Volume 67, Issue 1, 2023
Journal of Scientific Research
of
The Banaras Hindu University

IMPROVED ESTIMATORS OF POPULATION MEAN USING AUXILIARY VARIABLES IN POST-STRATIFICATION

Anamika Kumari*1,Upendra Kumar2,Rajesh Singh3

[1]Department of Statistics, Banaras Hindu University, anamikatiwari1410@gmail.com
[2]Department of Statistics, UP Autonomous College, ukumarupc@gmail.com
[3]Depart*ment of Statistics, Banaras Hindu University, rsinghstat@gmail.com

Abstract: **Through this paper, we present some improved estimators of population mean using auxiliary variables in post-stratification. We have derived the expressions for bias and mean square errors up to the first order of approximation and shown that the proposed estimators under optimum conditions are more efficient than other estimators taken in this paper. In an attempt to verify the efficiencies of proposed estimators theoretical results are supported by empirical study for which we have considered two datasets.**

***Index Terms:*Study variable, auxiliary variable, bias, mean square error and post-stratified sampling**

I. INTRODUCTION

Stratified sampling is a method of sampling from a population that can be divided into subpopulations known as strata. This method improves efficiency over simple random sampling when the population is not homogenous. In some cases, it is not possible to divide the population into strata before sampling then post-stratification comes forward. For example

• We cannot specify a population by age group until the census is conducted.

• It is impossible to stratify a population by gender if the sample is drawn using a telephonic interview.

For post-stratification, a simple random sample is drawn and then units are placed in strata. Many authors worked on the problem of post-stratification. Hansen et al (1953) were the first to tackle the problem of post-stratification. Holt and Smith (1979) showed that post-stratification is more efficient than stratification. Other works include the work of Agarwal and Panda (1993), Fuller (1966), Chouhan (2012), Bahl and Tuteja (1991), and Kish (1965) who have studied this technique of sampling.

We have utilized the information on an auxiliary variable in order to improve the efficiency of the estimator of the population mean. Cochran (1940) was the first to introduce a ratio estimator of the population mean. Shabbir and Gupta (2007), Singh et al. (2013), Singh et al. (2016), and Koyuncu and Kadilar (2009), have considered the problem of estimating the population mean Y by taking into consideration information on auxiliary variables. This article tries estimation of population mean in post-stratified sampling using the information on auxiliary variables.

Let a sample of size n is drawn from a size N using SRSWOR from a population $U = (U_1, U_2 \ \dots, U_N)$. After selecting the sample, we classify which unit in the sample 'n' belongs to h^{th} stratum such that $\sum_{h=1}^{L} n_h = n$. Here the assumption is that 'n' is so large that the possibility for $'n_h'$ being zero is very small.

Let (y_{hi}, x_{hi}) be the observation on i^{th} unit falling in the h^{th} stratum for the study variable Y and auxiliary variable X respectively. Let

$$\bar{Y}_h = \frac{1}{N}\sum_{i=1}^{N_h} y_{hi} ,$$

$$\bar{X}_h = \frac{1}{N}\sum_{i=1}^{N_h} x_{hi} ,$$

$$\bar{Y} = \frac{1}{N_h}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{L} W_h \bar{Y}_h ,$$

$$\bar{X} = \frac{1}{N_h}\sum_{h=1}^{L}\sum_{i=1}^{N_h} x_{hi} = \sum_{h=1}^{L} W_h \bar{X}_h ,$$

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} ,$$

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi} ,$$

where $W_h = \frac{N_h}{N}$.

I. Existing Estimators

The usual unbiased estimator of population mean $\bar{Y}$ in case of post-stratification is given by

$$\bar{y}_{ps} = \sum_{h=1}^{L} W_h \bar{y}_h \quad (2.1)$$

where $W_h = \frac{N_h}{N}$ is the weight of h^{th} stratum and N_h is assumed to be known (Holt and Smith (1979)).

