iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IX **Month of publication:** September 2026

DOI: https://doi.org/10.22214/ijraset.2026.84795

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IX Sep 2026- Available at www.ijraset.com

An Improved Data Science Based Ratio-Type Estimator for Population Mean in Stratified Random Sampling Using Multi-Auxiliary Information

Anupama Bhadauriya[1], Rakesh Prasad Sarang[2], Gajendra Singh Rajput[3]
[1, 2, 3]*Assistant Professor, School of Computing & Intelligent Systems (SOCIS), ITM University Gwalior Madhya Pradesh- Pin Code-474001,India*

Abstract: This article proposed an improved data science ratio-type estimator for estimating the population mean of a study variable using multiple auxiliary variables in stratified sampling. In this paper, we present a comprehensive view on data science including various types of estimating advanced analytics and population methods that can be applied multiple auxiliary and capabilities. The theoretical properties of the proposed estimator are investigated, and its bias and mean square error (MSE) are derived under suitable assumptions. The efficiency of the proposed estimator is compared with that of the estimator proposed by Sharma and Kumar. The relative efficiency of the estimators is examined both theoretically and numerically. A numerical illustration is also presented to demonstrate the performance of the proposed estimator and to assess its efficiency in comparison with the existing estimator. The results indicate the effectiveness of the proposed estimator in improving the precision of population mean estimation in stratified sampling when multiple auxiliary variables.
Keywords: Ratio Estimator, Multi Auxiliary Variables, Bias, MSE.

I. INTRODUCTION

The use of auxiliary information can increase or improve the precision of an estimator when study variable is highly correlated with auxiliary variable. The role of auxiliary information is most important in sampling survey. Auxiliary variable are commonly used in survey sampling in order to obtain improved designs and achieves higher precision in the estimates of some population parameter such as population mean, variance, population proportion and ratio of two population means are common in agriculture, economics, medicine and population study. Survey sampling describes various methods in literature under auxiliary variable at the estimation stage. This includes linear regression estimator, ratio estimator, product estimator, and difference estimator. The used of auxiliary information at the estimation stage and proposed a ratio estimator for the population mean. A ratio estimator is preferred when the correlation Coefficient between the study variate and the auxiliary variate is positive.

Kadilar and Cingi [5], Shabbir and Gupta [10], Haq and Shabbir have proposed estimators in stratified sampling using information on single Auxiliary variable [4].

Koyuncu, Kadilar andTailor et al. [5,13] proposed estimators of population mean using two auxiliary variables in stratified random sampling. The proposed improved estimator of population means using two auxiliary variables in stratified sampling.

Bahl and Kumar [1,2] proposed ratio and product type estimator for estimating the mean of the finite population using multi auxiliary variable. Sharma and Kumar proposed estimator using multi auxiliary variables in stratified sampling [11].

In this article Kadilar and cingi [5] estimator is used to define a new estimator using multi auxiliary variables and we have made comparison of our estimator with which is based on multi auxiliary variables.

Main aim of this article is to propose an Improved ratio type estimator for estimating the population mean of the study variable in stratified random sampling using multi auxiliary variable. Section 2 is devoted for the notation used in the paper, in Section 3 discussed the existing estimator [11]. In Section 4 the proposed estimator is derived along with its properties and optimized value. In Section 5 Comparison of the proposed estimator with existing estimator is being made and condition under which the proposed estimators perform in terms of efficiency is obtained. In Section 6 The numerical illustration is supported to the theoretical result. Paper is concluded in Section 7.

II. NOTATIONS

Consider a finite population $U = (U_1, U_2, \ldots\ldots U_n)$ of size N which are partitioned into l homogenous subgroups, called strata, such that h^{th} strata consist of N_h units, where $h = 1,2, \ldots, l$ and $\sum_{h=1}^{l} N_h = N$. From each stratum, we select a sample of size n_h, by simple random sampling without replacement such that $\sum_{h=1}^{l} n_h = n$, here n_h is the h^{th} stratum sample size. Let Y be the study variable and x_i, $(i = 1,2, \ldots\ldots, k)$are the auxiliary variables taking value y_{hj} and x_{hij}, $(h = 1,2, \ldots\ldots, l; j = 1,2, \ldots N_h)$. then, we define

$\bar{x}_{hi} = \frac{1}{n_h}\sum_{j=1}^{n_h} x_{hij}$; Sample mean of the observation taken from h^{th} stratum for variable x_i.

