REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 347–366
https://doi.org/10.57805/revstat.v21i3.406

The Effects of Ranking Error Models on Mean Estimators Based on Ranked Set Sampling

Authors: Sami Akdeniz
– The Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey
samiakdeniz19@gmail.com

Tugba Ozkal Yildiz ✉
– Department of Statistics, Dokuz Eylul University, Izmir, Turkey
tugba.ozkal@deu.edu.tr

Received: November 2020 Revised: February 2022 Accepted: February 2022

Abstract:

- Ranked Set Sampling (RSS) is a sampling method commonly used in recent years. This sampling method is especially useful for studies in medicine, agriculture, forestry and ecology. In this study, the widely used ranking error models in RSS literature are investigated. This study is aimed to explore the effects of ranking error models on the mean estimators based on RSS and some of its modified methods such as Extreme RSS (ERSS) and Percentile RSS (PRSS) for different distribution, set and cycle size in infinite population. Monte Carlo simulation study is conducted for this purpose. Additionally, the study is supported by real life data. It is observed that, RSS and some of its modified methods shows better results than Simple Random Sampling (SRS).

Keywords:

- *Ranked set sampling; ranking error models; relative efficiency; mean estimator; abalone dataset.*

AMS Subject Classification:

- 62P10, 62D99, 68U20.

✉ Corresponding author.

1. INTRODUCTION

RSS is developed as an alternative to SRS in order to estimate population parameters more efficiently where the measurement of sampling units is difficult or costly but the units are easier to rank. McIntyre [12] was the first to propose the use of RSS in the pasture research to estimate the mean amount of crops. Afterwards, Halls and Dell [11] used this method to estimate the mean weights of trees and plant leaves in pine forests located in the east of Texas. In order to compare the variances of the means obtained from RSS and SRS methods Evans [7] carried out a study on long leaf pine trees. The first mathematical theory of RSS in infinite population was developed by Takahasi and Wakimoto [24]. They also demonstrated that the estimator of the population mean obtained by RSS is unbiased and its variance is smaller than SRS when the errors in the ranking are ignored. Dell and Clutter [6] examined errors of ranking in RSS. They showed that the mean estimator of RSS is an unbiased estimator of population mean when ranking is imperfect. David and Levine [5] conducted a study to determine the effects of the errors in the ranking in RSS. The concept of concomitant variable for RSS which is an effective way to increase the accuracy of ranking was proposed by Stokes [22]. This variable should be highly correlated with the variable of interest. Also, Stokes [23] suggested RSS based variance estimator which is asmptotically unbiased and more efficient compared to SRS based variance estimator. In order to review other results and examples for RSS see these studies, Patil *et al.* [20] and Al-Omari and Bouza [1]. Also detailed information regarding theoretical and applicational studies based on RSS can be found in Chen *et al.* [4].

Ranking of the units in a set is made on the basis of the visual judgement of the researcher or a concomitant variable which has a strong correlation with the variable of interest. These ranking methods are defined as ranking error models. There are many studies in the literature that are focused on the modelling of ranking errors. Primarily, Dell and Clutter [6] developed a model including a term of random error for the observations. Later, Bohn and Wolfe [3] proposed a ranking error model based on the expected value of the difference between two order statistics. Fligner and MacEachern [9] used the principle of monotone likelihood ratio to model the ranking information in RSS. New class of models is presented for imperfect rankings, in a study carried out by Frey [10]. A calibration model is developed by Ozturk [17] to reduce the errors in the ranking for RSS. Besides, Ozturk [18] suggested inference techniques for ranked set sample data in the presence of judgement ranking errors. Alexandridis and Ozturk [2] developed robust statistical inference against imperfect ranking in a ranked set sample data obtained from a family of discrete distributions. By taking the ranking errors in RSS into account, Ozturk [19] obtained non-parametric maximum likelihood estimators.

The motivation of this study is to see the effects of ranking error models on the mean estimators of RSS and some of its modified methods and compare them with the mean estimator of SRS. For this purpose, a simulation study is conducted. In addition, an abalone data set is used to support the results of the simulation study.

This study consists of six sections. The first section includes the aim of the study and literarure review on RSS. The second section contains methodological background and detailed information about RSS, extreme RSS (ERSS) and percentile RSS (PRSS).

Ranking error models in RSS literature such as visual ranked set sampling (VRSS) and concomitant ranked set sampling (CRSS) are defined in section 3. In addition, a Monte Carlo simulation study is conducted to determine the effects of ranking error models on the MSE of the mean estimators based on RSS and some of its modified methods. Besides, abalone data set is used for comparing the results obtained from the simulation study in section 5. The final section contains the conclusions.

2. RANKED SET SAMPLING AND SOME OF ITS MODIFIED METHODS

2.1. Ranked Set Sampling

In recent years, RSS is a commonly used sampling method in literature. RSS was introduced by McIntyre [12] as an alternative sampling method to SRS in order to estimate the population parameters more efficiently. It is useful and preferable method due to several important factors. Set size and the relative costs of various operations such as sampling, ranking and measurement are the most important ones among these factors. Also RSS provides advantages due to its features such as the ability to work with finite or infinite populations and it does not require to measure all units in the selected sample in RSS.

There are two important parameters in RSS. These are the set size and the number of cycles which are denoted by n and m, respectively. The set size in RSS usually ranges from 2 to 5. Also, there are many studies available in the literature in which more sets are used. On the other hand, there is no limit for the number of cycles. RSS procedure is applied in 5 steps which are described as below:

1. Select a sample of size n^2 from the population of interest using SRS.

2. Divide this randomly chosen sample of size n^2 into n sets with size n.

3. Rank the units within each set via cost effective and straightforward measurement. This ranking can be made by using visual ranking method, a concomitant variable or other methods.

4. Select the smallest ranked unit from the first set, the second smallest ranked unit from the second set and the n-th smallest ranked unit from the n-th set for actual measurement of units.

5. This process is repeated m times, until maintaining the required sample size.

The following expression represents the RSS procedure for one cycle:

$$\begin{bmatrix} \boldsymbol{X_{1[1:n]}} \leq X_{1[2:n]} \leq X_{1[3:n]} \leq \cdots \leq X_{1[n:n]} \\ X_{2[1:n]} \leq \boldsymbol{X_{2[2:n]}} \leq X_{2[3:n]} \leq \cdots \leq X_{2[n:n]} \\ X_{3[1:n]} \leq X_{3[2:n]} \leq \boldsymbol{X_{3[3:n]}} \leq \cdots \leq X_{3[n:n]} \\ \vdots \\ X_{n[1:n]} \leq X_{n[2:n]} \leq X_{n[3:n]} \leq \cdots \leq \boldsymbol{X_{n[n:n]}} \end{bmatrix}.$$

Here, $X_{(i[k:n]j)}$ represents the unit which has the rank of k in the i-th set and j-th cycle where $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$. The obtained ranked set sample for n set and m cycle can be shown as

$$\begin{bmatrix} X_{[1]1} & X_{[2]1} & X_{[3]1} & \cdots & X_{[i]1} \\ X_{[1]2} & X_{[2]2} & X_{[3]2} & \cdots & X_{[i]2} \\ X_{[1]3} & X_{[2]3} & X_{[3]3} & \cdots & X_{[i]3} \\ \vdots & \vdots & & \vdots & \\ X_{[1]j} & X_{[2]j} & X_{[3]j} & \cdots & X_{[i]j} \end{bmatrix},$$

where $X_{[i]j}$ denotes the i-th ranked observation in the j-th cycle for i and j changing from 1 to n and m, respectively. The sets in RSS are random samples that are elements of the i-th set $X_{[i]1}, X_{[i]2}, ..., X_{[i]j}$ and each set has the same distribution function $F(x;\theta)$ and same probability density function $f(x;\theta)$, where $i = 1, 2, 3, ..., n$ and $j = 1, 2, 3, ..., m$.

The sample mean estimator of the population mean for RSS can be shown as

$$\bar{X}_{\text{RSS}} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} X_{[i]j}\,. \tag{2.1}$$

Also the variance of the mean estimator for RSS can be shown as

$$Var(\bar{X}_{\text{RSS}}) = \frac{\sigma_x^2}{mn} \left[1 - \sum_{i=1}^{n} \frac{(E(X_{[i]j}) - \mu_x)^2}{n\sigma_x^2} \right], \tag{2.2}$$

where, μ_x and σ_x^2 are the mean and the variance of the population of interest, respectively.

