-

A GENERALIZED STUDY IN SUCCESSIVE SAMPLING ON TWO OCCASIONS

MANOJ KUMAR SRIVASTAVA[1], NAMITA SRIVASTAVA[2]

[1]Institute of Social Sciences, Dr. B. R. Ambedkar University, Agra – 282002, INDIA.
[2]Department of Statistics, St. John's College, Agra – 282002, INDIA.
[1]mksiss87@gmail.com, [2]drnamita.sjc@gmail.com

ABSTRACT

Utilizing observations on two occasions for some units from the units selected at the first occasion, a generalized estimator representing a class of estimators including ratio and regression estimators as special cases, is considered and which is further combined with the usual mean per unit estimator based on freshly selected unmatched units from the population at the second occasion to get a combined generalized estimator of the population mean under successive sampling on two occasions. The results of some earlier estimators are shown to be the special cases of those of this combined generalized study in successive sampling on two occasions. Further, a comparative study of the combined generalized estimator is also made with some of the estimators in the literature regarding successive sampling on two occasions.

Keywords: Combined generalized estimator, successive sampling, two occasions, mean square error and efficiency.

1. INTRODUCTION

In repeated sample surveys the same characteristic is measured on the units of the given population over various occasions. The population is assumed to remain same except for some avoidable but insignificant changes which might occur over time. If we stick to only two occasions then while sampling for the second occasion, a portion of the sample taken on the previous occasion is retained and a few new units are selected from the population to have a complete sample. In practice the sample size on both occasions is assumed to be same. Given such information two separate estimators of the mean of the population characteristic are formulated. The first estimator is based on the information available on the units in sample on the previous occasion and the new information obtained at the second occasion on the common units. The second estimator is based on the information collected for the new units selected in the sample on the second occasion. Having obtain these two estimators in this way a combined generalized estimator

$$\bar{y}_f = \bar{y}_{n_2'}\, f\left(\frac{\bar{x}_{n_2'}}{\bar{x}_n}\right) + \beta_{21}(\bar{x}_n^a - \bar{x}_{n_2'}^a)$$

is proposed for the current population mean on successive sampling on two occasions. The weights assigned to the estimator are reciprocals of their respective variances.

Generally the regression estimator is used to formulate the first estimator from the information available out of 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 conditions may be more favourable for the use of ratio estimator. The use of ratio estimator is recommended under such conditions not only on efficiency grounds but also on account of ease and simplicity in its calculations.

Another advantage of using ratio estimator instead of regression estimator is that the optimum matched portion (i.e. portion which minimizes the variance of the pooled estimator) in case of ratio estimator is larger than optimum matched portion in case of the regression estimator. Since it is less expensive to obtain the information on the units in the matched portion, the economic efficiency of the

-

proposed estimator over the estimator using regression method might turn out to be higher in many practical situations.

2. FORMULATION OF THE ESTIMATOR

Let $U_1, U_2, \dots, U_N$ be a finite population of size N which is available for sampling over two occasions. A simple random sample of size n has been considered on the first occasion. A random subsample of size n_2' units is retained for use on the second occasion while a simple random sample without replacement of $n_2'' = n - n_2'$ units is drawn from the second occasion from the remaining population of $N - n$ units.
$\bar{Y}$: population mean of the study variable on the current occasion
$\bar{x}_n$: sample mean of the study variable on the first occasion
$\bar{x}_{n_2'}$:sample mean of the study variable over n_2' matched units(common to both the occasions) on the first occasion
$\bar{y}_{n_2'}$:sample mean of the study variable over n_2' matched units(common to both the occasions) on the second(current) occasion
$\bar{y}_{n_2''}$:sample mean of the study variable over n_2'' units drawn afresh on the second(current) occasion
S_1^2: population mean square of the study variable on the first occasion
S_2^2: population mean square of the study variable on the second occasion
β_{21}: regression coefficient of the study variable on the second occasion on the study variable on the first occasion