$\bar{y}_{h} = \frac{1}{n_h}\sum_{j=1}^{n_h} y_{hj}$; Sample mean of the observation taken from h^{th} stratum for variable y.

$\bar{X}_{hi} = \frac{1}{N_h}\sum_{j=1}^{N_h} X_{hij}$; h^{th} stratum mean of auxiliary variable x_i

$\bar{Y}_{h} = \frac{1}{N_h}\sum_{j=1}^{N_h} Y_{hj}$; h^{th} stratum mean for study variable y

$\bar{X}_i = \sum_{h=1}^{l} w_h \bar{X}_{hi}$; Population mean of the auxiliary variable x_i

$\bar{Y} = \sum_{h=1}^{l} w_h \bar{Y}_h$; Population mean of the study variable y

$\bar{x}_{sti} = \sum_{h=1}^{l} w_h \bar{x}_{hi}$; Unbiased estimator of the population mean $\bar{X}_i$

$\bar{y}_{st} = \sum_{h=1}^{l} w_h \bar{y}_h$; Unbiased estimator of the population mean $\bar{Y}$,

$S^2_{yh} = \frac{1}{N_h-1}\sum_{j=1}^{N_h}(Y_{hj} - \bar{Y}_h)^2$;Population mean square of h^{th} stratum for the study variable y

$S^2_{xhi} = \frac{1}{N_h-1}\sum_{j=1}^{N_h}(x_{hij} - \bar{X}_{hi})^2$; Population mean square of h^{th} stratum for the auxiliary variable x_i

$w_h = \frac{N_h}{N}$; Stratum weight of the h^{th} stratum.

III. REVIEW OF LITERATURE

Bahl and Tuteja proposed ratio- and product-type estimators for estimating the population mean by utilizing information from a single auxiliary variable. In reality such situations do occur when information is available in the form of auxiliary variable, which is highly correlated with study variable, for example, number of trees in an orchard and the yield of fruits[14].

However application of this technique presupposes the knowledge of strata size and the availability of sampling frames within strata. But in many practical situations, when we want to estimate the population mean $\boldsymbol{Y}$ of a variate y and consider it desirable to stratify the population, consisting of N units on the basis of the value of auxiliary character x but the information of x is unknown, the sampling frame for various strata and the strata weight $\boldsymbol{Wh} = \boldsymbol{Nh}\ \boldsymbol{N}$ (h=1,2,……L) are not known although the strata may be fixed in advance [9,10].

Simple random sampling technique has some shortcomings like less represen-tative of different sections of the population, administrative inconvenience and less efficiency in case of heterogeneous population. Literature reveals that ratio-cum-product estimator performs better than ratio and product type estimators in simple random sampling under certain conditions[15].

Subsequently, Bahl and Kumar [1,2] extended this work by developing ratio- and product-type estimators under simple random sampling for estimating the finite population mean using multiple auxiliary variables.

$$\hat{\bar{Y}}_{ReM} = \bar{y}\sum_{i=1}^{k} w_i \exp\left(\frac{\bar{X}_i - \bar{x}_i}{\bar{X}_i + \bar{x}_i}\right) \quad (1)$$

$$\widehat{\bar{Y}}_{PeM} = \bar{y}\sum_{i=1}^{k} w_i \exp\left(\frac{\bar{X}_i - \bar{X}_i}{\bar{X}_i + \bar{x}_i}\right) \quad (2)$$

And $w_1, w_2, \ldots\ldots, w_t$; are weights assigned to auxiliary variable such that $\sum_{i=1}^{k} w_i = 1$,
Sharma and Kumar (2020) proposed estimator in stratified sampling using multiple auxiliary variables for estimating the population mean $\bar{Y}$ as

$$t_{Rst} = \bar{y}_{st}\sum_{i=1}^{k} \lambda_i \exp\left(\frac{\bar{X}_i - \bar{x}_{sti}}{\bar{X}_i + \bar{x}_{sti}}\right) \quad (3)$$

$\lambda_1, \lambda_2, \ldots\ldots, \lambda_k$; are weights assigned to auxiliary variable such that $\sum_{i=1}^{k} \lambda_i = 1$,

