Volume 67, Issue 4, 2023

Journal of Scientific Research

of

The Banaras Hindu University

Improvement for Estimation of Population Mean in Post Stratification using Supplementary Variable.

M. Younis Shah[1]* and S.E.H. Rizvi

[*1]Division of Statistics and Computer Science, SKUAST- Jammu, syeedunis121@gmail.com

[2]Division of Statistics and Computer Science, SKUAST- Jammu, sehrizvistat@gmail.com

Abstract: **This investigation, considers the estimation for finite population mean utilizing knowledge from a supplementary variable in case of post stratification. We examine the suggested estimator's sample characteristics up to an approximation of order one. The ideal value of the constant in the suggested estimator has been found to have the lowest mean squared error (MSE). The suggested estimator is contrasted with alternative estimators. An empirical illustration verifies the theoretical results. We demonstrate the advantages of the suggested estimator over the competing estimators via an empirical example.**

Index Terms: **Post stratification, Auxiliary variable, Bias, Mean squared error, Efficiency**

I. INTRODUCTION

The application of stratified random sampling assures that the sizes and structure of sampling frames for every stratum are already defined. Whereas the total population size and the percentage of the unit that belongs to each stratum may be known in many existing system, it is possible that the sample frame for every stratum is neither available or would be costly and difficult to construct. We can't employ stratified random sampling in such types of situations. In order to resolve these difficulties, post stratification technique is applied, in which a sample of necessary size is first selected from the population employing simple random sampling, and it is then stratified using the stratification factor.

Initially, the post stratification idea was explained by Hansen *et al.* (1953). The classic Cochran (1940) ratio estimator was later investigated by Ige and Tripathi (1989) in the problem of post stratification. In the area of post stratification, contributions have been made by Jagers *et al.* (1985), Jagers (1986), Agrawal and Panda (1993), and Singh and Ruiz Espezo (2003). The characteristics of the post stratification product and ratio type exponential estimators by Bahl and Tuteja (1991), Tailor *et al*. (2017) were recently specified by Rather *et al.* (2022). Singh *et al.* (2020) developed an exponential type estimator for estimating population mean under simple random sampling without replacement and also utilize this estimator for missing data under varied imputation techniques and Kumari *et al*. (2023) presented some improved estimators of population mean using auxiliary variables in stratification.

II. REVIEW OF LITERATURE

Consider a population with a finite size of N that is split into L strata of sizes $N_1, N_2, \ldots, N_L$ such that $\sum_{h=1}^{L} N_h = N$. Assume that y is the study variable and that x is a supplementary variable. These variables should, respectively, be positively and negatively correlated with y. The study variable observation on the i^{th} unit of the h^{th} stratum will bey_{hi}, and the supplementary variable observation on the i^{th} unit of the h^{th} stratum will be x_{hi}. Population means for study variables y and x are represented by $\bar{Y}$ and $\bar{X}$, respectively, where h^{th} stratum means are represented by $\bar{Y}_h$ and $\bar{X}_h$ respectively. A sample of size n is drawn from the entire population using SRSWOR. It is indicated how much and which units belong to the h^{th} stratum after the SRS selection of the population. 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 presummated that n is so big that there is extremely little chance that n_h will be zero.

Sample mean is the most natural estimator and is unbiased for population means, defined as

$$\overline{Y}_{ps} = \sum_{h=1}^{L} w_h \overline{y}_h \tag{1}$$

Where $\frac{N_h}{N}$ is the weight of h^{th} stratum.

