Submitted to the Brazilian Journal of Probability and Statistics
arXiv: arXiv:0000.0000

Extreme-Cum-Median Ranked Set Sampling

Shakeel Ahmed[1] **and Javid Shabbir**[2]

Department of Statistics Quaid-i-Azam University Islamabad Pakistan[1 2]

Abstract. A mixture of Extreme Ranked Set Sampling (ERSS) and Median Ranked Set Sampling (MRSS) is introduced to obtain a more representative sample using three out of five number summary statistics [i.e. Minimum, Median and Maximum]. The proposed sampling scheme provides unbiased estimator of mean for symmetric population and gives moderate efficiency for both symmetric and asymmetric populations under perfect as well as imperfect rankings. Expressions for bias and asymptotic variance are presented. A simulation study is also conducted to observe the performance of the proposed estimator. Application of proposed sampling scheme is illustrated through a real life example.

Key words Median ranked set sampling, Extreme ranked set sampling, Ranking error, Imperfect ranking.

1 Introduction

Ranked Set Sampling (RSS), introduced by [**?**] for estimating the yield of herbage in agriculture researches, involves randomly selecting m sets of units, each of m units, from an infinite population of interest. The units in each set can assumed to be ranked visually or through some concomitant variable. From the first set of m units the lowest ranked unit is selected. From the second ranked set of m units second lowest unit is selected. The process is continued until the mth ranked unit is measured from the mth ranked set. This procedure is repeated r times to obtain $n = mr$ units.
[**?**] introduced a very essential statistical basis for the theory of RSS introduced by [**?**]. They showed that the sample mean under RSS is an unbiased estimator of the population mean and more efficient than the sample mean under simple random sampling with replacement (SRSWR) for same number of units. [**?**] proved that mean of the RSS is an unbiased estimator of the population mean, whether there are errors in ranking or not and more efficient than the mean of SRSWR. [**?**] showed that the variance estimator based on RSS is an asymptotically unbiased and more efficient than the usual variance estimator based on SRS, when the sample size is large. [**?**] analyzed the empirical distribution function based on RSS and proved that

it is an unbiased estimate of the underlying distribution function. [?] and [?] worked on error in ranking with application of RSS to real life situations. [?] suggested ERSS on basis of the fact that for sets containing a larger number of units, visual recognition of the ith ranked unit (other than the smallest and the largest) is difficult and expensive. [?] introduced the MRSS procedure to obtain a better representative units. After that [?] suggested quartile ranked set sampling (QRSS) for this purpose. Similarly, [?], [?] and [?] have written detail notes on ranked set sampling. [?] suggested ranked set sample by selecting extremes of the samples and called this sampling scheme as RSS(E). Later on, [?] have made a modification in RSS namely Independent extreme ranked set sampling (IERSS) in the case of both even as well as odd sample sizes. The main objective was to propose non-parametric estimators of the population mean using these two extremes RSS and are compared with mean estimator under SRSWR.
Fat-tailed distributions as the Cauchy distribution, students t-distribution, uniform distribution (and such other stable distributions with the exception of the normal distribution) units that deviate from the mean by five or more standard deviations (beyond 5-sigma limit) have higher frequency as appose to normal distribution where extreme units are less likely to occur. As a consequence, application of MRSS and ERSS for sampling from such distributions provides a non representative sample. For example, in estimating the ratio of width to length of petal and sepal of certain flowers (Iris flower data is used in Section 5) or a ratio of width to length of leaves on plants in a certain garden in agriculture experiments, use of ERSS or MRSS may provide a bad representation of population as these variables follow Cauchy distribution (Cauchy random variable is the ratio of two normal random variables). Similarly, many other phenomena in biological science and agriculture setting have fat tail distribution. In such situations application of these two sampling schemes simultaneously may provide more representative sample than their individual application.
To overcome the above stated deficiency in ERSS and MRSS, we propose an estimator of finite population mean by using a new ranked set sampling procedure obtained by combining ERSS and MRSS. The expressions for bias and variances under assumption of symmetric distribution are derived. We conduct a simulation study for purpose of comparison of EMRSS with ERSS and MRSS. We also give a numerical example before concluding the work.

