



# Estimation of Population Mean Using Auxiliary Attribute in The Presence Of Non-Response

Kamlesh Kumar<sup>1</sup> and Anupam Kumar<sup>2</sup>

<sup>1</sup>Department of Community Medicine, Rajkiya Medical College, Jalaun (Orai), UP, India

<sup>2</sup>Department of Statistics, PUC, Mizoram University, Aizawl, India

Received August 21, 2018, Revised February 10, 2019, Accepted March 13, 2019, Published May 1, 2019

**Abstract:** In this paper, exponential ratio and product type estimators for population mean of study character using known population proportion of auxiliary attribute in the presence of non-response have been proposed. The expressions for the mean square error of the proposed estimators have been obtained. The proposed estimators have been compared with the relevant estimators. The empirical studies have been done to demonstrate the efficiency of the proposed estimators over other relevant estimator.

**Keywords:** Auxiliary attribute, Mean square error, Non-response, Study character.

## 1. INTRODUCTION

It is well known that the estimators for population mean using information on auxiliary character give the better estimate in comparison to the usual estimator. Sometimes the auxiliary information is available in the form of attribute. Using information on auxiliary attribute we can also find the better estimate of population mean. For example, to estimate the person's weight, crop production, milk production and income, we can use gender, seed type, cow's breed and ownership of a house as auxiliary attributes. The research works for estimating the population mean using auxiliary attributes have been done by Naik and Gupta [1], Jhajj et al. [2], Shabbir and Gupta [3,4], Singh et.al [5], Abd-Elfattah et al. [6], Singh and Solanki [7] and Solanki and Singh [8].

In sample surveys, sometimes we do not get the complete information due to the occurrence of non-response. To reduce the effect of non-response in such situations, Hansen and Hurwitz [9] first suggested an unbiased estimator for estimating the population mean. Further, using the known population mean of auxiliary character, several research works for estimating the population mean in the presence of non-response have been done by Cochran [10], Rao [11,12], Khare and Srivastava [13,14], Singh et al. [15] and Kumar and Kumar [16].

In this paper, we have proposed exponential ratio and product type estimators for population mean using known population proportion of auxiliary attribute in the presence of non-response. The expressions for mean square error of the proposed estimators are obtained. The conditions under which the proposed estimators are more efficient in comparison to the relevant estimator are obtained. The empirical studies are also given to show the performance of the proposed estimators in comparison to the relevant estimators.

## 2. THE EARLIAR AND PROPOSED ESTIMATORS

Let  $Y_i$  and  $\phi_i : i=1\dots N$  denote the values of study character  $y$  and auxiliary attribute  $\phi$  for  $i^{th}$  unit of the population with population mean  $\bar{Y}$  and population proportion  $P$  respectively. Here  $\phi_i$  will take two values 1 and 0 if it possesses and does not possess the attribute respectively. The entire population is supposed to be divided into  $N_1$  responding and  $N_2$  non-responding units such that  $N = N_1 + N_2$ . Let  $P$  be the proportion of the units in the population possessing the attribute  $\phi$ ,  $P_1$  be the proportion of the units in the responding part of the population possessing the attribute  $\phi$  and  $P_2$  be the proportion of the units in the non-responding part of the population possessing the attribute



$\phi$ . Let a sample of size  $n (< N)$  is drawn from the population of size  $N$  by using simple random sampling without replacement (SRSWOR) method and it has been found that  $n_1$  units respond and  $n_2$  units do not respond in a sample of size  $n$  for the study character  $y$  and auxiliary attribute  $\phi$ . Further a subsample of size  $m = (n_2 / k, k > 1)$  is drawn from  $n_2$  non responding units by using SRSWOR method of sampling and collects the information on study character  $y$  and on auxiliary attribute  $\phi$  by personal interview. Using the technique of Hansen & Hurwitz [9], the estimators  $\bar{y}^*$  and  $p^*$  for the population mean  $\bar{Y}$  and for the population proportion  $P$  based on  $(n_1 + m)$  units are obtained as

$$\bar{y}^* = \frac{n_1}{n} \bar{y}_1 + \frac{n_2}{n} \bar{y}'_2 \quad (1)$$

and

$$p^* = \frac{n_1}{n} p_1 + \frac{n_2}{n} p'_2 \quad (2)$$

where  $(\bar{y}_1, \bar{y}'_2)$  are the means of study character  $y$  based on  $n_1$  responding units and  $m$  subsample units and  $(p_1, p'_2)$  are the proportions of the units possessing the attribute  $\phi$  in  $n_1$  responding units and in  $m$  subsample units respectively.

