VFAST Transactions on Mathematics
http://vfast.org/journals/index.php/VTM@ 2024 ISSN(e): 2309-0022, ISSN(p): 2411-6343
Volume 12, Number 1, January-June, 2024 pp: 336-348

Improved Class of Regression Type Mean Estimators in Simple Random Sampling Using Concomitant Variable: An Application in T-20 International Cricket

Mujeeb Hussain[1*], Syed Habib Shah[2], Sumayyia Azam[3], Ikram Ullah[3], Muhammad Irshad[3], Nisar Ullah[3]

[1]Department of Statistics, Government College Peshawar, Pakistan; [2]Institute of Numerical Sciences Kohat University of Science and Technology; [3]Department of Statistics, University of Peshawar,Pakistan

Keywords: *Auxiliary information, T20 cricket, score on ground, MSE, skewness* ***Subject Classification***: *62D05*

Journal Info:
Submitted:
May 15, 2024
Accepted:
June 16, 2024
Published:
June 30, 2024

Abstract In sample surveys, the information on supplementary variables is usually used to improve the efficiency of the estimators. The ratio, product, and regression estimators are examples. This study proposes new regression estimators for population mean estimation under simple random sampling using the data of Twenty20 (T-20) International Cricket. The estimators are proposed with the help of non-conventional location parameters of concomitant variable which includes Tri-Mean, Mid-Range, and coefficient of quartile deviation. The bias and mean squared error (MSE) of new estimators have been theoretically derived up to the first-order approximation. The theoretical findings were then compared with the existing estimators numerically through simulation study and real-life data. The numerical results reveal that the suggested estimators are more efficient than the estimators in the literature in all conditions.

***Correspondence Author Email Address:**
mujeebhussain.stat@gmail.com, cricsportsresearchgroup@gmail.com
DOI: 10.21015/vtm.v12i1.1835

1 Introduction

Practically, it is not easy to obtain complete information about a population. Therefore, conclusions are based on sample data for the whole population. Sampling is a technique that provides us with some data from the whole lot, with the help of probability theory. Among different sampling methods, the most common and easiest method is simple random sampling (SRS). In the sampling survey, with the variable of interest (y), some additional information (x) is collected to improve the efficiency of estimators which has a positive or negative correlation to the variable of interest. For this purpose, many researchers used auxiliary information for the estimation of parameters. These includes Kadilar and Cingi [7]; Shazad et al., [15]; Lone et al., [11].

Supplementary information is used to obtain precise estimates for the population mean with the help of ratio and regression estimators. Cochran [1] derived a new ratio estimator for the mean using auxiliary information. Singh and Espejo [16] and Singh and Singh [17] also proposed many efficient estimators with the help of different sampling techniques. Ratio estimators are used when there is a positive correlation between the auxiliary variable and the variable of interest. The important contributions in this area are made by Kadilar and Cingi [6], Shabbir and Gupta [14], Khoshnevisan et al., [9], Koyuncu and Kadilar [10], Khare and Khare [8], Mehta and Tailer [12], Garg and Pachori [4], Dash and Sunani [2], and Raja and Maqbool [13].

Subramani and Kumarapandiyan [18] started work by developing a modified ratio estimator for the population mean estimation of the study variable with the help of population deciles of the concomitant variable.

Recently Subzar *et al.* [20] also used population deciles and the coefficient of correlation of the auxiliary variables. On the other hand, Subzar *et al.* [21] proposed an efficient class of estimators by using the information of auxiliary variables of population deciles, median, and their linear combination with coefficient of correlation and coefficient of variation.

Auxiliary information also plays a vital role in the field of sports, particularly in cricket and, can be employed to estimate the efficiency of the average score (study variable) in T20 International Cricket. In such a scenario, the supplementary variables may be ground size, average score on the cricket ground, number of matches played, etc. The researcher can select different concomitant variables depending on the study and the availability of the variables.

Ratio or regression estimates employing auxiliary data have not yet been used in cricket literature while being often used in other study fields. This study makes a significant contribution to the use of advanced techniques in cricketing literature and represents arguably the first attempt to employ supplementary information in cricket. In the current study, an improved class of regression-type estimators is proposed using supplementary information from T-20 International cricket.