2 MRSS and ERSS based on $2m$ units

Consider an infinite population consisting variable of interest Y with distribution function $F(y)$ and probability density function $f(y)$ having mean μ and variance σ^2. Let $Y_{i(i)j}$ be the i-th unit from ith sample in jth cycle on the study variable Y. Then a ranked set sample of size $2m$ in jth $(j = 1, 2, ..., r)$ cycle is $\{Y_{1(1)j}, Y_{2(2)j}, ..., Y_{2m(2m)j}\}$.
The sample mean under RSS is given by

$$\bar{y}_{rss} = \frac{\sum_{j=1}^{r} \left(\sum_{i=1}^{2m} Y_{i(i)j} \right)}{2mr}. \tag{2.1}$$

It is easy to prove that $\bar{y}_{rss}$ is an unbiased estimator with variance

$$V(\bar{y}_{rss}) = \frac{1}{2mr} \left[\sigma^2 - \frac{1}{m} \sum_{i=1}^{2m} \wp_{(i)}^2 \right]. \tag{2.2}$$

Further $\{Y_{1(1)j}, Y_{2(1)j}, ..Y_{m(1)j}; Y_{m+1(2m)j}, ..., Y_{2m(2m)j}\}$ and $\{Y_{1(m)j}, Y_{2(m)j}, ..Y_{m(m)j}; Y_{m+1(m+1)j}, ..., Y_{2m(m+1)j}\}$ be an ERSS and MRSS of size $2m$ in cycle j respectively. The mean estimators under ERSS and MRSS are given by

$$\bar{y}_e = \frac{\sum_{j=1}^{r} \left(\sum_{i=1}^{m} Y_{i(1)j} + \sum_{i=m+1}^{2m} Y_{i(2m)j} \right)}{2mr}, \tag{2.3}$$

and

$$\bar{y}_m = \frac{\sum_{j=1}^{r} \left(\sum_{i=1}^{m} Y_{i(m)j} + \sum_{i=m+1}^{2m} Y_{i(m+1)j} \right)}{2mr} \tag{2.4}$$

with biases

$$E(\bar{y}_e - \mu) = \frac{\wp_{(1)} + \wp_{(2m)}}{2} \quad \text{and} \quad E(\bar{y}_m - \mu) = \frac{\wp_{(m)} + \wp_{(m+1)}}{2},$$

where

$$\wp_{(1)} = \mu_{(i)} - \mu, \quad \forall \quad i = 1, 2, ..., 2m.$$

It is straight forward to show that $\bar{y}_e$ and $\bar{y}_m$ are unbiased when the underlying distribution is symmetric about mean i.e. $\wp_{(2m)} = -\wp_{(1)}$ and $\wp_{(m+1)} = -\wp_{(m)}$. Assuming symmetric distribution we obtain variance of $\bar{y}_e$ and $\bar{y}_m$ as follows:

$$V(\bar{y}_m) = \frac{1}{2mr}\left[\sigma^2 - \frac{1}{2}\left(\wp^2_{(m)} + \wp^2_{(m+1)}\right)\right] \tag{2.5}$$

and

$$V(\bar{y}_e) = \frac{1}{2mr}\left[\sigma^2 - \frac{1}{2}\left(\wp^2_{(1)} + \wp^2_{(2m)}\right)\right]. \tag{2.6}$$

We can easily show that the estimators in Equations (2.3) and (2.4) are more efficient than mean estimator under SRSWR based on $2mr$ observations. Note that ERSS gives a sample consisting only extreme units in the data which ignore middle units of population in this way it is very sensitive to outliers. On the other hand MRSS includes only middle units of population in sample that may some time gives a less representative sample (in case of fat tailed distributions) although it gives an efficient estimate of population mean. A mixture of these two sampling schemes which may provide a more representative sample with moderate efficiency by including middle as well as extreme units is given in next section.

3 Extreme-cum-Median Ranked Set Sampling (EMRSS)

Taking motivation from [**?**] and [**?**], we provide a new sampling procedure, named Extreme-cum-Median Ranked Set Sampling (EMRSS), that provide a more representative sample than median and extreme ranked set sampling with moderate efficiency. This procedure works as follow:

1. Select $2m$ independent random samples each of size $2m$ and rank each sample in ascending order according to certain ranking mechanism.
2. In Step 2, we work as follow: (i) When m is odd, we initiate with selecting smallest ranked unit from first $(\frac{m-1}{2})$ samples, largest unit from next $(\frac{m-1}{2})$ samples and mth unit from mth sample which complete the application of ERSS. We continue the process by selecting $m+1$ unit from (m+1)th sample, mth unit from $(m+2)$ to $(\frac{m+1}{2}+m)$th samples and $(m+1)$th unit from last $(\frac{m-1}{2})$ samples to complete an MRSS from remaining samples.
   (ii) When m is even, we start with selection of smallest ranked unit from first $\frac{m}{2}$ samples and largest unit from next $\frac{m}{2}$ samples to complete the application of ERSS. From remaining m sets; we select mth unit from $(m+1)$ to $\frac{3m}{2}$th samples and (m+1)th unit from last $\frac{m}{2}$ samples.
3. Repeat Steps (1) and (2) r time to obtain $n = 2mr$ units.

