

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(12): 222-226
© 2025 Stats & Maths
www.mathsjournal.com
Received: 25-10-2025
Accepted: 28-11-2025

Housila P Singh
School of Studies in Statistics,
Vikram University, Ujjain,
Madhya Pradesh, India

Rajesh Tailor
School of Studies in Statistics,
Vikram University, Ujjain,
Madhya Pradesh, India

Bindu Chaudhry
School of Studies in Statistics,
Vikram University, Ujjain,
Madhya Pradesh, India

Corresponding Author:
Housila P Singh
School of Studies in Statistics,
Vikram University, Ujjain,
Madhya Pradesh, India

Enhanced estimation of population mean using post stratification sampling

Housila P Singh, Rajesh Tailor and Bindu Chaudhry

DOI: <https://www.doi.org/10.22271/math.2025.v10.i12c.2226>

Abstract

In the present paper, an improved estimator for the population mean is proposed using post-stratification sampling. The proposed estimator is examined up to the first order of approximation in terms of bias, mean square errors (MSE) and minimum MSE. The efficiency conditions are obtained by comparing the MSE of the proposed and existing estimators. An empirical investigation is also carried out and the observed positive findings clearly demonstrate the higher accuracy of the proposed estimate over the existing ones.

Keywords: Bias, mean squared error, post stratification, auxiliary information.

1. Introduction

Stratification is one of the most commonly used techniques in survey sampling. The application of stratified random sampling assumes that the sizes of strata and the sampling frame for each stratum are known beforehand. However, in many practical scenarios, while the overall population size and the proportion of units within different strata are available, the sampling frame for each stratum may be unavailable or difficult to compile due to cost and time constraints. In such cases, stratified random sampling cannot be directly implemented. To address this challenge, the post-stratification technique is used. In this approach, a sample of size n is first drawn using simple random sampling without replacement. This sample is then categorized into different strata and analyzed as a stratified sample. Several researchers have significantly contributed to the development of post-stratification methods, including Holt and Smith (1979) [3], Jagers *et al.* (1985) [6], Jagers (1986) [5], Ige and Tripathi (1989) [4], Agrawal and Panda (1993) [1], and Singh and Ruiz Espezo (2003) [9], Tailor *et al.* (2017) [10]. In this paper, we use the data from the auxiliary variable, based on post-stratification sampling, to propose an enhanced estimator for the population mean estimation of the study variable.

This paper objective is to:

- An enhanced estimator for the population mean under post-stratified sampling is proposed to compete with existing estimators.
- Comparing the proposed estimators' theoretical efficiency to that of the existing estimators
- Use real-world population to illustrate the proposed estimators.

The parts that follow provide further design for the current paper. The methodology and notations utilized are explained in Section 2. Section 3 discusses some existing estimators. In Section 4, we have proposed an enhanced estimator for population mean under post-stratified sampling, along with the bias, MSE and the minimum MSE expressions. The efficiency conditions are obtained in Section 5. Section 6 presents an empirical illustration of the efficiency of the proposed estimator using examples based on real populations. Finally, Section 7 concludes the paper.

2. Methodology and notations

Let $U = (U_1, U_2, \dots, U_N)$ be a finite population of size N . Using simple random sampling without replacement, a sample of size n is drawn from population U .

After selecting the sample, it is observed that which units belong to 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 probability of n_h being zero is very small.

Let x_{hi} be the observation on i^{th} unit that fall in h^{th} stratum for auxiliary variate x and y_{hi} be the observation on i^{th} unit that fall in h^{th} stratum for study variate y , then

- $\bar{X} = \sum_{h=1}^L W_h \bar{X}_h$: Population mean of auxiliary variate x ,
- $\bar{Y} = \sum_{h=1}^L W_h \bar{Y}_h$: Population mean of study variate y ,
- $\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$: h^{th} stratum mean for auxiliary variate x ,
- $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$: h^{th} stratum mean for study variate y

3. Existing estimators

In case of post-stratification, usual unbiased estimator of population mean $\bar{Y}$ is defined as

$$\bar{y}_{PS} = \sum_{h=1}^L W_h \bar{y}_h \quad (3.1)$$

Where $W_h = \frac{N_h}{N}$ is the weight of the h^{th} stratum and $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ is sample mean of n_h sample units that fall in the h^{th} stratum.

