Advanced Mathematical Models & Applications
Vol.9, No.3, 2024, pp.415-430
https://doi.org/10.62476/amma93415

DOUBLE EXCEPT EXTREME RANKED SET SAMPLING FOR ESTIMATING POPULATION MEAN

Mahmoud Zuhier Aldrabseh[1], **Mohd Tahir Ismail**[1*], **Amer Ibrahim Al-Omari**[2]

[1]School of Mathematical Science, Universiti Sains Malaysia, Pulau Pinang, Malaysia
[2]Department of Mathematics, Faculty of Science, Al al-Bayt University, Al-Mafraq, Jordan

Abstract. The ranked set sampling (RSS) method is an efficient sampling method that is used when the judgment ranking is easy while measuring the sampling units is expensive. This method has a higher efficiency than the commonly used simple random sampling (SRS) method. If this method is implemented in two stages, it will be called double-ranked set sampling (DRSS), which has a higher efficiency than RSS. Recently, the except extreme ranked set sampling (EERSS) method is proposed as a modification of the RSS. In the current study, we modified the EERSS and suggested the double-stage EERSS (DEERSS) method for estimating the population mean. The DEERSS estimator is compared with each DRSS, EERSS, RSS, and SRS counterparts. It has been shown that the DEERSS estimator is more efficient than all the competitor estimators considered in this study. Also, it is shown that this estimator is unbiased in the case of symmetrical distributions. Three data sets are employed to test the applicability and efficiency of the DEERSS method.

Keywords: Ranked set sampling, double ranked set sampling, population mean; mean estimate, relative efficiency, probability distribution, agricultural applications.

AMS Subject Classification: 62D05.

Corresponding author: Mohd Tahir Ismail, School of Mathematical Science, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia, Tel.: +60 16-414 3464, e-mail: *m.tahir@usm.my*

Received: 21 April 2024; Revised: 3 July 2024; Accepted: 9 September 2024; Published: 4 December 2024.

1 Introduction

In a lot of survey studies, particularly in agriculture, forestry, biology, and the environment, the challenge is to find a flexible, effective sampling design that reflects the reality of the population to be studied. However, sometimes measuring these elements is difficult and expensive, although it is easy to rank them by judgment before measuring them. Whether by eye, using a concomitant variable, or any other easy ranking method, to illustrate this, for example, if we assume that we want to measure the heights of trees in a forest, measuring each tree takes time and effort, whilst it is easy to make previous judgmental rankings before including the trees in the sample, in this case, the ranked set sampling and its modifications are useful. Ranked set sampling (RSS) is first introduced by McIntyre (1952) to estimate the mean yield in the pasture. This method is found to be more efficient than simple random sampling (SRS) in estimating the population mean (McIntyre, 1952; Takahasi & Wakimoto, 1968; Dell & Clutter, 1972). The RSS method can be performed by drawing m simple random samples from the target population each of size m, ordering the elements within each sample judgmentally or by using any in-expensive

How to cite (APA): Aldrabseh, M.Z., Ismail, M.T. & Al-Omari, A.I. (2024). Double except extreme ranked set sampling for estimating population mean. *Advanced Mathematical Models & Applications*, *9*(3), 415-430 https://doi.org/10.62476/amma93415

ranking, and lastly, measuring accurately the i^{th} unit within the i^{th} set for, $i = 1, 2, 3, \ldots, m$. These previous steps can be repeated c cycles to get a larger sample size $n = mc$. Takahasi & Wakimoto (1968) provided the theoretical properties for this method where the RSS estimator of the population mean $\bar{X}_{RSS} = \frac{1}{mc}\sum_{j=1}^{c}\sum_{i=1}^{m} X_{(i:m)j}$ is an unbiased estimator, where $X_{(i:m)j}$ denotes the i^{th} ordered statistic, with the cumulative distribution function (CDF), probability density function (PDF), mean, and variance are given, respectively, as follows

$$F_{(i:m)}(x) = \sum_{j=i}^{m} \binom{m}{j} [F(x)]^{j} [1 - F(x)]^{m-j}, \tag{1}$$

$$f_{(i:m)}(x) = m \binom{m-1}{i-1} f(x) [F(x)]^{i-1} [1 - F(x)]^{m-i}, \tag{2}$$

$$\mu_{(i:m)} = E\left(X_{(i:m)}\right) = \int_{-\infty}^{\infty} x f_{(i:m)}(x) dx, \tag{3}$$

and

$$\sigma^2_{(i:m)} = Var\left(X_{(i:m)}\right) = \int_{-\infty}^{\infty} \left(x - \mu_{(i:m)}\right)^2 f_{(i:m)}(x) dx, \tag{4}$$

where, $f(x)$ and $F(x)$ are the PDF and CDF of the variable of interest X, with mean μ and variance σ^2. The relative efficiency (RE) of the RSS estimator relative to the SRS estimator $(\bar{X})$ is given by

$$RE = \frac{Var\left(\bar{X}\right)}{Var\left(\bar{X}_{RSS}\right)} = \frac{1}{1 - \frac{1}{m\sigma^2}\sum_{i=1}^{m}\left(\mu_{(i:m)} - \mu\right)^2}. \tag{5}$$

Later, Dell & Clutter (1972) discussed this estimator with consideration of the errors in judgment ranking, Stokes (1980) developed the RSS method for estimating the population variance and proved that this estimator is asymptotically unbiased. It also has higher efficiency when the number of cycles is sufficiently large. To increase the efficiency of the RSS method for fixed sample size, Al-Saleh & Al-Kadiri (2000) introduced the double RSS (DRSS) for estimating the population mean. It is proven in their study that the DRSS estimator has higher efficiency than both the SRS and RSS estimators. Additionally, Al-Saleh & Al-Omari (2002) investigated the multistage RSS (MSRSS) method by applying RSS in more than two stages to increase the efficiency at fixed m. It is also proven that the efficiency of the MSRSS estimator is close to m^2 for uniform distribution.

Several modifications of the RSS are suggested by many authors, to name a few; Haq et al. (2016) introduced paired DRSS as an alternative method to DRSS by mixing the RSS and paired RSS, median DRSS by combining the median RSS and the RSS (Al-Mawan et al., 2018), robust ranked set sampling (LRSS) (Al-Nasser, 2007), its double LRSS (Al-Omari & Haq, 2019). Also, Khan et al. (2020) considered mixing the extreme RSS in some cycles with RSS in the remaining cycles, similarly, Hanandeh et al. (2022b) by mixing ERSS and MRSS with RSS, likewise, Hanandeh et al. (2022a) suggested two new RSS sampling methods by combining each RSS and Median RSS with Mini-Max RSS, and Yaparova (2024) suggested the double design of the moving extreme RSS. For more RSS designs refer to Haq et al. (2014) and Al-Nasser & Al-Omari (2018).

