singh (2001), Singh(2005), Singh and Priyanka (2006, 2008) used the auxiliary information on current or on both the occasions to estimate the population mean at the current occasions in two occasions successive sampling.

Generally the regression estimator is used to formulate the first estimator from the information available from the sample taken on the previous occasion and formed the matched (common) portion of the sample selected on the second occasion. However, in many practical situations it is more favourable to use the ratio estimator not only on the grounds of efficiency but also on account of ease and simplicity in its calculations.

Another advantage of using ratio estimator over regression estimator is that the optimum matched portion, the portion which minimizes the variance of the pooled estimator, is larger in case of ratio estimator than of optimum matched portion in case of the regression estimator.

Kulldr off (1963) had considered the question of cost in case of difference estimators. In this article ratio and regression estimators in successive sampling with two occasions under fixed cost and fixed efficiency are examined and suggestions are made for their possible adaption in different situations.

Let $U = U_1, U_2, \cdots, U_N$ be a finite population of size N which is available for sampling over two occasions. The characteristic is denoted by $x(y)$ on the first (second) occasion respectively. It is assumed that the population is considerably large. A simple random sample of size n is selected without replacement from the population on the first occasion. A random sub-sample of m units is retained (matched) out of n units selected at the first occasion for use on the second occasion while a fresh simple random sample of u units is selected without replacement from the population of $N-n$ units on the second occasion. Here u is different from $n-m$ as its value depends on the precision requirement or on the given budget.

On the basis of the information available on two occasions, some statistics are defined as follows:

$\bar{x}(n) = (1/n)\sum_{i=1}^{n} x_i$; $\bar{y}(u) = (1/u)\sum_{i=1}^{u} y_i$ are the sample means on the first occasion and unmatched portion of the sample drawn on the second occasion. Similarly, $\bar{x}(m) = (1/m)\sum_{i=1}^{m} x_i$, and $\bar{y}(m) = (1/m)\sum_{i=1}^{m} y_i$ are the sample means based on the matched portion of the sample on the first and second occasions respectively. The corresponding sample variances, covariance and the regression coefficient for the matched portion of the sample are given by:

$$S_x^2(m) = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x}(m))^2$$

$$S_y^2(m) = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y}(m))^2$$

$$S_{xy}(m) = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x}(m))(y_i - \bar{y}(m))$$

$$b(m) = \frac{S_{xy}(m)}{S_x^2(m)}$$

To estimate the population mean $T = \bar{y}$ on the second occasion, we use the combined estimator as regression estimator

$$\hat{T}_{lr} = \frac{u}{k+u}T(u) + \frac{k}{k+u}T(m)$$

where one estimator is $T(u)$ which is the ordinary sample mean $\bar{y}(u)$ on the basis of sample of size n drawn afresh on the second occasion and the second estimator is $T(m)$

$$T(m) = \bar{y}(m) + b(m)\big(\bar{x}(n) - \bar{x}(m)\big)$$

and

$$k = \left[\frac{1-\rho^2}{m} + \frac{\rho^2}{n}\right]^{-1}$$

and u is the number of unmatched units drawn on the second occasion. The variance of $\hat{T}_{lr}$ is given by

$$V\big(\hat{T}_{lr}\big) = \frac{1}{k+u}S_y^2$$

However, if we use the ratio estimator, the pooled estimator is given by

$$\hat{T}_R = \frac{u}{p+u}T(u) + \frac{p}{p+u}T(m)$$

where

$$T(u) = \bar{y}(u)$$

and

$$T(m) = \frac{\bar{y}(m)}{\bar{x}(m)}\bar{x}(n)$$

$$\frac{1}{p} = \frac{1-t}{m} + \frac{t}{n}$$

$$t = 2\rho\frac{\mathrm{CV}(X)}{\mathrm{CV}(Y)} - \left[\frac{\mathrm{CV}(X)}{\mathrm{CV}(Y)}\right]^2$$

and the variance of $\hat{T}_R$ is given by

$$V\big(\hat{T}_R\big) = \frac{1}{p+u}S_y^2$$

The important aspect in sampling on two occasions is the question of cost. It is generally more expensive to obtain information about the new units than about units already surveyed on the first occasions. It may be due to the availability of supplementary information on the first occasion that may reduce the travel cost etc on the second occasion, or due to the quick availability of information from the units on the second occasion due to the familiarity of the respondents will the subjects or due to the information on the second occasion may also be collected on the basis of mail questionnaire.

