Journal of Scientific Research
Banaras Hindu University, Varanasi
Vol. 62, 2018 : 203-212
ISSN : 0447-9483

ESTIMATION OF POPULATION MEAN USING TWO AUXILIARY VARIABLES IN SYSTEMATIC SAMPLING

Madhulika Mishra, B.P. Singh and *Rajesh Singh
*Corresponding Author, Email:- rsinghstart@gmail.com
Department of Statistics, Banaras Hindu University, Varanasi-221005, India.

Abstract

Through this manuscript, the problem of estimating the population mean in systematic sampling using information on two auxiliary variables has been addressed. The expressions for the mean square error of the proposed estimator have been calculated up to the first order of approximation. It has been shown that the proposed estimator is more efficient than other existing estimators for which an empirical study through a real data has also been carried out.

Keywords: Study variable, Auxiliary variable, Intra-class Correlation, Systematic sampling, Mean Squared Error, Efficiency, Bias.

Introduction

It is widely known in the context of literature of survey sampling that the efficiency of the estimators of the population parameters of the variable of interest can be increased by the use of auxiliary information related to auxiliary variable x which is highly correlated with the variable of interest y.

Hansel(1942) and Griffith(1945-1946) found systematic sampling to be more efficient and convenient in sampling certain natural populations like forest areas for estimating the volume of the timber and areas for estimating the volume of the timber and areas under different types of cover(Osborne 1942). Systematic sampling is generally more precise than simple random sampling and even more precise than stratified random sampling under certain specific conditions. Also it provides estimators which are more efficient than simple random sampling or stratified random sampling for certain types of populations see [Cochran(1946) , Gautschi(1957), Hajeck(1959). Various authors like Kushwaha and Singh (1989), Banarasi et al. (1993), Singh and Singh (1998), Singh et al. (2011), Singh and Solanki (2012), Singh and Jatwa (2012, Singh et al. (2012a), Singh et al. (2012b), Chaudhary et al. (2012), Verma et al. (2012), Verma and Singh (2014), Singh and Singh (2015) and Khan and Singh (2015) have discussed the problem of estimation of population mean using information on auxiliary variables in the context of systematic sampling. In this paper, we have proposed an estimator for the estimation of population mean using information on two auxiliary variables under the framework of systematic sampling.

Let us consider a finite population U of size N consisting of distinct and identifiable units U_1, U_2, U_3,…,U_N and number it from 1 to N units in some order. A random sample of size 'n' can be selected by the following manner; the first unit is randomly selected from the first 'k' units and then every k^{th} subsequent unit is selected till we get a sample of size 'n'. Thus there will be k samples each of size 'n' and observe the study variable y and auxiliary variable x for each and every unit selected in the sample.

$\overline{y}^* = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$ and $\overline{x}^* = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ are the unbiased estimators of the population mean

$\overline{Y} = \frac{1}{N}\sum_{j=1}^{N} y_{ij}$, $\overline{X} = \frac{1}{N}\sum_{j=1}^{N} x_{ij}$. Further let

$$\rho_x^* = \{1 + (n-1)\rho_x\},$$

$$\rho_y^* = \{1 + (n-1)\rho_y\}$$

$$\rho_z^* = \{1 + (n-1)\rho_z\}$$

where $\rho_y = \frac{E(y_{ij} - \overline{Y})(y_{ij'} - \overline{Y})}{E(y_{ij} - \overline{Y})^2}$

$$\rho_x = \frac{E(x_{ij} - \overline{X})(x_{ij'} - \overline{X})}{E(x_{ij} - \overline{X})^2}$$

$$\rho_z = \frac{E(z_{ij} - \overline{Z})(x_{ij'} - \overline{X})}{E(z_{ij} - \overline{Z})^2}$$

are the corresponding intra-class correlation coefficients for the study variable y and auxiliary variables x and z respectively.

Similarly, $\rho_{yx} = \frac{S_{yx}}{S_y S_x}$, $\rho_{yz} = \frac{S_{yz}}{S_y S_z}$, $\rho_{xz} = \frac{S_{xz}}{S_x S_z}$ are the correlation coefficients of the study and the auxiliary variables respectively.

Here S_y, S_x and S_z are the population standard deviations of the study variable y and the auxiliary variables x and z respectively. S_{xy}, S_{yz} and S_{zx} are the population covariances between y and x; y and z; x and z respectively. Also let C_y C_x and C_z are the population coefficient of variation of the study and the auxiliary variables respectively.

2. ESTIMATORS IN LITERATURE:

In this section, we consider several estimators of the finite population mean that are available in the sampling literature. The variance and mean squared error's (MSE's) of all the estimators considered here are obtained under the first order of approximation.