To understand our selection procedure, we present two examples as follow:

Example 1

To explain the EMRSS with odd set size, we take $m = 5$. The sampling layout is given below

$$\begin{bmatrix}
\underline{\mathbf{Y_{1(1)}}} & Y_{1(2)} & Y_{1(3)} & Y_{1(4)} & Y_{1(5)} & Y_{1(6)} & Y_{1(7)} & Y_{1(8)} & Y_{1(9)} & Y_{1(10)} \\
\underline{\mathbf{Y_{2(1)}}} & Y_{2(2)} & Y_{2(3)} & Y_{2(4)} & Y_{2(5)} & Y_{2(6)} & Y_{2(7)} & Y_{2(8)} & Y_{2(9)} & Y_{2(10)} \\
Y_{3(1)} & Y_{3(2)} & Y_{3(3)} & Y_{3(4)} & Y_{3(5)} & Y_{3(6)} & Y_{3(7)} & Y_{3(8)} & Y_{3(9)} & \underline{\mathbf{Y_{3(9)}}} \\
Y_{4(1)} & Y_{4(2)} & Y_{4(3)} & Y_{4(4)} & Y_{4(5)} & Y_{4(6)} & Y_{4(7)} & Y_{4(8)} & Y_{4(9)} & \underline{\mathbf{Y_{4(10)}}} \\
Y_{5(1)} & Y_{5(2)} & Y_{5(3)} & Y_{5(4)} & \underline{\mathbf{Y_{5(5)}}} & Y_{5(6)} & Y_{5(7)} & Y_{5(8)} & Y_{5(9)} & Y_{5(10)} \\
Y_{6(1)} & Y_{6(2)} & Y_{6(3)} & Y_{6(4)} & Y_{6(5)} & \underline{\mathbf{Y_{6(6)}}} & Y_{6(7)} & Y_{6(8)} & Y_{6(9)} & Y_{6(10)} \\
Y_{7(1)} & Y_{7(2)} & Y_{7(3)} & Y_{7(4)} & \underline{\mathbf{Y_{7(5)}}} & Y_{7(6)} & Y_{7(7)} & Y_{7(8)} & Y_{7(9)} & Y_{7(10)} \\
Y_{8(1)} & Y_{8(2)} & Y_{8(3)} & Y_{8(4)} & \underline{\mathbf{Y_{8(5)}}} & Y_{8(6)} & Y_{8(7)} & Y_{8(8)} & Y_{8(9)} & Y_{8(10)} \\
Y_{9(1)} & Y_{9(2)} & Y_{9(3)} & Y_{9(4)} & Y_{9(5)} & \underline{\mathbf{Y_{9(6)}}} & Y_{9(7)} & Y_{9(8)} & Y_{9(9)} & Y_{9(10)} \\
Y_{10(1)} & Y_{10(2)} & Y_{10(3)} & Y_{10(4)} & Y_{10(5)} & \underline{\mathbf{Y_{10(6)}}} & Y_{10(7)} & Y_{10(8)} & Y_{10(9)} & Y_{10(10)}
\end{bmatrix}$$

The above layout shows 10 ranked sets of observations each of size 10. The bold underlined units are selected for actual measurement.

