A New Estimator For Population Mean Using Two Auxiliary Variables in Stratified random Sampling

Rajesh Singh and Sachin Malik
Department of Statistics, Banaras Hindu University
Varanasi-221005, India
(sachinkurava999@gmail.com, rsinghistat@gmail.com)

Abstract

In this paper, we suggest an estimator using two auxiliary variables in stratified random sampling. The propose estimator has an improvement over mean per unit estimator as well as some other considered estimators. Expressions for bias and MSE of the estimator are derived up to first degree of approximation. Moreover, these theoretical findings are supported by a numerical example with original data.

Key words: Study variable, auxiliary variable, stratified random sampling, bias and mean squared error.

1. Introduction

The problem of estimating the population mean in the presence of an auxiliary variable has been widely discussed in finite population sampling literature. Out of many ratio, product and regression methods of estimation are good examples in this context. Diana [2] suggested a class of estimators of the population mean using one auxiliary variable in the stratified random sampling and examined the MSE of the estimators up to the kth order of approximation. Kadilar and Cingi [3], Singh et al. [7], Singh and Vishwakarma [8], Koyuncu and Kadilar [4] proposed estimators in stratified random sampling. Singh [9] and Perri [6] suggested some ratio cum product estimators in simple random sampling. Bahl and Tuteja [1] and Singh et al. [11] suggested some exponential ratio type estimators. In this chapter, we suggest some exponential-type estimators using the auxiliary information in the stratified random sampling.
Consider a finite population of size $N$ and is divided into $L$ strata such that $\sum_{h=1}^{L} N_h = N$ where $N_h$ is the size of $h^{th}$ stratum ($h=1,2,...,L$). We select a sample of size $n_h$ from each stratum by simple random sample without replacement sampling such that $\sum_{h=1}^{L} n_h = n$, where $n_h$ is the stratum sample size. A simple random sample of size $n_h$ is drawn without replacement from the $h^{th}$ stratum such that $\sum_{i=1}^{n_h} y_{hi} = y_{h}$, $x_{hi}$, $z_{hi}$ denote the observed values of $y$, $x$, and $z$ on the $i^{th}$ unit of the $h^{th}$ stratum, where $i=1, 2, 3...N_h$.

To obtain the bias and MSE, we write

$$
\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h = \bar{Y}(1 + e_0), \quad \bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h = \bar{X}(1 + e_1), \quad \bar{z}_{st} = \sum_{h=1}^{L} w_h \bar{z}_h = \bar{Z}(1 + e_2)
$$

Such that,

$$
E(e_0) = E(e_0) = E(e_0) = 0
$$

$$
V_{rs} = \sum_{h=1}^{L} w_h^{r+s+1} E \left[ \frac{\bar{y}_h - \bar{Y}}{\bar{Y}} \frac{\bar{x}_h - \bar{X}}{\bar{X}} \frac{\bar{z}_h - \bar{Z}}{\bar{Z}} \right]
$$

where,

$$
\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h, \quad \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}, \quad \bar{Y} = \frac{1}{N_h} \sum_{i=1}^{n_h} \bar{Y}_{hi}
$$

$$
Y = \bar{Y}_{st} = \sum_{h=1}^{L} w_h \bar{Y}_h, \quad w_h = \frac{N_h}{N}
$$

and

$$
V(\bar{y}_{st}) = \bar{Y}^2 V_{200}
$$

(1.1)

Similar expressions for $X$ and $Z$ can also be defined.
\[
E(e^2_0) = \frac{\sum_{h=1}^{L} W^2 f \cdot S^2_{y_h}}{Y^2} = V_{200},
\]

\[
E(e^2_2) = \frac{\sum_{h=1}^{L} W^2 f \cdot S^2_{z_h}}{Z^2} = V_{002},
\]

\[
E(e^2_0e_2) = \frac{\sum_{h=1}^{L} W^2 f \cdot S^2_{y_z}}{YZ} = V_{101},
\]

\[
E(e^2_0e_1) = \frac{\sum_{h=1}^{L} W^2 f \cdot S^2_{y_x}}{XY} = V_{110},
\]

\[
E(e_1e_2) = \frac{\sum_{h=1}^{L} W^2 f \cdot S^2_{x_z}}{XZ} = V_{011},
\]

where \( S^2_{y_h} = \sum_{i=1}^{N_h} \left( \frac{Y_h - \bar{Y}_h}{N_h - 1} \right)^2 \), \( S^2_{z_h} = \sum_{i=1}^{N_h} \left( \frac{Z_h - \bar{Z}_h}{N_h - 1} \right)^2 \)

\[ S^2_{y_z} = \sum_{i=1}^{N_h} \left( \frac{Z_h - \bar{Z}_h}{N_h - 1} \right) \left( \frac{Y_h - \bar{Y}_h}{N_h - 1} \right) \]

\[ S^2_{x_z} = \sum_{i=1}^{N_h} \left( \frac{X_h - \bar{X}_h}{N_h - 1} \right) \left( \frac{Z_h - \bar{Z}_h}{N_h - 1} \right) \]

and \( f_h = \frac{1}{n_h} - \frac{1}{N_h} \)

