Iraqi Journal of Statistical Sciences

http://stats.uomosul.edu.iq

Estimating the Population Mean in Stratified Median ranked Set Sampling Using Combined and Separate Regression with the Presence of Outliers

Ayman M. AL-Hadidi[1] Rikan A. Ahmed[2]

[1,2]Department of Statistics and Informatics, College of Computer Science and Mathematics University of Mosul, Mosul, Iraq

Article information

Article history:
Received October 26,2024
Revised: January,2025
Accepted April 5,2025
Available online: December 1,2025

Keywords:
Stratified median ranked set sample
Regression estimator
Combined regression estimator
separate regression estimator

Rikan A. Ahmed
rikan.ahmed@uomosul.edu.iq

Abstract

This research aims to demonstrate the high efficiency and accuracy in estimating the limited population mean through the estimates of the separate and combined stratified regression line based on the method of median ranked set sampling to choose a sample that is more representative of the community. With the problem of heterogeneity in the data and containing extreme values (outliers), it is recommended to use stratification of the community and draw samples using the method of sampling the middle ordered from these layers, which is known as the stratified median ranked set sampling (S_tMRSS), which is one of the modified ranked set sampling methods (RSS), where the mean square error (MSE) of the population mean estimator obtained in this way was compared with the MSE value of the population mean estimator obtained through regression estimates using the robust variance and covariance matrix (Minimum Covariance Determinant (MCD), Minimum Volume Ellipsoid (MVE)) to calculate the averages and using robust methods (Huber M, Huber MM, Least Median of Squares (LMS), Least Trimmed Squares (LTS)) to estimate the regression parameter. The simulation results show that the proposed estimator outperforms the robust estimators in most cases because it obtains the lowest values of the mean square error.

DOI: 10.33899/iqjoss.v22i2.54071,

1- Introduction

In order to obtain accurate and efficient estimates about the community. Efficient estimators can be developed by combining the research variable with the auxiliary variable or variables. It is common knowledge that the estimator's efficiency can be increased by making appropriate use of auxiliary data. In this context, estimators of ratios, products, and regressions are appropriate examples. In certain instances, the research variable may be difficult to quantify or prohibitively expensive. Yet, it can be readily ranked for little or no expense. Numerous methods for incorporating auxiliary data to create more effective estimators for the research variable are covered in the literature on ranked set sampling. Ranked set sampling is a logical approach to data collection that significantly improves estimation, as we will see in the results of this. The method of ranking units is based on the values of one of the auxiliary variables correlated to the variable of the study. Also, using complementary or auxiliary information on population units, the population is frequently divided into disconnected subpopulations (stratums) in survey sampling research. If the mean and variance of these subpopulations differ, a stratified sample will be used to create highly accurate population estimators. A stratified ranked set sample S_tRSS is a two-stage procedure that reduces sample variation. The first stage separates the population into fragmented groups (stratums), with ranked set samples of RSS selected from every stratum. It divides the

sample's total variation in this case into between- and within-stratum variations. The second stage divides the within-stratum variance into between- and within-ranking variations from each stratum.
(G.A. McIntyre 1952) was the first to suggest RSS as a technique for increasing the efficiency of the population mean estimator. The topic of calculating the population ratio of the two variables by using the RSS approach was examined by (H.M. Samawi and H.A. Muttlak in 1996). (Muttlak 1997) suggested the Median Ranked Set Sampling (MRSS), which reduces the errors in ranking and increases the efficiency over RSS with perfect ranking for some probability distribution functions. (H.M. Samawi and M.I. Siam 2003) calculated the results of the combined and separate ratio estimates using StRSS. (H.M. Samawi 2010) proposed stratified ranked set sampling (STRS) to create a more efficient estimate for a population mean. Stratified ranked set sampling has been employed by (V. L. Mandowara and N. Mehta 2014) to develop accurate kinds of ratio estimators. D.F. Linder, (H. Samawi, L. Yu, A. Chatterjee, Y. Huang, and R. Vogel 2015) studied the efficacy of stratified bivariate ranked set sampling SBVRSS and stratified simple random sampling StRSS in calculating the population mean using regression methods. Based on StRSS, (K. H. A. N. Lakhkar, J. Shabbir, and S. Gupta 2016) suggested a type separate ratio estimator of the finite population mean. (M. Saini and A. Kumar 2019) investigated the feasibility of employing auxiliary information to propose ratio estimators for the average population in stratified random sampling SRS and stratified ranked set sampling StRSS. (A.E. Cetin and N. Koyuncn 2020) utilized a simulation analysis using an actual data set to examine the suggested performance of ratio type estimators in many stratified ranked set sampling methods. (Ahmed, R.A., and Hussein, S.M. 2022) presented a proposal that aims to estimate the limited community mean of the main variable through stratified ranked set sampling STRSS through the modification that was made to the exponential estimator of the type of documents and the generalized product. (Mahdi, M.H., and Hussein, S.M. 2023) estimated the population mean in stratified random sampling using combined regression with the presence of outliers

2. Material and Methods

Which will be discussed in this research to estimate the linear regression model. Where, we will estimate the population mean of the separate regression estimator as well as the combined regression estimator, which were compared by (Cochrane 1967) using the stratified median ranked set sample proposed by (Muttlak 1997). These methods are used for accurate estimation and analysis of structured data. We will also highlight the importance of this estimator in terms of its accuracy and high efficiency.
For comparison purposes, the estimate used the population mean by the covariance method with the use of the robust covariance and covariance matrix using the minimum covariance determinant (MCD) estimator and the minimum volume ellipsoid (MVE) estimator to calculate the regression averages, and some robust methods (M, MM, LMS, LTS) for estimating the regression parameter in stratified random sampling in the presence of outliers in the data set.
The population mean will be estimated based on the linear regression of using two types of linear regression estimators: separate and combined regression estimators under stratified mean ordered set sampling (S_tMRSS), and thus we get an unbiased population mean estimator, and we also get less variance than the variance computed under simple random sampling (SRS) and stratified random sampling.
In Section 2, we describe the stratified median ranked set sampling method. In Section 3, we calculate the estimates of the separate and combined regression models using stratified median ranked set sampling. Section 4 discusses the steps of the simulation study conducted to compare the proposed method with robust methods and discusses the results of the study.

2.1 Stratified Median Ranked Set Sample "S_tMRSS"

The median ranked set sampling was suggested by (Muttlak 1997). In this method, only median units of the random sets are chosen as the sample for estimation of the population mean. For the odd set sizes, the $(\frac{m+1}{2})$th ranked units are chosen as the median of each set. For even set sizes, the $\left(\frac{m}{2}\right)$th ranked units are chosen from the first $(\frac{m}{2})$ sets, and the $(\frac{m+2}{2})$th ranked units are chosen from the remaining $(\frac{m}{2})$ sets. If necessary, the procedure can be repeated r times, and we have $n = mr$ a sample size.

And (Ibrahim et al. 2010) suggest S_tMRSS a procedure for estimating population mean; this procedure is as follows: Supposing our variable of interest Y belongs to a population of size N. We will divide our population into L mutually exclusive and collectively exhaustive strata with size $N_1, N_2, N_3, \dots, N_L$. Further, we will proceed as follows:

1) Select m_h^2 units from h-th stratum and divide them into m_h sets of each size m_h. Rank the units within each set in increasing scale.