The bias and MSE of above esimator is written as

$$B(t_{RSt}) = \bar{Y}\sum_{h=1}^{l} w_h^2 \delta_h \sum_{i=1}^{k} \lambda_i \frac{1}{\bar{X}_i}\left(\frac{3}{8} R_i S_{xhi}^2 - S_{yxhi}\right) \quad (4)$$

$$MSE(t_{Rst}) = \sum_{h=1}^{l} w_h{}^2 \delta_h (S_{yh}^2 + \frac{1}{4}\sum_{i=1}^{k} \lambda_i{}^2 R_i^2 S_{xhi}^2 - \sum_{i=1}^{k} \lambda_i R_i S_{yxhi}) \quad (5)$$

Differentiating (5) with respect to λ_i and equating to zero hence we have optimum value of λ_i as

$$\lambda_i = \frac{2S_{yxhi}}{R_i S_{xhi}^2} \quad (6)$$

Putting this value of λ_i in equation (5)

$$MSE(t_{Rst}) = \sum_{h=1}^{l} w_h{}^2 \delta_h (S_{yh}^2 - \sum_{i=1}^{k} \frac{S_{yxhi}^2}{S_{xhi}^2}) \quad (7)$$

In SRS Prasad (1989) proposed a ratio estimator as:

$$\bar{y}_p = K\bar{y}_R = K\frac{\bar{y}}{\bar{x}}\bar{X}$$

Where the coefficient K=$\frac{1+\gamma\rho C_y C_x}{\gamma C_y^2+1}$

In stratified Random Sampling, Kadilar and Cingi (2005) Suggested a ratio estimator

$\bar{y}_{stp} = K^{*}\bar{y}_{Rc}$

$$\bar{y}_{stp} = K^{*}\frac{\bar{y}_{st}}{\bar{x}_{st}}\bar{X} \quad (8)$$

IV. THE PROPOSED ESTIMATOR

Following estimator we proposed an estimator using multiple auxiliary variable defined as:
Kadilar and cingi (2005) estimator is used to define a new estimator using multi auxiliary variables.
We modify Kadilar and cingi [5,6] estimator using multi auxiliary variable

$$t_{stp} = Q\bar{y}_{st}\sum_{i=1}^{k} \lambda_i \frac{\bar{X}_i}{\bar{x}_{sti}} \quad (9)$$

Where λ_i $(i = 1,2,\ldots.k)$ are weight assigned to auxiliary variable such that $\sum_{i=1}^{k} \lambda_i = 1$

A. Bias and MSE of Proposed Estimator

Expanding the right hand side of (9) in terms of e's neglecting the term having power greater than two

$$t_{stp} = Q\bar{Y}(1+e_0)\sum_{i=1}^{k}\lambda_i \frac{\bar{X}_i}{\bar{X}_i(1+e_i)}$$

$$t_{stp} - \bar{Y} = Q\bar{Y} - \bar{Y} + \sum_{i=1}^{k}\lambda_i Q\bar{Y}\left(-e_i + \frac{e_i^2}{2} + e_0 - e_0 e_i + \frac{e_0 e_i^2}{2}\right) \tag{10}$$

Taking Expectation on both side of (10) and obtain the bias of t_{stp}

$$B(t_{stp}) = \bar{Y}(Q-1) + Q\bar{Y}\sum_{i=1}^{k}\lambda_i\left(\sum_{h=1}^{l} w_h^2\delta_h \frac{S_{xhi}^2}{\bar{X}_i^2} - \sum_{h=1}^{l}\frac{w_h^2\delta_h S_{yxhi}}{\bar{Y}\bar{X}_i}\right)$$

$$B(t_{stp}) = \bar{Y}(Q-1) + Q\bar{Y}\sum_{i=1}^{k}\frac{\lambda_i}{\bar{X}_i}\left(R_i S_{xhi}^2 - S_{yxhi}\right) \tag{11}$$

