Taylor & Francis
Taylor & Francis Group

Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Estimation of the mean using robust regression and probability proportional to size sampling

Mohamud Hussein Mohamud & Fartun Ahmed Mohamud

To cite this article: Mohamud Hussein Mohamud & Fartun Ahmed Mohamud (20 Jun 2025): Estimation of the mean using robust regression and probability proportional to size sampling, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2025.2516339

To link to this article: https://doi.org/10.1080/24754269.2025.2516339

Published online: 20 Jun 2025.

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2025.2516339

Taylor & Francis
Taylor & Francis Group

Estimation of the mean using robust regression and probability proportional to size sampling

Mohamud Hussein Mohamud and Fartun Ahmed Mohamud

Faculty of Economics, SIMAD University, Mogadishu, Somalia

ABSTRACT

In survey studies, mean estimation is a main issue, and regression estimators that use conventional regression coefficients are the preferred options. However, traditional estimates may exhibit undesirable behaviour when outliers are present in the data. For such a situation, robust regression tools are utilized. In this paper, inspired by recent developments, some new finite population mean estimators are proposed by utilizing the robust regression tools under probability proportional to size sampling with replacement scheme. Two real datasets are applied for measuring the percentage relative efficiency of the proposed estimators with respect to the traditional ordinary least square regression mean and adapted estimators. It is found that the proposed estimators are more efficient than the considered estimators. In light of this, the proposed estimators may be valuable and almost certainly increase the chance of obtaining additional accurate population mean estimates in the presence of outliers.

ARTICLE HISTORY
Received 21 October 2024
Accepted 1 June 2025

KEYWORDS
Proportional sampling; mean estimation; least squares regression; robust regression; percentage relative efficiency

1. Introduction

Regardless of whether it is used for descriptive or analytical purposes, the sample must be chosen using appropriate statistical methods in order to achieve the desired precision. Giving each unit of the population a specific non-zero probability is a crucial aspect of sample selection. This technique for gathering a sample is referred to as probability sampling or probability proportional to size (PPS) sampling. Consider the study variable to be the number of factories in a country, and the auxiliary variable to be the number of workers employed in these factories. The most commonly used is varying PPS sampling. The factories are chosen based on the number of employees. One strategy for reducing estimation error in a sampling design is to take selection probabilities that are inversely proportional to some carefully chosen auxiliary/supplementary variable. A well-chosen auxiliary/additional variable is roughly proportional to the study variable. The most significant benefit that comes from PPS sampling is that it increases the chance that the selection procedure will choose the most crucial units. The options include PPS sampling with replacement and PPS sampling without replacement, just like with conventional sampling designs. It is worth mentioning that Neyman (1934) introduced the idea of PPS sampling, and Hansen and Hurwitz (1943) developed the general theory of that sampling with replacement. For interesting details about PPS sampling,

CONTACT Mohamud Hussein Mohamud laalaabte@simad.edu.so Faculty of Economics, SIMAD University, Mogadishu, Somalia

readers may refer to Singh and Mangat (1996). We are interested in probability proportional to size sampling with replacement (PPSWR) to estimate the population mean for the study variable. Estimating the population mean for the study variable using an auxiliary variable or attribute is one of the most important measures of central tendency. Many effective ratio, product, exponential, and regression-type estimators are proposed by, among others, Cekim and Kadilar (2018), Zaman (2018, 2019a), Zaman and Toksoy (2019), Zaman and Kadilar (2019, 2020), Shahzad, Al-Noor, et al. (2021); Shahzad, Shahzadi, et al. (2021), and Ali, Ahmad, Shahzad and Al-Noor (2022). However, when unusual observations or outliers are present in the data, traditional estimates may exhibit undesirable behaviour. In the ordinary least squares (OLS) regression, outliers violate the assumption of normally distributed residuals. They have a tendency to distort the coefficients of least squares by wielding more influence than they deserve. One solution is to use robust tools, which are less sensitive to outliers. Robust regression is a less restrictive assumption-based alternative to regression least squares.