2.2. Extreme Ranked Set Sampling

ERSS is developed by Samawi *et al.* [21] to estimate the population parameters more efficiently than SRS with the same number of units by only using the minimum and maximum ranked units for n when it is even and, the median ranked unit when it is odd.

For example, when $n = 6$, extreme ranked set sample is given below:

$$\begin{bmatrix} \boldsymbol{X_{1[1:6]}} \le X_{1[2:6]} \le X_{1[3:6]} \le X_{1[4:6]} \le X_{1[5:6]} \le X_{1[6:6]} \\ \boldsymbol{X_{2[1:6]}} \le X_{2[2:6]} \le X_{2[3:6]} \le X_{2[4:6]} \le X_{2[5:6]} \le X_{2[6:6]} \\ \boldsymbol{X_{3[1:6]}} \le X_{3[2:6]} \le X_{3[3:6]} \le X_{3[4:6]} \le X_{3[5:6]} \le X_{3[6:6]} \\ X_{4[1:6]} \le X_{4[2:6]} \le X_{4[3:6]} \le X_{4[4:6]} \le X_{4[5:6]} \le \boldsymbol{X_{4[6:6]}} \\ X_{5[1:6]} \le X_{5[2:6]} \le X_{5[3:6]} \le X_{5[4:6]} \le X_{5[5:6]} \le \boldsymbol{X_{5[6:6]}} \\ X_{6[1:6]} \le X_{6[2:6]} \le X_{6[3:6]} \le X_{6[4:6]} \le X_{6[5:6]} \le \boldsymbol{X_{6[6:6]}} \end{bmatrix}.$$

Since the set size $n = 6$ is even, the actual measurement of units is made over the smallest ranked units ($X_{1[1:6]}$, $X_{2[1:6]}$, $X_{3[1:6]}$) from the first three sets and the largest ranked units ($X_{4[6:6]}$, $X_{5[6:6]}$, $X_{6[6:6]}$) from the last three sets, where $X_{i[m:n]}$ represents the m-th ranked unit in the i-th set for $i = 1, 2, ..., n$, $m = 1, 2, ..., n$. On the other hand, an example for odd set size, $n = 7$, is given below:

$$\begin{bmatrix} \boldsymbol{X_{1[1:7]}} \leq X_{1[2:7]} \leq X_{1[3:7]} \leq X_{1[4:7]} \leq X_{1[5:7]} \leq X_{1[6:7]} \leq X_{1[7:7]} \\ \boldsymbol{X_{2[1:7]}} \leq X_{2[2:7]} \leq X_{2[3:7]} \leq X_{2[4:7]} \leq X_{2[5:7]} \leq X_{2[6:7]} \leq X_{2[7:7]} \\ \boldsymbol{X_{3[1:7]}} \leq X_{3[2:7]} \leq X_{3[3:7]} \leq X_{3[4:7]} \leq X_{3[5:7]} \leq X_{3[6:7]} \leq X_{3[7:7]} \\ X_{4[1:7]} \leq X_{4[2:7]} \leq X_{4[3:7]} \leq X_{4[4:7]} \leq X_{4[5:7]} \leq X_{4[6:7]} \leq \boldsymbol{X_{4[7:7]}} \\ X_{5[1:7]} \leq X_{5[2:7]} \leq X_{5[3:7]} \leq X_{5[4:7]} \leq X_{5[5:7]} \leq X_{5[6:7]} \leq \boldsymbol{X_{5[7:7]}} \\ X_{6[1:7]} \leq X_{6[2:7]} \leq X_{6[3:7]} \leq X_{6[4:7]} \leq X_{6[5:7]} \leq X_{6[6:7]} \leq \boldsymbol{X_{6[7:7]}} \\ X_{7[1:7]} \leq X_{7[2:7]} \leq X_{7[3:7]} \leq \boldsymbol{X_{7[4:7]}} \leq X_{7[5:7]} \leq X_{7[6:7]} \leq X_{7[7:7]} \end{bmatrix}.$$

In this case, the actual measurement of units is made over the smallest ranked units ($X_{1[1:7]}$, $X_{2[1:7]}$, $X_{3[1:7]}$) from the first three sets and the largest ranked units ($X_{4[7:7]}$, $X_{5[7:7]}$, $X_{6[7:7]}$) from the following three sets. In addition, the fourth ranked unit ($X_{7[4:7]}$) is selected from the remaining set for the measurement where $X_{i[m:n]}$ represents the m-th ranked unit in the i-th set for $i = 1, 2, ..., n$, $m = 1, 2, ..., n$. For this case, the last unit corresponds to the median value of the last set in the sample.

For even set size, the mean estimator of ERSS is given by

$$\bar{X}_{\text{ERSS}} = \frac{1}{n}\left[\sum_{i=1}^{n/2} X_{2i-1[1:n]} + \sum_{i=1}^{n/2} X_{2i[n:n]}\right]. \tag{2.3}$$

Also, the variance of the mean estimator based on ERSS is given by

$$Var(\bar{X}_{\text{ERSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{n/2} Var(X_{2i-1[1:n]}) + \sum_{i=1}^{n/2} Var(X_{2i[n:n]})\right]. \tag{2.4}$$

For odd set size, the mean estimator of ERSS is given by

$$\bar{X}_{\text{ERSS}} = \frac{1}{n}\left[\sum_{i=1}^{(n-1)/2} X_{2i-1[1:n]} + \sum_{i=1}^{(n-1)/2} X_{2i[n:n]} + X_{n[(n-1/2):n]}\right]. \tag{2.5}$$

Also, the variance of the mean estimator based on ERSS is given by

$$Var(\bar{X}_{\text{ERSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{(n-1)/2} Var(X_{2i-1[1:n]}) + \sum_{i=1}^{(n-1)/2} Var(X_{2i[n:n]}) + Var(X_{n[((n+1)/2):n]})\right]. \tag{2.6}$$

2.3. Percentile Ranked Set Sampling

PRSS is suggested by Muttlak [13] to estimate the population parameters more efficiently than SRS with the same number of units by only using the $[p(n+1)]$-th and $[q(n+1)]$-th ranked units for n when it is even and, the median ranked unit when it is odd.

In this sampling method, p is denoted as the percentile value and takes value between 0 and 1, $(0 < p < 1)$. On the other hand, $q = 1 - p$ and $[p(n + 1)]$ and $[q(n + 1)]$ are rounded to the nearest integer. PRSS procedures are presented in the following examples.

Let the set size be $n = 6$, $p = 0.35$ and $q = 0.65$. Following the ranking process of units, the second ranked units from the first half of the sets ($X_{1[2:6]}$, $X_{2[2:6]}$, $X_{3[2:6]}$) and the fifth ranked units from the following three sets ($X_{4[5:6]}$, $X_{5[5:6]}$, $X_{6[5:6]}$) are selected:

$$\begin{bmatrix} X_{1[1:6]} \leq \boldsymbol{X_{1[2:6]}} \leq X_{1[3:6]} \leq X_{1[4:6]} \leq X_{1[5:6]} \leq X_{1[6:6]} \\ X_{2[1:6]} \leq \boldsymbol{X_{2[2:6]}} \leq X_{2[3:6]} \leq X_{2[4:6]} \leq X_{2[5:6]} \leq X_{2[6:6]} \\ X_{3[1:6]} \leq \boldsymbol{X_{3[2:6]}} \leq X_{3[3:6]} \leq X_{3[4:6]} \leq X_{3[5:6]} \leq X_{3[6:6]} \\ X_{4[1:6]} \leq X_{4[2:6]} \leq X_{4[3:6]} \leq X_{4[4:6]} \leq \boldsymbol{X_{4[5:6]}} \leq X_{4[6:6]} \\ X_{5[1:6]} \leq X_{5[2:6]} \leq X_{5[3:6]} \leq X_{5[4:6]} \leq \boldsymbol{X_{5[5:6]}} \leq X_{5[6:6]} \\ X_{6[1:6]} \leq X_{6[2:6]} \leq X_{6[3:6]} \leq X_{6[4:6]} \leq \boldsymbol{X_{6[5:6]}} \leq X_{6[6:6]} \end{bmatrix}.$$