Suppose $U = \{U_1, U_2, U_3, \ldots, U_N\}$ be a finite population of N identifiable units and y denotes the variable under study while x is the supplementary variable. Let the values of the study and concomitant variables be denoted by y_i and x_i respectively. A sample of size n is selected without replacement (WOR) from a population U. Let $\bar{x}$ and b be the means and regression coefficient respectively of sample observations corresponding to the population means $\bar{Y}$ and $\bar{X}$, and B be the regression coefficient. Let the standard deviations s_y and s_x be estimators of S_y and S_x respectively, while S_{yx} denotes population covariance between variables. Similarly, the coefficients of variation are denoted by C_y and C_x of study and supplementary variables respectively. ρ shows the correlation coefficient between variables, while β_1 and β_2 show population skewness and kurtosis respectively.

To derive biases and MSEs of the suggested estimators, let e_0 and e_1 be the error terms related to variables y and x respectively, and:

$$e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}, \quad e_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}, \quad \text{and} \quad E(e_0) = E(e_1) = 0,$$

$$E(e_0^2) = \frac{(1-f)}{n} C_y^2, \quad E(e_1^2) = \frac{(1-f)}{n} C_x^2, \quad E(e_0 e_1) = \frac{(1-f)}{n} \rho C_y C_x \quad \text{where} \quad f = \frac{n}{N}$$

$$Also, TM = \frac{Q_1 + 2Q_2 + Q_3}{4}, MR = \frac{x_{(1)} + x_{(N)}}{2}, QD = \frac{Q_3 - Q_1}{2}$$

be the Tri-Mean, Mid-Range and Quartile Deviation of x respectively.

2 Estimators in Literature

The traditional estimator u_0 for population mean and variance is given as

$$u_0 = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1}$$

and

$$\text{var}(u_0) = \frac{(1-f)}{n} \bar{Y}^2 C_y^2 \tag{2}$$

Kadilar and Cingi [6] suggested some modified ratio estimators with the help of known value of the correlation coefficient and kurtosis, as follows:

$$u_1 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + \rho)} \tag{3}$$

The bias and MSE of u_1 is given as:

$$\text{Bias}(u_1) \approx \frac{(1-f)}{n} \bar{Y} \left(\frac{\bar{X}}{\bar{X}\beta_2 + C_x} C_x \right)^2 \tag{4}$$

and

$$\text{MSE}(u_1) \approx \frac{(1-f)}{n} \bar{Y}^2 \left[C_y^2 (1 - \rho^2) + \left(\frac{\bar{X}}{\bar{X}\beta_2 + C_x} C_x \right)^2 \right] \tag{5}$$

Subzar et al. [19] suggested new regression-type estimators with the help of auxiliary information using tri mean, skewness, and population mid-range as follows:

$$u_2 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_1 + \text{TM})} (\bar{X}\beta_1 + \text{TM}) \tag{6}$$

$$u_3 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_1 + \text{MR})} (\bar{X}\beta_1 + \text{MR}) \tag{7}$$

The biases and MSEs of u_2 and u_3 are provided as:

$$\text{Bias}(u_2) \approx \frac{(1-f)}{n} \bar{Y} \left(\frac{\bar{X}\beta_1}{\bar{X}\beta_1 + \text{TM}} C_x \right)^2 \tag{8}$$

$$\text{Bias}(u_3) \approx \frac{(1-f)}{n}\bar{Y}\left(\frac{\bar{X}\beta_1}{\bar{X}\beta_1+\text{MR}}C_x\right)^2 \tag{9}$$

$$\text{MSE}(u_2) \approx \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2(1-\rho^2)+\left(\frac{\bar{X}\beta_1}{\bar{X}\beta_1+\text{TM}}C_x\right)^2\right] \tag{10}$$

$$\text{MSE}(u_3) \approx \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2(1-\rho^2)+\left(\frac{\bar{X}\beta_1}{\bar{X}\beta_1+\text{MR}}C_x\right)^2\right] \tag{11}$$

Ijaz et.al. [5] proposed regression-type estimators for population mean estimation using the parameters of the supplementary variables as described:

$$u_4 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\delta\bar{x}+(\text{QD})C_x}\left(\delta\bar{X}+(\text{QD})C_x\right) \tag{12}$$

$$u_5 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\delta\bar{x}+\bar{X}C_x)}\left(\delta\bar{X}+\bar{X}C_x\right) \tag{13}$$

$$u_6 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\delta\bar{x}+\bar{X})}\left(\delta\bar{X}+\bar{X}\right) \tag{14}$$

$$u_7 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\delta\bar{x}+(\text{QD})M_d}\left(\delta\bar{X}+(\text{QD})M_d\right) \tag{15}$$

$$u_8 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\delta\bar{x}+M_d)}\left(\delta\bar{X}+M_d\right) \tag{16}$$

$$u_9 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\delta\bar{x}+\text{QD})}\left(\delta\bar{X}+\text{QD}\right) \tag{17}$$

$$u_{10} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\delta\bar{x}+M_dC_x}\left(\delta\bar{X}+M_dC_x\right) \tag{18}$$