The variance of the estimator $\bar{y}_{ps}$ is given by

$$V(\bar{y}_{ps}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} W_h S_{yh}^2 + \frac{1}{n^2} \sum_{h=1}^{L} (1 - W_h) S_{yh}^2 \quad (2.2)$$

A separate ratio type estimator of population mean $\bar{Y}$ in the case of post-stratification is given as

$$t_1 = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{X}_h}{\bar{x}_h}\right) \quad (2.3)$$

The Bias and MSE of the estimator t_1 are obtained respectively

$$Bias(t_1) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} \bar{Y}_h \left(C_{xh}^2 - \rho_{xyh} C_{xh} C_{yh}\right) \quad (2.4)$$

$$MSE(t_1) = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\sum_{h=1}^{L} W_h S_{yh}^2 + \sum_{h=1}^{L} W_h R_{1h}^2 S_{xh}^2 - 2 \sum_{h=1}^{L} W_h R_{1h} S_{yxh}\right] \quad (2.5)$$

The separate product type estimator of the population mean $\bar{Y}$ in the case of post-stratification is given as

$$t_2 = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{x}_h}{\bar{X}_h}\right) \quad (2.6)$$

The Bias and MSE of the estimator t_2 are obtained as respectively

$$Bias\ (t_2) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} \bar{Y}_h \left(\rho_{xy}\ C_{xh} C_{yh}\right) \quad (2.7)$$

$$MSE(t_2) = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\sum_{h=1}^{L} W_h S_{yh}^2 + \sum_{h=1}^{L} W_h R_{1h}^2 S_{xh}^2 + 2 \sum_{h=1}^{L} W_h R_{1h} S_{yxh}\right] \quad (2.8)$$

where $R_{1h} = \frac{\bar{Y}_h}{\bar{X}_h}$.

II. PROPOSED ESTIMATORS

Motivated by Bhushan et al. (2017) we propose an improved exponential type estimator t_{g1} , improved log exponential estimator t_{g2}, a class of estimator t_p for estimating the population mean in post-stratified sampling as

- $$t_{g1} = \sum_{h=1}^{L} W_h \left[\bar{y}_h\, exp\left(\alpha_h \left(\frac{\bar{x}_h}{\bar{X}_h} - 1\right)\right)\right] \quad (3.1)$$
- $$t_{g2} = \sum_{h=1}^{L} W_h \left[\bar{y}_h exp\left(\beta_h \log \frac{\bar{x}_h}{\bar{X}_h}\right)\right] \quad (3.2)$$
- $$t_p = \sum_{h=1}^{L} W_h \left[\omega_{0h} \bar{y}_h + \omega_{1h} \bar{y}_h\, exp\left(\frac{\bar{X}_h - \bar{x}_h}{\bar{X}_h + \bar{x}_h}\right) + \omega_{2h} \bar{y}_h \left(1 + \log \frac{\bar{x}_h}{\bar{X}_h}\right)\right] \quad (3.3)$$

We have considered a class of estimator t_p such that $(t_{p0}, t_{p1}, t_{p2}) \in \Omega$, where Ω is a set of all possible estimators for estimating the population mean $\bar{Y}$. By definition, set Ω is a linear variety if

$$t_p = \omega_{0h} t_{p0} + \omega_{1h} t_{p1} + \omega_{2h} t_{p2} \quad \epsilon \Omega \quad (3.4)$$

$$for \sum_{i=0}^{2} \omega_{ih} = 1 \quad , \omega_{ih} \epsilon \mathcal{R} \quad (3.5)$$

where $\omega_{0h}, \omega_{1h}, and\ \omega_{2h}$ are the constants used for reducing the bias in the set of real numbers. $\mathfrak{R}$denotes the set of real numbers,

where t_{p0} is the usual unbiased estimator of the population mean.

$$t_{p0} = \sum_{h=1}^{L} W_h \bar{y}_h \quad (3.6)$$

Motivated by Bahl and Tuteja (1991)

$$t_{p1} = \sum_{h=1}^{L} W_h\, \bar{y}_h\, exp\left(\frac{\bar{X}_h - \bar{x}_h}{\bar{X}_h + \bar{x}_h}\right) \quad (3.7)$$

Motivated by Bhushan et al. (2015)

$$t_{p2} = \sum_{h=1}^{L} W_h\, \bar{y}_h \left(1 + \log \frac{\bar{x}_h}{\bar{X}_h}\right) \quad (3.8)$$