Example 2

Similarly, to explain the EMRSS with even set size, we take $m = 4$. The sampling layout is given by

$$\begin{bmatrix}
\underline{\mathbf{Y_{1(1)}}} & Y_{1(2)} & Y_{1(3)} & Y_{1(4)} & Y_{1(5)} & Y_{1(6)} & Y_{1(7)} & Y_{1(8)} \\
\underline{\mathbf{Y_{2(1)}}} & Y_{2(2)} & Y_{2(3)} & Y_{2(4)} & Y_{2(5)} & Y_{2(6)} & Y_{2(7)} & Y_{2(8)} \\
Y_{3(1)} & Y_{3(2)} & Y_{3(3)} & Y_{3(4)} & Y_{3(5)} & Y_{3(6)} & Y_{3(7)} & \underline{\mathbf{Y_{3(8)}}} \\
Y_{4(1)} & Y_{4(2)} & Y_{4(3)} & Y_{4(4)} & Y_{4(5)} & Y_{4(6)} & Y_{4(7)} & \underline{\mathbf{Y_{4(8)}}} \\
Y_{5(1)} & Y_{5(2)} & Y_{5(3)} & \underline{\mathbf{Y_{5(4)}}} & Y_{5(5)} & Y_{5(6)} & Y_{5(7)} & Y_{5(8)} \\
Y_{6(1)} & Y_{6(2)} & Y_{6(3)} & \underline{\mathbf{Y_{6(4)}}} & Y_{6(5)} & Y_{6(6)} & Y_{6(7)} & Y_{6(8)} \\
Y_{7(1)} & Y_{7(2)} & Y_{7(3)} & Y_{7(4)} & \underline{\mathbf{Y_{7(5)}}} & Y_{7(6)} & Y_{7(7)} & Y_{7(8)} \\
Y_{8(1)} & Y_{8(2)} & Y_{8(3)} & Y_{8(4)} & \underline{\mathbf{Y_{8(5)}}} & Y_{8(6)} & Y_{8(7)} & Y_{8(8)}
\end{bmatrix}$$

The above layout shows 8 ranked sets of units (sampling units) each of size 8. The bold underlined units are selected for actual measurement.
These layouts can be replicated r time, adding subscript j to $Y_{i(i)j}$ and by varying j from 1 to r, to obtain an EMRS sample having size $n = 2rm$. Let $\{Y_{1(1)j}, ...Y_{\frac{m-1}{2}(1)j}; Y_{\frac{m-1}{2}+1(2m)j}, ..Y_{m-1(2m)j}; Y_{m(m)j}; Y_{m+1(m+1)j}; Y_{m+2(m)j}, .., Y_{\frac{3m+1}{2}(m)j}; Y_{\frac{3m+1}{2}+1(m+1)j}..Y_{2m(m+1)j}\}$ and $\{Y_{1(1)j}, ...Y_{\frac{m}{2}(1)j}; Y_{\frac{m}{2}+1(2m)j}, ..., Y_{m(2m)j}; Y_{m+1(m)j}...Y_{\frac{3m}{2}(m)j}; Y_{\frac{3m}{2}+1(m)j}, ..Y_{m(2m)j}\}$ be an EMRSS of size

$2m$ in cycle j for odd and even m respectively. For each case, we have sample mean under EMRSS as follow

$$\bar{y}_{em}^{(o)} = \frac{1}{2rm}\sum_{j=1}^{r}\Bigg[\Bigg(\sum_{i=1}^{\frac{m-1}{2}} Y_{i(1)j} + \sum_{i=\frac{m-1}{2}+1}^{m-1} Y_{i(2m)j} + Y_{m(m)j}\Bigg) + \Bigg(Y_{m+1(m+1)j} + \sum_{i=m+2}^{\frac{3m+1}{2}} Y_{i(m)j} + \sum_{i=\frac{3(m+1)}{2}}^{2m} Y_{i(m+1)j}\Bigg)\Bigg] \tag{3.1}$$

and

$$\bar{y}_{em}^{(e)} = \frac{1}{2rm}\sum_{j=1}^{r}\Bigg[\Bigg(\sum_{i=1}^{\frac{m}{2}} Y_{i(1)j} + \sum_{i=\frac{m}{2}+1}^{m} Y_{i(2m)j}\Bigg) + \Bigg(\sum_{i=m+1}^{\frac{3m}{2}} Y_{i(m)j} + \sum_{i=\frac{3m}{2}+1}^{2m} Y_{i(m+1)j}\Bigg)\Bigg] \tag{3.2}$$

with

$$Bias(\bar{y}_{em}^{(o)}) = \frac{1}{2m}\Bigg[\Bigg(\frac{m-1}{2}\wp_{(1)} + \frac{m-1}{2}\wp_{(2m)} + \wp_{(m)}\Bigg) + \Bigg(\wp_{(m+1)} + \frac{m-1}{2}\wp_{(m)} + \frac{m-1}{2}\wp_{(m+1)}\Bigg)\Bigg] \tag{3.3}$$

and

$$Bias(\bar{y}_{em}^{(e)}) = \frac{1}{2m}\Bigg[\Bigg(\frac{m}{2}\wp_{(1)} + \frac{m}{2}\wp_{(2m)}\Bigg) + \Bigg(\frac{m}{2}\wp_{(m)} + \frac{m}{2}\wp_{(m+1)}\Bigg)\Bigg]. \tag{3.4}$$