Where $\delta = \frac{\text{QD}}{S_x}$. The bias and MSE of the estimator u_j in generalized form is provided as:

$$\text{Bias}(u_j) \approx \frac{(1-f)}{n}\bar{Y}(\psi_j^2C_x^2) \tag{19}$$

$$\text{MSE}(u_j) \approx \frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2+\psi_j^2C_x^2\right] \tag{20}$$

Where $j = 1, 2, \ldots, 7$

$$\psi_1 = \frac{\delta\bar{X}}{\delta\bar{X}+(\text{QD})C_x},\quad \psi_2 = \frac{\delta\bar{X}}{\delta\bar{X}+\bar{X}C_x},\quad \psi_3 = \frac{\delta\bar{X}}{\delta\bar{X}+\bar{X}},\quad \psi_4 = \frac{\delta\bar{X}}{\delta\bar{X}+(\text{QD})M_d}$$

$$\psi_5 = \frac{\delta\bar{X}}{\delta\bar{X}+M_d},\quad \psi_6 = \frac{\delta\bar{X}}{\delta\bar{X}+\text{QD}},\quad \psi_7 = \frac{\delta\bar{X}}{\delta\bar{X}+M_dC_x}$$

3 Proposed Estimator

Motivated by Kadilar and Cingi [6], Subzar et al. [19], and Ijaz et al. [5], some improved class of regression type estimators u_{Pm} are suggested for the mean of a population by using Tri-Mean, Mid-Range, and Quartile Deviation, given as:

$$u_{Pm} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(F_q\bar{x} + G_r)}((F_q\bar{X} + G_r)) \tag{21}$$

where

$$F_q = \left[\left(\frac{\text{QD}}{\text{TM}}\right), \left(\frac{\text{TM}}{\text{QD}}\right), \left(\frac{\text{TM}}{\text{MR}}\right), \left(\frac{\text{QD}}{\text{MR}}\right)\right]$$

and

$$G_r = [(\text{MR} \cdot \text{TM}), (\text{QD} \cdot \text{TM}), (\text{MR} \cdot \text{QD})]$$

$m = 1, 2, \ldots, 7$, $q = 1, 2, 3, 4$, $r = 1, 2, 3$

To find bias and mean square error, it is proceeded as:

$$u_{Pm} = \frac{\bar{Y}(1 + e_0) + b(\bar{X} - \bar{X}(1 + e_1))}{(F_q\bar{X}(1 + e_1) + G_r)}(F_q\bar{X} + G_r) \tag{22}$$

or

$$u_{Pm} = \frac{\bar{Y}(1 + e_0) - b\bar{X}e_1}{(F_q\bar{X} + G_r)\left[1 + \frac{F_q\bar{X}}{F_q\bar{X}+G_r}e_1\right]}(F_q\bar{X} + G_r) \tag{23}$$

$$u_{Pm} = \frac{\bar{Y}(1 + e_0) - b\bar{X}e_1}{1 + \theta_m e_1} \tag{24}$$

where $\theta_m = \frac{F_q\bar{X}}{F_q\bar{X}+G_r}$,

$$u_{Pm} = (\bar{Y}(1 + e_0) - b\bar{X}e_1)(1 + \theta_m e_1)^{-1} \tag{25}$$

$$u_{Pm} \approx \bar{Y}(1 + e_0 - \theta_m e_1 - \theta_m e_0 e_1 + \theta_m^2 e_1^2) - b\bar{X}(e_1 - \theta_m e_1^2) \tag{26}$$

$$(u_{Pm} - \bar{Y}) \approx \bar{Y}(e_0 - \theta_m e_1 - \theta_m e_0 e_1 + \theta_m^2 e_1^2) - b\bar{X}(e_1 - \theta_m e_1^2) \tag{27}$$