Using the results from Stephen (1945), the variance of $\bar{y}_{PS}$ to the first degree of approximation is obtained as

$$\text{Var}(\bar{y}_{PS}) = \theta \sum_{h=1}^L W_h S_{yh}^2 \quad (3.2)$$

Where $S_{yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$, $S_{xh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$ and $\theta = \left(\frac{1}{n} - \frac{1}{N}\right)$.

Ige and Tripathi (1989) [4] defined classical ratio type and product type estimators for estimating the population mean $\bar{Y}$ in case of post stratification as

$$\hat{\bar{Y}}_{PS}^R = \bar{y}_{PS} \left(\frac{\bar{X}}{\bar{x}_{PS}} \right) \quad (3.3)$$

And

$$\hat{\bar{Y}}_{PS}^P = \bar{y}_{PS} \left(\frac{\bar{x}_{PS}}{\bar{X}} \right) \quad (3.4)$$

Where $\bar{x}_{PS} = \sum_{h=1}^L W_h \bar{x}_h$ is the unbiased estimate of population mean in case of post stratification, $\bar{X} = \sum_{h=1}^L W_h \bar{X}_h$ is the known population mean of the auxiliary variate x and $\bar{x}_h$ is the mean of the sample of size n_h that fall in the h^{th} stratum.

To the first degree of approximation MSE of estimators $\hat{\bar{Y}}_{PS}^R$ and $\hat{\bar{Y}}_{PS}^P$ are respectively given by

$$\text{MSE}(\hat{\bar{Y}}_{PS}^R) = \theta \sum_{h=1}^L W_h (S_{yh}^2 + R^2 S_{xh}^2 - 2R S_{yxh}) \quad (3.5)$$

$$\text{MSE}(\hat{\bar{Y}}_{PS}^P) = \theta \sum_{h=1}^L W_h (S_{yh}^2 + R^2 S_{xh}^2 + 2R S_{yxh}) \quad (3.6)$$

Where $R = \frac{\bar{Y}}{\bar{X}}$

Tailor *et al* (2017) [10] suggested the following combined ratio-type and product type exponential estimators for population mean of $\bar{Y}$ in post-stratified sampling as

$$\hat{\bar{Y}}_{PS}^{RE} = \bar{y}_{PS} \exp \left[\frac{\bar{X} - \bar{x}_{PS}}{\bar{X} + \bar{x}_{PS}} \right] \quad (3.7)$$

$$\hat{\bar{Y}}_{PS}^{PE} = \bar{y}_{PS} \exp \left[\frac{\bar{x}_{PS} - \bar{X}}{\bar{x}_{PS} + \bar{X}} \right] \quad (3.8)$$

To the first degree of approximation, the MSEs of $\hat{\bar{Y}}_{PS}^{RE}$ and $\hat{\bar{Y}}_{PS}^{PE}$ are respectively given by

$$\text{MSE}(\hat{\bar{Y}}_{PS}^{RE}) = \theta \sum_{h=1}^L W_h \left(S_{yh}^2 + \frac{1}{4} R^2 S_{xh}^2 - R S_{yxh} \right) \quad (3.9)$$

$$\text{MSE}(\hat{\bar{Y}}_{PS}^{PE}) = \theta \sum_{h=1}^L W_h \left(S_{yh}^2 + \frac{1}{4} R^2 S_{xh}^2 + R S_{yxh} \right) \quad (3.10)$$

4. Proposed Estimator

Taking clue from Kumar and Siddiqui (2024) we suggest a class of estimator for population mean $\bar{Y}$ of y under post stratified sampling as

$$\bar{y}_{PS}^* = k_1 \bar{y}_{ps} \left(\frac{\bar{x}_{ps}}{\bar{X}} \right)^{k_2} \exp \left(\frac{\eta(\bar{X} - \bar{x}_{ps})}{\eta(\bar{X} - \bar{x}_{ps}) + 2\lambda} \right) \quad (4.1)$$