2) When 'm_h' is even, we start the selection of $(\frac{m_h}{2})^{th}$ ranked units from each of the first $\frac{m_h}{2}$ sets and $\{\frac{(m_h+2)}{2}\}^{th}$ ranked units from each of the next $\frac{m_h}{2}$ sets. When 'm_h' is odd, we select $\{\frac{(m_h+1)}{2}\}^{th}$ a ranked unit from all sets.
3) Repeat steps 1) and 2) in each stratum to obtain $m_1 + m_2 + \cdots + m_L = m$ units.
4) Repeat steps 1), 2) and 3) r times to obtain a sample of size $n = rm$.

Now assuming the sample size *m* is odd, then $\boldsymbol{S_tMRSS_o}$ represent **stratified median ranked set sampling**, where the items of $\boldsymbol{S_tMRSS_o}$ for main variable Y and the auxiliary variables X, and suppose that the ranking depends on the auxiliary variable X, as described below.

Note we will use some simple formulas to differentiate between perfect and imperfect ranking. In the case of perfect ranking, we will put the rank index in parentheses, say (i). Otherwise, in imperfect ranking, we will put the rank index in brackets, say [i]. The perfect ranking will always be for the variable X, since it is the known variable that can be used to estimate the variable Y.

$$\boldsymbol{Stratum\ 1}: \left\{ \begin{matrix} \left(Y_{1[1]\left[\frac{m_1+1}{2}\right]j}, X_{1(1)\left(\frac{m_1+1}{2}\right)j}\right)\left(Y_{1[2]\left[\frac{m_1+1}{2}\right]j}, X_{1(2)\left(\frac{m_1+1}{2}\right)j}\right), \dots \left(Y_{1[m_1]\left[\frac{m_1+1}{2}\right]j}, X_{1(m_1)\left(\frac{m_1+1}{2}\right)j}\right) \\ : j = 1,2,\dots,r \end{matrix} \right\}$$

$$\boldsymbol{Stratum\ 2}: \left\{ \begin{matrix} \left(Y_{2[1]\left[\frac{m_2+1}{2}\right]j}, X_{2(1)\left(\frac{m_2+1}{2}\right)j}\right)\left(Y_{2[2]\left[\frac{m_2+1}{2}\right]j}, X_{2(2)\left(\frac{m_2+1}{2}\right)j}\right), \dots \left(Y_{2[m_2]\left[\frac{m_2+1}{2}\right]j}, X_{2(m_2)\left(\frac{m_2+1}{2}\right)j}\right) \\ : j = 1,2,\dots,r \end{matrix} \right\}.$$

.
.

$$\boldsymbol{Stratum\ L}: \left\{ \begin{matrix} \left(Y_{L[1]\left[\frac{m_L+1}{2}\right]j}, X_{L(1)\left(\frac{m_L+1}{2}\right)j}\right)\left(Y_{L[2]\left[\frac{m_L+1}{2}\right]j}, X_{L(2)\left(\frac{m_L+1}{2}\right)j}\right), \dots \left(Y_{L[m_L]\left[\frac{m_L+1}{2}\right]j}, X_{L(m_L)\left(\frac{m_L+1}{2}\right)j}\right) \\ : j = 1,2,\dots,r \end{matrix} \right\}$$

And if the sample size *m* is even, then $\boldsymbol{S_tMRSS_e}$ represent stratified median ranked set sampling, let $k = \frac{m_h}{2}$, where items of $\boldsymbol{S_tMRSS_e}$ as follows.

$$\boldsymbol{Stratum\ 1}: \left\{ \begin{matrix} \left(Y_{1[1]\left[\frac{m_1}{2}\right]j}, X_{1(1)\left(\frac{m_1}{2}\right)j}\right)\left(Y_{1[2]\left[\frac{m_1}{2}\right]j}, X_{1(2)\left(\frac{m_1}{2}\right)j}\right) \dots \left(Y_{1[k]\left[\frac{m_1}{2}\right]j}, X_{1(k)\left(\frac{m_1}{2}\right)j}\right) \\ \left(Y_{1[k+1]\left[\frac{m_1+2}{2}\right]j}, X_{1(k+1)\left(\frac{m_1+2}{2}\right)j}\right)\left(Y_{1[k+2]\left[\frac{m_1+2}{2}\right]j}, X_{1(k+2)\left(\frac{m_1+2}{2}\right)j}\right) \\ \dots \left(Y_{1[m_1]\left[\frac{m_1+2}{2}\right]j}, X_{1(m_1)\left(\frac{m_1+2}{2}\right)j}\right) : j = 1,2,\dots,r \end{matrix} \right\}$$

$$\boldsymbol{Stratum\ 2}: \left\{ \begin{matrix} \left(Y_{2[1]\left[\frac{m_2}{2}\right]j}, X_{2(1)\left(\frac{m_2}{2}\right)j}\right)\left(Y_{2[2]\left[\frac{m_2}{2}\right]j}, X_{2(2)\left(\frac{m_2}{2}\right)j}\right) \dots \left(Y_{2[k]\left[\frac{m_2}{2}\right]j}, X_{2(k)\left(\frac{m_2}{2}\right)j}\right) \\ \left(Y_{2[k+1]\left[\frac{m_2+2}{2}\right]j}, X_{2(k+1)\left(\frac{m_2+2}{2}\right)j}\right)\left(Y_{2[k+2]\left[\frac{m_2+2}{2}\right]j}, X_{2(k+2)\left(\frac{m_2+2}{2}\right)j}\right) \\ \dots \left(Y_{2[m_2]\left[\frac{m_2+2}{2}\right]j}, X_{2(m_2)\left(\frac{m_2+2}{2}\right)j}\right) : j = 1,2,\dots,r \end{matrix} \right\}$$

$$\boldsymbol{Stratum\ L}: \left\{ \begin{matrix} \left(Y_{L[1]\left[\frac{m_L}{2}\right]j}, X_{L(1)\left(\frac{m_L}{2}\right)j}\right)\left(Y_{L[2]\left[\frac{m_L}{2}\right]j}, X_{L(2)\left(\frac{m_L}{2}\right)j}\right) \dots \left(Y_{L[k]\left[\frac{m_L}{2}\right]j}, X_{L(k)\left(\frac{m_L}{2}\right)j}\right) \\ \left(Y_{L[k+1]\left[\frac{m_L+2}{2}\right]j}, X_{L(k+1)\left(\frac{m_L+2}{2}\right)j}\right)\left(Y_{L[k+2]\left[\frac{m_L+2}{2}\right]j}, X_{L(k+2)\left(\frac{m_L+2}{2}\right)j}\right) \\ \dots \left(Y_{L[m_L]\left[\frac{m_L+2}{2}\right]j}, X_{L(m_L)\left(\frac{m_L+2}{2}\right)j}\right) : j = 1,2,\dots,r\} \end{matrix} \right\}$$

