



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 14    **Issue:** VII    **Month of publication:** July 2026

**DOI:** <https://doi.org/10.22214/ijraset.2026.84118>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Data Science Perspectives on Calibration Estimators for Finite Population Mean in Stratified Random Sampling Using Different Distance Measures

Anupama Bhadauriya<sup>1</sup>, Rakesh Prasad Sarang<sup>2</sup>, Anupam Kumar Srivastava<sup>3</sup>

<sup>1,2</sup>Department of Computer Science & Engineering, ITM University, Gwalior (MP), India

<sup>3</sup>Department of Computer Science and Engineering, S R Group of Institutions Jhansi (UP), India

**Abstract:** This article presents a data analytics framework that utilizes auxiliary information to improve estimation accuracy and decision-making in stratified sampling. Various analytical models based on different distance measures are developed and evaluated to optimize data-driven insights. The study discusses the theoretical foundations of the proposed estimation techniques and derives optimized weighting schemes under different distance measures. The performance of the developed methods is assessed through simulation studies and comparative analysis. Furthermore, a real-world dataset is analyzed to demonstrate the effectiveness, robustness, and practical applicability of the proposed data analytics approach. The results indicate that the suggested methodology provides superior accuracy and efficiency compared to conventional techniques, making it a valuable tool for modern data-driven applications.

**Keywords:** Calibration Estimator, Stratified Sampling, Distance Function, Data Analytics Framework, Auxiliary Information.

## I. INTRODUCTION

In survey sampling, the precision of the estimate of study variable can be increased by using the most popular stratified sampling technique.

The calibration-based estimation method helps in improving the survey estimates by means of auxiliary information through adjusting the initial design weights. A calibration estimator uses modified weights which are known as calibrated weights. These calibrated weights are determined by minimizing a given distance function to the initial design weights respecting a set of constraints associated with auxiliary information. In survey sampling many authors, such as Deville and Sarndal (1992), Estevao and Sarndal (2000), Arnab and Singh (2005), Farrell and Singh (2005), Kim and Park (2010) etc., defined some calibration estimators using different constraints. In stratified random sampling, calibration approach is used to get optimum strata weights. Tracy et al. (2003), Kim et al. (2007) and Koyuncu (2012) define some calibration estimators in stratified random sampling. Mouhamed et al. (2015), Koyuncu and Kadilar (2016), Clement and Enang (2017), Nidhi et al. (2017), Garg and Pachori (2020 etc., have contributed in the direction of developing some calibrated estimators for different population parameters using different calibration constraints under various sampling schemes. Koyuncu and Kadilar (2013) define calibration estimator using different distance measures in stratified sampling. Rai et al. (2020) worked on calibration based estimator using different distance measures in stratified sampling. Kris et al. (2026).

Next, we examine links to traditional statistical and design concepts such as the gulfs of interaction and parsimony. Important explainability methods, such as integrated gradients and learnt embeddings.

Lastly, we list unresolved issues and talk about how data science might help.

The main aim of this paper is to introduce calibration estimator under different distance measures using auxiliary information in stratified sampling scheme. We first discuss notations and calibration based estimators under different distance functions. Then we describe the simulation study using the real data of census (Uttar Pradesh, Series 10, Part 12B and District census hand book, AGRA). Finally, the conclusions of this study are presented.

## II. NOTATIONS AND CALIBRATION ESTIMATORS:

Consider a finite population  $U = \{U_1, U_2, \dots, U_N\}$  having  $N$  distinct and identifiable unit partitioned into  $L$  Strata. Let  $Y$  and  $X$  be the study and auxiliary variables taking values  $y_{hi}$  and  $x_{hi}$ , respectively for  $i^{th}$  unit ( $i = 1, 2, \dots, N$ ) in  $h^{th}$  stratum consisting of  $N_h$  units ( $h = 1, 2, \dots, L$ ) such that  $\sum_{h=1}^L N_h = N$ . Let  $n_h$  be the size of the sample for  $h^{th}$  stratum such that  $\sum_{h=1}^L n_h = n$ .

Let,

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h, \text{ where } \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

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

$W_h = \frac{N_h}{N}$  is the stratum weight. We can define similar expressions for  $\bar{x}_{st}$

$s_{yh}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$  and  $s_{xh}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$  be the sample variance of  $Y$  and  $X$  respectively in  $h^{th}$  stratum

corresponding to the population variances  $S_{yh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$  and  $S_{xh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$

$S_{yhx} = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h) (y_{hi} - \bar{Y}_h)$  Covariance between  $Y$  and  $X$ .

$\rho_{hx} = \frac{S_{yhx}}{S_{yh} * S_{xh}}, \rho_{hx} = \frac{S_{yhx}}{S_{yh} * S_{xh}}$  is the correlation coefficient of  $h^{th}$  stratum.