The motivation for this study is that much work is available with OLS and robust regression under simple random sampling, but no work is available with robust regression under PPS sampling. The rest of this paper is structured as follows. The formulation of the OLS based regression estimator under PPSWR and its theoretical mean square error are presented in Section 2. In Sections 3 and 4, respectively, adapted estimators using robust regression and members of the proposed class under PPSWR are offered. In Section 5, numerical illustrations that support the present problem in terms of percentage relative efficiency are provided. Finally, in Section 6, the conclusions are provided.

2. OLS based regression estimator

Let N be the population size and Y be the study variable with a limited population Φ having 1 to N recognizable units with values/observations $y_i, i \in \Phi$. Let (X, Z) be the auxiliary variables, which are employed to estimate the obscure mean of Y, to be precise $\overline{Y}$. Remember that auxiliary information can be made available to the entire population. The auxiliary variable Z is used as a size measure to determine the inclusion probabilities in PPSWR sampling. It helps in assigning larger selection probabilities to units with larger sizes, thus improving the efficiency of estimators by incorporating additional information related to the study variable. Suppose a sample of size n is selected based upon PPSWR. Let u_i and v_i be the study and auxiliary variables for the PPSWR where

$$u_i = \frac{y_i}{NP_i}, \quad v_i = \frac{x_i}{NP_i}.$$

Furthermore, let $\bar{u} = (\sum_{i=1}^{n} u_i)/N$ and $\bar{v} = (\sum_{i=1}^{n} v_i)/N$ be the sample means, $\overline{Y} = (\sum_{i=1}^{N} Y_i)/N$ and $\overline{X} = (\sum_{i=1}^{N} X_i)/N$ be the population means, and $S_u^2 = \sum_{i=1}^{N} P_i(u_i - \overline{Y})^2$ and $S_v^2 = \sum_{i=1}^{N} P_i(v_i - \overline{X})^2$ be the variances of the variables u and v, respectively, in which P_i denotes the inclusion probability of the ith unit under PPSWR sampling. As is mentioned above, P_i is proportional to the auxiliary size measure for the ith unit, Z_i, i.e., $P_i = Z_i/(\sum_{i=1}^{N} Z_i)$. The correlation between u and v is $\rho = (\sum_{i=1}^{N} P_i(u_i - \overline{Y})(v_i - \overline{X}))/S_u S_v$. Let $C_u = S_u/\overline{Y}, C_v = S_v/\overline{X}$ be the coefficient of variation of the variables u and v, respectively.

The traditional OLS based regression estimator is

$$\hat{\bar{y}}_{\text{reg}} = \bar{u} + b_{\text{OLS}}(\overline{X} - \bar{v}). \tag{1}$$

To attain MSE for $\hat{\bar{y}}_{\text{reg}}$, let's define

$$\eta_y = \frac{\bar{u} - \overline{Y}}{\overline{Y}}, \quad \eta_x = \frac{\bar{v} - \overline{X}}{\overline{X}},$$

such that

$$E(\eta_y) = E(\eta_x) = 0,$$
$$E(\eta_y^2) = \theta C_u^2, \quad E(\eta_x^2) = \theta C_v^2, \quad E(\eta_y \eta_x) = \theta \rho C_u C_v,$$

with $\theta = \frac{1}{N}$. Expand $\hat{\bar{y}}_{\text{reg}}$ in terms of η's as given below

$$\hat{\bar{y}}_{\text{reg}} = \overline{Y}\left(1 + \eta_y\right) - b_{\text{OLS}}\overline{X}\eta_x,$$
$$\hat{\bar{y}}_{\text{reg}} - \overline{Y} = \overline{Y}\left(1 + \eta_y\right) - b_{\text{OLS}}\overline{X}\eta_x - \overline{Y}.$$

After squaring the aforementioned expression, applying expectation, and simplifying the expression, the following MSE expression is obtained

$$\text{MSE}(\hat{\bar{y}}_{\text{reg}}) = \theta[S_u^2 + B_{\text{OLS}}^2 S_v^2 - 2\rho B_{\text{OLS}} S_u S_v], \tag{2}$$

where $B_{\text{OLS}} = \text{Cov}(X, Y)/\text{Var}(X)$ which denotes the theoretical population-level counterpart of b_{OLS}.