In (S_tMRSS) is indicated for the *j-th* cycle and the *h-th* stratum from the bivariate sample y_h and x_h using the notation $\{y_{h[i]j}, x_{h(i)j}\}$, be a set of two variables, where i-th judgment ordering in the *i-th* set for the study variable Y_h based on the *i-th* ranking of the *i-th* set of the auxiliary variable X_h at the *j-th* cycle of the *h-th* stratum, where $i = 1,2,\dots m_h$, $j = 1,2,\dots r$ and $h = 1,2,\dots,L$. So, under the (S_tMRSS) scheme, for the main variable Y and the auxiliary variables X, the finite population mean was estimated using (S_tMRSS) and has demonstrated that it is unbiased to the population mean and has lower than simple variance for a random sample (*SRS*) as shown below. The estimator of the mean population is known according to the following relationship in stratified median ranked sets sampling and the odd case:

$$\left.\begin{aligned} \bar{y}_{h[MRSS]o} &= \frac{1}{n_h}\sum_{j=1}^{r}\sum_{i=1}^{m_h} y_{h[i]\left[\frac{m_h+1}{2}\right]j} \\ \bar{x}_{h(MRSS)o} &= \frac{1}{n_h}\sum_{j=1}^{r}\sum_{i=1}^{m_h} x_{h(i)\left(\frac{m_h+1}{2}\right)j} \end{aligned}\right\} \tag{1}$$

Where:

$\bar{y}_{h[MRSS]o}$ The median ranked set sample mean for the main variable in stratum (h).

$\bar{x}_{h[MRSS]o}$ The median ranked set sample mean for the auxiliary variable in stratum (h)

$$\left.\begin{aligned} \bar{y}_{[S_tMRSS]_o} &= \sum_{h=1}^{L} W_h \bar{y}_{h(MRSS)o} \\ \bar{x}_{(S_tMRSS)_o} &= \sum_{h=1}^{L} W_h\, \bar{x}_{h(MRSS)o} \end{aligned}\right\} \tag{2}$$

Where:

$\bar{y}_{[S_tMRSS]_o}$ The median ranked set sample mean of the main variable after it has been pooled for all strata.

$\bar{x}_{(S_tMRSS)_o}$ The median ranked set sample mean of the auxiliary variable after it has been pooled for all strata.

W_h Weight of the stratum (h) relative to the population

$$\left.\begin{aligned} S^2_{y_{h[MRSS]o}} &= \frac{1}{n_h}\left[S^2_{y_h} - \frac{1}{m_h}\sum_{i=1}^{m_h}\left(\bar{y}_{h[i]\left[\frac{m_h+1}{2}\right]} - \bar{Y}_h\right)^2\right] \\ S^2_{x_{h(MRSS)o}} &= \frac{1}{n_h}\left[S^2_{x_h} - \frac{1}{m_h}\sum_{i=1}^{m_h}\left(\bar{x}_{h(i)\left(\frac{m_h+1}{2}\right)} - \bar{X}_h\right)^2\right] \\ S_{xy_{h(MRSS)o}} &= \frac{1}{n_h}\left[S_{xy_h} - \frac{1}{m_h}\sum_{i=1}^{m_h}(\bar{y}_{h[i]\left[\frac{m_h+1}{2}\right]} - \bar{Y}_h)(\,\bar{x}_{h(i)\left(\frac{m_h+1}{2}\right)} - \bar{X}_h)\right] \end{aligned}\right\} \tag{3}$$

Where:

$S^2_{y_{h[MRSS]o}}$ The median ranked set sample variance for the main variable of the stratum (h) $S^2_{x_{h[MRSS]o}}$ The median ranked set sample variance for the auxiliary variable of the stratum (h)

$S_{xy_{h(MRSS)o}}$ The median ranked set sample Covariance of main variable and auxiliary of the stratum (h).

$$\left.\begin{aligned} S^2_{y_{[S_tMRSS]o}} &= \sum_{h=1}^{L} W_h^2 \lambda_h S^2_{y_{h[MRSS]o}} \\ S^2_{x_{(S_tMRSS)o}} &= \sum_{h=1}^{L} W_h^2 \lambda_h S^2_{x_{h(MRSS)o}} \\ S_{xy_{(S_tMRSS)o}} &= \sum_{h=1}^{L} W_h^2 \lambda_h S_{xy_{h(MRSS)o}} \end{aligned}\right\} \tag{4}$$

$S^2_{y_{[S_tMRSS]o}}$ The median ranked set sample variance for the main variable after it has been pooled for all strata.

$S^2_{x_{[S_tMRSS]o}}$ The median ranked set sample variance for the auxiliary variable after it has been pooled for all strata.

$S_{xy_{(S_tMRSS)o}}$ The median ranked set sample Covariance of main variable and auxiliary after it has been pooled for all strata.

$\lambda_h = \frac{1-\frac{n_h}{N_h}}{n_h}$ It is the correction factor for the limited stratified population divided by the sample size for stratum h.

Now the estimators are defined as follows for the even case:

$$\left.\begin{aligned} \bar{y}_{h[MRSS]e} &= \frac{1}{n_h}\sum_{j=1}^{r}\left[\sum_{i=1}^{k_h} y_{h[i]\left[\frac{m_h}{2}\right]j} + \sum_{i=k_h+1}^{m_h} y_{h[i]\left[\frac{m_h+2}{2}\right]j}\right] \\ \bar{x}_{h(MRSS)e} &= \frac{1}{n_h}\sum_{j=1}^{r}\left[\sum_{i=1}^{k_h} x_{h(i)\left(\frac{m_h}{2}\right)j} + \sum_{i=k_h+1}^{m_h} x_{h(i)\left(\frac{m_h+2}{2}\right)j}\right] \end{aligned}\right\} \tag{5}$$

$$\left.\begin{aligned} \bar{y}_{[S_tMRSS]e} &= \sum_{h=1}^{L} W_h \bar{y}_{h[MRSS]e} \\ \bar{x}_{(S_tMRSS)e} &= \sum_{h=1}^{L} W_h \bar{x}_{h(MRSS)e} \end{aligned}\right\} \tag{6}$$

$$\left.\begin{aligned} S^2_{y_{h[MRSS]e}} &= \frac{1}{n_h}\left\{S^2_{y_h} - \frac{1}{2m_h}\left[\sum_{i=1}^{k_h}\left(\bar{y}_{h[i]\left[\frac{m_h}{2}\right]} - \bar{Y}_h\right)^2 + \sum_{i=k_h+1}^{m_h}\left(\bar{y}_{h[i]\left[\frac{m_h+2}{2}\right]} - \bar{Y}_h\right)^2\right]\right\} \\ S^2_{x_{h(MRSS)e}} &= \frac{1}{n_h}\left\{S^2_{x_h} - \frac{1}{2m_h}\left[\sum_{i=1}^{k_h}\left(\bar{x}_{h(i)\left(\frac{m_h}{2}\right)} - \bar{X}_h\right)^2 + \sum_{i=k_h+1}^{m_h}\left(\bar{x}_{h(i)\left(\frac{m_h+2}{2}\right)} - \bar{X}_h\right)^2\right]\right\} \\ S_{xy_{h(MRSS)e}} &= \frac{1}{n_h}\left\{S_{xy_h} - \frac{1}{2m_h}\left[\begin{aligned}&\sum_{i=1}^{k_h}\left(\bar{y}_{h[i]\left[\frac{m_h}{2}\right]} - \bar{Y}_h\right)\left(\bar{x}_{h(i)\left(\frac{m_h}{2}\right)} - \bar{X}_h\right) + \\ &\sum_{i=k_h+1}^{m_h}\left(\bar{y}_{h[i]\left[\frac{m_h+2}{2}\right]} - \bar{Y}_h\right)\left(\bar{x}_{h(i)\left(\frac{m_h+2}{2}\right)} - \bar{X}_h\right)\end{aligned}\right]\right\} \end{aligned}\right\} \tag{7}$$

