Kurdish Studies
October 2024
Volume: 12, No: 5, pp 1553 - 1562
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v12i5.3508

Novel Methods For Estimation Of Population Mean Using Auxiliary Information Under PPS Sampling: Application With Real And Simulated Data Sets

Manahil SidAhmed Mustafa[1], Sohaib Ahmad[2,*], Erum Zahid[3], Javid Shabbir[4], Saadia Masood[5]

[1]Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia
[2*]Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan
[3]Department of Applied Mathematics and Statistics, Institute of Space Technology, Pakistan
[4]Department of Statistics, University of Wah, Wah Cantt, Pakistan
[5]Department of Statistics, PMAS Arid Agriculture University, Rawalpindi, Pakistan
[1]Email: msida@ut.edu.sa
[2*]Email: sohaib_ahmad@awkum.edu.pk
[3]Email: erumzahid2@gmail.com
[4]Email: javidshabbir@gmail.com
[5]Email: saadia.masood@uaar.edu.pk

***Correspondence:** Sohaib Ahmad
*E-mail: sohaib_ahmad@awkum.edu.pk

Abstract
The use of simple random sampling (SRS) ensures that each and every unit in the population has an equal chance of being included in the sample. Due to the fact that it disregards the potential significance of the units size, SRS does not appear to be an appropriate strategy on the other hand when the units differ greatly in size. In these kinds of circumstances, it is possible that selecting units with the help of unequal probabilities instead of sampling with equal probability will result in more accurate estimations. In this method, the units are selected with a probability proportional to size (PPS) sampling that corresponds to a predetermined size measure. In order to estimate the finite population mean, this paper proposed a logarithmic-type estimator that relies on PPS sampling and simple random sampling. If the sampling technique is applied symmetrically across the population, biases can be minimized and the sample can be more accurately used to represent the entire population. We conducted an extensive numerical analysis and simulation study to evaluate the proposed estimator. We also provide visual representations of the results to show how well the proposed estimator works. In light of the numerical outcome, we see that the proposed estimator is more useful for estimating the population means using PPS sampling methods.

Keywords: PPS sampling, mean estimation, simulation, mean squared error, bias, efficiency.

Mathematical Subject Classification: 62D05

1. Introduction

A common issue with survey sampling is trying to estimate the finite population mean; many determinations have been made to increase the accuracy of these estimators. There is an enormous amount of information in the literature on various ways to incorporate the auxiliary variables using ratio, product, and regression-type estimations. In cases where many auxiliary variables are present, a comprehensive range of estimators have been proposed, most of which combine ratio, product, or regression estimators. Using the most effective statistical features, scientists have previously attempted to estimate kurtosis, variance, and coefficient of variation, among other population metrics. For this approach to work, it requires a population sample that is highly representative. If the target population is homogeneous, then simple random sampling (with or without replacement) can be used to choose units. It is also necessary to know the auxiliary variable's population parameters before estimating it using the ratio, product, or regression methods. By making suitable adjustments to the auxiliary variables, several authors have put forward various estimators. In order explore more into these research findings, the investigator may analyze the [10-18].

In many contexts, the size of the population might vary substantially. A medical study may alter the size of hospitals and the frequency of a certain diagnosis, for instance. A questionnaire focusing on household income may also yield different results for residences with varying numbers of siblings, leading to various probabilities of units. To address this unequal probability, we use PPS sampling. The probability of collecting data for each population sample unit is proportionate to an auxiliary variable in PPS, an unequal random sampling approach. Consider the case when we need to assess a province's population within a country; in this case, we would make use of the corresponding auxiliary variable to the study variable. For instance, examine (i) the total population of the provinces in the country (correlation with research variable = 0.97). (ii) the total number of families in each village in the province has a correlation coefficient of 0.99 with the research variable. In light of these facts, it appears that (ii) might demonstrate to be more useful when estimating. A number of estimators have been suggested by different researchers under PPS, which successfully adjust the auxiliary variables. The researcher has the capacity to examine this work by [19-40].

(i) It is the goal of this article to provide better estimators for the finite population mean under PPS using auxiliary variables.

(ii) The properties of the recommended estimator are disclosed up to the first order of approximation.

(iii) The uses of the suggested estimator are checked with the help of a simulation investigation and real data sets from different domains.

The article has been structure as follows. The methods and notations for this article are discussed in detailed in Section 2. Existing estimators were reviewed in Section 3. Section 4 outlined the suggested estimate for computing the populations mean using PPS sampling. Section 5 presents the empirical investigation. In Section 6, we present a simulation study that assesses the performance of our suggested estimator. Section 7 discusses the numerical results. In Section 8, the conclusion of the paper is given in detail.

2. Methodology

Let a population comprises $\mathbb{W} = \{\mathbb{W}_1, \mathbb{W}_2, \ldots, \mathbb{W}_N\}$. Let Y_i and $\{X_i, Z_i\}$ be the features of the population study variable and the auxiliary variables correspondingly. We select a sample of size *n* by using probability proportional to size sampling taking *Z* as size of the unity. i.e.

$\mathrm{P}_i = \frac{z_i}{\sum_{i=1}^{n} z_i}$, be the PPS sampling for gaining the components. To acquire a sample of size *n*, we employ the PPSWR method.

Define

$$\left[u_i = \frac{Y_i}{N\mathrm{P}_i}, v_i = \frac{X_i}{N\mathrm{P}_i}\right],$$

$$\left[\bar{u} = \frac{\sum_{i=1}^{n} u_i}{n} = \bar{y}_{pps}, \bar{v} = \frac{\sum_{i=1}^{n} v_i}{n} = \bar{x}_{pps}\right],$$

We evaluate the following error terms in order to determine the characteristics of the estimators, namely the bias and mean square error:

Let $\varrho_0 = \frac{\bar{u} - \bar{Y}}{\bar{Y}}$, $\varrho_1 = \frac{\bar{v} - \bar{X}}{\bar{X}}$, $E(\varrho_i) = 0$, $i = 0, 1$.