The calibration estimator under the stratified random sampling for population mean defined by Tracy et al. (2003) is given as

$$\bar{y}_{st}(\alpha) = \sum_{h=1}^L \Omega_h \bar{y}_h \tag{2.1}$$

## III. REVIEW OF LITERATURE

Nursel Koyuncu and Cem Kadilar (2013) worked on following distance function

$$L_1 = \sum_{h=1}^L \frac{(\Omega_h - W_h)^2}{W_h Q_h} \tag{2.2}$$

$$L_2 = \sum_{h=1}^L 2 \frac{(\sqrt{\Omega_h} - \sqrt{W_h})^2}{Q_h} \tag{2.3}$$

$$L_3 = \sum_{h=1}^L \frac{1}{Q_h} \left( \frac{\Omega_h}{W_h} - 1 \right)^2 \tag{2.4}$$

$$L_4 = \sum_{h=1}^L \frac{1}{Q_h} \left( \frac{\sqrt{\Omega_h}}{\sqrt{W_h}} - 1 \right)^2 \tag{2.5}$$

The simulation study shows that the calibration estimator using distance measure  $L_1$  are highly efficient than using distance measure  $L_3$

Rai et al. (2021) worked on calibration – based estimators using different distance measure under two auxiliary variables

$$D_1(\Omega_h, W_h) = \sum_{h=1}^L \frac{(\Omega_h - W_h)^2}{W_h Q_h} \tag{2.6}$$

$$D_2(\Omega_h, W_h) = 2 \sum_{h=1}^L \frac{(\sqrt{\Omega_h} - \sqrt{W_h})^2}{W_h Q_h} \tag{2.7}$$

$$D_3(\Omega_h, W_h) = \sum_{h=1}^L \frac{1}{Q_h} \left( \frac{\Omega_h}{W_h} - 1 \right)^2 \tag{2.8}$$

$$D_4(\Omega_h, W_h) = \sum_{h=1}^L \frac{1}{Q_h} \left( \frac{\sqrt{\Omega_h}}{\sqrt{W_h}} - 1 \right)^2 \tag{2.9}$$

$$D_5(\Omega_h, W_h) = \sum_{h=1}^L \frac{1}{Q_h} \left( -W_h \log \frac{\Omega_h}{W_h} + \Omega_h - W_h \right)^2 \tag{2.10}$$

We worked on following distance measures with help of Nursel Koyuncu and Cem Kadilar (2013), Rai et al. (2020):

$$L_{1p} = \sum_{h=1}^L \frac{(\Omega_h - W_h)^2}{W_h Q_h} \tag{2.11}$$

$$L_{2p} = \sum_{h=1}^L \frac{1}{Q_h} \left( \frac{\Omega_h}{W_h} - 1 \right)^2 \tag{2.12}$$

$$L_{3p} = \sum_{h=1}^L \frac{1}{Q_h} \left( -W_h \log \frac{\Omega_h}{W_h} + \Omega_h - W_h \right)^2 \tag{2.13}$$

$$L_{4p} = \sum_{h=1}^L 2 \frac{(\sqrt{\Omega_h} - \sqrt{W_h})^2}{Q_h} \tag{2.14}$$

and satisfy the following calibration constraints

$$\sum_{h=1}^L \Omega_h \bar{x}_h = \sum_{h=1}^L W_h \bar{X}_h \tag{2.15}$$

$$\sum_{h=1}^L \Omega_h \rho_{hx} = \sum_{h=1}^L W_h \rho_{hx} \tag{2.16}$$

Thus the calibration estimation problem reduce an optimization problem where

1) Case: 1 The Lagrange function for weights  $\Omega_h$  which satisfy the calibration equation (2.1) and minimize the loss function given in (2.11)

$$\phi_1 = \sum_{h=1}^L \frac{(\Omega_h - W_h)^2}{W_h Q_h} - 2\lambda_1 (\sum_{h=1}^L \Omega_h \bar{x}_h - \sum_{h=1}^L W_h \bar{X}_h) - 2\lambda_2 (\sum_{h=1}^L \Omega_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx})$$

where  $\lambda_1$  and  $\lambda_2$  are the Lagrange's multipliers differentiate above equation w.r.to  $\Omega_h$  and equating to zero for obtaining the value of  $\Omega_h$

$$\Omega_h = W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx}) \tag{2.17}$$

From equation (2.15)

$$\sum_{h=1}^L \{W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx})\} \bar{x}_h = \sum_{h=1}^L W_h \bar{X}_h$$

$$\sum_{h=1}^L \lambda_1 W_h Q_h \bar{x}_h^2 + \sum_{h=1}^L \lambda_2 W_h Q_h \rho_{hx} \bar{x}_h = \sum_{h=1}^L W_h \bar{X}_h - \sum_{h=1}^L W_h \bar{x}_h \tag{2.18}$$

