A new ratio type estimator for computation of population mean under post-stratification

K. UL ISLAM RATHER, M. IQBAL JEELANI, M. YOUNIS SHAH,
S. E. H. RIZVI AND M. SHARMA

Abstract

In this study, the difficulty of estimating the population mean in the situation of post-stratification is discussed. The case of post-stratification is presented for ratio-type exponential estimators of finite population mean. Mean-squared error of the proposed estimator is obtained up to the first degree of approximation. In the instance of post-stratification, the proposed estimator was compared with the existing estimators. An empirical study by using some real data and further, simulation study has been carried out to demonstrate the performance of the proposed estimator.

Mathematics Subject Classification 2000: 62D05
Keywords: post-stratification, auxiliary variable, exponential ratio estimator, mean squared error, efficiency.

1. INTRODUCTION

One of the most prevalent survey sample approaches is stratification. The use of stratified random sampling ensures that the size of each stratum as well as the sampling frame for each stratum is known ahead of time. However, in many practical cases, the overall population size and percentage of the unit that falls into distinct strata or stratum sizes are known, but a sampling frame for each stratum is either not accessible or is too expensive and time-consuming to produce. We can't use stratified random sampling in these scenarios. To overcome this problem, a post-stratification procedure is utilized, in which a sample of the requisite size is taken from the population using SRS and then stratified using the stratification factor. Many authors contributed in the field of post stratification including Holt and Smith (1979), Jagers et al. (1985), Jagers (1986), Ige and Tripathi (1989), Agrawal and Panda (1993), and Singh and Ruiz Espezo (2003).

We assume a finite population of size N which is divided into L strata of sizes $N_1, N_2, \dots, N_L$ such that $\sum_{h=1}^{L} N_h = N$. Let y be the study variate, and x be the

10.2478/jamsi-2022-0003
©2022 Khalid Ul Islam Rather, M. Iqbal Jeelani, M. Younis Shah, S. E. H. Rizvi, Manish Sharma
This is an open access article licensed under the Creative Commons Attribution Licence (http://creativecommons.org/licenses/by/4.0).
sciendo

auxiliary variate, positively and negatively correlated with study variate y, respectively. Let y_{hi} be the observation on i^{th} unit of h^{th} stratum for study variate, and x_{hi} be the observation on i^{th} unit of h^{th} stratum for auxiliary variate x. For study variate y and auxiliary variate x population means are denoted by $\bar{Y}$ and $\bar{X}$, respectively, while h^{th} stratum means are denoted by $\bar{Y}_h$ and $\bar{X}_h$ respectively. A sample of size n is drawn from the population using simple random sampling without replacement (SRSWR). It is observed that how many and which units belong to the h^{th} stratum. Let n_h be the size of the sample falling in, h^{th} stratum such that $\sum_{h=1}^{L} n_h = n$ here, it is assumed that n is so large that possibility of n_h being zero is very small.

After stratification, ratio and product estimators for population mean were discussed Ige and Tripathi (1989). Tailor investigated the ratio exponential estimator of Bahl and Tuteja (1991) in post-stratification Tailor *et al.* (2017). The authors are motivated to research ratio and ratio type exponential estimators in post-stratification as a result of their previous work. In post stratification, usual unbiased estimator for population mean is explained as.

$$\bar{y}_{ps} = \sum_{h=1}^{H} W_h \bar{y}_h \tag{1.1}$$

where, $\frac{N_h}{N}$ is the weight of h^{th} stratum and $\bar{y}_h$ is the sample mean of n_h sample units that fall in the h^{th} stratum.

