World Applied Sciences Journal 28 (9): 1193-1198, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.28.09.2778

Dual to Ratio cum Product Estimators of Finite Population Mean Using Auxiliary Attribute(s) in Stratified Random Sampling

Sachin Malik and Rajesh Singh

Department of Statistics, Banaras Hindu University, Varanasi-221005, India

Abstract: In this paper some new improved estimators are suggested using auxiliary attribute(s) in stratified random sampling. The expressions of the mean square error of the proposed estimators are derived. Also, the theoretical findings are supported by a numerical example.

Key words: Auxiliary information · Attributes · Stratified sampling · Bias · Mean squared error

INTRODUCTION

The problem of estimating the population mean in the presence of an auxiliary variable has been widely discussed in finite population sampling literature [1], 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 [2-8] suggested some ratio cum product estimators in simple random sampling [10, 11] suggested some exponential ratio type estimators. However in many practical situations instead of existence of auxiliary variables there exist some auxiliary attributes ϕ (say), which are highly correlated with the study variable y. Taking into consideration the point biserial correlation coefficient between auxiliary attribute ϕ and the study variable y, several authors including [12-17] envisaged large number of improved estimators for the population mean $\bar{Y}$ of the study variable y. There are some situations when in place of one auxiliary attribute, we have information on two qualitative variables. For illustration, to estimate the hourly wages we can use the information on marital status and region of residence [18]. Here we assume that both auxiliary attributes have significant point bi-serial correlation with the study variable and there is significant phi-correlation [19] between the two auxiliary attributes. Using point biserial correlation and phi-correlation [20-24] have proposed improved estimators for the population mean $\bar{Y}$ of the study variable y.

In planning surveys, stratified sampling has often proved as useful in improving the precision of un-stratified sampling strategies to estimate the finite population mean of the study variable, $\bar{Y}=\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}$.

Assume that the population U consist of L strata as $U=U_1, U_2,\ldots,U_L$. Here the size of the stratum U_h is N_h and the size of simple random sample in stratum U_h is n_h, where h=1, 2,…,L. In this study, under stratified random sampling without replacement scheme, we suggest estimators to estimate $\bar{Y}$ by considering the estimators in [25] and in [26]. 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. Let (y_{hi}, Ψ_{jhi}) (j=1,2) denote the values of the study variable (y) and the auxiliary attributes Ψ_j, (j=1,2) respectively in the h^{th} stratum.

Suppose

$\Psi_{jhi}=1$, if i^{th} unit in the stratum h possesses the attribute Ψ_j
=0, otherwise.

Corresponding Author: Rajesh Singh, Department of Statistics, Banaras Hindu University, Varanasi-221005, India.

Let $P_j = \sum_{h=1}^{L} W_h P_{jh}$ and $p_{jst} = \sum_{h=1}^{L} W_h p_{jh}$ (j=1, 2) are the population and sample proportions of units of the auxiliary attributes where $P_{jh} = \frac{A_{jh}}{N_h}$, $p_{jh} = \frac{a_{jh}}{n_h}$ and $A_{jh} = \sum_{i=1}^{N_h} \psi_{jhi}$ and $a_{jh} = \sum_{i=1}^{n_h} \psi_{jhi}$.

To estimate $\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h$, we assume that P_{jh} (j=1, 2) are known.

To obtain the bias and MSE of the proposed estimators, we use the following notations in the rest of the article:

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h = \bar{Y}(1+e_0)$$

$$p_{1st} = \sum_{h=1}^{L} W_h p_{1h} = P_1(1+e_1)$$

$$p_{2st} = \sum_{h=1}^{L} W_h p_{2h} = P_2(1+e_2)$$

such that,

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

Let

$$S_{yh}^2 = \sum_{i=1}^{N_h} \frac{\left(y_{hi} - \bar{Y}_h\right)^2}{N_h - 1}, S_{\psi_{jh}}^2 = \sum_{i=1}^{N_h} \frac{\left(\psi_{jhi} - P_{jh}\right)^2}{N_h - 1}$$

be the population variances of the study variable and the auxiliary attributes in the hth stratum, where,

$$\bar{Y}_h = \sum_{i=1}^{N_h} \frac{y_{hi}}{N_h},\ S_{y\psi_{jh}} = \sum \frac{\left(y_{hi} - \bar{Y}_h\right)\left(\psi_{jhi} - P_{jh}\right)}{N_h - 1} \quad \textit{and}$$

$$\rho_{y\psi_{jh}} = \frac{S_{y\psi_{jh}}}{S_{\psi_{jh}} S_{yh}}$$

be the population bi-covariance and point bi-serial correlation between the study variable and the auxiliary attributes respectively in the h^{th} stratum. Also $S_{\psi_1\psi_2 h} = \sum_{i=1}^{N_h} \frac{(\psi_{1hi} - P_{1h})(\psi_{2hi} - P_{2h})}{N_h - 1}$ and $\rho_{\psi_1\psi_2 h} = \frac{S_{\psi_1\psi_2 h}}{S_{\psi_{1h}} S_{\psi_{2h}}}$ be the population bi-covariance and phi-correlation coefficient between the auxiliary attributes in the h^{th} stratum respectively.