From equation (2.16)

$$\sum_{h=1}^L \{W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx})\} \rho_{hx} = \sum_{h=1}^L W_h \rho_{hx}$$

$$\sum_{h=1}^L \lambda_1 W_h Q_h \rho_{hx} \bar{x}_h + \sum_{h=1}^L \lambda_2 W_h Q_h \rho_{hx}^2 = \sum_{h=1}^L W_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx} \tag{2.19}$$

On solving equations (2.18) and (2.19) we get

$$\lambda_1 = \frac{\{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})\}}{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx})^2}$$

$$\lambda_2 = \frac{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) + \sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})}{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx})^2}$$

On substituting  $\lambda_1, \lambda_2$  in equation (2.17) we get

$$\Omega_h = W_h + \frac{A}{B} + \frac{C}{B}$$

$$\text{where } A = \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}), B = \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx})^2, C = \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) + \sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})$$

On substituting the  $\Omega_h$  in (2.1) we get

$$\bar{y}_{st}(n) = \sum_{h=1}^L W_h (\bar{y}_h + \beta_1 (\bar{X}_h - \bar{x}_h) + \beta_2 (\rho_{hx} - \rho_{hx})) \tag{2.20}$$

where

$$\beta_1 = \frac{W_h Q_h \bar{x}_h \bar{y}_h [\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \bar{x}_h \rho_{hx}^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) / \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h]}{(\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx}^2)^2}$$

$$\beta_2 = \frac{W_h Q_h \rho_{hx}^2 \bar{y}_h [\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx}^2 - \rho_{hx}^2) - \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) / \sum_{h=1}^L W_h Q_h \rho_{hx}^4 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx}^2)^2]}{(\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx}^2)^2}$$

2) Case: 2 The Lagrange function for weights  $\Omega_h$  which satisfy the calibration equation (2.1) and minimize the loss function given in (2.12)

$$\phi_2 = \sum_{h=1}^L \frac{1}{Q_h} \left(\frac{\Omega_h}{W_h} - 1\right)^2 - 2\lambda_1 (\sum_{h=1}^L \Omega_h \bar{x}_h - \sum_{h=1}^L W_h \bar{X}_h) - 2\lambda_2 (\sum_{h=1}^L \Omega_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx})$$

where  $\lambda_1$  and  $\lambda_2$  are the Lagrange's multipliers Differentiate above equation w.r.to  $\Omega_h$  and equating to zero for obtaining the value of  $\Omega_h$

$$\Omega_h = W_h + W_h^2 Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx}) \tag{2.21}$$

From equation (2.15)

$$\sum_{h=1}^L \{W_h + W_h^2 Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx})\} \bar{x}_h = \sum_{h=1}^L W_h \bar{X}_h$$

$$\sum_{h=1}^L \lambda_1 W_h^2 Q_h \bar{x}_h^2 + \sum_{h=1}^L \lambda_2 W_h^2 Q_h \rho_{hx}^2 \bar{x}_h = \sum_{h=1}^L W_h \bar{X}_h - \sum_{h=1}^L W_h \bar{x}_h \tag{2.22}$$

From equation (2.16)

$$\sum_{h=1}^L \{W_h + W_h^2 Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx})\} \rho_{hx} = \sum_{h=1}^L W_h \rho_{hx}$$

$$\sum_{h=1}^L \lambda_1 W_h^2 Q_h \rho_{hx} \bar{x}_h + \sum_{h=1}^L \lambda_2 W_h^2 Q_h \rho_{hx}^2 = \sum_{h=1}^L W_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx} \quad (2.23)$$

On solving equations (2.21) and (2.22) we get

$$\lambda_1 = \frac{\{\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\rho_{hx}^2 - \rho_{hx}^2)\}}{\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h^2 Q_h \bar{x}_h \rho_{hx})^2}$$

$$\lambda_2 = \frac{\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) + \sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})}{\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h^2 Q_h \bar{x}_h \rho_{hx})^2}$$

On substituting  $\lambda_1, \lambda_2$  in equation (2.21) we get

$$\Omega_h = W_h + \frac{A}{B} + \frac{C}{B}$$

$$\text{Where } A = W_h^2 Q_h \bar{x}_h [\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h^2 Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})]$$

$$B = \sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 - \sum_{h=1}^L (W_h^2 Q_h \bar{x}_h \rho_{hx})^2$$

$$C = W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) - \sum_{h=1}^L W_h^2 Q_h \rho_{hx} \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) /$$

On substituting the  $\Omega_h$  in (2.1) we get

$$\bar{y}_{st}(nk) = \sum_{h=1}^L W_h (\bar{y}_h + \beta_1 (\bar{X}_h - \bar{x}_h) + \beta_2 (\rho_{hx} - \rho_{hx}))$$

where

$$\beta_1 = \frac{W_h^2 Q_h \bar{x}_h \bar{y}_h [\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h^2 Q_h \bar{x}_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})]}{\sum_{h=1}^L W_h^2 Q_h \bar{x}_h \rho_{hx}^2}$$

$$\beta_2 = \frac{W_h^2 Q_h \rho_{hx}^2 \bar{y}_h [\sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) - \sum_{h=1}^L W_h^2 Q_h \rho_{hx} \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h)]}{\sum_{h=1}^L W_h^2 Q_h \rho_{hx}^2 \sum_{h=1}^L W_h^2 Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h^2 Q_h \bar{x}_h \rho_{hx})^2}$$