3. THE PROPOSED ESTIMATOR

To estimate the population mean $\bar{Y}_2$ on the current occasion, we have two estimators one is $\bar{y}_{n_2''}$ based on n_2'' units drawn afresh on the second occasion, and a generalized estimator that we propose

$$\bar{y}_f = \bar{y}_{n_2'}\, f\left(\frac{\bar{x}_{n_2'}}{\bar{x}_n}\right) + \beta_{21}(\bar{x}_n^a - \bar{x}_{n_2'}^a) \quad (1)$$

where f is some function that satisfies the condition that

$$f(u)|_{u=1} = 1 \quad (2)$$

β_{21} is the regression coefficient of the study variable on the second on the study variable on the first occasion, a is scaling constant.
This proposed estimator converts in the usual ratio, regression and product estimators in the following manner:

Case-I
If $f(u) = 1$ and $a = 1$, then $\bar{y}_f$ takes the form

$$\bar{y}_f = \bar{y}_{n_2'} + \beta_{21}(\bar{x}_n - \bar{x}_{n_2'}) \quad (3)$$

which is the usual regression estimator.

Case-II
If $f(u) = u$ and $a = 0$, then $\bar{y}_f$ takes the form

$$\bar{y}_f = \bar{y}_{n_2'}\, \frac{\bar{x}_{n2}'}{\bar{x}_n} \quad (4)$$

which is the product estimator.

Case-III
If $f(u) = u^{-1}$ and $a = 0$, then $\bar{y}_f$ takes the form

$$\bar{y}_f = \frac{\bar{y}_{n_2'}}{\bar{x}_{n_2'}}\, \bar{x}_n \quad (5)$$

which is the ratio estimator.Therefore, the proposed estimator is a general estimator that includes regression, ratio and product estimators and many more as its special cases.
It is important to mention the all above three cases satisfy the condition (2).

-

The estimators $\bar{y}_{n_2''}$ and $\bar{y}_f$ are unbiased and biased estimators of $\bar{Y}$ respectively.
The mean square error of the estimator $\bar{y}_f$, up to the first order of approximation and for large population size N, is given in the following theorem:

Theorem 3.1: Under the assumption that the population is large, the variance of the estimator $\bar{y}_{n_2''}$ and the mean square error of the estimator $\bar{y}_f$ up to the first order of approximation is

$$MSE(\bar{y}_f) = \frac{1}{n_2'}\bar{Y}^2C_2^2 + \left(\frac{1}{n_2'} - \frac{1}{n}\right)\bar{Y}^2C_2^2[\{f'(1)\}^2 + 2\rho f'(1)] + \left(\frac{1}{n_2'} - \frac{1}{n}\right)\bar{Y}^2C_2^2[a^2\rho_{21}^2\bar{X}^{2a-2} - 2a\rho_{21}^2\bar{X}^{a-1} - a\rho_{21}\bar{X}^{a-1}f'(1)] \quad (6)$$

The MSE(variance) of $\bar{y}_{n_2''}$ is

$$MSE(\bar{y}_{n_2''}) = \frac{S_2^2}{n_2''} = \frac{\bar{Y}C_2^2}{n_2''} \quad (7)$$

Proof:
Assume that the study variable is equally variable on both the occasions i.e. $S_1^2 = S_2^2$, and consider the large sample approximations

$$\bar{y}_{n_2'} - \bar{Y} = e_0'$$
$$\bar{x}_{n_2'} - \bar{X} = e_1'$$
$$\bar{x}_n - \bar{X} = e_1$$
$$s_{xy} - S_{xy} = e_2'$$
$$s_x^2 - S_x^2 = e_3'$$

Now expanding the function f by Taylor expansion and putting the above approximations

$$\bar{y}_f = \bar{y}_{n_2'} f\left(\frac{\bar{x}_{n_2'}}{\bar{x}_n}\right) + \beta_{21}(\bar{x}_n^a - \bar{x}_{n_2'}^a)$$

$$= (\bar{Y} + e_0')\left[1 + \left\{\left(1 + \frac{e_1'}{\bar{X}}\right)\left(1 + \frac{e_1}{\bar{X}}\right)^{-1} - 1\right\}f'(1)\right] + \beta_{21}\bar{X}^a\left\{\left(1 + \frac{e_1}{\bar{X}}\right)^a - \left(1 + \frac{e_1'}{\bar{X}}\right)^a\right\}$$

$$\bar{y}_f = (\bar{Y} + e_0')\left[1 + \left(\frac{e_1' - e_1}{\bar{X}}\right)f'(1)\right] + a\beta_{21}\bar{X}^a\left\{\frac{e_1 - e_1'}{\bar{X}}\right\}$$

$$\bar{y}_f - \bar{Y} = e_0' + \frac{\bar{Y}}{\bar{X}}(e_1' - e_1)f'(1) + a\beta_{21}\bar{X}^{a-1}(e_1 - e_1')$$