The mean square error (MSE) of the estimators  $\bar{y}^*$  and  $p^*$  are given as:

$$MSE(\bar{y}^*) = \frac{f}{n} S_y^2 + \frac{W_2(k-1)}{n} S_y'^2 \quad (3)$$

and

$$MSE(p^*) = \frac{f}{n} S_\phi^2 + \frac{W_2(k-1)}{n} S_\phi'^2, \quad (4)$$

where  $f = 1 - \frac{n}{N}$ ,  $W_2 = \frac{N_2}{N}$  and  $(S_y^2, S_y'^2)$  and  $(S_\phi^2, S_\phi'^2)$  are population mean squares of study character  $y$  and auxiliary attribute  $\phi$  for responding and non-responding part of the population.

In case when population proportion of auxiliary attribute is known, Singh et.al [5] proposed the exponential ratio and product type estimators for population mean of study character which are given as follows:

$$T_1 = \bar{y} \exp\left(\frac{P-p}{P+p}\right) \quad (5)$$

and

$$T_2 = \bar{y} \exp\left(\frac{p-P}{p+P}\right), \quad (6)$$

where  $\bar{y}$  is the sample mean based on  $n$  units and  $p$  is the proportion of the units possessing the attribute  $\phi$  in the sample of size  $n$  units.

Further Solanki and Singh [8] proposed the class of estimators for population mean of study character which is given as:

$$T_3 = \bar{y} \exp\left\{\alpha \left(\frac{P-p}{P+p}\right)\right\}, \quad (7)$$

where  $\alpha$  is a suitably chosen scalar.

In case when non-response occurs both on study character and on auxiliary attribute, following the Singh et.al [5] and Solanki and Singh [8] estimators, we propose conventional exponential ratio, exponential product and generalized estimators for population mean of study character in the presence of non-response which are given as follows:



$$t_1 = \bar{y}^* \exp\left(\frac{P-p^*}{P+p^*}\right), \quad t_2 = \bar{y}^* \exp\left(\frac{p^*-P}{p^*+P}\right) \quad (8)$$

and

$$t_3 = \bar{y}^* \exp\left\{\alpha_1 \left(\frac{P-p^*}{P+p^*}\right)\right\}, \quad (9)$$

where  $\alpha_1$  is a suitably chosen scalar.

If  $\alpha_1 = 0$  then  $t_3 = \bar{y}^*$ , if  $\alpha_1 = 1$  then  $t_3 = t_1$  and if  $\alpha_1 = -1$  then  $t_3 = t_2$ .

In case when non-response occurs only on study character, following the Singh et.al [5] and Solanki and Singh [8] estimators, we propose alternative exponential ratio, exponential product and generalized estimators for population mean of study character in the presence of non-response which are given as follows:

$$t_4 = \bar{y}^* \exp\left(\frac{P-p}{P+p}\right), \quad t_5 = \bar{y}^* \exp\left(\frac{p-P}{p+P}\right) \quad (10)$$

and

$$t_6 = \bar{y}^* \exp\left\{\alpha_2 \left(\frac{P-p}{P+p}\right)\right\} \quad (11)$$

where  $\alpha_2$  is a suitably chosen scalar.

If  $\alpha_2 = 0$  then  $t_6 = \bar{y}^*$ , if  $\alpha_2 = 1$  then  $t_6 = t_4$  and if  $\alpha_2 = -1$  then  $t_6 = t_5$ .

### 3. THE MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS

In order to derive the expression for the mean square error of the proposed estimators.

Let  $\bar{y}^* = \bar{Y}(1 + \varepsilon_0)$ ,  $p^* = P(1 + \varepsilon_1)$ ,  $p = P(1 + \varepsilon_2)$  such that  $E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0$ .

By using simple random sampling without replacement method of sampling, we have

$$E(\varepsilon_0^2) = \frac{1}{\bar{Y}^2} V(\bar{y}^*) = \frac{1}{\bar{Y}^2} \left[ \frac{f}{n} S_y^2 + \frac{W_2(k-1)}{n} S_y'^2 \right],$$

$$E(\varepsilon_1^2) = \frac{1}{P^2} V(p^*) = \frac{1}{P^2} \left[ \frac{f}{n} S_\phi^2 + \frac{W_2(k-1)}{n} S_\phi'^2 \right],$$

$$E(\varepsilon_2^2) = \frac{1}{P^2} V(p) = \frac{f}{n} \frac{S_\phi^2}{P^2}, \quad E(\varepsilon_0 \varepsilon_1) = \frac{1}{\bar{Y}P} COV(\bar{y}^*, p^*) = \frac{1}{\bar{Y}P} \left[ \frac{f}{n} S_{y\phi} + \frac{W_2(k-1)}{n} S_{y\phi}' \right],$$