This time, let $n = 7$, $p = 0.4$ and $q = 0.6$. Following the ranking process of units, the third ranked units from the first three sets ($X_{1[3:7]}$, $X_{2[3:7]}$, $X_{3[3:7]}$), the fifth ranked units from the following three sets ($X_{4[5:7]}$, $X_{5[5:7]}$, $X_{6[5:7]}$) and the median unit ($X_{7[4:7]}$) of the last set are selected:

$$\begin{bmatrix} X_{1[1:7]} \leq X_{1[2:7]} \leq \boldsymbol{X_{1[3:7]}} \leq X_{1[4:7]} \leq X_{1[5:7]} \leq X_{1[6:7]} \leq X_{1[7:7]} \\ X_{2[1:7]} \leq X_{2[2:7]} \leq \boldsymbol{X_{2[3:7]}} \leq X_{2[4:7]} \leq X_{2[5:7]} \leq X_{2[6:7]} \leq X_{2[7:7]} \\ X_{3[1:7]} \leq X_{3[2:7]} \leq \boldsymbol{X_{3[3:7]}} \leq X_{3[4:7]} \leq X_{3[5:7]} \leq X_{3[6:7]} \leq X_{3[7:7]} \\ X_{4[1:7]} \leq X_{4[2:7]} \leq X_{4[3:7]} \leq X_{4[4:7]} \leq \boldsymbol{X_{4[5:7]}} \leq X_{4[6:7]} \leq X_{4[7:7]} \\ X_{5[1:7]} \leq X_{5[2:7]} \leq X_{5[3:7]} \leq X_{5[4:7]} \leq \boldsymbol{X_{5[5:7]}} \leq X_{5[6:7]} \leq X_{5[7:7]} \\ X_{6[1:7]} \leq X_{6[2:7]} \leq X_{6[3:7]} \leq X_{6[4:7]} \leq \boldsymbol{X_{6[5:7]}} \leq X_{6[6:7]} \leq X_{6[7:7]} \\ X_{7[1:7]} \leq X_{7[2:7]} \leq X_{7[3:7]} \leq \boldsymbol{X_{7[4:7]}} \leq X_{7[5:7]} \leq X_{7[6:7]} \leq X_{7[7:7]} \end{bmatrix}.$$

For even set size, the mean estimator of PRSS is obtained as

$$\bar{X}_{\text{PRSS}} = \frac{1}{n}\left[\sum_{i=1}^{n/2} X_{i[a:n]} + \sum_{i=(n/2)+1}^{n} X_{i[b:n]}\right]. \tag{2.7}$$

Also, the variance of the mean estimator based on PRSS is obtained as

$$Var(\bar{X}_{\text{PRSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{n/2} Var(X_{i[a:n]}) + \sum_{i=(n/2)+1}^{n} Var(X_{i[b:n]})\right]. \tag{2.8}$$

For odd set size, the mean estimator of PRSS is obtained as

$$\bar{X}_{\text{PRSS}} = \frac{1}{n}\left[\sum_{i=1}^{(n-1/2)} X_{i[a:n]} + \sum_{i=((n-1)/2)+1}^{n-1} X_{i[b:n]} + X_{i[((n-1)/2):n]})\right]. \tag{2.9}$$

Also, the variance of the mean estimator based on PRSS is obtained as

$$Var(\bar{X}_{\text{PRSS}}) = \frac{1}{n^2}\left[\sum_{i=1}^{(n-1)/2} Var(X_{i[a:n]}) \sum_{i=((n-1)/2)+1}^{n-1} Var(X_{i[b:n]}) + Var(X_{i[(n+1/2):n]})\right], \tag{2.10}$$

where $a = [p(n+1)]$ and $b = [q(n+1)]$.

3. RANKING ERROR MODELS

3.1. Visual Ranked Set Sampling

Visual judgement ranking is firstly noted by McIntyre [12] to estimate the mean amount of products. This ranking method is a subjective ranking method since the ranking of units in the set is based on the personal judgement of the researcher. The reliability of visual ranking depends on the knowledge and experience of the researcher based on the subject of study and also on the materials used to rank the units.

Modelling the i-th visual score V_i was suggested by Dell and Clutter [6]. This model is given as follows:

$$V_i = X_i + \tau_i\,, \tag{3.1}$$

where

- V_i : i-th visual judgement order statistic,
- X_i : i-th true order statistic,
- τ_i : i-th random error term where $\tau_i \sim \text{iid}(0, \sigma_\tau^2)$ and X_i's are mutually independent of τ_i's.

In RSS, visual ranking process can be defined as follows:

1. Generate $V_i = X_i + \tau_i$ with $\tau_i \sim \text{iid}(0, \sigma_\tau^2)$ where X_i's and τ_i's are mutually independent.

2. Rank the visual scores ($V_1, V_2, ..., V_n$) from the lowest to the highest.

3. In the last step select the sampling unit corresponding to the r-th visual score (V_r) and measure the $X_{[r]}$ value for this unit.

This method is called Visual Ranked Set Sampling (VRSS). The correlation between visual judgement order statistic (V) and true order statistic (X) is computed by the following equation proposed by Nahhas *et al.* [14, 15]:

$$\rho_{xv} = \frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_\tau^2}}\,. \tag{3.2}$$

3.2. Concomitant Ranked Set Sampling

In RSS, another method used to rank the units in the set is concomitant variable (Y) based ranking which is suggested by Stokes [22]. The concomitant variable (Y) is a variable that has a high correlation with the variable of interest (X). The accuracy of the ranking is increased by using this variable. As an example, to estimate the mean weight of a certain number of fish belonging to a population, a researcher may use a concomitant variable, such as fish size, which has a high correlation with the fish weight.

David and Levine [5] were the first to study concomitant variable (Y). Detailed information and some limiting assumptions for concomitant variable (Y) were developed by Stokes [22] in order to determine its effects on RSS. These assumptions are given as follows:

- There is a linear relationship between concomitant variable (Y) and the variable of interest (X).
- Standardized concomitant variable (Y) and the standardized variable of interest (X) have identical distribution.

Concomitant based ranking can be modelled as

$$X_i = \mu_x + \frac{\rho_{xy}\sigma_x}{\sigma_y}(Y_i - \mu_y) + \tau_i\,, \tag{3.3}$$

where

μ_x : the mean of the variable of interest (X),
σ_x : the standard deviation of the variable of interest (X),
μ_y : the mean of the concomitant variable (Y),
σ_y : the standard deviation of the concomitant variable (Y),
ρ_{xy} : the correlation between the variable of interest (X) and concomitant variable (Y),
X_i : the i-th observation on the variable of interest (X),
Y_i : the i-th observation on the concomitant variable (Y),
τ_i : i-th random error term.

The random error term is independent identically distributed (iid) with mean 0 and variance σ_τ^2 and τ_i's and Y_i's are mutually independent. The stepwise period of ranking the units in the set with respect to the concomitant variable is given below:

1. Generate Equation (3.3) where τ_i's and Y_i's are mutually independent.

2. The Y_i's are ranked from the lowest to the highest to obtain the Y_i order statictics $Y_1 \leq \cdots \leq Y_n$.

3. Select the r-th correctly ranked order statistic Y_r and measure the r-th true order statistic $X = X_r$ from the sampling unit.

This method is defined as Concomitant Ranked Set Sampling (CRSS) method.

4. A MONTE CARLO SIMULATION STUDY

Our basic goal in this simulation study is to investigate the effects of ranking error models on the mean estimators based on RSS, ERSS and PRSS. For this reason, bias and MSE of the mean estimators are computed and compared with MSE of mean estimator based on SRS for different set and cycle sizes, distributions and ranking error models such as VRSS and CRSS in infinite population. The simulation study is performed via R Project with 10000 repetitions. In the simulation study:

- The population of the variable of interest (X) and concomitant variable (Y) are generated from $N(0,1)$ (symmetric), $Uniform(0,1)$ (symmetric), $Exp(1)$ (right skewed) and $Gamma(4,2)$ (right skewed) distributions with size (N) 10000.
- Set sizes (n) are determined to be 2, 3, 4 and 5. Also cycle sizes (m) are determined to be 5 and 10.
- Four sampling methods are used. These sampling methods are SRS, RSS, ERSS and PRSS. (In this study, p and q value for PRSS is determined as 0.4 and 0.6, respectively. $p = 0.4$ and $q = 0.6$ values were used in the simulation study since they offer the best results for PRSS.)
- For CRSS, the correlation values between the variable of interest (X) and concomitant variable (Y) ρ_{xy} are determined as 0.95, 0.75, 0.50 and 0.25. (The same values for ρ_{xv} and ρ_{xy} were used in the simulation study.)
- For VRSS, the random error term $\tau_i \sim N(0, \sigma_\tau^2)$. For the distributions used in the simulation study, the ρ_{xv} values corresponding to σ_τ^2 were calculated by Equation (3.2). These values are given in the table below.

Table 1: The values of σ_τ^2 corresponding to ρ_{xv} for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$.

ρ_{xv}	σ_τ^2			
	$N(0,1)$	$Uniform(0,1)$	$Exp(1)$	$Gamma(4,2)$
0.95	0.108	0.009	0.108	1.7285
0.75	0.778	0.0649	0.778	12.4444
0.50	3	0.25	3	48
0.25	15	1.25	15	240

The bias and mean squared error (MSE) of an estimator $\hat{\theta}$ of a parameter θ formulas given below are used in the simulation study:

$$Bias(\hat{\theta}) = \hat{\theta} - \theta, \tag{4.1}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2. \tag{4.2}$$

Note that, θ represents the population mean (μ) and $\hat{\theta}$ represents the mean estimators of population mean based on SRS ($\bar{X}_{\text{SRS}}$), RSS ($\bar{X}_{\text{RSS}}$), ERSS ($\bar{X}_{\text{ERSS}}$) and PRSS ($\bar{X}_{\text{PRSS}}$), respectively. The performance of the mean estimators of RSS, ERSS and PRSS is compared with respect to SRS in terms of relative efficiency criteria. The relative efficiency formulas given below are used:

$$RE_1(\bar{X}_{\text{RSS}}, \bar{X}_{\text{SRS}}) = \frac{MSE(\bar{X}_{\text{SRS}})}{MSE(\bar{X}_{\text{RSS}})}, \tag{4.3}$$

$$RE_2(\bar{X}_{\text{ERSS}}, \bar{X}_{\text{SRS}}) = \frac{MSE(\bar{X}_{\text{SRS}})}{MSE(\bar{X}_{\text{ERSS}})}, \tag{4.4}$$

$$RE_3(\bar{X}_{\text{PRSS}}, \bar{X}_{\text{SRS}}) = \frac{MSE(\bar{X}_{\text{SRS}})}{MSE(\bar{X}_{\text{PRSS}})}. \tag{4.5}$$

The comparisons of the mean estimators are constructed in terms of bias, mean squared error and relative efficiency for different correlation levels, variances of the random error term, set and cycle sizes. The results of the simulation study with 10000 repetitions are presented in tables.

Table 2 and Table 3 show bias values of mean estimators in VRSS and CRSS. The results indicate that:

- For symmetric distributions, the bias values obtained from mean estimators of RSS, ERSS and PRSS are close to 0. This means the mean estimators of RSS, ERSS and PRSS are unbiased estimators of population mean for symmetric distributions.
- For right skewed distributions, the bias values obtained from the mean estimator of RSS are close to symmetric distributions. On the other hand, for right skewed distributions, the bias values obtained from mean estimators of ERSS and PRSS are far from 0 when the set size increases. This means the mean estimators of ERSS and PRSS are biased estimators of population mean when the set size increases.

Table 2: Bias values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in VRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xv}=0.95$			$\rho_{xv}=0.75$			$\rho_{xv}=0.50$			$\rho_{xv}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	−0.001	−0.001	0.003	−0.001	0.000	−0.006	−0.003	0.002	0.000	0.002	−0.003	0.000
		3	0.003	−0.001	0.002	0.000	−0.003	−0.004	0.000	0.000	0.001	−0.003	0.000	0.000
		4	0.001	0.003	−0.001	0.001	−0.001	0.002	−0.001	−0.001	−0.001	0.002	0.002	−0.004
		5	0.000	0.000	0.001	0.000	0.001	−0.001	−0.004	−0.001	−0.001	−0.004	−0.002	0.002
	10	2	−0.003	0.000	0.003	−0.003	0.002	−0.001	0.004	0.001	−0.004	−0.002	0.001	0.001
		3	−0.002	−0.002	0.000	0.001	−0.001	0.001	0.000	−0.003	0.000	0.000	0.001	0.001
		4	0.000	0.000	0.002	0.000	−0.001	0.000	0.002	0.001	0.000	0.002	−0.002	0.000
		5	0.000	0.001	0.001	0.000	0.001	−0.001	0.000	−0.002	−0.001	0.001	0.001	0.000
$Uniform(0,1)$	5	2	−0.002	0.000	−0.001	−0.001	0.000	0.001	0.000	0.000	0.001	0.000	0.000	−0.001
		3	0.000	0.000	0.000	−0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.001
		4	0.000	0.000	−0.001	0.000	−0.001	−0.001	0.001	0.000	0.000	0.001	0.001	−0.001
		5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	10	2	0.000	0.000	0.000	−0.002	0.000	−0.001	0.001	0.000	0.001	0.000	0.000	−0.001
		3	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001
		4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	0.000
		5	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	−0.001	0.000	0.000	0.000	0.000
$Exp(1)$	5	2	−0.004	0.006	0.005	−0.002	−0.002	0.006	0.005	0.004	−0.001	−0.001	−0.003	0.004
		3	0.000	0.002	−0.178	−0.003	0.001	−0.149	−0.002	−0.003	−0.094	−0.002	−0.001	−0.033
		4	−0.002	0.179	−0.177	−0.001	0.150	−0.154	0.001	0.092	−0.092	−0.001	0.032	−0.028
		5	−0.004	0.159	−0.160	−0.001	0.136	−0.139	0.001	0.082	−0.087	0.002	0.029	−0.024
	10	2	0.003	0.001	−0.001	0.000	−0.002	0.003	0.001	0.002	0.002	0.000	−0.001	−0.001
		3	−0.001	−0.001	−0.176	0.000	−0.002	−0.152	−0.001	0.000	−0.091	0.001	0.000	−0.028
		4	−0.001	0.177	−0.178	0.000	0.150	−0.153	0.000	0.092	−0.091	0.000	0.032	−0.027
		5	0.000	0.158	−0.159	0.000	0.137	−0.137	0.000	0.084	−0.082	0.000	0.027	−0.026
$Gamma(4,2)$	5	2	0.001	0.005	−0.032	0.000	−0.025	0.004	−0.016	−0.010	0.013	−0.018	−0.012	0.005
		3	0.005	−0.003	−0.399	0.000	−0.002	−0.371	0.006	−0.005	−0.239	0.014	0.009	−0.064
		4	0.006	0.401	−0.392	0.005	0.359	−0.375	0.014	0.230	−0.205	0.001	0.062	−0.061
		5	−0.003	0.347	−0.350	0.005	0.320	−0.319	−0.012	0.205	−0.201	−0.003	0.055	−0.064
	10	2	0.009	−0.008	0.011	−0.009	−0.003	−0.002	0.002	−0.006	0.011	0.016	−0.008	−0.005
		3	0.009	0.002	−0.380	0.007	−0.001	−0.365	0.005	−0.002	−0.222	−0.008	−0.005	−0.060
		4	0.002	0.387	−0.390	0.005	0.354	−0.371	−0.006	0.227	−0.221	0.003	0.060	−0.064
		5	−0.003	0.350	−0.349	−0.001	0.329	−0.325	−0.002	0.186	−0.194	0.001	0.062	−0.053