- The variance of the classical unbiased estimator $\bar{y}^*$ i.e., sample mean is defined as,

$$V(\bar{y}^*) = \theta\bar{Y}^2\rho_y^*C_y^2 \tag{2.1}$$

- Swain(1964), proposed a ratio estimator in systematic sampling as,

$$\bar{y}_R^* = \bar{y}^* * \left(\frac{\bar{X}}{\bar{x}^*}\right) \tag{2.2}$$

The mean squared error of the estimator $\bar{y}_R^*$ is given by

$$MSE(\bar{y}_R^*) = \theta\bar{Y}^2\left[\rho_y^*C_y^2 + \rho_x^*C_x^2\left(1 - 2k\sqrt{\rho^{**}}\right)\right] \tag{2.3}$$

where $\rho^{**} = \dfrac{\rho_y^*}{\rho_{x*}}$ and $k = \rho_{yx}\dfrac{C_y}{C_x}$.

- The classical product estimator in systematic sampling by Shukla (1971) is defined as,

$$\bar{y}_P^* = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{X}}\right) \tag{2.4}$$

The mean squared error of the estimator $\bar{y}_P^*$ is given by

$$MSE(\bar{y}_P^*) = \theta\bar{Y}^2\left[\rho_y^*C_y^2 + \rho_x^*C_x^2\left(1 + 2k\sqrt{\rho^{**}}\right)\right] \tag{2.5}$$

- Singh et al. (2011) proposed ratio-product type exponential estimators and are given by-

$$\bar{y}^{*}_{Re} = \bar{y}^{*}\exp\left(\frac{\bar{X}-\bar{x}^{*}}{\bar{X}+\bar{x}^{*}}\right) \tag{2.6}$$

$$\bar{y}^{*}_{Pe} = \bar{y}^{*}\exp\left(\frac{\bar{x}^{*}-\bar{X}}{\bar{x}^{*}+\bar{X}}\right) \tag{2.7}$$

The mean squared error of the estimators $\bar{y}^{*}_{Re}$ and $\bar{y}^{/*}_{Pe}$ are given as

$$MSE(\bar{y}^{*}_{Re}) = \theta\bar{Y}^{2}\left[\rho^{*}_{y}C^{2}_{y} + \frac{\rho^{*}_{x}C^{2}_{x}}{4}\left(1-4k\sqrt{\rho^{**}}\right)\right] \tag{2.8}$$

$$MSE(\bar{y}^{*}_{Pe}) = \left[\theta\bar{Y}^{2}\left[\rho^{*}_{y}C^{2}_{y} + \frac{\rho^{*}_{x}C^{2}_{x}}{4}\left(1+4k\sqrt{\rho^{**}}\right)\right]\right. \tag{2.9}$$

Singh et al (2009) proposed a ratio-cum- product type exponential estimator given by,

$$t^{*}_{RPe} = \bar{y}^{*}\exp\left(\frac{\bar{X}-\bar{x}^{*}}{\bar{X}+\bar{x}^{*}}\right)\exp\left(\frac{\bar{z}^{*}-\bar{Z}}{\bar{z}^{*}+\bar{Z}}\right) \tag{2.10}$$

The mean square error of the estimator $\bar{y}^{*}_{RPe}$ is given by

$$MSE(t^{*}_{RPe}) = \theta\bar{Y}^{2}\left[\rho^{*}_{y}C^{2}_{y} + \rho^{*}_{x}C^{2}_{x}\left(1-k\sqrt{\rho^{**}}\right) + \rho^{*}_{z}C^{2}_{z}\left(1-k\sqrt{\rho^{**}_{1}}\right) + k^{*}C^{2}_{z}\sqrt{\rho^{**}_{2}}\,\rho^{*}_{z}\right] \tag{2.11}$$

3. PROPOSED ESTIMATOR:

For estimating unknown population mean $\bar{Y}$ of the study variable we propose a class of estimator as follows:

$$t = w_{0}t_{0} + w_{1}t_{1} + w_{2}t_{2} \tag{3.1}$$

where, $t_{0} = \bar{y}^{*}$, $t_{1} = \bar{y}^{*}\left(\frac{\bar{X}}{\bar{x}^{*}}\right)\left(\frac{\bar{z}^{*}}{\bar{Z}}\right)$ and $t_{2} = \bar{y}^{*}\exp\left(\frac{\bar{X}-\bar{x}^{*}}{\bar{X}+\bar{x}^{*}}\right)\exp\left(\frac{\bar{z}^{*}-\bar{Z}}{\bar{z}^{*}+\bar{Z}}\right)$

Here, $t_{o}, t_{1}, t_{2} \in W$

W denotes the set of all possible estimators for estimating the population mean $\overline{Y}$. By definition, the set W is a linear variety if

$$t = \sum_{i=0}^{2} w_i t_i \in W \tag{3.2}$$

$$\text{Such that } \sum_{i=0}^{2} w_i = 1 \text{ and } w_i \in R \tag{3.3}$$

where, $w_i (i = 0,1,2)$ denotes the constants used for reducing the bias in the class of estimators.