$E(\varrho_0^2) = \frac{C_u^2}{n}$, $E(\varrho_1^2) = \frac{C_v^2}{n}$, $E(\varrho_0 \varrho_1) = \lambda \rho_{uv} C_u C_v$,

$C_u = \frac{S_u}{\bar{Y}}$, $C_v = \frac{S_v}{\bar{X}}$, be the population C.V such that

$S_u = \sqrt{\sum_{i=1}^{N} \mathrm{P}_i (u_i - \bar{Y})^2}$, $S_u = \sqrt{\sum_{i=1}^{N} \mathrm{P}_i (v_i - \bar{X})^2}$

Let $\rho_{uv} = \frac{\sum_{i=1}^{N} \mathrm{P}_i (u_i - \bar{Y})(v_i - \bar{X})}{S_u S_v}$.

3. Existing estimators

This section provides a detailed analysis of many standard estimators that are currently in use for mean estimate under PPS sampling.

1. The usual estimator with variance under PPS, are given by:

$$T_{1pps} = \bar{u} \quad (1)$$

$$\mathrm{Var}(T_{1pps}) = \lambda \bar{Y}^2 C_u^2 \quad (2)$$

2. The ratio estimator with bias and MSE under PPS, are given by:

$$T_{2pps} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right) \quad (3)$$

$\mathrm{Bias}(T_{2pps}) \cong \bar{Y} \lambda [C_u^2 - \rho_{uv} C_u C_v]$,

$$\mathrm{MSE}(T_{2pps}) \cong \bar{Y}^2 \lambda [C_u^2 + C_v^2 - 2\rho_{uv} C_u C_v] \quad (4)$$

3. [6] suggested a product estimator with bias and MSE, are given by:

$$T_{3pps} = \bar{u}\left(\frac{\bar{v}}{\bar{X}}\right) \quad (5)$$

$\mathrm{Bias}(T_{3pps}) \cong \bar{Y} \lambda \rho_{uv} C_u C_v$,

$$\mathrm{MSE}(T_{3pps}) \cong \bar{Y}^2 \lambda [C_u^2 + C_v^2 + 2\rho_{uv} C_u C_v]. \quad (6)$$

4. The regression estimator, is given by:

$$T_{4pps} = \bar{u} + \psi_1(\bar{X} - \bar{v}), \quad (7)$$

where ψ_1 is constant. The value of ψ_1 is given by:

$\psi_{1(opt)} = \frac{\rho_{uv} S_u}{S_v}$

The variance of T_{4pps} at $\psi_{1(opt)}$ is given by:

$$\mathrm{Var}(T_{4pps}) \cong \lambda \bar{Y}^2 C_u^2 (1 - \rho_{uv}^2) \quad (8)$$

5. [7] suggested the following estimators:

$$T_{5pps} = \bar{u} \exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}}\right), \quad (9)$$

and

$$T_{6pps}=\bar{u}\ \exp\left(\frac{\bar{v}-\bar{X}}{\bar{v}+\bar{X}}\right). \quad (10)$$

The biases and MSEs of T_{5pps}, T_{6pps} are given by:

$\text{Bias}(T_{5pps}) \cong \lambda\,\bar{Y}\left(\frac{3}{8}C_v^2-\frac{1}{2}\rho_{uv}C_uC_v\right)$

$\text{MSE}(T_{5pps}) \cong \lambda\,\bar{Y}^2\left(C_u^2+\frac{1}{4}C_v^2-\rho_{uv}C_uC_v\right),$ (11)

$\text{Bias}(T_{6pps}) \cong \lambda\,\bar{Y}\left(\rho_{uv}C_uC_v-\frac{1}{4}C_v^2\right),$

$\text{MSE}(T_{6pps}) \cong \lambda\,\bar{Y}^2\left[C_u^2+\frac{1}{4}C_v^2+\rho_{uv}C_uC_v\right]$ (12)

6. [1] suggested ratio type estimator with bias and MSE are given by:

$T_{7pps}=(1-\Omega_1)\bar{u}+\bar{u}\left(\frac{\bar{X}}{\bar{v}}\right)$ (13)

$\text{Bias}(T_{7pps}) \cong \lambda\,\bar{Y}\left(\frac{\Omega_1}{2}C_v^2-\Omega_1\rho_{uv}C_uC_v\right)$

$\text{MSE}(T_{7pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_1^2C_v^2-2\Omega_1\rho_{uv}C_uC_v]$ (14)

where

$\Omega_1=\frac{\rho_{uv}C_v}{C_v}$

7. [2] suggested improved ratio type estimator using coefficient of variation with bias and MSE, are given by:

$$T_{8pps}=\bar{u}\ \exp\left(\frac{\bar{X}-C_v}{\bar{v}+C_v}\right) \quad (15)$$

$\text{Bias}(T_{8pps}) \cong \lambda\,\bar{Y}(\Omega_2^2C_v^2-\Omega_2\rho_{uv}C_uC_v)$

$\text{MSE}(T_{8pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_2^2C_v^2-2\Omega_2\rho_{uv}C_uC_v]$ (16)

$\Omega_2=\frac{\bar{X}}{\bar{X}+C_v}$

8. [3] recommended ratio estimator using C.V and correlation among the study and the auxiliary variables, with bias and MSE are given by:

$T_{9pps}=\left(\frac{\bar{u}}{\bar{v}C_v+\rho_{uv}}\right)[\bar{X}C_v+\rho_{uv}]$ (17)

$\text{Bias}(T_{9pps}) \cong \lambda\bar{Y}(\Omega_3^2C_v^2-\Omega_3\rho_{uv}C_uC_v)$

$\text{MSE}(T_{9pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_3^2C_v^2-2\Omega_3\rho_{uv}C_uC_v]$ (18)

$\Omega_3=\frac{\bar{X}C_v}{\bar{X}C_v+\rho_{uv}}$

9. [4] suggested ratio estimator using third quartile with bias and MSE are given by:

$T_{10pps}=\bar{u}\left(\frac{\bar{X}+Q_{3(x)}}{\bar{v}+Q_{3(x)}}\right)[\bar{X}C_v+\rho_{uv}]$ (19)