Table 3: Bias values for $N(0,1)$, *Uniform*$(0,1)$, *Exp*(1) and *Gamma*$(4,2)$ in CRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xy}=0.95$			$\rho_{xy}=0.75$			$\rho_{xy}=0.50$			$\rho_{xy}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	0.001	0.003	0.001	0.004	−0.004	0.000	−0.006	−0.005	0.003	−0.001	0.001	0.002
		3	0.000	0.001	−0.006	0.001	−0.001	0.005	−0.004	−0.003	−0.002	−0.002	0.002	−0.002
		4	0.002	−0.003	−0.003	0.002	−0.001	0.000	0.000	0.001	0.005	0.001	0.001	0.004
		5	0.001	0.005	−0.002	−0.002	0.004	0.000	−0.001	0.000	−0.004	−0.002	0.005	−0.001
	10	2	0.000	−0.002	0.000	0.003	0.001	−0.001	−0.003	−0.002	0.001	0.000	0.002	−0.007
		3	0.001	0.001	0.003	0.003	0.001	−0.001	0.000	0.001	−0.005	−0.003	0.000	−0.004
		4	0.000	0.001	0.002	0.003	−0.001	−0.002	0.001	0.003	0.001	0.000	0.000	0.001
		5	0.001	0.003	0.005	0.000	0.003	0.004	0.000	0.008	−0.002	0.000	0.005	0.003
Uniform$(0,1)$	5	2	−0.002	0.000	0.000	0.001	0.000	0.000	−0.001	0.001	0.000	0.000	0.000	−0.001
		3	0.000	−0.001	0.001	−0.001	−0.001	0.001	0.001	0.000	0.000	0.000	0.000	−0.001
		4	0.000	0.001	−0.001	−0.001	−0.001	−0.001	0.000	−0.001	0.001	−0.001	0.000	0.000
		5	0.000	0.000	0.000	0.000	0.001	−0.001	0.001	0.000	−0.001	0.000	0.000	0.002
	10	2	−0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	0.000	−0.002	−0.001
		3	0.000	0.001	−0.001	0.000	0.000	0.001	0.000	0.000	0.000	−0.001	0.000	−0.002
		4	0.000	−0.001	0.000	0.000	0.000	0.002	0.000	−0.001	0.000	0.000	0.001	−0.001
		5	0.000	0.000	0.000	−0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000
Exp(1)	5	2	−0.002	0.001	0.001	0.001	0.001	0.004	−0.001	−0.001	−0.002	−0.002	0.000	−0.001
		3	0.002	0.003	−0.152	0.002	0.002	−0.094	0.001	0.002	−0.046	−0.003	0.002	0.001
		4	0.002	0.152	−0.149	−0.003	0.094	−0.099	0.001	0.045	−0.039	−0.004	0.006	0.001
		5	0.000	0.134	−0.135	0.000	0.086	−0.081	−0.001	0.029	−0.032	−0.001	0.012	0.002
	10	2	−0.003	−0.002	0.001	0.000	−0.001	0.003	−0.001	0.000	0.001	−0.002	0.003	0.000
		3	0.001	0.000	−0.148	−0.002	0.002	−0.096	0.000	0.002	−0.045	0.001	0.002	−0.002
		4	0.000	0.151	−0.153	0.002	0.045	−0.088	−0.001	0.040	−0.047	0.001	0.007	−0.002
		5	0.000	0.143	−0.133	0.000	0.029	−0.079	0.002	0.034	−0.038	−0.001	0.009	0.000
Gamma$(4,2)$	5	2	0.005	−0.004	−0.001	0.010	0.010	0.011	0.003	0.003	0.005	−0.018	0.002	0.016
		3	0.010	0.003	−0.324	0.003	0.008	−0.193	0.001	0.002	−0.110	0.007	0.005	−0.054
		4	0.004	0.320	−0.330	0.008	0.203	−0.235	0.007	0.104	−0.043	−0.003	−0.004	−0.023
		5	0.002	0.278	−0.285	0.006	0.180	−0.182	0.003	0.064	−0.068	0.015	0.010	−0.013
	10	2	−0.005	0.009	−0.003	−0.014	−0.007	0.005	−0.013	0.006	0.001	−0.004	0.009	−0.006
		3	0.002	−0.001	−0.346	−0.007	0.006	−0.205	0.008	−0.015	−0.095	−0.009	0.011	−0.014
		4	−0.004	0.335	−0.334	0.003	0.209	−0.199	−0.007	0.079	−0.098	0.005	0.014	−0.030
		5	0.004	0.275	−0.301	−0.005	0.195	−0.185	−0.006	0.086	−0.081	−0.004	0.022	−0.014

Table 4 and Table 5 show MSE values of the mean estimators in VRSS and CRSS. The results indicate that:

- Based on Table 4 and Table 5, the smallest and the highest MSE values were obtained from *Uniform*$(0,1)$ and *Gamma*$(4,2)$, respectively.
- MSE values obtained from $N(0,1)$ are less than *Exp*(1).

Table 4: MSE values for $N(0,1)$, *Uniform*$(0,1)$, *Exp*(1) and *Gamma*$(4,2)$ in VRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xv}=0.95$			$\rho_{xv}=0.75$			$\rho_{xv}=0.50$			$\rho_{xv}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	0.072	0.072	0.071	0.085	0.082	0.082	0.092	0.092	0.091	0.096	0.099	0.099
		3	0.038	0.038	0.033	0.048	0.049	0.046	0.058	0.059	0.057	0.065	0.064	0.066
		4	0.025	0.028	0.021	0.034	0.036	0.033	0.044	0.044	0.041	0.048	0.049	0.049
		5	0.017	0.019	0.015	0.026	0.027	0.024	0.033	0.034	0.033	0.039	0.039	0.039
	10	2	0.035	0.036	0.036	0.041	0.041	0.041	0.045	0.046	0.046	0.048	0.048	0.048
		3	0.019	0.019	0.017	0.024	0.024	0.023	0.029	0.030	0.028	0.032	0.033	0.032
		4	0.012	0.014	0.011	0.017	0.018	0.016	0.021	0.021	0.021	0.024	0.024	0.024
		5	0.008	0.010	0.007	0.013	0.013	0.012	0.017	0.017	0.016	0.019	0.019	0.019
Uniform$(0,1)$	5	2	0.006	0.006	0.006	0.007	0.007	0.007	0.008	0.008	0.008	0.008	0.008	0.008
		3	0.003	0.003	0.004	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005
		4	0.002	0.001	0.002	0.003	0.002	0.003	0.004	0.003	0.004	0.004	0.004	0.004
		5	0.001	0.001	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
	10	2	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.004
		3	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003
		4	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		5	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002
Exp(1)	5	2	0.077	0.077	0.079	0.084	0.085	0.087	0.095	0.095	0.092	0.096	0.096	0.100
		3	0.045	0.044	0.058	0.052	0.052	0.057	0.061	0.058	0.054	0.063	0.065	0.060
		4	0.029	0.071	0.049	0.037	0.072	0.048	0.044	0.063	0.042	0.048	0.055	0.045
		5	0.020	0.053	0.039	0.028	0.054	0.038	0.034	0.047	0.034	0.038	0.043	0.036
	10	2	0.040	0.039	0.039	0.043	0.043	0.044	0.046	0.047	0.046	0.049	0.050	0.048
		3	0.021	0.022	0.045	0.026	0.026	0.041	0.030	0.030	0.031	0.033	0.033	0.030
		4	0.014	0.052	0.041	0.018	0.047	0.036	0.022	0.035	0.025	0.024	0.027	0.023
		5	0.010	0.039	0.032	0.014	0.037	0.029	0.017	0.028	0.020	0.019	0.021	0.019
Gamma$(4,2)$	5	2	1.166	1.181	1.161	1.348	1.360	1.370	1.494	1.456	1.480	1.576	1.572	1.578
		3	0.629	0.622	0.661	0.802	0.803	0.834	0.942	0.952	0.911	1.066	1.035	1.010
		4	0.412	0.672	0.480	0.556	0.750	0.634	0.692	0.801	0.672	0.758	0.784	0.742
		5	0.291	0.460	0.350	0.416	0.579	0.463	0.544	0.618	0.536	0.606	0.624	0.605
	10	2	0.599	0.583	0.590	0.659	0.670	0.664	0.737	0.738	0.761	0.786	0.778	0.770
		3	0.321	0.319	0.403	0.401	0.402	0.481	0.479	0.475	0.477	0.511	0.515	0.503
		4	0.207	0.390	0.315	0.274	0.448	0.372	0.347	0.438	0.360	0.385	0.402	0.372
		5	0.145	0.299	0.238	0.213	0.344	0.291	0.269	0.324	0.279	0.308	0.326	0.302

Table 5: MSE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in CRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xy}=0.95$			$\rho_{xy}=0.75$			$\rho_{xy}=0.50$			$\rho_{xy}=0.25$		
			RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS	RSS	ERSS	PRSS
$N(0,1)$	5	2	0.073	0.071	0.072	0.081	0.082	0.081	0.091	0.090	0.091	0.101	0.096	0.097
		3	0.038	0.039	0.032	0.047	0.050	0.045	0.059	0.059	0.056	0.064	0.064	0.063
		4	0.024	0.026	0.021	0.034	0.036	0.032	0.043	0.042	0.041	0.050	0.048	0.047
		5	0.016	0.020	0.014	0.026	0.026	0.024	0.033	0.033	0.032	0.038	0.037	0.037
	10	2	0.035	0.036	0.032	0.040	0.039	0.040	0.045	0.046	0.045	0.050	0.048	0.049
		3	0.018	0.018	0.016	0.024	0.023	0.022	0.029	0.028	0.028	0.032	0.032	0.030
		4	0.011	0.013	0.010	0.016	0.018	0.015	0.021	0.022	0.020	0.024	0.024	0.023
		5	0.008	0.009	0.007	0.013	0.013	0.011	0.016	0.017	0.016	0.018	0.018	0.018
$Uniform(0,1)$	5	2	0.006	0.006	0.006	0.007	0.007	0.007	0.008	0.007	0.008	0.008	0.008	0.008
		3	0.003	0.003	0.004	0.004	0.004	0.004	0.005	0.005	0.005	0.006	0.005	0.005
		4	0.002	0.002	0.002	0.003	0.003	0.003	0.004	0.003	0.004	0.004	0.004	0.004
		5	0.001	0.001	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
	10	2	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		3	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
		4	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		5	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002
$Exp(1)$	5	2	0.074	0.076	0.075	0.088	0.085	0.089	0.092	0.087	0.088	0.091	0.101	0.007
		3	0.044	0.041	0.049	0.052	0.052	0.051	0.065	0.061	0.054	0.065	0.066	0.005
		4	0.028	0.061	0.040	0.035	0.054	0.038	0.044	0.051	0.042	0.004	0.004	0.003
		5	0.019	0.045	0.031	0.027	0.041	0.029	0.034	0.036	0.033	0.003	0.003	0.003
	10	2	0.037	0.038	0.039	0.042	0.041	0.042	0.045	0.048	0.047	0.004	0.004	0.004
		3	0.021	0.021	0.035	0.024	0.025	0.030	0.030	0.030	0.028	0.002	0.002	0.002
		4	0.013	0.042	0.032	0.018	0.030	0.021	0.022	0.024	0.021	0.002	0.002	0.002
		5	0.010	0.035	0.024	0.013	0.025	0.017	0.018	0.021	0.017	0.001	0.001	0.001
$Gamma(4,2)$	5	2	1.173	1.138	1.165	1.363	1.317	1.313	1.453	1.490	1.415	1.579	1.564	1.624
		3	0.638	0.630	0.618	0.804	0.775	0.768	0.940	0.950	0.901	1.037	1.062	1.054
		4	0.413	0.587	0.418	0.555	0.660	0.549	0.700	0.709	0.677	0.776	0.742	0.737
		5	0.290	0.408	0.312	0.423	0.483	0.425	0.525	0.562	0.543	0.589	0.586	0.617
	10	2	0.590	0.571	0.605	0.681	0.633	0.667	0.719	0.724	0.751	0.788	0.782	0.796
		3	0.304	0.308	0.375	0.396	0.399	0.400	0.468	0.478	0.465	0.508	0.510	0.498
		4	0.203	0.363	0.277	0.277	0.343	0.292	0.343	0.365	0.330	0.391	0.395	0.363
		5	0.145	0.246	0.210	0.209	0.270	0.226	0.276	0.287	0.277	0.302	0.316	0.308

Table 6 and Table 7 show RE values of mean estimators in VRSS and CRSS. RE values obtained from the simulation study which are greater than 1 mean that RSS, ERSS or PRSS are more efficient than SRS:

- RE values obtained from symmetric distributions give better results than right skewed distributions.

Table 6: RE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in VRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xv}=0.95$			$\rho_{xv}=0.75$			$\rho_{xv}=0.50$			$\rho_{xv}=0.25$		
			RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3
$N(0,1)$	5	2	1.422	1.387	1.412	1.146	1.223	1.236	1.077	1.101	1.119	1.044	1.010	1.020
		3	1.746	1.754	1.975	1.356	1.398	1.464	1.165	1.146	1.178	1.036	1.026	1.006
		4	1.989	1.819	2.381	1.453	1.401	1.484	1.118	1.131	1.194	1.027	1.036	1.011
		5	2.323	2.028	2.646	1.587	1.461	1.626	1.214	1.195	1.197	1.006	1.020	1.009
	10	2	1.417	1.386	1.389	1.212	1.185	1.225	1.153	1.079	1.076	1.057	1.033	1.012
		3	1.750	1.764	1.940	1.377	1.385	1.453	1.137	1.109	1.228	1.024	1.005	1.020
		4	2.086	1.881	2.361	1.448	1.407	1.622	1.162	1.191	1.173	1.062	1.052	1.038
		5	2.379	2.111	2.673	1.509	1.471	1.589	1.184	1.152	1.218	1.064	1.028	1.011
$Uniform(0,1)$	5	2	1.417	1.394	1.455	1.257	1.228	1.232	1.099	1.104	1.101	1.009	1.027	1.032
		3	1.830	1.879	1.521	1.458	1.429	1.193	1.140	1.160	1.081	1.050	1.044	1.029
		4	2.181	2.928	1.852	1.524	1.876	1.314	1.164	1.303	1.105	1.055	1.052	0.990
		5	2.511	3.119	2.112	1.636	1.922	1.419	1.246	1.325	1.134	1.067	1.036	1.037
	10	2	1.477	1.447	1.411	1.242	1.224	1.240	1.081	1.060	1.062	1.016	0.992	1.031
		3	1.857	1.785	1.492	1.426	1.442	1.157	1.128	1.176	1.072	1.072	1.015	1.013
		4	2.218	2.801	1.755	1.530	1.879	1.318	1.162	1.257	1.045	1.046	1.085	1.050
		5	2.474	3.067	2.052	1.670	2.000	1.417	1.187	1.275	1.152	1.044	1.007	1.009
$Exp(1)$	5	2	1.296	1.229	1.278	1.188	1.190	1.168	1.031	1.074	1.121	1.071	1.017	1.005
		3	1.539	1.550	1.140	1.267	1.306	1.203	1.111	1.136	1.298	1.066	1.008	1.126
		4	1.692	0.701	1.022	1.384	0.696	1.044	1.134	0.793	1.193	1.060	0.910	1.159
		5	1.953	0.778	1.066	1.422	0.757	1.068	1.158	0.839	1.193	1.036	0.946	1.114
	10	2	1.266	1.272	1.270	1.174	1.135	1.188	1.068	1.076	1.101	1.022	1.028	1.058
		3	1.596	1.594	0.737	1.221	1.279	0.811	1.119	1.106	1.081	1.019	1.001	1.088
		4	1.780	0.479	0.621	1.370	0.537	0.688	1.142	0.735	0.994	1.010	0.917	1.078
		5	1.968	0.515	0.625	1.430	0.548	0.694	1.228	0.724	0.994	1.027	0.962	1.088
$Gamma(4,2)$	5	2	1.391	1.344	1.371	1.186	1.179	1.152	1.073	1.107	1.067	1.046	1.038	1.006
		3	1.718	1.718	1.595	1.307	1.342	1.306	1.141	1.098	1.198	0.984	1.016	1.066
		4	1.953	1.188	1.624	1.469	1.075	1.255	1.154	1.004	1.208	1.036	1.036	1.091
		5	2.211	1.420	1.840	1.555	1.094	1.365	1.179	1.035	1.194	1.048	1.039	1.096
	10	2	1.344	1.384	1.355	1.191	1.186	1.222	1.119	1.065	1.043	1.046	1.020	1.076
		3	1.630	1.666	1.339	1.352	1.342	1.084	1.126	1.110	1.102	1.041	1.013	1.057
		4	1.923	1.019	1.273	1.456	0.884	1.078	1.137	0.930	1.105	1.004	0.999	1.067
		5	2.221	1.071	1.328	1.486	0.942	1.097	1.207	1.012	1.129	1.026	0.991	1.044

Table 7: RE values for $N(0,1)$, $Uniform(0,1)$, $Exp(1)$ and $Gamma(4,2)$ in CRSS based on RSS, ERSS and PRSS.