$$\left.\begin{aligned} S^2_{y_{[S_tMRSS]e}} &= \sum_{h=1}^{L} W_h^2 \lambda_h S^2_{y_{h[MRSS]e}} \\ S^2_{x_{(S_tMRSS)e}} &= \sum_{h=1}^{L} W_h^2 \lambda_h S^2_{x_{h(MRSS)e}} \\ S_{xy_{(S_tMRSS)e}} &= \sum_{h=1}^{L} W_h^2 \lambda_h S^2_{xy_{h(MRSS)e}} \end{aligned}\right\} \tag{8}$$

3. A Proposed Method for Estimating the Population Mean using the Linear Regression Model in A Stratified Ordered Mean Sampling Framework.

3.1. Regression Estimator using S_tMRSS in Separate Case

In this estimator, a median ranked set sampling regression estimator is used at the stratum level, meaning that for each stratum separately, a vector of the combined regression coefficients $b_{h(MRSS)}$ is estimated. The regression estimates for each stratum are then combined by calculating a weighted average, using the relative sizes of the strata as weights.

This type of estimation is used when there is sufficient evidence to assume that the true regression coefficients (B_h) in all strata are heterogeneous among themselves.

The regression model for each stratum of population can be represented (when ranking is on a variable X_i) as follows:

$$\bar{Y}_{h(MRSS)} = \bar{y}_{h[MRSS]} + b_{h(MRSS)}(\bar{X}_h - \bar{x}_{h(MRSS)}) \tag{9}$$

Then the population mean of the separate stratums regression after pooling is estimated with the following formula:

$$\left.\begin{aligned} \bar{Y}_{S_tMRSS(s)} &= \sum_{h=1}^{L} W_h \bar{Y}_{h(MRSS)} \\ \bar{Y}_{S_tMRSS(s)} &= \sum_{h=1}^{L} W_h\left[\bar{y}_{h[MRSS]} + b_{h(MRSS)}(\bar{X}_h - \bar{x}_{h(MRSS)})\right] \end{aligned}\right\} \tag{10}$$

We can see from formula (10) that $\bar{Y}_{S_tMRSS(s)}$ is also an unbiased estimate of the population mean $\bar{Y}$.

The variance of the estimator $\bar{Y}_{S_tMRSS(s)}$ is as follows:

$$V(\hat{\bar{Y}}_{S_tMRSS(s)}) = \sum_{h=1}^{L} \frac{W_h^2(1-f_h)}{n_h} [S_{y_h}^2 + b_h^2 S_{x_h}^2 - 2b_h S_{xy_h} - \frac{1}{m_h} \sum_{i=1}^{m_h} \left(M_{yh[i]} - M_{xh(i)}\right)^2] \tag{11}$$

The value of b that makes the variance as small as possible is:

$$b_{h(MRSS)} = \frac{S_{xy_{h(MRSS)}}}{S_{x_{h(MRSS)}}^2} \tag{12}$$

The first order of approximation to the MSE ($\bar{Y}_{S_tMRSS(s)}$) is given by

$$\mathrm{MSE}(\hat{\bar{Y}}_{S_tMRSS(s)}) = \sum_{h=1}^{L} W_h^2 \lambda_h [S_{y_h}^2 + b_h^2 S_{x_h}^2 - 2b_h S_{xy_h} - \frac{1}{m_h} \sum_{i=1}^{m_h} \left(M_{yh[i]} - M_{xh(i)}\right)^2] \tag{13}$$

Where:

$b_h = \frac{S_{xy_{h(MRSS)}}}{S_{x_{h(MRSS)}}^2}$ is obtained by the classic covariance matrix.

As is known, the median sampling contains two methods for selecting the sample from the subgroups m_h, depending on the size of the subgroup m_h, whether it is even or odd.

The odd case $M_{yh[i]o} = (\bar{y}_{h[i]\left[\frac{m+1}{2}\right]} - \bar{Y}_h)$,

$M_{xh(i)o} = (\bar{x}_{h(i)\left(\frac{m+1}{2}\right)} - \bar{X}_h)$

The even case $M_{yh[i]e} = \sum_{i=1}^{j_h} \left(\bar{y}_{h[i]\left[\frac{m}{2}\right]} - \bar{Y}_h\right)^2 + \sum_{i=j_h+1}^{m_h} (\bar{y}_{h[i]\left[\frac{m+2}{2}\right]} - \bar{Y}_h)^2$,

$M_{xh(i)e} = \sum_{i=1}^{j_h} \left(\bar{x}_{h(i)\left(\frac{m}{2}\right)} - \bar{X}_h\right)^2 + \sum_{i=j_h+1}^{m_h} (\bar{x}_{h(i)\left(\frac{m+2}{2}\right)} - \bar{X}_h)^2$

3.2. Regression Estimator using S_tMRSS in Combined Case

A combined regression estimator is a statistical method used in stratified random sampling that allows more accurate estimates of population parameters by integrating information from different regression models. (William G. Cochran, 1977)& (Ibrahim, O. S., & Mohammed, M. J. 2022) . The standard regression estimator is modified to create the combined regression estimator, where each stratum's regression coefficients are computed independently. The population coefficient is then computed overall by combining the coefficients estimated for each stratum. Better population parameter estimation is possible with this method, particularly when there are notable variations in variables of interest between strata. In conclusion, by taking into account the variance between strata, the combined regression estimator is a helpful tool in stratified random sampling that can increase the accuracy of estimating population parameters. Furthermore, it is advantageous when there are strong correlations between the variables of interest and the stratification variables, as ignoring these connections could result in skewed results.

The formula for estimating the average combined regression using median stratified ranked set sample is as follows:

$$\hat{\bar{Y}}_{S_tMRSS(c)} = \bar{y}_{[S_tMRSS]} + b_{c(S_tMRSS)}\left(\bar{X} - \bar{x}_{(S_tMRSS)}\right) \tag{14}$$

We can see from formula (14) that $\bar{Y}_{S_tMRSS(c)}$ is also an unbiased estimate of the population mean $\bar{Y}$.