$$E(\varepsilon_0 \varepsilon_2) = \frac{1}{\bar{Y}P} COV(\bar{y}^*, p) = \frac{1}{\bar{Y}P} \frac{f}{n} S_{y\phi},$$

where  $S_{y\phi} = \rho_{y\phi} S_y S_\phi$ ,  $S_{y\phi}' = \rho_{y\phi}' S_y' S_\phi'$  and  $(\rho_{y\phi}, \rho_{y\phi}')$  are point bi-serial correlation coefficients between study character  $y$  and auxiliary attribute  $\phi$  for responding and not responding units of the population.

The expressions for the *MSE* of the proposed estimators  $t_1, t_2, t_3, t_4, t_5$  and  $t_6$  up to the terms of order  $n^{-1}$  are given as follows:



$$MSE(t_1) = MSE(\bar{y}^*) + \bar{Y}^2 \left[ A \left\{ \frac{1}{4} C_p^2 - C_{yp} \right\} + B \left\{ \frac{1}{4} C_p'^2 - C'_{yp} \right\} \right], \quad (12)$$

$$MSE(t_2) = MSE(\bar{y}^*) + \bar{Y}^2 \left[ A \left\{ \frac{1}{4} C_p^2 + C_{yp} \right\} + B \left\{ \frac{1}{4} C_p'^2 + C'_{yp} \right\} \right], \quad (13)$$

$$MSE(t_3) = MSE(\bar{y}^*) + \bar{Y}^2 \left[ A \left\{ \frac{1}{4} \alpha_1^2 C_p^2 - \alpha_1 C_{yp} \right\} + B \left\{ \frac{1}{4} \alpha_1^2 C_p'^2 - \alpha_1 C'_{yp} \right\} \right], \quad (14)$$

$$MSE(t_4) = MSE(\bar{y}^*) + \bar{Y}^2 \left[ A \left\{ \frac{1}{4} C_p^2 - C_{yp} \right\} \right], \quad (15)$$

$$MSE(t_5) = MSE(\bar{y}^*) + \bar{Y}^2 \left[ A \left\{ \frac{1}{4} C_p^2 + C_{yp} \right\} \right] \quad (16)$$

and

$$MSE(t_6) = MSE(\bar{y}^*) + \bar{Y}^2 \left[ A \left\{ \frac{1}{4} \alpha_2^2 C_p^2 - \alpha_2 C_{yp} \right\} \right] \quad (17)$$

Differentiating equations (14) and (17) with respect to  $\alpha_1$  and  $\alpha_2$  and equating it to zero, we get the optimum values of  $\alpha_1$  and  $\alpha_2$  as

$$\alpha_{1opt} = \frac{2 \{ AC_{yp} + BC'_{yp} \}}{\{ AC_p^2 + BC_p'^2 \}} \quad (18)$$

$$\text{and } \alpha_{2opt} = \frac{2\rho_{y\phi} C_y}{C_p}. \quad (19)$$

By putting the optimum values of  $\alpha_1$  and  $\alpha_2$  in equations (14) and (17), we get the minimum  $MSE$  of the estimators  $t_3$  and  $t_6$  as:

$$MSE(t_3)_{\min} = MSE(\bar{y}^*) - \bar{Y}^2 \left[ \frac{\{ AC_{yp} + BC'_{yp} \}^2}{\{ AC_p^2 + BC_p'^2 \}} \right] \quad (20)$$

and

$$MSE(t_6)_{\min} = MSE(\bar{y}^*) - \bar{Y}^2 A \rho_{y\phi}^2 C_y^2, \quad (21)$$

where  $A = \left( \frac{1}{n} - \frac{1}{N} \right)$ ,  $B = \frac{W_2(k-1)}{n}$ ,  $C_{yp} = \rho_{y\phi} C_y C_p$ ,  $C'_{yp} = \rho'_{y\phi} C'_y C'_p$ ,  $C_y = \frac{S_y}{\bar{Y}}$ ,  $C_p = \frac{S_\phi}{P}$ ,  $C'_y = \frac{S'_y}{\bar{Y}}$ ,

$$C'_p = \frac{S'_\phi}{\bar{X}}, \quad S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2, \quad S_\phi^2 = \frac{1}{N-1} \sum_{i=1}^N (\phi_i - P)^2, \quad S_y'^2 = \frac{1}{N_2-1} \sum_{i=1}^{N_2} (Y_i - \bar{Y}_2)^2,$$

$S_\phi'^2 = \frac{1}{N_2-1} \sum_{i=1}^{N_2} (\phi_i - P_2)^2$  and  $(\bar{Y}_2, P_2)$  are population mean and population proportion of study character  $y$  and auxiliary attribute  $\phi$  for the non-responding units ( $N_2$ ) of the population.