2. REPLACEMENT POLICY AND CHOICE OF ESTIMATORS UNDER FIXED COST

Let us consider the cost function structure

$$c = c_0 + c_1 m + c_2 u \qquad (1)$$

where c is the total cost, c_0 is the overhead cost, c_1 and c_2 are the unit costs for matched and unmatched units respectively. This cost function may be written as

$$\frac{c-c_0}{c_2} = R = \frac{c_1}{c_2}m + u = dm + u$$

Generally, $d \le 1$. Now, the problem is to minimize the variance of $\hat{T}_R$ or of $\hat{T}_{lr}$ or to maximize $\Phi = \big[V\big(\hat{T}_R\big)\big]^{-1} = (p+u)/S_y^2$ or $\Phi = \big[V\big(\hat{T}_{lr}\big)\big]^{-1} = (k+u)/S_y^2$ subject to fixed R for the choice of matched portion m. Thus for ratio estimator we defferentiatiate

$$\Phi = \left[\frac{mn}{n(1-t)+mt} + R - dm\right]\frac{1}{S_y^2} \qquad (2)$$

with respect to m and equating it to zero, yields the following result:

Theorem 2.1. Given R as a fixed cost of the survey the optimum matched portion and the optimum number of unmatched units drawn on the second occasion for the ratio estimator $\hat{T}_R$ for estimating the mean at the current occasion, so that $V\big(\hat{T}_R\big)$ is minimum, is (3)

Case I. $0 \le d \le 1 - t$

$$m_{opt} = n = p_{opt}$$

$$u_{opt} = R - dn$$

$$V_{min}(\hat{T}_R) = \frac{S_y^2}{R + n(1-d)}$$

Case II. $0 \le 1 - t \le d$

$$m_{opt} = \frac{n}{t}\left[\left\{\frac{1-t}{d}\right\}^{1/2} - (1-t)\right]$$

$$u_{opt} = R - dm_{opt} \qquad (4)$$

$$p_{opt} = \frac{n}{t}\left[1 - \{d(1-t)\}^{1/2}\right]$$

$$V_{min}\big(\hat{T}_R\big) = \frac{S_y^2}{R + \frac{n}{t}[1 - \{d(1-t)\}^{1/2}]^2}$$

Theorem 2.2. Given R as a fixed cost of the survey and the use of the regression estimator $\hat{T}_{lr}$ for estimating population mean at the current occasion, the optimum matched portion and the optimum number of the unmatched units drawn on the second occasion, so that $V(\hat{T}_{lr})$ is minimum, is

Case I. $0 \le d \le 1-\rho^2$

$$m_{opt} = n = k_{opt}$$

$$u_{opt} = R - dn$$

$$V_{min}(\hat{T}_{lr}) = \frac{S_y^2}{R + n(1-d)}$$

Case II. $0 \le 1-\rho^2 \le d$

$$m_{opt} = \frac{n}{\rho^2}\left[\left\{\frac{1-\rho^2}{d}\right\}^{1/2} - (1-\rho^2)\right]$$

$$u_{opt} = R - d\, m_{opt}$$

$$p_{opt} = \frac{n}{\rho^2}\left[1 - \{d(1-\rho^2)\}^{1/2}\right]$$

$$V_{min}(\hat{T}_{lr}) = \frac{S_y^2}{R + \frac{n}{\rho^2}[1 - \{d(1-\rho^2)\}^{1/2}]^2}$$

Theorem 2.3. Under the conditions of theorem 2.1 and 2.2,

Case I if $d \le 1-\rho^2$ and $d \le 1-t$ the estimators $\hat{T}_{lr}$ and $\hat{T}_R$ are equally efficient.

Case II if $0 \le 1-t \le d$ and $0 \le 1-\rho^2 \le d$ and assuming the coefficient of variation is stable over time (the characteristic X and Y are are of similar nature) (Murthy, 1977), $t = 2\rho - 1$ and if $\rho \ge 0.5$

$$\left[d^{1/2} - (a+b)\right]\left[d^{1/2} - (a-b)\right] \le 0$$

where

$$a = \frac{(1-\rho)^{1/2}}{(1-\rho)^2}\left[2^{1/2}\rho^2 - (2\rho-1)(1+\rho)^{1/2}\right]$$

$$b = \frac{1}{(1-\rho)^2}\left[a^2 - (1-\rho)^4\right]^{1/2}$$

$\hat{T}_{lr}$ is more efficient than $\hat{T}_R$.

Case III if $1-\rho^2 \le d \le 1-t, \hat{T}_{lr}$ is more efficient than $\hat{T}_R$.