Since $\bar{Y}_{S_tMRSS(c)}$ is an estimator from a stratified median ranked set sample of the variable $\bar{y}_{h[MRSS]} + b_c(\bar{X}_h - \bar{x}_{h(MRSS)})$ the variance of the estimator is as follows:

$$V(\hat{\bar{Y}}_{S_tMRSS(c)}) = \sum_{h=1}^{L} \frac{W_h^2}{n_h^2}(1-f_h)[S_{y_h}^2 + b_c^2 S_{x_h}^2 - 2b_c S_{xy_h} - \frac{1}{m_h}\sum_{i=1}^{m_h}\left(M_{yh[i]} - M_{xh(i)}\right)^2] \qquad (15)$$

The value of b that makes the variance as small as possible is:

$$b_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h s_{xy_{h(MRSS)}}}{\sum_{h=1}^{L} W_h^2 \lambda_h S_{x_{h(MRSS)}}^2} \qquad (16)$$

The first order of approximation to the MSE ($\bar{Y}_{S_tMRSS(c)}$) is given by

$$MSE(\hat{\bar{Y}}_{S_tMRSS(c)}) = \sum_{h=1}^{L} \frac{W_h^2}{n_h^2}(1-f_h)[S_{y_h}^2 + b_c^2 S_{x_h}^2 - 2b_c S_{xy_h} - \frac{1}{m_h}\sum_{i=1}^{m_h}\left(M_{yh[i]} - M_{xh(i)}\right)^2] \qquad (17)$$

Where:

$b_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h s_{xy_{h(MRSS)}}}{\sum_{h=1}^{L} W_h^2 \lambda_h S_{x_{h(MRSS)}}^2}$ is obtained by the classic covariance matrix.

Here also there are two cases for the size of the subset (m_h), the odd case:

$M_{yh[i]o} = (\bar{y}_{h[i]\left[\frac{m+1}{2}\right]} - \bar{Y}_h)$, $M_{xh(i)o} = (\bar{x}_{h(i)\left(\frac{m+1}{2}\right)} - \bar{X}_h)$.

The even case:

$M_{yh[i]e} = \sum_{i=1}^{k_h}\left(\bar{y}_{h[i]\left[\frac{m}{2}\right]} - \bar{Y}_h\right)^2 + \sum_{i=k_h+1}^{m_h}(\bar{y}_{h[i]\left[\frac{m+2}{2}\right]} - \bar{Y}_h)^2$,

$M_{xh(i)e} = \sum_{i=1}^{k_h}\left(\bar{x}_{h(i)\left(\frac{m}{2}\right)} - \bar{X}_h\right)^2 + \sum_{i=k_h+1}^{m_h}(\bar{x}_{h(i)\left(\frac{m+1}{2}\right)} - \bar{X}_h)^2$.

3.3. Outliers

In statistics, the outliers problem is among the most significant and traditional issues. One of the primary issues with regression estimation techniques may be the occurrence of outliers in the data set, which alter the outcome of the statistical analysis of the data. that the outliers are the observations that are very different from the other observations that are supposed to be the result of a different mechanism. (Dick, 2022) and (TALAL, A. R., & SALIM, O. 2024)

3.4. Methods For Estimating the Robust Covariance and Covariance Matrix

Researchers have developed methods to estimate robust variance and covariance matrices in the presence of outliers. Two such methods are Minimum Volume Ellipsoid (MVE) and Minimum Covariance Determinant (MCD). These methods are particularly useful in stratified sampling, where data are collected from distinct population subgroups with different characteristics. To learn more about the two previous methods and the steps to apply them, see, for example: (Hawkins, D.M. 1980), (Rousseeuw, P.J. 1985), (Hubert, M. and Engelen, S. 2004), (Rousseeuw, P.J. and Hubert, M., 2018), (Zaman, T., & Bulut, H. 2020) and (Mahdi, M.H. and Hussein, S.M.2023), (Ibrahim, O. S., & Mohammed, M. J. 2022) .

3.5. Some Robust Methods for Estimating the Linear Regression Parameter

Several statistical methods or approaches have been proposed to estimate the linear regression parameter, so that the estimation methods have become very numerous, and the goal of the estimation process is to obtain the best and most efficient results in an attempt to represent the community well.

Conventional estimation methods fail to deal with data that include the presence of outliers, as they have an undesirable effect on the results. Hence, the importance of using alternative methods to conventional methods, called robust estimation methods, emerged, as they are insensitive to outliers if they exist in the data. (Robert G. Staudte, and Simon J. Sheather. 1990), (Whitaker, R.T. 2001). Some of these robust methods will be used in the simulation study chapter. The most important of these methods are: Method M Which was suggested by (Huber, P. J. 1973). Method MM This method has been suggested by (Yohai, V. J. 1987). Least Trimmed Squares Regression Method (LTS) which was discussed by (P. J. Rousseeuw and A. M. Leroy 1987), (A. Marazzi 1993) and (P. Rousseeuw and M. Hubert 1997), And Least Median of Squares Regression Method (LMS) proposed by (Rousseeuw, P. J. 1984) ,(SuraA. Mohammad & OsamaB. Hannon 2020) and (Mahammad Mahmoud Bazid & TahaHussein Ali 2025).

Statistical Analysis

4. Simulation Study and Discussion of Results

4.1 Simulation Preparation

In this aspect, we used the R language program (version (4.2.1)) to conduct the statistical analysis to achieve the study objectives by comparing the proposed estimator with the robust estimators and testing the best of them based on the standard of the mean squares of error.

The simulation experiments were described as follows:

1. One explanatory variable for the population X_i was determined according to the normal distribution $X_i \sim N(3{,}2)$ with a size of 1200 observations and adopting different pollution ratios (10%, 15%, 20%, 25%), where the simple regression model was used: $Y = 2 + 4X + \varepsilon$
2. The data was classified into five strata based on the x values. The strata are as follows: First stratum when $X_i < 0.5$, Second stratum when $0.5 \leq X_i < 2$, the third stratum when $2 \leq X_i < 3.8$, the fourth stratum when $3.8 \leq X_i < 5.5$, the fifth stratum when $X_i \geq 5.5$, and the strata sizes were as follows ($N_1 = 115,\ N_2 = 235,\ N_3 = 455, N_4 = 266, N_5 = 129$), With different strata sizes, the proportional distribution can be used to maintain the stability of the sample proportion in each part of the population. If the stratum (h) contains (N_h) units, then the sample sizes are calculated according to the proportional distribution formula: $n_h = \frac{n\,N_h}{N}$
3. Five different sizes of the hypothetical samples were selected (100, 250, 400, 600) from the total population (N=1200).
4. Generate random errors according to the normal distribution of non-outliers' data $\varepsilon_i \sim N(0.1)$, negative outliers' data $\varepsilon_i \sim N(-20.4)$, and positive outliers' data $\varepsilon_i \sim N(20.4)$.

The (S_tMRSS) estimator in Tables (1) and (2) achieved more efficiency for the combined regression estimator and separate regression estimator to estimate the population mean at all assumed sample sizes (n = 100, 250, 400, 600) and for all proportions outliers, because it has a lower MSE value than the traditional covariance matrix estimator, which is sensitive to the presence of outliers, which strongly affects the accuracy of the estimated covariance matrix, and also because it has a lower MSE value than the robust covariance matrices estimators (MCD, MVE) when estimating the regression coefficient β_c by robust methods (M, MM, LMS, LTS).