Squaring both side of equation (10) and neglecting term of e's having power greater than two, we have

$$\left(t_{stp} - \bar{Y}\right)^2 = \left[Q\bar{Y} - \bar{Y} + \sum_{i=1}^{k}\lambda_i Q\bar{Y}\left(-e_i + \frac{e_i^2}{2} + e_0 - e_0 e_i + \frac{e_0 e_i^2}{2}\right)\right]^2$$

$$E\left(t_{stp} - \bar{Y}\right)^2 = (Q-1)^2\bar{Y}^2 + Q\bar{Y}^2\sum_{i=1}^{k}\lambda_i{}^2 E\left(e_i^2 + e_0^2 - 2e_i e_0\right)$$

$$E\left(t_{stp} - \bar{Y}\right)^2 =$$

$$(Q-1)^2\bar{Y}^2 + Q\bar{Y}^2\sum_{i=1}^{k}\lambda_i{}^2\left[\sum_{h=1}^{l}\frac{w_h^2\delta_h S_{xhi}^2}{\bar{X}_i^2} \quad \sum_{h=1}^{l}\frac{w_h^2\delta_h S_{yh}^2}{\bar{Y}^2} - 2\sum_{h=1}^{l} w_h^2\frac{\delta_h S_{yxhi}}{\bar{Y}\bar{X}_i}\right]$$

$$MSE\left(t_{stp}\right) = (Q-1)^2\bar{Y}^2 + Q^2\left[\sum_{h=1}^{l} w_h^2\delta_h\left(S_{yh}^2 + \sum_{i=1}^{k}\lambda_i^2 R_i^2 S_{xhi}^2 - 2\sum_{i=1}^{k}\lambda_i R_i S_{yxhi}\right)\right] \tag{12}$$

B. Optimum value of K^* and λ_i

Differentiate equation (12) w.r.to $K^*$and equating to zero for obtaining the optimum value of Q

$$Q = \frac{\bar{Y}^2}{\bar{Y}^2 + \left[\sum_{h=1}^{l} w_h^2\delta_h\left(S_{yh}^2 + \sum_{i=1}^{k}\lambda_i^2 R_i^2 s_{xh}^2 - 2\sum_{i=1}^{k}\lambda_i R_i S_{yxhi}\right)\right]} \tag{13}$$

Again Differentiate equation (12) w.r.to $\lambda_i$and equating to zero for obtaining the optimum value of λ_i

$$\lambda_i = \frac{S_{yxhi}}{R_i S_{xhi}^2} \tag{14}$$

Put this value of λ_i in equation (12), the minimum value of MSE is given by

$$MSE\left(t_{stp}\right)_{min} = (Q-1)^2\bar{Y}^2 + Q^2\sum_{h=1}^{l}\left[w_h^2\delta_h\left\{S_{yh}^2 - \sum_{i=1}^{k}\left(\frac{S_{yxhi}^2}{S_{xhi}^2}\right)\right\}\right] \tag{15}$$

V. EFFICIENCY COMPARISONS

On comparing the MSE of our proposed estimator (t_{stp}) with Sharma and Kumar (2020) estimator that of we obtained the following condition

$$\text{Put } A = \sum_{h=1}^{l}\left[w_h^2\delta_h\left\{S_{yh}^2 - \sum_{i=1}^{k}\left(\frac{S_{yxhi}^2}{S_{xhi}^2}\right)\right\}\right]$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IX Sep 2026- Available at www.ijraset.com

$\mathrm{MSE}(t_{stp}) < MSE(t_{Rst})$
$A < (Q-1)^2\bar{Y}^2 - Q^2A$
$A - (Q^*-1)^2\bar{Y}^2 - Q^2A > 0$
$Q^2A - A + (Q-1)^2\bar{Y}^2 > 0$
$A(Q^2-1) + (Q^*-1)^2\bar{Y}^2 > 0$
$(Q^*-1)[(Q+1) + (Q-1)\bar{Y}^2] > 0$
From this condition it is clear that if
$$Q > \frac{\bar{Y}^2 - A}{\bar{Y}^2 + A} \tag{16}$$

The suggested estimator is more efficient than the Sharma and Kumar (2020) estimator. When we examine the condition (16) in detail, we see that this condition is always satisfied. Therefore, we can say that the suggested estimator is more efficient than existing estimator in all conditions [12,13].