$$E\left(e_0^2\right) = \frac{\sum_{h=1}^{L} W_h^2 f_h S_{yh}^2}{\bar{Y}^2} = V_1,\ E\left(e_1^2\right) = \frac{\sum_{h=1}^{L} W_h^2 f_h S_{\psi_1 h}^2}{P_1^2} = V_2,$$

$$E\left(e_2^2\right) = \frac{\sum_{h=1}^{L} W_h^2 f_h S_{\psi_2 h}^2}{P_2^2} = V_3,\ E\left(e_0 e_1\right) = \frac{\sum_{h=1}^{L} W_h^2 f_h S_{y\psi_1 h}^2}{\bar{Y} P_1} = V_4,$$

$$E(e_0 e_2) = \frac{\sum_{h=1}^{L} W_h^2 f_h S_{y\psi_2 h}^2}{\bar{Y} P_2} = V_5, \text{and} E(e_1 e_2) = \frac{\sum_{h=1}^{L} W_h^2 f_h S_{\psi_1\psi_2 h}^2}{P_1 P_2} = V_6$$

The combined ratio and the combined product estimators are, respectively, defined as

$$t_1 = \bar{y}_{st}\left(\frac{P_1}{p_{1st}}\right) \tag{1.1}$$

$$t_2 = \bar{y}_{st}\left(\frac{P_2}{p_{2st}}\right) \tag{1.2}$$

$$t_3 = \bar{y}_{st} \exp\left(\frac{P_1 - p_{1st}}{P_1 + p_{1st}}\right) \tag{1.3}$$

$$t_4 = \bar{y}_{st} \exp\left(\frac{p_{2st} - P_2}{p_{2st} + P_2}\right) \tag{1.4}$$

and the MSE of t_1 and t_2 to the first degree of approximation are, respectively, given by

$$MSE(t_1) = \bar{Y}^2 \left[V_1 + V_2 - 2V_4\right] \tag{1.5}$$

$$MSE(t_2) = \overline{Y}^2 [V_1 + V_2 + 2V_4] \quad (1.6)$$

$$MSE(t_3) = \overline{Y}^2 \left[V_1 + \frac{V_2}{4} - V_4\right] \quad (1.7)$$

$$MSE(t_4) = \overline{Y}^2 \left[V_1 + \frac{V_2}{4} + V_4\right] \quad (1.8)$$

Following [27], we considered a simple transformation as

$$u_i = A - p_i; \quad (i = 1,2,...,N)$$

$$\overline{U} = A - P_1$$

where, A is a scalar to be chosen. $\bar{u}_{st} = A - p_{1st}$, Using this transformation, the ratio estimator in the stratified random sampling is defined as

$$t_5 = \bar{y}_{st}\left(\frac{\bar{u}_{st}}{\overline{U}}\right) \quad (1.9)$$

This is dual to combined ratio estimator in stratified random sampling.

The exact expression for MSE of t_3 is given by

$$MSE(t_5) = \overline{Y}^2 \left[V_1 + \lambda^2 V_2 - \lambda 2 V_4\right] \quad (1.10)$$

$$\text{where,} \lambda = \frac{p_{1st}}{A - p_{1st}}$$

In some survey situations, information on a second auxiliary variable, P_2, correlated negatively with the study variable, Y, is readily available. To estimate $\overline{Y}$, [26] considered ratio-cum-product estimator as

$$t_6 = \bar{y}\left(\frac{P_1}{p_1}\right)\left(\frac{P_2}{p_2}\right) \quad (1.11)$$

In that case, following [9] we use $t_{p1} = p_{1st} + \alpha_1(P_1 - p_{1st})$ and $t_{p2} = p_{2st} + \alpha_2(P_2 - p_{2st})$ instead of p_{1st} and p_{2st} respectively. Where, α_1 and α_2 are constants that makes the MSE minimum.