By writing in simplest form

$$Bias(\bar{y}_{em}^{(o)}) = \frac{1}{2m}\Bigg[\frac{m-1}{2}\Bigg(\wp_{(1)} + \wp_{(2m)}\Bigg) + \frac{m+1}{2}\Bigg(\wp_{(m)} + \wp_{(m+1)}\Bigg)\Bigg]$$

and

$$Bias(\bar{y}_{em}^{(e)}) = \frac{1}{4}\Bigg[\Bigg(\wp_{(1)} + \wp_{(2m)}\Bigg) + \Bigg(\wp(m+1)\Bigg)\Bigg],$$

where $\wp_{(i)} = \mu_{(i)} - \mu$ and $\mu_{(i)}$ represents population mean of ith order statistic for $(i = 1, 2, ..., 2m)$. Here, $\bar{y}_{em}^{(o)}$ and $\bar{y}_{em}^{(e)}$ will be unbiased only if Y has symmetric distribution i.e. $\wp_{(m+1)} = -\wp_{(m)}$and $\wp_{(2m)} = -\wp_{(1)}$. We obtain variance of $\bar{y}_{em}^{(o)}$ and $\bar{y}_{em}^{(e)}$ assuming that the underlying distribution is symmetric about μ.

$$\begin{aligned}
Var(\bar{y}_{em}^{(o)}) =& \frac{1}{(2rm)^2}\sum_{j=1}^{r}\Bigg[\Bigg(\sum_{i=1}^{\frac{m-1}{2}}\sigma^2_{(1)} + \sum_{i=\frac{m-1}{2}+1}^{m-1}\sigma^2_{(2m)} + \sigma^2_{(m)}\Bigg) + \Bigg(\sigma^2_{(m+1)} \\
&+ \sum_{i=m+2}^{\frac{3m+1}{2}}\sigma^2_{(m)} + \sum_{i=\frac{3(m+1)}{2}}^{2m}\sigma^2_{(m+1)}\Bigg)\Bigg] \\
=& \frac{1}{(2rm)^2}\sum_{j=1}^{r}\Bigg[\Bigg\{\sum_{i=1}^{\frac{m-1}{2}}\Big(\sigma^2 - \wp^2_{(1)}\Big) + \sum_{i=\frac{m-1}{2}+1}^{m-1}\Big(\sigma^2 - \wp^2(2m)\Big), \\
&+ \Big(\sigma^2 - \wp^2(m)\Big)\Bigg\} + \Bigg\{\Big(\sigma^2 - \wp^2_{(m+1)}\Big) + \sum_{i=m+2}^{\frac{3m+1}{2}}\Big(\sigma^2 - \wp^2_{(m)}\Big) \\
&+ \sum_{i=\frac{3(m+1)}{2}}^{2m}\Big(\sigma^2 - \wp^2_{(m+1)}\Big)\Bigg\}\Bigg].
\end{aligned}$$

Simplest form of variance is

$$\begin{aligned}
Var(\bar{y}_{em}^{(o)}) =& \frac{1}{2rm}\sigma^2 - \frac{1}{4rm^2}\Bigg[\frac{m-1}{2}\Big(\wp^2_{(1)} + \wp^2_{(2m)}\Big) + \frac{m+1}{2}\Big(\wp^2_{(m)} + \wp^2_{(m+1)}\Big)\Bigg] \\
=& Var(\bar{y}_{srs}) - \frac{1}{4rm^2}\Bigg[\frac{m-1}{2}\Big(\wp^2_{(1)} + \wp^2_{(2m)}\Big) + \frac{m+1}{2}\Big(\wp^2_{(m)} + \wp^2_{(m+1)}\Big)\Bigg].
\end{aligned} \tag{3.5}$$

From Equations (2.5) and (3.5), we notice that $Var(\bar{y}_{em}^{(o)}) \leq Var(\bar{y}_m)$ if

$$\wp^2_{(1)} + \wp^2_{(2m)} \leq (\wp^2_{(m)} + \wp^2_{(m+1)}). \tag{3.6}$$

Further from Equations (2.6) and (3.5), we observe that $Var(\bar{y}_{em}^{(o)}) \leq Var(\bar{y}_e)$ if

$$\wp^2_{(1)} + \wp^2_{(2m)} \geq (\wp^2_{(m)} + \wp^2_{(m+1)}). \tag{3.7}$$

Inequalities (3.6) and (3.7) reveal the conditions in which we prefer EMRSS over MRSS and ERSS respectively. Similarly

$$
\begin{aligned}
Var(\bar{y}^{(e)}_{em}) =& \frac{1}{(2rm)^2}\sum_{j=1}^{r}\Bigg[\Bigg(\sum_{i=1}^{\frac{m}{2}}\sigma^2_{(1)} + \sum_{i=\frac{m}{2}+1}^{m}\sigma^2_{(2m)}\Bigg) + \Bigg(\sum_{i=m+2}^{\frac{3m}{2}}\sigma^2_{(m)} + \sum_{i=\frac{3m}{2}}^{2m}\sigma^2_{(m+1)}\Bigg)\Bigg],\\
=& \frac{1}{(2rm)^2}\sum_{j=1}^{r}\Bigg[\Bigg\{\sum_{i=1}^{\frac{m}{2}}\Big(\sigma^2 - \wp^2_{(1)}\Big) + \sum_{i=\frac{m}{2}+1}^{m-1}\Big(\sigma^2 - \wp^2(2m)\Big)\Bigg\}\\
&+\Bigg\{\sum_{i=m+2}^{\frac{3m}{2}}\Big(\sigma^2 - \wp^2_{(m)}\Big) + \sum_{i=\frac{3m}{2}}^{2m}\Big(\sigma^2 - \wp^2_{(m+1)}\Big)\Bigg\}\Bigg].
\end{aligned}
$$

Simplifying, we have

$$
\begin{aligned}
Var(\bar{y}^{(e)}_{em}) =& \frac{1}{2rm}\sigma^2 - \frac{1}{4rm^2}\Bigg[\frac{m-1}{2}\Big(\wp^2_{(1)} + \wp^2_{(2m)}\Big) + \frac{m+1}{2}\Big(\wp^2_{(m)} + \wp^2_{(m+1)}\Big)\Bigg]\\
=& Var(\bar{y}_{srs}) - \frac{1}{4rm^2}\Bigg[\frac{m}{2}\Big(\wp^2_{(1)} + \wp^2_{(2m)}\Big) + \frac{m}{2}\Big(\wp^2_{(m)} + \wp^2_{(m+1)}\Big)\Bigg],
\end{aligned}
\tag{3.8}
$$

where $Var(\bar{y}_{srs}) = \frac{1}{2rm}\sigma^2$ is the variance of $\bar{y}_{srs}$ with $n = 2rm$. This shows the superiority of EMRSS over SRSWR with same sample size in terms of efficiency.
From Equations (2.5) and (3.8), we notice that $Var(\bar{y}^{(o)}_{em}) \leq Var(\bar{y}_m)$ if

$$\wp^2_{(1)} + \wp^2_{(2m)} \leq (\wp^2_{(m)} + \wp^2_{(m+1)}). \tag{3.9}$$

Further from Equations (2.6) and (3.8), we see that $Var(\bar{y}^{(o)}_{em}) \leq Var(\bar{y}_e)$ if

$$\wp^2_{(1)} + \wp^2_{(2m)} \geq (\wp^2_{(m)} + \wp^2_{(m+1)}). \tag{3.10}$$

Inequalities (3.9) and (3.10) show the conditions in which we prefer EMRSS over MRSS and ERSS respectively.

4 Empirical Study

For purpose of the efficiency and bias comparisons, we conduct a simulation study by generating a hypothetical population of size 1000 on two variables Y (the study variable) and X (a concomitant variable correlated with Y).