In order to find the Bias and MSE of the proposed estimator, we consider the following relative error terms,

$$\in_0 = \frac{\overline{y}^* - \overline{Y}}{\overline{Y}} \;,\; \in_1 = \frac{\overline{x}^* - \overline{X}}{\overline{X}} \;,\; \in_2 = \frac{\overline{z}^* - \overline{Z}}{\overline{Z}}$$

Such that $E(\in_0) = E(\in_1) = E(\in_2) = 0$

$$E(\in_0^2) = \theta\rho_y^* C_y^2 \;,\; E(\in_1^2) = \theta\rho_x^* C_x^2 \;,\; E(\in_2^2) = \theta\rho_z^* C_z^2$$

$$E(\in_0 \in_1) = \theta k C_x^2 \sqrt{\rho_y^* \rho_x^*} \;\; E(\in_0 \in_2) = \theta k^* C_z^2 \sqrt{\rho_y^* \rho_z^*} \;\; E(\in_0 \in_2) = \theta k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*}$$

where $k = \dfrac{\rho_{yx} C_y}{C_x}$, $k^* = \dfrac{\rho_{yz} C_y}{C_z}$, $k^{**} = \dfrac{\rho_{xz} C_x}{C_z}$, $\theta = \dfrac{N-1}{Nn}$

Expressing the estimator t in terms of $\in$'s we get

$$t = \overline{Y}(1+\in_0)\left[w_0 + w_1(1+\in_1)^{-1}(1+\in_2) + w_2 \exp\{-\in_1 (2+\in_1)^{-1}\} \exp\{(2+\in_2)^{-1}\}\right] \tag{3.4}$$

By expanding the above equation (3.4) and keeping terms only up to first of approximation $E's$, we can write

$$t = \overline{Y}(1+\in_0)\left[1 - \in_1\left(w_1 + \frac{w_2}{2}\right) + \in_2\left(w_1 + \frac{w_2}{2}\right) + \in_1^2\left(w_1 + \frac{w_2}{4}\right) - \in_1\in_2\left(w_1 + \frac{w_2}{4}\right)\right] \tag{3.5}$$

Now, subtracting $\overline{Y}$ from both the sides of equation (3.5) and then taking expectation of both sides, the bias of the estimator t is obtained up to the first order of approximation as:

$$Bias(t)=\bar{Y}\left[\left(w_1+\frac{w_2}{2}\right)\left(\theta k^* C_z^2\sqrt{\rho_y^*\rho_z^*}-\theta k C_x^2\sqrt{\rho_y^*\rho_x^*}\right)+\left(w_1+\frac{w_2}{4}\right)\right.$$
$$\left.\left(\theta\rho_x^* C_x^2-\theta k^{**} C_z^2\sqrt{\rho_y^*\rho_z^*}\right)\right] \quad (3.6)$$

From (3.5) we get

$$t-\bar{Y}=\bar{Y}\left[\in_0+\left(w_1+\frac{w_2}{2}\right)(\in_2-\in_1)(1+\in_0)\right] \quad (3.7)$$

Squaring both the sides and then taking expectation we get

$$MSE(t)=\bar{Y}^2\theta\left[\rho_y^* C_y^2+Q^2\left\{\rho_x^* C_x^2+\rho_z^* C_z^2-2k^{**}C_z^2\sqrt{\rho_x^*\rho_z^*}\right\}+2Q\left\{k^* C_z^2\sqrt{\rho_y^*\rho_z^*}\right.\right.$$
$$\left.\left.-kC_x^2\sqrt{\rho_y^*\rho_x^*}\right\}\right] \quad (3.8)$$

Where $w_1+\frac{w_2}{2}=Q$ (3.9)

The MSE of the estimator t is minimum when

$$Q=\frac{k^* C_z^2\sqrt{\rho_y^*\rho_z^*}-kC_x^2\sqrt{\rho_y^*\rho_x^*}}{\rho_x^* C_x^2+\rho_z^* C_z^2-2k^{**}C_z^2\sqrt{\rho_x^*\rho_z^*}} \quad (3.10)$$

