# LEVERAGING AUXILIARY VARIABLES: ADVANCING MEAN ESTIMATION THROUGH CONDITIONAL AND UNCONDITIONAL POST-STRATIFICATION

G.R.V. Triveni[1] and Faizan Danish[2]

•

[1,2]Department of Mathematics, School of Advanced Sciences, VIT-AP University, Inavolu, Beside AP Secretariat, Amaravati AP-522237, India

[1]trivenigullinkala@gmail.com ,[2]danishstat@gmail.com

**Abstract**

*This article presents a novel class of estimators designed for post-stratification to estimate the mean of a study variable using information from auxiliary variables. Through a rigorous examination of bias and Mean Square Error (MSE), we demonstrate the potential to improve estimation accuracy up to the first order of approximation. We also thoroughly explore both Conditional and Unconditional post-stratification properties, enhancing our understanding of the estimator's performance. To assess the effectiveness of our proposed estimator, we conduct a comprehensive numerical illustration. The results affirm its superiority over existing estimators in both Conditional and Unconditional Post-stratification scenarios, exhibiting the highest Percentage Relative Efficiency. Additionally, graphical analysis reveals that Conditional post- stratification outperforms Unconditional post-stratification. These findings underscore the significant practical value of our proposed estimator in enhancing the accuracy of mean estimation in post-stratification studies. By accurately estimating population parameters, our novel class of estimators contributes to more informed decision-making in various fields of study. The utilization of auxiliary variables allows for better utilization of available information and leads to more reliable and robust conclusions. Overall, the novel class of estimators introduced in this article represents a valuable contribution to the field of post-stratification. As researchers continue to explore and apply these estimators, they have the potential to revolutionize data analysis methods, becoming indispensable tools for survey and research design. The improvements in estimation accuracy brought about by these estimators are particularly crucial in situations where reliable data is scarce or challenging to obtain, making them invaluable for decision-makers and researchers alike. With the increased accuracy and efficiency of our proposed estimators, they provide a pathway for better resource allocation, cost-effective decision-making, and improved policy formulation. Policymakers and researchers can confidently rely on these estimators to produce more accurate results and achieve better outcomes in various domains. In conclusion, the novel class of estimators for post-stratification presented in this article opens up new avenues for advancing statistical estimation methods. The fusion of auxiliary variables with traditional post-stratification techniques represents a powerful approach to enhance estimation accuracy. Embracing and incorporating these estimators into research practices will undoubtedly bring us closer to making data-driven decisions that have a meaningful impact on society.*

**Keywords:** Conditional post stratification, Unconditional post stratification, Mean square error.

## I. Introduction

A common statistical method used in research studies to increase the precision and representativeness of survey data is post-stratification. It entails breaking down the sample population into discrete subgroups according to certain traits or factors, such age, gender, income level, or geography. Researchers can reduce potential biases and improve the generalizability of the results by stratifying the population to make sure that each subgroup is appropriately represented in the sample. After stratifying the sample, researchers can determine the population parameters by giving each subgroup the proper weights depending on its relative size. This method enables researchers to account for differences and generate more accurate and trustworthy estimates when the sample does not properly reflect the makeup of the target population. By taking into account population changes and enhancing the precision of inferential statistics, post-stratification is a useful tool that improves the validity and robustness of study findings.

Auxiliary variable information is used in various types of literature to estimate population mean or variance. The significance of post-stratification and the proper framework for statistical inference were covered in [1]. By utilizing auxiliary data and empirical research, [2] proposed estimators for population mean and shown that the proposed estimator outperformed the alternative. For judgement post-stratification, [3] offered an alternative estimate that regularly beats the usual non-parametric mean estimator, and they noted a decrease in Mean Square Error (MSE) in their proposed estimator compared to the standard estimator. [4] proposed a class of estimators and compared them with a few already in use. They came to the conclusion that their suggested class of estimators performed well based on a numerical investigation. In post-stratified sampling, [5] created a new family of combined estimators of the population mean, and the outcomes are empirically demonstrated. Exponential estimators are later proposed in post-stratification by [6], and its bias and MSE equations are obtained. The theoretical findings are further supported by a numerical analysis. [7] suggested an estimator of the population mean utilizing information from an auxiliary variable and demonstrated the superiority of the proposed estimator over others through comparison analysis. Additionally, [8] constructed a generalized class of estimators for population variance and demonstrated the effectiveness of the suggested estimator through a numerical investigation. Through empirical research, [9] demonstrated the superiority of the suggested estimator and developed a new family of exponential estimators. The ratio and product type exponential estimators were improved in the case of post-stratification by [10]. They demonstrated that the suggested estimator performed more effectively after stratification than unbiased, ratio, and product estimators. Additionally, [11] raises the issue of estimating a population proportion in a decision following stratification. They conducted Monte Carlo simulation research to evaluate the performance of proportion estimators. According to [12], a family of Ratio estimators and the formulations for bias and an MSE are constructed in the case of the non-response issue. It is demonstrated by numerical analysis that the suggested estimator has reduced MSE values. A novel class of estimators was recently developed [13], and by numerical analysis, under ideal circumstances, the proposed class of estimators outperformed the previously taken into consideration existing estimators.[14] suggested some post-stratification enhanced estimators. They demonstrated the effectiveness of the proposed estimator using two real data sets. In this paper, a class of estimators for estimating population mean under post-stratification has been developed.