Now squaring both the sides and taking expectation gives means square error of $\bar{y}_f$ under the assumption $S_1^2 = S_2^2$, $C_1 = C_2$

$$MSE(\bar{y}_f) = \frac{1}{n_2'}\bar{Y}^2C_2^2 + \left(\frac{1}{n_2'} - \frac{1}{n}\right)\bar{Y}^2C_2^2[\{f'(1)\}^2 + 2\rho f'(1)] + \left(\frac{1}{n_2'} - \frac{1}{n}\right)\bar{Y}^2C_2^2[a^2\rho_{21}^2\bar{X}^{2a-2} - 2a\rho_{21}^2\bar{X}^{a-1} - a\rho_{21}\bar{X}^{a-1}f'(1)]$$

Theorem 3.2: The combined estimator of $\bar{Y}$ based on the matched sample common to both the occasions and fresh sample on the second occasion is

$$\hat{T} = \varphi\bar{y}_{n_2''} + (1 - \varphi)\bar{y}_f \quad (9)$$

with

$$MSE(\hat{T}) = \varphi^2 MSE(\bar{y}_{n_2''}) + (1 - \varphi)^2 MSE(\bar{y}_f) \quad (10)$$

when $MSE(\bar{y}_{n_2''})$ and $MSE(\bar{y}_f)$ are taken from theorem 3.1.

4. MINIMUM MEAN SQUARE ERROR OF THE PROPOSED ESTIMATOR

Since the $MSE(\hat{T})$ is a function of φ, therefore, $MSE(\hat{T})$ is minimized with respect to φ to get the optimum values

-

$$\varphi_{opt} = \frac{MSE(\bar{y}_f)}{MSE(\bar{y}_{n_2''}) + MSE(\bar{y}_f)} \quad (11)$$

and

$$MSE(\hat{T})_{\varphi_{opt}} = \frac{MSE(\bar{y}_{n_2''})MSE(\bar{y}_f)}{MSE(\bar{y}_{n_2''}) + MSE(\bar{y}_f)} \quad (12)$$

Theorem 4.1: Under the assumption that $S_1^2 = S_2^2$

$$\varphi_{opt} = \frac{\mu(1-\mu c)}{(1-\mu^2 c)} \quad (13)$$

$$MSE(\hat{T})_{\varphi_{opt}} = \frac{1-\mu c}{1-\mu^2 c}\frac{\bar{Y}^2 C_2^2}{n} \quad (14)$$

where

$$-c = [\{f'(1)\}^2 + 2\rho f'(1) + a^2\rho_{21}^2\bar{X}^{2a-2} - 2a\rho_{21}^2\bar{X}^{a-1} - a\rho_{21}\bar{X}^{a-1}f'(1)]$$

Proof: Under the assumption $S_1^2 = S_2^2$, $MSE(\bar{y}_f)$ is given by

$$MSE(\bar{y}_f) = \frac{1}{n_2'}\bar{Y}^2C_2^2 + \left(\frac{1}{n_2'} - \frac{1}{n}\right)\bar{Y}^2C_2^2[\{f'(1)\}^2 + 2\rho f'(1)] + \left(\frac{1}{n_2'} - \frac{1}{n}\right)\bar{Y}^2C_2^2[a^2\rho_{21}^2\bar{X}^{2a-2} - 2a\rho_{21}^2\bar{X}^{a-1} - a\rho_{21}\bar{X}^{a-1}f'(1)]$$

