

IMPROVED ESTIMATOR OF FINITE POPULATION MEAN USING AUXILIARY ATTRIBUTE IN STRATIFIED RANDOM SAMPLING

Hemant K. Verma, Prayas Sharma and [†]Rajesh Singh

Department of Statistics, Banaras Hindu University, Varanasi-221005

[†]Corresponding author

Abstract

The present study discuss the problem of estimating the finite population mean using auxiliary attribute in stratified random sampling. In this paper taking the advantage of point bi-serial correlation between the study variable and auxiliary attribute, we have improved the estimation of population mean in stratified random sampling. The expressions for Bias and Mean square error have been derived under stratified random sampling. In addition, an empirical study has been carried out to examine the merits of the proposed estimator over the existing estimators.

Key words: Efficiency, auxiliary attribute, mean square error, ratio estimator, regression estimator, stratified random sampling.

1. Introduction

In survey sampling, it is usual to make use of the auxiliary information at the estimation stage in order to improve the precision or accuracy of an estimator of unknown population parameter of interest. Ratio, product and regression methods of estimation are good examples in this context. Many authors including Upadhyaya and Singh (1999), Kadilar and Cingi (2005), Khoshnevisan et al. (2007), Singh et al. (2008a, b, c), Singh et al. (2009) and Singh and Kumar (2011) suggested estimators using known population parameters of an auxiliary variable. But there may be many practical situations when auxiliary information is not available directly but is qualitative in nature, that is, auxiliary information is available in the form of an attribute. For example the height of a person may depend on the fact that whether the person is male or female. The efficiency of a Dog may depend on the particular breed of that Dog. In these situations by taking the advantage of point bi-serial correlation between the study variable y and the auxiliary attribute ϕ along with the prior knowledge of the population parameter of auxiliary attribute, the estimators of population parameter of interest can be constructed.

Taking into consideration the point bi-serial correlation between auxiliary attribute and study variable, several authors including Naik and Gupta (1996), Singh et al. (2007), Shabbir and Gupta (2007), Singh et al. (2010), Abd-Elfattah et al. (2010), Singh and Solanki (2013), Malik and Singh (2013 a, b), Sharma et al. (2013 a, b), Verma et al. (2013) proposed improved estimators of population mean. In this paper we propose an estimator using the auxiliary attribute in stratified random sampling.

Let, the size of the population is N and it is stratified into L strata. Let h^{th} stratum containing N_h units, where $h=1, 2, 3, \dots, L$ such that $\sum_{h=1}^L N_h = N$. A sample of size n_h is drawn using simple random sampling without replacement (SRSWOR) from the h^{th} stratum such that $\sum_{h=1}^L n_h = n$. Let y_{hi} and φ_{hi} denote observed values of study variable Y and auxiliary attribute respectively, on the i^{th} unit of the h^{th} stratum, where $i=1, 2, 3, \dots, N_h$ and $h=1, 2, 3, \dots, L$. We note that, if the i^{th} unit of the population possesses attribute φ then $\varphi_{hi}=1$ otherwise 0. Let $A = \sum_{i=1}^N \varphi_i$, $A_h = \sum_{i=1}^{N_h} \varphi_{hi}$, $a = \sum_{i=1}^n \varphi_i$ and $a_h = \sum_{i=1}^{n_h} \varphi_{hi}$ denote the total number of units in the population, population stratum h , sample and sample stratum h possessing attribute φ .

Let $P = \frac{A}{N}$, $P_h = \frac{A_h}{N_h}$ and $p_h = \frac{a_h}{n_h}$ denote the proportion of units in population, h^{th} population stratum and h^{th} sample stratum, respectively.

The usual ratio and regression estimator when auxiliary variable is attribute, are respectively given by

$$t_1 = \bar{y}_{st} \left(\frac{P}{p_{st}} \right) \quad (1.1)$$

$$t_2 = \bar{y}_{st} + b(P - p_{st}) \quad (1.2)$$

where $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$ and

$$b = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{\varphi y h}}{\sum_{h=1}^L W_h^2 \gamma_h S_{\varphi h}^2} \quad \text{is sample regression coefficient of } y \text{ on } x.$$

Bias and MSE expression of t_1 and t_2 are respectively given by

$$\text{Bias}(t_1) = \bar{Y}[V_2 - V_3] \quad (1.3)$$

$$\text{Bias}(t_2) = \beta P[V_4 - V_5] \quad (1.4)$$

$$\text{MSE}(t_1) = \bar{Y}^2[V_1 + V_2 - 2V_3] \quad (1.5)$$

$$\text{MSE}(t_2) = \bar{Y}^2 V_1 + \beta^2 P^2 V_2 - 2\beta P \bar{Y} V_3 \quad (1.6)$$

2. Proposed Estimator

Motivated by Solanki et al.(2013), we propose an estimator using information on attribute in stratified sampling as

$$t_s = \left[\bar{y}_{st} + b(P - p_{st}) \right] \left(\frac{aP + (1-a)p_{st}}{ap_{st} + (1-a)P} \right) \quad (2.1)$$

where a is constant and b is regression coefficient.