Using Stephen's (1945) findings, its variance is given by

$$V\left(\overline{Y}_{ps}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{yh}^2 + \frac{1}{n^2}\sum_{h=1}^{L}\left(1 - w_h\right) S_{yh}^2 \tag{2}$$

Ratio estimator described by Ige and Tripathi (1989) as

$$\overline{Y}_R = \overline{y}_{ps}\left(\frac{\overline{X}}{\overline{x}_{ps}}\right), \quad \text{Where} \quad \overline{X} = \sum_{h=1}^{L} w_h \overline{X}_h,$$

$$\overline{y}_{ps} = \sum_{h=1}^{L} w_h \overline{y}_h, \quad \text{and} \quad \overline{x}_{ps} = \sum_{h=1}^{L} w_h x_h,$$

Bias and MSE of $\overline{Y}_R$ is expressed as

$$Bias\left(\overline{Y}_R\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\overline{X}}\sum_{h=1}^{L} w_h\left(R_1 S_{xh}^2 - S_{yxh}\right) \tag{3}$$

$$MSE\left(\overline{Y}_R\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h\left(S_{yh}^2 + R_1^2 S_{xh}^2 - 2R_1 S_{yxh}\right) \tag{4}$$

Where $R_1 = \frac{\overline{Y}}{\overline{X}}$

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

$$\left(\overline{Y}_{RC}\right) = \overline{y}_{ps}\exp\left(\frac{\overline{x}_{ps} - \overline{X}}{\overline{x}_{ps} + \overline{X}}\right) \tag{5}$$

Bias and MSE of $\overline{Y}_{RC}$ is expressed as

$$Bias\left(\overline{Y}_{RC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\overline{X}}\sum_{h=1}^{L} w_h\left(\frac{3}{8}R_1 S_{xh}^2 + \frac{1}{2}S_{yxh}\right) \tag{6}$$

$$MSE\left(\overline{Y}_{Rc}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h\left(S_{yh}^2 + \frac{1}{4}R_1^2 S_{xh}^2 - R_1 S_{yxh}\right) \tag{7}$$

$$\left(\overline{Y}_{PC}\right) = \overline{y}_{ps}\exp\left(\frac{\overline{X} - \overline{x}_{ps}}{\overline{X} + \overline{x}_{ps}}\right) \tag{8}$$

Bias and MSE of $\overline{Y}_{PC}$ is expressed as

$$Bias\left(\overline{Y}_{PC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\overline{X}}\sum_{h=1}^{L} w_h\left(\frac{3}{8}R_1 S_{xh}^2 - \frac{1}{2}S_{yxh}\right) \tag{9}$$

(2.3)

$$MSE\left(\overline{Y}_{Pc}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h\left(S_{yh}^2 + \frac{1}{4}R_1^2 S_{xh}^2 + R_1 S_{yxh}\right) \tag{10}$$

Rather *et al*. (2022) described a new ratio type estimator as

$$\left(\overline{Y}_{RK}\right) = \overline{y}_{ps}\left[K + \left(1 - K\right)\exp\left(\frac{\overline{X} - \overline{x}_{ps}}{\overline{X} + \overline{x}_{ps}}\right)\right] \tag{11}$$

Bias and MSE of $\overline{Y}_{RK}$ is expressed as

$$Bias\left(\overline{Y}_{PC}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\overline{X}}\sum_{h=1}^{L} w_h\left(\frac{3}{8}R_1 S_{xh}^2 - \frac{1}{2}S_{yxh}\right) \tag{12}$$

$$MSE\left(\overline{Y}_{Pc}\right) = f\sum_{h=1}^{L} w_h\left(S_{yh}^2 + 2\left(\frac{w_h S_{yxh}^2}{w_h S_{xh}^2}\right)\right) \tag{13}$$

III. PROPOSED ESTIMATOR

As a result of Kumar *et al*. (2019) inspiration, we have developed a new ratio as,

$$\left(\overline{Y}_{YS}\right) = \overline{y}_h\left[\eta_h\left(2 - \frac{\overline{X}_h}{\overline{x}_h}\right) + \left(1 - \eta_h\right)\left(2 - \frac{\overline{x}_h}{\overline{X}_h}\right)\right] \tag{14}$$

Where η_h is an unknown constant whose value is to be estimated later.