Where $k_j, j = 1, 2$ are suitable chosen constants and η and λ are either real amount or parametric values associated with auxiliary variable/ study variable

To obtain the bias and MSE of the proposed class of estimator $\bar{y}_{PS}^*$ we write

$$\bar{y}_{ps} = \bar{Y}(1 + e_0), \bar{x}_{ps} = \bar{X}(1 + e_1)$$

And to the first degree of approximation

$$E(e_0^2) = \frac{\theta}{\bar{Y}^2} \sum_{h=1}^L W_h S_{yh}^2,$$

$$E(e_1^2) = \frac{\theta}{\bar{X}^2} \sum_{h=1}^L W_h S_{xh}^2,$$

$$E(e_0 e_1) = \frac{\theta}{\bar{Y}\bar{X}} \sum_{h=1}^L W_h S_{yxh}$$

Now expressing $\bar{y}_{PS}^*$ in terms of e_0 and e_1 we have

$$\begin{aligned} \bar{y}_{PS}^* &= k_1 \bar{Y}(1 + e_0)(1 + e_1)^{k_2} \exp \left(\frac{-\eta \bar{X} e_1}{2(\eta \bar{X} + \lambda) + \eta \bar{X} e_1} \right) \\ &= k_1 \bar{Y}(1 + e_0)(1 + e_1)^{k_2} \exp \left\{ \frac{-\tau e_1}{2} \left(1 + \frac{\tau}{2} e_1 \right)^{-1} \right\} \end{aligned} \quad (4.2)$$

$$\text{Where } \tau = \frac{\eta \bar{X}}{\eta \bar{X} + \lambda}$$

Expanding the right hand side of (4.2), multiplying out and neglecting terms of e_i 's having power greater than two we have

$$\bar{y}_{PS}^* = \bar{Y} k_1 \left[1 + \left(k_2 - \frac{\tau}{2} \right) (e_1 + e_0 e_1) + \left(\frac{k_2(k_2 - 1)}{2} - \frac{k_2}{2} \tau + \frac{3}{8} \tau^2 \right) e_1^2 \right]$$

Or

$$(\bar{y}_{PS}^* - \bar{Y}) = \bar{Y} \left[k_1 \left\{ 1 + \left(k_2 - \frac{\tau}{2} \right) (e_1 + e_0 e_1) + \left(\frac{k_2(k_2 - 1)}{2} - \frac{k_2}{2} \tau + \frac{3}{8} \tau^2 \right) e_1^2 \right\} - 1 \right] \quad (4.3)$$

Taking expectation of both sides of (4.3) we get the bias of t as

$$B(\bar{y}_{PS}^*) = \bar{Y} \left[k_1 \left\{ 1 + \left(k_2 - \frac{\tau}{2} \right) \frac{\theta}{\bar{Y}\bar{X}} \sum_{h=1}^L W_h S_{yxh} + \left(\frac{k_2(k_2 - 1)}{2} - \frac{k_2}{2} \tau + \frac{3}{8} \tau^2 \right) \frac{\theta}{\bar{X}^2} \sum_{h=1}^L W_h S_{xh}^2 \right\} - 1 \right] \quad (4.4)$$

Squaring both sides of (4.3) and neglecting terms of e_i 's having power greater than two we have

$$\begin{aligned} (\bar{y}_{PS}^* - \bar{Y})^2 &= \bar{Y}^2 \left[1 + k_1^2 \left\{ 1 + 2e_0 + 2 \left(k_2 - \frac{\tau}{2} \right) e_1 + e_0^2 + 4 \left(k_2 - \frac{\tau}{2} \right) e_0 e_1 + (k_2(2k_2 - 1) - 2k_2 \tau + \tau^2) e_1^2 \right\} - 2k_1 \left\{ 1 + e_0 + \right. \right. \\ &\quad \left. \left(k_2 - \frac{\tau}{2} \right) (e_1 + e_0 e_1) + \left(\frac{k_2(k_2 - 1)}{2} - \frac{k_2}{2} \tau + \frac{3}{8} \tau^2 \right) e_1^2 \right\} \right] \end{aligned} \quad (4.5)$$