## II. Terminology

Consider a finite population $\chi = \{1,2,\dots N\}$ of size N is stratified into K strata with $h^{th}$ stratum each of which has $N_h$ units such that $\sum_{h=1}^{K} N_h = N$. With the use of Simple random sampling without replacement, a sample with the dimension $n_h$ is taken from $h^{th}$ stratum. Let d serve as study(dependent) variable and i serve as auxiliary (independent) variable. We have used the following notations:

- $\bar{D} = \sum_{h=1}^{K} W_h \bar{D}_h$ is the population mean of study variable
- $\bar{I} = \sum_{h=1}^{K} W_h \bar{I}_h$ is the population mean of auxiliary variable
- $\bar{D}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} d_{hi}$ and $\bar{I}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} i_{hi}$ are the stratum means of study and auxiliary variables
- $S_{dh}^2 = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}(d_{hi} - \bar{D}_h)^2$ is the variance of study variable at $h^{th}$ stratum
- $S_{ih}^2 = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}(i_{hi} - \bar{I}_h)^2$ is the variance of auxiliary variable at $h^{th}$ stratum
- $S_{dih} = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}(d_{hi} - \bar{D}_h)((i_{hi} - \bar{I}_h)$ is the covariance at $h^{th}$ stratum
- $C_{dh}^2 = \frac{1}{\bar{D}^2(N_h-1)}\sum_{i=1}^{N_h}(d_{hi} - \bar{D}_h)^2$ be the square of coefficient of variation of d

- $C_{ih}^2 = \frac{1}{\bar{I}^2(N_h-1)}\sum_{i=1}^{N_h}(i_{hi}-\bar{I}_h)^2$ be the square of coefficient of variation of i
- $\rho_{dih} = \frac{\frac{1}{N_h-1}\sum_{i=1}^{N_h}(d_{hi}-\bar{D}_h)(i_{hi}-\bar{I}_h)}{(\bar{D}_h * C_{dh})(\bar{I}_h * I_{dh})}$ be the correlation coefficient of d and i.
- $W_h = \frac{N_h}{N}$ represents stratum weight.

I. Properties of Estimators in Unconditional Post-Stratification

To derive bias and mean squared error (MSE), we write
$\bar{d}_h = \bar{D}_h(1+e_{0h})$, $\bar{i}_h = \bar{I}_h(1+e_{1h})$,
$e_0 = \frac{\sum_{h=1}^{K} W_h\bar{D}_h e_{0h}}{\bar{D}}$ and $e_1 = \frac{\sum_{h=1}^{K} W_h\bar{I}_h e_{1h}}{\bar{I}}$

Where

$e_{0h} = \frac{\bar{d}_h-\bar{D}_h}{\bar{D}_h}$ and $e_{1h} = \frac{\bar{i}_h-\bar{I}_h}{\bar{I}_h}$

$E\,(e_{0h}) = E\,(e_{1h}) = 0$

$$E(e_{0h}^2) = \left[\frac{1}{nW_h}-\frac{1}{N_h}\right]C_{dh}^2$$

$$E(e_{1h}^2) = \left[\frac{1}{nW_h}-\frac{1}{N_h}\right]C_{ih}^2$$

$$E\,(e_{0h}e_{1h}) = \left[\frac{1}{nW_h}-\frac{1}{N_h}\right]\rho_{dih}C_{ih}C_{dh}.$$

we will find the expected values of error terms as

$$E\,(e_0) = E\left(\frac{\sum_{h=1}^{K} W_h F_h(y)e_{0h}}{F(y)}\right) = \frac{1}{F(y)}\left(\sum_{h=1}^{K} W_h F_h(y)E(e_{0h})\right) = 0$$

Similarly,

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

$$E\,(e_0^2) = E\left(\frac{\sum_{h=1}^{K} W_h\bar{D}_h^2 e_{0h}}{\bar{D}}\right)^2 = \frac{1}{\bar{D}^2}\sum_{h=1}^{K} W_h^2\bar{D}_h^2 E({e^2}_{0h})$$

$$= \frac{1}{\bar{D}^2}\sum_{h=1}^{K} W_h^2\bar{D}_h^2\left[\frac{1}{nW_h}-\frac{1}{N_h}\right]C_{dh}^2$$

$$= \frac{1}{\bar{D}^2}\sum_{h=1}^{K} W_h^2\bar{D}_h^2\left[\frac{1}{nW_h}-\frac{1}{N_h}\right]C_{dh}^2$$

$$= \frac{1}{\bar{D}^2}\left[\frac{1}{n}-\frac{1}{N}\right]\sum W_h S_{dh}^2 = V_D(\text{say})$$

Similarly,

$$E\,(e_1^2) = \frac{1}{\bar{I}^2}\left[\frac{1}{n}-\frac{1}{N}\right]\sum_{h=1}^{K}\sum W_h S_{ih}^2 = V_I$$

$$E\,(e_0 e_1) = \frac{1}{\bar{D}\bar{I}}\left[\frac{1}{n}-\frac{1}{N}\right]\sum_{h=1}^{K} W_h S_{dih} = V_{DI} \quad (1)$$

II. Properties of Estimators in Conditional Post-Stratification

$$E_1(e_0^2) = \frac{1}{\bar{D}^2}\sum_{h=1}^{K} W_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_{dh}^2 = V_{1D}(\text{say})$$

$$E_1(e_1^2) = \frac{1}{\bar{I}^2}\sum_{h=1}^{K} W_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_{ih}^2 = V_{1I}$$

$$E_1(e_0 e_1) = \frac{1}{\bar{D}\bar{I}} \sum_{h=1}^{K} W_h \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{dih} = V_{1DI} \quad (2)$$

## III. Estimators in Literature

We write the following estimators in terms of Unconditional case in post-stratification as

a. The usual unbiased estimator of population mean $\bar{D} = \sum_{h=1}^{K} W_h \bar{D}_h$ is given by

$$u_1 = \bar{d}_{ps} = \sum_{h=1}^{K} W_h \bar{d}_h \quad (3)$$

Using the results from Stephen (1945), the variances of $\bar{d}_{ps}$ to the first degree of approximation is given by

For Unconditional post-stratification,

$$\mathrm{Var}(u_{1a}) = \left[\frac{1}{n} - \frac{1}{N}\right] \sum_{h=1}^{K} \sum W_h S_{dh}^2 + \frac{1}{n^2} \sum_{h=1}^{K} (1 - W_h) S_{dh}^2 \quad (4)$$