Due to its straightforward mathematics and easy execution, OLS regression is undoubtedly one of the most popular techniques, but because of its high sensitivity to outliers, its results become unreliable. Robust regression methods, in contrast, are designed to cope with the potential presence of outliers.

3. Adapted estimators using robust regression under PPSWR

In the circumstance that OLS presumptions are broken, robust regression is employed. The robust-regression tools perform better in these situations because outliers are given less weight, leading to more accurate results.

For population mean estimation, Zaman and Bulut (2019) constructed a class based on some robust regression tools including least absolute deviations (LAD), least median of squares (LMS), least trimmed squares (LTS), M estimation (M-Huber(HBM)), M-Hampel (HPM), M-Tukey (TKY)), and MM-Huber (HMM). The primary objectives regarding each tool are as follows.

- LAD: Minimize the absolute residuals.
- LMS: Minimize the median of squared residuals.
- ITS: After arranging the squared errors, OLS is run by utilizing observations based on the first (smallest) specified errors.
- M estimation: Minimize residuals' objective function (say q) that is constructed under some conditions. Numerous q functions can be found in the literature, including Huber (1964, 1973), Hampel (1971), and Tukey (1977).
- MM (Modified M) estimation: Obtain an estimator that has a high breakdown value and is more efficient with the help of weighted least squares. Readers with an interest

Table 1. Adapted estimators under PPSWR.

i	Estimators	MSE
1	$\bar{y}_1 = \bar{u} + b_{\text{LAD}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_1) = \theta[S_u^2 + B_{\text{LAD}}^2 S_v^2 - 2\rho B_{\text{LAD}} S_u S_v]$
2	$\bar{y}_2 = \bar{u} + b_{\text{LMS}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_2) = \theta[S_u^2 + B_{\text{LMS}}^2 S_v^2 - 2\rho B_{\text{LMS}} S_u S_v]$
3	$\bar{y}_3 = \bar{u} + b_{\text{LTS}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_3) = \theta[S_u^2 + B_{\text{LTS}}^2 S_v^2 - 2\rho B_{\text{LTS}} S_u S_v]$
4	$\bar{y}_4 = \bar{u} + b_{\text{HBM}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_4) = \theta[S_u^2 + B_{\text{HBM}}^2 S_v^2 - 2\rho B_{\text{HBM}} S_u S_v]$
5	$\bar{y}_5 = \bar{u} + b_{\text{HPM}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_5) = \theta[S_u^2 + B_{\text{HPM}}^2 S_v^2 - 2\rho B_{\text{HPM}} S_u S_v]$
6	$\bar{y}_6 = \bar{u} + b_{\text{TKY}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_6) = \theta[S_u^2 + B_{\text{TKY}}^2 S_v^2 - 2\rho B_{\text{TKY}} S_u S_v]$
7	$\bar{y}_7 = \bar{u} + b_{\text{HMM}}(\overline{X} - \bar{v})$	$\text{MSE}(\bar{y}_7) = \theta[S_u^2 + B_{\text{HMM}}^2 S_v^2 - 2\rho B_{\text{HMM}} S_u S_v]$

in MM estimation can refer to Yohai (1987) for more information. Moreover, for recent interesting papers using robust regression tools with mean estimation, one can refer to Zaman (2019b), Zaman and Bulut (2020), Zaman et al. (2021), Ali et al. (2021); Ali, Ahmad, Shahzad, Al-Noor and Hanif (2022), and Shahzad et al. (2022).

Motivated by estimators under simple random sampling of Zaman and Bulut (2019) and Ali et al. (2021), the adapted class of robust regression mean estimators under PPSWR is

$$\bar{y}_i = \bar{u} + b_i(\overline{X} - \bar{v}) \quad \text{for } i = 1, \ldots, 7, \tag{3}$$

where all the seven adapted estimators $i = 1, \ldots, 7$ rely respectively on the aforementioned robust regression coefficients, i.e., LAD, LMS, LTS, HBM, HPM, TKY, and HMM.

