# **LOG-TYPE ESTIMATORS FOR ESTIMATING POPULATION MEAN IN SYSTEMATIC SAMPLING IN THE PRESENCE OF NON-RESPONSE**

# **Rajesh Singh and Sakshi Rai\***

\*Corresponding Author \*E Mail: sakshi.rai@bhu.ac.in Department of Statistics, Banaras Hindu University, Varanasi, India

> Received October 12, 2018 Modified April 10, 2019 Accepted May 05, 2019

#### **Abstract**

 This paper utilizes the information of auxiliary variable for estimating population mean in systematic sampling under the effect of non-response in study variable. Along with the proposed estimators, the paper also discusses some existing estimators. The expressions of mean squared errors of proposed estimators are derived up to the first order of approximation and it is observed that the efficiencies of proposed estimators are better than the existing estimators. To ratify the results, an empirical study has been performed taking two data sets.

**Key Words:** Systematic Sampling, Auxiliary Variable, Non-Response, Mean Square Error, Efficiency.

## **1. Introduction**

 Auxiliary information is the necessary information about the population either available in advance or can be obtained through conducting a small survey aside or in parallel with the main survey by investing the cost as minimum as possible. The main purpose of involving auxiliary variable with the study variable is to increase the precision of the estimators but the important thing is that there should be correlation between the auxiliary variable and the study variable. Ratio, Product, Regression estimators are the examples in this reference. Many authors have used this methodology to increase the precision of the estimators in various sampling techniques, some of them are Cochran (1977), Kadilar & Cingi (2004), Singh et al. (2009), Singh, Kumar and Smarandache (2010), Singh and Kumar (2011), Malik and Singh (2013), Singh and Malik (2014) and Singh et al. (2018).

 Systematic sampling is scheme of drawing samples from the population conveniently. It ensures equal probability for each unit in the population to be included in the sample by selecting the very first unit randomly and the next units will get automatically selected following a predetermined pattern. Let N be the population size and the units are numbered from 1 to N. For drawing the sample of size n first unit is selected randomly from 1 to k and the next units will be selected at a predetermined interval k such that  $N = nk$ . Then the sampled unit will be as-



 Due to its simplicity and better efficiency, many authors have shown their interest in this sampling scheme in their research works. Some of them are Singh et al. (2011), Singh et al. (2012), Singh et al. (2013), Khan and Singh (2015), Pal and Singh (2017) etc.

 Non-response is a phenomenon of failing to collect the required data or information from the unit selected in the sample. In survey sampling, it is a known fact that data can't be received from every sampled unit. There are many factors such as surveying agency's official status, the visiting time of the enumerator to the sampled unit, the nature of the information required, the length of the questionnaire or schedule etc. that are responsible for the non-response to creep in the data. Hansen and Hurwitz (1946) were the first who discussed the problem of non-response in mailed questionnaire. Verma et al. (2014) used this concept of non-response under systematic sampling scheme. Let N be classified into two classes i.e. respondent class and nonrespondent class, such that

 $N = N_1 + N_2$ 

 $N_1$  are the units who have responded and  $N_2$  are those who have not responded to the questionnaire or schedule at the first attempt and n be the sample size such that  $n_1$  and  $n_2$  are the sampled unit from the respondent class  $N_1$  and non-respondent class  $N_2$ respectively, such that

$$
n = n_1 + n_2
$$

Now let us draw a subsample of size  $w_2$  from non-respondent class  $n_2$  to be interviewed again such that

 $n_2 = w_2 h \forall h > 1$ Here  $N_1$  and  $N_2$  are not known then unbiased estimators of  $N_1$  and  $N_2$  are  $\widehat{N_{1*}} = \left(\frac{n_1}{n}\right)N$  and  $\widehat{N_{2*}} = \left(\frac{n_2}{n}\right)N$ 

The unbiased estimator of population mean is given by

$$
\overline{y}_{i*}^* = \frac{(n_1 \overline{y}_{n1} + n_2 \overline{y}_{w2})}{n}
$$