2. Estimators in literature

In order to have an estimate of the study variable \( y \), assuming the knowledge of the population proportion \( P \), Naik and Gupta [5] and Singh et al. [11] respectively proposed following estimators

\[
t_1 = \bar{y}_{st} \left( \frac{\bar{X}}{X_{st}} \right)
\]

(2.1)

\[
t_2 = \bar{y}_{st} \exp \left( \frac{\bar{X} - X_{st}}{X + X_{st}} \right)
\]

(2.2)

The MSE expressions of these estimators are given as
\[ \text{MSE}(t_1) = \bar{Y}^2 \left[ V_{200} + V_{020} - 2V_{110} \right] \tag{2.3} \]

\[ \text{MSE}(t_2) = \bar{Y}^2 \left[ V_{200} + \frac{V_{020}}{4} - V_{110} \right] \tag{2.4} \]

When the information on the two auxiliary variables is known, Singh [10] proposed some ratio cum product estimators in simple random sampling to estimate the population mean of the study variable \( y \).

Motivated by Singh [10] and Singh et al. [7], Singh and kumar propose some estimators in stratified sampling as

\[ t_3 = \bar{y}_{st} \exp \left[ \frac{X - \bar{x}_{st}}{\bar{X} + x_{st}} \right] \exp \left[ \frac{Z - \bar{z}_{st}}{\bar{Z} + z_{st}} \right] \tag{2.5} \]

\[ t_4 = \bar{y}_{st} \exp \left[ \frac{\bar{x}_{st} - \bar{X}}{x_{st} + \bar{X}} \right] \exp \left[ \frac{\bar{z}_{st} - \bar{Z}}{z_{st} + \bar{Z}} \right] \tag{2.6} \]

\[ t_5 = \bar{y}_{st} \exp \left[ \frac{\bar{X} - \bar{x}_{st}}{X + x_{st}} \right] \exp \left[ \frac{\bar{Z} - \bar{z}_{st}}{Z + z_{st}} \right] \tag{2.7} \]

\[ t_6 = \bar{y}_{st} \exp \left[ \frac{\bar{x}_{st} - \bar{X}}{x_{st} + \bar{X}} \right] \exp \left[ \frac{\bar{Z} - \bar{z}_{st}}{Z + z_{st}} \right] \tag{2.8} \]

The MSE equations of these estimators can be written as

\[ \text{MSE}(t_3) = \bar{Y}^2 \left[ V_{200} + \frac{V_{020}}{4} + \frac{V_{002}}{4} - V_{110} - V_{101} + \frac{V_{011}}{2} \right] \tag{2.9} \]

\[ \text{MSE}(t_4) = \bar{Y}^2 \left[ V_{200} + \frac{V_{020}}{4} + \frac{V_{002}}{4} + V_{110} + V_{101} + \frac{V_{011}}{2} \right] \tag{2.10} \]

\[ \text{MSE}(t_5) = \bar{Y}^2 \left[ V_{200} + \frac{V_{020}}{4} + \frac{V_{002}}{4} - V_{110} + V_{101} - \frac{V_{011}}{2} \right] \tag{2.11} \]
\[
\text{MSE}(t_6) = \bar{Y}^2 \left[ \frac{V_{200} + V_{020}}{4} + \frac{V_{002} + V_{110} - V_{441}}{2} \right]
\]

(2.12)

When there are two auxiliary variables, the regression estimator of \( \bar{Y} \) will be
\[
t_7 = \bar{y}_{st} + b_{1h} (\bar{X} - \bar{x}_{st}) + b_{2h} (\bar{Z} - \bar{z}_{st})
\]

(2.13)

Where \( b_{1h} = \frac{s_{yx}}{s_x^2} \) and \( b_{2h} = \frac{s_{yz}}{s_z^2} \). Here \( s_x^2 \) and \( s_z^2 \) are the sample variances of \( x \) and \( z \) respectively, \( s_{yx} \) and \( s_{yz} \) are the sample covariance’s between \( y \) and \( x \) and between \( z \) respectively. The MSE expression of this estimator is:
\[
\text{MSE}(t_7) = \sum_{h=1}^{l} W_h^2 f_h s_{yh}^2 \left( 1 - \rho_{yhx}^2 - \rho_{yhz}^2 + 2 \rho_{yhx} \rho_{yhz} \rho_{xh} \right)
\]