Where $\mathrm{Var}(u_{1a})$ is the Unconditional variance of post stratified estimator $\bar{d}_{ps}$.

and $S_{dh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (d_{hi} - \bar{D}_h)^2$

For Conditional post- stratification,

$$\mathrm{Var}(u_{1b}) = \sum_{h=1}^{K} W_h^2 \left[\frac{1}{n_h} - \frac{1}{N_h}\right] S_{dh}^2 \quad (5)$$

$\mathrm{Var}(u_{1b})$ is the Conditional variance of post stratified estimator $\bar{d}_{ps}$.

b. The Ratio estimator for population mean according to Naik and Gupta [2] is given by

$$u_2 = \bar{d}_{ps} \left(\frac{\bar{I}}{\bar{i}_{ps}}\right)$$

Where $\bar{i}_{ps} = \sum_{h=1}^{K} W_h \bar{i}_h$

Up to the first degree of approximation, the MSE of estimator $u_2$ is given by

$$\mathrm{MSE}\,(u_2) = \bar{D}^2 \left[V_D + V_I \left[1 - 2\left(\frac{V_{DI}}{V_I}\right)\right]\right] \quad (6)$$

c. The usual product estimator is given by

$$u_3 = \bar{d}_{ps} \left(\frac{\bar{i}_{ps}}{\bar{I}}\right)$$

Up to the first degree of approximation, the MSE of estimator $u_3$ is given by

$$\mathrm{MSE}\,(u_3) = \bar{D}^2 \left[V_D + V_I \left[1 + 2\left(\frac{V_{DI}}{V_I}\right)\right]\right] \quad (7)$$

d. The Usual regression estimator for $\bar{D}$ is given by

$$u_4 = \bar{d}_{ps} + b_{ps}(\bar{I} - \bar{i}_{ps})$$

The MSE of estimator $u_4$ is given by

$$\mathrm{MSE}\,(u_4) = \bar{D}^2 V_D (1 - \xi_{DI}^2) \quad (8)$$

Where $\xi_{DI}^2 = \frac{V_{DI}^2}{V_D V_I}$

e. Koyuncu [7] proposed a class of estimators as

$$u_5 = \left[r_1\bar{d}_{ps} + r_2(\bar{I} - \bar{i}_{ps})\right]\left(\frac{a_{ps}\bar{I}+b_{ps}}{a_{ps}\bar{i}_{ps}+b_{ps}}\right)$$

Its MSE is given by

$$\mathrm{MSE}\,(u_5) = \bar{D}^2[1 + r_1^2A_1 + r_2^2A_2 + 2r_1r_2A_3 - 2r_1A_4 - 2r_2A_5] \tag{9}$$

Where $A_1 = 1 + V_D + \varphi_{ps}V_I\left[3\varphi_{ps} - 4\left(\frac{V_{DI}}{V_I}\right)\right]$

$$A_2 = \frac{V_I}{R^2}$$

$$A_3 = \frac{V_I}{R}\left(2\varphi_{ps} - \frac{V_{DI}}{V_I}\right)$$

$$A_4 = 1 + \varphi_{ps}V_I\left(\varphi_{ps} - \frac{V_{DI}}{V_I}\right)$$

$A_5 = \left(\frac{V_I}{R}\right)\varphi_{ps}$, $R=\frac{\bar{D}}{\bar{I}}$, $\varphi_{ps} = \frac{a_{ps}\bar{I}}{a_{ps}\bar{I}+b_{ps}}$,

$r_1 = \frac{A_2A_4-A_3A_5}{A_1A_2-A_3^2}$ and $r_2 = \frac{A_1A_5-A_3A_4}{A_1A_2-A_3^2}$.

f. Sharma and Singh [9] proposed exponential type estimators as

Ratio type exponential estimator

$$u_{6a} = \bar{d}_{ps}\exp\left(\frac{\bar{I} - \bar{i}_{ps}}{\bar{I} + \bar{i}_{ps}}\right)$$

Product type exponential estimator as

$$u_{6b} = \bar{d}_{ps}\exp\left(\frac{\bar{i}_{ps} - \bar{I}}{\bar{i}_{ps} + \bar{I}}\right)$$

$$u_{6c} = \bar{d}_{ps}\exp\left(\frac{\alpha(\bar{I} - \bar{i}_{ps})}{\bar{I} + \bar{i}_{ps}}\right)$$

Where α being a suitable constant

The MSEs of the above estimators as

$$\mathrm{MSE}\,(u_{6a}) = \bar{D}^2\left\{V_D + \frac{V_I}{4}\left[1 - 4\left(\frac{V_{DI}}{V_I}\right)\right]\right\}$$

$$\mathrm{MSE}\,(u_{6b}) = \bar{D}^2\left\{V_D + \frac{V_I}{4}\left[1 + 4\left(\frac{V_{DI}}{V_I}\right)\right]\right\}$$

$$\mathrm{MSE}\,(u_{6c}) = \bar{D}^2\left\{V_D + \frac{\alpha V_I}{4}\left[\alpha - 4\left(\frac{V_{DI}}{V_I}\right)\right]\right\} \tag{10}$$

Where $\alpha = 2 * \left(\frac{V_{DI}}{V_I}\right)$

g. Sharma and Singh [9] suggested a class of estimators as

$$u_7 = \left[r_1\bar{d}_{ps} + r_2(\bar{I} - \bar{i}_{ps})\right]\exp\left(\frac{a_{ps}(\bar{I} - \bar{i}_{ps})}{a_{ps}(\bar{I} + \bar{i}_{ps}) + 2b_{ps}}\right)$$

Where $a_{ps}$, $b_{ps}$ are either real numbers or the functions of the auxiliary variable.