#### **Notations and Terminology**

Let  $(y_{ii}, x_{ii})$  be the pair of sample units of the study variable and auxiliary variable respectively, where  $(i= 1, 2, \ldots, k)$  and  $(j= 1, 2, \ldots, n)$  denoting  $j<sup>th</sup>$  unit in the  $i<sup>th</sup>$ 

sample drawn under systematic sampling technique. Mean of the study variable and auxiliary variable for  $i^{th}$  sampled unit is given by

$$
\overline{y}_{i*} = \left(\frac{\sum_{j=1}^{n} y_{ij}}{n}\right) \& \overline{x}_{i*} = \left(\frac{\sum_{j=1}^{n} x_{ij}}{n}\right), (i=1,2,\ldots,k)
$$

The non-response is seen only on the study variable and the auxiliary variable is kept free from it. Taking a sub sample from the sampled observation of nonrespondent unit then an unbiased estimator of population mean  $\overline{Y}$  is given as-

$$
\overline{y}_{i*}^{*} = \frac{(n_1 \overline{y}_{n1} + n_2 \overline{y}_{w2})}{n}
$$
  
The error terms are defined as-  

$$
(\overline{y} - \overline{Y})
$$

$$
\epsilon_0 = \frac{V_{i*} - I}{\overline{Y}} \tag{1.1}
$$

$$
\epsilon_1 = \frac{(x_{i*} - X)}{\overline{X}}
$$
\n
$$
\epsilon_0^* = \frac{(\overline{y}_{i*}^* - \overline{Y})}{\overline{Y}}
$$
\n(1.2)

The expectations of the error terms are -

$$
E(\epsilon_0) = E(\epsilon_1) = E(\epsilon_0^*) = 0
$$
\n
$$
E(\epsilon_0^2) = \theta \rho_y^* C_y^2
$$
\n
$$
E(\epsilon_1^2) = \theta \rho_x^* C_x^2
$$
\n(1.6)

$$
E(\epsilon_0 \epsilon_1) = \theta \left( \sqrt{\rho_y^* \rho_x^*} \right) \rho C_y C_x
$$
\n
$$
E(\epsilon_0^* \epsilon_1^2) = \theta \rho_t^* C^2 + \frac{(h-1)}{(h-1)} M S^2, \quad h > 1
$$
\n(1.8)

$$
E(\epsilon_0^*) = \theta \rho_y c_y + \frac{\eta * (\overline{Y})^2}{n * (\overline{Y})^2}, \quad i \ge 1
$$
\n
$$
E(\epsilon_0^* \epsilon_1) = \theta \left(\sqrt{\rho_y^* \rho_x^*}\right) \rho C_y C_x \tag{1.9}
$$

Where, 
$$
\theta = \frac{(N-1)}{Nn}
$$
  
\n $\rho_y^* = (1 + (n-1)\rho_y)$   
\n $\rho_x^* = (1 + (n-1)\rho_x)$ 

$$
M = \frac{\tilde{N}_2}{N}
$$

$$
\omega = \rho \frac{C_y}{N}
$$

$$
\omega = \rho \frac{\overline{c}_x}{C_x}
$$

$$
\gamma = \sqrt{\frac{\rho_y^*}{\rho_x^*}}
$$

$$
C_y = \frac{S_y}{Y}
$$

$$
C_x = \frac{3x}{X}
$$

 $\partial = (\sqrt{\rho_y^* \rho_x^*}) \rho C_y C_x$ 

$$
S_{xy} = \frac{1}{(N-1)}\sum_{i=1}^{k}\sum_{j=1}^{n} (x_{ij} - \overline{X})(y_{ij} - \overline{Y})
$$

# **2. Existing Estimators**

 This section lists some existing estimators under systematic sampling scheme following with incorporating non-response and without incorporating non- response along with their mean square error expressions.