Taking advantage of known results, along with some straightforward algebra, and avoiding tedious or pointless calculations, we can present the MSE expressions of the proposed class of estimators, up to order n^{-1}, as

$$\text{MSE}(\bar{y}_i) = \theta[S_u^2 + B_i^2 S_v^2 - 2\rho B_i S_u S_v] \quad \text{for } i = 1, \ldots, 7. \tag{4}$$

The members of the adapted class are presented in Table 1.

4. Proposed estimators using robust regression under PPSWR

Survey statisticians have discovered that PPS sampling is useful for selecting units from the population as well as estimating parameters of interest, particularly when the survey is large in size and includes multiple attributes. Our aim is to provide estimators for the population mean by combining the robust regression tools under PPSWR sampling.

By extending the idea of adapted estimators, we propose the class of robust regression mean estimators under PPSWR as

$$\bar{y}_{N_i} = a\bar{u} + b_i(\overline{X} - \bar{v}) \quad \text{for } i = 1, \ldots, 7. \tag{5}$$

Table 2. Proposed estimators under PPSWR.

i	Estimators	MSE
1	$\bar{y}_{N_1} = a\bar{u} + b_{\text{LAD}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_1}) = \theta B^2_{\text{LAD}} S^2_v (1-\rho^2)$
2	$\bar{y}_{N_2} = a\bar{u} + b_{\text{LMS}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_2}) = \theta B^2_{\text{LMS}} S^2_v (1-\rho^2)$
3	$\bar{y}_{N_3} = a\bar{u} + b_{\text{LTS}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_3}) = \theta B^2_{\text{LTS}} S^2_v (1-\rho^2)$
4	$\bar{y}_{N_4} = a\bar{u} + b_{\text{HBM}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_4}) = \theta B^2_{\text{HBM}} S^2_v (1-\rho^2)$
5	$\bar{y}_{N_5} = a\bar{u} + b_{\text{HPM}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_5}) = \theta B^2_{\text{HPM}} S^2_v (1-\rho^2)$
6	$\bar{y}_{N_6} = a\bar{u} + b_{\text{TKY}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_6}) = \theta B^2_{\text{TKY}} S^2_v (1-\rho^2)$
7	$\bar{y}_{N_7} = a\bar{u} + b_{\text{HMM}}(\overline{X} - \bar{v})$	$\text{MSE}_{\min}(\bar{y}_{N_7}) = \theta B^2_{\text{HMM}} S^2_v (1-\rho^2)$

Using Taylor series of expansion, MSE of the proposed estimator is defined as

$$\begin{aligned} h(\bar{v},\bar{u}) - h(\overline{X},\overline{Y}) &= \left[\frac{\partial h(\bar{v},\bar{u})}{\partial \bar{v}}\right]_{\overline{X},\overline{Y}} (\bar{v}-\overline{X}) + \left[\frac{\partial h(\bar{v},\bar{u})}{\partial \bar{u}}\right]_{\overline{X},\overline{Y}} (\bar{u}-\overline{Y}), \\ &= \left[\frac{\partial h(a\bar{u} + b_i(\overline{X}-\bar{v}))}{\partial \bar{v}}\right]_{\overline{X},\overline{Y}} (\bar{v}-\overline{X}) \\ &\quad + \left[\frac{\partial h(a\bar{u} + b_i(\overline{X}-\bar{v}))}{\partial \bar{u}}\right]_{\overline{X},\overline{Y}} (\bar{u}-\overline{Y}). \end{aligned} \tag{6}$$

Now, by partially differentiating the first and second terms of Equation (6) w.r.t. $\bar{u}$ and $\bar{v}$, respectively, we obtain

$$h(\bar{v},\bar{u}) - h(\overline{X},\overline{Y}) = -b_i(\bar{v}-\overline{X}) + a(\bar{u}-\overline{Y}). \tag{7}$$