(2.14)

3. The proposed estimator

We suggest using the ratio estimator given in equation (2.5) instead of estimator given in equation (2.13). By this way, we obtain the following estimator
\[
t_p = \bar{y}_{st} \exp \left[ \frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}} \right]^{m_1} \exp \left[ \frac{\bar{Z} - \bar{z}_{st}}{\bar{Z} + \bar{z}_{st}} \right]^{m_2} + b_{1h} (\bar{X} - \bar{x}_{st}) + b_{2h} (\bar{Z} - \bar{z}_{st})
\]

(3.1)

Expressing equation (3.1) in terms of e’s, we have
\[
t_p = \bar{Y} \left( 1 + e_0 \right) \left\{ \exp \left( \frac{-e_1}{2 + e_1} \right)^{m_1} \exp \left( \frac{-e_2}{2 + e_2} \right)^{m_2} \right\} - b_{1h} \bar{X} e_1 - b_{2h} \bar{Z} e_2
\]

\[
= \bar{Y} \left[ 1 + e_0 - \frac{m_1 e_1}{2} + \frac{m_1^2 e_1^2}{4} - \frac{m_2 e_2}{2} + \frac{m_1 m_2 e_1 e_2}{4} + \frac{m_1^2 e_1^2}{4} - \frac{m_2 e_0 e_2}{2} - \frac{m_1 e_0 e_1}{2} \right]
\]

\[
- b_{1h} e_1 \bar{X} - b_{2h} e_2 \bar{Z}
\]

(3.2)
Squaring both sides of (3.2) and neglecting the term having power greater than two, we have

\[
(t_p - \bar{Y})^2 = \left[ Y \left( \frac{-m_1e_1}{2} - \frac{m_2e_2}{2} \right) - b_{1h}e_1\bar{X} - b_{2h}e_2\bar{Z} \right]^2
\] (3.3)

Taking expectations of both the sides of (3.3), we have the mean squared error of \( t_p \) up to the first degree of approximation as

\[
\text{MSE}(t_p) = \bar{Y}^2 [V_{200} + P_1] + P_2 - \bar{Y}P_3
\] (3.4)

Where,

\[
P_1 = \frac{m_1^2V_{020} + m_2^2V_{002}}{4} + \frac{m_1m_2V_{011} - m_1V_{110} - m_2V_{101}}{2}
\]

\[
P_2 = B_{1h}^2V_{020} + B_{2h}^2V_{002} + 2B_{1h}B_{2h}V_{011}
\]

\[
P_3 = -2B_{1h}V_{110} - 2B_{2h}V_{101} + m_1B_{1h}V_{020} + m_1B_{2h}V_{011} + m_2B_{1h}V_{011} + m_2B_{2h}V_{002}
\] (3.5)

\[
W_h^2f_h\rho_{ysh}S_{ysh}S_{xh} = \sum_{h=1}^{L} W_h^2f_h\rho_{ysh}S_{ysh}S_{xh}
\]

\[
B_{1h} = \sum_{h=1}^{L} W_h^2f_h\rho_{ysh}S_{ysh}S_{xh}
\]

\[
B_{2h} = \sum_{h=1}^{L} W_h^2f_h\rho_{ysh}S_{ysh}S_{xh}
\]

The optimum values of \( m_1 \) and \( m_2 \) will be

\[
m_1 = \frac{4\left[B_{1h}V_{011}V_{002} + B_{2h}V_{020}^2 - B_{1h}V_{020}V_{002} - B_{2h}V_{011}V_{002}\right]}{Y[V_{020}V_{002} - V_{011}^2]}
\]

\[
m_2 = \frac{4\left[B_{1h}V_{011}V_{020} + B_{2h}V_{011}^2 - B_{1h}V_{011}V_{020} - B_{2h}V_{002}V_{020}\right]}{Y[V_{020}V_{002} - V_{011}^2]}
\] (3.6)

Putting optimum values of \( m_1 \) and \( m_2 \) from (3.6), we obtained min MSE of proposed estimator \( t_p \).