Taking expectation of both sides of (4.5) we get the MSE of $\bar{y}_{PS}^*$ as

$$MSE(\bar{y}_{PS}^*) = \bar{Y}^2 [1 + k_1^2 B_1 - 2k_1 B_2] \quad (4.6)$$

Where,

$$B_1 = \left[1 + \frac{\theta}{R^2 \bar{X}^2} \left\{ \sum_{h=1}^L W_h S_{yh}^2 + 4R \left(k_2 - \frac{\tau}{2} \right) \sum_{h=1}^L W_h S_{yxh} + (k_2(2k_2 - 1) - 2k_2 \tau + \tau^2) R^2 \sum_{h=1}^L W_h S_{xh}^2 \right\} \right],$$

$$B_2 = \left[1 + \frac{\theta}{R \bar{X}^2} \left\{ \left(k_2 - \frac{\tau}{2} \right) \sum_{h=1}^L W_h S_{yxh} + \left(\frac{k_2(k_2 - 1)}{2} - \frac{k_2}{2} \tau + \frac{3}{8} \tau^2 \right) R \sum_{h=1}^L W_h S_{xh}^2 \right\} \right],$$

Minimizing the $MSE(\bar{y}_{PS}^*)$ against k_1 provides the optimum value of k_1 as

$$k_{1(opt)} = \frac{B_2}{B_1} \quad (4.7)$$

Using the value of $k_{1(opt)}$ in the $MSE(\bar{y}_{PS}^*)$ we get the minimum MSE of $\bar{y}_{PS}^*$ as

$$MSE(\bar{y}_{PS}^*)_{min} = \bar{Y}^2 \left(1 - \frac{B_2^2}{B_1}\right) \quad (4.8)$$

Putting $k_1 = 1$ in (4.6) we have

$$MSE(\bar{y}_{PS}) = \theta \left[\sum_{h=1}^L W_h S_{yh}^2 + 2R \left(k_2 - \frac{\tau}{2}\right) \sum_{h=1}^L W_h S_{yxh} + \left(k_2 - \frac{\tau}{2}\right)^2 R^2 \sum_{h=1}^L W_h S_{xh}^2 \right] \quad (4.9)$$

Minimizing the MSE of $\bar{y}_{PS}^*$ against k_2 we get the optimum value of k_2 as

$$k_{2(opt)} = \left(\frac{\tau}{2} - \frac{\sum_{h=1}^L W_h S_{yxh}}{R \sum_{h=1}^L W_h S_{xh}^2} \right) \quad (4.10)$$

Substitution of (4.10) in (4.9) yield the minimum MSE of $\bar{y}_{PS}^*$ (at $k_1 = 1$) as

$$MSE(\bar{y}_{PS}^*)_{min} = \theta \left(\sum_{h=1}^L W_h S_{yh}^2 - \frac{\sum_{h=1}^L W_h S_{yxh}}{\sum_{h=1}^L W_h S_{xh}^2} \right) \quad (4.11)$$

Which equals to the minimum MSE of the difference estimator $\bar{y}_d = \bar{y}_{PS} + d(\bar{X} - \bar{x}_{PS})$, d being a constant

5. Efficiency conditions

By comparing the $MSE(\bar{y}_{PS}^*)_{min}$ to the variance/MSE/minimum MSE of the estimators we get the conditions under which suggested estimator $\bar{y}_{PS}^*$ is better than existing estimators.