# **2.1 Without Incorporating Non-Response**

i) The variance expression of usual unbiased estimator in systematic sampling is given by  $\frac{1}{2}$ 

$$
V(y_{u*})
$$
  
=  $\theta \overline{Y}^2 \rho_y^* C_y^2$  (2.1.1)

ii) The usual ratio estimator and its MSE expression is given by-

$$
\overline{y}_{R*} = \frac{y_{i*}}{\overline{x}_{i*}} \overline{X}
$$
  
\n
$$
mse(\overline{y}_{R*}) = \overline{Y}^2 \theta \left( \rho_y^* C_y^2 + \rho_x^* C_x^2 (1 - 2\omega \gamma) \right)
$$
\n(2.1.2)

iii) The linear regression estimator and its MSE is given by-

$$
\overline{y}_{tr*} = \overline{y}_{i*} + \beta(\overline{X} - \overline{x}_{i*})
$$
  

$$
mse(\overline{y}_{tr*}) = \theta \overline{Y}^2 \rho_y^* C_y^2 (1 - \rho^2)
$$
 (2.1.3)

Where,

 $\beta = \frac{s_{xy}}{s_x^2}$ 

iv) The expression of Verma et al estimators and their MSE's are given by-

$$
t_{1*} = (\bar{y}_{i*}) \left( \frac{\bar{x} - \alpha(\bar{x} - \bar{x}_{i*})}{\bar{x}_{i*} + \alpha(\bar{x} - \bar{x}_{i*})} \right)
$$

$$
t_{2*} = (\bar{y}_{i*}) \left( 2 - \left( \frac{\bar{x}_{i*}}{\bar{x}} \right)^g \right)
$$

$$
mse(t_{1*}) =
$$
  
\n
$$
\theta \bar{Y}^{2} \left( \rho_{y}^{*} C_{y}^{2} + (1 - 2 \alpha_{opt})^{2} \rho_{x}^{*} C_{x}^{2} - 2 (1 - 2 \alpha_{opt}) \sqrt{\rho_{y}^{*} \rho_{x}} \rho C_{y} C_{x} \right)
$$
  
\n
$$
\alpha_{opt} = \frac{1}{2} \left( 1 - \left( \sqrt{\frac{\rho_{y}^{*}}{\rho_{x}^{*}}} \right) \left( \rho \frac{C_{y}}{C_{x}} \right) \right)
$$
\n(2.1.4)

$$
mse(t_{2*}) =
$$
  
\n
$$
\theta \overline{Y}^{2} (g_{opt}^{2} \rho_{x}^{*} C_{x}^{2} + \rho_{y}^{*} C_{y}^{2} - 2g_{opt} (\sqrt{\rho_{y}^{*} \rho_{x}^{*}}) \rho C_{y} C_{x})
$$
  
\n
$$
g_{opt} = \left( \sqrt{\frac{\rho_{y}^{*}}{\rho_{x}^{*}}} \right) (\rho \frac{c_{y}}{c_{x}})
$$
\n(2.1.5)

## **2.2 With Incorporating Non-Response**

 Since the non-response is seen only on study variable Y then the existing estimators take the form-

i)The variance of usual unbiased estimator is given as-

$$
V(\overline{y}_{u*}^{*})
$$
  
=  $\theta \overline{Y}^{2} \rho_{y}^{*} C_{y}^{2} + \frac{(h-1)}{n} M S_{y(2)}^{2}$  (2.2.1)

ii) The usual ratio estimator and its MSE expression is given by-<br> $\frac{\pi^*}{\pi^*}$ 

$$
\overline{y}_{R*}^* = \frac{\overline{y}_{i*}}{\overline{x}_{i*}} \overline{X}
$$
  
\n
$$
mse(\overline{y}_{R*}^*) = \overline{Y}^2 \theta \left( \rho_y^* C_y^2 + \rho_x^* C_x^2 (1 - 2\omega \gamma) \right)
$$
  
\n
$$
+ \frac{(h-1)}{n} M S_{y(2)}^2
$$
\n(2.2.2)

iii) The expression of linear regression estimator and its MSE is given by-

$$
\overline{y}_{lr*}^{*} = \overline{y}_{l*}^{*} + \beta(\overline{X} - \overline{x}_{l*})
$$
\n
$$
mse(\overline{y}_{lr*}^{*}) = \theta \overline{Y}^{2} \rho_{y}^{*} C_{y}^{2} (1 - \rho^{2})
$$
\n
$$
+ \frac{(h-1)}{n} MS_{y(2)}^{2}
$$
\nWhere (2.2.3)