VI. NUMERICAL EXAMPLE

Source of data: The population, is taken from the Census of India 2011(Uttar Pradesh , Series 10, Part 12B and District census hand book, AGRA).
Email:dcoutp.rgi@censusindia.gov.inandWebsite:http:/www.censusindia.gov.in
Population:1 For this we considered the 2011census Data which is relates to the total number of agricultural laborers, total area total population, and total numbers of cultivators of 45 villages of Khandauli block of Agra districts (U.P). We take the numbers of Agricultural laborers in villages as y, the total area of villages as x_1, the total population of villages as x_2, and the total numbers of Cultivators as $x_{3.}$
The whole population of 45 villages stratified in to 5 strata according to the Area. So the strata become as under

Table 1

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

Table2. Data Statistics

Stratum no.	$\bar{Y}_h$	$\bar{X}_{h1}$	$\bar{X}_{h2}$	$\bar{X}_{h3}$	N_h	n_h	W_h
1	112.09	196.4	1973.28	169.76	21	10	0.467
2	175.9	413.411	3799.3	270.8	10	5	0.23
3	149.83	672.3316	4026.33	327.67	6	3	0.14
4	232.2	844.266	5129.2	389.2	5	3	0.12
5	545	1445.97	15594	545	3	2	0.067

$$\bar{X}_1 = 463.3719911, \bar{X}_2 = 3911.508, \bar{X}_3 = 262.6662222, \bar{Y} = 173.5082222, K^* = 0.995300903$$
Where $\bar{X}_1, \bar{X}_2, \bar{X}_3$ Population mean of the auxiliary variable $x_i; i = 1,2,3$

Population:2 For this we considered the 2011census 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 x_1, the total population of villages as x_2, and the total numbers of Cultivators as $x_{3.}$
The whole population of 55 villages stratified in to 5 strata according to the Area. So the strata become as under:

Table 3

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

Table4. Data Statistics

Stratum no.	$\overline{Y}_h$	$\overline{X}_{h1}$	$\overline{X}_{h2}$	$\overline{X}_{h3}$	N_h	n_h	W_h
1	72.38	117.67	1217.06	117.61	31	15	0.567
2	158.81	324.96	2452.54	214.27	11	5	0.2
3	194.4	550.23	2855	288	5	3	0.0909
4	250.6	777.97	5089.66	413.57	3	2	0.545
5	478.2	1173.2	8952.2	713	5	3	0.0909

$$\bar{X}_1 = 330.425, \overline{X}_2 = 2527.49, \overline{X}_3 = 222.7016, \bar{Y} = 147.37, K^* = 0.9982$$

Where $\bar{X}_1, \overline{X}_2$, $\overline{X}_3$ Population mean of the auxiliary variable x_i; $i = 1,2,3$

Table 5 MSE of population 1 and 2

Population	Existing Estimators (t_{rst})	Proposed Estimators (t_{stp})
population 1	142.1346985	141.4667938
population 2	39.98	39.90