3) Case: 3 The Lagrange function for weights  $\Omega_h$ , which satisfy the calibration equation (2.1) and minimize the loss function given in (2.13)

$$\phi_3 = \sum_{h=1}^L \frac{1}{Q_h} \left( -W_h \log \frac{\Omega_h}{W_h} + \Omega_h - W_h \right)^2 - 2\lambda_1 (\sum_{h=1}^L \Omega_h \bar{x}_h - \sum_{h=1}^L W_h \bar{x}_h) - 2\lambda_2 (\sum_{h=1}^L \Omega_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx})$$

where  $\lambda_1$  and  $\lambda_2$  are the Lagrange's multipliers Differentiate above equation w.r.to  $\Omega_h$  and equating to zero for obtaining the value of  $\Omega_h$

$$\Omega_h = W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx}) \quad (2.24)$$

From equation (2.15)

$$\sum_{h=1}^L \{W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx})\} \bar{x}_h = \sum_{h=1}^L W_h \bar{x}_h$$

$$\sum_{h=1}^L \lambda_1 W_h Q_h \bar{x}_h^2 + \sum_{h=1}^L \lambda_2 W_h Q_h \rho_{hx} \bar{x}_h = \sum_{h=1}^L W_h \bar{x}_h - \sum_{h=1}^L W_h \bar{x}_h \quad (2.25)$$

From equation (2.16)

$$\sum_{h=1}^L \{W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 \rho_{hx})\} \rho_{hx} = \sum_{h=1}^L W_h \rho_{hx}$$

$$\sum_{h=1}^L \lambda_1 W_h Q_h \rho_{hx} \bar{x}_h + \sum_{h=1}^L \lambda_2 W_h Q_h \rho_{hx}^2 = \sum_{h=1}^L W_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx} \quad (2.26)$$

On solving equations (2.25) and (2.26) we get

$$\lambda_1 = \frac{\{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})\}}{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h Q_h \rho_{hx})^2}$$

$$\lambda_2 = \frac{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) + \sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})}{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h Q_h \rho_{hx})^2}$$

On substituting  $\lambda_1, \lambda_2$  in equation (2.24) we get

$$\Omega_h = W_h + \frac{A}{B} + \frac{C}{B}$$

where

$$A = W_h Q_h \bar{x}_h \left[ \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) \right]$$

$$B = \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - \sum_{h=1}^L (W_h Q_h \rho_{hx})^2$$

$$C = \left[ W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) - \sum_{h=1}^L W_h Q_h \rho_{hx} \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) \right]$$

The value of  $\Omega_h$  is equal to  $\chi^2$  distance function (case 1)

4) Case: 4 The Lagrange function for weights  $\Omega_h$ , which satisfy the calibration equation (2.1) and minimize the loss function given in (2.14)

$$\phi_4 = \sum_{h=1}^L 2 \frac{(\sqrt{\Omega_h} - \sqrt{W_h})^2}{Q_h} - 2\lambda_1 (\sum_{h=1}^L \Omega_h \bar{x}_h - \sum_{h=1}^L W_h \bar{X}_h) - 2\lambda_2 (\sum_{h=1}^L \Omega_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx})$$

where  $\lambda_1$  and  $\lambda_2$  are the Lagrange's multipliers Differentiate above equation w.r.to  $\Omega_h$  and equating to zero for obtaining the value of  $\Omega_h$

$$\Omega_h = W_h + W_h Q_h (2\lambda_1 \bar{x}_h + 2\lambda_2 \rho_{hx}) \tag{2.27}$$

Substituting the value of  $\Omega_h$  in equation (2.15) & (2.16)

$$\sum_{h=1}^L \{W_h + W_h Q_h (2\lambda_1 \bar{x}_h + 2\lambda_2 \rho_{hx})\} \bar{x}_h = \sum_{h=1}^L W_h \bar{X}_h$$

$$\sum_{h=1}^L \lambda_1 W_h Q_h \bar{x}_h^2 + \sum_{h=1}^L \lambda_2 W_h Q_h \rho_{hx} \bar{x}_h = \frac{\sum_{h=1}^L W_h \bar{x}_h - \sum_{h=1}^L W_h \bar{x}_h}{2} \tag{2.28}$$