 Where  $\beta = \frac{S_{xy}}{S_x^2}$ 

iv) The expression of Verma et al estimators and their MSE's are given by-

$$
t_{1*}^{*} = (\bar{y}_{t*}^{*}) \left( \frac{\bar{x} - \alpha(\bar{x} - \bar{x}_{t*})}{\bar{x}_{t*} + \alpha(\bar{x} - \bar{x}_{t*})} \right)
$$
  
\n
$$
t_{2*}^{*} = (\bar{y}_{t*}^{*}) \left( 2 - \left( \frac{\bar{x}_{t*}}{\bar{x}} \right)^{g} \right)
$$
  
\n
$$
mse(t_{1*}^{*}) = \theta \bar{Y}^{2} \left( \rho_{y}^{*} C_{y}^{2} + (1 - 2\alpha_{opt})^{2} \rho_{x}^{*} C_{x}^{2} - 2(1 - 2\alpha_{opt}) \sqrt{\rho_{y}^{*} \rho_{x}^{*}} \rho C_{y} C_{x} \right) +
$$
  
\n
$$
\frac{(h-1)}{n} M S_{y(2)}^{2}
$$
  
\n
$$
\alpha_{opt} = \frac{1}{2} \left( 1 - \left( \sqrt{\frac{\rho_{y}^{*}}{\rho_{x}^{*}}} \right) \left( \rho \frac{C_{y}}{C_{x}} \right) \right)
$$
  
\n(2.2.4)

$$
mse(t_{2*}^{*}) =
$$
  
\n
$$
\theta \overline{Y}^{2} (g_{opt}^{2} \rho_{x}^{*} C_{x}^{2} + \rho_{y}^{*} C_{y}^{2} - 2g_{opt} (\sqrt{\rho_{y}^{*} \rho_{x}^{*}}) \rho C_{y} C_{x}) + \frac{(h-1)}{n} M S_{y(2)}^{2}
$$
 (2.2.5)  
\n
$$
g_{opt} = \left(\sqrt{\frac{\rho_{y}^{*}}{\rho_{x}^{*}}}\right) (\rho \frac{c_{y}}{c_{x}})
$$

## **3. Proposed Estimators**

 In this section adopting Mishra et al. (2017) log type estimators for population mean in systematic sampling scheme under two cases i.e. with non-response and without non-response and their mean square error is derived up to the first order of approximation.

$$
t_{S1*}
$$
  
=  $\overline{y} + \varphi \log \left( \frac{\overline{x}}{\overline{x}} \right)$ ,  $\forall \varphi$  is a constant  
 $t_{S2*}$  (3.1)

$$
= \overline{y}(l_1 + 1) + l_2 \log\left(\frac{\overline{x}}{\overline{x}}\right), \forall l_1, l_2 \text{ are constants}
$$
\n(3.2)

# **3.1 Without Incorporating Non-Response**

The MSE of  $t_{S1*}$  and  $t_{S2*}$  are derived up to the first order of approximation and indicated under equation  $(3.1.1)$  and  $(3.1.2)$  respectively.

i) min 
$$
mse(t_{S1*}) = \overline{Y}^2 \theta \rho_y^* C_y^2 (1
$$
  
\t\t\t $-\rho^2$ ) \t\t(3.1.1)  
\t\t\t $\forall \varphi_{opt} = -\overline{Y} \gamma \omega$   
\t\t\tii)  
\t\t\tmin  $mse(t_{S2*}) = a l_{1opt}^2 + bl_{2opt}^2 + 2cl_{1opt} + el_{2opt} + dl_{1opt}l_{2opt} + c$   
\t\t\t $(3.1.2)$   
\t\t\t $a = \overline{Y}^2 (1 + \theta \rho_y^* C_y^2)$   
\t\t\t $b = \theta \rho_x^* C_x^2$   
\t\t\t $c = \overline{Y}^2 \theta \rho_y^* C_y^2$   
\t\t\t $d = 2\overline{Y} \theta \partial - \overline{Y} \theta \rho_x^* C_x^2$   
\t\t\t $e = 2\overline{Y} \theta \partial$   
\t\t\t $l_{1opt} = \frac{(4bc - ed)}{(d^2 - 4ab)}$ ,  $l_{2opt} = \frac{(2ae - 2cd)}{(d^2 - 4ab)}$ 