To obtain the Bias and Mean square errors expression of (2.1), we use large sample approximation

$$\bar{y}_{st} = \bar{Y}(1 + e_0), \quad p_{st} = P(1 + e_1), \quad s_{\phi h}^2 = S_{\phi h}^2(1 + e_2), \quad s_{\phi y h}^2 = S_{\phi y h}^2(1 + e_3),$$

$$b = \beta(1 + e_3)(1 + e_2)^{-1},$$

$$e_2 = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi h}^2 e_{2h}}{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi h}^2}, \quad e_3 = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi y h} e_{3h}}{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi y h}}$$

such that $E(e_0) = E(e_1) = E(e_2) = E(e_3) = 0$.

$$E(e_0^2) = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{y h}^2}{\bar{Y}^2} = V_1, \quad E(e_1^2) = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi h}^2}{P^2} = V_2,$$

$$E(e_0 e_1) = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi y h}}{\bar{Y} P} = V_3, \quad E(e_1 e_2) = \frac{\sum_{h=1}^L W_h^2 \gamma_h \mu_{30h} a_h}{P} = V_4,$$

$$E(e_1 e_3) = \frac{\sum_{h=1}^L W_h^2 \gamma_h \mu_{21h} a_h^*}{P} = V_5$$

$$\text{where, } S_{y h}^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}, \quad S_{\phi y h}^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(p_{hi} - P_h)}{N_h - 1}, \quad \gamma_h = \frac{1}{n_h} - \frac{1}{N_h}$$

$$a_h = \frac{N_h^2 W_h \gamma_h}{(N_h - 1)(N_h - 2) \sum_{h=1}^L W_h^2 \gamma_h S_{\psi h}^2},$$

$$a_h^* = \frac{N_h^2 W_h \gamma_h}{(N_h - 1)(N_h - 2) \sum_{h=1}^L W_h^2 \gamma_h S_{\phi y h}}$$

$$\mu_{rsh} = \frac{1}{N_h} \sum (y_{hi} - \bar{Y}_h)^r (p_{hi} - P_h)^s, \quad \beta = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi y h}}{\sum_{h=1}^L W_h^2 \gamma_h S_{\phi h}^2}$$

Expressing (2.1) in terms of e's we have

$$t_s = \left[\bar{Y}(1 + e_0) - P\beta e_1(1 + e_3)(1 + e_2)^{-1} \right] (1 + (1 - a)e_1)(1 + ae_1)^{-1} \quad (2.2)$$

We assume that $|e_2| < 1$ and $|ae_1| < 1$ so that the term $(1 + e_2)^{-1}$ and $(1 + ae_1)^{-1}$ can be expanded.

Expanding the right hand side of (2.2) up to the first order of approximation, we have

$$t_s - \bar{Y} = \bar{Y} \left[(1 - 2a)e_1 - a(1 - 2a)e_1^2 + e_0 + (1 - 2a)e_0e_1 \right] - \beta P \left[e_1 + (1 - 2a)e_1^2 - e_1e_2 + e_1e_3 \right] \quad (2.3)$$

Taking expectation of both sides of (2.3), we get the Bias expression of t_s as

$$\text{Bias}(t_s) = \bar{Y} \left[(1 - 2a)V_3 - a(1 - 2a)V_2 \right] - \beta P \left[(1 - 2a)V_2 - V_4 + V_5 \right] \quad (2.4)$$

Squaring both sides of (2.3) and neglecting the terms e's having power greater than two, we have

$$(t_s - \bar{Y})^2 = \left[(1 - 2a)^2 e_1^2 + e_0^2 + 2(1 - 2a)e_0e_1 \right] + \beta^2 P^2 e_1^2 - 2\beta P \bar{Y} \left[(1 - 2a)e_1^2 + e_0e_1 \right] \quad (2.5)$$

Taking expectations of both sides of (2.4), we get the MSE of t_s given by

$$\text{MSE}(t_s) = \bar{Y} \left[V_1 + (1 - 2a)^2 V_2 + 2(1 - 2a)V_3 \right] + \beta^2 P^2 V_2 - 2\beta P \bar{Y} \left[(1 - 2a)V_2 + V_3 \right] \quad (2.6)$$

Differentiating (2.6) with respect to a and equating to zero, we get the optimum value of a as

$$a^* = \frac{\bar{Y}(V_2 + V_3) - P\beta V_2}{2\bar{Y}V_2} \quad (2.7)$$

By substituting the optimum value of a in (2.6) we get the min MSE of t_s .

3. Empirical Study

To illustrate the efficiency of suggested estimators in the application, we consider the data concerning the number of teachers as the study variable (y) and for auxiliary attribute we use number of students classifying more or less than 750, in both primary and secondary schools as auxiliary variable for 923 districts at 6 regions (as 1:Marmara 2:Aegean 3:Mediterranean 4:Central Anatolia 5:Black Sea 6:East and Southeast Anatolia) in Turkey in 2007 (source: The Turkish Republic Ministry of Education). The summary statistics of the data are given in Table 2. We used Neyman allocation for allocating the samples to different strata (Cochran, 1977).