From equation (2.16)

$$\sum_{h=1}^L \{W_h + W_h Q_h (2\lambda_1 \bar{x}_h + 2\lambda_2 \rho_{hx})\} \rho_{hx} = \sum_{h=1}^L W_h \rho_{hx}$$

$$\sum_{h=1}^L \lambda_1 W_h Q_h \rho_{hx} \bar{x}_h + \sum_{h=1}^L \lambda_2 W_h Q_h \rho_{hx}^2 = \frac{\sum_{h=1}^L W_h \rho_{hx} - \sum_{h=1}^L W_h \rho_{hx}}{2} \tag{2.29}$$

On solving equations (2.28) and (2.29) we get

$$\lambda_1 = \frac{\{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h)/4 - \sum_{h=1}^L W_h Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})/4\}}{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx})^2}$$

$$\lambda_2 = \frac{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h)/4 + \sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx})/4}{\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx})^2}$$

Putting the value of  $\lambda_1, \lambda_2$  in equation (2.27)

$$\Omega_h = W_h + \frac{A}{B} + \frac{C}{B}$$

$$A = \frac{1}{2} W_h Q_h \bar{x}_h \left[ \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) \right]$$

$$B = \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - \left( \sum_{h=1}^L W_h \bar{x}_h Q_h \rho_{hx} \right)^2$$

$$C = \frac{1}{2} W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) - \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h)$$

On substituting the  $\Omega_h$  in (2.1) we get

$$\bar{y}_{st}(t) = \sum_{h=1}^L W_h \left( \bar{y}_h + \frac{1}{2} \beta_1 (\bar{X}_h - \bar{x}_h) + \frac{1}{2} \beta_2 (\rho_{hx} - \rho_{hx}) \right) \tag{2.30}$$

where

$$\beta_1 = \frac{W_h Q_h \bar{x}_h \bar{y}_h [\sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) - \sum_{h=1}^L W_h Q_h \rho_{hx} \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) / \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h Q_h \bar{x}_h \rho_{hx}^2)^2]}{(\sum_{h=1}^L W_h Q_h \bar{x}_h \rho_{hx}^2)^2}$$

$$\beta_2 = \frac{W_h Q_h \rho_{hx}^2 \bar{y}_h [\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \sum_{h=1}^L W_h (\rho_{hx} - \rho_{hx}) - \sum_{h=1}^L W_h Q_h \rho_{hx} \bar{x}_h \sum_{h=1}^L W_h (\bar{X}_h - \bar{x}_h) / \sum_{h=1}^L W_h Q_h \rho_{hx}^2 \sum_{h=1}^L W_h Q_h \bar{x}_h^2 - (\sum_{h=1}^L W_h Q_h \bar{x}_h \rho_{hx}^2)^2]}{(\sum_{h=1}^L W_h Q_h \bar{x}_h \rho_{hx}^2)^2}$$

#### IV. SIMULATION STUDY

To study the performance of the proposed estimator we use the following data set

Source of Data: Population, is taken from the census of India 2011(Uttar Pradesh , Series 10, Part 12B and District census hand book, AGRA).

Email: [dco-utp.rgi@censusindia.gov.in](mailto:dco-utp.rgi@censusindia.gov.in) and Website: <http://www.censusindia.gov.in>

1) Population: 1 The considered data relates to total area of 45 villages of Khandauli block at Agra districts (U.P). We consider the numbers of agricultural laborers in villages as study variable *Y* and the total area of villages as auxiliary variable *X*

We divided the whole population of 45 villages is divided in to 5 strata according to the area. Accordingly we have:

Strata	Area in Hectare
1	(1-4400) (21 Villages)
2	(4400-8400) (10 Villages)
3	(8400-12400) (6 Villages)
4	(12400-16500) (5 Villages)
5	(16500-20900) (3 Villages)

Table (1) parametric values of the population (1)

Population	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5
<i>N</i> = 45	<i>N</i> <sub>1</sub> = 21	<i>N</i> <sub>2</sub> = 10	<i>N</i> <sub>3</sub> = 6	<i>N</i> <sub>4</sub> = 5	<i>N</i> <sub>5</sub> = 3
<i>n</i> = 23	<i>n</i> <sub>1</sub> = 10	<i>n</i> <sub>2</sub> = 5	<i>n</i> <sub>3</sub> = 3	<i>n</i> <sub>4</sub> = 3	<i>n</i> <sub>5</sub> = 2
$\bar{Y}$ = 173.508	$\bar{Y}_1$ = 112.09	$\bar{Y}_2$ = 175.9	$\bar{Y}_3$ = 149.83	$\bar{Y}_4$ = 232.2	$\bar{Y}_5$ = 545
$\bar{X}$ = 463.37	$\bar{X}_1$ = 196.4	$\bar{X}_2$ = 413.411	$\bar{X}_3$ = 672.33	$\bar{X}_4$ = 844.26	$\bar{X}_5$ = 1445.97