Adapting t_6 to the stratified random sampling, the ratio cum product estimator using two auxiliary attributes can be defined as

$$t_7 = \bar{y}_{st}\left(\frac{P_1}{p_{1st}}\right)\left(\frac{p_{2st}}{P_2}\right) \quad (1.12)$$

The approximate MSE of this estimator is

$$MSE(t_7) = \overline{Y}^2 \left[V_1 + V_2 + V_3 + 2(V_5 - V_4 - V_6)\right] \quad (1.13)$$

Suggested Estimators: Motivated by [26], we propose the following product estimator for the stratified random sampling scheme as

$$t_8 = \bar{y}_{st}\left(\frac{\bar{u}}{\overline{U}}\right)\left(\frac{p_{2st}}{P_2}\right) \quad (2.1)$$

Expressing t_8 in terms of e's, we can write (2.1) as

$$t_8 = \overline{Y}(1 + e_0)(1 - e_1)(1 + e_2)$$

The MSE(t_8) to the first order of approximation, is given as

$$MSE(t_8) = \overline{Y}^2 \left[V_1 + ?^2 V_2 + V_3 - 2(?V_4 - V_5 + \lambda V_6)\right] \quad (2.2)$$

And this MSE equation is minimized for

$$\lambda = \frac{V_4 + V_6}{V_2} \quad (2.3)$$

Note that the corresponding A is

$$A_{opt} = \frac{(1\text{-}\lambda_{opt})P_1}{\lambda_{opt}} \quad (2.4)$$

By putting the optimum value of λ in (12), we can obtain the minimum MSE equation for the first proposed estimator t_8,

Following [7], we proposed a family of dual to ratio cum estimators as

$$t_9 = \bar{y}_{st}\left(\frac{p^*_{1st}}{P_1}\right)^{\theta_1}\left(\frac{P_2}{p^*_{2st}}\right)^{\theta_2} \quad (2.5)$$

where,

$$p^*_{1st} = (1 + g_h P_{1h}) - g_h p_{1h}, \; p^*_{2st} = (1 + g_h P_{2h}) - g_h p_{2h}$$

To obtain the MSE of the estimator t_9, we can write

$$\bar{y}_{st} = \bar{Y}\left(1+e_0\right)$$

$$p^*_{1st} = P_1\left(1+e^*_1\right)$$

$$p^*_{2st} = P_2\left(1+e^*_2\right)$$

Expressing (2.5) in terms of e's, we have

$$t_9 = \bar{Y}\left(1+e_0\right)\left(1+e^*_1\right)^{\theta_1}\left(1+e^*_2\right)^{-\theta_2} \tag{2.6}$$

Expanding the right hand side of (2.6), to the first order of approximation, we get

$$t_9 = \bar{Y}\left[\left[e_0 + \theta_1 e_0 e^*_1 - \theta_2 e_0 e^*_2 + \theta_1 e^*_1 - \theta_2 e^*_2 - \theta_1\theta_2 e^*_1 e^*_2 + \frac{\theta_1\left(\theta_1 - 1\right)}{2} e^{*2}_1 + \frac{\theta_2\left(\theta_2 + 1\right)}{2} e^{*2}_2\right]\right.$$

Squaring both sides of (16) and then taking expectation, we obtain the MSE of the second proposed estimator, t_9 to the first order approximation, as

$$\mathrm{MSE}(t_9) = \bar{Y}^2\left[V^*_1 + \theta_1^2 V^*_2 + \theta_2^2 V^*_3 + 2\left(\theta_1 V^*_4 - \theta_1\theta_2 V^*_6 - \theta_2 V^*_5\right)\right] \tag{2.7}$$

This MSE equation is minimized for the optimum values of θ_1 and θ_2 given by

$$\theta_1 = \frac{V^*_5 V^*_6 - V^*_3 V^*_4}{V^*_2 V^*_3 - V^{*2}_6} \quad \text{and} \quad \theta_2 = \frac{V^*_2 V^*_5 - V^*_4 V^*_6}{V^*_2 V^*_3 - V^{*2}_6}$$

Putting these values of θ_1 and θ_2 in MSE(t_9), given in (2.7), we can obtain the minimum MSE of the second proposed estimator, (t_9).