In order to obtain the expressions of Bias and MSE of suggested estimators, we assume that

$$\bar{y}_h = \bar{Y}_h (1 + \epsilon_{0h})$$
$$\bar{x}_h = \bar{X}_h (1 + \epsilon_{1h})$$
$$E(\epsilon_{0h}) = E(\epsilon_{1h}) = 0$$

$$E(\epsilon_{0h}^2) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right) C_{yh}^2$$

$$E(\epsilon_{1h}^2) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right) C_{xh}^2$$

$$E(\epsilon_{0h}\epsilon_{1h}) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right) \rho_{yxh} C_{yh} C_{xh}$$

$$C_{yh} = \frac{S_{yh}}{\bar{Y}_h} \quad , \quad C_{xh} = \frac{S_{xh}}{\bar{X}_h}$$

$$S_{yh}^2 = \frac{1}{N_h - 1} \sum_{h=1}^{L} (Y_{hi} - \bar{Y}_h)$$

$$S_{xh}^2 = \frac{1}{N_h - 1} \sum_{h=1}^{L} (X_{hi} - \bar{X}_h)$$

Using above expressions, we get

$$Bias(t_{p0}) = B_o = 0 \quad (3.9)$$

$$Bias(t_{p1}) = B_1 = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\frac{\bar{Y}_h}{8}\left(3C_{xh}^2 - 4\rho_{xyh}C_{xh}C_{yh}\right) \quad (3.10)$$

$$Bias(t_{p2}) = B_2 = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h\left(\rho_{xyh}C_{xh}C_{yh} - \frac{C_{xh}^2}{2}\right) \quad (3.11)$$

Expressing estimator t_{g1} defined by equation (3.1) in terms of ϵ's we get

$$t_{g1} = \sum_{h=1}^{L}W_h\left[\bar{Y}_h(1 + \epsilon_{0h})\, exp\left(\alpha_h\left(\frac{\bar{X}_h(1+\epsilon_{1h})}{\bar{X}_h} - 1\right)\right)\right] \quad (3.12)$$

The Bias of the estimator t_{g1} is given by

$$Bias(t_{g1}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h\left(\frac{{\alpha_h}^2C_{xh}^2}{2} + \alpha_h\rho_{xy}\ C_{xh}C_{yh}\right) \quad (3.13)$$

The MSE of the estimator t_{g1} is given by

$$MSE(t_{g1}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h^2W_h\left[C_{yh}^2 + 2\alpha_h\rho_{yxh}C_{yh}C_{xh} + {\alpha_h}^2C_{xh}^2\right] \quad (3.14)$$

In order to find out the minimum MSE for t_{g1} we partially differentiate equation (3.14) wrt α_h and equating to zero we get

$$\alpha_h^* = -\frac{\rho_{xyh}C_{yh}}{C_{xh}} \quad (3.15)$$

Putting the optimum value of α_h in equation (3.14) we get minimum MSE of t_{g1} as

$$\min MSE = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h^2W_hC_{yh}^2\left(1 - \rho_{yxh}^2\right) \quad (3.16)$$

Expressing estimator t_{g2} defined by equation (3.2) in terms of ϵ's we get

$$t_{g2} = \sum_{h=1}^{L}W_h\left[\bar{Y}_h(1 + \epsilon_{0h})\, exp\left(\beta_h \log\frac{\bar{X}_h(1+\epsilon_{1h})}{\bar{X}_h}\right)\right] \quad (3.17)$$

The Bias of the estimator t_{g2} is given by

$$Bias(t_{g2}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h\left(\frac{({\beta_h}^2 - \beta_h)C_{xh}^2}{2} + \beta_h\rho_{xyh}C_{xh}C_{yh}\right) \quad (3.18)$$

The MSE of the estimator t_{g2} is given by

$$MSE(t_{g2}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h^2W_h\left[C_{yh}^2 + 2\beta_h\rho_{yxh}C_{yh}C_{xh} + {\beta_h}^2C_{xh}^2\right] \quad (3.19)$$

In order to find out the minimum MSE for t_{g2} we partially differentiate equation (3.19) wrt β_h and equating to zero we get

$$\beta_h^* = -\frac{\rho_{xyh}C_{yh}}{C_{xh}} \quad (3.20)$$

Putting the optimum value of β_h in equation (3.19) we get minimum MSE of t_{g2} as

$$\min MSE = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h^2W_hC_{yh}^2\left(1 - \rho_{yxh}^2\right) \quad (3.21)$$