In addition to the application in estimating the population mean, the RSS sampling method and its different designs are used in many aspects, such as, Zamanzade & Al-Omari (2016) developed a new neoteric RSS method for estimating the population variance in addition to the population mean. While Al-Omari & Al-Nasser (2018) utilized the multistage median RSS method for estimating the population ratio. Correspondingly, Khan et al. (2020) discussed estimating the population median using the mixture RSS method. Abdallah et al. (2022) recommended new maximum likelihood (ML) estimators for the distribution function (CDF) and the reliability using LRSS and robust extreme RSS methods. Al-Omari & Abdallah (2023) used both the moving extreme RSS and the minimax RSS methods to estimate the distribution function, and Shang & Yan (2024) suggested ML and least square estimates for reliability estimation based on RSS when strength and stress have Kumaraswamy and Weibull distributions. In contrast, Mahdizadeh & Zamanzade (2022) discussed the interval estimation under RSS. Additional applications are discussed by Al-Omari (2010, 2016).

A recent study by Aldrabseh & Ismail (2023) investigated a new except extremes RSS (EERSS) method for estimating the population mean by excluding the maximum and minimum ranks. Their study shows that the EERSS estimator of the population mean is unbiased under the symmetry assumption. Also, based on their simulation results, this method is more efficient than SRS and RSS in estimating the population mean. For this reason and based on the results of previous designs implemented in two stages, the researchers expect that the double design of EERSS will be more efficient than the EERSS design.

The remainder of the paper is organized as follows: The sampling methods considered for comparisons are explained in Section 2. The proposed sampling design is described in Section 3. The estimator of the population mean is defined in Section 4. In Section 5, the results of relative efficiency for some selected distributions are conducted. An illustration of the sampling method using real data sets is given in Section 6. Finally, the conclusion is stated in Section 7.

2 Sampling designs

This section explains the steps for implementing the sampling designs considered for comparison. It also explains the population mean estimators under these methods.

2.1 Except extremes ranked set sampling design

The EERSS design is suggested by Aldrabseh & Ismail (2023) as a modification of the RSS method by excluding the extremes from the selected ranks. It is implemented as follows:

1. Draw m random sets, each set of size $m+2$.
2. Order the units within each set judgmentally for the variable of interest.
3. Select the $(i+1)^{th}$ ranked unit from the i^{th} set, for $i = 1, 2, 3 \ldots, m$.
4. The Steps (1-3) can be repeated c times to get a sample of size $n = cm$.

Let $X_{(2:m+2)}, X_{(3:m+2)}, \cdots, X_{(m+1:m+2)}$ denote the except extreme ranked variables. Then the estimator based on this method is given by

$$\bar{X}_{EERSS} = \frac{1}{m} \sum_{i=2}^{m+1} X_{(i:m+2)}, \tag{6}$$

this estimator proved to be an unbiased estimator of the population mean for symmetrical distribution, but we can't say that it is unbiased for asymmetrical distribution. So, the mean of this estimator is given by

$$E\left(\bar{X}_{EERSS}\right) = \frac{1}{m} \sum_{i=2}^{m+1} \mu_{(i:m+2)}, \tag{7}$$

with mean square error

$$MSE\left(\bar{X}_{EERSS}\right) = \frac{1}{m^2} \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} + \left(E\left(\bar{X}_{EERSS}\right) - \mu\right)^2, \tag{8}$$

therefore the relative efficiency of the EERSS estimator $\bar{X}_{EERSS}$ relative to the SRS estimator $(\bar{X}_{SRS})$ is given by

$$RE = \frac{Var\left(\bar{X}_{SRS}\right)}{MSE\left(\bar{X}_{EERSS}\right)}. \tag{9}$$

For more information about the DRSS method, refer to Aldrabseh & Ismail (2023).

2.2 Double ranked set sampling design

The DRSS design suggested by Al-Saleh & Al-Kadiri (2000) is a modification of the RSS method by applying RSS in two stages. It is implemented as follows:

1. Randomly select m groups of samples from the target population, each group of size m^2, with m sets each containing m elements.
2. Apply the RSS method to each group. These yields m ranked set samples, each of size m.
3. Apply the RSS design again on the new matrix to get a double-ranked set sample of size m.
4. The Steps (1-3) can be repeated c cycles to get a sample of size $n = cm$.

Let X denote the random variable of interest with mean μ and variance σ^2, and $Y_{1j}, Y_{2j}, \cdots, Y_{mj}$ be the double ranked set sample of variables, and $j = 1, 2, \ldots, c$ be the cycle number. The variable $Y_{ij} = i^{th} - Min\left\{X_{j(1:m)}, X_{j(2:m)}, \cdots, X_{j(m:m)}\right\}$, having the CDF $G_{(i:m)}(x)$, mean $\mu^*_{(i:m)}$ and variance $\sigma^{2*}_{(i:m)}$. Then, the DRSS unbiased estimator is defined by

$$\bar{Y}_{DRSS} = \frac{1}{mc}\sum_{j=1}^{c}\sum_{i=1}^{m} Y_{ij}, \tag{10}$$

with variance

$$Var\left(\bar{Y}_{DRSS}\right) = \frac{1}{cm^2}\sum_{i=1}^{m}\sigma^{2*}_{(i:m)} = \frac{\sigma^2}{mc} - \frac{1}{m^2c}\sum_{i=1}^{m}\left(\mu^*_{(i:m)} - \mu\right)^2, \tag{11}$$

thus, the relative efficiency of the $\bar{Y}_{DRSS}$ with respect to the SRS estimator $(\bar{X}_{SRS})$ is given as follows

$$RE = \frac{Var\left(\bar{X}_{SRS}\right)}{Var\left(\bar{Y}_{DRSS}\right)} = \frac{1}{1 - \frac{1}{m\sigma^2}\sum_{i=1}^{m}\left(\mu^*_{(i:m)} - \mu\right)^2}. \tag{12}$$

For more information about the DRSS method, refer to Al-Saleh & Al-Kadiri (2000).

3 Proposed sampling design

In this section, we present a new sampling design by implementing the EERSS method in two stages. It is called double EERSS (DEERSS), and it is described in the following steps:

1. Randomly select $(m+2)$ groups from the target population, each of size $m \times (m+2)$ units, and divide each group into m sets.
2. Apply the EERSS method on each of the $(m+2)$ groups. This step yields $(m+2)$ except extreme ranked set samples each of size m.

3. For each of the $(m+2)$ except extreme ranked set samples obtained in Step [2], move the units that have the same $(i+1)$ rank from all samples to the new i^{th} sets, for $i = 1, 2, \ldots, m$. This yields m sets each of size $(m+2)$.

4. Re-apply the EERSS again on the m sets obtained in Step [3] to get a double except extreme ranked set sample of size m.