4. Efficiency comparison

In this section, the conditions for which the proposed estimator \( t_p \) is better than \( y_{st}, t_1, t_2, t_3 \),
The variance is given by

\[ V(\bar{Y}_{st}) = \bar{Y}^2 V_{200} \tag{4.1} \]

To compare the efficiency of the proposed estimator with the existing estimator, from (4.1) and (2.3), (2.4), (2.9), (2.10), (2.11), (2.12) and (2.14), we have

\[ V(\bar{Y}_{st}) - \text{MSE}(t_p) = \bar{Y}^2 P_1 - P_2 - \bar{Y}P_3 \geq 0 \tag{4.2} \]

\[ \text{MSE}(t_1) - \text{MSE}(t_p) = \bar{Y}^2 \left[ V_{020} - 2V_{110} \right] - \bar{Y}^2 P_1 - P_2 + \bar{Y}P_3 \geq 0 \tag{4.3} \]

\[ \text{MSE}(t_2) - \text{MSE}(t_p) = \bar{Y}^2 \left[ \frac{V_{020}}{4} - V_{110} \right] - \bar{Y}^2 P_1 - P_2 + \bar{Y}P_3 \geq 0 \tag{4.4} \]

\[ \text{MSE}(t_3) - \text{MSE}(t_p) = \bar{Y}^2 \left[ \frac{V_{020}}{4} + \frac{V_{002}}{4} - V_{110} - V_{101} + \frac{V_{011}}{2} \right] - \bar{Y}^2 P_1 - P_2 + \bar{Y}P_3 \geq 0 \tag{4.5} \]

\[ \text{MSE}(t_4) - \text{MSE}(t_p) = \bar{Y}^2 \left[ \frac{V_{020}}{4} + \frac{V_{002}}{4} + V_{110} + V_{101} + \frac{V_{011}}{2} \right] - \bar{Y}^2 P_1 - P_2 + \bar{Y}P_3 \geq 0 \tag{4.6} \]

\[ \text{MSE}(t_5) - \text{MSE}(t_p) = \bar{Y}^2 \left[ \frac{V_{020}}{4} + \frac{V_{002}}{4} - V_{110} + V_{101} + \frac{V_{011}}{2} \right] - \bar{Y}^2 P_1 - P_2 + \bar{Y}P_3 \geq 0 \tag{4.7} \]

\[ \text{MSE}(t_6) - \text{MSE}(t_p) = \bar{Y}^2 \left[ \frac{V_{020}}{4} + \frac{V_{002}}{4} + V_{110} - V_{101} + \frac{V_{011}}{2} \right] - \bar{Y}^2 P_1 - P_2 + \bar{Y}P_3 \geq 0 \tag{4.8} \]

Using (4.2) - (4.8), we conclude that the proposed estimator outperforms than the estimators considered in literature.

**5. Empirical study**

In this section, we use the data set in Koyuncu and Kadilar [4]. The population statistics are given in Table 3.2.1. In this data set, the study variable (Y) is the number of teachers, the first
auxiliary variable (X) is the number of students, and the second auxiliary variable (Z) is the number of classes in both primary and secondary schools.

**Table 5.1: Data statistics**

|   | $N_i$ | $n_i$ | $S_{yi}$     | $\bar{Y}_i$ | $S_{xi}$     | $S_{yx_i}$ | $\rho_{yx_i}$ | $\beta_2(x_i)$ |
|---|-------|-------|--------------|-------------|--------------|-------------|---------------|----------------|
| 1 | 127   | 31    | 883.835      | 703.74      | 30486.751    | 25237153.52 | 0.936         | 4.593          |
| 2 | 117   | 21    | 644          | 413         | 15180.760    | 9747942.85  | 0.996         | 18.543         |
| 3 | 103   | 29    | 1033.467     | 573.17      | 23094.141    | 28294397.04 | 0.983         | 15.446         |
| 4 | 170   | 38    | 644          | 267.03      | 23094.141    | 15864573.97 | 0.989         | 21.947         |
| 5 | 205   | 22    | 711.723      | 393.84      | 23094.141    | 15864573.97 | 0.965         | 23.114         |
| 6 | 201   | 39    | 74.703       | 306.1518    | 12997.59     | 28294397.04 | 0.994         |                |
$\beta_2(y_1) = 2.158$  \hspace{1cm} $\beta_2(y_2) = 16.392$  \hspace{1cm} $\beta_2(y_3) = 14.979$

$\beta_2(y_4) = 12.167$  \hspace{1cm} $\beta_2(y_5) = 21.008$  \hspace{1cm} $\beta_2(y_6) = 20.254$

$S_{z1} = 555.5816$  \hspace{1cm} $S_{z2} = 365.4576$  \hspace{1cm} $S_{z3} = 612.9509$

$S_{z4} = 458.0282$  \hspace{1cm} $S_{z5} = 260.8511$  \hspace{1cm} $S_{z6} = 397.0481$