2) Population:2 For this we considered the 2011 census data which is relates to the total number of agricultural laborers, total area total population, and total numbers of cultivators of 55 villages of Etmadpur block of Agra districts (U.P). We take the numbers of Agricultural laborers in villages as *Y*, the total area of villages as auxiliary variable *X*. The whole population of 55 villages stratified in to 5 strata according to the Area. So the strata become as under

Strata	Area in Hectare
1	(1-3647) (31 Villages)
2	(3647-7222) (11 Villages)
3	(7222-9973) (5 Villages)
4	(9973-12307) (3 Villages)
5	(12307-18173) (5 Villages)

Table 2. Parametric values of the population (2)

Population	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5
<i>N</i> = 55	<i>N</i> <sub>1</sub> = 31	<i>N</i> <sub>2</sub> = 11	<i>N</i> <sub>3</sub> = 5	<i>N</i> <sub>4</sub> = 3	<i>N</i> <sub>5</sub> = 5
<i>n</i> = 31	<i>n</i> <sub>1</sub> = 15	<i>n</i> <sub>2</sub> = 5	<i>n</i> <sub>3</sub> = 3	<i>n</i> <sub>4</sub> = 2	<i>n</i> <sub>5</sub> = 3
$\bar{Y}$ = 147.4	$\bar{Y}_1$ = 72.38	$\bar{Y}_2$ = 158.81	$\bar{Y}_3$ = 194.4	$\bar{Y}_4$ = 250.6	$\bar{Y}_5$ = 478.2
$\bar{X}$ = 330.43	$\bar{X}_1$ = 117.67	$\bar{X}_2$ = 324.96	$\bar{X}_3$ = 550.23	$\bar{X}_4$ = 777.97	$\bar{X}_5$ = 1173.2

We calculated empirical mean square error and relative efficiency using following formulas:

$$MSE \bar{y}_{st}(\alpha) = \frac{1}{50} \sum_{i=1}^{50} \left[ \sum_{h=1}^5 W_h (\bar{y}_h + \beta_1 (\bar{X}_h - \bar{x}_h) + \beta_2 (\rho_{hx} - \rho_{hx})) - \bar{Y} \right]^2$$

where  $\alpha = n, nk, t$

Table: 3 Mean square error of estimators

Estimators	Population 1	Population 2
$\bar{y}_{st}(n)$	32068843	24677991
$\bar{y}_{st}(nk)$	336977041	99861561
$\bar{y}_{st}(t)$	9395651	20555521

It should be mentioned that distance function  $L_{3p}$  is equal to  $L_{1p}$ . The simulation study shows that calibration estimator using distance measure  $L_{4p}$  are more efficient than using distance measure  $L_{2p}, L_{1p}$ .

### V. CONCLUSION

In this study we worked on new weights using different distance measures in stratified sampling. The performance of distance measures are compared with a simulation study.

### REFERENCES

- [1] Arnab, R., Singh, S., (2005), A note on variance estimation for the generalized regression predictor, Australian and New Zealand Journal of Statistics, 47, 2, 231–234.
- [2] Clement, E. P., and E. I. Enang. (2017), on the efficiency of ratio estimator over the regression estimator, communications in Statistics: Theory and Methods, 46 (11), 5357–67. doi:10.1080/03610926.2015.1100741.
- [3] Deville, J.C., Sarndal, C.E., (1992), Calibration estimators in survey sampling, Journal of the American Statistical Association, 87, 376-382.
- [4] Estevao, V.M., Sarndal, C.E. (2000), a functional form approach to calibration, Journal of Official Statistics, 16, 379-399.
- [5] Farrell, P.J., Singh, S. (2005), model assisted higher order calibration of estimators of variance, Australian and New Zealand Journal of Statistics, 47, 3, 375–383.
- [6] Kim, J. K., Park, M. (2010), calibration estimation in survey sampling, international statistical review, 78, 1, 21-29.
- [7] Kim, J.M., Sungur, E.A., Heo T.Y. (2007), calibration approach estimators in stratified sampling, Statistics and Probability Letters, 77, 1, 99-103.
- [8] Koyuncu, N. (2012), Application of Calibration Method to Estimators in Sampling Theory, Hacettepe University Department of Statistics, PhD. Thesis.
- [9] Koyuncu, N., and C. Kadilar. (2013), Calibration estimator using different distance measures in stratified random sampling, International Journal of Modern Engineering Research, 3 (1), 415–419.
- [10] Koyuncu, N., and C. Kadilar. (2016), Calibration weighting in stratified random sampling. Communications in Statistics: Simulation and Computation, 45 (7), 2267–75. doi:10.1080/03610918.2014.901354.
- [11] Mouhamed, A. M., A. A. Ei-Sheikh, and H. A. Mohamed, (2015), A new calibration estimator of stratified random sampling, Applied Mathematical Sciences, 9 (35),1735–1744. doi:10.12988/ams.2015.514.
- [12] Neha Garg & Menakshi Pachori (2020) Use of coefficient of variation in calibration estimation of population mean in stratified sampling, Communications in Statistics Theory and Methods, 49:23, 5842-5852,
- [13] Nidhi, Sisodia, B. V. S., Singh, S. Singh. and S. K. (2017), Calibration approach estimation of the mean in stratified sampling and stratified double sampling, Communications in Statistics: Theory and Methods, 46 (10),4932–4942, doi:10.1080/03610926.2015.1091083.
- [14] Pumputis, D. (2005), calibrated estimators under different distance measures. Proceedings of the Workshop on Survey Sampling Theory and Methodology, 137–141.
- [15] Rai, P.K., Singh, A., Qasim, M. (2020), calibration based Estimators using different distance measures under two Auxiliary Variables: A Comparative Study , Journal of Modern Applied Statistical Methods, Vol. 19, No. 1
- [16] Sarndal, C. E., B. Swensson, and J. Wretman (1992), Model assisted survey sampling, New York, Springer Verlag.
- [17] Singh, S., and R. Arnab.(2011), On Calibration of design weights, METRON: International Journal of Statistics, LXIX 69 (2),185–205.,doi:10.1007/BF03263556.
- [18] Tracy, D. S., S. Singh, and R. Arnab.(2003), Note on calibration in stratified and double sampling, Survey Methodology, 29 (1),99–104.
- [19] Kris Sankaran, (2026), Data Science Principles for Interpretable and Explainable AI, Statistical aspects of Trustworthy Machine Learning, Volume 24, Issue 1 pp. 26–52.