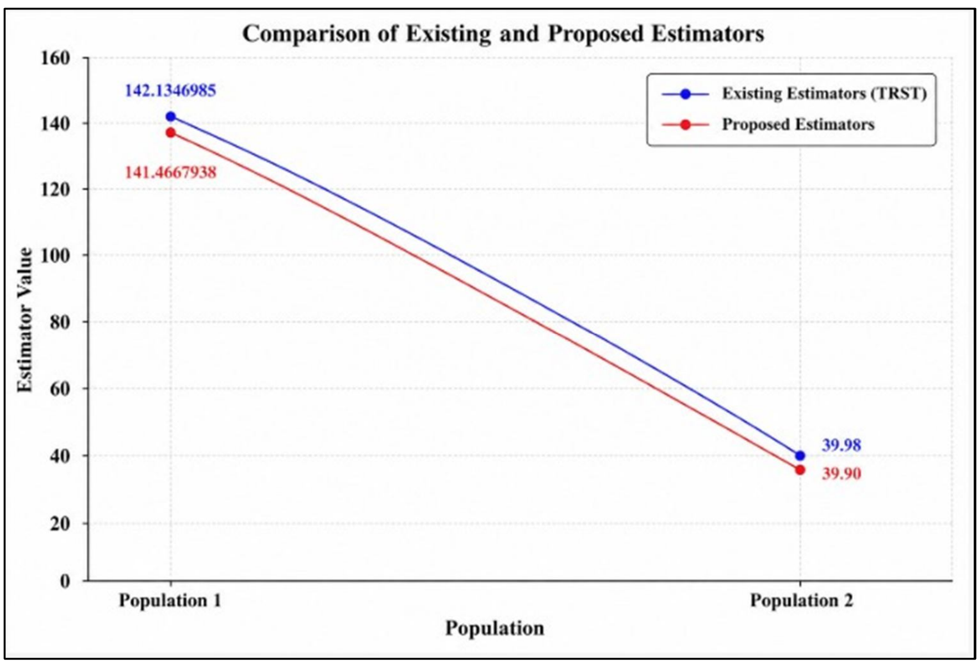


Fig. 5.1 MSE of population

From these value, it is seen that the MSE value of the proposed estimator is smaller than existing estimator. It is an expected result from population 1 since $Q = 0.995300903 > \frac{\bar{Y}^2 - A}{\bar{Y}^2 + A}$ =0.990601806 as mentioned in sec 5.

It is an expected result from population 2 $Q = 0.9981 > \frac{\bar{Y}^2 - A}{\bar{Y}^2 + A}$ =0.9963 as mentioned in sec 5.