$$\overline{Y}_{YS} = \overline{Y}_h\left(1 + \varepsilon_0\right)\left[\eta_h\left(2 - \frac{\overline{X}_h}{\overline{X}_h\left(1 + \varepsilon_1\right)}\right) + \left(1 - \eta_h\right)\left(2 - \frac{\overline{X}_h\left(1 + \varepsilon_1\right)}{\overline{X}_h}\right)\right] \tag{15}$$

We use the standard approximation given below for studying the properties of the proposed estimator as,

$$\varepsilon_0 = \frac{1}{\overline{Y}_h}\sum_{h=1}^{L} w_h \overline{y}_h \quad \text{, and} \quad \varepsilon_1 = \frac{1}{\overline{X}_h}\sum_{h=1}^{L} w_h \overline{x}_h \text{ ,}$$

Such that,

$$E(\varepsilon_0) = E(\varepsilon_1) = 0 \quad , \quad E(\varepsilon_0^2) = \frac{1}{\overline{Y}_h^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{yh}^2$$

$$E(\varepsilon_1^2) = \frac{1}{\overline{X}_h^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{xh}^2$$

$$E(\varepsilon_0 \varepsilon_1) = \frac{1}{\overline{X}_h^2 \overline{Y}_h^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{yxh}^2$$

Expanding eq. (15) in terms of errors, we have

$$\overline{Y}_{YS} - \overline{Y}_h = \overline{Y}_h \begin{bmatrix} \varepsilon_0 + (2\eta_h - 1)\varepsilon_1 + \\ (2\eta_h - 1)\varepsilon_0 \varepsilon_1 - \eta_h \varepsilon_1^2 \end{bmatrix} \qquad (16)$$

Taking expectation on both sides of eq. (16) and utilizing standard results of expectations, we have the bias of $\overline{Y}_{YS}$ as

$$Bias(\overline{Y}_{YS}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h \frac{1}{\overline{X}_h}\begin{bmatrix} (2\eta_h - 1)S_{yxh} \\ -\eta_h R_h S_{xh}^2 \end{bmatrix}$$

$$Bias(\overline{Y}_{YS}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h \frac{1}{\overline{X}_h}\begin{bmatrix} \eta_{1h} S_{yxh} \\ -\eta_h R_h S_{xh}^2 \end{bmatrix} \qquad (17)$$

Where $\quad \eta_{1h} = (2\eta_h - 1)$

From eq. (16), we have

$$\overline{Y}_{YS} - \overline{Y}_h = \overline{Y}_h \left[\varepsilon_0 + (2\eta_h - 1)\varepsilon_1\right]$$

$$\overline{Y}_{YS} - \overline{Y}_h = \overline{Y}_h \left[\varepsilon_0 + \eta_{1h}\varepsilon_1\right] \qquad (18)$$

Taking expectation on both sides and squaring eq. (18), we have

$$(\overline{Y}_{YS} - \overline{Y}_h)^2 = \overline{Y}_h^2\left[\varepsilon_0^2 + \eta_{1h}^2 \varepsilon_1^2 + 2\eta_{1h}\varepsilon_0 \varepsilon_1\right] \qquad (19)$$

We derive the MSE of developed estimator by taking the expectation of equation (19) as

$$MSE(\overline{Y}_{YS}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h \begin{bmatrix} S_{yh}^2 + \eta_{1h}^2 R_h^2 S_{xh}^2 \\ + 2\eta_{1h} R_h S_{yxh} \end{bmatrix} \qquad (20)$$

The MSE $\overline{Y}_{YS}$ is least for,

$$\frac{\partial MSE(\overline{Y}_{YS})}{\partial \eta_{1h}} = 0 \text{, gives}$$

$$\eta_{1h_{Opt}} = -\frac{\sum_{h=1}^{L} W_h S_{yxh}}{\sum_{h=1}^{L} W_h R_1 S_{xh}^2}$$