Applying expectation,

$$E(u_{Pm} - \bar{Y}) \approx \bar{Y}\left(-\theta_m\frac{(1-f)}{n}\rho_{yx}C_yC_x + \theta_m^2\frac{(1-f)}{n}C_x^2\right) - B\bar{X}\left(-\theta_m\frac{(1-f)}{n}C_x^2\right) \tag{28}$$

$$\text{Bias}(u_{Pm}) \approx \frac{(1-f)}{n}\left[\bar{Y}(\theta_m^2 C_x^2 - \theta_m\rho_{yx}C_yC_x) + B\bar{X}\theta_m C_x^2\right] \tag{29}$$

$$\text{Bias}(u_{Pm}) \approx \frac{(1-f)}{n}\left[\bar{Y}(\theta_m^2 C_x^2 - \theta_m\rho C_yC_x) + \frac{\bar{Y}\rho C_y}{\bar{X}C_x}\bar{X}C_x^2\theta_m\right] \tag{30}$$

where $B = \frac{S_{xy}}{S_x^2} = \frac{\bar{Y}\bar{X}\rho C_yC_x}{\bar{X}^2C_x^2} = \frac{\bar{Y}\rho C_y}{\bar{X}C_x}$.

The bias of the suggested class of estimators up to the first degree of approximation is given by:

$$\text{Bias}(u_{Pm}) \approx \frac{(1-f)}{n}\bar{Y}(\theta_m^2 C_x^2) \tag{31}$$

To find mean square error, squaring equation (27) both sides,

$$(u_{Pm} - \bar{Y})^2 \approx \left[\bar{Y}(e_0 - \theta_m e_1 - \theta_m e_0 e_1 + \theta_m^2 e_1^2) - b\bar{X}(e_1 - \theta_m e_1^2)\right]^2 \tag{32}$$

$$(u_{Pm} - \bar{Y})^2 \approx \bar{Y}^2(e_0^2 + \theta_m^2 e_1^2 - 2\theta_m e_0 e_1) - b^2\bar{X}^2 e_1^2 - 2\bar{Y}\bar{X}b(e_0 e_1 - \theta_m e_1^2) \quad (33)$$

Applying expectation to both sides and solving the above equation, the MSE is given as:

$$\text{MSE}(u_{Pm}) \approx \frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2 + \theta_m^2 C_x^2\right] \quad (34)$$

4 Efficiency Comparison

In this section, theoretical comparisons are made between the existing estimators and the suggested estimators of the population mean. Using MSEs, the conditions that show that the proposed estimators are more efficient than the existing estimators are derived and given as follows:

Condition (i)

$MSE(u_{Pm}) < \text{Var}(u_0)$, if

$$\frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2 + \theta_m^2 C_x^2\right] < \frac{(1-f)}{n}\bar{Y}^2 C_y^2$$

$$\theta_m^2 < \frac{\rho^2 C_y^2}{C_x^2}$$

Condition (ii)

$MSE(u_{Pm}) < MSE(u_1)$

$$\frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2 + \theta_m^2 C_x^2\right] < \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2(1-\rho^2) + \left(\frac{\bar{X}}{\bar{X}\beta_2 + C_x}C_x\right)^2\right]$$

$$\theta_m < \frac{\bar{X}}{\bar{X}\beta_2 + C_x}C_x$$

Condition (iii)

$MSE(u_{Pm}) < MSE(u_2)$

$$\frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2 + \theta_m^2 C_x^2\right] < \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2(1-\rho^2) + \left(\frac{\bar{X}\beta_1}{\bar{X}\beta_1 + \text{TM}}C_x\right)^2\right]$$

$$\theta_m < \frac{\bar{X}\beta_1}{\bar{X}\beta_1 + \text{TM}}C_x$$

Condition (iv)

$MSE(u_{Pm}) < MSE(u_3)$

$$\frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2 + \theta_m^2 C_x^2\right] < \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2(1-\rho^2) + \left(\frac{\bar{X}\beta_1}{\bar{X}\beta_1 + \text{MR}}C_x\right)^2\right]$$

$$\theta_m < \frac{\bar{X}\beta_1}{\bar{X}\beta_1 + \text{MR}}C_x$$

Condition (v)

$MSE(u_{Pm}) < MSE(u_j)$

$$\frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2+\theta_m^2C_x^2\right] < \frac{(1-f)}{n}\bar{Y}^2\left[(1-\rho^2)C_y^2+\psi_j^2C_x^2\right]$$

$$\theta_m < \psi_j$$

5 Numerical Study

The performance of the suggested estimators is evaluated and compared with the existing estimators in the literature also numerically using real data sets, and simulation studies. In the real data study, three data sets related to T-20 International Cricket, available on the "ESPN Cricinfo website" [3], are used. The data is taken from all international grounds where T-20 cricket played from 2005 to 2023. Different were study and auxiliary variables are considered. The description of the study and auxiliary variables in the three data sets is given in Section 5.2.