Expressing estimator t_p defined in the equation (3.3) in terms of ϵ's we get

$$t_p = \sum_{h=1}^{L}W_h\,\bar{Y}_h\left[1 + \epsilon_{0h} - \left(\frac{\omega_{1h}}{2} - \omega_{2h}\right)\epsilon_{1h} + \left(\frac{3\omega_{1h}}{8} - \frac{\omega_{2h}}{2}\right)\epsilon_{1h}^2 - \left(\frac{\omega_{1h}}{2} - \omega_2\right)\epsilon_{0h}\epsilon_{1h}\right] \quad (3.22)$$

Taking expectations of both sides of (3.22) and then subtracting $\bar{Y}$ from both sides, we get the bias of the estimator, t_p, up to the first order of approximation as

$$Bias(t_p) = E\left[\sum_{h=1}^{L}W_h\,\bar{Y}_h\left[\epsilon_{0h} - \omega_h\epsilon_{1h} + \frac{1}{8}(5\omega_{1h} - \omega_h)\epsilon_{1h}^2 - \omega_h\epsilon_{0h}\epsilon_{1h}\right]\right] \quad (3.23)$$

where

$\omega_h = \frac{\omega_{1h}}{2} - \omega_{2h} \Rightarrow 2\omega_h = \omega_{1h} - 2\omega_{2h}$. (3.24)

From (3.22), we have

$$t_p - \bar{Y}_h \cong \sum_{h=1}^{L}W_h\,\bar{Y}_h(\epsilon_{0h} - \omega_h\epsilon_{1h}) \quad (3.25)$$

Squaring both sides of (3.25) and then taking expectations, we get the MSE of the estimator t_p , up to the first order of approximation as

$$MSE(t_p) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h^2W_h\left[C_{yh}^2 + 2\omega_h\ \rho_{yxh}C_{yh}C_{xh} + {\omega_h}^2C_{xh}^2\right] \quad (3.26)$$

In order to find out the minimum MSE for t_p we partially differentiate equation (3.26) wrt ω_h and equating to zero we get

$$\omega_h^* = \frac{\rho_{xyh}C_{yh}}{C_{xh}} = H\ (say) \quad (3.27)$$

Putting the optimum value of ω_h in equation (3.26) we get minimum MSE of t_p as

$$\min MSE = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\bar{Y}_h^2W_hC_{yh}^2\left(1 - \rho_{yxh}^2\right) \quad (3.28)$$

From (3.24) and (3.27), we have

$$\omega_{1h} - 2\omega_{2h} = 2H \quad (3.29)$$

From (3.27) and (3.29), we have only two equations in three unknowns. It is not possible to find unique values of $\omega_{ih}(i = 0,1,2)$. In order to get unique values for ω_{ih}, we shall impose the linear restriction

$$\omega_{0h}B_0 + \omega_{1h}B_1 + \omega_{2h}B_2 = 0 \quad (3.30)$$

Equations (3.5), (3.29), and (3.30) can be written in matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -2 \\ 0 & B_1 & B_2 \end{bmatrix} \begin{bmatrix} \omega_{0h} \\ \omega_{1h} \\ \omega_{2h} \end{bmatrix} = \begin{bmatrix} 1 \\ 2H \\ 0 \end{bmatrix} \quad (3.31)$$

Using (3.31) we get the unique values of ω_{ih} as

$$\omega_{0h} = \frac{T_0}{T_3}, \qquad \omega_{1h} = \frac{T_1}{T_3}, \qquad \omega_{2h} = \frac{T_2}{T_3}$$

where

$T_0 = B_2(1 - 2H) + 2B_1(1 + H)$

$T_1 = 2HB_2$

$T_2 = -2HB_1$

$T_3 = 2B_1 + B_2$.

III. Empirical Study

In this section, we compare the performance of the estimators with the other estimators considered in this article using two population data sets are as follows

Population I- [Source: National horticulture Board] (Lone and Tailor (2015))

Y: Productivity (MT/Hectare)

X: Production in '000 Tons

Constant	Stratum 1	Stratum 2
N_h	10	10
n_h	4	4
$\bar{Y}_h$	1.7	3.67
$\bar{X}_h$	10.41	289.14
S_{yh}	0.5	1.41
S_{xh}	3.53	111.61
S_{yxh}	1.60	144.87

Population II- [Source: Chouhan (2012)]

Y: Snowy days

X: Rainy days and

Constant	Stratum 1	Stratum 2
N_h	10	10
n_h	4	4
$\bar{Y}_h$	149.7	102.6
$\bar{X}_h$	142.8	91
S_{yh}	13.46	12.60
S_{xh}	6.09	1.57
S_{yxh}	18.44	23.30