#### 4. COMPARISON OF THE PROPOSED ESTIMATORS WITH RELEVANT ESTIMATORS

Comparing the proposed conventional estimators ( $t_1, t_2, t_3$ ) and alternative estimators ( $t_4, t_5, t_6$ ) with relevant estimator  $\bar{y}^*$ .

$$MSE(t_1) < MSE(\bar{y}^*)$$

$$\text{If } \rho_{y\phi} > \frac{1}{4} \frac{C_p}{C_y} \quad \text{and} \quad \rho'_{y\phi} > \frac{1}{4} \frac{C'_p}{C'_y}$$

$$MSE(t_2) < MSE(\bar{y}^*)$$

$$\text{If } \rho_{y\phi} < -\frac{1}{4} \frac{C_p}{C_y} \quad \text{and} \quad \rho'_{y\phi} < -\frac{1}{4} \frac{C'_p}{C'_y}$$

$$MSE(t_3) < MSE(\bar{y}^*)$$

$$\text{If } 0 < \alpha_1 < \frac{4\{AC_{yp} + BC'_{yp}\}}{\{AC_p^2 + BC_p'^2\}}$$

and

$$MSE(t_4) < MSE(\bar{y}^*) \quad \text{If } \rho_{y\phi} > \frac{1}{4} \frac{C_p}{C_y}$$

$$MSE(t_5) < MSE(\bar{y}^*) \quad \text{If } \rho_{y\phi} < -\frac{1}{4} \frac{C_p}{C_y}$$

$$MSE(t_6) < MSE(\bar{y}^*) \quad \text{If } 0 < \alpha_2 < \frac{4\rho_{y\phi}C_y}{C_p}$$

#### 5. EMPIRICAL STUDIES

##### 5.1 Data Set I

Source: Sinha [17]

One hundred nine village population of urban area under police station-Baria, Tahsil – Champua, Orissa has been taken under consideration from district Census Handbook, 1981, Orissa, published by Govt. of India. The first 25% villages (i.e. 27 villages) have been considered as non-responding group of the population. In this data set, we have considered the study character  $y$  and auxiliary attribute  $\phi$  as  $y$  -number of literate persons in the village and  $\phi$  - number of literate persons greater than equal to 150. The values of parameters are as follows:

$$\begin{aligned} \bar{Y} &= 145.30, & P &= 0.3211, & S_y &= 111.3835, & S_\phi &= 0.4691, & C_y &= 0.7666, & C_p &= 1.4608, \\ \bar{Y}_2 &= 189.52, & P_2 &= 0.5185, & S'_y &= 121.204, & S'_\phi &= 0.5092, & C'_y &= 0.6395, & C'_p &= 0.9820, \\ \rho_{y\phi} &= 0.742, & \rho'_{y\phi} &= 0.672 \end{aligned}$$

TABLE 1. RELATIVE EFFICIENCY (RE) OF THE PROPOSED ESTIMATORS WITH RESPECT TO  $\bar{y}^*$  ( $N=109, n=30$ )

Estimators	1/k		
	1/4	1/3	1/2
	RE	RE	RE
$\bar{y}^*$	100.00	100.00	100.00
$t_1$	159.03	165.58	177.07



$t_3$	129.64	138.84	156.33
$t_4$	164.75	172.99	187.84
$t_6$	133.10	143.74	164.47

## 5.2 Data Set II

Source: Sinha [17]

A list of 70 villages in a tehsil of India along with their population in 1981 and cultivated area in the same year is given. The first 30% villages (i.e. 21 villages) have been considered as non-responding group of the population. In this data set, we have considered the study character  $y$  and auxiliary attribute  $\phi$  as  $y$  - cultivated areas (in acres) in the village and  $\phi$  -total population of village greater than equal to 1100. The values of parameters are as follows:

$$\begin{aligned} \bar{Y} &= 981.29, & P &= 0.6000, & S_y &= 613.3558, & S_\phi &= 0.4934, & C_y &= 0.6253, & C_p &= 0.8224, \\ \bar{Y}_2 &= 12168, & P_2 &= 0.7619, & S'_y &= 430.3177, & S'_\phi &= 0.4364, & C'_y &= 0.3536, & C'_p &= 0.5728, \\ \rho_{y\phi} &= 0.596, & \rho'_{y\phi} &= 0.690. \end{aligned}$$