5. The Steps [1-4] can be repeated c cycles to get a double except extreme ranked set sample of size $n = mc$.

Example 1. *Consider $m = 2$, and $c = 1$. In this case, we must randomly draw $m+2 = 4$ groups each containing $m \times (m+2) = 8$ units, divide each group into $m = 2$ sets, and then apply the EERSS method on each of the $m+2 = 4$ groups as follows:*

$$G_1 \Rightarrow \begin{Bmatrix} X^1_{1(1:4)} & \mathbf{X^1_{1(2:4)}} & X^1_{1(3:4)} & X^1_{1(4:4)} \\ X^1_{2(1:4)} & X^1_{2(2:4)} & \mathbf{X^1_{2(3:4)}} & X^1_{2(4:4)} \end{Bmatrix} \Rightarrow \begin{Bmatrix} \mathbf{X^1_{1(2:4)}} & \mathbf{X^1_{2(3:4)}} \end{Bmatrix}$$

$$G_2 \Rightarrow \begin{Bmatrix} X^2_{1(1:4)} & \mathbf{X^2_{1(2:4)}} & X^2_{1(3:4)} & X^2_{1(4:4)} \\ X^2_{2(1:4)} & X^2_{2(2:4)} & \mathbf{X^2_{2(3:4)}} & X^2_{2(4:4)} \end{Bmatrix} \Rightarrow \begin{Bmatrix} \mathbf{X^2_{1(2:4)}} & \mathbf{X^2_{2(3:4)}} \end{Bmatrix}$$

$$G_3 \Rightarrow \begin{Bmatrix} X^3_{1(1:4)} & \mathbf{X^3_{1(2:4)}} & X^3_{1(3:4)} & X^3_{1(4:4)} \\ X^3_{2(1:4)} & X^3_{2(2:4)} & \mathbf{X^3_{2(3:4)}} & X^3_{2(4:4)} \end{Bmatrix} \Rightarrow \begin{Bmatrix} \mathbf{X^3_{1(2:4)}} & \mathbf{X^3_{2(3:4)}} \end{Bmatrix}$$

$$G_4 \Rightarrow \begin{Bmatrix} X^4_{1(1:4)} & \mathbf{X^4_{1(2:4)}} & X^4_{1(3:4)} & X^4_{1(4:4)} \\ X^4_{2(1:4)} & X^4_{2(2:4)} & \mathbf{X^4_{2(3:4)}} & X^4_{2(4:4)} \end{Bmatrix} \Rightarrow \begin{Bmatrix} \mathbf{X^4_{1(2:4)}} & \mathbf{X^4_{2(3:4)}} \end{Bmatrix}$$

For each of the $m+2 = 4$ except extreme ranked set samples obtained in the previous steps, move the units with the same rank from all sets to have new i^{th} sets, for $i = 1, 2$. This yield $m = 2$ sets each of size $m+2 = 4$.

$$S \Rightarrow \begin{Bmatrix} X^1_{1(2:4)} & X^2_{1(2:4)} & X^3_{1(2:4)} & X^4_{1(2:4)} \\ X^1_{2(3:4)} & X^2_{2(3:4)} & X^3_{2(3:4)} & X^4_{2(3:4)} \end{Bmatrix}$$

Now, reapply the EERSS again on the 2 sets to get a double except extreme ranked set samples of size 2.

$$S^* \Rightarrow \begin{Bmatrix} Z^1_{(1:4)} & \mathbf{Z^1_{(2:4)}} & Z^1_{(3:4)} & Z^1_{(4:4)} \\ Z^2_{(1:4)} & Z^2_{(2:4)} & \mathbf{Z^2_{(3:4)}} & Z^2_{(4:4)} \end{Bmatrix} \Rightarrow \begin{Bmatrix} \mathbf{Z^1_{(2:4)}} & \mathbf{Z^2_{(3:4)}} \end{Bmatrix}$$

where

$$Z^1_{(2:4)} = 2^{nd} - Min \left\{ X^1_{1(2:4)}, X^2_{1(2:4)}, X^3_{1(2:4)}, X^4_{1(2:4)} \right\}$$

and

$$Z^1_{(3:4)} = 3^{rd} - Min \left\{ X^1_{2(3:4)}, X^2_{2(3:4)}, X^3_{2(3:4)}, X^4_{2(3:4)} \right\}.$$

For simplicity and in the general case of m, let us denote the DEERSS ordered variables by $Z_2, Z_3, \ldots, Z_{m+1}$. That is to say, $Z_i = (i)^{th} - Min \left\{ X^1_{j(i:m+2)}, X^2_{j(i:m+2)}, \ldots, X^{m+2}_{j(i:m+2)} \right\}$. Therefore, the CDF and PDF of the variable Z_i are defined, respectively, as follows:

$$\begin{aligned} H_{(i:m+2)}(x) &= P(Z_i \le x) \\ &= P\left(at\ least\ i\ of\ the\ \left\{ X^1_{i(i:m+2)}, X^2_{i(i:m+2)}, \ldots, X^{m+2}_{i(m+2:m+2)} \right\} \le x \right) \\ &= \sum_{s=i}^{m+2} \binom{m+2}{s} \left[F_{(i:m+2)}(x) \right]^s \left[1 - F_{(i:m+2)}(x) \right]^{m+2-s}, \end{aligned} \tag{13}$$

and

$$h_{(i:m+2)}(x) = \frac{(m+2)!}{(i-1)!\,(m+2-i)!} f_{(i:m+2)}(x) \left[F_{(i:m+2)}(x)\right]^{i-1} \times \left[1 - F_{(i:m+2)}(x)\right]^{m+2-i}, \tag{14}$$

where $f_{(i:m+2)}(x)$ and $F_{(i:m+2)}(x)$ are given in equations [1 - 2], respectively. Thus, the mean and variance of these double-ordered statistics are given, respectively, as follows:

$$\mu^{**}_{(i:m+2)} = \int_{-\infty}^{\infty} x h_{(i:m+2)}(x) dx, \tag{15}$$

and

$$\sigma^{2**}_{(i:m+2)} = \int_{-\infty}^{\infty} \left(x - \mu^{**}_{(i:m+2)}\right)^2 h_{(i:m+2)}(x) dx. \tag{16}$$

See David & Nagaraja, (2004) and Arnold et al. (2008).

Example 2. *The CDF of the DEERSS variables in the case of m=2 is determined as follows: Firstly, we have to determine at the beginning the distribution of the ordered variables in the first stage as*

$$F_{(2:4)}(x) = \sum_{j=2}^{4} \binom{4}{j} [F(x)]^j [1 - F(x)]^{4-j} = 6[F(x)]^2 [1 - F(x)]^2 + 4[F(x)]^3 [1 - F(x)]^1 + [F(x)]^4,$$

and

$$F_{(3:4)}(x) = \sum_{j=3}^{4} \binom{4}{j} [F(x)]^j [1 - F(x)]^{4-j} = 4[F(x)]^3 [1 - F(x)]^1 + [F(x)]^4.$$

After that, the CDFs of the ordered variables are determined in the second stage as follows:

$$H_{(2:4)}(x) = \sum_{j=2}^{4} \binom{4}{j} \left[F_{(2:4)}(x)\right]^j \left[1 - F_{(2:4)}(x)\right]^{4-j} = 6\left[F_{(2:4)}(x)\right]^2 \left[1 - F_{(2:4)}(x)\right]^2 + 4\left[F_{(2:4)}(x)\right]^3 \left[1 - F_{(2:4)}(x)\right]^1 + \left[F_{(2:4)}(x)\right]^4,$$

and

$$H_{(3:4)}(x) = \sum_{j=3}^{4} \binom{4}{j} \left[F_{(3:4)}(x)\right]^j \left[1 - F_{(3:4)}(x)\right]^{4-j} = 4\left[F_{(3:4)}(x)\right]^3 \left[1 - F_{(3:4)}(x)\right]^1 + \left[F_{(3:4)}(x)\right]^4.$$

Using these CDFs, we can easily determine the PDFs of the considered double-ordered statistics. Thus, their moments can be easily evaluated.