To reach the best effective estimate, compare the proposed method with the robust methods, and demonstrate its efficiency, the mean square error values were calculated for all the previously mentioned estimates according to Tables (1) and (2).

$$MSE = \frac{1}{R}\sum_{i=1}^{R}(\hat{\hat{\bar{Y}}}_i - \bar{Y})^2 \qquad i = 1,2,\dots,R \tag{18}$$

R: The number of repetitions for each experiment represents (500) samples.

$\hat{\hat{\bar{Y}}}_i$: The estimated value of the population means of the dependent variable.

$\bar{Y}$: The true value of the population means of the dependent variable.

4.2. Simulation Results

There are two aspects of our findings:

-Population mean for separate regression estimator.

Table (1):Values of (MSE) to estimate the population mean by separate regression and percentages of extreme values (10%, 15%, 20%, and 25%) .

n	methods	The outliers rate is 10%			The outliers rate is 15%			The outliers rate is 20%			The outliers rate is 25%		
		Classic	MCD	MVE	Classic	MCD	MVE	Classic	MCD	MVE	Classic	MCD	MVE
100	OLS	0.41174	0.03961	0.03559	0.58441	0.03555	0.04423	0.76798	0.13279	0.15582	0.93763	0.32521	0.39236
	M	0.37948	0.03324	0.03011	0.56810	0.03152	0.03014	0.74276	0.07680	0.07223	0.90550	0.13542	0.14980
	MM	0.36899	0.02908	0.02565	0.56257	0.02365	0.02291	0.74683	0.06430	0.06099	0.88632	0.11432	0.12944
	LTS	0.37067	0.03484	0.03059	0.56225	0.02905	0.02588	0.73438	0.07288	0.06640	0.89102	0.12026	0.13700
	LMS	0.37344	0.03781	0.03436	0.55946	0.02905	0.02741	0.72002	0.06017	0.05803	0.87673	0.12578	0.13969
	$S_tMRSS_{(o)}$	0.00815			0.01301			0.02745			0.05630		
	$S_tMRSS_{(e)}$	0.01318			0.01907			0.04100			0.05748		
250	OLS	0.13908	0.01709	0.01112	0.20729	0.01770	0.01185	0.29373	0.01057	0.01138	0.34943	0.00891	0.01156
	M	0.13526	0.01676	0.01071	0.20363	0.01754	0.01167	0.28326	0.00999	0.01144	0.34389	0.00803	0.01187
	MM	0.13508	0.01650	0.01050	0.20331	0.01704	0.01170	0.28158	0.01042	0.01092	0.34197	0.00668	0.01032
	LTS	0.13541	0.01686	0.01093	0.20318	0.01775	0.01166	0.28236	0.01015	0.01232	0.34180	0.00719	0.01085
	LMS	0.13520	0.01693	0.01108	0.20352	0.01711	0.01171	0.28182	0.01042	0.01211	0.34244	0.00730	0.01044
	$S_tMRSS_{(o)}$	0.00358			0.00576			0.01100			0.01656		
	$S_tMRSS_{(e)}$	0.00530			0.00755			0.01430			0.02677		
400	OLS	0.06881	0.01348	0.00935	0.10986	0.01594	0.01119	0.14531	0.00698	0.00830	0.17717	0.00328	0.00328
	M	0.06702	0.01355	0.00959	0.10884	0.01577	0.01018	0.14337	0.00690	0.00820	0.17539	0.00323	0.00321
	MM	0.06696	0.01345	0.00946	0.10891	0.01548	0.01017	0.14322	0.00674	0.00802	0.17530	0.00312	0.00308
	LTS	0.06702	0.01357	0.00957	0.10866	0.01581	0.01080	0.14362	0.00698	0.00852	0.17502	0.00319	0.00314
	LMS	0.06720	0.01332	0.00934	0.10887	0.01555	0.01064	0.14329	0.00736	0.00879	0.17509	0.00323	0.00322
	$S_tMRSS_{(o)}$	0.00215			0.00345			0.00738			0.01304		
	$S_tMRSS_{(e)}$	0.00339			0.00473			0.00823			0.01575		
600	OLS	0.03586	0.01213	0.00771	0.04873	0.01415	0.00745	0.07001	0.00320	0.00354	0.09581	0.00155	0.00158
	M	0.03564	0.01217	0.00797	0.04793	0.01399	0.00742	0.06945	0.00317	0.00370	0.09466	0.00155	0.00159
	MM	0.03566	0.01211	0.00794	0.04787	0.01382	0.00748	0.06945	0.00315	0.00355	0.09453	0.00151	0.00154
	LTS	0.03560	0.01207	0.00780	0.04778	0.01412	0.00756	0.06967	0.00326	0.00368	0.09446	0.00151	0.00154
	LMS	0.03571	0.01189	0.00782	0.04814	0.01406	0.00764	0.06950	0.00335	0.00320	0.09447	0.00156	0.00160
	$S_tMRSS_{(o)}$	0.00165			0.00234			0.00479			0.00920		
	$S_tMRSS_{(e)}$	0.00230			0.00362			0.00537			0.01107		

Population mean for combined regression estimator.

Table (2):Values of (MSE) to estimate the population mean by combined regression and percentages of extreme values (10%, 15%, 20%, and 25%) .