# **3.2 With Incorporating Non-Response**

The MSE of  $t_{S1*}^*$  and  $t_{S2*}^*$  are derived up to the first order of approximation and indicated under equation (3.2.1) and (3.2.2) respectively.

$$
\begin{aligned}\n\text{i})\text{min}\,mse(t_{51*}^*) &= \\
\theta \overline{Y}^2 \rho_y^* C_y^2 (1 - \rho^2) + \frac{(h-1)}{n} M S_{y(2)}^2 \\
\forall \,\varphi_{opt} &= -\overline{Y} \omega \gamma\n\end{aligned} \tag{3.2.1}
$$

ii)minmse(t<sub>52\*</sub><sup>\*</sup>) = 
$$
(a_{11} - \overline{Y}^2) + l_{1opt}^2 a_{11} + 2l_{1opt} a_{13} + l_{2opt}^2 a_{22} + 2l_{2opt} a_{23} - 2l_{1opt}l_{2opt} a_{12}
$$
  
\n
$$
a_{11} = \overline{Y}^2 + \theta \overline{Y}^2 \rho_y^* C_y^2 + \frac{(h-1)}{n} M S_{y(2)}^2
$$
\n
$$
a_{12} = \frac{\theta \overline{Y} \rho_x^* C_x^2}{2} - \overline{Y} \theta \theta
$$
\n
$$
a_{13} = \theta \overline{Y}^2 \rho_y^* C_y^2 + \frac{(h-1)}{n} M S_{y(2)}^2
$$
\n
$$
a_{22} = \theta \rho_x^* C_x^2
$$
\n
$$
a_{23} = \overline{Y} \theta \theta
$$
\n
$$
l_{1opt} = \frac{(a_{13} a_{22} + a_{23} a_{12})}{(a_{12}^2 - a_{11} a_{22})}
$$
\n
$$
l_{2opt} = \frac{(a_{11} a_{23} + a_{13} a_{12})}{(a_{12}^2 - a_{11} a_{22})}
$$

### **4. Empirical Study**

 In this section we have taken two data sets to make a comparison of efficiency among proposed estimators and existing estimators for population mean adopting systematic sampling scheme under two sections i.e. in first section accounting nonresponse on the study variable and in second section keeping both the auxiliary variable and study variable free from non-response.

Population 1: [Tailor et al. (2013)]

The data is as:

 $N = 15$ ,  $n = 3$ ,  $Y = 80$ ,  $X = 44.47$ ,  $C_y = 0.56$ ,  $C_x = 0.28$  $\rho = 0.9848,$   $\rho_y = 0.6652,$   $\rho_x = 0.707$ Population 2: [Singh and Chaudhary (1986), P. 177]

Let y= Cultivated area in acres in 1974 census. x= Cultivated area in acres in 1971 census. The data is as:  $N = 128$ ,  $n = 25$ ,  $\overline{Y} = 853.5$ ,  $\overline{X} = 3243.352$ ,  $C_y = 0.6729$ ,  $C_x =$ 0.6045  $\rho = 0.8311$ ,  $\rho_y = 0.036$ ,  $\rho_x = 0.076$ 



**Population 1** 



**Population 2** 

**Table 1: PRE's of existing and proposed estimators incorporating the nonresponse on study variable** 

	$\overline{y}_{u*}$	$t_{S1*}$	$t_{S2*}$	$\overline{y}_{R*}$	$\overline{y}_{lr*}$	$t_{1*}$	$t_{2*}$
Pop 1	100	3314.665	7159.617	389.621		3314.65 3314.665	3314.665
Pop 2	100	323.339	344.745	259.941	323.339	323.339	323.339

**Table 2: PRE's of existing and proposed estimators without incorporating nonresponse** 

#### **5. Conclusion**

 Table 1 depicts the impact of incorporating non-response on study variable with systematic sampling scheme results in the decrease of the efficiency of estimators as the non-response rate M and the value of constant h increases the value of percentage relative efficiency of the proposed as well as the existing estimators decreases. It can also be seen that the efficiency of  $t_{51}^*$  is better than usual unbiased estimator, ratio estimator and equivalent to the liner regression estimator as well as Verma et al's estimators and the efficiency of the proposed estimator  $t_{s2}^*$  is better among all mentioned existing estimators.