$\text{Bias}(T_{10pps}) \cong \lambda\bar{Y}(\Omega_4^2C_v^2-\Omega_4\rho_{uv}C_uC_v)$

$\text{MSE}(T_{10pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_4^2C_v^2-2\Omega_4\rho_{uv}C_uC_v]$ (20)

$\Omega_4=\frac{\bar{X}}{\bar{X}+Q_{3(v)}}$

10. [4] suggested ratio estimator using first quartile with bias and MSE are given by:

$T_{11pps}=\bar{u}\left(\frac{\bar{X}+Q_{1(v)}}{\bar{v}+Q_{1(v)}}\right)[\bar{X}C_v+\rho_{uv}]$ (21)

$\text{Bias}(T_{11pps}) \cong \lambda\bar{Y}(\Omega_5^2C_v^2-\Omega_5\rho_{uv}C_uC_v)$

$\text{MSE}(T_{11pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_5^2C_v^2-2\Omega_5\rho_{uv}C_uC_v]$ (22)

$\Omega_5=\frac{\bar{X}}{\bar{X}+Q_{1(v)}}$

11. [5] suggested ratio estimator using median of the auxiliary variables with bias and MSE are given by:

$T_{12pps}=\bar{u}\left(\frac{\bar{X}C_v+M_x}{\bar{v}C_v+M_x}\right)$ (23)

$\text{Bias}(T_{12pps}) \cong \lambda\bar{Y}(\Omega_6^2C_v^2-\Omega_6\rho_{uv}C_uC_v)$

$\text{MSE}(T_{12pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_6^2C_v^2-2\Omega_6\rho_{uv}C_uC_v]$ (24)

$\Omega_6=\frac{\bar{X}\,C_v}{\bar{X}C_v+M_x}$

12. [5] suggested ratio estimator using quartile deviation with bias and MSE are given by:

$T_{13pps}=\bar{u}\left(\frac{\bar{X}+Q.D_x}{\bar{v}+Q.D_x}\right)$ (25)

$\text{Bias}(T_{13pps}) \cong \lambda\bar{Y}(\Omega_7^2C_v^2-\Omega_7\rho_{uv}C_uC_v)$

$\text{MSE}(T_{13pps}) \cong \lambda\,\bar{Y}^2[C_u^2+\Omega_7^2C_v^2-2\Omega_7\rho_{uv}C_uC_v]$ (26)

$Q.D_x=\frac{Q_3-Q_1}{2}$, $\Omega_7=\frac{\bar{X}}{\bar{X}+Q.D_x}$

13. [5] suggested estimator using quartile average with bias and MSE are given by:

$T_{14pps}=\bar{u}\left(\frac{\bar{X}+Q.Da_{(x)}}{\bar{v}+Q.D_{xa(x)}}\right)$ (27)

$\text{Bias}\left(T_{14pps}\right) \cong \lambda\bar{Y}(\Omega_8^2 C_v^2 - \Omega_8\rho_{uv}C_uC_v)$

$\text{MSE}\left(T_{14pps}\right) \cong \lambda\,\bar{Y}^2[C_u^2 + \Omega_8^2C_v^2 - 2\Omega_8\rho_{uv}C_uC_v]$ (28)

$Q.Da_{(x)}=\frac{Q_3+Q_1}{2}$, $\Omega_8=\frac{\bar{X}}{\bar{X}+Q.Da_{(x)}}$

4. Suggested estimator

In most cases, the precision of population parameter estimation is enhanced by including auxiliary information, which reduces the Mean Square Error (MSE) and boost the percent relative efficiency (PREs). This improvement is predicated on a proper function being applied with care and consideration. This articles aims to suggest an improved estimator for estimation of population mean under PPS, which is given by:

$$T_{prop(pps)}=\left[\bar{u}\left(\frac{\bar{X}}{\bar{v}}\right)+\Omega_{11}\bar{u}\exp\left(\frac{\bar{X}-\bar{v}}{\bar{X}+\bar{v}}\right)+\Omega_{12}\bar{u}\left\{1+\log\left(\frac{\bar{X}}{\bar{v}}\right)\right\}\right] \quad (29)$$

In (29), the Ω_{11} and Ω_{12} are constants.

By expressing (29), we got the simplified form, which is given by:

$$T_{prop(pps)}-\bar{Y}=\bar{Y}(-\varrho_1+\varrho_1^2)+\bar{Y}\Omega_{11}\left(1-\frac{\varrho_1}{2}+\frac{\varrho_1^2}{2}\right)+\bar{Y}\Omega_{12}\left(1-\frac{\varrho_1}{2}+\frac{\varrho_1^2}{2}\right)+\bar{Y}(\varrho_0-\varrho_0\varrho_1)+\bar{Y}\Omega_{11}\left(\varrho_0-\frac{\varrho_0\varrho_1}{2}\right)+\bar{Y}\Omega_{12}(\varrho_0-\varrho_0\varrho_1) \quad (30)$$

Taking expectation on both sides, we got bias of $T_{prop(pps)}$, which is given by:

$$\text{Bias}\left(T_{prop(pps)}\right)=\bar{Y}\left[\lambda C_v^2-\lambda\rho_{uv}C_uC_v+\Omega_{11}\left(1+\frac{3}{8}\lambda C_v^2-\frac{1}{2}\lambda\rho_{uv}C_uC_v\right)+\Omega_{11}\left(1+\frac{1}{2}\lambda C_v^2-\lambda\rho_{uv}C_uC_v\right)\right]$$

Squaring and taking expectation of (30), after simplification, we got the MSE expressions, which is given by:

$$\text{MSE}\left(T_{prop(pps)}\right)=\bar{Y}^2[A_{11}+\Omega_{11}^2B_{11}+\Omega_{12}^2C_{11}+2\Omega_{11}D_{11}+2\Omega_{12}F_{11}+2\Omega_{11}\Omega_{12}G_{11}] \quad (31)$$

Where

$A_{11}=[\lambda C_u^2+\lambda C_v^2-2\lambda\rho_{uv}C_uC_v]$, $B_{11}=1+\lambda[C_v^2+C_u^2-2\lambda\rho_{uv}C_uC_v]$,