Table 3.1: Data Statistics

Values	Stratum (h)					
	1	2	3	4	5	6
N_h	127	117	103	170	205	201
n_h	31	21	29	38	22	39
\bar{Y}_h	707.74	413	573.17	424.66	267.03	393.84
S_{yh}	883.835	944.922	1033.467	810.585	403.654	711.723
$S_{\phi h}$	0.213	0.159	0.253	0.316	0.284	0.218
P_h	0.952	0.974	0.932	0.888	0.912	0.950
$S_{\phi y h}$	25.267	9.982	37.453	44.625	21.04	18.66
$\beta_2(\phi_h)$	16.922	35.579	10.34	4.231	6.675	15.56

Table 3.2: MSE and Percentage relative efficiency of estimators

Estimator	MSE	PRE with respect to \bar{y}_{st}
\bar{y}_{st}	70061.37	100
t_1	2189.33	101.83
t_2	2185.53	102.00
t_s	2185.54	102.00

4. Conclusion

From theoretical discussion and empirical study we conclude that the proposed estimator t_s under optimum conditions performs better than usual ratio estimator when auxiliary variable is an attribute under stratified sampling and it the minimum MSE is similar to that of usual regression estimator. The relative efficiencies and MSE of various estimators are listed in Table 3.2.

References :

1. Abd-Elfattah, A.M., Sherpeny E.A., Mohamed S.M., Abdou O.F. (2010): Improvement estimating the population mean in simple random sampling using information on auxiliary attribute. *Applied mathematics and computation*.
2. Cochran, W.G. (1977): *Sampling techniques*. Third U.S. edition. Wiley eastern limited, 325.
3. Kadilar,C. and Cingi,H. (2005): A New Ratio Estimators in Stratified Random Sampling. *Commun. Stat. Theo. And Meth.*34, 597-602.
4. Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2007): A general family of estimators for estimating population means using known value of some population parameter(s). *Far east journal of theoretical statistics*, 22(2), 181-191.
5. Malik, S. and Singh, R. (2013a): An improved estimator using two auxiliary attributes. *Appli. Math. Compt.*, 219, 10983-10986.
6. Malik, S. and Singh, R. (2013b): A family of estimators of population means using information on point bi-serial and phi correlation coefficient. *IJSE* 10(1), 75-89.
7. 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. Agr. Stat.*, 48(2), 151-158.
8. Shabbir, J. and Gupta, S., (2007): On estimating the finite population mean with known population proportion of an auxiliary variable, *Pakistan Journal of Statistics* 23 (1) 1-9.
9. Sharma, P., Verma, H., Sanaullah, A. and Singh, R. (2013a): Some exponential ratio-product type estimators using information on auxiliary attributes under second order approximation. *IJSE*, 12(3), 58-66.
10. Sharma, P., Singh, R. and Kim, J. M. (2013b): Study of some improved ratio type estimators using information on auxiliary attributes under second order approximation. *Jour. Sci. Res.*, 57, 138-146.
11. Solanki, R. S. and Singh, H.P. (2013): Improved estimation of population mean in using population proportion of an auxiliary character. *Chilean journal of Statistics*. 4 (1), 3-17.
12. Solanki, R. S., Singh, H. P. and Pal S. K. (2013): Improved Estimation of Finite Population Mean in Sample Surveys. *Columbia International Publishing Journal of Advanced Computing*, 1: 70-78.
13. Singh, R., Cauhan, P., Sawan, N., and Smarandache, F.,(2007): *Auxiliary information and a priori values in construction of improved estimators*. Renaissance High Press.
14. Singh, R., Kumar, M. and Smarandache, F., (2008a): Almost unbiased estimator for estimating population mean using known value of some population parameter(s). *Pak. J. Stat. Oper. Res.*, 4(2) 63-76.

15. Singh, R., Chauhan, P. and Sawan, N.(2008b): On linear combination of Ratio-product type exponential estimator for estimating finite population mean. *Statistics in Transition*,9(1),105-115.
16. Singh, R., Chauhan, P., Sawan, N. and Smarandache,F. (2008c): Ratio estimators in simple random sampling using information on auxiliary attribute. *Pak. J. Stat. Oper. Res.*,4,1,47-53.
17. Singh, R., Chauhan, P., Sawan, N. and Smarandache,F.(2009): Improvement in estimating the population mean using exponential estimator in simple random sampling. *Intern. Jour. Of Stat. and Econ. (BSE)*,3,13-18.
18. Singh, R. and Kumar, M. (2011): A note on transformations on auxiliary variable in survey sampling. *MASA*, 6:1, 17-19.
19. Singh, R., Kumar, M. and Smarandache, F., (2010): Ratio estimators in simple random sampling when study variable is an attribute. *World Applied Sciences Journal* 11 (5) pp 586-589.
20. Upadhyaya, L.N. and Singh, H. P. (1999): An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. *Vikram Mathematical Journal*, 19, 14-17.
21. Verma, H. K., Singh, R. and Smarandache, F. (2013): Some improved estimators of population mean using information on two auxiliary attributes. *Educational Publishing & Journal of Matter Regularity (Beijing)*, pp. 17-24.