Theoretical Efficiency Comparisons: In this section, we first compare the efficiency between the first proposed estimator t_8, with and the classical combined estimator, t_9, as follows:

$$\mathrm{MSE}(t_7) < V\left(\bar{y}_{st}\right)$$

$$\bar{Y}^2\left[V_1 + \lambda^2 V_2 + V_3 - 2\left(\lambda V_4 - V_5 + \lambda V_6\right)\right] < \bar{Y}^2 V_1$$

$$\frac{A_1}{2A_2} < 1 \tag{3.1}$$

Where, $A_1 = \lambda^2 V_2 + V_3$ and $A_2 = \lambda V_4 - V_5 + \lambda V_6$

If the condition (3.1) is satisfied, the first proposed estimator t_8, performs better than the classical combined estimator.

We also find the condition under which the second proposed estimator, t_9, performs better than the classical combined estimator in theory as follows:

$$MSE(t_9) < V\left(\bar{y}_{st}\right)$$

$$\bar{Y}^2\left[V_1^* + \theta_1^2 V_2^* + \theta_2^2 V_3^* + 2\left(\theta_1 V_4^* - \theta_1\theta_2 V_6^* - \theta_2 V_5^*\right)\right] < \bar{Y}^2 V_1$$

$$\frac{A_3}{A_4} < 1 \tag{3.2}$$

where, $A_3 = \theta_1^2 V_2^* + \theta_2^2 V_3^* - 2\theta_1\theta_2 V_6^*$ and $A_4 = ?_2 V_5^* - ?_1 V_4^*$

Empirical Study: Source: [See (28)]

We randomly select a sample of size n_h from each stratum by using the Neyman allocation and consider the first 10%, 20% and 30% values in each stratum as non-response for W_{h2} =0.1, W_{h2} =0.2 and W_{h2}=0.3 respectively.

The population consists of village wise complete enumeration and data are obtained in 1951 and 1961 censuses for a Tehsil. The area of village is used to stratify the population into three strata.

Let y be the cultivated area in the village in hectares in 1951 and Ψ_{jhi}(j=1,2) are given below

Let

Ψ_{11i}=1, if i^th village in the stratum 1has area greater than 550 hectares =0, otherwise.

Ψ_{12i}=1, if i^th village in the stratum 2 has area greater than 1300 hectares =0, otherwise and

Ψ_{13i}=1, if i^th village in the stratum 3has area greater than 2500 hectares =0, otherwise.

Let

Ψ_{21i}=1, if i^th village in the stratum 1has number of cultivators greater than 550 =0, otherwise.

Ψ_{22i}=1, if i^th village in the stratum 2has number of cultivators greater than 700 =0, otherwise and

Ψ_{23i}=1, if i^th village in the stratum 3has number of cultivators greater than 1500 =0, otherwise.

Data is presented in Table 4.1 and results are given in Table 4.2.

Table 4.1: Data Statistics

Stratum no.	1	2	3
N_h	43	45	40
n_h	10	12	18
$\bar{Y}_h$	397.1425	760.1795	1234.6180
P_{1h}	0.5814	0.4444	0.4000
P_{2h}	0.39535	0.5333	0.4000
S_{yh}^2	39975.0569	61455.990	172425.900
$S_{\psi_{1h}}^2$	0.2492	0.2525	0.2462
$S_{\psi_{2h}}^2$	0.2448	0.2545	0.2462
$\rho_{y\psi_{1h}}$	0.6922	0.3750	0.5057
$\rho_{y\psi_{2h}}$	0.3956	0.2847	0.3261
$\rho_{\psi_1\psi_2 h}$	0.2040	0.3884	0.5833

Table 4.2: Percent Relative Efficiencies (PRE) of estimators

Estimators	PRE's
$\bar{y}_{st}$	100
t_1	309.854
t_2	15.125
t_3	125.232
t_4	31.358
t_5	58.694
t_7	26.482
t_8	71.753
$MSE(t_9)_{min}$	428.773

We compute percent relative efficiencies as

$$PRE = \frac{var\left(\bar{y}_{st}\right)}{MSE\left(\bar{Y}_i\right)} x100, \qquad (\text{ i=1, 2, 3})$$

CONCLUSION

Superiority of the proposed estimator is established theoretically by the conditions derived in section 3. Results in Table 4.2 confirm this superiority numerically using the previously used data set.

REFERENCES

1. Diana, G., 1993. A class of estimators of the population mean in stratified random sampling Statistica, 53(1): 59-66.
2. Singh, H.P. and G.K. Vishwakarma, 2005. Combined Ratio-Product Estimator of Finite Population Mean in Stratified Sampling. Metodologia de Encuestas, 8: 35-44.