Its MSE is given by

$$MSE\ (u_7) = \overline{D}^2\ [1 + r_1^2 B_1 + r_2^2 B_2 + 2r_1 r_2 B_3 - 2r_1 B_4 - 2r_2 B_5] \quad (11)$$

Where $B_1 = 1 + V_D + \varphi_{ps} V_I \left[\varphi_{ps} - 2\left(\frac{V_{DI}}{V_I}\right)\right]$

$B_2 = \frac{V_I}{R^2}$

$B_3 = \frac{V_I}{R}\left(\varphi_{ps} - \frac{V_{DI}}{V_I}\right)$

$B_4 = 1 + \varphi_{ps}\frac{V_I}{8}\left(3\varphi_{ps} - 4\frac{V_{DI}}{V_I}\right)$

$B_5 = \left(\frac{V_I}{2R}\right)\varphi_{ps}$

$r_1 = \frac{B_2B_4 - B_3B_5}{B_1B_2 - B_3^2}$ and $r_2 = \frac{B_1B_5 - B_3B_4}{B_1B_2 - B_3^2}$.

h. Singh et al. [13] suggested another class of estimators for population mean as

$$u_8 = \left[r_1\overline{d}_{ps} + r_2 \exp\left(\frac{\delta a_{ps}(\overline{I} - \overline{i}_{ps})}{a_{ps}(\overline{I} + \overline{i}_{ps}) + 2b_{ps}}\right)\right]\left(\frac{a_{ps}\overline{I} + b_{ps}}{a_{ps}\overline{i}_{ps} + b_{ps}}\right)^{\eta}$$

Where $(\delta, \eta)$are constants belongs to real numbers like (-1,0,1).

$$MSE\ (u_8) = \overline{D}^2\ [1 + r_1^2 C_1 + r_2^2 C_2 + 2r_1 r_2 C_3 - 2r_1 C_4 - 2r_2 C_5] \quad (12)$$

Where, $C_1 = 1 + V_D - 4\eta\varphi_{ps}V_{DI} + \eta(2\eta + 1)\varphi_{ps}^2 V_I$

$C_2 = \frac{1}{\overline{I}^2 R^2}[\,1 + \theta(2\theta + 1)\varphi_{ps}^2 V_I]$

$C_3 = \frac{1}{\overline{I}R}\left[1 + \frac{(\eta + \theta)(\eta + \theta + 1)}{2}\varphi_{ps}^2 V_I - (\eta + \theta)\varphi_{ps}V_{DI}\right]$

$C_4 = 1 + \varphi_{ps}\frac{\eta}{2}\left(\frac{(\eta + 1)}{2}\varphi_{ps}V_I - 2V_{DI}\right)$

$C_5 = \frac{1}{\overline{I}R}\left[1 + \frac{\theta(\theta+1)}{2}\varphi_{ps}^2 V_I\right]$

$\theta = (2S_{dih} - 1)/2$ , $r_1 = \frac{C_2C_4 - C_3C_5}{C_1C_2 - C_3^2}$ and $r_2 = \frac{C_1C_5 - C_3C_4}{C_1C_2 - C_3^2}$.

We have written the above considered pre-existing estimators in Unconditional case. If we change the expectations of error terms like in equation (2), we get the estimators in Conditional case.

## IV. Suggested class of estimators in post stratification

We propose a class of estimators for population mean $\overline{D}$ as

$$u_{prop} = \left[r_1\overline{d}_{ps} + r_2\left(\overline{I} - \overline{i}_{ps}\right) - r_3\overline{d}_{ps}\left(\frac{a_{ps}\overline{I} + b_{ps}}{a_{ps}\overline{i}_{ps} + b_{ps}}\right)\right]\exp\left(\frac{\overline{I} - \overline{i}_{ps}}{\overline{I} + \overline{i}_{ps}}\right)^{\eta} \quad (13)$$

Where $(r_1, r_2, \eta)$ are suitable constants and $(a_{ps}, b_{ps})$ are either constants or functions of auxiliary variable.

Expressing the equation (13) in terms of $e_{0h}$ and $e_{1h}$, we have

$$u_{prop} = \left[ r_1 \sum_{h=1}^{K} W_h \overline{D}_h \, (1+e_{0h}) + r_2 \left( \sum_{h=1}^{K} W_h \overline{I}_h - \sum_{h=1}^{K} W_h \overline{I}_h \, (1+e_{1h}) \right) - r_3 \sum_{h=1}^{K} W_h \overline{D}_h \, (1+e_{0h})(1 + \varphi_{ps} e_1)^{-1} \right] \exp \left( \frac{\sum_{h=1}^{K} W_h \overline{I}_h - \sum_{h=1}^{K} W_h \overline{I}_h (1+e_{1h})}{\sum_{h=1}^{K} W_h \overline{I}_h + \sum_{h=1}^{K} W_h \overline{I}_h (1+e_{1h})} \right)^{\eta}$$

Where $\varphi_{ps} = \frac{a_{ps}\overline{I}}{a_{ps}\overline{I}_{ps} + b_{ps}}$

$$= \left[ r_1 \overline{D}(1+e_0) + r_2 \left( \overline{I} - \overline{I}(1+e_1) \right) - r_3 \overline{D}(1+e_0)(1+\varphi_{ps}e_1)^{-1} \right] \exp \left( \frac{\overline{I} - \overline{I}(1+e_1)}{\overline{I} + \overline{I}(1+e_1)} \right)^{\eta}$$

$$= \left[ r_1 \overline{D}(1+e_0) - r_2 \overline{I}\, e_1 - r_3 \overline{D}(1+e_0)(1 - \varphi_{ps} e_1 + \varphi_{ps}^2 e_1^2 \right] \exp \left( \frac{-e_1}{2+e_1} \right)^{\eta}$$

$$= \left[ r_1 \overline{D}(1+e_0) - r_2 \overline{I}\, e_1 - r_3 \overline{D}(1 + e_0 - \varphi_{ps} e_1 - \varphi_{ps} e_0 e_1 + \varphi_{ps}^2 e_1^2) \right] \left[ 1 - \frac{\eta}{2} e_1 + \frac{3}{8} \eta e_1^2 \right]$$

$$u_{prop} - \overline{D} = \overline{D} \left\{ \left[ (r_1 - 1) + r_1 e_0 - r_2 m\, e_1 - r_3 (1 + e_0 - \varphi_{ps} e_1 - \varphi_{ps} e_0 e_1 + \varphi_{ps}^2 e_1^2) \right] \left[ 1 - \frac{\eta}{2} e_1 + \frac{3}{8} \eta e_1^2 \right] \right\} \tag{14}$$