Using the results from Stephen (1945), the variance is given as

$$Var\left(\hat{\bar{Y}}_{ps}\right) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} W_h S_{yh}^2 + \frac{1}{n^2} \sum_{h=1}^{L} (1 - W_h) S_{yh}^2 \tag{1.2}$$

Ige and Tripathi (1989) defined classical ratio type estimator in post-stratification, as

$$\hat{\bar{Y}}_R = \bar{y}_{ps} \left(\frac{\bar{X}}{\bar{x}_{ps}}\right) \tag{1.3}$$

where

$$\bar{X} = \sum_{h=1}^{L} W_h \bar{X}_h \; and \quad \bar{y}_{ps} = \sum_{h=1}^{L} W_h \bar{y}_h , \qquad \bar{x}_{ps} = \sum_{h=1}^{L} W_h \bar{x}_h$$

Upto the fda, the bias and MSE statement of $\hat{\bar{Y}}_{ps}^{R}$ is defined as

$$B\left(\hat{\bar{Y}}_R\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\bar{X}}\sum_{h=1}^{L} W_h\left(R_1 S_{xh}^2 - S_{yxh}\right) \tag{1.4}$$

And

$$MSE\left(\hat{\bar{Y}}_R\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\left(S_{yh}^2 + R_1^2 - 2R_1 S_{yxh}\right) \tag{1.5}$$

where $R_1 = \frac{\bar{Y}}{\bar{X}}$

Tailor *et al.* (2017) defined Bahl and Tuteja (1991) ratio type exponential estimator as

$$\hat{\bar{Y}}_{RC} = \bar{y}_{ps} exp\left(\frac{\bar{x}_{ps} - \bar{X}}{\bar{x}_{ps} + \bar{X}}\right) \tag{1.6}$$

Up to the fda, the bias and MSE assertion of $\hat{\bar{Y}}_{RC}$ are respectively given by

$$B\left(\hat{\bar{Y}}_{RC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\bar{X}}\sum_{h=1}^{L} W_h\left(\frac{3}{8}R_1 S_{xh}^2 + \frac{1}{2}S_{yxh}\right) \tag{1.7}$$

$$MSE\left(\hat{\bar{Y}}_{RC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\left(S_{yh}^2 + \frac{1}{4}R_1^2 S_{xh}^2 - R_1 S_{yxh}\right) \tag{1.8}$$

$$\hat{\bar{Y}}_{PC} = \bar{y}_{ps} exp\left(\frac{\bar{X} - \bar{x}_{ps}}{\bar{X} + \bar{x}_{ps}}\right) \tag{1.9}$$

Up to the fda, the bias and MSE assertion of $\hat{\bar{Y}}_R$ are respectively given by

$$B\left(\hat{\bar{Y}}_{PC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\bar{X}}\sum_{h=1}^{L} W_h\left(\frac{3}{8}R_1 S_{xh}^2 - \frac{1}{2}S_{yxh}\right) \tag{1.10}$$

$$MSE\left(\hat{\bar{Y}}_{PC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\left(S_{yh}^2 + \frac{1}{4}R_1^2 S_{xh}^2 - R_1 S_{yxh}\right) \tag{1.11}$$

2. PROPOSED ESTIMATOR

Motivated by Subramani (2016), we have proposed the following general ratio type estimator of population mean of study variable under post stratification.

$$\hat{\bar{Y}}_{RK} = \bar{y}_{PS}\left[k + (1-k)exp\left(\frac{\bar{X} - \bar{x}_{ps}}{\bar{X} + \bar{x}_{ps}}\right)\right] \tag{2.1}$$

where k is constant and we write,

To obtain the MSE of $\hat{\bar{Y}}_{RK}$, write

$$\mathrm{e}_0 = \frac{1}{\bar{\mathrm{Y}}}\sum_{\mathrm{h}=1}^{\mathrm{L}} \mathrm{W_h}\bar{\mathrm{Y}}_{\mathrm{h}} \text{ and } e_1 = \frac{1}{\bar{\mathrm{X}}}\sum_{\mathrm{h}=1}^{\mathrm{L}} \mathrm{W_h}\bar{\mathrm{X}}_{\mathrm{h}}$$

such that

$E(e_0) = E(e_1) = 0$, and

$$E(e_0^2) = \frac{1}{\bar{\bar{Y}}^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{\mathrm{h}=1}^{l} W_{\mathrm{h}}\, S_{yh}^2$$

$$E(e_1^2) = \frac{1}{\bar{X}^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\, S_{xh}^2,$$

$$E(e_0 e_1) = \frac{1}{\bar{\bar{Y}}\bar{X}}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\, S_{yxh}$$