The values of X are generated using different distributions mentioned in caption of Tables 1-8. After that Y is computed such that $Y = (\rho \times X) + e$, where e is normaly distributed error term, having zero mean and unit variance, independent of Y. The procedure of RSS, MRSS, ERSS and EMRSS is applied to estimate population mean of Y. The process is replicated 10,000 times to see the long-run behavior of the mean estimators. The results obtained from simulation study are presented in Tables 1-8. We compute the absolute bias (AB) and relative efficiency (RE) of the means estimator ($\bar{y}_\bullet$) for different ranked set sampling schemes with respect to mean estimator under SRSWR as follows:

$$AB = |\frac{\sum_{j=1}^{100,000} (\bar{y}_\bullet - \mu)}{100,000}| \quad \text{and} \quad RE = \frac{Var(\bar{y}_{srs})}{Var(\bullet)}. \tag{4.1}$$

Note: In Tables $1-4$, ranking is performed on the study variable Y itself while in Tables 5-8 ranking is performed on X (imperfect ranking).

Table 1 *AB and RE under Normal(5, 1)*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0005	0.0006	0.0167	3.1314	4.1016	2.4311	5.1323
	4	0.0012	0.002	0.015	4.009	5.1792	2.9046	5.5618
	5	0.0001	0.0017	0.0097	4.7232	6.9291	2.8897	7.5682
5	6	0.0013	0.0021	0.0094	5.6941	7.9063	3.4649	7.7904
	7	0	0.0017	0.0068	6.2995	9.6398	3.2775	9.4978
	8	0.0009	0.0019	0.007	7.1629	10.3364	3.9064	9.8288
	3	0.0003	0.0005	0.0169	3.1476	4.0856	2.4488	5.3077
	4	0.001	0.0015	0.0149	4.0046	5.3378	2.868	5.736
10	5	0.0001	0.0013	0.0099	4.6909	6.7885	2.898	7.6062
	6	0.0012	0.0019	0.0094	5.4645	7.5657	3.2978	7.5436
	7	0.0001	0.0017	0.0069	6.4544	9.3844	3.2447	9.2297
	8	0.001	0.0019	0.0069	7.1998	10.2368	3.8461	9.6834

Table 2 *AB and RE under Weibull(1.5,5)*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0785	0.1092	0.0836	2.9376	4.069	2.0198	5.4133
	4	0.0899	0.1488	0.085	3.6049	5.3921	2.2392	5.4063
	5	0.0961	0.1779	0.0518	4.2907	6.5971	2.2453	6.8137
5	6	0.101	0.2062	0.051	5.1071	7.958	2.3236	6.9271
	7	0.1033	0.2276	0.0297	5.7251	9.0975	2.4035	7.743
	8	0.1054	0.2478	0.029	6.3733	10.6377	2.4268	7.7882
	3	0.0785	0.1089	0.0847	2.9511	4.081	2.0599	5.3258
	4	0.0899	0.1483	0.0851	3.7027	5.358	2.2152	5.4629
10	5	0.096	0.1789	0.0511	4.287	6.7012	2.2989	6.5905
	6	0.1009	0.2057	0.0509	4.9923	8.028	2.3056	6.8929
	7	0.1034	0.2275	0.0295	5.6519	9.089	2.3642	7.5508
	8	0.1057	0.2477	0.029	6.536	10.8502	2.4188	7.6387

Table 3 *AB and RE under Uniform(0, 1)*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0146	0.0155	0.1171	3.1987	4.278	2.2933	5.4687
	4	0.01	0.0051	0.085	4.157	5.3486	2.876	5.628
	5	0.0176	0.0233	0.0734	4.8964	7.139	2.8616	7.9177
5	6	0.0142	0.0023	0.0489	5.7951	7.6916	3.3781	7.7221
	7	0.0154	0.0209	0.0531	6.2704	10.1769	3.2807	9.3923
	8	0.018	0.0002	0.0302	7.0944	10.091	3.9091	9.6747
	3	0.0143	0.0154	0.117	3.2652	4.2963	2.4024	5.4438
	4	0.0098	0.0048	0.0865	4.1352	5.2994	2.9538	5.7113
10	5	0.017	0.0223	0.0712	4.6022	6.9861	2.7285	7.4647
	6	0.0139	0.0027	0.0485	5.5044	7.5863	3.389	8.0759
	7	0.0158	0.0225	0.0524	6.2808	9.8752	3.2368	9.4496
	8	0.0178	0.0003	0.0296	7.3374	10.1235	3.9387	9.8072