5.1 Performance criteria

The efficiency of existing and suggested estimators is evaluated using two criteria as follows:

1. The MSE of the estimators.
2. The PRE of the proposed estimator u_{Pm} and considered estimators are computed by:

$$\text{PRE} = \frac{\text{Var}(u_0)}{\text{MSE}(u_j) \text{ and } \text{MSE}(u_{Pm})} \times 100 \tag{35}$$

5.2 Real Data Analysis

The description of real data sets and numerical analysis are given below.

Data Set 1

Y: Score on the ground by both teams

X: Number of matches played on ground

$N = 70$, $n = 20$, $\bar{Y} = 3668.852$, $\bar{X} = 12.51$, $\rho = 0.996$

$S_y = 4115.506$, $S_x = 14.749$, $\beta_1 = 3.014$, $\beta_2 = 10.77$

$Q_1 = 4.0$, $Q_2 = 7.0$, $Q_3 = 14.5$

$X_1 = 1$, $X_n = 87$

Data Set 2

Y: Score on the ground by both teams

X: Wickets fall during the whole match

$N = 70$, $n = 20$, $\bar{Y} = 3668.852$, $\bar{X} = 162.806$, $\rho = 0.99$

$S_y = 4115.506$, $S_x = 191.039$, $\beta_1 = 2.883$, $\beta_2 = 9.983$

$Q_1 = 52$, $Q_2 = 95.37$, $Q_3 = 188.57$

$X_1 = 12.73$, $X_n = 1113.99$

Data Set 3

Y: 1st innings score on ground

X: Wickets fall during the whole match in 1st innings

$N = 70$, $n = 25$, $\bar{Y} = 1943.03$, $\bar{X} = 82.77$, $\rho = 0.986$

$S_y = 2175.76$, $S_x = 96.45$, $\beta_1 = 0.274$, $\beta_2 = 8.71$

$Q_1 = 25.66$, $Q_2 = 49.33$, $Q_3 = 98.98$

$X_1 = 6.82$, $X_n = 541.99$

Using the above data, the numerical values of different conditions, MSEs, and PREs are given in Table 1 and Table 2.

5.3 Simulation Study

Simulations of 20000 are performed using a bivariate normal distribution with theoretical means of [y,x] $\mu = [2, 3]$ and variances equal to 1. The coefficient of correlation between the study and the concomitant variable is 0.9. The covariance matrix is:

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

Sample sizes $n = 30, 60$, and 100 are selected by simple random sampling without replacement (WOR). MSEs and PREs of the existing and proposed estimators are computed and presented in Table 3.

Comments and Discussion

The present study introduces new regression estimators for estimating the finite population mean under simple random sampling using auxiliary information in the field of T20 International Cricket.

Some general findings are:

1. The results provided in Table 1 show that the PRE of all the proposed estimators u_{Pm} is greater than 100, which indicates that it performs better then the existing estimators.
2. It is also observed from Table 2 that the proposed estimators satisfy all the efficiency conditions and have small MSEs as compared to considered estimators in all three types of data sets. Therefore, the performance of the suggested estimator u_{Pm} is the best among all other considered estimators.
3. The suggested estimator u_{P5} has the smallest MSE as compared to all existing and proposed estimators for all data sets.
4. The results of the simulation study provided in Table 3 reveal that all the suggested estimators have minimum MSE and maximum PRE, then all the competitor estimators.
5. In the simulation study, it is evident that the proposed estimator u_{P5} has the highest PREs (i.e. 517.0765, 505.1173, and 506.1031) for all three samples.
6. The results of the suggested estimators of the simulation study in Table 3 also indicate that, as the sample size increases, the MSE of all the estimators decreases.