$C_{11}=1+\lambda[2C_v^2+C_u^2-4\lambda\rho_{uv}C_uC_v]$, $D_{11}=\lambda\left(C_v^2-\frac{1}{5}\rho_{uv}C_uC_v+\frac{3}{2}C_v^2\right)$

$F_{11}=\lambda[2C_v^2+C_u^2-3\lambda\rho_{uv}C_uC_v]$, $G_{11}=1+\lambda\left[\frac{11}{8}C_v^2+C_u^2-3\lambda\rho_{uv}C_uC_v\right]$

Differentiating (31), with respect to Ω_{11} and Ω_{12}, we got the optimum values, which is given by:

$\Omega_{11}=\frac{F_{11}G_{11}-D_{11}C_{11}}{B_{11}C_{11}-G_{11}^2}$

$\Omega_{12}=\frac{D_{11}G_{11}-B_{11}F_{11}}{B_{11}C_{11}-G_{11}^2}$

Putting the values of Ω_{11} and Ω_{12} in (31):

$$\text{MSE}\left(T_{prop(pps)_{min}}\right)=\bar{Y}^2\left[A_{11}+\left(\frac{P_{11}}{Q_{11}}\right)\right] \quad (32)$$

Where

$P_{11}=B_{11}(F_{11}G_{11}-D_{11}C_{11})^2+C_{11}(D_{11}G_{11}-B_{11}F_{11})^2+2(F_{11}G_{11}-D_{11}C_{11})D_{11}(B_{11}C_{11}-G_{11}^2)+2(D_{11}G_{11}-B_{11}F_{11})F_{11}(B_{11}C_{11}-G_{11}^2)+2G_{11}^2(F_{11}G_{11}-D_{11}C_{11})(D_{11}G_{11}-B_{11}F_{11})$,

$Q_{11}=(B_{11}C_{11}-G_{11}^2)^2$.

5. Numerical study

Here we use a few real data sets to compare the proposed and current estimators. Each estimator is evaluated by calculating its PRE and MSE. Details and a brief summary of the datasets being considered are provided below:

Population-I: [**Source:** [9]]

Y=No. of IDPs in USA in 1996, *X*= No. of IDPs in USA in 1995, *Z*= Number of IDPs in USA in 1994.

Population-II: [**Source:** [8]]

Y=No of beds allocated for Covid patients, *X*= Beds allocated for Covid, *Z*= Beds used by patient,

Population-III: [Source: [6]]

Y=Agricultural maize in 1964, *X*=agriculture maize in 1963, R_{X1}=rank of X_1.

Table 1: Summary of Populations I-IV

Parameters	I	II	III
N	51	36	34
n	12	8	8
f	0.08333333	0.125	0.125
$\bar{Y}$	13903.24	660.1389	199.4412
$\bar{X}$	15464.06	215.6389	208.8824
$\bar{u}$	61284.25	3645.545	817.6256
$\bar{v}$	64373.61	1308.779	864.0967
C_u	0.02931779	0.481517	0.07564295
C_v	0.04196459	0.4155077	0.07429192
ρ_{uv}	0.735276	0.1726467	0.9018024
S_u	1796.718	317.8681	61.84761
S_v	2571.768	89.59963	60.74297
C_u^2	0.0008595326	0.2318587	0.005721856
C_v^2	0.001761027	0.1726467	0.005519289

Table 2: MSE and PRE based on real data

	I		II		III	
Estimator	**MSE**	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
T_{1pps}	12630.02	100	13845.63	100	28.44963	100
T_{2pps}	10851.7	116.3875	13069.11	105.9417	5.496653	517.5809
T_{3pps}	33217.47	38.0222	71356.55	19.40345	106.2875	26.76667
T_{4pps}	9305.647	135.7242	6360.257	217.6899	5.313038	535.4681
T_{5pps}	9389.713	134.5091	6365.571	217.5081	10.11253	281.3305
T_{6pps}	20572.6	61.39241	35509.3	38.99158	60.50795	47.018
T_{7pps}	11175.6	113.0142	10570.78	130.9802	18.32737	155.2303
T_{8pps}	9607.106	131.4653	12672.59	109.2566	24.85712	114.4526
T_{9pps}	9383.771	134.5942	7775.361	178.0706	5.316287	535.1409
T_{10pps}	12176.87	103.7214	8063.308	171.7116	26.44194	107.5928
T_{11pps}	10934.19	115.5094	11507.6	120.3173	20.52	138.6434
T_{12pps}	12195.43	103.5635	8597.544	161.0417	28.0879	101.2878
T_{13pps}	12130.32	104.1194	8336.823	166.0781	26.23207	108.4536
T_{14pps}	12215.5	103.3934	7821.731	177.015	26.61554	106.891
$T_{prop(pps)_{min}}$	8072.972	156.4481	2437.415	568.0458	4.854198	586.083