Putting this value of Q in equation (3.8), we get the minimum value for the MSE of the estimator t which is by,

$$\min MSE(t)=\bar{Y}^2\theta\left[\rho_y^* C_y^2+3\frac{\left(k^* C_z^2\sqrt{\rho_y^*\rho_z^*}-kC_x^2\sqrt{\rho_y^*\rho_x^*}\right)^2}{\rho_x^* C_x^2+\rho_z^* C_z^2-2k^{**}C_z^2\sqrt{\rho_x^*\rho_z^*}}\right] \quad (3.11)$$

From equation (3.3) and (3.9) there are two equations and three unknown. It is not possible to find the unique values for w_i's , $i = 0,1,2$. In order to get unique values of w_i's , we impose the linear restriction as,

$$\sum_{i=0}^{2} w_i B(t_i)=0 \quad (3.12)$$

Where $B(t_i)$ (i=0, 1, 2) denotes the Bias in the i^{th} estimator.

Equations (3.3), (3.9) and (3.12) can be written in the matrix form as,

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1/2 \\ B(t_0) & B(t_1) & B(t_2) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 \\ Q \\ 0 \end{bmatrix} \tag{3.13}$$

Solving (3.13) we get the unique values of w_0 , w_1 and w_2 as,

$$w_0 = \frac{2A_1A_2 + A_3A_4}{2A_1 - A_2}$$

$$w_1 = (1 - A_2) + \frac{(1 - A_2)A_3}{2A_1 - A_2}$$

$$w_2 = \frac{-2(1 - A_2)A_3}{2A_1 - A_2}$$

Here,

$$A_1 = \left[1 + \frac{\theta\rho_x^* C_x^2}{4} - \frac{\theta k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*}}{4} - \frac{\theta k C_x^2 \sqrt{\rho_y^* \rho_x^*}}{2} + \frac{\theta k^* C_z^2 \sqrt{\rho_y^* \rho_z^*}}{2}\right]$$

$$A_2 = 1 - \left(\frac{k^* C_z^2 \sqrt{\rho_y^* \rho_z^*} - k C_x^2 \sqrt{\rho_y^* \rho_x^*}}{\rho_x^* C_x^2 + \rho_z^* C_z^2 - 2k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*}}\right)$$

$$A_3 = \left(\theta k^* C_z^2 \sqrt{\rho_y^* \rho_z^*} - \theta k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*} - \theta k C_x^2 \sqrt{\rho_y^* \rho_x^*} + \theta\rho_x^* C_x^2\right)$$

$$A_4 = 2\left(\frac{k^* C_z^2 \sqrt{\rho_y^* \rho_z^*} - k C_x^2 \sqrt{\rho_y^* \rho_x^*}}{\rho_x^* C_x^2 + \rho_z^* C_z^2 - 2k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*}}\right) - 1$$

4. EMPIRICAL STUDY-

To examine the merits of the proposed estimator over the other existing estimators at optimum conditions, we have considered natural population data set from the literature. The source of population is given by, Population (Source: Tailor et al. (2013))

$N = 15$, $n = 3$, $\overline{X} = 44.47$, $\overline{Y} = 80$, $\overline{Z} = 48.40$, $C_y = 0.56$, $C_x = 0.28$, $C_z = 0.43$

$S_y^2 = 2000$

$S_x^2 = 149.55$, $S_z^2 = 427.83$, $S_{yx} = 538.57$, $S_{yz} = -902.86$, $S_{xz} = -241.06$,

$\rho_{yx} = 0.9848$,

$\rho_{yz} = -0.9760$, $\rho_{xz} = -0.9530$, $\rho_y = 0.6652$, $\rho_x = 0.707$, $\rho_z = 0.5487$

Table-1: MSE's and Percent Relative Efficiencies (PRE's) of the estimators w.r.to $\overline{y}^*$

ESTIMATOR	MSE	PRE
$\overline{y}^*$	1455.13	100.00
$\overline{y}_R^*$	373.47	389.62
$\overline{y}_P^*$	3290.45	44.22
$\overline{y}_{Re}^*$	820.09	177.43
$\overline{y}_{Pe}^*$	1044.42	139.32
$\overline{y}_{RPe}^*$	235.61	617.61
t	26.19	5556.86

The Percent Relative Efficiencies (PRE's) of the estimators with respect to the usual unbiased estimator $\overline{y}^*$ are obtained from the following mathematical formula.

$$PRE(ESTIMATOR) = \frac{V(y^y)}{MSE(ESTIMATOR)} \times 100$$

5. CONCLUSION

In this paper we have proposed a class of estimator for the population mean in systematic sampling using information on two auxiliary variables. The MSE of the proposed estimator has been derived up to first order of approximation. Further we have used empirical approach for comparing the efficiency of the proposed estimator with other estimators for which we have used known natural population dataset, see Tailor et al (2013) and the results have been shown above in the Table 1. From the Table 1, it is clear that the proposed estimator turns out to be more efficient as compared to other estimators because of smaller value of MSE and higher value of

PRE. So it is obviously more desirable to use the proposed estimator in practical surveys.