Squaring and taking expectation to both sides, we obtain the MSE of the proposed estimator as

$$\text{MSE}(\bar{y}_{N_i}) = \theta\left[B_i^2 S_v^2 + a^2 S_u^2 - 2B_i a S_{vu}\right] \quad \text{for } i = 1,\dots,7, \tag{8}$$

where $S_{vu} = \text{Cov}(\bar{v},\bar{u}) = E[(\bar{v}-\mu_v)(\bar{u}-\mu_u)]$. Partially differentiating Equation (8) to obtain the optimum value that minimizes the MSE of $\bar{y}_{N_i}$, we get

$$a^{\text{opt}} = B_i \frac{S_{vu}}{S_u^2} \quad \text{for } i = 1,\dots,7. \tag{9}$$

Substituting the value of Equation (8), the minimum MSE of the proposed estimator is given by

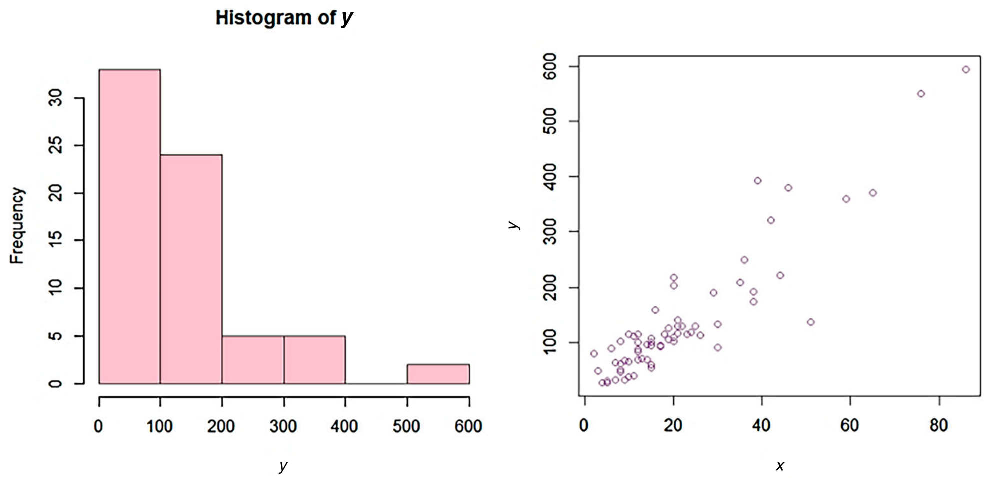
$$\text{MSE}_{\min}(\bar{y}_{N_i}) = \theta B_i^2 S_v^2 (1-\rho^2) \quad \text{for } i = 1,\dots,7. \tag{10}$$

The members of the proposed class are presented in Table 2.

5. Numerical illustrations

Numerical illustrations are performed using two population datasets in order to evaluate the advantages of the proposed class of estimators over the existing ones.

Population 1: Data belong to 69 serially numbered villages (Singh & Mangat, 1996) of Punjab's Doraha development district (India) with the variables, net irrigated area (in hectares),



(a) Non-normality (b) Outliers

Figure 1. Data visualization of Population 1. (a) Non-normality and (b) Outliers.

number of tube wells, and number of tractors. The study variable Y is the number of tube wells and the auxiliary variable X is the number of tractors. Furthermore, we consider another auxiliary variable which is the size measure of net irrigated area, denoted by Z. Using PPSWR sampling, we choose a random sample of 10 villages.