TABLE 2. RELATIVE EFFICIENCY (RE) OF THE PROPOSED ESTIMATORS WITH RESPECT TO  $\bar{y}^*$  ( $N=70, n=25$ )

Estimators	1/k		
	1/4	1/3	1/2
	RE	RE	RE
$\bar{y}^*$	100.00	100.00	100.00
$t_1$	149.41	150.49	151.99
$t_3$	126.30	131.75	140.06
$t_4$	150.70	151.69	153.07
$t_6$	126.66	132.20	140.66

## 6. CONCLUSION

From table 1 and 2 it has been observed that for the fixed sample size  $n$  and different values of  $k$ , the proposed estimators  $t_1$  and  $t_3$  are most efficient than the usual estimator  $\bar{y}^*$  and the proposed estimators  $t_4$  and  $t_6$  are most efficient than the corresponding estimators  $t_1$  and  $t_3$ . So, on the basis of theoretical and empirical studies, we may conclude that the proposed estimators  $t_1, t_3, t_4$  and  $t_6$  using auxiliary attribute are more efficient than the usual estimator  $\bar{y}^*$ .

## REFERENCES

- [1] V. D., Naik and P. C. Gupta, "A note on estimating of mean with known population of an auxiliary character," Jour. Ind. Soc. Ag. Stat., vol.48, pp. 151-158, 1996.
- [2] H. S. Jhaji, M. K. Sharma and L. K. Grover, "A family of estimators of population mean using information on auxiliary attribute," Pak. Jour. Stat., vol.22, pp. 43-50, 2006.
- [3] J. Shabbir and S. Gupta, "On estimating the finite population mean with known population proportion of an auxiliary variable," Pak. Jour. Stat., vol.23, pp. 1-9, 2007.
- [4] J. Shabbir and S. Gupta, "Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes," Hacett. Jour. Math. Stat., vol.39, pp. 121-129, 2010.
- [5] R. Singh, P. Chouhan, N. Sawan and F. Smarandache, "Ratio-product Type Exponential Estimator for Estimating Finite Population Mean Using Information on Auxiliary Attribute," Renaissance High Press, USA, pp. 18-32, 2007.



- [6] A. M. Abd-Elfattah, E. A. El-Sherpieny, S. M. Mohamed and O. F. Abdou, "Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute," *Appl. Math. Comp.*, vol.215, pp. 4198-4202, 2010.
- [7] H. P. Singh and R. S. Solanki, "Improved estimation of population mean in simple random sampling using information on auxiliary attribute," *Appl. Math. Comp.*, vol.218, pp. 7798-7812, 2012.
- [8] R. S. Solanki and H. P. Singh, "Improved estimation of population mean using population proportion of an auxiliary character," *Chil. Jour. Stat.* vol.4(1), pp. 3-17, 2013.
- [9] M. H. Hansen and W. N. Hurwitz, "The problems of non-response in sample surveys," *Jour. Amer. Stat. Assoc.*, vol.41, pp. 517-529, 1946.
- [10] W. G. Cochran, *Sampling techniques*, 3rd edn. Wiley, New York, 1977.
- [11] P. S. R. S. Rao, "Ratio estimation with sub sampling the non-respondents," *Survey Methodology*, vol.12, pp. 217-230, 1986.
- [12] P. S. R. S. Rao, "Ratio and regression estimates with sub sampling the non- respondents", Paper presented at a special contributed session of the International Statistical Association Meeting, Sept. 2-16, Tokyo, Japan. 1987.
- [13] B. B. Khare and S. Srivastava, "Transformed product type estimators for population mean in presence of soft-core observations," *Proc. Math. Soci. B.H.U.*, vol.12, pp. 29-34, 1996.
- [14] B. B. Khare and S. Srivastava, "Transformed ratio type estimators for the population mean in the presence of non-response," *Communi. Stat.-Theory and Methods*, vol.26, pp. 1779-1791, 1997.
- [15] R. Singh, M. Kumar, M. K. Chaudhary and F. Smarandache, "Estimation of population mean in presence of non-response using exponential estimator", Paper published in the Book *Multispace & Multistructure Neutrosophic Transdisciplinarity* vol.4, pp. 758-768, 2010.
- [16] K. Kumar and M. Kumar "Improved exponential ratio and product type estimators for population mean in the presence of non-response," *Bullen. Math. Stats. Rese.*, vol.5, pp. 68-76, 2017.
- [17] R. R. Sinha, "Estimation of ratio of two population means using auxiliary attributes in presence of non-response," *Int. Jour. Rec. Tren. Life Sci. Maths* vol.1(2), pp. 01-07, 2014.