3. Singh, R., M. Kumar, M.K. Chaudhary and C. Kadilar, 2009. Improved Exponential Estimator in Stratified Random Sampling. Pak. Jour. of Stat. and Oper. Res. 5(2): 67-82.
4. Koyuncu, N. and C. Kadilar, 2008. Ratio and product estimators in stratified random sampling. J. Statist. Plann. Inference. 139(8): 2552-2558.
5. Koyuncu, N. and C. Kadilar, 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.
6. Singh, R. and M. Kumar, 2012. Improved Estimators of Population Mean Using Two Auxiliary Variables in Stratified random Sampling. Pak. Jour. of Stat. and Oper. Res., 8(1): 65-72.
7. Malik, S. and R. Singh, 2012. Some improved multivariate ratio type estimators using geometric and harmonic means in stratified random sampling. ISRN Prob. and Stat. Article ID 509186, doi:10.5402/2012/509186.
8. Singh, M.P., 1965. On the estimation of ratio and product of the population parameters. Sankhya, B, 27: 321-328.
9. Perri, P.F., 2007. Improved ratio-cum-product type estimators. Statist. Trans., 8: 51-69.
10. Bahl, S. and R.K. Tuteja, 1991. Ratio and product type exponential estimator. Jour. of Inf. and Opt. Sci. XII, I: 159-163.
11. Singh, R., P. Cauhan, N. Sawan and F. Smarandache, 2007. Auxiliary information and a priory values in construction of improved estimators. Renaissance High press.
12. Naik, V.D. and P.C. Gupta, 1996. A note on estimation of mean with known population proportion of an auxiliary character. Jour. of the Ind. Soc. of Agr. Stat., 48(2): 151-158.
13. Jhajj, H.S., M.K. Sharma and L.K. Grover, 2006. A family of estimators of population mean using information on auxiliary attribute. Pak. Journal of Stat., 22(1): 43-50.
14. Shabbir, J. and S. Gupta, 2007. A new estimator of population mean in stratified sampling, Commun. Stat. Theo. Meth., 35: 1201-1209.
15. Singh, R., P. Chauhan, N. Sawan and F. Smarandache, 2008. Ratio estimators in simple random sampling using information on auxiliary attribute. Pak. J. Stat. Oper. Res., 4(1): 47-53.
16. Abd-Elfattah, A.M. E.A. El-Sherpieny, S.M. Mohamed and O.F. Abdou, 2010. Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Appl. Math. and Comp. doi:10.1016/j.amc.2009.12.041.
17. Singh, H.P. and R.S. Solanki, 2012. Improved estimation of population mean in simple random sampling using information on auxiliary attribute. Appl. Math. and Comp., 218: 7798-7812.
18. Gujarati, D.N. and Sangeetha, 2007. Basic Econometrics. Tata McGraw - Hill.
19. Udny, Y.G., 1912. On the Methods of Measuring Association between Two Attributes. J. the Roy. Stat. Soc., 75: 579-642.
20. Shabbir, J. and S. Gupta, 2010. Estiamtion of the finite population mean in two phase sampling when auxiliary variables are attributes. Hacet. J. Math. and Stat., 39(1): 121-129.
21. Hanif, M., I. Haq and M.Q. Shahbaz, 2009. On a new family of estimators using multiple auxiliary attributes. W. A. S. J., 7(11): 1419-1422.
22. Hanif, M., I. Haq and M.Q. Shahbaz, 2010. Ratio Estimators using Multiple Auxiliary Attributes. W. A. S. J., 8(1): 133-136.
23. Butt, N.S., S. Kamal and M.Q. Shabaz, 2011. Multivaraite estimators for two phase sampling. W.A.S.J. 3(10): 2216-2223.
24. Malik, S. and R. Singh, 2013. A Family Of Estimators Of Population Mean Using Information On Point Bi-Serial And Phi-Correlation Coefficient. Int. Jour. Of Stat. Econ. (10)1.
25. Plikusas, A., 2008. Some overview of the ratio type estimators In: Workshop on survey sampling theory and methodology, Statistics Estonia.
26. Malik, S. and R. Singh, 2012. A Family of estimators of population mean using information on two auxiliary attribute. WASJ.
27. Tracy, D.S., H.P. Singh and R. Singh, 1996. An alternative to the ratio-cum-product estimator in sample surveys. Jour. of Statist. Plann. and Infere., 53: 375-387.
28. Srivenkataramana, T. and D.S. Tracy, 1981. Extending product method of estimation to positive correlation case in surveys. Austral. J. Statist. 23, 1981, 95-100.
29. Murthy, M.N., 1967. Sampling Theory and Methods. Statistical Publishing Society, Calcutta.