Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666,
Vol. 5, Issue 1 (2012): 73-82

A GENERAL FAMILY OF RATIO-TYPE ESTIMATORS IN SYSTEMATIC SAMPLING

***Rajesh Singh,* Sachin Malik, *Manoj K. Chaudhary,**
***Hemant K. Verma and **A.A. Adewara**
*Department of Statistics, Banaras Hindu University, Varanasi-221005, India
**Department of Statistics, University of Ilorin,Ilorin, Kwara State, Nigeria
(Received March 13, 2012)

Abstract

In this paper, we have adapted Gupta and Shabbir (2008) estimator in systematic sampling using auxiliary information. For the family of estimators, under systematic sampling the expression of mean square error (MSE) up to the first order approximations is derived. The family of estimators in its optimum case is discussed. Also, an empirical study is carried out to show the properties of the proposed estimators.

Keywords: Auxiliary Variable, Systematic Sampling, Ratio Estimator, Mean Square Error, Efficiency.

1. Introduction

There are some natural populations like forest etc., where it is not possible to apply easily the simple random sampling or other sampling schemes for estimating the population characteristics. In such situations, one can easily implement the method of systematic sampling for selecting a sample from the population. Systematic sampling has the advantage of selecting the whole sample with just random start. Estimation in systematic sampling has been discussed in detail by Lahiri (1954), Gautschi (1957), Hajeck (1959), Madow, and Madow (1944) and Cochran (1946) Madow, W. G. and Madow, L.H. (1944) Madow, W. G. and Madow, L.H. (1944). Use of auxiliary information in construction of estimators is considered by Kushwaha and Singh (1989), Banarasi et. al. (1993), Singh and Singh (1998) and Singh et al. (2012). We introduced the following terminology to discuss the estimators.

Let y be the study variable and x be the auxiliary variable defined on a finite population $U = (U_1, U_2 ...., U_N)$. Here, we assume N=nk, where n and k are positive integers. Let (y_{ij}, x_{ij}); i=1,2,...,k; j=1,2,...,n denote the value of j^{th} unit in the i^{th} sample. The systematic sample means

$$\bar{y}^* = \frac{1}{n}\sum_{j=1}^{n} y_{ij}, \qquad \bar{x}^* = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$$

are unbiased estimators of the population means $(\bar{Y}, \bar{X})$ of (y, x), respectively.

If $e_0 = \frac{(\bar{y}^* - \bar{Y})}{\bar{Y}}$ and $e_1 = \frac{(\bar{x}^* - \bar{X})}{\bar{X}}$

Then, we have

$E(e_0) = E(e_1) = 0,$ and

$E(e_0^2) = \theta(1+(n-1)\rho_y)C_y^2, \qquad E(e_1^2) = \theta(1+(n-1)\rho_x)C_x^2$

$E(e_0 e_1) = \theta(1+(n-1)\rho_y)^{\frac{1}{2}}(1+(n-1)\rho_x)^{\frac{1}{2}}\rho C_y C_x,$

where $\theta = \dfrac{N-1}{Nn},$

$$\rho_x = \frac{E(x_{ij}-\overline{X})(x_{ij'}-\overline{X})}{E(x_{ij}-\overline{X})^2}, \qquad \rho_y = \frac{E(y_{ij}-\overline{Y})(y_{ij'}-\overline{Y})}{E(y_{ij}-\overline{Y})^2},$$

$$\rho = \frac{E(x_{ij}-\overline{X})(y_{ij}-\overline{Y})}{\left(E(x_{ij}-\overline{X})^2(y_{ij}-\overline{Y})^2\right)^{\frac{1}{2}}} \text{ and } \rho^* = \frac{\{1+(n-1)\rho_y\}}{\{1+(n-1)\rho_x\}}.$$

Where (C_y, C_x) are the coefficients of variation of the variates (y, x) respectively. It is assumed that the population mean $\overline{X}$ of the auxiliary variable is known.