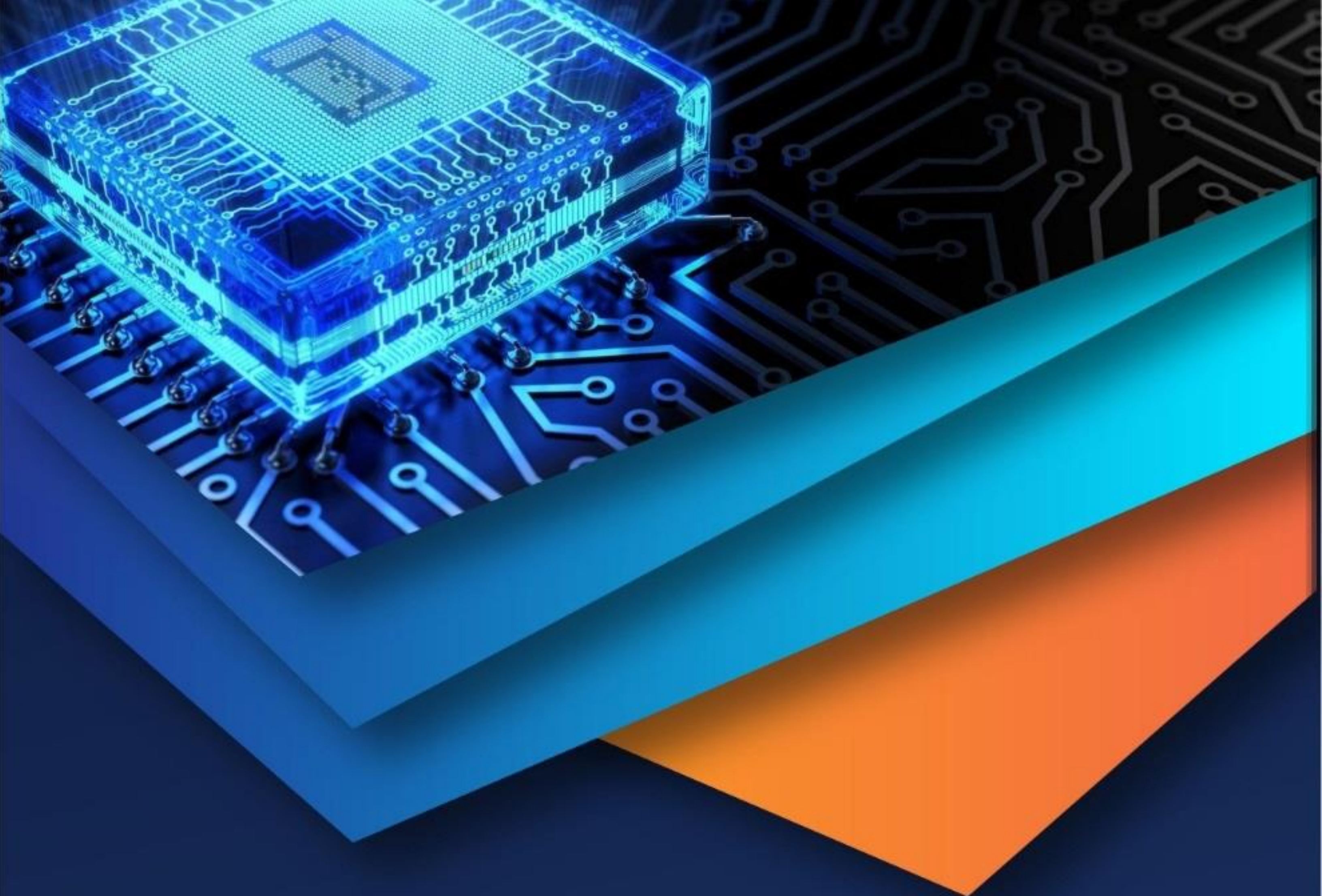
VII. CONCLUSIONS

An improved ratio-type estimator for estimating the population mean in stratified random sampling has been proposed by utilizing multiple auxiliary variables. The proposed estimator is developed based on the estimator of Kadilar and Cingi (2005), and its mean square error (MSE) is derived theoretically. The efficiency of the proposed estimator is then compared with that of the estimator proposed by Sharma and Kumar (2020) through their respective MSE expressions. The theoretical comparison demonstrates that the proposed estimator has a lower MSE than the Sharma and Kumar (2020) estimator under the considered conditions, indicating its superior efficiency.

The theoretical findings are further supported by a numerical illustration based on two populations. For Population 1, the value of the existing estimator is 142.1347, whereas the proposed estimator gives a lower value of 141.4668. Similarly, for Population 2, the existing estimator has a value of 39.9800, while the proposed estimator gives 39.9000. These numerical results are consistent with the theoretical findings and demonstrate the improved performance and efficiency of the proposed estimator over the existing estimator.

REFERENCES

[1] Bahl, S., & Kumar, M. (2000). Estimation of finite population mean using multi-auxiliary variables. Journal of Statistical Management Systems, 3(1), 67–74.
[2] Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimator. Journal of Information and Optimization Sciences, 12(1), 159–164.
[3] Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.
[4] Haq, A., & Shabbir, J. (2013). Improved family of ratio estimators in simple and stratified random sampling. Communications in Statistics—Theory and Methods, 42(5), 782–799.
[5] Kadilar, C., & Cingi, H. (2005). A new ratio estimator in stratified random sampling. Communications in Statistics—Theory and Methods, 34, 597–602.
[6] Kadilar, C., & Cingi, H. (2003). Ratio estimator in stratified random sampling: Family of estimators of population mean using two auxiliary variables in stratified random sampling. Biometrical Journal, 45(2), 218–225.
[7] Koyuncu, N., & Kadilar, C. (2009). Family of estimators of population mean using two auxiliary variables in stratified random sampling. Communications in Statistics—Theory and Methods, 38(14), 2552–2558.
[8] Prasad, B. (1989). Some improved ratio type estimators of population mean and ratio in finite population sample survey. Communications in Statistics—Theory and Methods, 18(1), 379–392.
[9] Saini, M., & Bahl, S. (2012). Estimation of population mean in two stage design using double sampling for stratification and multi-auxiliary information. International Journal of Computer Applications, 47(9), 17–21.
[10] Shabbir, J., & Gupta, S. (2006). A new estimator of population mean in stratified sampling. Communications in Statistics—Theory and Methods, 35, 1201–1209.
[11] Sharma, V., & Kumar, S. (2020). Simulation study of ratio type estimators in stratified random sampling using multi-auxiliary information. Thailand Statistician, 18(3), 281–289.
[12] Singh, H. P., & Kakran, M. S. (1993). A modified ratio estimator using known coefficient of kurtosis of an auxiliary character (Revised version submitted to Journal of the Indian Society of Agricultural Statistics). New Delhi, India.
[13] Singh, R., & Kumar, M. (2012). Improved estimators of population mean using two auxiliary variables in stratified random sampling. Pakistan Journal of Statistics and Operation Research, 8(1), 65–72.
[14] Sisodia, B. V. S., & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. Journal of the Indian Society of Agricultural Statistics, 33, 13–18.
[15] Tailor, R., Chauhan, S., & Garg, N. (2012). A ratio-cum-product estimator of population mean in stratified random sampling using two auxiliary variables. Statistica, 72(3), 287–297.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY