Improved Estimators of Population Mean Using Two Auxiliary Variables in Stratified Random Sampling

Rajesh Singh
Department of Statistics
Banaras Hindu University (U.P.), India
rsinghstat@yahoo.com

Mukesh Kumar
Department of Statistics
Banaras Hindu University (U.P.), India
mukesh.stat@gmail.com

Abstract

An exponential family of estimators, which use the information of two auxiliary variables in the stratified sampling, is proposed to estimate the population mean of the variable under study. The mean-squared error of the suggested family of estimators are derived under large sample approximation. The family of estimators in its optimum case is carried out to show the properties of the proposed estimators.

Keywords: Auxiliary information, Study variate, Stratified random sampling, Mean square error, Exponential estimator.

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 (1993) suggested a class of estimators of the population mean using one auxiliary variable in the stratified random sampling and examined the mean square Error (MSE) of the estimators up to the k^{th} order of approximation. Kadilar and Cingi (2003), Singh, H.P. et. al. (2008), Singh and Vishwakarma (2008), Singh, R.et al. (2008), Koyuncu and Kadilar (2009) proposed estimators in stratified random sampling. Singh (1967) and Perri (2007) suggested some ratio cum product estimators in simple random sampling. Bahl and Tuteja (1991) and Singh, R. et. al. (2008, 2009) suggested some exponential ratio type estimators. Ghosh (1958) and Rao (1977) have suggested estimators in stratified random sampling with multiple characteristics. In this study, we suggest some exponential-type estimators using the auxiliary information in the stratified random sampling when two auxiliary variables are available.

Consider a finite population U= $(U_1,U_2,U_3 \ldots U_N)$ of size N and let y, x and z, respectively, be the study and two auxiliary variables associated with each Uj (j=1,2,3..N) of the population. Let the population of size, N, is stratified into L strata with h^{th} stratum containing N_h units, where h=1,2,3,…L such that $\sum_{h=1}^{L} N_h$ = N.

A simple random sample of size n_h is drawn without replacement from the h_{th} stratum such that $\sum_{h=1}^{L} n_h = n$. Let (y_{hi}, 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

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

$\bar{z}_{st} = \sum_{h=1}^{L} w_h \bar{z}_h = \bar{Z}(1 + e_2)$.

Such that,

$E(e_0) = E(e_1) = E(e_2) = 0$.

$$V_{rst} = \sum_{h=1}^{L} w_h^{r+s+t} \frac{E\left[(\bar{y}_h - \bar{Y}_h)^r (\bar{x}_h - \bar{X}_h)^s (\bar{z}_h - \bar{Z}_h)^t\right]}{\bar{Y}^r \bar{X}^s \bar{Z}^t} \qquad (1.1)$$

where,

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

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

and

$$V(\bar{y}_{st}) = \bar{Y}^2 V_{200} \qquad (1.2)$$

Similar expressions for X and Z can also be defined.

Using (1.1), we can write

$$E(e_0^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{yh}^2}{\bar{Y}^2} = V_{200}, \qquad E(e_1^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xh}^2}{\bar{X}^2} = V_{020},$$

$$E(e_2^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{zh}^2}{\bar{Z}^2} = V_{002}, \qquad E(e_0 e_1) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xyh}}{\bar{X}\bar{Y}} = V_{110},$$

$$E(e_1 e_2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xzh}}{\bar{X}\bar{Z}} = V_{011}, \qquad E(e_0 e_2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xzh}}{\bar{X}\bar{Z}} = V_{101},$$

where

$$S_{yh}^2 = \frac{\sum_{i=1}^{N_h} (\bar{y}_h - \bar{Y}_h)^2}{N_h - 1}, \qquad S_{xh}^2 = \frac{\sum_{i=1}^{N_h} (\bar{x}_h - \bar{X}_h)^2}{N_h - 1},$$

$$S_{zh}^2 = \frac{\sum_{i=1}^{N_h} (\bar{z}_h - \bar{Z}_h)^2}{N_h - 1}, \qquad S_{xyh} = \frac{\sum_{i=1}^{N_h} (\bar{x}_h - \bar{X}_h)(\bar{y}_h - \bar{Y}_h)}{N_h - 1},$$

$$S_{yzh} = \frac{\sum_{i=1}^{N_h} (\bar{y}_h - \bar{Y}_h)(\bar{z}_h - \bar{Z}_h)}{N_h - 1} \; S_{xzh} = \frac{\sum_{i=1}^{N_h} (\bar{x}_h - \bar{X}_h)(\bar{z}_h - \bar{Z}_h)}{N_h - 1}$$

$$\gamma_h = \frac{1 - f_h}{n_h}.$$

2. Adapted Estimators

When the information on the two auxiliary variables is known, Singh (1965, 1967) proposed some ratio cum product estimators in simple random sampling to estimate the population mean of the study variable y.

Motivated by Singh (1965, 1967) and Singh, R. et. al. (2009), we propose some estimators in stratified sampling as

$$t_1 = \bar{y}_{st}\exp\left[\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right]\exp\left[\frac{\bar{Z}-\bar{z}_{st}}{\bar{X}+\bar{z}_{st}}\right] \tag{2.1}$$

$$t_2 = \bar{y}_{st}\exp\left[\frac{\bar{x}_{st}-\bar{X}}{\bar{x}_{st}+\bar{X}}\right]\exp\left[\frac{\bar{z}_{st}-\bar{Z}}{\bar{z}_{st}+\bar{Z}}\right] \tag{2.2}$$

$$t_3 = \bar{y}_{st}\exp\left[\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right]\exp\left[\frac{\bar{z}_{st}-\bar{Z}}{\bar{z}_{st}+\bar{Z}}\right] \tag{2.3}$$

$$t_4 = \bar{y}_{st}\exp\left[\frac{\bar{x}_{st}-\bar{X}}{\bar{x}_{st}+\bar{X}}\right]\exp\left[\frac{\bar{Z}-\bar{z}_{st}}{\bar{X}+\bar{z}_{st}}\right] \tag{2.4}$$

The MSE equations of these estimators can be written as

$$MSE(t_1) = \bar{Y}^2\left(V_{200}+\frac{V_{020}}{4}+\frac{V_{002}}{4}-V_{110}-V_{101}+\frac{V_{011}}{2}\right) \tag{2.5}$$

$$MSE(t_2) = \bar{Y}^2\left(V_{200}+\frac{V_{020}}{4}+\frac{V_{002}}{4}+V_{110}+V_{101}+\frac{V_{011}}{2}\right) \tag{2.6}$$

$$MSE(t_3) = \bar{Y}^2\left(V_{200}+\frac{V_{020}}{4}+\frac{V_{002}}{4}-V_{110}+V_{101}-\frac{V_{011}}{2}\right) \tag{2.7}$$

$$MSE(t_4) = \bar{Y}^2\left(V_{200}+\frac{V_{020}}{4}+\frac{V_{002}}{4}+V_{110}-V_{101}-\frac{V_{011}}{2}\right) \tag{2.8}$$

3. Suggested Estimators

$$t_i^* = \bar{y}_{st}\exp\left[\frac{(A_{st}^*+b_{st})-(a_{st}^*+b_{st})}{(A_{st}^*+b_{st})+(a_{st}^*+b_{st})}\right]^{\alpha_1}\exp\left[\frac{(c_{st}^*+d_{st})-(C_{st}^*+d_{st})}{(c_{st}^*+d_{st})+(C_{st}^*+d_{st})}\right]^{\alpha_2} \quad (i = 1,2,3,4) \tag{3.1}$$

Let $A_{st}^* = \sum_{h=1}^{L} w_h \bar{X}_h A_h$ and $a_{st}^* = \sum_{h=1}^{L} w_h \bar{x}_h A_h$,
where A_h may be some population information of the first auxiliary variable for h^{th} stratum similarly, we define
$C_{st}^* = \sum_{h=1}^{L} w_h \bar{Z}_h B_h$ and $c_{st}^* = \sum_{h=1}^{L} w_h \bar{z}_h B_h$
where B_h may be some population information of the second auxiliary variable for h^{th} stratum.