The usual ratio, product and regression estimators of the population mean $\overline{Y}$ based on a systematic sample of size n, respectively, be defined as

$$\overline{y}_R^* = \frac{\overline{y}^*}{\overline{x}^*}\overline{X} \tag{1.1}$$

$$\overline{y}_P^* = \frac{\overline{y}^*\overline{x}^*}{\overline{X}} \tag{1.2}$$

$$\overline{y}_{lr}^* = \overline{y}^* + b(\overline{X} - \overline{x}^*) \tag{1.3}$$

where b = $\dfrac{s_{xy}}{s_x^2}$, and $\overline{y}^*, \overline{x}^*$ are estimators of population means $\overline{Y}$ of study variable and $\overline{X}$ of auxiliary variable based on the systematic sample of size n units. s_x^2 is the population variance of the auxiliary variate and s_{xy} is the population covariance between auxiliary variate and variate of interest .

The mean square errors (MSE's) of $\overline{y}_R^*$, $\overline{y}_P^*$ and $\overline{y}_{lr}^*$ are respectively, given by

$$MSE(\overline{y}_R^*) = \theta\overline{Y}^2\{1+(n-1)\rho_X\}\left[\rho^{*2}C_Y^2 + (1-2K_1\rho^*)C_X^2\right] \tag{1.4}$$

$$MSE\left(\overline{y}_P^*\right)=\theta\overline{Y}^2\left\{1+(n-1)\rho_X\right\}\left[\rho^{*2}C_Y^2+\left(1+2K_1\rho^*\right)C_X^2\right] \tag{1.5}$$

$$MSE\left(\overline{y}_{lr}^*\right)=\theta\overline{Y}^2\left\{1+(n-1)\rho_X\right\}\left[C_Y^2-K_1^2C_X^2\right]\rho^{*2} \tag{1.6}$$

where $K_1=\rho\dfrac{C_Y}{C_X}$,

2. Proposed Family of Estimators

Motivated by Gupta and Shabbir (2008), we propose the following general class of ratio-type estimators in systematic sampling as

$$t_p^*=\left[w_1\overline{y}^*+w_2\left(\overline{X}-\overline{x}^*\right)\right]\frac{\overline{X}}{\overline{x}^*} \tag{2.1}$$

where w_1 and w_2 are constants whose values are to determined later.

Expressing (2.1) in terms of e’s, we have

$$t_p^*=\left[w_1\overline{Y}\left(1+e_0\right)-w_2\overline{X}e_1\right]\left(1+e_1\right)^{-1} \tag{2.2}$$

We assume that $|e_1|$<1, so that the term $(1+e_1)^{-1}$ is expandable. Expanding the right hand side of (2.2), we have

$$t_p^*=\left[w_1\overline{Y}\left(1+e_0\right)-w_2\overline{X}e_1\right]\left[1-e_1+e_1^2-...\right]$$

$$=\left[w_1\overline{Y}\left(1+e_0\right)-w_2\overline{X}e_1-w_1\overline{Y}\left(e_1+e_0e_1\right)+w_2\overline{X}e_1^2+w_1\overline{Y}e_1^2\right] \tag{2.3}$$

Neglecting terms of e’s having power greater than two we have

$$t_p^*=\left[w_1\overline{Y}\left(1+e_0\right)-w_2\overline{X}e_1-w_1\overline{Y}\left(e_1+e_0e_1\right)+w_2\overline{X}e_1^2+w_1\overline{Y}e_1^2\right] \tag{2.4}$$

Subtracting $\overline{Y}$ from both sides of (2.4) and then squaring and neglecting terms of e’s having power greater than two, we have

$$\begin{aligned}\left(t_p^*-\overline{Y}\right)^2 &= \overline{Y}^2+w_1^2\overline{Y}^2\left(1+2e_0-2e_1+e_0^2+3e_1^2-4e_0e_1\right)+w_2^2\overline{X}^2e_1^2\\ &\quad -2w_1w_2\overline{Y}\overline{X}\left(e_1+e_0e_1-2e_1^2\right)-2w_1\overline{Y}^2\left(1+e_0-e_1-e_0e_1+e_1^2\right)\\ &\quad +2w_2\overline{Y}\overline{X}\left(e_1-e_1^2\right)\end{aligned} \tag{2.5}$$