Thus the minimum $MSE(\hat{T})$ is given by

$$MSE(\hat{T})_{opt} = \frac{MSE(\bar{y}_{n_2''})MSE(\bar{y}_f)}{MSE(\bar{y}_{n_2''}) + MSE(\bar{y}_f)} = \frac{1-\mu c}{1-\mu^2 c}\frac{\bar{Y}^2C_2^2}{n}$$

Remark: If there is complete matching i.e. $\mu = 0$ $(n_2' = n, n_2'' = 0)$

$$MSE(\hat{T})_{opt} = \frac{S_2^2}{n} = MSE(\bar{y}_f)|_{n_2'=n}$$

and that if there is no matching i.e. $\mu = 1$ $(n_2' = 0, n_2'' = n)$

$$MSE(\hat{T})_{opt} = \frac{S_2^2}{n} = MSE(\bar{y}_f)|_{n_2''=n}$$

Note that in both the extreme situations $MSE(\hat{T})_{opt}$ is identical. Thus, one should choose an optimal value of μ so that $MSE(\hat{T})_{opt}$ attains its minimum value.

5. OPTIMUM REPLACEMENT POLICY

To decide the optimum value of μ i.e sampling fraction to be taken at the current occasion to estimate $\bar{Y}$ by $\hat{T}$, we minimize $MSE(\hat{T})_{opt}$ with respect to μ

$$\frac{\partial}{\partial\mu}MSE(\hat{T})_{opt} = 0$$

$$-(1-\mu^2 c) - (1-\mu c)(-2\mu c) = 0$$

$$-c + \mu^{2c} + 2\mu c - 2\mu^2 c^2 = 0$$

$$-\mu^2c^2 + 2\mu c - c = 0$$

$$\mu = \frac{2 \pm \sqrt{(4-4c)}}{2c}$$

$$\mu = \frac{1-\sqrt{(1-c)}}{c}\frac{1+\sqrt{(1-c)}}{1+\sqrt{(1-c)}}$$

$$\mu = \frac{1}{(1+\sqrt{(1-c)})} \quad (15)$$

-

The optimum variance is given by

$$MSE(\hat{T})_{opt}\Big|_{\mu=\mu_0} = \left(1+\sqrt{1-c}\right)\frac{(\bar{Y}C_2)^2}{2n} \qquad (16)$$

Percent relative efficiency of $\hat{T}$ with respect to $\bar{y}_n$ the sample mean of observations on the n units drawn on the second occasion

$$E = \frac{MSE(\bar{y}_n)}{MSE(\hat{T})}\times 100 = \frac{2}{1+\sqrt{1-c}}\times 100 \qquad (17)$$

Special cases of the generalized estimator and related results

(i) $f(u) = 1$ and $a = 1$ we have $\bar{y}_f$ as a regression estimator.

In this case $f'(1) = 0$ and gives

$$MSE(\bar{y}_f) = \left(\frac{1}{n_2'}-\frac{1}{N}\right)S_2^2 + \left(\frac{1}{n_2'}-\frac{1}{n}\right)[\beta_{21}^2 S_1^2 - 2\beta_{21}S_{12}]$$

$$= \left(\frac{1}{n_2'}-\frac{1}{N}\right)S_2^2 + \left(\frac{1}{n_2'}-\frac{1}{n}\right)\rho^2 S_2^2$$

$$= \left(\frac{1}{n_2'}-\frac{1}{n}\right)S_2^2 + \left(\frac{1}{n}-\frac{1}{N}\right)S_2^2 + \left(\frac{1}{n_2'}-\frac{1}{n}\right)\rho_{21}^2 S_2^2$$

$$= \left(\frac{1}{n_2'}-\frac{1}{n}\right)S_2^2(1-\rho_{21}^2) + \left(\frac{1}{n}-\frac{1}{N}\right)S_2^2$$

$$-c = [\{f'(1)\}^2 + 2\rho f'(1) + a^2\rho_{21}^2\bar{X}^{2a-2} - 2a\rho_{21}^2\bar{X}^{a-1} - a\rho_{21}\bar{X}^{a-1}f'(1)]$$

$$= \rho_{21}\{\rho_{21} - 2\rho_{21}\} = -\rho_{21}^2$$

This gives

$$\mu_{op} = \frac{1}{1+\sqrt{1-\rho_{21}^2}}$$

$$MSE(\hat{T})_{opt}\Big|_{\mu=\mu_0} = \left(1+\sqrt{1-\rho_{21}^2}\right)\frac{(\bar{Y}C_2)^2}{2n}$$

$$E = \frac{MSE(\bar{y}_n)}{MSE(\hat{T})}\times 100 = \frac{2}{1+\sqrt{1-\rho_{21}^2}}\times 100$$