$$MSE(\bar{y}_{PS}^*)_{min} < Var(\bar{y}_{PS}) = \frac{B_2^2}{B_1} > 1 - \frac{\theta}{\bar{Y}^2} \left[\sum_{h=1}^L W_h S_{yh}^2 \right] \quad (5.1)$$

$$MSE(\bar{y}_{PS}^*)_{min} < MSE(\hat{\bar{y}}_{PS}^R) = \frac{B_2^2}{B_1} > 1 - \frac{\theta}{\bar{Y}^2} \left[\sum_{h=1}^L W_h (S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{yxh}) \right] \quad (5.2)$$

$$MSE(\bar{y}_{PS}^*)_{min} < MSE(\hat{\bar{y}}_{PS}^P) = \frac{B_2^2}{B_1} > 1 - \frac{\theta}{\bar{Y}^2} \left[\sum_{h=1}^L W_h (S_{yh}^2 + R^2 S_{xh}^2 + 2RS_{yxh}) \right] \quad (5.3)$$

$$MSE(\bar{y}_{PS}^*)_{min} < MSE(\hat{\bar{y}}_{PS}^{RE}) = \frac{B_2^2}{B_1} > 1 - \frac{\theta}{\bar{Y}^2} \left[\sum_{h=1}^L W_h \left(S_{yh}^2 + \frac{1}{4} R^2 S_{xh}^2 - RS_{yxh} \right) \right] \quad (5.4)$$

$$MSE(\bar{y}_{PS}^*)_{min} < MSE(\hat{\bar{y}}_{PS}^{PE}) = \frac{B_2^2}{B_1} > 1 - \frac{1}{\bar{Y}^2} \left[\sum_{h=1}^L W_h \left(S_{yh}^2 + \frac{1}{4} R^2 S_{xh}^2 + RS_{yxh} \right) \right] \quad (5.5)$$

From (5.1) to (5.5) it follows that the proposed estimator $\bar{y}_{PS}^*$ is more efficient than the estimators $\hat{\bar{y}}_{PS}^R$, $\hat{\bar{y}}_{PS}^P$, $\hat{\bar{y}}_{PS}^{RE}$ and $\hat{\bar{y}}_{PS}^{PE}$ as long as the conditions (5.1), (5.2), (5.3), (5.4) and (5.5) are satisfied respectively.

6. Numerical study

To strengthen the theoretical establishment, we have conducted an empirical investigation on a real population. The descriptive statistics of this population are presented in Table 1. The optimum values of η and λ are computed and the ranges of η (for fixed values of λ) and λ (for fixed values of η) are also computed and the results are presented in Tables 2, 3 and 4, respectively. In these tables, for different fixed values of η (such as = 0.25, 0.5, 1) and varying values of λ (such as $\lambda = -1, -2, -3, -4, -5$) are considered to evaluate the performance of the proposed estimator $\bar{y}_{PS}^*$. It is observed that under these combinations, the proposed estimator $\bar{y}_{PS}^*$ performs more efficiently than the existing estimators.

Population: Onyeka (2012, 65-78) ^[11] Y = cumulative grade point average (during the 2008/2009 academic session) X = performance in a general Statistics pretest examination (during the 2008/2009 academic session)

Table 1: Data Statistics

Population	Males = Stratum 1	Female = Stratum 2
N= 96	N ₁ = 72	N ₂ = 24
n =20	n ₁ = 8	n ₂ = 12
$\bar{X} = 68.13$	$\bar{X}_1 = 68.11$	$\bar{X}_2 = 68.17$
$\bar{Y} = 2.44$	$\bar{Y}_1 = 2.44$	$\bar{Y}_2 = 2.46$
$S_x = 7.03$	$S_{x1} = 7.28$	$S_{x2} = 6.36$
$S_x^2 = 49.37$	$S_{x1}^2 = 52.97$	$S_{x2}^2 = 40.41$
$S_y = 0.57$	$S_{y1} = 0.33$	$S_{y2} = 0.35$
$S_y^2 = 0.33$	$S_{y1}^2 = 3.26$	$S_{y12}^2 = 3.43$
$S_{yx} = 3.26$	$S_{yx1} = 3.43$	$S_{yx2} = 2.75$

Table 2: PRE of different estimators with respect to usual unbiased estimator $\bar{y}_{PS}$