Expressing t_i^* in terms of e's, we have

$$t_i^* = \bar{Y}(1+e_0)\exp\left[\frac{-\theta_1}{2}e_1^*\left(1+\frac{\theta_1}{2}e_1^*\right)^{-1}\right]^{\alpha_1}\exp\left[\frac{\theta_2}{2}e_2^*\left(1+\frac{\theta_2}{2}e_2^*\right)^{-1}\right]^{\alpha_2} \tag{3.2}$$

where,

$$\theta_1 = \frac{A_{st}^{*}}{A_{st}^{*} + b_{st}}, \quad \theta_2 = \frac{C_{st}^{*}}{C_{st}^{*} + d_{st}}$$

$$e_1^{*} = \frac{a_{st}^{*} - A_{st}^{*}}{A_{st}^{*}} = \frac{\sum_{h=1}^{L} w_h A_h(\bar{x}_h - \bar{X}_h)}{A_{st}^{*}}$$

and

$$e_2^{*} = \frac{c_{st}^{*} - C_{st}^{*}}{C_{st}^{*}} = \frac{\sum_{h=1}^{L} w_h B_h(\bar{x}_h - \bar{X}_h)}{C_{st}^{*}}$$

It is obvious that $E(e_1^{*}) = E(e_2^{*}) = 0$.

Expending the right hand side of (2.10) and retaining the terms up to first order of approximation, we get

$$(t_i^{*} - \bar{Y}) \cong \bar{Y}[e_0 - \frac{1}{2}(\alpha_1\theta_1 e_1^{*} - \alpha_2\theta_2 e_2^{*}) + \frac{1}{4}(\alpha_1\theta_1^2 e_1^{*2} - \alpha_2\theta_2^2 e_2^{*2})$$

$$+\alpha_1(\alpha_1 - 1)\frac{\theta_1^2 e_1^{*2}}{4} + \alpha_2(\alpha_2 - 1)\frac{\theta_2^{*2} e_2^{*2}}{4} \qquad (3.3)$$

Squaring (3.3) to the first order of approximation and then taking expectation, we have -

$$MSE(t_i^{*}) \cong \bar{Y}^2[V_{200}^{*} + \frac{1}{4}\{\alpha_1^2\theta_1^{*2}V_{020}^{*} + \alpha_2^2\theta_2^{*2}V_{002}^{*}\}$$

$$-\alpha_1\theta_1^{*}V_{110}^{*} + \alpha_2\theta_2^{*}V_{101}^{*} - \frac{1}{2}\alpha_1\alpha_2\theta_1^{*}\theta_2^{*}V_{011}^{*}] \qquad (3.4)$$

where,

$$\theta_1^{*} = \frac{\bar{X}_{st}}{A_{st}^{*} + b_{st}}, \qquad \theta_2^{*} = \frac{\bar{Z}_{st}}{C_{st}^{*} + d_{st}},$$

$$V_{rst}^{*} = \sum_{i=1}^{L} w^{r+s+t}(A_h)^s(B_h)^t \frac{E[(\bar{y}_h - \bar{Y}_h)^r(\bar{x}_h - \bar{X}_h)^s(\bar{z}_h - \bar{Z}_h)^t]}{\bar{Y}^r\bar{X}^s\bar{Z}^t}.$$

On differentiating (3.4) with respect to α_1 and α_2, respectively we get the optimum values of α_1 and α_2 as:

$$\alpha_1^{*} = \frac{2(V_{002}^{*}V_{110}^{*} - V_{101}^{*}V_{011}^{*})}{\theta_1^{*}\left(V_{020}^{*}V_{002}^{*} - V_{011}^{*\,2}\right)} \qquad (3.5)$$

and

$$\alpha_2^{*} = \frac{2(V_{011}^{*}V_{110}^{*} - V_{101}^{*}V_{020}^{*})}{\theta_2^{*}\left(V_{020}^{*}V_{002}^{*} - V_{011}^{*\,2}\right)} \qquad (3.6)$$

Putting optimum values of α_1^* and α_2^* from (3.5) and (3.6) in (3.4) we obtain min. MSE of t_i^* as-

$$\mathrm{MSE}(t_i^*)\min = \bar{Y}^2\left(V_{200}^* - \frac{V_{002}^* V_{110}^{*2} + V_{020}^* V_{110}^{*2} - 2V_{110}V_{101}V_{011}}{V_{020}^* V_{002}^* - V_{011}^{*2}} \right) \quad (3.7)$$

4. Efficiency comparison

In this section, firstly, we compare the MSE of proposed estimator t_i^*(i=1,2,3,4) with V ($\bar{y}_{st}$).