4 Estimation of population mean under DEERSS

In this section, the estimator using DEERSS is defined, along with its expectation, variance, and mean square error. In addition to its bias and relative efficiency, to make comparisons with the estimators under SRS, RSS, EERSS, and DRSS. Let $Z_{1j}, Z_{2j}, \ldots, Z_{mj}$ denote the DEERSS observations, where $j = 1, 2, \cdots, c$ is the cycle number, then the DEERSS estimator of the population mean μ is given by

$$\bar{Z}_{DEERSS} = \frac{1}{mc}\sum_{j=1}^{c}\sum_{i=2}^{m+1} Z_{j(i:m+2)}, \tag{17}$$

with mean

$$E\left(\bar{Z}_{DEERSS}\right) = \frac{1}{m}\sum_{i=2}^{m+1}\mu^{**}_{(i:m+2)}. \tag{18}$$

Theorem 1. *Let $f(x)$ be a continuous symmetric distribution with the population mean μ. Then, the DEERSS estimator $\bar{Z}_{DEERSS}$ is an unbiased estimator of the population mean μ.*

Proof. This theorem can be proved by the symmetry fact that $\mu^{(k)}_{(i:n)} = (-1)^k\,\mu^{(k)}_{(n-i+1:n)}$ (Arnold et al., 2008). Where, $\mu^{(k)}_{(i:n)}$ is the k^{th} moment around the mean μ, of the i^{th} ordered statistics, and $\mu^{(k)}_{(n-i+1:n)}$ is the k^{th} moment around the mean μ of the $(n-i+1:n)^{th}$ ordered statistics. This means, for example at the $k=1$, The moment for each ordered statistic on the left side is equal in magnitude to the corresponding ordered statistic on the right side, with a difference in its location relative to the mean. Therefore,

$$\mu^{**}_{(2:m+2)} - \mu = \mu - \mu^{**}_{(m+1:m+2)} \Longleftrightarrow \mu^{**}_{(2:m+2)} + \mu^{**}_{(m+1:m+2)} = 2\mu,$$

similarly,

$$\mu^{**}_{(3:m+2)} + \mu^{**}_{(m:m+2)} = 2\mu \ldots, Median = \mu,$$

thus

$$E\left(\bar{Z}_{DEERSS}\right) = \frac{1}{m}\sum_{i=2}^{m+1}\mu^{**}_{(i:m+2)} = \frac{1}{m}m\mu = \mu.$$

The proof is complete. □

Now, the variance of the $\bar{Z}_{DEERSS}$ is given by

$$\begin{aligned} Var\left(\bar{Z}_{DEERSS}\right) &= \frac{1}{m^2c}\sum_{i=2}^{m+1}\sigma^{2**}_{(i:m+2)} \\ &= \frac{\sigma^2}{cm} - \frac{1}{cm^2}\left[\sigma^{2**}_{(1:m+2)} + \sigma^{2**}_{(m+2:m+2)} - 2\sigma^2 + \sum_{i=1}^{m+2}\left(\mu^{**}_{(i:m+2)} - \mu\right)^2\right], \end{aligned} \tag{19}$$

thus, the bias and mean square error (MSE) of the estimator are given, respectively, as follows

$$bias = \frac{1}{m}\sum_{i=2}^{m+1}\mu^{**}_{(i:m+2)} - \mu = \frac{1}{m}\sum_{i=2}^{m+1}\left(\mu^{**}_{(i:m+2)} - \mu\right), \tag{20}$$

and

$$MSE\left(\bar{Z}_{DEERSS}\right) = \frac{1}{m^2c}\sum_{i=2}^{m+1}\sigma^{2**}_{(i:m+2)} + \frac{1}{m^2}\left[\sum_{i=2}^{m+1}\left(\mu^{**}_{(i:m+2)} - \mu\right)\right]^2. \tag{21}$$

Therefore, the relative efficiency of the $\bar{Z}_{DEERSS}$ with respect to the $\bar{X}_{SRS}$ is given by

$$RP = \frac{Var\left(\bar{X}_{SRS}\right)}{MSE\left(\bar{Z}_{DEERSS}\right)} = \frac{\sigma^2}{\frac{1}{m}\sum_{i=2}^{m+1}\sigma^{2**}_{(i:m+2)} + \frac{c}{m}\left[\sum_{i=2}^{m+1}\left(\mu^{**}_{(i:m+2)} - \mu\right)\right]^2}. \tag{22}$$

To evaluate each of $\mu^{**}_{(i:m+2)}$ and $\sigma^{2**}_{(i:m+2)}$, follow example 2 to determine the densities of the considered double ordered variables, and then use the equations [15-16] to evaluate them. Therefore, the quantities of the bias, MSE, and RE can be easily calculated.

5 Results of comparisons

In this section, comparisons are made with the other considered sampling methods. The *RE* and the *bias* of the DEERSS estimator are calculated based on the formulas [22, 20]. It includes calculating the RE and the bias of the DEERSS estimator for some symmetric distributions ($Uniform\,(0,1)$, $Normal\,(0,1)$, $Student\ (4)$, $Logistic\,(5,2)$ and $Beta\,(3,3)$) . Also, for some asymmetric, $Exp\,(1)$, $Beta\,(5,2)$, $Gamma\,(2,3)$, $HalfNormal\,(2)$, $Rayliegh\,(1)$, $Weibull\,(1,1)$, and $\chi^2(5)$. The *RE* and the *bias* are evaluated at different values of $m = 2, 3, 4, 5$ and 6. Each of the corresponding DRSS, EERSS, and RSS estimators to the SRS estimator is evaluated to make the comparisons. All calculations are done using Mathematica 13.3 Software. The exact values are obtained in all cases except for each of the distributions ($Normal(0,1)$, $Student(4)$, $Logistic(5,2)$, $HalfNormal\,(2)$ and $\chi^2(5)$), where in these cases, the numerical methods are used in the calculations.

Tables 1–5 report the results of the RE for some symmetric and asymmetric distributions at different values of $m = 2, 3, 4, 5$, and 6. Table 6 reports the results of the bias of both the DEERSS and EERSS estimators in the case of asymmetrical distributions at the same considered values of $m = 2, 3, 4, 5, 6$. To help readers understand Tables 1 - 6, the results of RE and bias are also summarized in Figures 1 and 2, respectively. Based on these results, we conclude that

Table 1: Relative efficiency based on each of DEERSS, DRSS, EERSS, and RSS to SRS at m=2

		Sampling Method			
	Distribution	DEERSS	DRSS	EERSS	RSS
Symmetrical	$Uniform\,(0,1)$	5.5135	1.9231	2.08333	1.5000
	$Normal\,(0,1)$	7.6322	1.7852	2.7743	1.4669
	$Student\,(4)$	11.4227	1.6325	3.9304	1.3719
	$Logistic\,(5,2)$	8.9134	1.7057	3.1637	1.4367
	$Beta\,(3,3)$	6.6482	1.8525	2.4641	1.4880
Asymmetrical	$Exp\,(1)$	5.4157	1.5158	2.8235	1.3333
	$Beta\,(5,2)$	6.3461	1.7946	2.5031	1.4639
	$Gamma\,(2,3)$	6.1847	1.6290	2.7849	1.3913
	$HN\,(2)$	5.8353	1.7209	2.51071	1.4298
	$Rayleigh\,(1)$	6.7504	1.7742	2.6225	1.4576
	$Weibull\,(1,1)$	5.4157	1.5158	2.8235	1.3333
	$\chi^2(5)$	6.4041	1.6563	2.7138	1.4658

Table 2: Relative efficiency based on each of DEERSS, DRSS, EERSS, and RSS to SRS at m=3

		Sampling Method			
	Distribution	DEERSS	DRSS	EERSS	RSS
Symmetrical	$Uniform\,(0,1)$	8.6601	3.0256	2.5200	2.0000
	$Normal\,(0,1)$	10.6285	2.6633	3.2972	1.9138
	$Student\,(4)$	14.6872	2.2201	4.5910	1.6853
	$Logistic\,(5,2)$	11.8680	2.4133	3.7281	1.8381
	$Beta\,(3,3)$	9.6625	2.8284	2.9498	1.9682
Asymmetrical	$Exp\,(1)$	6.2899	2.0236	2.9605	1.6364
	$Beta\,(5,2)$	8.8055	2.6805	2.9332	1.9146
	$Gamma\,(2,3)$	7.6468	2.2664	3.0854	1.7532
	$HN\,(2)$	7.7143	2.4992	2.8586	1.8409
	$Rayleigh\,(1)$	9.2352	2.6176	3.0724	1.8975
	$Weibull\,(1,1)$	6.2899	2.0236	2.96053	1.6364
	$\chi^2(5)$	8.0583	2.3277	3.2130	1.9128

Table 3: Relative efficiency based on each of DEERSS, DRSS, EERSS, and RSS to SRS at m=4

		Sampling Method			
	Distribution	DEERSS	DRSS	EERSS	RSS
Symmetrical	$Uniform\,(0,1)$	12.5345	4.2808	2.9697	2.5000
	$Normal\,(0,1)$	13.9469	3.5264	3.8038	2.3469
	$Student\,(4)$	17.8545	2.7730	5.1922	1.9627
	$Logistic\,(5,2)$	14.9492	3.1199	4.2587	2.2164
	$Beta\,(3,3)$	13.1741	3.9025	3.4323	2.4433
Asymmetrical	$Exp\,(1)$	7.5720	2.5232	3.0509	1.9200
	$Beta\,(5,2)$	11.7075	3.6364	3.3407	2.3565
	$Gamma\,(2,3)$	9.4538	2.9066	3.3301	2.0958
	$HN\,(2)$	9.9942	3.3172	3.1674	2.2393
	$Rayleigh\,(1)$	12.0753	3.5105	3.4957	2.3251
	$Weibull\,(1,1)$	7.7520	2.5232	3.0509	1.9200
	$\chi^2(5)$	10.0422	3.0066	3.6931	2.3471

Table 4: Relative efficiency based on each of DEERSS, DRSS, EERSS, and RSS to SRS at m=5

		Sampling Method			
	Distribution	DEERSS	DRSS	EERSS	RSS
Symmetrical	$Uniform\,(0,1)$	17.1110	5.6705	3.4286	3.0000
	$Normal\,(0,1)$	17.5611	4.4556	4.2979	2.7702
	$Student\,(4)$	20.9197	3.2983	5.7497	2.2150
	$Logistic\,(5,2)$	18.1418	3.8252	4.7642	2.5783
	$Beta\,(3,3)$	17.1639	5.0609	3.9120	2.9145
Asymmetrical	$Exp\,(1)$	9.2997	3.0160	3.1212	2.1898
	$Beta\,(5,2)$	15.1196	4.6510	3.7302	2.7920
	$Gamma\,(2,3)$	11.6941	3.5484	3.5391	2.4244
	$HN\,(2)$	12.7912	4.1663	3.44759	2.6284
	$Rayleigh\,(1)$	15.3289	4.4431	3.8975	2.7436
	$Weibull\,(1,1)$	9.2997	3.0160	3.1212	2.1898
	$\chi^2(5)$	12.4487	3.6906	4.1579	2.7721

Table 5: Relative efficiency based on each of DEERSS, DRSS, EERSS, and RSS to SRS at m=6

		Sampling Method			
	Distribution	DEERSS	DRSS	EERSS	RSS
Symmetrical	$Uniform\,(0,1)$	22.3703	7.1815	3.8942	3.5000
	$Normal\,(0,1)$	21.4502	5.4155	4.7820	3.1857
	$Student\,(4)$	23.8812	3.8011	6.2731	2.4484
	$Logistic\,(5,2)$	21.4331	4.5295	5.2501	2.9276
	$Beta\,(3,3)$	21.6172	6.2941	4.3893	3.3829
Asymmetrical	$Exp\,(1)$	11.5398	3.5036	3.1827	2.4490
	$Beta\,(5,2)$	19.0897	5.7169	4.1050	3.2227
	$Gamma\,(2,3)$	14.4431	4.1913	3.7240	2.7423
	$HN\,(2)$	16.2038	5.0413	3.7061	3.0100
	$Rayleigh\,(1)$	19.0417	5.4093	4.2816	3.1551
	$Weibull\,(1,1)$	11.5398	3.5036	3.1827	2.4490
	$\chi^2(5)$	15.3507	4.3786	4.6103	3.1899

Table 6: Bias of the mean estimator based on each of DEERSS and EERSS at $m = 2, 3, 4, 5$, and 6

	$\lvert Bias\,(\hat{\mu}_{DEERSS})\rvert$				
Distribution	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
$Exp\,(1)$	0.1958	0.1547	0.1215	0.0947	0.0729
$Beta\,(5,2)$	0.0130	0.0100	0.0076	0.0058	0.0043
$Gamma\,(2,3)$	0.0689	0.0546	0.0430	0.0336	0.0260
$HN\,(2)$	0.1544	0.1205	0.0933	0.0715	0.0539
$Rayleigh\,(1)$	0.0491	0.0391	0.0309	0.0242	0.0188
$Weibull\,(1,1)$	0.1958	0.1547	0.1215	0.0947	0.0729
$\chi^2(5)$	0.4172	0.3310	0.2610	0.2043	0.1580
	$\lvert Bias\,(\hat{\mu}_{EERSS})\rvert$				
	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
$Exp\,(1)$	0.1667	0.1611	0.1542	0.1471	0.1405
$Beta\,(5,2)$	0.0110	0.0106	0.0101	0.0096	0.0091
$Gamma\,(2,3)$	0.0586	0.0567	0.0543	0.0518	0.0495
$HN\,(2)$	0.1310	0.1265	0.1208	0.1151	0.1097
$Rayleigh\,(1)$	0.0418	0.0405	0.0388	0.0370	0.0354
$Weibull\,(1,1)$	0.1667	0.1611	0.1542	0.1471	0.1405
$\chi^2(5)$	0.0231	0.0224	0.0214	0.0205	0.0196

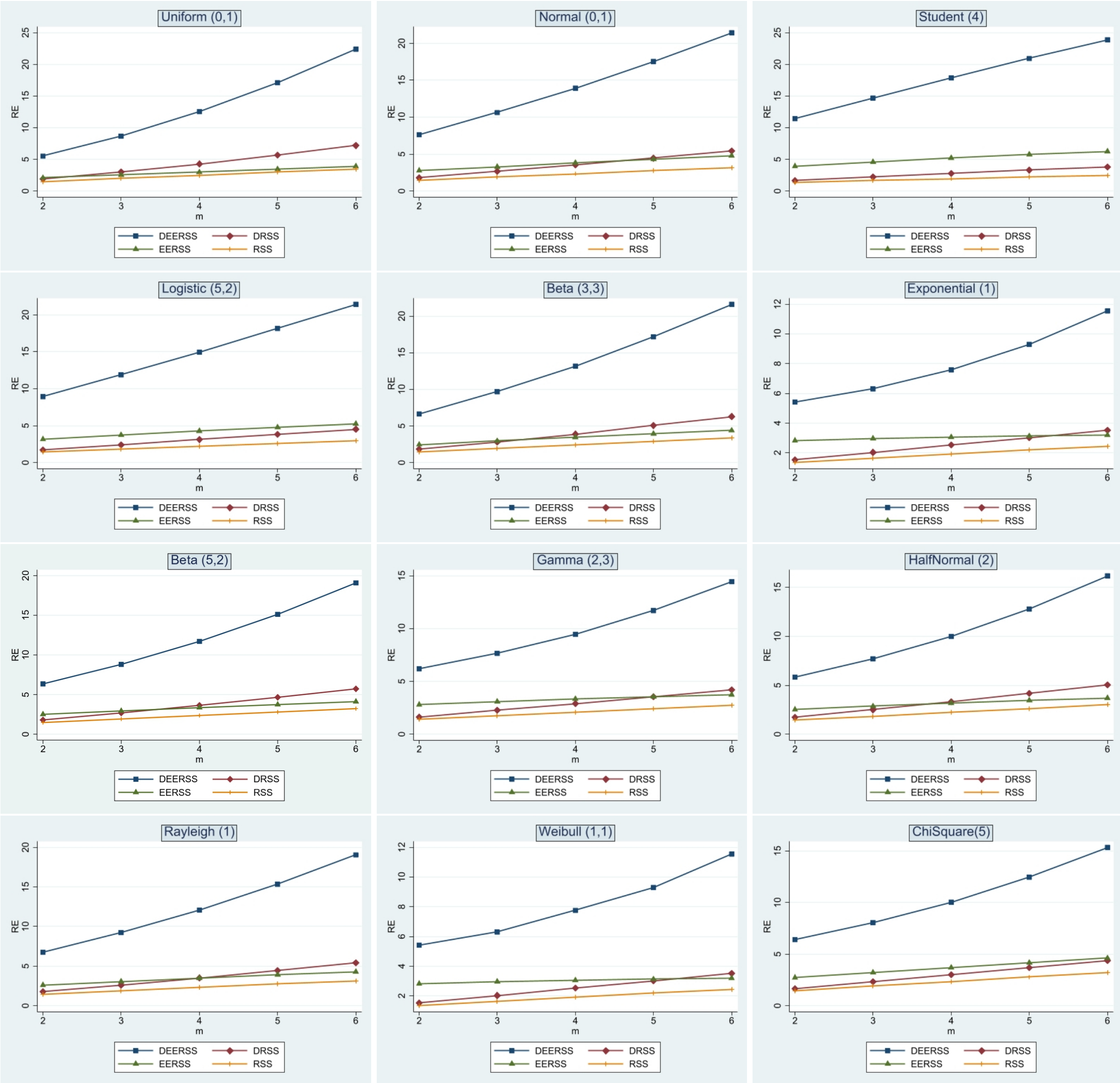

Figure 1: Relative efficiency based on each of DEERSS, DRSS, EERSS, and RSS to SRS at $m = 2, 3, 4, 5, 6$

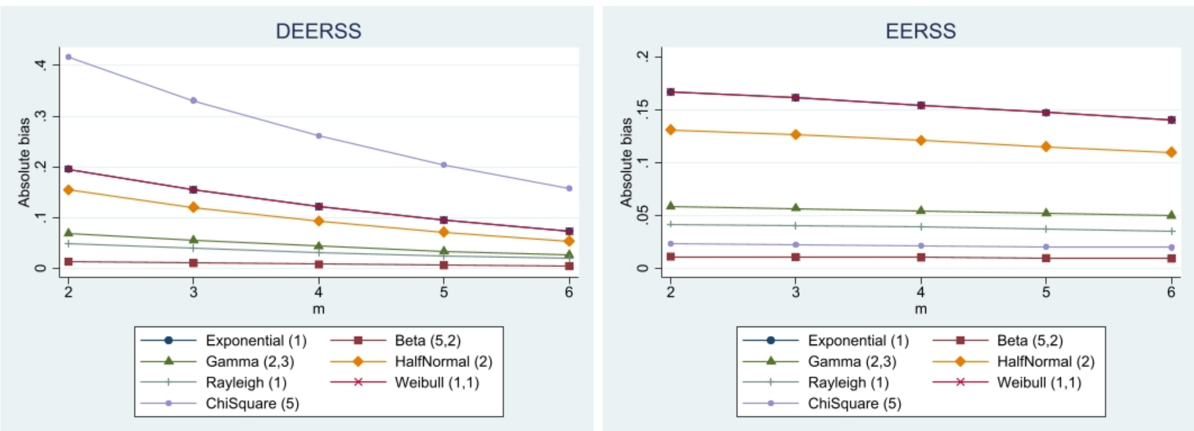

Figure 2: Absolute bias of the estimator based on each of DEERSS and EERSS at $m = 2, 3, 4, 5, 6$