Where $m = \frac{\overline{I}}{\overline{D}}$

$$u_{prop} - \overline{D} = \overline{D} \left\{ (r_1 - 1) + e_0 (r_1 - r_3) - r_3 + e_1 \left( -r_2 m + r_3 \varphi_{ps} - \frac{\eta(r_1 - 1)}{2} - \frac{\eta r_3}{2} \right) + e_1^2 \left( \frac{3\eta(r_1 - 1)}{8} + r_2 \left( \frac{m\eta}{2} \right) - r_3 \varphi_{ps}^2 - r_3 \frac{\eta \varphi_{ps}}{2} - r_3 \frac{3\eta}{8} \right) + e_0 e_1 \left( -r_1 \frac{\eta}{2} + r_3 \varphi_{ps} + r_3 \frac{\eta}{3} \right) \right\}$$

By taking expectation on both sides of equation (14), we get bias as

$$\text{Bias}\,(u_{prop}) = \overline{D} \left\{ (r_1 - 1) - r_3 + V_I \left[ \frac{3\eta(r_1-1)}{8} + r_2 \left( \frac{m\eta}{2} \right) - r_3 \left( \varphi_{ps}^2 + \frac{\eta \varphi_{ps}}{2} + \frac{3\eta}{8} \right) \right] + V_{DI} \left( -r_1 \frac{\eta}{2} + r_3 \varphi_{ps} + r_3 \frac{\eta}{3} \right) \right\}$$

By taking square on both sides of equation (14), we have

$$\left( u_{prop} - \overline{D} \right)^2 = \overline{D}^2 \left\{ \left( (r_1 - 1) - r_3 \right)^2 + e_0^2 (r_1 - r_3)^2 + e_1^2 \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} + \frac{\eta r_3}{2} \right)^2 - 2 e_0 e_1 \left[ (r_1 - r_3) \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} + \frac{\eta r_3}{2} \right) \right] + 2((r_1 - 1) + r_3) \left[ e_0 (r_1 - r_3) - e_1 \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} - \frac{\eta r_3}{2} \right) \right] + e_1^2 \left[ \left( \frac{3\eta(r_1-1)}{8} + r_2 \left( \frac{m\eta}{2} \right) - r_3 \varphi_{ps}^2 - r_3 \frac{\eta \varphi_{ps}}{2} - r_3 \frac{3\eta}{8} \right) \right] + e_0 e_1 \left( -r_1 \frac{\eta}{2} + r_3 \varphi_{ps} + r_3 \frac{\eta}{3} \right) \right\} \tag{15}$$

By considering expectation on both sides of equation (15), we get MSE as

$$\text{MSE}\,(u_{prop}) = \overline{D}^2 \left\{ \left( (r_1 - 1) - r_3 \right)^2 + V_D (r_1 - r_3)^2 + V_I \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} - \frac{\eta r_3}{2} \right)^2 - 2 V_{DI} \left[ (r_1 - r_3) \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} - \frac{\eta r_3}{2} \right) \right] + V_I \left[ \left( \frac{3\eta(r_1-1)}{8} + r_2 \left( \frac{m\eta}{2} \right) - r_3 \varphi_{ps}^2 - r_3 \frac{\eta \varphi_{ps}}{2} - r_3 \frac{3\eta}{8} \right) \right] + V_{DI} \left( -r_1 \frac{\eta}{2} + r_3 \varphi_{ps} + r_3 \frac{\eta}{3} \right) \right\}$$

$$= \overline{D}^2 \left\{ (r_1 - 1)^2 + (r_3)^2 - 2 r_3 (r_1 - 1) + V_D (r_1 - r_3)^2 + V_I \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} - \frac{\eta r_3}{2} \right)^2 - 2 V_{DI} \left[ (r_1 - r_3) \left( r_2 m - r_3 \varphi_{ps} + \frac{\eta(r_1-1)}{2} - \frac{\eta r_3}{2} \right) \right] + V_I \left[ \left( \frac{3\eta(r_1-1)}{8} + r_2 \left( \frac{m\eta}{2} \right) - r_3 \varphi_{ps}^2 - r_3 \frac{\eta \varphi_{ps}}{2} - r_3 \frac{3\eta}{8} \right) \right] + \right.$$

$V_{DI}\left(-r_1\frac{\eta}{2}+r_3\varphi_{ps}+r_3\frac{\eta}{3}\right)\Big\}$

We rewrite the above equation as

$$\text{MSE }(u_{prop})=\overline{D}^2\left[1-\frac{3\eta}{8}V_I+\vartheta_1 r_1+\vartheta_2 r_1^2+\vartheta_3 r_2+\vartheta_4 r_2^2+\vartheta_5 r_3+\vartheta_6 r_3^2+\vartheta_7 r_1 r_2+\vartheta_8 r_1 r_3+\vartheta_9 r_2 r_3\right] \quad (16)$$

Where $\vartheta_1=-2+\frac{3\eta}{8}V_I+\eta V_{DI}-\frac{\eta^2}{2}V_I-\frac{\eta}{2}V_{DI}$