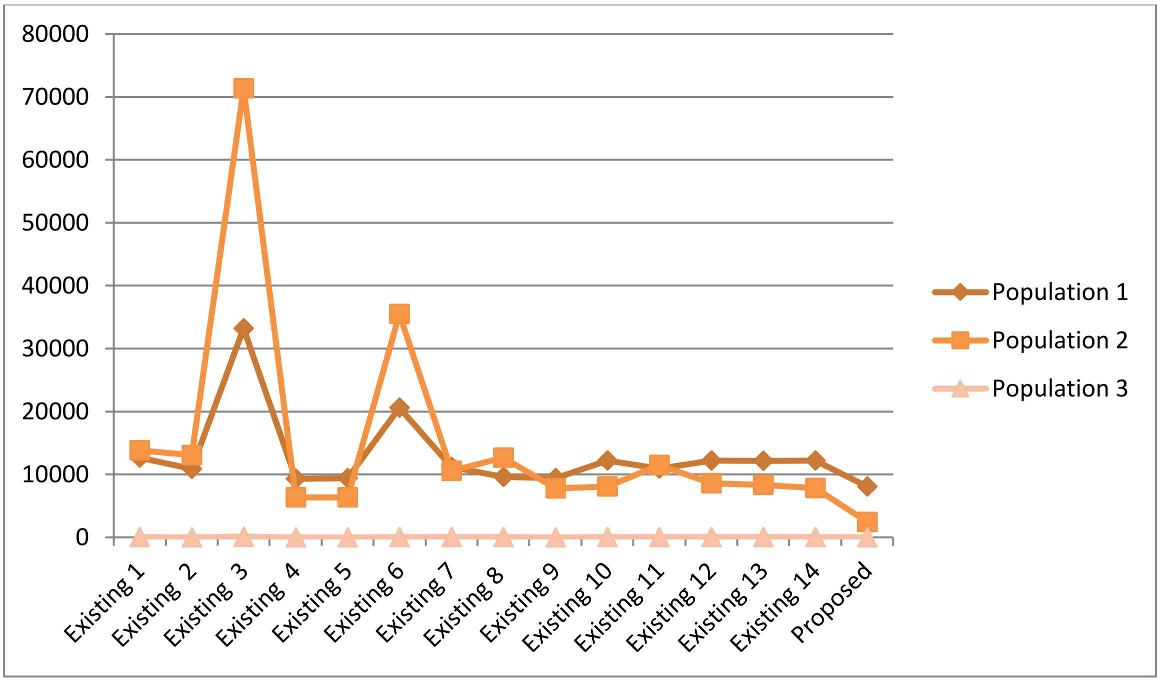

Figure 1: MSEs using actual data sets

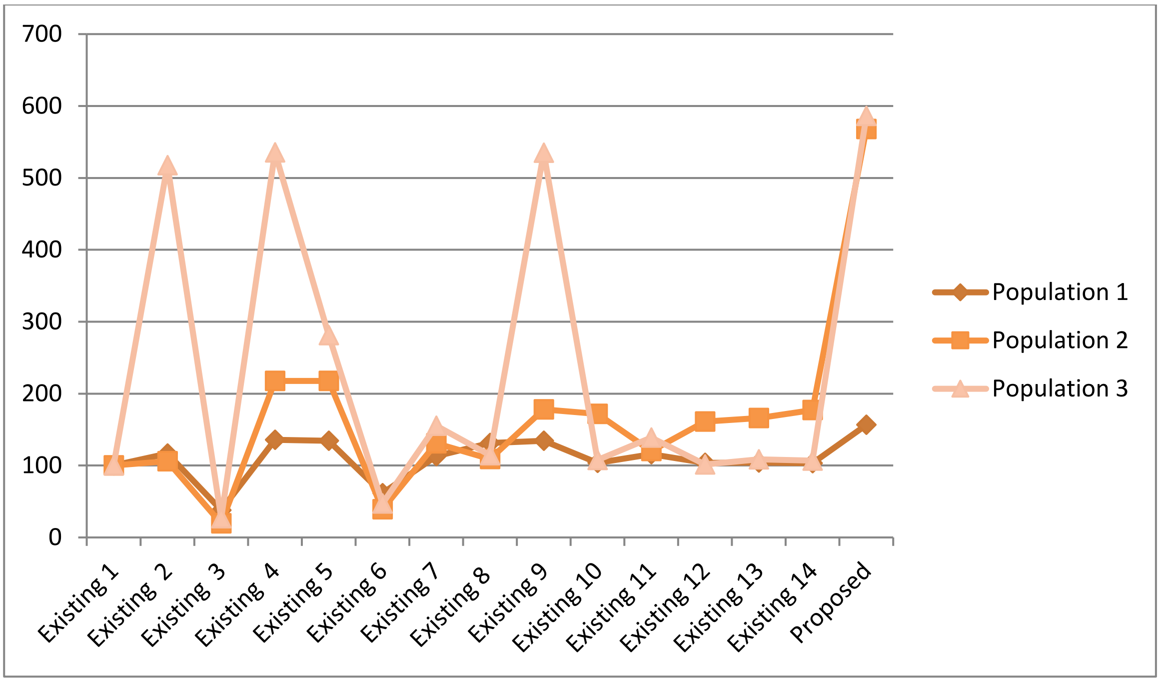

Figure 2: PREs using actual data sets

6. Simulation analysis

A multivariate distribution has been used to construct three populations, each with 5,000 individuals. The means of the population and matrices of covariance are presented below.

Population-I:

$$\mu_1=\begin{bmatrix}500\\500\\500\end{bmatrix}$$

and

$$\sum=\begin{bmatrix}1000 & 800 & 810\\800 & 850 & 820\\810 & 820 & 840\end{bmatrix}$$

Population-II:

$$\mu_1=\begin{bmatrix}500\\500\\500\end{bmatrix}$$

and

$$\sum=\begin{bmatrix}1200 & 870 & 820\\870 & 900 & 800\\820 & 800 & 740\end{bmatrix}$$

Population-III:

$$\mu_1=\begin{bmatrix}500\\500\\500\end{bmatrix}$$

and

$$\sum=\begin{bmatrix}400 & 270 & 220\\270 & 500 & 300\\220 & 300 & 200\end{bmatrix}$$

Following populations are used in simulation study algorithm to investigate the performance of the proposed estimator $T_{prop(pps)_{min}}$ over the existing estimators T_{ipps} where *i=1,2,3,…,14.*

i. The mean square error (MSE) of the estimator is given by:

MSE$(\bar{y}_i)=\frac{1}{k}\sum_{j=1}^{k}\left(y_{ij}-\bar{y}\right)$, where k=5000.

ii. The percentage relative efficiency (PRE) of estimators as compared to the usual estimator is defined by

PRE= $\frac{Var(T_{1pps})}{MSE(T_i)}$ ×100, where $\left(i = 1,2,3,\ldots,14, T_{prop(pps)_{min}}\right)$.