Next if we consider

$f(u) = u$ and $a = 0$

we get $\bar{y}_f$ as a ratio estimator. In this $f'(1) = 1$. This gives

$$MSE(\bar{y}_f) = \left(\frac{1}{n_2'}-\frac{1}{N}\right)S_2^2 + \left(\frac{1}{n_2'}-\frac{1}{n}\right)[R^2 S_2^2 + 2RS_{12}]$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)S_2^2 + \left(\frac{1}{n_2'}-\frac{1}{n}\right)[R^2 S_1^2 + 2RS_{12} + S_2^2]$$

$$-c = [\{f'(1)\}^2 + 2\rho f'(1) + a^2\rho_{21}^2\bar{X}^{2a-2} - 2a\rho_{21}^2\bar{X}^{a-1} - a\rho_{21}\bar{X}^{a-1}f'(1)]$$

$$-c = 1 + 2\rho$$

gives

$$\mu_{opt} = \frac{1}{1+\sqrt{2+2\rho_{21}}}$$

$$MSE(\hat{T})_{opt}\Big|_{\mu=\mu_0} = \left(1+\sqrt{2+2\rho_{21}}\right)\frac{(\bar{Y}C_2)^2}{2n}$$

$$E = \frac{MSE(\bar{y}_n)}{MSE(\hat{T})}\times 100 = \frac{2}{1+\sqrt{2+2\rho_{21}}}\times 100$$

Similarly, in the case when

$f(u) = u^{-1}$ and $a = 0$

we get the product estimator and

$$f^{-1}(u)|_{u=1} = -1$$

-

This yields

$$V(\bar{y}_f) = \left(\frac{1}{n_2'} - \frac{1}{N}\right) S_2^2 + \left(\frac{1}{n_2'} - \frac{1}{n}\right) [R^2 S_1^2 - 2RS_{12}]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_2^2 + \left(\frac{1}{n_2'} - \frac{1}{n}\right) [R^2 S_1^2 - 2RS_{12} + S_2^2]$$

$$-c = [\{f'(1)\}^2 + 2\rho f'(1) + a^2 \rho_{21}^2 \bar{X}^{2a-2} - 2a\rho_{21}^2 \bar{X}^{a-1} - a\rho_{21} \bar{X}^{a-1} f'(1)]$$

$$-c = 1 - 2\rho$$

gives

$$\mu_{opt} = \frac{1}{1 + \sqrt{2 - 2\rho_{21}}}$$

$$MSE(\hat{T})_{opt}\Big|_{\mu=\mu_0} = \left(1 + \sqrt{2 - 2\rho_{21}}\right) \frac{(\bar{Y}C_2)^2}{2n}$$

$$E = \frac{MSE(\bar{y}_n)}{MSE(\hat{T})} \times 100 = \frac{2}{1 + \sqrt{2 - 2\rho_{21}}} \times 100$$

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. **Patterson, H.D**., (1950). Sampling on successive occasions with partial replacement of units. *J. Royal Statist.Soc. Series* B 12, pp.241-255.
3. **Ray, S.K. and Singh R. Karan**(1979). A generalized ratio product estimator difference (RPD) estimator in double sampling, *JISAS*, vol 31, pp83-88,
4. **Singh R. Karan** (1982). A generalized double sampling estimator for the ratio and product of population parameters, vol 20, pp 39-49, JISA.
5. **Sukhatme P**.V. et al(1984). Sampling Theory of Surveys with Applications, Iowa state university press, (USA) and Indian Society of Agricultural Statistics, New Delhi.
6. **Tripathi, T.P., Sinha, S.K.P** (1976). Estimation of ratio on successive occasions, Paper presented at the *Symposium on Recent Developments in Survey Methodology*, held at Indian Statistical Institute.