Table 1: value of ω_{ih} $(i = 0,1,2)$

Constant	Population 1	Population 2
ω_{0h}	-5.121954	-1.49730
ω_{1h}	5.217966	2.29976
ω_{2h}	0.903988	0.19754

ssses (PRE) of the Estimators

Estimator	Population 1		Population 2	
	MSE	PRE	MSE	PRE
t_{ps}	0.1014	100	15.4031	100
t_1	0.0141	719.1489	12.9134	119.2799
t_2	0.3291	30.8113	19.7536	77.9761
t_{g1}	0.0130	780	11.9391	129.0139
t_{g2}	0.0130	780	11.9391	129.0139
t_p	0.0130	780	11.9391	129.0139

Table2 shows the MSE and PRE of the estimators $t_{ps}, t_1, t_2, t_{g1}, t_{g2}, t_p$ for both the population 1 & 2. It can be seen that the suggested estimators perform better than the existing estimators $t_{ps}, t_1 \& t_2$. Talking of suggested estimators, all the estimators t_{g1}, t_{g2}, t_p have same MSE.

IV. Conclusion

In this article, we have proposed estimators for the population mean in post-stratified sampling using the information on auxiliary variables. The expressions for Bias and MSE of the suggested estimators have been derived up to the first order of approximation. An empirical approach for comparing the efficiency of the proposed estimators with other estimators has been used. For comparison we have used known natural population datasets, see National Horticulture Board and Chouhan (2012).The results have been shown in Table 2. Table 2 shows that the proposed estimators turn out to be more efficient as compared to the other estimators for both populations because of the higher value of PRE.

References

Agarwal, M.C., & Panda, K.B. (1993). An efficient estimator in post-stratification. *Metron, 51(3-4),* 179-187

Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of information and optimization sciences, 12(1),* p. 159- 164

Bhushan, S., Gupta, R., and Pandey, S.K. (2015). Some Log-Type Classes of Estimators using Multiple Auxiliary Information. *International Journal of Scientific Engineering and Research,* ISSN 2347-3878.

Bhushan, S., Misra, P.K., and Yadav, S.K. (2017). A Generalised Class of Estimators for Population Mean Using Auxiliary Information on an Attribute and an Auxiliary Variable. *International Journal of Computational & Applied Mathematics*, ISSN 1819-4966 vol 12, 11-28.

Chouhan S, (2012). Improved estimation of parameters using auxiliary information in sample surveys. *Ph. D. Thesis, Vikram University, Ujjain, M.P., India.*

Cochran, W. G. (1940). The estimation of the yield of cereal experiments by sampling for the ratio of gain to total produce, *Journal of Agricultural Science*, 30, 262–275.

Cochran, W.G. (1977).Sampling Techniques. *Wiley Eastern Limited*.

Fuller, W., (1966). Estimation employing post-strata. *J. Amer. Statist. Assoc.,* 61, 1172−1183.

Hansen, M.H., and Hurwitz, W.N. (1953). Sample survey methods and theory. *Vol. I & II, John Wiley & Sons, New York.*

Holt, D., Smith, T. M. F., (1979). Post-stratification. *J. Roy. Statist. Soc., 142, A,* 33−46.

Kish, L., (1965). Survey Sampling. John Wiley & Sons, New York.

Koyuncu, N., Kadilar, C. (2009). Family of Estimators of Population Mean Using Two Auxiliary Variables in Stratified Random Sampling. *Communication in Statistics- Theory and Methods 38(14):*2398-2417

Lone, H.A., & Tailor, R. (2015). Improved separate ratio and product exponential type estimators in the case of post-stratification. *Statistics in Transition new series, 1(16),* 53-64.

Shabbir, J., Gupta, S. (2007). On Estimating the finite population mean with known population proportion of an auxiliary variable. *Pakistan Journal of Statistics* 23(1).

Sharma, P., and Singh, R. (2013).Efficient Estimators of Population Mean in Stratified Random Sampling Using Auxiliary Attribute. *World Applied Sciences Journal 27 (12)*, ISSN 1818-4952.

Singh, P., Adichwal, N.K., and Singh, R. (2016). Estimation of Finite Population Mean Using Auxiliary Attribute in Sample Surveys. *J Adv. Res Appl Math Stat; 1(2):* 39-44.

***