Table 3: MSE and PRE based on simulated data

	I		II		III	
Estimator	**MSE**	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
T_{1pps}	0.4923419	100	0.7099783	100	0.7095295	100
T_{2pps}	0.1019253	483.0421	0.08038802	883.1892	0.1986832	357.1161
T_{3pps}	1.754176	28.06684	2.52683	28.09759	1.8502470	38.34782
T_{4pps}	0.1007482	488.6854	0.07984345	889.213	0.1682158	421.7971
T_{5pps}	0.1882064	261.5968	0.2467755	287.7021	0.3753724	189.0202
T_{6pps}	1.014332	48.53854	1.469996	48.29796	1.201155	59.07063
T_{7pps}	0.3210197	153.3681	0.4342943	163.4786	0.4727048	150.0999
T_{8pps}	0.468335	105.126	0.6685326	106.1995	0.6890009	102.9795
T_{9pps}	0.1020816	482.3025	0.08433732	841.8318	0.2283451	310.7268
T_{10pps}	0.419834	117.2706	0.6021761	117.9021	0.6360148	111.5587
T_{11pps}	0.2645815	186.033	0.3669585	193.4765	0.4683221	151.5046
T_{12pps}	0.4867002	101.1592	0.7002404	101.3907	0.704722	100.6822
T_{13pps}	0.4129308	119.2311	0.5918615	119.9568	0.628896	112.8214
T_{14pps}	0.4256358	115.6721	0.6108373	116.2303	0.6419799	110.5221
$T_{prop(pps)_{min}}$	0.08595412	572.7962	0.02592395	2738.697	0.0552664	1283.834

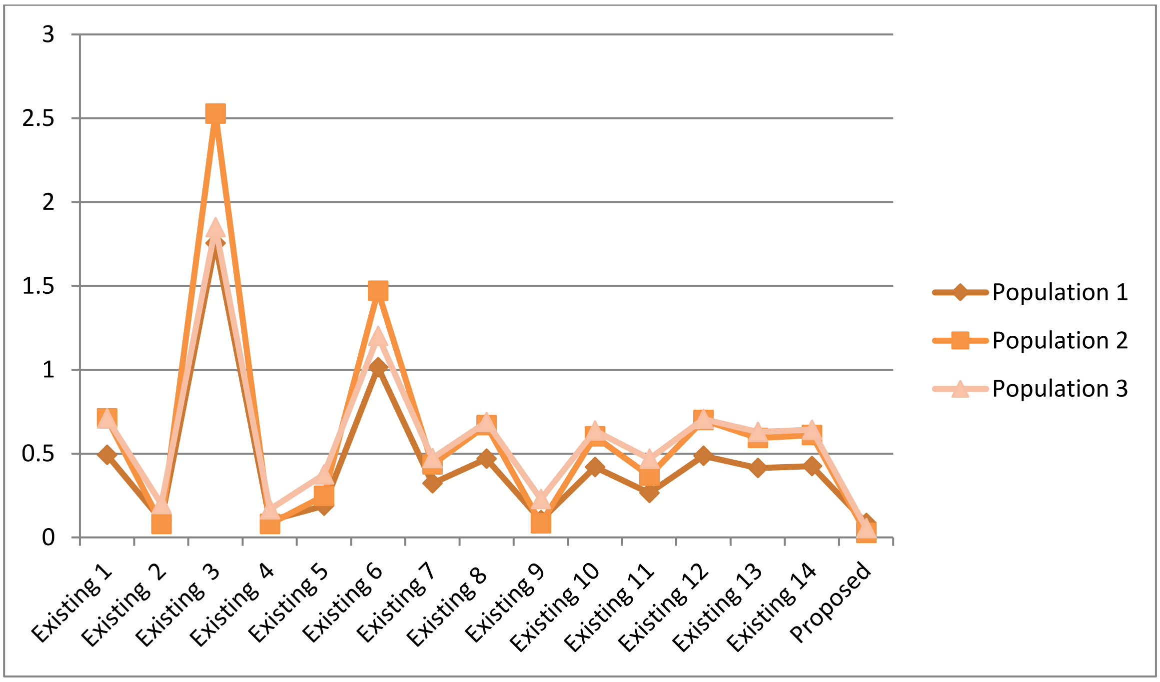

Figure 3: MSEs of estimators using simulation data sets

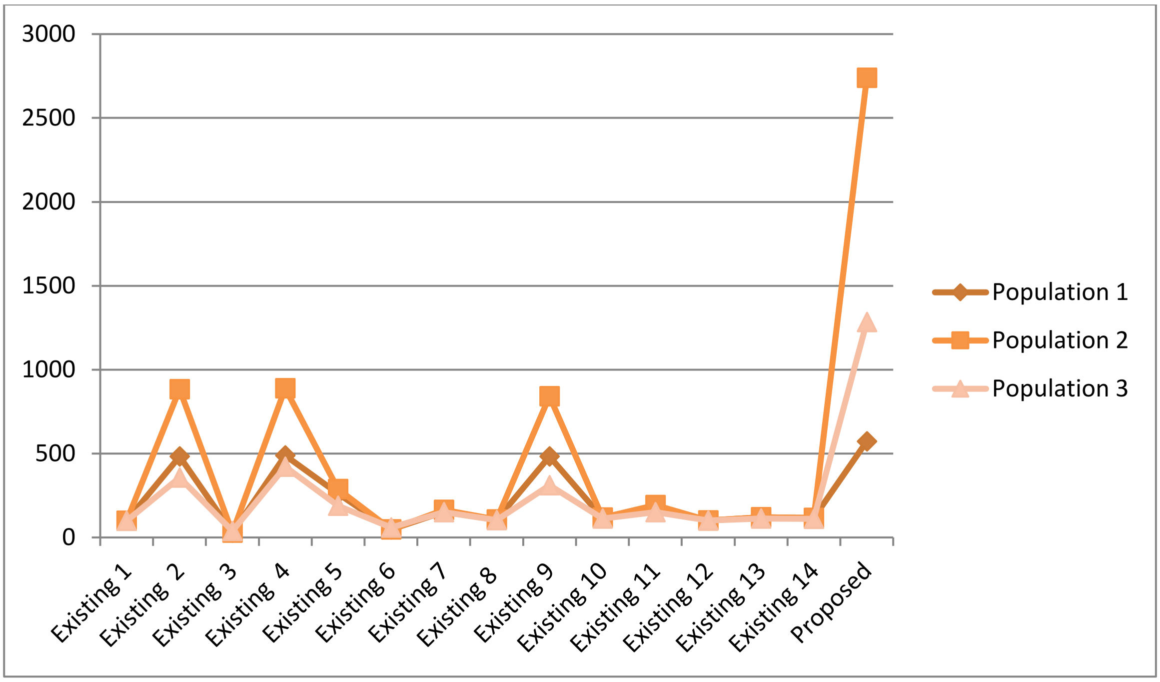

Figure 4: PREs of estimators using simulation data sets

7. Discussion

This article introduced a logarithmic-type estimator specifically for the purpose of computing the finite population mean of an investigated variable. By including information from an auxiliary variable, this estimator improves the mean estimate's accuracy. In order to make survey results more reliable, this study highlights the need of mean estimation. We not only analyzed the existing variance estimators in the literature and derived their mean squared errors (MSEs), but we also suggested an estimator and calculated its bias and MSEs. To determine the performance and effectiveness of the estimators, including the proposed estimator, we calculated their MSE and PRE values using a real data set. The MSE value is low as compared to the other, as seen in Table 2. The high PRE further proves that the proposed estimator is significant. Additionally, simulation simulations allow us to prove that the proposed estimator is effective. The suggested estimator is more effective than the existing estimators, as shown in Table 3. Here, we used real-world data sets to run a simulation and generated data following a normal distribution. Then, for every estimator that was considered and suggested, we calculated its MSE and PRE. We also visualize the efficiency of estimators with the help of line graphs. For more details see Figures 1-4. From Tables 2 and 3, we can observe that the MSE values differ for different estimators and parameter combinations. For example, estimator $T_{prop(pps)_{min}}$ appears to generate less variable and more accurate estimates due to its relatively minimum MSE values in comparison to other estimators. The T_{3pps} has very high MSE values, which is indicative of poor accuracy when compared to other estimators.

8. Conclusion

In this article we have recommended a robust estimator for estimation of population mean using auxiliary variable under PPS sampling. We compared the suggested estimators with existing ones using real data sets to find out how well they worked and how much better they were. The robustness of the proposed estimator was also investigated in a simulation exercise that covered a range of scenarios. According on the results of the simulations and the empirical data sets, the proposed estimator outperformed the existing ones. Therefore, the recommended estimator is ideal for employing in future investigations. The proposed estimator can be used to better estimate of population mean in a wide variety of contexts, such as agriculture, biological sciences, business, medicine, etc. The present work can be easily comprehensive to estimate population proportion based on simple random sampling. Another extension of the current work is to estimate population mean using calibration approach under simple random sampling.

Data availability
The datasets are available on reasonable request.
Conflict of interest
The authors declares no conflict of interest

References

1. Chakrabarty, R. P. (1979). Some ratio-type estimators. *Journal of the Indian Society of Agricultural Statistics*, *25*, 49-57.
2. Sisodia, B. V. S., & Dwivedi, V. K. (1981). Modified ratio estimator using coefficient of variation of auxiliary variable. *Journal-Indian Society of Agricultural Statistics*, *33*(2).
3. Kadilar, C., & Cingi, H. (2007). An improvement in estimating the population mean by using the correlation coefficient. *Quality control and applied statistics*, *52*(4), 443-446.

a. I. Al-Omari, A. A. Jemain, and K. Ibrahim, "New ratio estimators of the mean using simple random sampling and ranked set sampling methods," Investig. Oper. **30**, 97–108 (2009).