$$\mathrm{MSE}(t_i^*) < V(\bar{y}_{st})$$

$$\bar{Y}^2\left| V_{200}^* + \frac{1}{4}\{\alpha_1^2\theta_1^{*2}V_{020}^* + \alpha_2^2\theta_2^{*2}V_{002}^*\} - \alpha_1\theta_1^*V_{110}^* + \alpha_2\theta_2^*V_{101}^* - \frac{1}{2}\alpha_1\alpha_2\theta_1^*\theta_2^*V_{011}^* \right| < \bar{Y}^2 V_{200}^*$$

Estimator t_i^* (i=1,2,3,4) performs better than $\bar{y}_{st}$, whenever

$$\frac{2A}{B} > 1 \quad (4.1)$$

where, $A = \alpha_1\alpha_2\theta_1^*\theta_2^*V_{011}^* + 2\alpha_1\theta_1^*V_{110}^* - 2\alpha_2\theta_2^*V_{101}^*$

$B = \alpha_1^2\theta_1^{*2}V_{020}^* + \alpha_2^2\theta_2^{*2}V_{002}^*$

Secondly, we find the condition under which $\mathrm{MSE}(t_i^*)\min$ performs better than V (y_{st})

$\mathrm{MSE}(t_i^*)\min < V(\bar{y}_{st})$

$$\bar{Y}^2\left(V_{200}^* - \frac{V_{002}^* V_{110}^{*2} + V_{020}^* V_{110}^{*2} - 2V_{110}V_{101}V_{011}}{V_{020}^* V_{002}^* - V_{011}^{*2}} \right) < \bar{Y}^2 V_{200}^*$$

$\mathrm{MSE}(t_i^*)\min$ is less than V (y_{st}), whenever

$$\frac{C}{D} > 0 \quad (4.2)$$

where,

$C = V_{002}^* V_{110}^{*2} + V_{020}^* V_{110}^{*2} - 2V_{110}V_{101}V_{011}$

and $D = V_{020}^* V_{002}^* - V_{011}^{*2}$.

5. Empirical study

For empirical study we use the data set which is earlier used by Koyuncu and Kadilar (2009).

Data Statistics

N_1=127	N_2=117	N_3=103
N_4=170	N_5=205	N_6=201
n_1=31	n_2=21	n_3=29
n_4=38	n_5=22	n_6=39
S_{y1}=883.835	S_{y2}=644	S_{y3}= 1033.467
S_{y4} = 810.585	S_{y5}=403.654	S_{y6}=711.723
$\overline{Y}_1$=703.74	$\overline{Y}_2$= 413	$\overline{Y}_3$ = 573.17
$\overline{Y}_4$ = 424.66	$\overline{Y}_5$= 267.03	$\overline{Y}_6$ = 393.84
S_{x1}=30486.751	S_{x2}=15180.760	S_{x3}=27549.697
S_{x4}=18218.931	S_{x5}=8997.776	S_{x6}=23094.141
$\overline{X}_1$=20804.59	$\overline{X}_2$=9211.79	$\overline{X}_3$=14309.30
$\overline{X}_4$=9478.85	$\overline{X}_5$= 5569.95	$\overline{X}_6$=12997.59
S_{xy1}=25237153.52	S_{xy2}=9747942.85	S_{xy3}=28294397.04
S_{xy1}= 14523885.53	S_{xy1}=3393591.75	S_{xy6}=15864573.97
ρ_{xy1}= 0.936	ρ_{xy2}= 0.996	ρ_{xy3}=0.994
ρ_{xy4}= 0.983	ρ_{xy5}=0.989	ρ_{xy6}= 0.965
$\beta_2(x_1)$= 4.593	$\beta_2(x_2)$= 18.543	$\beta_2(x_3)$ =15.446
$\beta_2(x_4)$ =10.162	$\beta_2(x_5)$ = 21.947	$\beta_2(x_6)$ =23.114
$\beta_2(y_1)$=2.158	$\beta_2(y_2)$= 16.392	$\beta_2(y_3)$=14.979
$\beta_2(y_4)$= 12.167	$\beta_2(y_5)$=21.088	$\beta_2(y_6)$= 20.254
S_{z1}= 555.5816	S_{z2} = 365.4576	S_{z3}=612.9509281
S_{z4}= 458.0282	S_{z5} = 260.8511	S_{z6}= 397.0481
$\bar{Z}_1$= 498.28	$\bar{Z}_2$= 318.33	$\bar{Z}_3$ = 431.36
$\bar{Z}_4$= 498.28	$\bar{Z}_5$ = 227.20	$\bar{Z}_6$= 313.71
S_{yz1}= 480688.2	S_{yz2}= 230092.8	S_{yz1}= 623019.3
S_{yz1}= 364943.4	S_{yz1}= 101539	S_{yz1}= 277696.1
S_{xz1}= 15914648	S_{xz2}= 5379190	S_{xz3}= 164900674.56
S_{xz4}= 8041254	S_{xz5}= 2144057	S_{xz1}= 8857729
ρ_{yz1}= 0.978914	ρ_{yz2} = 0.9762	ρ_{y3}= 0.983511
ρ_{yz4}= 0.982958	ρ_{yz5}= 0.964342	ρ_{yz1}= 0.982689
$\beta_2(z_1)$=2.314926	$\beta_2(z_2)$ = 11.19093	$\beta_2(z_3)$= 10.78635
$\beta_2(z_4)$= 8.624111	$\beta_2(z_5)$ = 9.720886	$\beta_2(z_6)$= 14.40696