Distribution	m	n	$\rho_{xy}=0.95$			$\rho_{xy}=0.75$			$\rho_{xy}=0.50$			$\rho_{xy}=0.25$		
			RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3	RE_1	RE_2	RE_3
$N(0,1)$	5	2	1.351	1.398	1.361	1.188	1.236	1.180	1.115	1.060	1.072	1.024	1.020	0.993
		3	1.739	1.728	2.028	1.401	1.360	1.449	1.118	1.147	1.144	1.061	0.998	1.044
		4	2.050	1.890	2.302	1.452	1.388	1.511	1.166	1.185	1.220	1.048	1.043	1.041
		5	2.395	2.074	2.676	1.509	1.562	1.620	1.194	1.154	1.206	1.040	1.033	1.064
	10	2	1.405	1.402	1.511	1.213	1.222	1.224	1.082	1.062	1.107	0.999	1.043	1.038
		3	1.740	1.726	1.962	1.345	1.382	1.430	1.147	1.149	1.152	1.043	1.019	1.061
		4	2.127	1.882	2.372	1.558	1.317	1.588	1.188	1.138	1.223	1.029	0.995	1.041
		5	2.349	2.062	2.735	1.559	1.529	1.686	1.182	1.137	1.216	1.065	1.068	1.041
$Uniform(0,1)$	5	2	1.474	1.408	1.442	1.193	1.251	1.208	1.079	1.106	1.093	1.022	1.000	1.047
		3	1.804	1.789	1.509	1.388	1.386	1.233	1.116	1.163	1.085	0.994	1.035	1.022
		4	2.202	2.670	1.844	1.497	1.648	1.340	1.186	1.206	1.101	1.043	1.025	1.029
		5	2.470	2.921	2.153	1.559	1.698	1.374	1.168	1.254	1.120	1.053	1.056	1.015
	10	2	1.447	1.414	1.416	1.214	1.197	1.225	1.097	1.094	1.102	1.008	1.022	1.022
		3	1.752	1.866	1.545	1.393	1.369	1.275	1.108	1.092	1.104	1.055	1.021	0.992
		4	2.177	2.755	1.802	1.417	1.601	1.308	1.132	1.214	1.117	1.031	1.044	1.067
		5	2.513	2.923	2.182	1.569	1.693	1.415	1.208	1.199	1.158	1.041	1.022	0.995
$Exp(1)$	5	2	1.323	1.284	1.297	1.167	1.186	1.182	1.099	1.077	1.102	1.017	1.010	1.040
		3	1.526	1.501	1.347	1.288	1.293	1.387	1.018	1.110	1.221	1.034	1.027	1.009
		4	1.705	0.802	1.286	1.362	0.944	1.312	1.101	1.009	1.197	1.028	1.011	1.017
		5	1.912	0.882	1.271	1.381	0.986	1.324	1.112	1.079	1.182	1.021	0.991	1.050
	10	2	1.384	1.281	1.232	1.165	1.177	1.145	1.097	1.064	1.040	1.023	0.998	0.998
		3	1.539	1.551	0.912	1.321	1.224	1.077	1.097	1.102	1.138	0.997	1.070	1.042
		4	1.801	0.586	0.754	1.313	0.810	1.150	1.166	0.970	1.087	1.039	1.028	0.996
		5	1.939	0.614	0.818	1.480	0.826	1.077	1.143	0.979	1.162	1.054	0.961	1.023
$Gamma(4,2)$	5	2	1.370	1.364	1.374	1.180	1.198	1.231	1.121	1.063	1.053	1.022	0.999	0.967
		3	1.681	1.688	1.773	1.300	1.357	1.371	1.122	1.120	1.147	1.024	1.007	1.053
		4	1.926	1.373	1.907	1.427	1.259	1.471	1.180	1.097	1.116	1.018	1.075	1.052
		5	2.201	1.530	2.040	1.481	1.284	1.505	1.204	1.138	1.175	1.043	1.062	1.068
	10	2	1.336	1.366	1.338	1.190	1.227	1.198	1.086	1.089	1.066	1.041	1.017	0.988
		3	1.741	1.729	1.442	1.358	1.371	1.330	1.103	1.145	1.146	1.067	1.047	1.032
		4	1.989	1.081	1.510	1.443	1.171	1.356	1.143	1.097	1.220	1.001	1.020	1.087
		5	2.165	1.260	1.573	1.532	1.204	1.380	1.111	1.104	1.199	1.047	1.000	1.016

According to the results obtained from the simulation study:

For VRSS:

- When the number of set size increases, relative efficiency increases.
- When the variance of random error term (σ_τ^2) increases, the relative efficiency decreases. On the other hand, when the correlation between visual judgement order statistic (V) and true order statistic (X) decreases, the relative efficiency decreases.

For CRSS:

- When the number of set size increases, relative efficiency increases.
- When the correlation between the variable of interest (X) and the concomitant variable (Y) increases, relative efficiency increases.

For VRSS and CRSS:

- The number of cycles didn't cause a regular increase or decrease in relative efficiency for VRSS and CRSS. For this reason, exact comment can not be made about the effect of number of cycles on relative efficiency.
- In both visual and concomitant based ranking methods MSE decreases when set size and number of cycles increase.
- MSE increases as the variance of the error term increases in visual ranking and as the correlation between the concomitant variable (Y) and the variable of interest (X) variable decreases in concomitant based ranking.
- In both visual and concomitant based ranking methods MSE values obtained from right skewed distributions are greater than the MSE values obtained from symmetric distributions.
- In both visual and concomitant based ranking methods, the bias, MSE and RE values from mean estimators based on RSS, ERSS and PRSS for symmetric distributions and right skewed distributions are similar.

5. REAL DATA APPLICATION

Abalone is a common name given to a group of small to very large sea snails, marine gastropod molluscs which are the member of Haliotidae family [8]. Age of an abalone can be determined by making some physical measurements which, in advance, include cutting and staining of the shell. After the staining process, the rings become clear and they are counted under a microscope to obtain age information. Estimating the age of abalone includes difficult, costly and time-consuming physical measurements. Therefore, it forces us to use alternative measurement techniques. A new physical measurement method which is easier than the others in estimating the age of abalone is proposed by Nash *et al.* [16]. This data set is taken from `https://archive.ics.uci.edu/ml/datasets/abalone` [25]. Abalone dataset includes 4177 samples with 9 variables. Information about these variables are given in the table below:

Table 8: Descriptions of abalone dataset.

Variable	Data Type	Measurement Unit of Data	Description
Length	Continuous	mm	Longest shell measurement
Diameter	Continuous	mm	Perpendicular to length
Height	Continuous	mm	With meat in shell
Whole weight	Continuous	gr	Whole abalone
Shucked weight	Continuous	gr	Weight of meat
Viscera weight	Continuous	gr	Gut weight (after bleeding)
Shell weight	Continuous	gr	After being died
Rings	Integer	—	+1.5 gives the age in years
Sex	Nominal	—	Male, Female and Infant

Rings variable is selected as the variable of interest (X). For concomitant based ranking, Shell weight (Y_1) and Shucked weight (Y_2) are determined as concomitant variables. The correlations between variable of interest and concomitant variables are given in table below:

Table 9: Correlations between variable of Interest (X) and concomitant variables (Y's) in abalone dataset for CRSS.

Variable of Interest (X)	Concomitant Variable (Y)	Correlations
Rings	Shell Weight	0.627
	Shucked Weight	0.420

The results obtained from abalone dataset using CRSS are given in Table 10 and Table 11, respectively.

Table 10: MSE(bias) values for CRSS based on RSS, ERSS and PRSS.

ρ_{xy}	m	n	RSS	ERSS	PRSS
0.627	5	2	0.891 (−0.029)	0.887 (−0.023)	0.909 (−0.013)
		3	0.558 (−0.019)	0.554 (−0.000)	0.543 (−0.113)
		4	0.395 (−0.004)	0.433 (0.127)	0.383 (−0.119)
		5	0.306 (−0.009)	0.329 (−0.103)	0.295 (0.099)
	10	2	0.443 (−0.002)	0.449 (0.003)	0.453 (−0.015)
		3	0.279 (0.001)	0.278 (−0.001)	0.273 (0.120)
		4	0.201 (0.001)	0.219 (0.119)	0.202 (0.116)
		5	0.152 (−0.002)	0.173 (−0.107)	0.152 (0.093)
0.420	5	2	0.979 (0.064)	0.951 (0.009)	0.981 (−0.006)
		3	0.610 (−0.008)	0.624 (−0.010)	0.751 (0.320)
		4	0.440 (−0.001)	0.518 (−0.314)	0.571 (0.309)
		5	0.346 (−0.004)	0.403 (−0.271)	0.450 (0.269)
	10	2	0.485 (−0.000)	0.467 (−0.013)	0.479 (0.003)
		3	0.305 (−0.008)	0.306 (−0.004)	0.429 (0.329)
		4	0.236 (0.003)	0.313 (0.324)	0.342 (0.319)
		5	0.176 (−0.040)	0.242 (−0.276)	0.258 (0.270)

Table 11: RE values for CRSS based on RSS, ERSS and PRSS.