$\vartheta_2=1+V_D+\frac{\eta^2}{4}V_I-\eta V_{DI}$

$\vartheta_3=-m\,\eta V_I+\frac{m\eta}{2}V_I$

$\vartheta_4=m^2V_I$

$\vartheta_5=2+\frac{\eta^2}{2}V_I+\varphi_{ps}\,\eta V_I-\varphi_{ps}^2V_I-\frac{\eta}{2}\varphi_{ps}V_I-\frac{3\eta}{8}V_I+\eta V_{DI}+\varphi_{ps}V_{DI}+\frac{\eta}{3}V_{DI}$

$\vartheta_6=1+V_D+\varphi_{ps}^2V_I+\frac{\eta^2}{4}V_I+\varphi_{ps}V_I-2\varphi_{ps}V_{DI}-\eta V_{DI}$

$\vartheta_7=m\,\eta V_I-2mV_{DI}$

$\vartheta_8=-2-V_D-\frac{\eta^2}{2}V_I-\varphi_{ps}\,\eta V_I+2\varphi_{ps}V_{DI}+2\eta V_{DI}$

$\vartheta_9=-2m\varphi_{ps}V_I-m\eta V_I+2mV_{DI}$

To get the values of $r_1, r_2$ and $r_3$, differentiate equation (16) with respect to $r_1, r_2$ and $r_3$ and equate them to zero. We get,

$r_3=\frac{(2\vartheta_2\vartheta_3-\vartheta_1\vartheta_7)(2\vartheta_4\vartheta_8-\vartheta_7\vartheta_9)-(\vartheta_5\vartheta_7-\vartheta_3\vartheta_8)(\vartheta_7{}^2-4\vartheta_2\vartheta_4)}{(\vartheta_7\vartheta_8-2\vartheta_2\vartheta_9)(2\vartheta_4\vartheta_8-\vartheta_7\vartheta_9)-(\vartheta_7{}^2-4\vartheta_2\vartheta_4)(\vartheta_8\vartheta_9-2\vartheta_6\vartheta_7)}=\theta_1(\text{say})$

$r_2=\frac{(2\vartheta_2\vartheta_3-\vartheta_1\vartheta_7)-\theta_1(\vartheta_7\vartheta_8-2\vartheta_2\vartheta_9)}{(\vartheta_7{}^2-4\vartheta_2\vartheta_4)}=\theta_2(\text{say})$

$r_1=\frac{-(\vartheta_7\theta_2+\vartheta_8\theta_1+\vartheta_1)}{2\vartheta_2}$

## V. Efficiency comparison

Theoretically, we establish the following criteria to assess the effectiveness of suggested estimator and those estimators taken into consideration in the literature.

By comparing equations (4) and (16), we have

$\text{MSE }(u_{prop})-\text{MSE}(u_{1a})<0$

$\overline{D}^2\left[1-\frac{3\eta}{8}V_I+\vartheta_1 r_1+\vartheta_2 r_1^2+\vartheta_3 r_2+\vartheta_4 r_2^2+\vartheta_5 r_3+\vartheta_6 r_3^2+\vartheta_7 r_1 r_2+\vartheta_8 r_1 r_3+\vartheta_9 r_2 r_3\right]<\left[\frac{1}{n}-\frac{1}{N}\right]\sum_{h=1}^{K}\sum W_h S_{dh}^2+\frac{1}{n^2}\sum_{h=1}^{K}(1-W_h)\,S_{dh}^2$

By comparing equations (5) and (16),

$\text{MSE }(u_{prop})-\text{MSE}(u_{1b})<0$

$\overline{D}^2\left[1-\frac{3\eta}{8}V_I+\vartheta_1 r_1+\vartheta_2 r_1^2+\vartheta_3 r_2+\vartheta_4 r_2^2+\vartheta_5 r_3+\vartheta_6 r_3^2+\vartheta_7 r_1 r_2+\vartheta_8 r_1 r_3+\vartheta_9 r_2 r_3\right]<\sum_{h=1}^{K}W_h^2\left[\frac{1}{n_h}-\frac{1}{N_h}\right]S_{dh}^2$

By comparing equations (6) and (16),

$\text{MSE }(u_{prop})-\text{MSE }(u_2)<0$

$\left[1-\frac{3\eta}{8}V_I+\vartheta_1 r_1+\vartheta_2 r_1^2+\vartheta_3 r_2+\vartheta_4 r_2^2+\vartheta_5 r_3+\vartheta_6 r_3^2+\vartheta_7 r_1 r_2+\vartheta_8 r_1 r_3+\vartheta_9 r_2 r_3\right]<\left[V_D+V_I\left[1-2\left(\frac{V_{DI}}{V_I}\right)\right]\right]$

By comparing equations (7) and (16),

MSE ($u_{prop}$) − MSE ($u_3$) < 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<\left[V_D+V_I\left[1+2\left(\frac{V_{DI}}{V_I}\right)\right]\right]$$

By comparing equations (8) and (16),

MSE ($u_{prop}$) − MSE ($u_4$) < 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<V_D(1-\xi_{DI}^2)$$

By comparing equations (9) and (16),

MSE ($u_{prop}$) − MSE ($u_5$) < 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<[1+r_1^2A_1+r_2^2A_2+2r_1r_2A_3-2r_1A_4-2r_2A_5]$$

By comparing equations (10) and (16), we have

MSE ($u_{prop}$) − MSE ($u_{6a}$)< 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<\left\{V_D+\frac{V_I}{4}\left[1-4\left(\frac{V_{DI}}{V_I}\right)\right]\right\}$$

MSE ($u_{prop}$) < MSE ($u_{6b}$)

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<\left\{V_D+\frac{V_I}{4}\left[1+4\left(\frac{V_{DI}}{V_I}\right)\right]\right\}$$

MSE ($u_{prop}$) − MSE ($u_{6c}$) < 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<\left\{V_D+\frac{Av_I}{4}\left[\alpha-4\left(\frac{V_{DI}}{V_I}\right)\right]\right\}$$