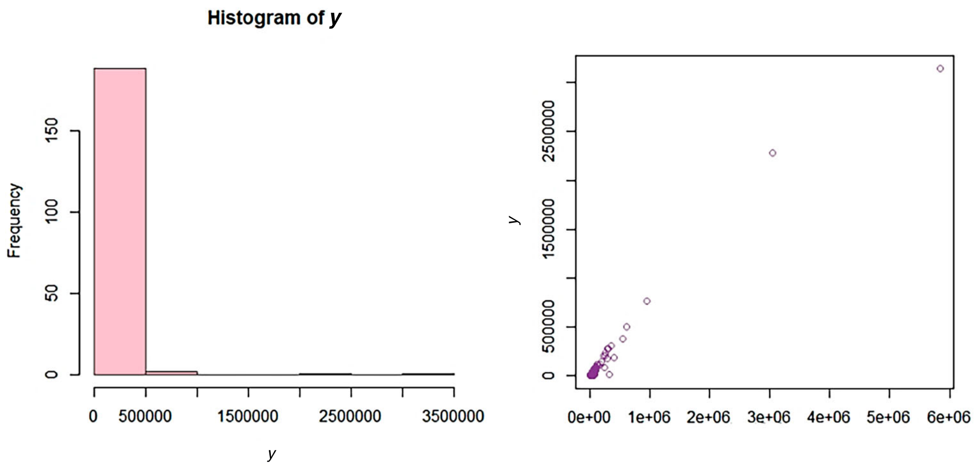
Population 2: COVID-19 data is considered for Africa, Asia, Europe, and North America from January 22, 2020 to August 23, 2020, (Source: https://www.worldometers.info/corona virus), where, across these regions, $Y =$ total recoveries, $X =$ total cases, and $Z =$ total population.

The histograms of Y in Figures 1 and 2 for Populations 1 and 2 display non-normality, and the scatter plots reveal the presence of outliers, making data perfect for assessing our proposed estimators. Furthermore, Table 3 gives some descriptive statistics corresponding to each population. Tables 4 and 5 listed the estimators' percentage relative efficiency (PRE) obtained for each population, where

$$\mathrm{PRE}(T, \bar{y}_{N_i}) = \frac{\mathrm{MSE}(T)}{\mathrm{MSE}(\bar{y}_{N_i})} \times 100 \quad \text{for } T = \hat{\bar{y}}_{\mathrm{reg}} \text{ or } \bar{y}_i \quad \text{and} \quad i = 1, \ldots, 7. \tag{11}$$

It is clear from the values of PRE that proposed estimators based on PPSWR are much more efficient than the OLS regression and adapted estimators. In other words, PRE values are much greater than 100 in all cases with two populations. It can also be observed that the largest values of the PRE of all proposed estimators are with respect to adapted estimators $\bar{y}_2$ and $\bar{y}_3$ respectively that depend on LMS and LTS robust regression tools. Furthermore, the values of PRE of proposed estimators with respect to all adapted estimators are larger than that with respect to the OLS regression estimator.

In particular, with respect to OLS regression and adapted estimators, it is evident from Table 1 that the proposed estimators $\bar{y}_{N_5}$ and $\bar{y}_{N_4}$ followed by $\bar{y}_{N_6}$ introduced the largest values of the PRE while it is evident from Table 2 that $\bar{y}_{N_6}$ and $\bar{y}_{N_7}$ followed by $\bar{y}_{N_1}$ introduced the largest values of the PRE.



(a) Non-normality (b) Outliers

Figure 2. Data visualization of Population 2. (a) Non-normality and (b) Outliers.

Table 3. Descriptive statistics corresponding to Populations 1 and 2.

Statistics	Population 1	Population 2
N	69	193
n	10	20
$\overline{Y}$	135.26	60279.26
$\overline{X}$	21.23	91424.59
S_u^2	2119.15	9.15e9
S_v^2	63.93	2.32e10
ρ	0.57	0.95
B_{OLS}	3.30	0.60
B_{LAD}	4.94	0.84
B_{LMS}	5.38	0.94
B_{LTS}	5.31	0.93
B_{HBM}	3.84	0.85
B_{HPM}	3.62	0.88
B_{TKY}	4.01	0.84
B_{HMM}	4.05	0.84

Table 4. PRE of $\bar{y}_{N_i}$ w.r.t. $\hat{\bar{y}}_{\mathrm{reg}}$ and $\bar{y}_i$ in Population 1.