Conclusion

The present study introduces new regression estimators for estimating the finite population mean under simple random sampling using auxiliary information in the field of T-20 International Cricket. Different

Table 1. MSE and PRE of existing and proposed estimators

Estimators	Parameters	MSE/PRE	Data 1	Data 2	Data 3
u_0	–	MSE	626903.4	626903.4	127877.6
		PRE	100.00	100.00	100.00
u_1	–	MSE	19348.66	19348.73	5375.005
		PRE	3240.035	3240.024	2379.116
u_2	–	MSE	466098.2	466102.5	15112.74
		PRE	134.5003	134.499	846.1577
u_3	–	MSE	154175.4	154176.8	4362.519
		PRE	406.6169	406.6134	2931.279
u_4	ψ_1	MSE	133896.9	133896.7	28464.15
		PRE	468.1986	468.1992	449.2585
u_5	ψ_2	MSE	49874.26	49874.07	11932.83
		PRE	1256.968	1256.973	1071.645
u_6	ψ_3	MSE	60039.75	60039.87	14059.49
		PRE	1044.147	1044.145	909.547
u_7	ψ_4	MSE	12529.19	12529.19	3595.998
		PRE	5003.544	5003.544	3556.11
u_8	ψ_5	MSE	110987.1	110987.5	24554.47
		PRE	564.8433	564.8415	520.7916
u_9	ψ_6	MSE	157696.7	157697.3	33092.29
		PRE	397.5374	397.5358	386.4273
u_{10}	ψ_7	MSE	92762.86	92762.62	20881.12
		PRE	675.8129	675.8146	612.4079
u_{P1}	(F_1, G_1)	MSE	12477.35	12477.35	3557.241
		PRE	5024.332	5024.332	3594.854
u_{P2}	(F_1, G_3)	MSE	12480.28	12480.28	3559.509
		PRE	5023.152	5023.152	3592.564
u_{P3}	(F_2, G_3)	MSE	9146.485	12505.61	3576.714
		PRE	6854.036	5012.978	3575.283
u_{P4}	(F_3, G_1)	MSE	5033.742	12475.56	3555.677
		PRE	12454.02	5025.053	3596.436
u_{P5}	(F_4, G_1)	MSE	5017.169	12475.45	3555.582
		PRE	12495.16	5025.096	3596.532
u_{P6}	(F_4, G_2)	MSE	5796.809	12480.28	3559.509
		PRE	10814.63	5023.152	3592.564
u_{P7}	(F_4, G_3)	MSE	5033.742	12475.56	3555.677
		PRE	12454.02	5025.053	3596.436

auxiliary information is used from cricket, which includes wickets that fall during the whole match, wickets that fall in different innings, and the number of matches played on the ground. Non-conventional location parameters of auxiliary information are used, which consist of tri-mean, mid-range, quartile deviation, etc.

Table 2. MSE and PRE for evaluating the performance of the suggested and competing estimators.

			Estimators										
conditions	proposed	Data	u_0	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}
i	u_{P1}	1	614426.1	6871.31	453620.9	141698.1	121419.6	37396.91	47562.4	51.84	98509.75	145219.4	80285.51
		2	614426.1	6871.38	453625.2	141699.5	121419.4	37396.72	47562.52	51.84	98510.15	145220	80285.27
		3	124320.4	1817.764	11555.5	805.278	24906.91	8375.589	10502.25	38.757	20997.23	29535.05	17323.88
ii	u_{P2}	1	614423.1	6868.38	453617.9	141695.1	121416.6	37393.98	47559.47	48.91	98506.82	145216.4	80282.58
		2	614423.1	6868.45	453622.2	141696.5	121416.4	37393.79	47559.59	48.91	98507.22	145217	80282.34
		3	124318.1	1815.496	11553.23	803.01	24904.64	8373.321	10499.98	36.489	20994.96	29532.78	17321.61
iii	u_{P3}	1	621869.7	14314.92	461064.5	149141.7	128863.2	44840.52	55006.01	7495.448	105953.4	152663	87729.12
		2	614427.8	6873.17	453626.9	141701.2	121421.1	37398.51	47564.31	53.63	98511.94	145221.7	80287.06
		3	124321.9	1819.328	11557.06	806.842	24908.47	8377.153	10503.81	40.321	20998.79	29536.61	17325.44
iv	u_{P4}	1	620070.5	12515.76	459265.3	147342.5	127064	43041.36	53206.85	5696.289	104154.2	150863.8	85929.96
		2	614415.8	6861.17	453614.9	141689.2	121409.1	37386.51	47552.31	41.63	98499.94	145209.7	80275.06
		3	124312.9	1810.273	11548.01	797.787	24899.42	8368.098	10494.76	31.266	20989.74	29527.56	8393.56
v	u_{P5}	1	621886.2	14331.49	461081	149158.2	128879.7	44857.09	55022.58	7512.021	105969.9	152679.5	87745.69
		2	614428	6873.28	453627.1	141701.4	121421.3	37398.62	47564.42	53.74	98512.05	145221.9	80287.17
		3	124322	1819.423	11557.16	806.937	24908.57	8377.248	10503.91	40.416	20998.89	29536.71	17325.54
vi	u_{P6}	1	621106.6	13551.85	460301.4	148378.6	128100.1	44077.45	54242.94	6732.381	105190.3	151899.9	86966.05
		2	614423.1	6868.45	453622.2	141696.5	121416.4	37393.79	47559.59	48.91	98507.22	145217	80282.34
		3	124318.1	1815.496	11553.23	803.01	24904.64	8373.321	10499.98	36.489	20994.96	29532.78	17321.61
vii	u_{P7}	1	621869.7	14314.92	461064.5	149141.7	128863.2	44840.52	55006.01	7495.448	105953.4	152663	87729.12
		2	614427.8	6873.17	453626.9	141701.2	121421.1	37398.51	47564.31	53.63	98511.94	145221.7	80287.06
		3	124321.9	1819.328	11557.06	806.842	24908.47	8377.153	10503.81	40.321	20998.79	29536.61	17325.44