References

Banarasi., Kushwaha, S. N. S. and Kushwaha, K.S. (1993). A class of ratio, product and difference (RPD) estimators in systematic sampling, Microelectronics Reliability 33(4), 455–457.

Chaudhary, M. K., Malik, S., Singh, J., and Singh, R. (2012): A general family of estimators for estimating population mean in systematic sampling using auxiliary information in the presence of missing observations. Jour. Raj. Stat. Assoc. 1(2), 1-8.

Chaudhary, M.K., Malik, S., Singh, R. and Florentin, S. (2013): Use of auxiliary information for estimating population mean in systematic sampling under non-response.ON IMPROVEMENT IN ESTIMATING POPULATION PARAMETER(S) USING AUXILIARY INFORMATION. Educational Publishing & Journal of Matter Regularity (Beijing).

Cochran, W. G. 1946. Relative efficiency of systematic and stratified random samples for a certain class of populations. Ann. Math. Stat., 17, 164–177.

Gautschi, W. 1957. Some remarks on systematic sampling. Ann. Math. Stat., 28, 385–394.

Griffth, A. L. 1945–1946. The efficiency of enumerations. Forest-Research Institute, Dehra Dun. Indian Forest Leaflets 83–93.

Hajeck, J. 1959. Optimum strategy and other problems in probability sampling. Casopis pro Pestovani Matematiky 84, 387–423.

Hansel, A. A. 1942. Estimation of volume of timber stands by strip sampling. Ann. Math. Stat., 13, 179–206.

Khan, M. and Singh, R. (2015) : Estimation of population mean in chain ratio-type estimator under systematic sampling. Jour. of Prob. and Stats., 1-5. http://dx.doi.org/10.1155/2015/248374

Kushwaha, K. S. and Singh, H. P. (1989). Class of almost unbiased ratio and product estimators in systematic sampling, Journal of the Indian Society of Agricultural Statistics 41(2), 193-205.

Obsborne, J. G. 1942. Sampling errors of systematic and random surveys of cover-types areas. J. Am. Stat. Assoc., 37, 256–264.

Shukla, N. D. (1971). Systematic sampling and product method of estimation, in Proceeding of all India Seminar on Demography and Statistics. B.H.U. Varanasi: India.

Singh, H. P. and Jatwa, N. K. (2012). A class of exponential type estimators in systematic sampling, Economic Quality Control 27, 195–208

Singh, H. P. and Singh, R. (1998). Almost unbiased ratio and product type estimators in systematic sampling, Questiio 22(3), 403–416.

Singh, H. P. and Solanki, R. S. (2012). An efficient class of estimators for the population mean using auxiliary information in systematic sampling, Journal of Statistical Theory and Practice 6(2), 274–285.

Singh, H. P., Tailor, R. and Jatwa, N. K. (2011). Modified ratio and product estimators for population mean in systematic sampling, Journal of Modern Applied Statistical Methods 10(2), 424-435.

Singh, H. P., Tailor, R. and Jatwa, N. K. (2011). Modified ratio and product estimators for population mean in systematic sampling, Journal of Modern Applied Statistical Methods 10(2), 424-435.

Singh, R., Malik, S., Singh, V. K. (2012a) : An improved estimator in systematic sampling. Jour. Of Scie. Res., 56, 177-182.

Singh, R., Malik, S., Chaudhary, M.K., Verma, H. and Adewara, A. A. (2012b): A general family of ratio type- estimators in systematic sampling. Jour. Reliab. and Stat. Stud.,5(1), 73-82).

Singh, V.K. and Singh, R. (2015) : Some improved suggestions in circular systematic sampling under fully and partially respondent case. JRSS 8(1), 115-134.

Swain, A. K. P. C. (1964). The use of systematic sampling ratio estimate, Journal of the Indian Statistical Association 2, 160-164.

Tailor, T., Jatwa, N. K . and Singh, H. P. (2013). A ratio-cum-product estimator of finite population mean in systematic sampling, Statistics in Transition 14(3), 391-398.

Verma, H, Singh, R. D. and Singh, R. (2012) : A general class of regression type estimators in systematic sampling under non-response. Octogon Mat. Mag. 20(2), 542-550.