Estimators	$\bar{y}_{N_1}$	$\bar{y}_{N_2}$	$\bar{y}_{N_3}$	$\bar{y}_{N_4}$	$\bar{y}_{N_5}$	$\bar{y}_{N_6}$	$\bar{y}_{N_7}$
$\hat{\bar{y}}_{\mathrm{reg}}$	135.95	114.51	117.36	224.32	253.29	206.01	202.37
$\bar{y}_1$	152.34	128.31	131.50	251.36	283.82	230.83	226.76
$\bar{y}_2$	162.39	136.78	140.18	267.94	302.54	246.06	241.72
$\bar{y}_3$	160.75	135.39	138.76	265.23	299.48	243.57	239.27
$\bar{y}_4$	137.76	116.03	118.92	227.31	256.66	208.74	205.06
$\bar{y}_5$	136.57	115.03	117.89	225.34	254.44	206.94	203.29
$\bar{y}_6$	139.05	117.12	120.03	229.42	259.05	210.69	206.97
$\bar{y}_7$	139.37	117.38	120.31	229.95	259.65	211.17	207.45

6. Conclusions

In this paper, we introduce a new and improved class of estimators for mean estimation based on robust regression tools under PPSWR when data is contaminated by outliers, drawing

Table 5. PRE of $\bar{y}_{N_i}$ w.r.t. $\hat{\bar{y}}_{reg}$ and $\bar{y}_i$ in Population 2.

Estimators	$\bar{y}_{N_1}$	$\bar{y}_{N_2}$	$\bar{y}_{N_3}$	$\bar{y}_{N_4}$	$\bar{y}_{N_5}$	$\bar{y}_{N_6}$	$\bar{y}_{N_7}$
$\hat{\bar{y}}_{reg}$	105.83	104.83	105.19	104.85	101.19	106.25	106.19
$\bar{y}_1$	146.50	117.64	118.57	143.91	134.31	147.60	147.44
$\bar{y}_2$	234.29	188.14	189.63	230.16	214.81	236.05	235.81
$\bar{y}_3$	230.45	185.05	186.51	226.38	211.27	232.18	231.93
$\bar{y}_4$	152.23	122.24	123.21	149.54	139.56	153.37	153.21
$\bar{y}_5$	176.65	141.85	142.97	173.53	161.95	177.97	177.79
$\bar{y}_6$	144.16	115.76	116.68	141.61	132.16	145.24	145.09
$\bar{y}_7$	144.49	116.03	116.94	141.94	132.46	145.57	145.42

inspiration from Zaman and Bulut (2019) and Ali et al. (2021). Outliers represent observations that stand out as being inconsistent with the majority of the data and have a significant impact on mean estimation. Two real datasets containing outliers are used for comparing the proposed estimators with traditional OLS regression and seven adapted estimators. According to numerical results of percentage relative efficiency, the proposed estimators based on all robust regression tools provide more efficient results than the other comparative estimators. For both datasets, the proposed estimator $\bar{y}_{N_6}$ based on the M-Tukey function performed well among the top three estimators for estimating the mean population.

Authorship contribution

Mohamud Hussein conceived the presented idea. Mohamud and Fartun performed the statistical analysis, computation. Mohamud Hussein took the lead in crafting the introduction and literature review sections. Fartun did the methodology section, interpretation and conclusion. Additionally, he verified the analytical methods and computations employed throughout the research, ensuring the robustness and accuracy of the study's findings. Both authors discussed the results and contributed to the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors, which is funded by SIMAD University.

Data availability

The data is included within the study for finding the results.

ORCID

Mohamud Hussein Mohamud http://orcid.org/0009-0002-8321-6044

References

Ali, N., Ahmad, I., Hanif, M., & Shahzad, U. (2021). Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information. *Communications in Statistics - Theory and Methods*, *50*(4), 979–992. https://doi.org/10.1080/03610926.2019.1645857

Ali, Nasir, Ahmad, Ishfaq., Shahzad, U., & Al-Noor, N. H. (2022). Some improved estimators for the mean estimation under stratified sampling by using transformations. *Journal of Science and Arts*, *22*(2), 265–288. https://doi.org/10.46939/j.sci.arts

Ali, N., Ahmad, I., Shahzad, U., Al-Noor, N. H., & Hanif, M. (2022). Robust regression type estimators to determine the population mean under simple and two-stage random sampling techniques. *Electronic Journal of Applied Statistical Analysis*, *15*(2), 357–373.