**SPECIAL ISSUE**

Appendix

R Calculations

```
install.packages("xlsx")
library(xlsx)
library(readxl)
N1<-21;N2<-10;N3<-6;N4<-5;N5<-3
N<-45
n1<-10;n2<-5;n3<-3;n4<-3;n5<-2
n<-23
Q<-1;W1<-N1/N;W2<-N2/N;W3<-N3/N;W4<-N4/N;W5<-N5/N
data<-read_excel("C:/Users/dell pc/Desktop/new data.xlsx")
head(data)
data1<-as.data.frame(data)
newdata<-data.frame(data1,g=c(rep(1,21),rep(2,10),rep(3,6),rep(4,5),rep(5,3)))
dim(data1)
head(newdata)
X<-split(newdata,newdata$g)
y<-lapply(seq_along(X), function(x)as.data.frame(X[[x]]),1:2)
A1<-y[[1]]
A2<-y[[2]]
A3<-y[[3]]
A4<-y[[4]]
A5<-y[[5]]
X1<-mean(A1$X);X2<-mean(A2$X);X3<-mean(A3$X);X4<-mean(A4$X);X5<-mean(A5$X)
S1X<-sd(A1$X);S2X<-sd(A2$X);S3X<-sd(A3$X);S4X<-sd(A4$X);S5X<-sd(A5$X)
y1<-mean(A1$Y);y2<-mean(A2$Y);y3<-mean(A3$Y);y4<-mean(A4$Y);y5<-mean(A5$Y)
S1y<-sd(A1$Y);S2y<-sd(A2$Y);S3y<-sd(A3$Y);S4y<-sd(A4$Y);S5y<-sd(A5$Y)
Mys<-W1*y1+W2*y2+W3*y3+W4*y4+W5*y5
S1XY<-cov(A1$X,A1$Y);S2XY<-cov(A2$X,A2$Y);S3XY<-cov(A3$X,A3$Y);S4XY<-cov(A4$X,A4$Y);S5XY<-
cov(A5$X,A5$Y)
r1<-cor(A1$X,A1$Y);r2<-cor(A2$X,A2$Y);r3<-cor(A3$X,A3$Y);r4<-cor(A4$X,A4$Y);r5<-cor(A5$X,A5$Y)
ybart<-NA; MSEt<-NA; ybartn<-NA; MSEn<-NA; ybartnk<-NA; MSEnk<-NA

for(i in 1:50){
sam<-sample(1:21,10,replace=F)
sam1<-sample(1:10,5,replace=F)
sam2<-sample(1:6,3,replace=F)
sam3<-sample(1:5,3,replace=F)
sam4<-sample(1:3,2,replace=F)
sam11<-sample(A1$X,10);sam12<-sample(A2$X,5);sam13<-sample(A3$X,3);sam14<-sample(A4$X,3);sam15<-sample(A5$X,2)
x11<-mean(sam11);x12<-mean(sam12);x13<-mean(sam13);x14<-mean(sam14);x15<-mean(sam15)
x11s<-sd(sam11);x12s<-sd(sam12);x13s<-sd(sam13);x14s<-sd(sam14);x15s<-sd(sam15)
sam21<-sample(A1$Y,10);sam22<-sample(A2$Y,5);sam23<-sample(A3$Y,3);sam24<-sample(A4$Y,3);sam25<-sample(A5$Y,2)
y11<-mean(sam21);y12<-mean(sam22);y13<-mean(sam23);y14<-mean(sam24);y15<-mean(sam25)
r11<-cor(sam11,sam21);r12<-cor(sam12,sam22);r13<-cor(sam13,sam23);r14<-cor(sam14,sam24);r15<-cor(sam15,sam25)
Q1<-Q*W1*r11^4+Q*W2*r12^4+Q*W3*r13^4+Q*W4*r14^4+Q*W5*r15^4
Q2<-Q*W1*y11*x12+Q*W2*y21*x22+Q*W3*y31*x32+Q*W4*y41*x42+Q*W5*y51*x52
Q3<-Q*W1*r11^2*x12+Q*W2*r12^2*x22+Q*W3*r13^2*x32+Q*W4*r14^2*x42+Q*W5*r15^2*x52
```