Table 3. MSEs and PREs of different estimators using simulation study

Estimators	n = 30		n = 60		n = 100	
	MSE	PRE	MSE	PRE	MSE	PRE
u_0	0.029686	100.00	0.013262	100.00	0.006995	100.00
u_1	0.006524	455.0164	0.003063	432.9709	0.001577	443.6744
u_2	0.006595	450.1478	0.002976	445.598	0.001604	436.1892
u_3	0.006747	440.0165	0.003020	439.226	0.001555	459.8696
u_4	0.016854	176.1341	0.007471	177.3068	0.003942	177.4276
u_5	0.011884	249.8113	0.005318	249.3934	0.002806	249.2849
u_6	0.007868	377.3179	0.003559	372.6918	0.001877	372.5662
u_7	0.009124	325.3738	0.004101	322.6989	0.002169	322.5127
u_8	0.007889	376.3243	0.003568	371.7387	0.001882	371.6109
u_9	0.013399	221.5505	0.005979	221.8109	0.003154	221.774
u_{10}	0.011916	249.1407	0.005332	248.7399	0.002813	248.6328
			$\{F_1, (G_1, G_3)\}$			
u_{P1}	0.005749	516.3388	0.002629	504.4382	0.001384	505.3999
u_{P2}	0.006396	464.1188	0.002912	455.4071	0.001535	455.6804
			$\{F_2, G_3\}$			
u_{P3}	0.006074	488.7188	0.002709	489.6599	0.001428	489.7321
			$\{F_3, G_1\}$			
u_{P4}	0.006293	471.772	0.002867	462.6418	0.001511	462.9775
			$\{F_4, G_r\}$			
u_{P5}	0.005741	517.0765	0.002626	505.1173	0.001382	506.1031
u_{P6}	0.006396	464.1188	0.002912	455.4071	0.001535	455.6804
u_{P7}	0.006293	471.772	0.002867	462.6418	0.001511	462.9775

Expressions for bias and mean square error of the suggested estimators are obtained up to the first-order approximation. The new suggested estimators are compared with the existing estimators using MSE and PRE criteria. The results reveal that all the suggested estimators are efficient as their MSEs are less than the competing estimators. Moreover, the suggested estimator u_{P5} has the largest PRE, then all existing and other proposed estimators. Hence, the proposed estimators can be applied to appropriate real-world situations and obtain better and more efficient results as compared to the usual and other competing population mean estimators.

Recommendations

It is recommended that this study can be extended by developing new estimators utilizing other sampling schemes.

Data Availability

The data used to support the findings of the study are available within this article.

Acknowledgment

The authors express appreciation to the editor and reviewers for their valuable and positive comments/suggestions which certainly have improved the presentation and quality of the paper.

Author Contributions

Mujeeb Hussain: Conceptualization, Supervision, Data. **Syed Habib Shah:** Data curation, Software, Visualization. **Sumayyia Azam:** Investigation, Removed all the grammatical mistakes. **Muhammad Ikramullah:** Analyzed and interpreted, Implementation, Investigation. **Muhammad Irshad:** Writing - Original draft preparation, Validation. **Nisar Ullah:** Methodology, Improved language, Editing.

Compliance with Ethical Standards

The authors declare that they have no conflicts of interest. Additionally, this article does not involve studies with human participants or animals conducted by any of the authors. Furthermore, informed consent was taken from all individual participants involved in the study.