The proposed estimator $\hat{\bar{Y}}_{\mathrm{RK}}$ expressed in terms of e_i,s

$$\hat{\bar{Y}}_{RK} = \bar{Y}(1+e_0)\left\{k + (1-k)exp\left(\frac{-\bar{X}e_1}{2\bar{X}+\bar{X}e_1}\right)\right\}$$

$$\hat{\bar{Y}}_{RK} = \bar{Y}(1+e_0)\left\{k + (1-k)exp\left(\frac{-e_{1h}}{(2+e_{1h})}\right)\right\}$$

$$\hat{\bar{Y}}_{RK} = \bar{Y}(1+e_0)\left\{k + (1-k)exp\left(\frac{-e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right)\right\} \quad (2.2)$$

Expanding the right hand side of (2.2) and retaining terms up to the second power of *e*'s,

$$\hat{\bar{Y}}_{RK} = \bar{Y}(1+e_0)\left\{k + (1-k)exp\left(\frac{-e_1}{2} - \frac{e_1^2}{4}\right)\right\}$$

$$MSE\left(\hat{\bar{Y}}_{RK}\right) = \left\{E(e_0^2) + k^2E\left(\frac{e_1^2}{4}\right) - kE(e_0e_1)\right\}$$

$$MSE\left(\hat{\bar{Y}}_{RK}\right) = f\sum_{h=1}^{L} W_h\left(S_{yh}^2 + \frac{1}{4}k^2R_1^2S_{xh}^2 - kR_1S_{yxh}\right) \quad (2.3)$$

2.1. Optimality of κ

Obtain the optimum *k* to minimize $MSE\left(\hat{\bar{Y}}_{RK}\right)$. Differentiating $MSE\left(\hat{\bar{Y}}_{RK}\right)$ with respect to *k* and equating the derivative to zero, optimum value of *k* is given by

$$k^* = 2\sum_{h=1}^{L}\frac{W_hS_{yxh}}{W_hR_1S_{xh}^2} \quad (2.4)$$

Substituting the value of k^* in (2.3), we get the minimum value of $MSE\left(\hat{\bar{Y}}_{RK(min)}\right)$

$$MSE\left(\hat{\bar{Y}}_{RK(min)}\right) = f\sum_{h=1}^{L}\left[S_{yh}^2 + 2\left(\frac{W_hS_{yxh}^2}{W_hS_{xh}^2}\right)\right] \quad (2.5)$$

3. EFFICIENCY COMPARISON

In this section, the performance of the proposed estimator has been demonstrated over the existing estimators in the literature as follows:

From (1.2) and (2.5)

I. $\left[Var\left(\hat{\bar{Y}}_{ps}\right) - MSE\left(\hat{\bar{Y}}_{RK(min)}\right)\right] > 0$

$$\frac{1}{n^2}\sum_{h=1}^{L}(1-W_h)\,S_{yh}^2 - \sum_{h=1}^{L}W_h\left(\frac{2S_{yxh}^2}{S_{xh}^2}\right) > 0 \tag{3.1}$$

From (1.5) and (2.5)

II. $\left[MSE\left(\hat{\bar{Y}}_{R}\right) - MSE\left(\hat{\bar{Y}}_{RK(min)}\right)\right] > 0$

$$\sum_{h=1}^{L}W_h\left(R_1^2 - 2R_1S_{yxh}\right) - \sum_{h=1}^{L}W_h\left(\frac{2S_{yxh}^2}{S_{xh}^2}\right) > 0 \tag{3.2}$$

From (1.8) and (2.5)

III. $\left[MSE\left(\hat{\bar{Y}}_{\mathrm{RC}}\right) - \mathrm{MSE}\left(\hat{\bar{Y}}_{\mathrm{RK(min)}}\right)\right] > 0$

$$\sum_{h=1}^{L}W_h\left(\frac{1}{4}R_1^2S_{xh}^2 - R_1S_{yxh}\right) - \sum_{h=1}^{L}W_h\left(\frac{2S_{yxh}^2}{S_{xh}^2}\right) > 0 \tag{3.3}$$

From (1.11) and (2.5)

IV. $\left[MSE\left(\hat{\bar{Y}}_{PC}\right) - MSE\left(\hat{\bar{Y}}_{RK(min)}\right)\right] > 0$

$$\sum_{h=1}^{L}W_h\left(\frac{1}{4}R_1^2S_{xh}^2 + R_1S_{yxh}\right) - \sum_{h=1}^{L}W_h\left(\frac{2S_{yxh}^2}{S_{xh}^2}\right) > 0 \tag{3.4}$$

It is observed that $\hat{\bar{Y}}_{RK(min)}$ is always more efficient than the traditional estimators $\hat{\bar{Y}}_{ps}$, $\hat{\bar{Y}}_{R}$, $\hat{\bar{Y}}_{\mathrm{RC}}$ and $\hat{\bar{Y}}_{PC}$, because the conditions from (3.1) to (3.4) are always satisfied.

4. EMPIRICAL STUDY

A data set is considered to exhibits the performance of the suggested estimator; we use some real-life populations. Description of the populations is given below:

Population I [Source: Murthy (1967), p. 228]

y : Output,

x: Fixed capital

Constant	Stratum I	Stratum II
N_h	05	05
n_h	4	4
$\bar{Y}_h$	1925	3115.60
$\bar{X}_h$	214.40	333.80
S_{yh}	615.92	340.38
S_{xh}	74.87	66.35
S_{yxh}	39360.68	22356.50

Population II [Source: Murthy (1967), p. 228]

y : Output,

x: Fixed capital

Constant	Stratum I	Stratum II
N_h	05	05
n_h	2	2
$\bar{Y}_h$	1925	3115.60
$\bar{X}_h$	214.40	333.80
S_{yh}	615.92	340.38
S_{xh}	74.87	66.35
S_{yxh}	39360.68	22356.50

Population III

[Source: Japan Meteorological Society. Retrieved from the World Wide Web http://www.data.jma.go.jp/obd/stats/data/en/index.html**]**

Constant	Stratum I	Stratum II
N_h	10	10
n_h	4	4
$\bar{Y}_h$	1629.99	2035.96
$\bar{X}_h$	149.7	102.6
S_{yh}	32.305	103.26
S_{xh}	74.872	66.35
S_{yxh}	-1072.8	-655.25

Table 4.1 Mean square errors (MSEs) of the estimators for Population I:

Estimators	MSE
$V(\bar{y}_{ps})$	10059.07
$\widehat{\bar{Y}}_R$	2580.72
$\widehat{\bar{Y}}_{RC}$	15929.74
$\widehat{\bar{Y}}_{PC}$	1740.528
$\widehat{\bar{Y}}^*_{RK}$	**1432.67**

Table 4.2 PRE of Estimators for Population I:

Estimators	PRE
$V(\bar{y}_{ps})$	100
$\widehat{\bar{Y}}_R$	389.78
$\widehat{\bar{Y}}_{RC}$	63.15
$\widehat{\bar{Y}}_{PC}$	577.93
$\widehat{\bar{Y}}^*_{RK}$	**702.12**

Table 4.2, revealed the percent relative efficiencies (PRE) of estimators for population I. It is observed that the proposed exponential ratio type estimator $\bar{Y}_{RK}$ proved to be the best estimator in the sense of having highest percent relative efficiency than usual unbiased estimators$V(\bar{y}_{ps})$, $\widehat{\bar{Y}}_R$, $\widehat{\bar{Y}}_{RC}$, $\widehat{\bar{Y}}_{PC}$ for the population I.

Table 4.3 Mean square errors (MSEs) of the estimators for Population II:

Estimators	**MSE**
$V(\bar{y}_{ps})$	52617.18
$\widehat{\bar{Y}}_R$	15483.09
$\widehat{\bar{Y}}_{RC}$	95578.39
$\widehat{\bar{Y}}_{PC}$	10443.09
$\widehat{\bar{Y}}^*_{RK}$	**8595.62**

Table 4.4 PRE of Estimators for Population II:

Estimators	**PRE**
$V(\bar{y}_{ps})$	100
$\widehat{\bar{Y}}_R$	339.84
$\widehat{\bar{Y}}_{RC}$	55.05
$\widehat{\bar{Y}}_{PC}$	503.85
$\widehat{\bar{Y}}^*_{RK}$	**612.14**

Table 4.4, revealed the percent relative efficiencies (PRE) of estimators for population II. It is observed that the proposed exponential ratio type estimator $\bar{Y}_{RK}$ proved to be the best estimator in the sense of having highest percent relative

efficiency than usual unbiased estimators $V(\bar{y}_{ps})$, $\widehat{\bar{Y}}_R$, $\widehat{\bar{Y}}_{RC}$, $\widehat{\bar{Y}}_{PC}$ for the population II.

Table 4.5 Mean square errors (MSEs) of the estimators for Population III:

Estimators	**MSE**
$V(\bar{y}_{ps})$	530.48
$\widehat{\bar{Y}}_R$	2709.97
$\widehat{\bar{Y}}_{RC}$	112.85
$\widehat{\bar{Y}}_{PC}$	1304.72
$\widehat{\bar{Y}}^*_{RK}$	**109.90**

Table 4.6 PRE of Estimators for Population III:

Estimators	**PRE**
$V(\bar{y}_{ps})$	100
$\widehat{\bar{Y}}_R$	19.58
$\widehat{\bar{Y}}_{RC}$	470.06
$\widehat{\bar{Y}}_{PC}$	40.66
$\widehat{\bar{Y}}^*_{RK}$	**482.67**

Table 4.4, revealed the percent relative efficiencies (PRE) of estimators for population III. It is observed that the proposed exponential ratio type estimator $\bar{Y}_{RK}$ proved to be the best estimator in the sense of having highest percent relative efficiency than usual unbiased estimators $V(\bar{y}_{ps})$, $\widehat{\bar{Y}}_R$, $\widehat{\bar{Y}}_{RC}$, $\widehat{\bar{Y}}_{PC}$ for the population III.

5. SIMULATION STUDY

We did a simulation research in this section to explore the qualities of proposed estimators. We use a bivariate normal distribution to produce finite populations of size N=10000 in the simulation study. The samples were created from a bivariate normal distribution using the R program's mvrnorm function. In the simulation, we considered $\mu_x = 2, \mu_y = 4,$ we have computed mean square errors (MSEs) and percent relative efficiencies (PREs) of estimators with respect to $V(\bar{y}_{ps})$ and displayed in Table 5.1- Table 5.2. From the Tables we can say that proposed estimator $\widehat{\bar{Y}}^*_{RK}$ is the most efficient estimator than the existing estimators in literature for this simulation study.

Table 5.1 MSE and PRE Values of Estimators for population I

$\rho_{yx1} : \rho_{yx2}$	*MSE*	*PRE*
0.75:0.88	10059.07	100
	4249.29	236.72
	15095.45	66.64
	2574.82	390.67
	2551.59	**294.23**
0.80:0.93	10059.07	100
	3459.59	290.76
	15490.3	64.94
	7179.96	461.43
	2022.02	**497.48**
0.85:0.98	10059.07	100
	2669.88	376.76
	15885.16	63.32
	1785.11	563.50
	1492.46	**673.99**

Table 5.2 MSE and PRE Values of Estimators for population II

$\rho_{yx1} : \rho_{yx2}$	*MSE*	*PRE*
0.75:0.88	52617.17	100
	25494.61	206.39
	90572.64	58.09
	15448.85	340.59
	15309.35	**343.69**
0.80:0.93	52617.17	100
	20756.34	253.50
	92941.77	56.61
	13079.72	402.28
	12131.87	**431.71**
0.85:0.98	52617.17	100
	16018.07	318.49
	95310.91	55.206
	10710.59	491.26
	8954.38	**587.61**

6. CONCLUSION

In this article, an exponential ratio type estimator has been proposed under post stratification. The mathematical form of the estimator has been derived and its condition of efficiencies has been formulated with respect to some existing estimators from literature. For comparing the efficiencies of proposed estimator with some existing estimators, we utilized some real-life populations for estimating population mean. The results from these populations show that our proposed estimator performs efficiently as compared to existing estimators. We also observe that efficiency of proposed estimator increases when the correlation between study and auxiliary variable increases by simulation study. Therefore, it is recommended to use proposed estimator for estimating population mean under post stratification.

ACKNOWLEDGMENT

Authors are thankful to the referee for his valuable suggestions regarding the improvement of the paper.

REFERENCES

Agrawal, M. C., Panda, K. B. (1993). An efficient estimator in post stratification. *Metron* 51:179–188.

Bahl, S., Tuteja, R. K. (1991). Ratio and product type exponential estimator. *Inf. Optim. Sci.* 12:159–163.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley Eastern Limited.

Holt, D., Smith, T. M. F. (1979). Post-stratification. *J. R. Stat. Soc.* 142:33–46.

Ige, A. F., Tripathi, T. P. (1989). Estimation of population mean using post-stratification and auxiliary information. *Abacus* 18:265–276.

Jagers, P. (1986). Post stratification against bias in sampling. *Int. Statist. Rev.* 55:159–167.

Jagers, P., Oden, A., Trulsson, L. (1985). Post stratification and ratio estimation: Usages of auxiliary information in survey sampling and opinion polls. *Int. Statist. Rev.* 53:221–238.

Murthy, M. N. (1967). *Sampling Theory and Methods* (p. 228). Calcutta: Statistical Publishing Society.

Singh, H. P., Ruiz Espejo, M. (2003). *Improved post stratified estimation*. Bulletin of the International Statistical Institute, 54th session, contributed papers, vol. LX, Two Books, Book 2, pp. 341–342.

Singh, R., Mangat, N. S . (1996). *Elements of Survey Sampling* (Vol. , p. 219). London: Kluwer Academic.

Stephan, F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Stat.* 16:50–61.

Subramani, J. (2016). A new median based ratio estimator for estimation of the finite population mean, Statistics in Transition New Series, 17, 4:1-14.

Tailor R., Tailor R., and Chouhan S. (2017). Improved Ratio- and Product-Type Exponential Estimators for Population Mean in Case of Post-Stratification. *Communications in Statistics - Theory and Methods,* Vol 46, No. 21, 10387-10393.

Tailor R., Mehta P. (2019). A Ratio and Ratio Exponential Estimator for Finite Population Mean in Case of Post-Stratification. *Journal of Statistics Applications & Probability*. 8(3):241-246.

Khalid Ul Islam Rather
Division of Statistics and Computer Science, Faculty of Basic Sciences,
Main Campus SKUAST-J, Jammu, J&K, India, 180009.
Email: khalidstat34@gmail.com (corresponding author)

M. Iqbal Jeelani
Division of Statistics and Computer Science, Faculty of Basic Sciences,
Main Campus SKUAST-J, Jammu, J&K, India, 180009.

M. Younis Shah
Division of Statistics and Computer Science, Faculty of Basic Sciences,
Main Campus SKUAST-J, Jammu, J&K, India, 180009.

S. E. H. Rizvi
Division of Statistics and Computer Science, Faculty of Basic Sciences,
Main Campus SKUAST-J, Jammu, J&K, India, 180009.

Manish Sharma
Division of Statistics and Computer Science, Faculty of Basic Sciences,
Main Campus SKUAST-J, Jammu, J&K, India, 180009.