$\overline{Z}_1 = 498.28$  \hspace{1cm} $\overline{Z}_2 = 318.33$  \hspace{1cm} $\overline{Z}_3 = 431.36$

$\overline{Z}_4 = 498.28$  \hspace{1cm} $\overline{Z}_5 = 227.20$  \hspace{1cm} $\overline{Z}_6 = 313.71$

$S_{yz1} = 480688.2$  \hspace{1cm} $S_{yz2} = 230092.8$  \hspace{1cm} $S_{yz3} = 623019.3$

$S_{yz4} = 36493.4$  \hspace{1cm} $S_{yz5} = 101539$  \hspace{1cm} $S_{yz6} = 277696.1$

$S_{xz1} = 15914648$  \hspace{1cm} $S_{xz2} = 5379190$  \hspace{1cm} $S_{xz3} = 16490067456$

$S_{xz4} = 8041254$  \hspace{1cm} $S_{xz5} = 214457$  \hspace{1cm} $S_{xz6} = 8857729$

$\rho_{yz1} = 0.978$  \hspace{1cm} $\rho_{yz2} = 0.976$  \hspace{1cm} $\rho_{yz3} = 0.983$

$\rho_{yz4} = 0.982$  \hspace{1cm} $\rho_{yz5} = 0.964$  \hspace{1cm} $\rho_{yz6} = 0.982$

$\beta_2(z_1) = 2.314$  \hspace{1cm} $\beta_2(z_2) = 11.190$  \hspace{1cm} $\beta_2(z_3) = 10.786$

$\beta_2(z_4) = 8.624$  \hspace{1cm} $\beta_2(z_5) = 9.720$  \hspace{1cm} $\beta_2(z_6) = 14.406$

We have computed the pre relative efficiency (PRE) of different estimators of $\overline{Y}_{st}$ with respect to $\overline{y}_{st}$ and complied in table 5.2:
Table 5.2: Percent Relative Efficiencies (PRE) of estimator

| S. No. | Estimators | PRE’S   |
|--------|------------|---------|
| 1      | $\bar{y}_{st}$ | 100     |
| 2      | $t_1$      | 1029.46 |
| 3      | $t_2$      | 370.17  |
| 4      | $t_3$      | 2045.43 |
| 5      | $t_4$      | 27.94   |
| 6      | $t_5$      | 126.41  |
| 7      | $t_6$      | 77.21   |
| 8      | $t_7$      | 2360.54 |
| 9      | $t_p$      | 4656.35 |
6. Conclusion

In this paper, we proposed a new estimator for estimating unknown population mean of study variable using information on two auxiliary variables. Expressions for bias and MSE of the estimator are derived up to first degree of approximation. The proposed estimator is compared with usual mean estimator and other considered estimators. A numerical study is carried out to support the theoretical results. In the table 5.2, the proposed estimator performs better than the usual sample mean and other considered estimators.

References

[1] Bahl, S. and Tuteja, R.K. (1991): Ratio and product type exponential estimator. Infrm. and Optim. Sci., XII, I, 159-163.

[2] Diana, G. (1993). A class of estimators of the population mean in stratified random sampling. Statistica 53(1):59–66.

[3] Kadilar, C. and Cingi, H. (2003): Ratio Estimators in Straitified Random Sampling. Biometrical Journal 45 (2003) 2, 218-225.

[4] Koyuncu, N. and Kadilar, C. (2009) : Family of estimators of population mean using two auxiliary variables in stratified random sampling. Comm. In Stat. – Theory and Meth., 38:14, 2398-2417.

[5] Naik, V.D and Gupta, P.C., 1996: A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agri. Stat., 48(2), 151-158.

[6] Perri, P. F. (2007). Improved ratio-cum-product type estimators. Statist. Trans. 8:51–69.

[7] Singh, R., Kumar, M., Singh,R.D.,and Chaudhary, M.K.(2008): Exponential Ratio Type Estimators in Stratified Random Sampling. Presented in International Symposium On Optimisation and Statistics (I.S.O.S).-2008 held at A.M.U., Aligarh, India, during 29-31 Dec 2008.
[8] Singh, H., P. and Vishwakarma, G. K. (2008): A family of estimators of population mean using auxiliary information in stratified sampling. Communication in Statistics Theory and Methods, 37(7), 1038-1050.

[9] Singh, M. P. (1965): On the estimation of ratio and product of population parameter. Sankhya 27 B, 321-328.

[10] Singh, M.P. (1967): Ratio-cum-product method of estimation. Metrika 12, 34-72.

[11] Singh, R., Chauhan, P., Sawan, N. and Smarandache,F. (2007): Auxiliary information and a priory values in construction of improved estimators. Renaissance High press.