Taking expectations of both sides of (2.5), the MSE of t to the first order of approximation is

$$\begin{aligned}MSE\left(t_p^*\right) &= \overline{Y}^2+w_1^2\overline{Y}^2\left(1+\theta\left(C_0^2+3C_1^2-4C_0C_1\right)\right)+w_2^2\overline{X}\theta C_1^2\\ &\quad -2w_1w_2\overline{Y}\overline{X}\theta\left(C_0C_1-2C_1^2\right)-2w_1\overline{Y}^2\left(1-\theta\left(C_0C_1-C_1^2\right)\right)\\ &\quad -2w_2\theta\overline{Y}\overline{X}C_1^2\end{aligned} \tag{2.6}$$

$$MSE\left(t_p^*\right) = \overline{Y}^2 + w_1^2\overline{Y}^2 A_1 + w_2^2\overline{X}^2 A_2 - 2w_1 w_2 \overline{X}\overline{Y} A_3 - 2w_1 \overline{Y}^2 A_4 - 2w_2 \overline{Y}\overline{X} A_5 \tag{2.7}$$

where,

$$\left.\begin{aligned} A_1 &= 1+\theta\left[C_0^2 + 3C_1^2 - 4C_0C_1\right] \\ A_2 &= \theta C_1^2 \\ A_3 &= \theta\left[C_0C_1 - 2C_1^2\right] \\ A_4 &= 1-\theta\left[C_0C_1 - C_1^2\right] \\ A_5 &= \theta C_1^2 \end{aligned}\right\}$$

and

$$\left.\begin{aligned} C_0^2 &= \left(1+(n-1)\rho_y\right)C_y^2 \\ C_1^2 &= \left(1+(n-1)\rho_x\right)C_x^2 \\ C_0C_1 &= \left(\frac{1+(n-1)\rho_y}{1+(n-1)\rho_x}\right)^{\frac{1}{2}} \rho C_y C_x \end{aligned}\right\} \tag{2.8}$$

Partially differentiating (2.7) with respect to w_1 and w_2, respectively ,we have

$$\begin{bmatrix} \overline{Y}^2 A_1 & -A_3\overline{Y}\overline{X} \\ -A_3\overline{Y}\overline{X} & A_2\overline{X}^2 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \overline{Y}^2 A_4 \\ \overline{Y}\overline{X} A_5 \end{bmatrix} \tag{2.9}$$

Solving (2.9), we get the optimum values of w_1 and w_2 as

$$w_1^* = \frac{A_2A_4 + A_3A_5}{A_1A_2 - A_3^2} \text{ and } w_2^* = \frac{\overline{Y}\left(A_1A_5 + A_3A_4\right)}{\overline{X}\left(A_1A_2 - A_3^2\right)}.$$

Substituting optimum values of w_1 and w_2 in equation (2.7) we get the minimum MSE of t as

$$\min.\ MSE(t_p^*) = \overline{Y}^2\left[1 - \frac{\left(A_2A_4^2 + 2A_3A_4A_5 + A_1A_5^2\right)}{\left(A_1A_2 - A_3^2\right)}\right] \tag{2.10}$$

3. Empirical Study

In the support of theoretical results, we have considered the data given in Murthy (1967, p. 131-132). These data are related to the length and timber volume for ten blocks of the blacks mountain experimental forest. The value of intraclass correlation coefficients ρ_X and ρ_Y have been given approximately equal by Murthy (1967, p. 149) and Kushwaha and Singh (1989) for the systematic sample of size 16 by enumerating all possible systematic samples after arranging the data in ascending order of strip length. The particulars of the population are given below:

$N = 176$, $n = 16$, $\bar{Y} = 282.6136$, $\bar{X} = 6.9943$,

$S_Y^2 = 24114.6700$, $S_X^2 = 8.7600$, $\rho = 0.8710$.