- For a fixed value of m and the same parent distribution, the RE differs from one sampling method to another.
- In all sampling methods DEERSS, DRSS, EERSS, and RSS, all values of RE are more than one, which means that all these estimators are more precise than the SRS estimator.
- For a fixed value of m and the same sampling method, the RE varies from one distribution to another. For instance, the RE of the DEERSS estimator at $m = 2$ is the smallest for the $Exp\,(1)$ distribution; it is 5.5135, and the largest for the $Student\,(4)$ distribution; is 11.4227. Also, for a fixed value of m, it cannot be said that the efficiency of symmetric distributions is always greater than that of asymmetric distributions.
- For the same parent distribution and sampling method, the values of the RE increase in m for all considered sampling methods. To explain that in the case of $Uniform\,(0, 1)$ at $m = 2, 4$, and 6, the REs of the DEERSS estimator, respectively, are 5.5135, 12.5345, and 22.3703, the REs of the DRSS estimator, respectively, are 1.9231, 4.2808, and 7.1815, the REs of the EERSS estimator, respectively, are 2.08333, 2.9697, and 3.8942, the REs of the RSS estimator, respectively, are 1.5, 2.5 and 3.5. Also, we see from the results of the RE that each DEERSS and DRSS increases faster than each EERSS and RSS (Figure 1).
- For fixed m and at the same parent distribution, the DEERSS estimator has the largest RE in all cases considered in the study (Figure 1).
- For fixed m and at the same parent distribution, the DEERSS estimator has the largest RE in all cases considered in the study. This means that DEERSS is the best estimator among all considered estimators. There is a significant difference that can be seen in the results.
- All of these are strong indications of the ability of the DEERSS to reduce the sample size to be measured.
- The results of the bias show that the estimators in both methods, RSS and DRSS, are unbiased in the case of symmetrical and asymmetrical distributions, while the DEERSS and EERSS estimators are only unbiased for the symmetrical distribution. This is consistent with the results of Theorem 1.
- For $m = 2$, the bias of the DEERSS estimator is greater than that of the EERSS estimator for all asymmetrical distributions considered in this study. In contrast, for $m > 2$, the bias of the DEERSS estimator is smaller than that of the EERSS estimator for all asymmetrical distributions considered in this study, except for the $\chi^2(5)$ distribution (Table 6). Additionally, the bias values of both DEERSS and EERSS estimators decrease as m increases (Figure 2).

6 Applications

In this section, three real-life examples will be studied. This is to verify the applicability of the sampling method introduced in this study and to validate the relative efficiency results in the previous section. Data sets about some of the most important agricultural indicators were obtained from Word Bank Data websites. The populations are described as follows: The first data set presents the percent of arable land out of the total land in 258 countries for the year 2021 (World Bank Data, 2022a). While the second data set presents the percent of production from agriculture, fishing, and forestry out of the gross domestic product, for 262 countries in 2021 (World Bank Data, 2022b). Lastly, the third data set represents the percent of forest areas out of the total area in 266 countries for the year 2021 (World Bank Data, 2022c).

Before analyzing the real data sets, the missing and outlier values were treated using the MATLAB 2023b program. Table 7 shows the descriptive statistics for the three data sets after cleaning data, where the mean of the percent of agricultural land (Data set 1), the percent of agricultural production (Data set 2), and the percent of forest land (Data set 3) are, respectively, 37.2522, 9.1496, and 31.3706. Figure 3 shows that the first data set is approximately symmetric, while the second and third data sets are non-symmetric. The skewness values in Table 7 support these results.

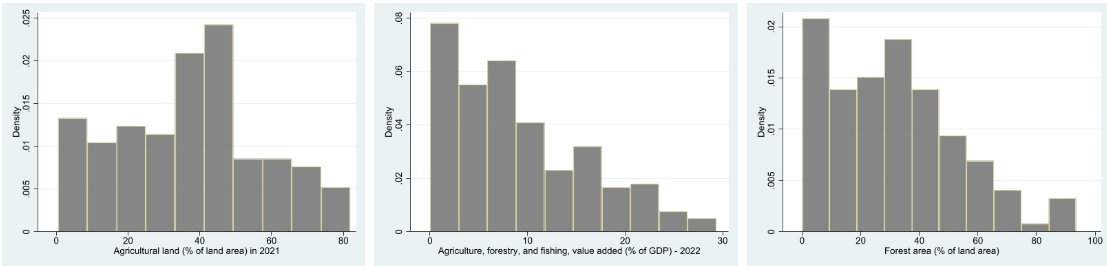

Figure 3: Histogram of the real data sets

Table 7: Descriptive statistics of the real data sets

	N	Min	Max	Mean	Q1	Q2	Q3	Var	Skewness
Pop1	258	0.5000	81.8925	37.2522	20.6258	38.5132	48.8254	419.3468	0.1162
Pop2	266	0.0281	29.3329	9.1496	3.3905	7.4930	13.6199	48.2194	0.8057
Pop3	262	0.0000	93.5033	31.3706	11.7228	30.3187	45.8381	504.4981	0.5676

To illustrate the applicability of the DEERSS method and its efficiency compared to SRS, the estimator of the population mean was calculated using the DEERSS method. The RE and bias measures were used for illustration. Both RE and bias were calculated using the formulas (22, 20) via MATLAB 2023b, with 1,000,000 repetitions.

The RE of the estimators compared to SRS and their bias values are presented in Tables 8 and 9, and illustrated in Figures 4 and 5, respectively. It is clear from these results that the RE of $\bar{Z}_{DEERSS}$ are the largest among all estimators and increase in m (Figure 4). These results are also consistent with the results in the previous section.

Table 8: Relative efficiency of the mean estimator for the real data sets based on each of DEERSS, DRSS, EERSS, and RSS to SRS at m=2,3,4 and 5

Population	Sampling Method	m=2	m=3	m=4	m=5
Population 1	DEERSS	6.5449	9.3161	13.9303	21.8861
	DRSS	1.8520	2.8661	4.0545	5.4764
	EERSS	2.5488	3.0723	3.5960	4.1329
	RSS	1.4961	1.9933	2.5124	3.0332
Population 2	DEERSS	4.9146	7.6038	11.9296	17.815
	DRSS	1.7811	2.7140	3.8173	5.0937
	EERSS	2.2806	2.6338	2.9750	3.3201
	RSS	1.4562	1.9164	2.3846	2.8774
Population 3	DEERSS	6.4558	9.4905	13.4400	18.5413
	DRSS	1.8172	2.7572	3.8193	5.0602
	EERSS	2.5344	3.0304	3.5130	3.9934
	RSS	1.4776	1.9465	2.4352	2.9280

Table 9: Absolute bias of the estimators for the real data sets based on each of DEERSS, DRSS, EERSS, and RSS at m=2,3,4 and 5

Population	Sampling Method	m=2	m=3	m=4	m=5
Population 1	DEERSS	0.4238	0.4922	0.4194	0.3473
	DRSS	0.0054	0.0048	0.0016	0.0027
	EERSS	0.2171	0.2166	0.2217	0.2465
	RSS	0.0285	0.0095	0.0005	0.0026
Population 2	DEERSS	0.9563	0.6553	0.4649	0.3173
	DRSS	0.0013	0.0033	0.0004	0.0005
	EERSS	0.8006	0.7730	0.7258	0.6829
	RSS	0.0016	0.0019	0.0043	0.0019
Population 3	DEERSS	1.8166	1.5992	1.3788	1.1932
	DRSS	0.0039	0.0049	0.0072	0.0007
	EERSS	1.4441	1.4247	1.3893	1.3516
	RSS	0.0068	0.0141	0.0037	0.0058