Cekim, H. O., & Kadilar, C. (2018). New families of unbiased estimators in stratified random sampling. *Journal of Statistics and Management Systems*, *21*(8), 1481–1499. https://doi.org/10.1080/09720510.2018.1530176

Hampel, Frank R (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, *42*(6), 1887–1896. https://doi.org/10.1214/aoms/1177693054

Hansen, M. H., & Hurwitz, W. N. (1943). On the theory of sampling from a finite population. *The Annals of Mathematical Statistics*, *14*(4), 333–362. https://doi.org/10.1214/aoms/1177731356

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Mathematical Statistics*, *1*(5), 799–821.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, *97*(4), 558–625. https://doi.org/10.2307/2342192

Shahzad, U., Al-Noor, N. H., Hanif, M., & Sajjad, I. (2021). An exponential family of median based estimators for mean estimation with simple random sampling scheme. *Communications in Statistics - Theory and Methods*, *50*(20), 4890–4899. https://doi.org/10.1080/03610926.2020.1725828

Shahzad, U., Al-Noor, N. H., Hanif, M., Sajjad, I., & Anas, M. M. (2022). Imputation based mean estimators in case of missing data utilizing robust regression and variance-covariance matrices. *Communications in Statistics - Simulation and Computation*, *51*(8), 4276–4295. https://doi.org/10.1080/03610918.2020.1740266

Shahzad, U., Shahzadi, S., Afshan, N., Al-Noor, N. H., Alilah, D. A., Hanif, M., & Anas, M. M. (2021). Poisson regression-based mean estimator. *Mathematical Problems in Engineering*, *22*, 1–6. https://doi.org/10.1155/2021/9769029

Singh, R., & Mangat, N. S. (1996). *Elements of Survey Sampling*. Springer Science & Business Media.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, *15*(2), 642–656. https://doi.org/10.1214/aos/1176350366

Zaman, T. (2018). New family of estimators using two auxiliary attributes. *International Journal of Advanced Research in Engineering and Management*, *4*(11), 11–16.

Zaman, T. (2019a). Efficient estimators of population mean using auxiliary attribute in stratified random sampling. *Advances and Applications in Statistics*, *56*(2), 153–171. https://doi.org/10.17654/AS056020153

Zaman, T. (2019b). Improvement of modified ratio estimators using robust regression methods. *Applied Mathematics and Computation*, *348*(3), 627–631. https://doi.org/10.1016/j.amc.2018.12.037

Zaman, T., & Bulut, H. (2019). Modified ratio estimators using robust regression methods. *Communications in Statistics - Theory and Methods*, *48*(8), 2039–2048. https://doi.org/10.1080/03610926.2018.1441419

Zaman, T., & Bulut, H. (2020). Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling. *Communications in Statistics - Theory and Methods*, *49*(14), 3407–3420. https://doi.org/10.1080/03610926.2019.1588324

Zaman, T., Dunder, E., Audu, A., Alilah, D. A., Shahzad, U., & Hanif, M. (2021). Robust regression-ratio-type estimators of the mean utilizing two auxiliary variables: A simulation study. *Mathematical Problems in Engineering*, *27*(3), 1–9. https://doi.org/10.1155/2021/6383927

Zaman, T., & Kadilar, C. (2019). Novel family of exponential estimators using information of auxiliary attribute. *Journal of Statistics and Management Systems*, *22*(8), 1499–1509. https://doi.org/10.1080/09720510.2019.1621488

Zaman, T., & Kadilar, C. (2020). On estimating the population mean using auxiliary character in stratified random sampling. *Journal of Statistics and Management Systems*, *23*(8), 1415–1426. https://doi.org/10.1080/09720510.2020.1723924

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.