$$\begin{aligned}
 Q4 &<- Q*W1*r11^2*y11+Q*W2*r12^2*y21+Q*W3*r13^2*y31+Q*W4*r14^2*y41+Q*W5*r15^2*y51 \\
 Q5 &<- Q*W1*x12^2+Q*W2*x22^2+Q*W3*x32^2+Q*W4*x42^2+Q*W5*x52^2 \\
 Q6 &<- Q*W1*r11^2+Q*W2*r12^2+Q*W3*r13^2+Q*W4*r14^2+Q*W5*r15^2 \\
 b1n[i] &<- (Q2*Q1-Q3*Q2)/(Q5*Q1-Q6^2) \\
 b2n[i] &<- (Q5*Q4-Q4*Q3)/(Q5*Q1-Q6^2) \\
 ybartn[i] &<- W1*y11+W2*y21+W3*y31+W4*y41+W5*y51+b1n[i]*(W1*(X1-x12)+(W2*(X2-x22))+(W3*(X3-x32))+(W4*(X4-x42))+(W5*(X5-x52)))+b2n[i]*(W1*(r1^2-r11^2)+W2*(r2^2-r12^2)+W3*(r3^2-r13^2)+W4*(r4^2-r14^2)+W5*(r5^2-r15^2)) \\
 MSEn[i] &<- (ybart[i]-Mys)^2 \\
 ybart[i] &<- W1*y11+W2*y12+W3*y13+W4*y14+W5*y15+1/2*(b1n[i]*(W1*(X1-x11)+(W2*(X2-x12))+(W3*(X3-x13))+(W4*(X4-x14))+(W5*(X5-x15)))+b2n[i]*(W1*(r1^2-r11^2)+W2*(r2^2-r12^2)+W3*(r3^2-r13^2)+W4*(r4^2-r14^2)+W5*(r5^2-r15^2))) \\
 MSEt[i] &<- (ybart[i]-Mys)^2 \\
 Q7 &<- Q*W1^2*r11^2+Q*W2^2*r12^2+Q*W3^2*r13^2+Q*W4^2*r14^2+Q*W5^2*r15^2 \\
 Q8 &<- Q*W1^2*y11*x12+Q*W2^2*y21*x22+Q*W3^2*y31*x32+Q*W4^2*y41*x42+Q*W5^2*y51*x52 \\
 Q9 &<- Q*W1^2*r11^2*x12+Q*W2^2*r12^2*x22+Q*W3^2*r13^2*x32+Q*W4^2*r14^2*x42+Q*W5^2*r15^2*x52 \\
 Q10 &<- Q*W1^2*r11^2*y11+Q*W2^2*r12^2*y21+Q*W3^2*r13^2*y31+Q*W4^2*r14^2*y41+Q*W5^2*r15^2*y51 \\
 Q11 &<- Q*W1^2*x12^2+Q*W2^2*x22^2+Q*W3^2*x32^2+Q*W4^2*x42^2+Q*W5^2*x52^2 \\
 Q12 &<- Q*W1^2*r11^2+Q*W2^2*r12^2+Q*W3^2*r13^2+Q*W4^2*r14^2+Q*W5^2*r15^2 \\
 b1nk[i] &<- (Q8*Q7-Q9*Q8)/(Q11*Q7-Q12^2) \\
 b2nk[i] &<- (Q11*Q10-Q10*Q9)/(Q11*Q7-Q12^2) \\
 ybartnk[i] &<- W1*y11+W2*y21+W3*y31+W4*y41+W5*y51+b1nk[i]*(W1*(X1-x12)+(W2*(X2-x22))+(W3*(X3-x32))+(W4*(X4-x42))+(W5*(X5-x52)))+b2nk[i]*(W1*(r1^2-r11^2)+W2*(r2^2-r12^2)+W3*(r3^2-r13^2)+W4*(r4^2-r14^2)+W5*(r5^2-r15^2)) \\
 MSEnk[i] &<- (ybartnk[i]-Mys)^2
 \end{aligned}$$



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)