3. REPLACEMENT POLICY AND CHOICE OF ESTIMATORS UNDER FIXED VARIANCE

Let us fix the precision of the estimators $\hat{T}_R$ and $\hat{T}_{lr}$ at a given level by fixing $p + u = z^0 = p + k$ in their variances. We minimize the cost function

$$R = dm + u = dm + z^0 - p$$

$$= dm + z^0 - \frac{mn}{n(1-t) + mt} \quad (11)$$

subject to $p + u = z^0$. By equating the derivative of (11) with respect to m and equating it to zero, we get the following result:

Theorem 3.1 Assuming the precision of the ratio estimator $\hat{T}_R$ is fixed by fixing $p + u = z^0$, the optimum matched portion for $\hat{T}_R$ is

(9)

Case I: $0 \le d \le 1-t$

$$m_{opt} = n = p_{opt}$$

$$R_{min} = z^0 - n(1-d)$$

Case II: $0 \le 1-t \le d$

$$m_{opt} = \frac{n}{t}\left[\left(\frac{1-t}{d}\right)^{1/2} - (1-t)\right]$$

$$p_{opt} = \frac{n}{t}\left[1 - \{d(1-t)\}^{1/2}\right] \quad (10)$$

$$R_{min} = dm_{opt} + z^0 - p_{opt}$$

$$= z^0 - \frac{n}{t}\left[1 - \{d(1-t)\}^{1/2}\right]^2$$

Theorem 3.2. Assuming the precision of the regression estimator $\hat{T}_{lr}$ is fixed by fixing $k + u = z^0$, the optimum matched portion for $\hat{T}_{lr}$ is given by

Case I. $0 \le d \le 1-\rho^2$

$$m_{opt} = n = k_{opt}$$

The minimum cost is given by

$$R_{min} = z^0 - n(1-d)$$

Case II $0 \le 1-\rho^2 \le d$

$$m_{opt} = \frac{n}{\rho^2}\left[\left(\frac{1-\rho^2}{d}\right)^{1/2} - (1-\rho^2)\right]$$

$$k_{opt} = \frac{n}{\rho^2}\left[1 - \{d(1-\rho^2)\}^{1/2}\right]$$

The minimum cost is given by

$$R_{min} = z^0 - \frac{n}{\rho^2}\left[1 - \{d(1-\rho^2)\}^{1/2}\right]^2$$

Theorem 3.3. Under the conditions of theorem 3.1 and 3.2,

Case I. if $d \le 1-\rho^2$ and $d \le 1-t$,

minimum cost for $\hat{T}_R$ and $\hat{T}_{lr}$ are equal.

Case II. if $d \le 1-\rho^2$ and $d \ge 1-t$, it is less expensive to record $\hat{T}_{lr}$ as compared $\hat{T}_R$ if

$$\frac{1}{t}\left[1 - \{d(1-t)\}^{1/2}\right]^2 \le \frac{1}{\rho^2}\left[1 - \{d(1-\rho^2)\}^{1/2}\right]^2$$

The same kind of argument as given in theorem show that $\hat{T}_{lr}$ is less expensive.

Case III. $1-\rho^2 \leq d \leq 1-t$

It is more economical to record $\hat{T}_{lr}$ as compared to $\hat{T}_R$.

The efficiencies and cost for observing the estimators $\hat{T}_R$ and $\hat{T}_{lr}$ depend upon the values of d and ρ^2. These estimators are equally efficient and bear same cost in recording them if $d \leq 1-\rho^2$ and $d \leq 1-t$. It is thus, when ρ is high and d is small, to prefer the ratio estimator $\hat{T}_R$ over regression estimator $\hat{T}_{lr}$ on the grounds of simplicity in calculating $\hat{T}_R$. In other cases, $\hat{T}_{lr}$ is more efficient and less costly than $\hat{T}_R$. The relative efficiency and cost of the two estimators can be worked out to see whether the gain efficiency and cost of $\hat{T}_{lr}$ is marginal or substantial. If it is marginal $\hat{T}_R$ is preferred to $\hat{T}_{lr}$ otherwise $\hat{T}_{lr}$ is preferred.

REFERENCES

[1] **Biradar, R.S., Singh, H.P**. (2001): Successive sampling using auxiliary information on both occasions, *Cal. Statist. Assoc. Bull.* **51** : 243-251.

[2] **Chaturvedi, D. K., Tripathi, T.P. (1983)**: Estimation of population ratio on two occasions using multivariate auxiliary information. *J. Ind. Statist.Assoc.* 21:113-120.

[3] **Gonzalez, M.E., Ogus, J.L., Shapiro, O.G., and Tepping, B.J.** (1975), "Standards for discussion and presentation of errors in survey and census data " in *Journal of American Statistical Association*, **70**(351), part II, 5-23.

[4] **Lepkovski, (1998),** "Presentation of Sampling Errors", Methods of Survey Sampling/Applied Sampling-Lecture at Statistics South Africa.

[5] **Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., Ashok, C. (1984),** *Sampling Theory of Surveys with Applications*, Indian Society of Agricultural Statistics, New Delhi, India.