Table1: Percent Relative Efficiencies (PRE) of estimators

Estimator	α_1	α_2	A_h	B_h	PRE
$\bar{y}_{st}$	0	0	1	1	100
t_1	1	-1	1	1	2045.439
t_2	-1	1	1	1	27.948
t_3	1	1	1	1	126.419
t_4	-1	-1	1	1	77.219
t_1^*	0.0215	0.140416	$\beta_2(x)$	$\beta_2(z)$	101.155
t_2^*	0.6893	-0.9946	C_{xh}	C_{zh}	256.181
t_3^*	0.5835	-1.7739	1	1	2854.597
t_4^*	0.7533	-1.5332	ρ_{xy}	ρ_{yz}	7251.178

Conclusion

From Table 1, we observe that the estimator t_4^* which uses correlation coefficient as A_h and B_h has the smallest MSE value among their own family of estimators. When we further examine Table 1, we see that all exponential ratio-type estimators performs better than usual mean estimator $\bar{y}_{st}$.

References

1. Bahl. S. and Tuteja, R.K. (1991): Ratio and Product Type Exponential Estimator. Infrm. and Optim. Sci., XIII, 159-163.
2. Diana, G. (1993): A class of estimators of the population mean in stratified random sampling. Statistica 53(1), 59–66.
3. Ghosh, S.P. (1958). A Note on stratified random sampling with multiple characteristics. Calc. Stat. Assoc. Bull., 8, (30-31), 81-90.
4. Kadilar, C., Cingi, H. (2003): Ratio estimators in stratified random sampling. Biometr. J.45, 218–225.
5. Koyuncu, N. and Kadilar, C. (2009): Family of Estimators of Population Mean Using Two Auxiliary Variables in Stratified Random Sampling. Commun.in Statist. Theo. and Meth., 38: 2398–2417.
6. Perri, P. F. (2007): Improved ratio-cum-product type estimators. Statist. Trans. 8:51–69.
7. Rao, T.J. (1977). Optimum allocation of sample size and prior distributions: A review. Intern. Stat. Rev., 45, 173-179.
8. Singh, H. P., Vishwakarma, G. K. (2008): A family of estimators of population mean using auxiliary information in stratified sampling. Commun. Statist. Theo. Meth.37, 1038–1050.

9. Singh, H. P., Tailor, R., Singh, S., Kim, J. M. (2008): A modified estimator of population mean using power transformation. Statist.Pap.49:37–58.
10. Singh, M. P. (1965): On the estimation of ratio and product of the population parameters. Sankhya *B* 27:231-328.
11. Singh, M. P. (1967). Ratio cum product method of estimation. Metrika 12: 34–42.
12. Singh, R., Cauhan, P., Sawan, N. and Smarandache, F. (2007): Auxiliary information and a priori values in construction of improved estimators. Renaissance High press.
13. Singh, R., Kumar, M., Singh, R.D., and Chaudhary, M.K. (2008): Exponential Ratio Type Estimators in Stratified Random Sampling. (Proceedings of International Symposium on Optimisation and Statistics (I.S.O.S.), 2008 held at A.M.U., Aligarh, India, during 29-31 Dec 2008).
14. Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2009): Improvement in estimating the population mean using exponential estimator in simple random sampling. Bull. of Stat. and Econ., 3, 13-18.