The reduced form of $MSE(\bar{Y}_{YS(min)})$ is obtained by Substituting the value of $\eta_{1h_{Opt}}$ in (20),

$$MSE(\overline{Y}_{YS(\min)}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{h} W_h \left[S_{yh}^2 - \left(\frac{S_{yxh}^2}{S_{xh}^2}\right)\right] \qquad (21)$$

IV. EFFICIENCY COMPARISON

From (2), (4), (7), (10), (13) and (21), it is determined that the developed estimator would be more precise than other if

I. $\quad [Var(\bar{Y}_{ps}) - MSE(\bar{Y}_{YS(min)})] > 0$

$$\frac{1}{n^2}\sum_{h=1}^{L}(1 - w_h)S_{yh}^2 - \sum_{h=1}^{L} w_h \left(\frac{S_{yxh}^2}{S_{xh}^2}\right) > 0 \qquad (22)$$

II $\quad [MSE(\bar{Y}_R) MSE(\bar{Y}_{YS(min)})] > 0$

$$\sum_{h=1}^{L} w_h (R_1^2 - 2R_1 S_{yxh}) - \sum_{h=1}^{L} w_h \left(\frac{S_{yxh}^2}{S_{xh}^2}\right) > 0 \qquad (23)$$

III. $\quad [MSE(\bar{Y}_{RC}) - MSE(\bar{Y}_{YS(min)})] > 0$

$$\sum_{h=1}^{L} w_h \left(\frac{1}{4}R_1^2 S_{xh}^2 - R_1 S_{yxh}\right) - \sum_{h=1}^{L} w_h \left(\frac{S_{yxh}^2}{S_{xh}^2}\right) > 0 \qquad (24)$$

IV. $\quad [MSE(\bar{Y}_{PC}) - MSE(\bar{Y}_{YS(min)})] > 0$

$$\sum_{h=1}^{L} w_h \left(\frac{1}{4}R_1^2 S_{xh}^2 + R_1 S_{yxh}\right) - \sum_{h=1}^{L} w_h \left(\frac{S_{yxh}^2}{S_{xh}^2}\right) > 0 \qquad (25)$$

V. $\quad [MSE(\bar{Y}_{RK}) - MSE(\bar{Y}_{YS(min)})] > 0$

$$\left[2\left(\frac{w_h S_{yxh}^2}{w_h S_{xh}^2}\right) - \sum_{h=1}^{L} w_h \left(\frac{S_{yxh}^2}{S_{xh}^2}\right)\right] > 0 \qquad (26)$$

Because the requirements from (22) to (26) are always met, it is seen that $\bar{Y}_{YS(min)}$ is always more efficient than the classical estimators $\bar{Y}_{ps}$, $\bar{Y}_R$, $\bar{Y}_{RC}$, $\bar{Y}_{PC}$ and $\bar{Y}_{RK}$

V. EMPIRICAL STUDY

We will take into account three natural population datasets to evaluate the merits of the developed estimator. Below is a description of the populations:

Dataset I [Source: Japan Meteorological Society. Webhttp://www.data.jma.go.jp/obd/stats/data/en/index.html**]**

Constants	Stratum I	Stratum II
N_h	10	10
n_h	04	04
$\bar{Y}_h$	142.80	102.60
$\bar{X}_h$	149.70	91.0
S_{yh}^2	37.08	158.76
S_{xh}^2	181.17	43.16
S_{yxh}	18.44	23.30

Dataset II- [Source: Johnston et al. (1972), p. 171]

Constants	Stratum I	Stratum II
N_h	05	05
n_h	02	02
$\bar{Y}_h$	45.6	58.6
$\bar{X}_h$	37.6	46.4
S_{yh}^2	21. 0	15.84
S_{xh}^2	5.04	10.24
S_{yxh}	4.04	3.44

Dataset III- [Source: National Horticulture Board]