Estimators	PRE $(\bar{y}^*)$
$\bar{y}^*$	100.00
$\bar{y}_R^*$	397.55
$\bar{y}_P^*$	34.07
$\bar{y}_{lr}^*$	1374.91
t_p^*	1374.91

Table 3.1: PRE of different estimators with respect to $\bar{y}^*$

4. Non-Response

We assume that the non-response is observed only on study variable and auxiliary variable is free from non-response. Using Hansen-Hurwitz (1946) technique of sub-sampling of non-respondents, the estimator of population mean $\bar{Y}$, can be defined as

$$\bar{y}^{**} = \frac{n_1 \bar{y}_{n1} + n_2 \bar{y}_{h_2}}{n} \tag{4.1}$$

where $\bar{y}_{n1}$ and $\bar{y}_{h_2}$ are, respectively the means based on n_1 respondent units from the systematic sample of n units and sub-sample of h_2 units selected from n_2 non-respondent units in the systematic sample. The estimator of population mean $\bar{X}$ of auxiliary variable based on the systematic sample of size n units, is given by

$$\bar{x}^{*} = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \quad , \qquad (i = 1,2,...,k) \tag{4.2}$$

Obviously, $\bar{y}^{**}$ and $\bar{x}^{*}$ are unbiased estimators. The variance expressions for the estimators $\bar{y}^{**}$ and $\bar{x}^{*}$ are, respectively, given by

$$V\left(\bar{y}^{**}\right) = \theta\{1 + (n-1)\rho_Y\}S_Y^2 + \frac{L-1}{n}KS_{Y2}^2 \tag{4.3}$$

$$V\left(\bar{x}^{*}\right) = \theta\{1 + (n-1)\rho_X\}S_X^2 \tag{4.4}$$

where ρ_Y and ρ_X are the correlation coefficients between a pair of units within the systematic sample for the study and auxiliary variables respectively. S_Y^2 and S_X^2 are respectively the mean squares of the entire group for study and auxiliary variable. S_{Y2}^2 be the mean square of non-response group under study variable, K is the non-response rate in the population and $L = \frac{n_2}{h_2}$.

The ratio, product and regression estimators defined in equation (1.1), (1.2) and (1.3) under non-response can be respectively, written as

$$\bar{y}_R^{**} = \frac{\bar{y}^{**}}{\bar{x}^{*}}\bar{X} \tag{4.5}$$

$$\bar{y}_P^{**} = \frac{\bar{y}^{**}\bar{x}^{*}}{\bar{X}} \tag{4.6}$$

$$\bar{y}_{lr}^{**} = \bar{y}^{**} + b(\bar{X} - \bar{x}^{*}) \tag{4.7}$$

The MSE expressions for these estimators are respectively written as-

$$MSE\left(\bar{y}_R^{**}\right) = \theta\bar{Y}^2\left\{1 + (n-1)\rho_X\right\}\left[\rho^{*2}C_Y^2 + \left(1 - 2K_1\rho^{*}\right)C_X^2\right] + \frac{L-1}{n}W_2S_{Y2}^2 \tag{4.8}$$

$$MSE\left(\bar{y}_P^{**}\right) = \theta\bar{Y}^2\left\{1 + (n-1)\rho_X\right\}\left[\rho^{*2}C_Y^2 + \left(1 + 2K_1\rho^{*}\right)C_X^2\right] + \frac{L-1}{n}W_2S_{Y2}^2 \tag{4.9}$$

$$MSE\left(\bar{y}_{lr}^{**}\right) = \theta\bar{Y}^2\{1 + (n-1)\rho_X\}\left[C_Y^2 - K_1^2C_X^2\right]\rho^{*2} + \frac{L-1}{n}W_2S_{Y_2}^2 \tag{4.10}$$

5. Proposed Estimator under Non-Response

The estimator t in systematic sampling when the study variable y is having non –response, is given by

$$t^{**} = \left[l_1 \overline{y}^{**} + l_2 (\overline{X} - \overline{x}^{*}) \right] \left(\frac{\overline{X}}{\overline{x}^{*}} \right)^{p} \tag{5.1}$$

where l_1, l_2 and p are constants.