 While in Table 2, it can be seen that without incorporating non-response on study variable as well as auxiliary variable the pattern of the efficiency of proposed estimator  $t_{\text{S1*}}$  is better than usual unbiased estimator, ratio estimator and equivalent to the liner regression estimator as well as Verma et al's estimators while the efficiency of proposed estimator  $t_{s2*}$  is better among the mentioned existing estimators.

Hence the proposed estimators can be used in practice.

#### **References**

- 1. Cochran, W. G. (1977). Sampling Techniques:  $3<sup>rd</sup>$  Ed. New York: Wiley.
- 2. Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys, Journal of the American Statistical Association, 41(236), p. 517-529.
- 3. Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling, Applied Mathematics and Computation, 151(3), p. 893-902.
- 4. Khan, M. and Singh, R. (2015). Estimation of population mean in chain ratiotype estimator under systematic sampling, Journal of Probability and Statistics, Volume 2015, Article ID 248374, http://dx.doi.org/10.1155/2015/ 248374, 5 pages.
- 5. Malik, S. and Singh, R. (2013) : An improved estimator using two auxiliary attributes, Appli. Math. Compt., 219, p. 10983-10986.
- 6. Mishra, P. and Singh, R. (2017): A new log-product-type estimator using auxiliary information, Jour. Sci. Res., 61 (1&2), p. 179-183.
- 7. Pal, S. K. and Singh, H. P. (2017). Estimation of finite population mean using auxiliary information in systematic sampling, International Journal of System Assurance Engineering and Management, 8(2), p. 1392-1398.
- 8. Singh, D. and Chaudhary, F. S. (1986). Theory and analysis of sample survey designs, John Wiley & Sons.
- 9. Singh, R. and Kumar, M. (2011). A note on transformations on auxiliary variable in survey sampling, Mod. Assis. Stat. Appl., 6(1), p. 17-19.
- 10. Singh, R., Kumar, M., Chaudhary, M.K. and Kadilar, C. (2009). Improved exponential estimator in stratified random sampling, Pak. J. Stat. Oper. Res. 5(2), p. 67-82.
- 11. Singh, R., Kumar, M. and Smarandache, F. (2010): Ratio estimators in simple random sampling when study variable is an attribute, WASJ, 11(5), p. 586- 589.
- 12. Singh, R. and Malik, S. (2014). Improved estimator of population variance using information on auxiliary attribute in simple random sampling, Appl. Math. Comp., 235, p. 43-49.
- 13. Singh, R., Malik, S., Chaudhary, M. K.,Verma, H. K. and Adewara, A. A. (2012). A general family of ratio-type estimators in systematic sampling, Journal of Reliability and Statistical Studies, 5(1), p. 73-82.
- 14. Singh, R., Malik, S. and Singh, V. K. (2013). An improved estimator in systematic sampling. arXiv preprint arXiv:1307.5145.
- 15. Singh, P., Singh, R. and Bouza, C.N. (2018). Effect of measurement error and non-response on estimation of population mean, Revista Investigacion Operacional, 39(1), p. 108-120.
- 16. 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), p. 424-435.
- 17. Tailor, R., Jatwa, N. K. and Tailor, R. (2014). Ratio-cum-product estimator of population mean in systematic sampling using known parameters of auxiliary variates, Journal of Reliability and Statistical Studies, 7(2), p. 129-138.
- 18. Verma, H. K., Singh, R. D. and Singh, R. (2014). Some improved estimators in systematic sampling under non-response, National Academy Science Letters, 37(1), p. 91-95.