Author Information

ORCID:

Mujeeb Hussain: 0009-0007-5971-2685
Sumayyia Azam: 0009-0006-1350-0225
Ikram Ullah: 0009-0007-8801-232X
Muhammad Irshad: 0009-0000-4656-7772
Nisar Ullah: 0009-0009-1317-1208

References

[1] Cochran, W. G. [1940], 'The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce', *The journal of agricultural science* **30**(2), 262–275.

[2] Dash, P. and Kalyani, S. [2019], 'Estimation of population proportion by using calibration estimator', *Int. J. Agricult. Stat. Sci.* **15**(1), 249–254.

[3] ESPN Cricinfo [2023], 'Statistics'. viewed 15 July 2023, retrieved from http://stats.espncricinfo.com/ci/engine/stats/index.html.

[4] Garg, N. and Pachori, M. [2019], 'Calibration estimation in population mean in stratified sampling using coefficient of skewness', *Int. J. Agricult. Stat. Sci.* **15**(1), 211–219.

[5] Ijaz, M., Zaman, T., Bulut, H., Ullah, A. and Asim, S. M. [2020], 'An improved class of regression estimators using the auxiliary information', *Journal of Science and Arts* **20**(4), 789–800.

[6] Kadilar, C. and Cingi, H. [2004], 'Ratio estimators in simple random sampling', *Applied mathematics and computation* **151**(3), 893–902.

[7] KADILAR, C., CİNGİ, H. and ÇINGI, H. [2006], 'An improvement in estimating the population mean by using the correlation coefficient', *hacettepe Journal of Mathematics and Statistics* **35**(1), 103–109.

[8] Khare, B. and Khare, S. [2019], 'On the utilization of known coefficient of variation and preliminary test of significance in the estimation of population mean', *Int. J. Agricult. Stat. Sci.* **15**(1), 35–37.

[9] Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. [2007], 'A general family of estimators for estimating population mean using known value of some population parameter(s)', *Far East J. Theoret. Statist.* **22**, 181–191.

[10] Koyuncu, N. and Kadilar, C. [2009], 'Ratio and product estimators in stratified random sampling', *Journal of statistical planning and inference* **139**(8), 2552–2558.

[11] Lone, S. A., Subzar, M. and Sharma, A. [2021], 'Enhanced estimators of population variance with the use of supplementary information in survey sampling', *Mathematical Problems in Engineering* .

[12] Mehta, P. and Tailor, R. [2019], 'Estimation of ratio of two population means in stratified random sampling', *Int. J. Agricult. Stat. Sci.* **15**(1), 199–203.

[13] Raja, T. A. and Maqbool, S. [2021], 'On modified ratio estimator using a new linear combination', *International Journal of Agricultural and Statistical Sciences* **17**(1), 209–211.

[14] Shabbir, J. and Gupta, S. [2006], 'A new estimator of population mean in stratified sampling', *Communications in Statistics-Theory and Methods* **35**(7), 1201–1209.

[15] Shahzad, U., Ahmad, I., Oral, E., Hanif, M. and Almanjahie, I. M. [2021], 'Estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling', *Mathematical Population Studies* **28**(3), 176–199.

[16] Singh, H. P. and Espejo, M. R. [2003], 'On linear regression and ratio–product estimation of a finite population mean', *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**(1), 59–67.

[17] Singh, V. K. and Singh, R. [2014], 'Performance of an estimator for estimating population mean using simple and stratified random sampling', *SOP transactions on Statistics and Analysis* **1**(1), 1–8.

[18] Subramani, J. and Kumarapandiyan, G. [2012], 'A class of modified ratio estimators using deciles of an auxiliary variable', *International Journal of Statistical Application* **2**(6), 101–107.

[19] Subzar, M., Maqbool, S., Raja, T. A. and Abid, M. [2018], 'Ratio estimators for estimating population mean in simple random sampling using auxiliary information', *Applied Mathematics and Information Sciences Letters* **6**(3), 123–130.

[20] Subzar, M., Maqbool, S., Raja, T. A. and Shabeer, M. [2017], 'A new ratio estimators for estimation of population mean using conventional location parameters', *World Applied Sciences Journal* **35**(3), 377–384.

[21] Subzar, M., Raja, T. A., Maqbool, S. and Nazir, N. [2016], 'New alternative to ratio estimator of population mean', *International Journal of Agricultural and Statistical Sciences* **1**, 221–225.