By comparing equations (11) and (16),

MSE ($u_{prop}$) − MSE ($u_7$) < 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<[1+r_1^2B_1+r_2^2B_2+2r_1r_2B_3-2r_1B_4-2r_2B_5]$$

By comparing equations (12) and (16),

MSE ($u_{prop}$) − MSE ($u_8$) < 0

$$\left[1-\frac{3\eta}{8}V_I+\vartheta_1r_1+\vartheta_2r_1^2+\vartheta_3r_2+\vartheta_4r_2^2+\vartheta_5r_3+\vartheta_6r_3^2+\vartheta_7r_1r_2+\vartheta_8r_1r_3+\vartheta_9r_2r_3\right]<[1+r_1^2C_1+r_2^2C_2+2r_1r_2C_3-2r_1C_4-2r_2C_5]$$

## VI. Empirical study

We use information from the Ministry of Education of the Turkish Republic from 2007 on the number of teachers as the study variable (d) and the number of students classifying more or less than 750 in primary and secondary schools as the auxiliary attribute (i) for 923 districts across 6 regions (as 1: Marmara) 2, Atlantic, 3, Mediterranean, and 4, Central Anatolia Black Sea 5 and 6: East and Southeast Anatolia). Table 1 provides the data's summary statistics. We used Neyman allocation to place the samples in different strata.

The functions of auxiliary variable which we used in numerical calculation are:

$\sum W_h\, C_{ih} = 0.266448$, $\sum W_h\, S_{ih}$ =0.246447 and $\sum W_h\, \rho_{dih} = 0.145833$

In the case of unconditional post stratification, Table.2 shows the MSE values for our suggested estimator and the other estimators that were taken into consideration, together with the PRE values. It has been noted that the proposed estimator exhibits the maximum relative efficiency. It is also same in case of Conditional post stratification by observing Table.3.

**Table 1.** *Data Descriptive Statistics*

| Stratum no. | $N_h$ | $n_h$ | $\bar{D}_h$ | $\bar{I}_h$ | $S_{dh}$ | $S_{ih}$ | $S_{dih}$ | $\beta_{2(ih)}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 127 | 31 | 703.74 | 0.952 | 883.835 | 0.213 | 25.267 | 16.922 |
| 2 | 117 | 21 | 413 | 0.974 | 644.922 | 0.159 | 9.982 | 35.579 |
| 3 | 103 | 29 | 573.17 | 0.932 | 1033.467 | 0.253 | 37.453 | 10.34 |
| 4 | 170 | 38 | 424.66 | 0.888 | 810.585 | 0.316 | 44.625 | 4.231 |
| 5 | 205 | 22 | 267.03 | 0.912 | 403.654 | 0.284 | 21.04 | 6.675 |
| 6 | 201 | 39 | 393.84 | 0.95 | 711.723 | 0.218 | 18.66 | 15.56 |

**Table 2**. *Unconditional case: MSE and PRE values of existing estimators and proposed estimator.*

| S. No. | Estimator | MSE value | Percentage Relative Efficiency |
|---|---|---|---|
| 1. | $u_{1a}$ | 2539.82 | 100 |
| 2. | $u_2$ | 2400.58 | 105.80 |
| 3. | $u_3$ | 2629.67 | 96.58 |
| 4. | $u_4$ | 2398.50 | 105.89 |
| 5. | $u_5$ | 2397.75 | 105.92 |
| 6. | $u_{6a}$ | 2405.89 | 105.57 |
| 7. | $u_{6b}$ | 2520.44 | 100.77 |
| 8. | $u_{6c}$ | 2398.50 | 105.89 |
| 9. | $u_7$ | 2397.75 | 105.93 |
| 10. | $u_8$ | 1931.21 | 131.51 |
| 11. | $u_{prop}$ | 1714.78 | 148.11 |

**Table 3**. *Conditional case: MSE and PRE values of existing estimators and proposed estimator.*

| S. No. | Estimator | MSE value | Percentage Relative Efficiency |
|---|---|---|---|
| 1. | $u_{1b}$ | 2229.27 | 100 |
| 2. | $u_2$ | 1638.65 | 136.10 |
| 3. | $u_3$ | 2983.40 | 77.32 |
| 4. | $u_4$ | 846.88 | 263.23 |
| 5. | $u_5$ | 846.78 | 263.26 |
| 6. | $u_{6a}$ | 1913.52 | 116.50 |
| 7. | $u_{6b}$ | 2585.89 | 86.21 |
| 8. | $u_{6c}$ | 2204.72 | 101.11 |
| 9. | $u_7$ | 846.78 | 263.26 |
| 10. | $u_8$ | 519.11 | 429.44 |
| 11. | $u_{prop}$ | 68.54 | 3252.51 |

We may conclude that conditional post stratification outperformed unconditional post stratification by comparing MSE and PRE values in Tables.2 and 3.