ρ_{xy}	m	n	RE_1	RE_2	RE_3
0.627	5	2	1.181	1.153	1.180
		3	1.215	1.266	1.280
		4	1.298	1.192	1.384
		5	1.358	1.268	1.425
	10	2	1.203	1.150	1.118
		3	1.245	1.242	1.278
		4	1.269	1.163	1.268
		5	1.334	1.178	1.320
0.420	5	2	1.062	1.081	1.058
		3	1.135	1.086	0.900
		4	1.158	1.006	0.889
		5	1.193	1.020	0.885
	10	2	1.050	1.119	1.063
		3	1.104	1.133	0.800
		4	1.095	0.810	0.756
		5	1.138	0.855	0.822

Suppose that the ρ_{xv} values are 0.627 and 0.420, respectively. For VRSS, we need to find the value of standard deviation of the Rings variable (X). This value is $\sqrt{\sigma_x^2} = \sigma_x = \sqrt{10.395} = 3.224$. Then, we need to find the values of σ_τ^2 corresponding to ρ_{xv}. We use Equation (3.2) to obtain the values of σ_τ^2 corresponding to ρ_{xv}. These values are given in the table below:

Table 12: The values of σ_τ^2 corresponding to ρ_{xv} for Rings variable in abalone dataset.

ρ_{xv}	σ_τ^2
0.627	16.048
0.420	48.536

Table 13: MSE(bias) values for VRSS based on RSS, ERSS and PRSS.

ρ_{xv}	m	n	RSS	ERSS	PRSS
0.627	5	2	0.921 (−0.008)	0.939 (0.001)	0.934 (−0.004)
		3	0.567 (0.011)	0.573 (−0.021)	0.537 (−0.256)
		4	0.415 (0.002)	0.560 (0.247)	0.398 (−0.247)
		5	0.321 (0.005)	0.421 (0.219)	0.318 (−0.231)
	10	2	0.463 (−0.009)	0.456 (−0.001)	0.450 (−0.001)
		3	0.287 (0.002)	0.291 (−0.004)	0.299 (0.250)
		4	0.205 (−0.001)	0.305 (0.247)	0.224 (−0.251)
		5	0.156 (−0.004)	0.235 (0.224)	0.188 (−0.231)
0.420	5	2	0.973 (−0.002)	0.986 (−0.005)	0.986 (−0.016)
		3	0.635 (−0.003)	0.639 (0.004)	0.583 (−0.152)
		4	0.480 (0.010)	0.536 (0.142)	0.441 (−0.143)
		5	0.375 (0.005)	0.423 (0.131)	0.350 (−0.122)
	10	2	0.511 (0.005)	0.496 (−0.017)	0.481 (0.007)
		3	0.328 (0.006)	0.325 (0.002)	0.310 (−0.140)
		4	0.234 (0.003)	0.279 (0.145)	0.231 (−0.138)
		5	0.179 (−0.002)	0.216 (0.127)	0.181 (−0.124)

Table 14: RE values for VRSS based on RSS, ERSS and PRSS.

ρ_{xv}	m	n	RE_1	RE_2	RE_3
0.627	5	2	1.128	1.093	1.093
		3	1.247	1.293	1.293
		4	1.286	0.954	1.312
		5	1.293	0.994	1.277
	10	2	1.108	1.144	1.178
		3	1.176	1.204	1.119
		4	1.259	0.842	1.128
		5	1.291	0.865	1.095
0.420	5	2	1.067	1.044	1.056
		3	1.067	1.088	1.201
		4	1.059	0.952	1.183
		5	1.104	0.968	1.224
	10	2	0.994	1.010	1.106
		3	1.036	1.074	1.114
		4	1.098	0.914	1.122
		5	1.152	0.971	1.145

6. CONCLUSION

In this study, we aimed to use ranking error models (VRSS and CRSS) to compare the bias and MSE of the mean estimators based on RSS and some of its modified methods such as ERSS and PRSS.

For this reason the effects of ranking errors in RSS and in some of its modified methods are examined in the simulation study. In this study, it is deduced that ranking errors may occur depending on the ranking method used. In VRSS, σ_τ^2 and ρ_{xv} change depending on the researcher's knowledge, experience and materials used in the study. The greater knowledge of researcher involved in the study and the use of more appropriate materials would yield a higher accuracy in the ranking. On the other hand, for CRSS, the accuracy of the ranking depends on the correlation between the variable of interest (X) and the concomitant variable (Y) and the distribution of (X, Y). Generally, when $\rho_{xy} \geq 0.5$, the error in the ranking decreases and the accuracy of the ranking increases. Thus, better results can be achieved by minimizing the error in the ranking. The application is performed using abalone data set in order to support the simulation study performed in the section 5. It is seen that similar results were obtained in real data application and simulation study. It is observed that, RSS and some of its modified methods such as ERSS and PRSS methods show better results than SRS.

REFERENCES

[1] Al-Omari, A.I. and Bouza, C.N. (2014). Review of ranked set sampling: modifications and applications, *Investigacion Operacional*, **35**(3), 215–241.

[2] Alexandridis, R. and Ozturk, O. (2009). Robust inference against imperfect ranking in ranked set sampling, *Communications in Statistics Theory and Methods*, **38**(12), 2126–2143.

[3] Bohn, L.L. and Wolfe, D.A. (2014). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the mannwhitney-wilcoxon statistic, *Journal of the American Statistical Association*, **89**(425), 168–176.

[4] Chen, Z.; Bai, Z. and Sinha, B. (2003). *Ranked Set Sampling: Theory and Applications*, Springer Science & Business Media, New York.

[5] David, H. and Levine, D. (1972). Ranked set sampling in the presence of judgment error, *Biometrics*, **28**, 553–555.

[6] Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics background, *Biometrics*, 545–555.

[7] Evans, M. (1967). Application of ranked set sampling to regeneration surveys in areas direct-seeded to longleaf pine, *Master of Forestry Dissertation*, Louisiana State University, Baton Rouge, LA.

[8] Fleming C.A. (1952). Notes on the genus Haliotis (Mollusca). A new subgenus from New Zealand and a new species from the late Cenozoic of Ohope, Bay of Plenty, *Transactions of the Royal Society of New Zealand*, **101**, 229–232.

[9] Fligner, M.A. and MacEachern, S.N. (2006). Nonparametric two-sample methods for ranked-set sample data, *Journal of the American Statistical Association*, **101**(475), 1107–1118.

[10] Frey, J.C. (2007). New imperfect rankings models for ranked set sampling, *Journal of Statistical planning and Inference*, **37**(4), 1433–1445.

[11] Halls, L.K. and Dell, T.R. (1966). Trial of ranked-set sampling for forage yields, *Forest Science*, **12**(1), 22–26.

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

[13] Muttlak, H.A. (2003). Modified ranked set sampling methods, *Pakistan Journal of Statistics – All Series*, **19**(3), 315–324.

[14] Nahhas, R.W.; Wolfe, D.A. and Chen, H. (2002). Ranked set sampling: cost and optimal set size, *Biometrics*, **58**(4), 964–971.

[15] Nahhas, R.W.; Wolfe, D.A. and Chen, H. (2004). Ranked set sampling: ranking error models and estimation of visual judgment error variance, *Biometrical Journal*, **46**(2), 255–263.

[16] Nash, W.J.; Sellers, T.L.; Talbot, S.R.; Cawthorn, A.J. and Ford, W.B. (1994). The population biology of abalone. Technical report. Sea Fisheries Division, Australia, *Biometrics*, **48**, 1–69.

[17] Ozturk, O. (2007). Statistical inference under a stochastic ordering constraint in ranked set sampling, *Nonparametric Statistics*, **19**(3), 131–144.

[18] Ozturk, O. (2008). Inference in the presence of ranking error in ranked set sampling, *Canadian Journal of Statistics*, **36**(4), 577–594.

[19] Ozturk, O. (2010). Nonparametric maximum-likelihood estimation of within-set ranking errors in ranked set sampling, *Journal of Nonparametric Statistics*, **22**(7), 823–840.

[20] Patil, G.; Sinha, A. and Taillie, C. (1999). Ranked set sampling: a bibliography, *Biometrical Journal*, **38**(5), 577–586.

[21] Samawi, H.M.; Ahmed, M.S. and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling, *Environmental and Ecological Statistics*, **6**(1), 91–98.

[22] Stokes, S. (1977). Ranked set sampling with concomitant variables, *Communications in Statistics – Theory and Methods*, **6**(12), 1207–1211.

[23] Stokes, S. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, 35–42.

[24] Takahasi, K. and 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.

[25] The abalone dataset. Retrieved June 1, 2019, available from: `https://archive.ics.uci.edu/ml/datasets/abalone`