N	methods	The outliers rate is 10%			The outliers rate is 15%			The outliers rate is 20%			The outliers rate is 25%		
		Classic	MCD	MVE	Classic	MCD	MVE	Classic	MCD	MVE	Classic	MCD	MVE
100	OLS	0.38755	0.01966	0.02002	0.57328	0.06364	0.05355	0.72829	0.08917	0.09003	0.92534	0.17157	0.18119
	M	0.38387	0.01851	0.01891	0.65268	0.06306	0.05332	0.72158	0.08531	0.08666	0.92041	0.16777	0.17662
	MM	0.38378	0.01846	0.01887	0.56274	0.06326	0.05350	0.72215	0.08583	0.08713	0.92068	0.16878	0.17785
	LTS	0.38421	0.01865	0.01901	0.56262	0.06315	0.05338	0.72227	0.08598	0.08723	0.92062	0.16910	0.17818
	LMS	0.38384	0.01865	0.01901	0.56237	0.06317	0.05342	0.72217	0.08581	0.08708	0.92083	0.16920	0.17830
	$S_tMRSS_{(o)}$	0.00859			0.01142			0.02343			0.04184		
	$S_tMRSS_{(e)}$	0.01387			0.01741			0.03602			0.04735		
250	OLS	0.14155	0.00657	0.00654	0.19786	0.03627	0.03251	0.26459	0.04684	0.04710	0.33567	0.07568	0.07950
	M	0.14016	0.00645	0.00644	0.19648	0.03469	0.03122	0.26355	0.04535	0.04559	0.33432	0.07354	0.07794
	MM	0.14017	0.00645	0.00644	0.19653	0.03469	0.03121	0.26364	0.04570	0.04593	0.33441	0.07432	0.07859
	LTS	0.14011	0.00648	0.00642	0.19651	0.03460	0.03111	0.26364	0.04546	0.04572	0.33415	0.07397	0.07842
	LMS	0.14014	0.00645	0.00641	0.19649	0.03477	0.03131	0.26366	0.04548	0.04572	0.33439	0.07404	0.07842
	$S_tMRSS_{(o)}$	0.00341			0.00534			0.00939			0.01605		
	$S_tMRSS_{(e)}$	0.00562			0.00777			0.01343			0.02125		
400	OLS	0.06629	0.00332	0.00356	0.10620	0.03464	0.02688	0.14041	0.03525	0.04062	0.17873	0.06903	0.06205
	M	0.06585	0.00324	0.00354	0.10574	0.03301	0.02586	0.14063	0.03353	0.03876	0.17879	0.06744	0.06017
	MM	0.06585	0.00324	0.00353	0.10573	0.03298	0.02584	0.14065	0.03385	0.03912	0.17879	0.06800	0.06079
	LTS	0.06585	0.00324	0.00359	0.10573	0.03272	0.02565	0.14060	0.03357	0.03882	0.17881	0.06787	0.06050
	LMS	0.06588	0.00322	0.00354	0.10569	0.03288	0.02579	0.14062	0.03379	0.03906	0.17880	0.06784	0.06055
	$S_tMRSS_{(o)}$	0.00236			0.00336			0.00626			0.01164		
	$S_tMRSS_{(e)}$	0.00383			0.00568			0.00876			0.01465		
600	OLS	0.03876	0.00182	0.00181	0.07151	0.02268	0.02136	0.07151	0.02768	0.02625	0.08988	0.06207	0.03166
	M	0.03842	0.00173	0.00174	0.07129	0.02125	0.02048	0.07129	0.02585	0.02472	0.08935	0.06029	0.03076
	MM	0.03841	0.00173	0.00174	0.07129	0.02127	0.02049	0.07129	0.02609	0.02494	0.08934	0.06095	0.03106
	LTS	0.03854	0.00179	0.00176	0.07131	0.02123	0.02021	0.07131	0.02583	0.02469	0.08934	0.06088	0.03103
	LMS	0.03853	0.00174	0.00174	0.07129	0.02118	0.02054	0.07129	0.02609	0.02495	0.08935	0.06093	0.03101
	$S_tMRSS_{(o)}$	0.00165			0.00217			0.00424			0.00754		
	$S_tMRSS_{(e)}$	0.00250			0.00409			0.00608			0.00941		

5. Discussion

In Table (1), a comparison was made according to the mean square error (MSE) criterion to estimate the population mean using the separate regression method between the proposed estimation method (S_tMRSS) to estimate the regression parameter ($b_{h(MRSS)}$) and the arithmetic means ($\bar{y}_{h[MRSS]}, \bar{x}_{h[MRSS]}$) and the robust estimation methods (Huber M, Huber MM, LTS, LMS) and traditional (OLS) to estimate the regression parameter (b_c) with the use of robust arithmetic means of the estimators (MCD, MVE) and the traditional arithmetic means. From these comparisons, we noticed that the proposed estimation method (S_tMRSS) gives more efficient results in most cases than the other estimators. as these values were estimated when applying Formula (18)

In Table (2), a comparison was made according to the mean square error (MSE) criterion to estimate the population mean using the combined regression method between the proposed estimation method (S_tMRSS) to estimate the regression parameter ($b_{c(MRSS)}$) and the arithmetic means ($\bar{y}_{[S_tMRSS]}, \bar{x}_{[S_tMRSS]}$) and the robust estimation methods (Huber M, Huber MM, LTS, LMS) and traditional (OLS) to estimate the regression parameter (b_c) with the use of robust arithmetic means of the estimators (MCD, MVE) and the traditional arithmetic means. From these comparisons, we noticed that the proposed estimator ($S_t MRSS$) gives more efficient results in all cases than the other estimators. as these values were estimated when applying Formula (18)

6.Conclusion

From the results in Table (1), the proposed estimator ($S_t MRSS$) achieved high efficiency when used to estimate the regression parameter ($b_{h(MRSS)}$) to estimate the population mean by separate stratified regression ($\bar{Y}_{S_t MRSS(s)}$) and by using the arithmetic means for the stratified median ranked set sampling ($\bar{y}_{h[MRSS]}$, $\bar{x}_{h[MRSS]}$), at the assumed sample sizes ($n = 100, 250, 400, 600$) and the percentages of outliers (10%, 15%, 20%) compared to the studied estimation methods, with the exception of the high percentage of 25%. The robust estimators slightly outperformed, as the (MSE) values of Huber MM method outperforming the (MSE) values of the proposed estimator.

As for the results of Table (2), the proposed estimator ($S_t MRSS$) achieved high efficiency when used to estimate the regression parameter ($b_{c(S_t MRSS)}$) to estimate the population mean using the stratified joint regression method ($\hat{\bar{Y}}_{S_t MRSS(c)}$) and using the arithmetic means of the stratified median ranked set sampling ($\bar{y}_{[S_t MRSS]}$, $\bar{x}_{[S_t MRSS]}$), at all assumed sample sizes ($n = 100, 250, 400, 600$) and the percentages of the abnormal section (10%, 15%, 20%, 25%) compared to the studied estimation methods.

We also conclude that the proposed estimator ($S_t MRSS$) is more efficient when using odd subsets (m_h) compared to the case of using even subsets for the same estimator because it obtained the smallest values of mean square errors (MSE).

Acknowledgment

The authors are very grateful to the University of Mosul, College of Computer Science and Mathematics for their provided facilities, which helped improve this work's quality.

Conflict of interest

The author has no conflict of interest.

References

1. Ahmed, R. A., & Hussein, S. M. (2022). Generalized modified ratio-cum-product kind exponentially estimator of the populations mean in stratified ranked set sample. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 1137-1149, Doi.org/10.22075/ijnaa.2022.5657.
2. Çetin, A. E., & Koyuncu, N. (2020). Estimation of population mean under different stratified ranked set sampling designs with simulation study application to BMI data. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 69(1), 560-575, Doi 10.31801/cfsuasmas.649909
3. Cochran, W. G. (1977), Sampling Techniques, Third Edition, John Wiley & Sons
4. Dell T.R. and Clutter J.L. (1972). Ranked set sampling theory with order statistics background, Biometrics, 545–555, Doi.org/10.2307/2556166
5. Dick J. Brus. (2022). " Spatial Sampling with R “. CRC Press is an imprint of Taylor & Francis Group, LLC, Boca Raton London New-York, Doi: 10.18637/jss.v105.i03
6. Hawkins, D. (1980). Identification of outliers. (Vol. 11). London: Chapman and Hall, https://link.springer.com/book/10.1007/978-94-015-3994-4
7. Hubert, M., & Engelen, S. (2004). Fast cross validation for high breakdown resampling algorithms.
8. Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The annals of statistics*, 799-821.Dio jstor.org/stable/2958283
9. Ibrahim, K., Syam, M., & Al-Omari, A. I. (2010). Estimating the population mean using stratified median ranked set sampling. Applied Mathematical Sciences, 4(47), 2341-2354, https://www.m-hikari.com/ams/ams-2010/
10. Ibrahim, O. S., & Mohammed, M. J. (2022). Use The Robust RFCH Method With A Polychoric Correlation Matrix In Structural Equation Modeling When You Are Ordinal Data 19(2),24-35 , http://creativecommons.org/licenses/by/4.0/