To obtain the expression for mean square error, we use large sample approximation

$$\overline{y}^{**} = \overline{Y}(1 + e_0^{*})$$

$$\overline{x}^{*} = \overline{X}(1 + e_1)$$

such that $E(e_0^{*}) = E(e_1) = 0$

$$E(e_0^{*2}) = \frac{V(\overline{y}^{**})}{\overline{Y}^2} = \theta\{1 + (n-1)\rho_Y\} C_Y^2 + \frac{L-1}{n} K \frac{S_{Y2}^2}{\overline{Y}^2},$$

$$E(e_1^2) = \frac{V(\overline{x}^{*})}{\overline{X}^2} = \theta\{1 + (n-1)\rho_X\} C_X^2,$$

and $E(e_0^{*} e_1) = \theta\{1 + (n-1)\rho_Y\}^{1/2} \{1 + (n-1)\rho_X\}^{1/2} \rho C_Y C_X.$

Expressing (5.1) in terms of e's, we have

$$t^{**} = \left[l_1 \overline{Y}(1 + e_0^{*}) - l_2 \overline{X} e_1 \right] (1 + e_1)^{-p} \tag{5.2}$$

We assume that $|e_1| < 1$, so that the term $(1 + e_1)^{-1}$ is expandable.

$$\begin{aligned} MSE(t^{**}) = {} & \overline{Y}^2 + l_1^2 \overline{Y}^2 \left(1 + \theta\left(C_0^2 + (2p^2 + p) C_1^2 - 4pC_0C_1\right) + \frac{L-1}{n} K \frac{S_{Y2}^2}{\overline{Y}^2} \right) \\ & + l_2^2 \overline{X}^2 \theta C_1^2 - 2 l_1 l_2 \overline{Y}\,\overline{X} \theta \left(C_0 C_1 - 2pC_1^2\right) \\ & - 2 l_1 \overline{Y}^2 \left(1 - \theta\left(pC_0C_1 - \frac{p(p+1)}{2} C_1^2 \right) \right) - 2 l_2 p \theta \overline{Y}\,\overline{X} C_1^2 \end{aligned} \tag{5.3}$$

$$\begin{aligned} MSE(t^{**}) = {} & \overline{Y}^2 + l_1^2 \overline{Y}^2 A_1^{*} + l_2^2 \overline{X}^2 A_2^{*} - 2 l_1 l_2 \overline{X}\,\overline{Y} A_3^{*} - 2 l_1 \overline{Y}^2 A_4^{*} \\ & - 2 l_2 \overline{Y}\,\overline{X} A_5^{*} + l_1^2 \overline{Y}^2 A_6^{*} \end{aligned} \tag{5.4}$$

Where,
$$\left.\begin{aligned}
A_1^* &= 1+\theta\left[C_0^2+(2p^2+p)C_1^2-4pC_0C_1\right]\\
A_2^* &= \theta C_1^2\\
A_3^* &= \theta\left[C_0C_1-2pC_1^2\right]\\
A_4^* &= 1-\theta\left[pC_0C_1-\frac{p(p+1)}{2}C_1^2\right]\\
A_5^* &= p\theta C_1^2\\
A_6^* &= \frac{(L-1)}{n}K\frac{S_{y2}^2}{\overline{Y}^2}
\end{aligned}\right\}$$

Partially differentiating (5.4) with respect to l_1 and l_2, respectively, we have

$$\begin{bmatrix} \overline{Y}^2A_1^*+\dfrac{(L-1)}{n}KS_{y2}^2 & -A_3^*\overline{Y}\overline{X} \\ A_3^*\overline{Y}\overline{X} & -A_2^*\overline{X}^2 \end{bmatrix}\begin{bmatrix} l_1 \\ l_2 \end{bmatrix}=\begin{bmatrix} \overline{Y}^2A_4^* \\ -\overline{Y}\overline{X}A_5^* \end{bmatrix} \tag{5.5}$$

Solving (5.5) we get the optimum values of l_1 and l_2 as

$$l_1^*=\frac{A_2^*A_4^*+A_3^*A_5^*}{A_1^*A_2^*+A_2^*A_6^*-A_3^{*2}}$$

And

$$l_2^*=\frac{\overline{Y}\left(A_1^*A_5^*+A_3^*A_4^*\right)}{\overline{X}\left(A_1^*A_2^*+A_2^*A_6^*-A_3^{*2}\right)}.$$

Substituting optimum values of l_1 and l_2 in equation (5.4) we get the minimum MSE of the proposed estimator t**.