Constants	Stratum I	Stratum II
N_h	10	10
n_h	04	04
$\bar{Y}_h$	1.70	3.67
$\bar{X}_h$	10.41	289.14
S_{yh}^2	0.50	1.41
S_{xh}^2	3.53	111.61
S_{yxh}	1.60	144.87

Table 5.1.MSEs and PREs of Estimators for Dataset I:

Estimators	MSE	PRE
$V(\bar{y}_{ps})$	20.196	100
$\bar{Y}_R$	34.390617	58.72532
$\bar{Y}_{RC}$	21.157902	95.45369
$\bar{Y}_{PC}$	29.668909	68.07126
$\bar{Y}_{RK}$	18.80736	107.3835
$\boldsymbol{\bar{Y}_{YS}}$	**17.254091**	**117.0505**

Table5.2.MSEs and PREs of Estimators for Dataset II:

Estimators	MSE	PRE
$V(\bar{y}_{ps})$	3.10837	100
$\bar{Y}_R$	3.802266	81.75059
$\bar{Y}_{RC}$	2.843463	109.3165
$\bar{Y}_{PC}$	3.560877	87.29239
$\bar{Y}_{RK}$	2.689753	115.5636
$\boldsymbol{\bar{Y}_{YS}}$	**2.54326**	**122.2201**

Table5.3.MSEs and PREs of Estimators for Dataset III:

Estimators	MSE	PRE
$V(\bar{y}_{ps})$	0.08524	100
$\bar{Y}_R$	0.037271	228.706
$\bar{Y}_{RC}$	0.023031	370.1045
$\bar{Y}_{PC}$	0.219963	38.75208
$\bar{Y}_{RK}$	0.01941	439.1633
$\boldsymbol{\bar{Y}_{YS}}$	**0.00328**	**2598.868**

The proposed and traditional estimators were taken into consideration over the population mean reported in tables 5.1 to 5.3 above. The tables also display the MSE and PRE. The PRE values for suggested estimator as compared to other conventional estimators reveal that the suggested estimator $\bar{Y}_{YS}$ is superior to the competing estimators with least MSE for datasets I to III.

VI. CONCLUSION

The population mean of the research variable can be estimated using auxiliary data by using a new ratio type estimator under post stratification. The suggested estimators, such as bias and means square error, are attained roughly to the degree of one. Under efficiency conditions that are inferred, the recommended estimator outperforms the existing estimators. Furthermore, using some real populations, an empirical research is conducted in support of the theoretical findings. According to empirical results, the recommended estimator is much superior to the current estimators by having a lower MSE and a higher PRE. As a result, we heartily advise using our suggested estimator for experimental surveys rather than any of the well-known estimators covered in this study.

ACKNOWLEDGMENT

The authors appreciate the referee's insightful comments on how to strengthen the manuscript.

REFERENCES

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

Bahl, S. and 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.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953) Sample Survey Methods and techniques

http://www.data.jma.gr.jb/obd/stats/data/en/index.html (Official website of Japan Meteorological society).

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

Johnston, J. (1972). Econometric Methods, 2nd ed., Tokyo: McGraw-Hill.

Kumari, A.,Upendra, K, and Singh, R. (2023). Improved estimators of population mean using auxiliary variables in post stratification. *Journal of Scientific Research of the Banaras Hindu University*, 67 (1).

R. Tailor and S. Chouhan. (2017). Improved Ratio- and Product-Type Exponential Estimators for Population Mean in Case of Post-Stratification. *Communications in Statistics – Theory and Methods.* 46 (21), 10387-10393.

R. Tailor and P. Mehta. (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.

Rather, K.U.I., Jeelani, M. I., Shah, M. Y., Rizvi, S.E.H. and Sharma, M. *Journal of applied mathematics statistics and informatics*, 18(1), 29–42, 2022.

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

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.

Singh, R., Mishra, P., Ahmed, A and Supriya, k. (2020). Exponential Type Estimator for Estimating Finite Population Mean. *International Journal of Computational and Theoretical Statistics*, 7 (1).