η	λ	PRE ($\bar{y}_{PS}^*, \bar{y}_{PS}$)	Estimator	PRE ($\bar{y}_{PS}$)
0.25	-1	335.25	$\hat{\bar{Y}}_{PS}^R$	209.10
0.25	-2	407.41	$\hat{\bar{Y}}_{PS}^P$	52.21
0.25	-3	545.62	$\hat{\bar{Y}}_{PS}^{RE}$	144.94
0.25	-4	916.14	$\hat{\bar{Y}}_{PS}^{PE}$	71.002
0.25	-5	5139.10		

Table 3: PRE of different estimators with respect to usual unbiased estimator $\bar{y}_{PS}$

η	λ	PRE ($\bar{y}_{PS}^*$)	Estimator	PRE ($\bar{y}_{PS}$)
0.5	-1	310.73	$\hat{\bar{Y}}_{PS}^R$	209.10
0.5	-2	335.25	$\hat{\bar{Y}}_{PS}^P$	52.21
0.5	-3	366.43	$\hat{\bar{Y}}_{PS}^{RE}$	144.94
0.5	-4	407.41	$\hat{\bar{Y}}_{PS}^{PE}$	71.002
0.5	-5	463.66		

Table 4: PRE of different estimators with respect to usual unbiased estimator $\bar{y}_{PS}$

η	λ	PRE ($\bar{y}_{PS}^*$)	Estimator	PRE ($\bar{y}_{PS}$)
1	-1	300.35	$\hat{\bar{Y}}_{PS}^R$	209.10
1	-2	310.73	$\hat{\bar{Y}}_{PS}^P$	52.21
1	-3	322.30	$\hat{\bar{Y}}_{PS}^{RE}$	144.94
1	-4	335.25	$\hat{\bar{Y}}_{PS}^{PE}$	71.002
1	-5	349.84		

7. Conclusions

This paper proposes an enhanced estimator for the population mean using post stratified sampling. The bias, mean square error, and minimum mean square error of the proposed estimator are obtained mathematically to the approximation of first order. By comparing the minimum MSE of the proposed estimator with the MSE of the existing estimators, the efficiency conditions are determined. A numerical illustration using real population data along with empirical results, demonstrate that the proposed estimator outperforms conventional estimators.

References

1. Agrawal MC, Panda KB. An efficient estimator in post-stratification. *Metron*. 1993;5(3-4):179-188.
2. Bahl S, Tuteja RK. Ratio and product type exponential estimator. *J Inf Optim Sci*. 1991;12(1):159-163.
3. Holt D, Smith TMF. Post-stratification. *J R Stat Soc Ser A*. 1979;142:33-46.
4. Ige AF, Tripathi TP. Estimation of population mean using post-stratification and auxiliary information. *Abacus*. 1989;18(2):265-276.
5. Jagers P. Post-stratification against bias in sampling. *Int Stat Rev*. 1986;54(2):159-167.
6. Jagers P, Oden A, Trulsson L. Post-stratification and ratio estimation: usage of auxiliary information in survey sampling and opinion polls. *Int Stat Rev*. 1985;53(3):221-238.
7. Kumar A, Siddiqui AS. Enhanced estimation of population mean using simple random sampling. *Res Stat*. 2024;2(1):1-9.
8. Stephan F. The expected value and variance of the reciprocal and other negative powers of a positive Bernoulli variate. *Ann Math Stat*. 1945;16:50-61.
9. Singh HP, Ruiz Espejo M. Improved post-stratified estimation. In: *Proceedings of the 54th Session of the International Statistical Institute*; vol LX, Book 2; 2003, p. 341-342.
10. Tailor R, Tailor R, Chouhan S. Improved ratio- and product-type exponential estimators for population mean in case of post-stratification. *Commun Stat Theory Methods*; 2017, p. 1-7.
11. Onyeka AC. Estimation of population mean in post-stratified sampling using known value of some population parameters. *Stat Transit*. 2012;13(1):65-78.