4. Subramani, J., & Kumarapandiyan, G. (2012). Modified ratio estimators for population mean using function of quartiles of auxiliary variable. *Bonfring International Journal of Industrial Engineering and Management Science*, *2*(2), 19-23.
5. Murthy, M. N. (1967). *Sampling theory and methods.* Calcutta-35: Statistical Publishing Society, 204/1, Barrackpore Trunk Road, India.
6. Bahl, S and Tuteja, R. (1991). Ratio and Product type exponential estimators. *Journal of Information of Optimization Sciences,* 12(1): 159-164.
7. Punjab Bureau of Statistics (PBS), Punjab, Pakistan (2021-2022).
8. Singh, S. (2003). *Advanced Sampling Theory With Applications: How Michael"" Selected"" Amy* (Vol. 2). Springer Science & Business Media.
9. Raja, T. A., & Maqbool, S. (2024). Ratio Type Estimators for Estimating Population Mean by Using Some Auxiliary Information. *International Journal of Agriculture Sciences, ISSN*, 0975-3710.
10. Kumar, A., & Siddiqui, A. S. (2024). Enhanced estimation of population mean using simple random sampling. *Research in Statistics*, *2*(1), 2335949.
11. Iqbal, K., Muslim Raza, S. M., Butt, M. M., Ahmad, H., & Askar, S. (2024). On exploring the generalized mixture estimators under simple random sampling and application in health and finance sector. *AIP Advances*, *14*(1).
12. Gupta, R., Ali, M., Yadav, S. K., Kumar, G., & Kumar, S. (2024). Improved Estimation Of The Population Mean By Integrating Additional Auxiliary Parameters. *Korean Journal of Physiology and Pharmacology*, *28*(1), 205-210.
13. Lawson, N. (2022). An improved class of population mean estimators by utilizing some prior information in simple random sampling using Searl's approach. *Lobachevskii Journal of Mathematics*, *43*(11), 3376-3383.
14. Kumar, A., Bhushan, S., Mustafa, M. S., Aldallal, R., Aljohani, H. M., & Almulhim, F. A. (2024). Novel imputation methods under stratified simple random sampling. *Alexandria Engineering Journal*, *95*, 236-246.
15. Kumar, A., Siddiqui, A. S., Mustafa, M. S., Hussam, E., Aljohani, H. M., & Almulhim, F. A. (2024). Mean estimation using an efficient class of estimators based on simple random sampling: Simulation and applications. *Alexandria Engineering Journal*, *91*, 197-203.
16. Zaman, T. (2018). New family of estimators using two auxiliary attributes. *International Journal of Advanced Research I Engineering & Management (IJAREM)*, *4*(11), 11-16.
17. Zaman, T., & Toksoy, E. (2019). Improvement in estimating the population mean in simple random sampling using information on two auxiliary attributes and numerical application in agricultural engineering. *Fresenius Environmental Bulletin*, *28*(6), 4584-4590.
18. Ahmad, S., Shabbir, J., Zahid, E., & Aamir, M. (2023). Improved family of estimators for the population mean using supplementary variables under PPS sampling. *Science Progress*, *106*(2), 00368504231180085.
19. Ahmad, S., Zahid, E., Shabbir, J., Aamir, M., & Onyango, R. (2023). Enhanced estimation of the population mean using two auxiliary variables under probability proportional to size sampling. *Mathematical Problems in Engineering*, *2023*.
20. Singh, H. P., Mishra, A. C., & Pal, S. K. (2018). Improved estimator of population total in PPS sampling. *Communications in Statistics-Theory and Methods*, *47*(4), 912-934.
21. Sinha, R. R., & Khanna, B. (2022). Estimation of population mean under probability proportional to size sampling with and without measurement errors. *Concurrency and Computation: Practice and Experience*, *34*(18), e7023.
22. Olayiwolla, O. M., Apantaku, F. S., & Wale-Orojo, O. A. (2019). Probability proportional to size (PPS) method to enhance efficiency of estimator in two stage sampling. *Ann Comput Sci Ser*, *17*, 311-315.
23. Wang, J., Ahmad, S., Arslan, M., Lone, S. A., Abd Ellah, A. H., Aldahlan, M. A., & Elgarhy, M. (2023). Estimation of finite population mean using double sampling under probability proportional to size sampling in the presence of extreme values. *Heliyon*, *9*(11).
24. Latpate, R., Kshirsagar, J., Kumar Gupta, V., Chandra, G., Latpate, R., Kshirsagar, J., ... & Chandra, G. (2021). Probability proportional to size sampling . *Advanced sampling methods*, 85-98.
25. Ahmad, S., & Shabbir, J. (2018). Use of extreme values to estimate finite population mean under PPS sampling scheme. *Journal of Reliability and Statistical Studies*, 99-112.
26. Ahmad, S., Hussain, S., Aamir, M., Yasmeen, U., Shabbir, J., & Ahmad, Z. (2021). Dual use of auxiliary information for estimating the finite population mean under the stratified random sampling scheme. *Journal of Mathematics*, *2021*(1), 3860122.
27. Hussain, S., Ahmad, S., Saleem, M., & Akhtar, S. (2020). Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. *Plos one*, *15*(9), e0239098.
28. Hussain, S., Ahmad, S., Akhtar, S., Javed, A., & Yasmeen, U. (2020). Estimation of finite population distribution function with dual use of auxiliary information under non-response. *Plos one*, *15*(12), e0243584.
29. Ahmad, S., Hussain, S., Ullah, K., Zahid, E., Aamir, M., Shabbir, J., ... & Alajlan, W. (2022). A simulation study: Improved ratio-in-regression type variance estimator based on dual use of auxiliary variable under simple random sampling. *PLoS One*, *17*(11), e0276540.

30. Ahmad, S., Hussain, S., & Ahmad, S. (2021). Finite population distribution function estimation using auxiliary information under simple random sampling. *Statistics, Computing and Interdisciplinary Research*, *3*(1), 29-38.
31. Ahmad, S., Shabbir, J., Aamir, M., Hussain, S., & Ullah, K. (2024). Improved mean estimation for a finite population with twofold auxiliary variables under simple random sampling. *International Journal of Applied and Computational Mathematics*, *10*(1), 31.
32. Ahmad, S., Hussain, S., Aamir, M., Khan, F., Alshahrani, M. N., & Alqawba, M. (2022). Estimation of finite population mean using dual auxiliary variable for non-response using simple random sampling. *Aims Mathematics*, *7*(3), 4592-4613.
33. Hussain, S., Zichuan, M., Hussain, S., Iftikhar, A., Asif, M., Akhtar, S., & Ahmad, S. (2020). On estimation of distribution function using dual auxiliary information under nonresponse using simple random sampling. *Journal of Probability and Statistics*, *2020*(1), 1693612.
34. Iftikhar, A., Hongbo, S., Hussain, S., Khan, S., Alnssyan, B., Shabbir, J., ... & Aamir, M. (2022). A New Exponential Factor-Type Estimator for Population Distribution Function Using Dual Auxiliary Variables under Stratified Random Sampling. *Mathematical Problems in Engineering*, *2022*(1), 7673945.
35. Semary, H. E., Ahmad, S., Hamdi, W. A., Albalawi, O., Elbatal, I., Chesneau, C., & Almarzouki, S. M. (2024). Improved novel estimation for estimation of population distribution function using auxiliary information under stratified sampling strategy. *Journal of Radiation Research and Applied Sciences*, *17*(3), 101011.
36. Ahmad, S., Adichwal, N. K., Aamir, M., Shabbir, J., Alsadat, N., Elgarhy, M., & Ahmad, H. (2023). An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling. *Scientific Reports*, *13*(1), 21444.
37. Ahmad, S., Ullah, K., Zahid, E., Shabbir, J., Aamir, M., Alshanbari, H. M., & El-Bagoury, A. A. A. H. (2023). A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling. *Scientific Reports*, *13*(1), 5415.
38. Ahmad, S., & Shabbir, J. (2024). Proportion Estimation Using Enhanced Class of Estimators Under Simple Random Sampling: Application with Real Data Sets and Simulation. *Measurement: Interdisciplinary Research and Perspectives*, 1-13.
39. Ahmad, S., Masood, S., Alomair, A. M., & Alomair, M. A. (2024). Improved modified estimator for estimation of median using auxiliary information under simple random sampling. *Scientific Reports*, *14*(1), 16504.
40. Semary, H. E., Ahmad, S., Elbatal, I., Chesneau, C., Elgarhy, M., & Almetwally, E. M. (2024). New comprehensive class of estimators for population proportion using auxiliary attribute: Simulation and an application. *Alexandria Engineering Journal*, *99*, 130-136.