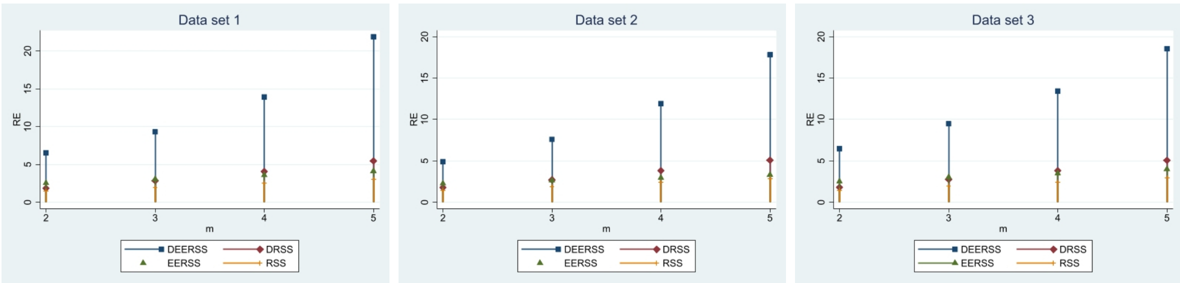

Figure 4: Relative efficiency of the mean estimator for the three data sets based on DEERSS, DRSS, EERSS, and RSS to SRS at $m = 2, 3, 4, 5$

Figure 5: Absolute bias of the mean estimator for the three data sets based on DEERSS, DRSS, EERSS, and RSS at $m = 2, 3, 4, 5$

7 Conclusions

This study introduced a new sampling method to estimate the population mean. It is proven that the estimator using this method is unbiased for symmetrical distributions. The results show that this estimator has a higher efficiency than its counterparts based on DRSS, EERSS, and RSS. Three real data sets are used to illustrate the new estimator. Therefore, if the assumption of easy judgment ranking in the two stages is achieved, then using this method to estimate the population mean will be more effective than the other considered estimators.

This study focused on estimating the population mean. This method can be used in future applications, such as estimating the population median, quartiles, variance, proportion, ratio, and reliability.

References

Abdallah, M.S., Al-Omari, A.I., Alotaibi, N., Alomani, G.A., & Al-Moisheer, A. (2022). Estimation of distribution function using l ranked set sampling and robust extreme ranked set sampling with application to reliability. *Computational Statistics, 37*(5), 2333-2362.

Al-Mawan, N.M., Rady, E., & Rashwan, N. (2018). Median double ranked set sampling method. *Journal of Advances in Mathematics, 14*(01).

Al-Nasser, A.D. (2007). L ranked set sampling: A generalization procedure for robust visual sampling. *Communications in Statistics—Simulation and Computation R, 36*(1), 33-43.

Al-Nasser, A.D., Al-Omari, A.I. (2018). Minimax ranked set sampling. *Investigacion Operacional, 39*(4), 560-571.

Al-Omari, A.I. (2010). Estimation of the population median of symmetric and asymmetric distributions using double robust extreme ranked set sampling. *Investigacison Operacional, 31*(3), 199-207.

Al-Omari, A.I. (2016). Quartile ranked set sampling for estimating the distribution function. *Journal of the Egyptian Mathematical Society, 24*(2), 303-308.

Al-Omari, A.I., Abdallah, M.S. (2023). Estimation of the distribution function using moving extreme and minimax ranked set sampling. *Communications in Statistics-Simulation and Computation, 52*(5), 1909-1925.

Al-Omari, A.I., Al-Nasser, A.D. (2018). Ratio estimation using multistage median ranked set sampling approach. *Journal of Statistical Theory and Practice*, 12, 512-529.

Al-Omari, A.I., Haq, A. (2019). A new sampling method for estimating the population mean. *Journal of Statistical Computation and Simulation, 89*(11), 1973-1985.

Al-Saleh, M.F., Al-Kadiri, M.A. (2000). Double-ranked set sampling. *Statistics & Proba- bility Letters, 48*(2), 205-212.

Al-Saleh, M.F., Al-Omari, A.I. (2002). Multistage ranked set sampling. *Journal of Statistical planning and Inference, 102*(2), 273-286.

Aldrabseh, M.Z., Ismail, M.T. (2023). New modification of ranked set sampling for estimating population mean. *Journal of Statistical Computation and Simulation*, pages 1-13.

Arnold, B.C., Balakrishnan, N., & Nagaraja, H.N. (2008). A first course in order statistics. SIAM.

Chen, M., Chen, W.-X., & Yang, R. (2024). Double moving extremes ranked set sampling design. *Acta Mathematicae Applicatae Sinica, English Series, 40*(1), 75-90.

David, H.A., Nagaraja, H.N. (2004). *Order statistics*. John Wiley & Sons.

Dell, T., Clutter, J. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, pages 545-555.

World Bank Data (2022a). Agricultural land (percent of land area). `https://data.worldbank.org/indicator/AG.LND.AGRI.ZS?view=chart`.

World Bank Data (2022b). Agriculture, forestry, and fishing, value added (percent of gdp). `https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?view=chart`.

World Bank Data (2022c). Forest area (percent of land area). `https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?view=chart`.

Hanandeh, A., Al-Nasser, A.D., & Al-Omari, A.I. (2022a). New double stage ranked set sampling for estimating the population mean. *Electronic Journal of Applied Statistical Analysis, 15*(2), 463-478.

Hanandeh, A.A., Al-Nasser, A.D., & Al-Omari, A.I. (2022b). New mixed ranked set sampling variations. *Stat, 11*(1), e417.

Haq, A., Brown, J., Moltchanova, E., & Al-Omari, A.I. (2014). Mixed ranked set sampling design. *Journal of Applied Statistics, 41*(10), 141-2156.

Haq, A., Brown, J., Moltchanova, E., & Al-Omari, A.I. (2016). Paired double-ranked set sampling. *Communications in Statistics-Theory and Methods, 45*(10), 2873-2889.

Khan, Z., Ismail, M., & Samawi, H. (2020). Mixture ranked set sampling for estimation of population mean and median. *Journal of Statistical Computation and Simulation, 90*(4), 573-585.

Mahdizadeh, M., Zamanzade, E. (2022). On interval estimation of the population mean in ranked set sampling. *Communications in Statistics: Simulation and Computation, 51*(5), 2747-2768.

McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research, 3*(4), 385-390.

Shang, L.-F., Yan, Z.-Z. (2024). Reliability estimation stress-strength dependent model based on copula function using ranked set sampling. *Journal of Radiation Research and Applied Sciences, 17*(1), 100811.

Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics*, pages 35-42.

Takahasi, K., Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratiled by means of ordering. *Annals of the Institute of Statistical Mathematics, 20*(1), 1-31.

Zamanzade, E., Al-Omari, A.I. (2016). New ranked set sampling for estimating the population mean and variance. *Hacettepe Journal of Mathematics and Statistics, 45*(6), 1891-1905.