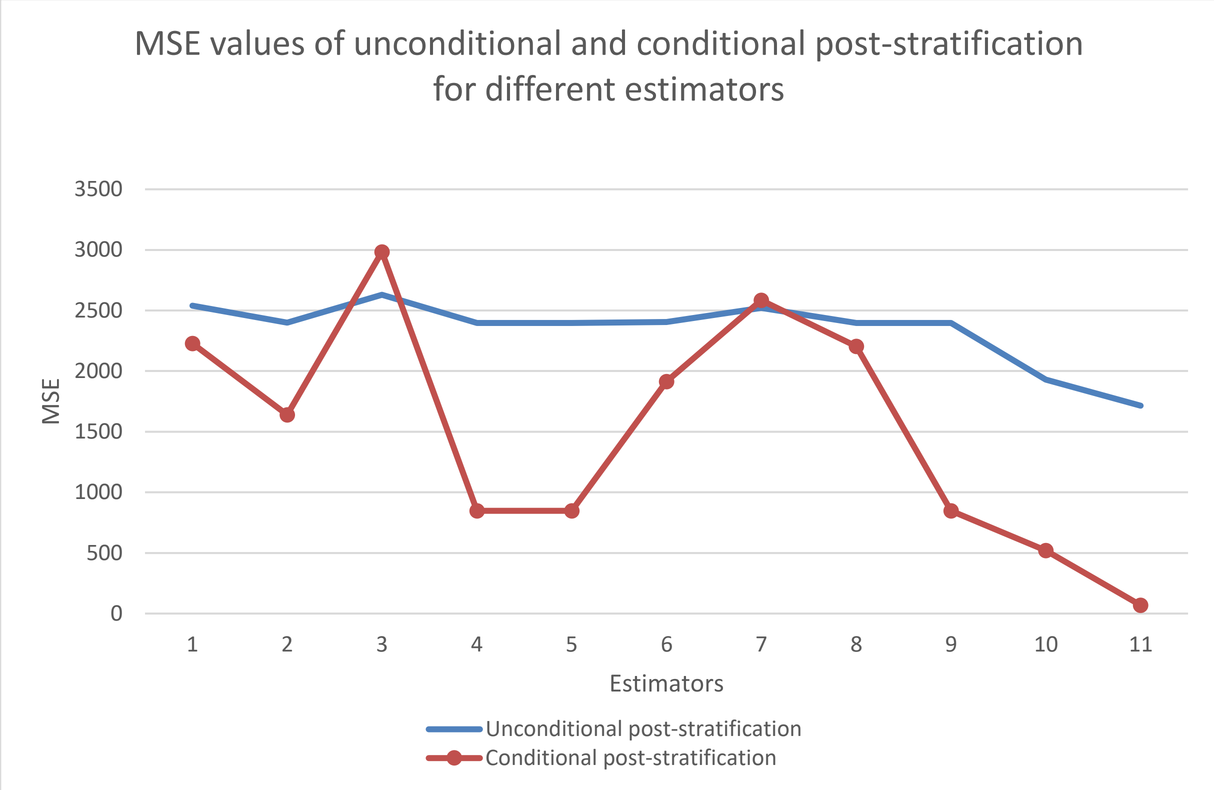


**Figure 1**. *MSE values for both Unconditional and Conditional Post-stratified estimators*.

We have presented the estimators in Table 2 and 3 in the above figure graphically. A line that represents the MSE values of Unconditional post stratified estimators can be seen in the picture together with a line with markers that represents the MSE values of conditional post stratified estimators. According to the graphic, conditional post stratified estimators have lower MSE values than unconditional post stratified estimators.

## VII. Conclusion

In this research paper, we introduced a novel class of estimators and derived their Mean Square Error (MSE). We also investigated existing estimators and considered two cases in post-stratification: conditional and unconditional. Through a real data analysis, we computed the MSE and Percentage Relative Efficiency (PRE) values for all estimators presented in this study. The results, as shown in Table 2 and 3, clearly demonstrate that our proposed estimator exhibits the highest relative efficiency compared to the other estimators considered. Furthermore, we observed from the figure 1 that conditional post-stratification outperformed unconditional post-stratification in our analysis. These findings highlight the potential of our proposed estimators for enhancing mean estimation accuracy in post-stratification studies.

## References

[1] Holt, D. and Smith, T. F. (1979). Post stratification. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 142(1):33-46.

[2] Naik, V. D. and Gupta, P. C. (1996). A note on estimation of mean with known population proportion of an auxiliary character. *Journal of Indian Society of Agricultural Stat*istics, *48*(2):151-158.

[3] Frey, J. and Feeman, T. G. (2012). An improved mean estimator for judgment post-stratification. *Computational Statistics and Data Analysis*, 56(2):418-426.

[4] Singh, H. P. and Solanki, R. S. (2012). Improved estimation of population mean in simple random sampling using information on auxiliary attribute. *Applied Mathematics and Computation*, 218(15):7798-7812.

[5] Onyeka, A. C. (2012). Estimation of population mean in post-stratified sampling using known value of some population parameter (s). *Statistics in Transition. New Series*, 13(1): 65-78.

[6] Onyeka, A. C., Nlebedim, V. U. and Izunobi, C. H. (2013). Exponential Estimators of Population Mean in Post-Stratified Sampling using Known Value of Some Population Parameters. *Global Journal of Science Frontier Research Mathematics and Decision Sciences*, *13*(9).

[7] Koyuncu, N. (2013). Improved estimation of population mean in stratified random sampling using information on auxiliary attribute. In *Proceeding of 59-th ISI world Statistics Congress, Hong Kong, China, August* (pp. 25-30).

[8] Singh, H. P. and Solanki, R. S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statistical papers*, 54:479-497.

[9] Sharma, P. and Singh, R. (2013). Efficient estimator of population mean in stratified random sampling using auxiliary attribute. *World Applied Sciences Journal*, *27*(12):1786-1791.

[10] 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*, 46(21):10387-10393.

[11] Zamanzade, E. and Wang, X. (2017). Estimation of population proportion for judgment post-stratification. *Computational Statistics and Data Analysis*, *112*:257-269.

[12] Shahzad, U., Hanif, M., Koyuncu, N. and Luengo, A. G. (2019). A family of ratio estimators in stratified random sampling utilizing auxiliary attribute alongside the nonresponse issue. *Journal of Statistical Theory and Applications*, *18*(1):12-25.

[13] Singh, H. P., Gupta, A., and Tailor, R. (2023). Efficient class of estimators for finite population mean using auxiliary attribute in stratified random sampling. *Scientific Reports*, *13*(1):10253.

[14] Kumari, A., Kumar, U. and Singh, R. (2023). Improved estimators of population mean using auxiliary variables in post-stratification. *Journal of Scientific Research*, *67*(1):94-98.