6. Empirical Study

For numerical illustration, we have considered the data given in Murthy (1967, p. 131-132). The data are based on length (X) and timber volume (Y) for 176 forest strips. Murthy (1967, p.149) and Kushwaha and Singh (1989) reported the values of intra class correlation coefficients ρ_X and ρ_Y approximately equal for the systematic sample of size 16 by enumerating all possible systematic samples after arranging the data in ascending order of strip length. The details of population parameters are:

N = 176, n = 16, $\overline{Y}$ = 282.6136, $\overline{X}$ = 6.9943,

S_Y^2 = 24114.6700, S_X^2 = 8.7600, ρ = 0.8710,

$S_{Y2}^2=\frac{3}{4}S_Y^2$ = 18086.0025.

Table 6.1 shows the percentage relative efficiency (PRE) of t^{**} (optimum) and $\bar{y}_{lr}^{**}$ with respect to $\bar{y}^{**}$ for the different choices of K and L.

K	L	PRE of t(optimum) with respect to $\bar{y}^{**}$ (when p=1)	PRE of $\bar{y}_{lr}^{**}$ with respect to $\bar{y}^{**}$
0.1	2.0	438.9431	407.4884
	2.5	435.7211	404.1824
	3.0	432.5694	400.9468
	3.5	429.4858	397.7794
0.2	2.0	432.5694	400.9468
	2.5	426.4682	394.6779
	3.0	420.6229	388.6647
	3.5	415.018	382.8921
0.3	2.0	426.4682	394.6779
	2.5	417.7913	385.7493
	3.0	409.6395	377.3458
	3.5	401.9679	369.4225
0.4	2.0	420.6229	388.6647
	2.5	409.6395	377.3458
	3.0	399.5103	366.881
	3.5	390.1422	357.1773

Table 6.1: PRE of t^{**} (optimum) and $\bar{y}_{lr}^{**}$ with respect to $\bar{y}^{**}$

Conclusion

Both the theoretical and numerical comparisons show that the proposed estimator t^{**} is more efficient than the usual mean, Swain (1964) estimator, Shukla (1971) estimator and regression estimator in systematic sampling. Also, the proposed estimator t in case of presence of non-response in study variable performs better than other estimators considered here.

Acknowledgement

The authors would like to thank the referees for their useful suggestions on a previous draft of the paper.

References

1. Banarasi, Kushwaha, S.N.S. and Kushwaha, K.S**.** (1993). A class of ratio, product and difference (RPD) estimators in systematic sampling, Microelectron. Reliab., 33 (4), p. 455–457
2. Cochran, W.G. (1946). Relative efficiency of systematic and stratified random samples for a certain class of population, Ann. Math. Statist., 17, p. 164–177.

3. Gautschi, W. (1957). Some remarks on systematic sampling, Ann. Math. Statist., 28, p. 385-394.
4. Gupta, S. and Shabbir, J. (2008). On improvement in estimating the population mean in simple random sampling, J. Appl. Statist., 35(5), p. 559-566.
5. Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. Jour. of The Amer. Stat. Assoc., 41, p. 517-529.
6. Hajeck, J. (1959). Optimum strategy and other problems in probability sampling, Casopis pro Pestovani Matematiky, 84, p. 387–423.
7. Kushwaha, K. S. and Singh, H.P. (1989). Class of almost unbiased ratio and product estimators in systematic sampling. Jour. Ind. Soc. Ag. Statistics, 41(2), p. 193–205.
8. Lahiri, D. B. (1954): On the question of bias of systematic sampling, Proceedings of World Population Conference, 6, p. 349-362.
9. Madow, W. G. and Madow, L.H. (1944). On the theory of systematic sampling, Ann. Math. Statist., 15, p. 1-24.
10. Murthy, M.N. (1967). Sampling theory and methods. Statistical Publishing Society, Calcutta.
11. Shukla, N. D. (1971). Systematic sampling and product method of estimation, Proceedings of all India Seminar on Demography and Statistics, B.H.U.,Varanasi, India.
12. Singh, R. and Singh, H. P. (1998). Almost unbiased ratio and product type-estimators in systematic sampling. Questiio, 22 (3), p. 403-416.
13. Singh, R., Malik, S. and Singh V. K. (2012). An improved estimator in systematic sampling, Jour. of Scie. Res., 56, p. 177-182.
14. Swain, A.K.P.C. (1964): The use of systematic sampling in ratio estimate, Jour. Ind. Stat. Assoc., 2 (213), p. 160-164.