11. Khan, L., Shabbir, J., & Gupta, S. (2016). Unbiased ratio estimators of the mean in stratified ranked set sampling. Hacettepe Journal of Mathematics and Statistics, 46(6), 1151-1158, Doi 10.15672/HJMS.201610814857
12. Lynne Stokes, S. (1977). Ranked set sampling with concomitant variables. Communications in Statistics-Theory and Methods, 6(12), 1207-1211, Doi.org/10.1080/03610927708827563
13. Linder, D. F., Samawi, H., Yu, L., Chatterjee, A., Huang, Y., & Vogel, R. (2015). On stratified bivariate ranked set sampling for regression estimators. Journal of Applied Statistics, 42(12), 2571-2583, Doi.org/10.1080/02664763.2015.1043868
14. Mahdi, M. H., & Hussein, S. M. (2023). Estimating the Population Mean in Stratified Random Sampling Using Combined Regression with the Presence of Outliers. *Journal of Economics and Administrative Sciences*, *29*(136), 70-80, http://jeasiq.uobaghdad.edu.iq
15. Mandowara, V. L., & Mehta, N. (2014). Modified ratio estimators using stratified ranked set sampling. Hacettepe Journal of Mathematics and Statistics, 43(3), 461-471, https://dergipark.org.tr/en/download/article-file/649202
16. Mahammad Mahmoud Bazid & TahaHussein Ali (2025). Estimating Outliers Using the Iterative Method in Partial Least Squares Regression Analysis for Linear Models,22(1), 88-100 , Doi 10.33899/iqjoss.2025.187757
17. Marazzi, A. (1993). *Algorithms, routines, and S-functions for robust statistics*. CRC Press.
18. McIntyre, G.A. (1952). A method of unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research, 3, 385- 390, Doi.org/10.1071/AR9520385
19. Muttlak H. A. (1997). Median ranked set sampling. J. Appl. Statist. Sci, 6:245–255, Doi 10.33899/rtyew.22575.124789
20. Nadia, H., & Mohammad, A. A. (2013). Model of robust regression with parametric and nonparametric methods. *Mathematical Theory and Modeling*, *3*(5), 27-39, https://d1wqtxts1xzle7.cloudfront.net/34634542
21. Rousseeuw, P.J., Leroy, A. M. (1987) " Robust Regression and Outliers Detection " John Wiley, New York.
22. Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. Mathematical Statistics and Applications, https://wis.kuleuven.be/stat/robust/papers/publications-1985/rousseeuw-multivariateestimationhighbreakdown-1985.pdf
23. Rousseeuw, P. J., & Hubert, M. (2018). Anomaly detection by robust statistics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(2), e1236, Doi: 10.1002/widm.1236
24. Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, *79*(388), 871-880, https://wis.kuleuven.be/stat/robust/papers/publications-1984/rousseeuw-leastmediansquaresregression-jasa-1984.pdf
25. Samawi, H. M., & Tawalbeh, E. M. (2002). Double median ranked set sample: comparing to other double ranked samples for mean and ratio estimators. Journal of Modern Applied Statistical Methods, 1, 428-442, http://digitalcommons.wayne.edu/jmasm/vol1/iss2/52
26. Samawi H. M. and Siam M. I. (2003). Ratio estimation using stratified ranked set sample. METRON, 61(1), pp. 75-90 , https://d1wqtxts1xzle7.cloudfront.net/83615321/Ratio_estimation_using_stratified_ranked20220409-9064-159rn8b.pdf
27. Samawi H.M. (2010)., Stratified Ranked Set Sample, Cambridge Scholars Publishing https://ncbae.edu.pk/wp-content/uploads/2018/02/2
28. Saini, M., & Kumar, A. (2019). Ratio estimators using stratified random sampling and stratified ranked set sampling. Life Cycle Reliability and Safety Engineering, 8, 85-89, https://link.springer.com/article/10.1007/s41872-018-0046-8
29. Staudte, R. G., & Sheather, S. J. (1990). Robust estimation and testing. John Wiley & Sons. *Inc., New York*.
30. SuraA. Mohammad & OsamaB. Hannon (2020). Using Logistic Regression with Time-Stratified Method for Air Pollution Datasets Forecasting, 17(1), 19-28 , DOI: 10.33899/IQJOSS.2020.165444
31. Takahasi, K., & Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. Annals of the institute of statistical mathematics, 20(1), 1-31, https://www.ism.ac.jp/editsec/aism/
32. Thanoon Y. Thanoon , Robiah Adnan & Zarina Mohd Khalid (2023), Analysis of Two Populations Dichotomous Data in Latent Variable Models Using Bayesian Approach, 20(2), 57-78, DOI: 10.33899/IQJOSS.2023.0181176
33. Whitaker, R. T. (2001, November). Maximum Likelihood Surface Estimation. In *Proceedings of the Vision Modeling and Visualization Conference 2001* (pp. 463-464), Doi 10.1109/TPAMI.2002.1039208
34. Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, 642-656, C:/Users/HP/Downloads/1176350366.pdf
35. 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://www.tandfonline.com/loi/lsta20

تقدير متوسط المجتمع في معاينة المجموعة المرتبة الوسطية الطبقية باستخدام الانحدار المنفصل والمشترك مع وجود القيم الشاذة

ايمن محمود الحديدي و د. ريكان عبد العزيز احمد

قسم الإحصاء والمعلوماتية، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق

الخلاصة : يهدف هذا البحث الى بيان الكفاءة العالية والدقة في تقدير متوسط المجتمع المحدود من خلال تقديرات خط الانحدار الطبقي المنفصل والمشترك بالاعتماد على اسلوب معاينة المجموعة المرتبة المتوسطة لاختيار عينة أكثر تمثيلا للمجتمع. ومع وجود مشكلة عدم التجانس في البيانات واحتوائها على قيم متطرفة (قيم شاذة)، يوصى باستخدام التقسيم الطبقي للمجتمع وسحب عينات بأسلوب المعاينة المرتبة الوسطية من هذه الطبقات والتي تعرف بمعاينة المجموعة المرتبة المتوسطة الطبقية (S_tMRSS) وهي احدى اساليب معاينة المجموعة المرتبة المعدلة، حيث تمت مقارنة متوسط مربعات الخطأ **MSE** لمقدر متوسط المجتمع الذي تم الحصول عليه بهذه الطريقة مع قيمة **MSE** لمقدر متوسط المجتمع الذي تم الحصول عليه من خلال تقديرات الانحدار باستخدام مصفوفة التباين والتباين المشترك الحصينة (**Minimum Covariance Determinant (MCD), Minimum Volume Ellipsoid (MVE)**) لحساب المتوسطات واستخدام الطرق الحصينة (**Huber M, Huber MM, Least Median of Squares (LMS), Least Trimmed Squares (LTS)**) لتقدير معلمة الانحدار لتظهر النتائج ان المقدر المقترح يتفوق على المقدرات الحصينة في اغلب الحالات لحصوله على اقل القيم من متوسط مربعات الخطأ.

الكلمات المفتاحية: عينة مجموعة مرتبة متوسطة طبقية مقدر الانحدار مقدر الانحدار المشترك مقدر الانحدار المنفصل.