Table 4 *AB and RE under Gamma(4, 3)*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0138	0.016	0.0563	3.1642	4.2259	2.3145	5.4877
	4	0.0156	0.0206	0.0541	3.7758	5.5858	2.5987	5.848
	5	0.0183	0.0235	0.0367	4.7468	7.1835	2.8996	7.7715
5	6	0.0195	0.0257	0.0356	5.5917	8.7557	3.0517	7.9716
	7	0.02	0.0258	0.0273	6.2134	10.0244	3.2978	9.5799
	8	0.0203	0.0275	0.0275	7.1932	11.8527	3.6392	10.2764
	3	0.0132	0.0153	0.0566	3.154	4.3525	2.3174	5.5388
	4	0.0163	0.0206	0.0535	3.9316	5.92	2.6061	5.9407
10	5	0.0182	0.0233	0.0363	4.5348	7.065	2.7109	7.3477
	6	0.0192	0.0251	0.0356	5.4887	8.6775	3.0781	7.8993
	7	0.0201	0.0258	0.0279	6.2698	10.3035	3.2456	9.6882
	8	0.0202	0.0275	0.0275	6.839	11.6733	3.4981	9.6977

Table 5 *AB and RE under Normal(5, 1) with imperfect ranking*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0002	0.001	0.0132	1.8096	1.9305	1.5829	2.65
	4	0.0008	0.0071	0.0288	1.863	2.1748	1.602	2.0936
5	5	0.0012	0.002	0.009	1.9614	2.0876	1.7928	2.9789
	6	0	0.0107	0.0199	2.0653	2.4369	1.7639	2.1932
	7	0.0017	0.003	0.0075	2.0555	2.2628	1.7817	3.224
	8	0.0003	0.0135	0.0139	2.2421	2.6114	1.8578	2.3055
	3	0.0006	0.0006	0.0135	1.7623	1.8514	1.6346	2.6866
	4	0.0014	0.0068	0.029	1.8341	2.1584	1.568	2.0833
10	5	0.0012	0.002	0.0091	2.0345	2.1322	1.8166	3.0383
	6	0.0014	0.0068	0.029	1.8341	2.1584	1.568	2.0833
	7	0.0002	0.0109	0.0198	2.1389	2.4775	1.7388	2.2435
	8	0.0002	0.0135	0.0139	2.2837	2.5841	1.8925	2.3354

Table 6 *AB and RE under Weibull(5, 1.5) with Imperfect ranking*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0841	0.1239	0.0793	2.4961	3.4896	1.6411	4.387
	4	0.1149	0.1949	0.1624	3.2778	4.6891	2.1122	2.8033
5	5	0.0935	0.2161	0.0386	3.3486	4.9917	1.6698	4.8715
	6	0.1252	0.2705	0.1709	4.0992	6.3877	2.3904	3.3865
	7	0.0984	0.2842	0.008	4.1034	6.1822	1.7583	5.3783
	8	0.1299	0.3251	0.1792	5.1031	7.7224	2.552	3.7354
	3	0.0838	0.1243	0.0799	2.435	3.5872	1.6485	4.2576
	4	0.1143	0.1947	0.1628	3.1553	4.632	2.0637	2.7638
10	5	0.0936	0.2157	0.0386	3.4231	5.0596	1.7014	4.9701
	6	0.125	0.2695	0.1713	4.1554	6.3588	2.3668	3.3209
	7	0.0978	0.2841	0.0078	4.0036	6.1803	1.6812	5.2811
	8	0.1299	0.3246	0.1791	5.0198	7.6715	2.5471	3.7639

Table 7 *AB and RE under Uniform(0, 1) with imperfect ranking*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0312	0.0287	0.0251	1.1184	1.0891	1.141	1.6567
	4	0.027	0.0121	0.0937	1.0647	1.1231	1.0464	1.2201
5	5	0.0588	0.0229	0.0133	1.183	1.127	1.1892	1.6825
	6 4	0.0465	0.0123	0.0775	1.0932	1.2055	1.0445	1.1847
	7	0.0862	0.0143	0.034	1.1585	1.1304	1.2202	1.7499
	8	0.0654	0.0172	0.0734	1.1374	1.1648	1.0341	1.1664
	3	0.0343	0.0267	0.0237	1.1078	1.1051	1.1606	1.7018
	4	0.0294	0.013	0.0915	1.0996	1.1335	1.0065	1.236
10	5	0.0615	0.0221	0.0151	1.0909	1.1033	1.1259	1.6208
	6	0.0478	0.0128	0.0799	1.0952	1.1904	1.0461	1.1982
	7	0.0866	0.014	0.0335	1.1697	1.1294	1.2165	1.7533
	8	0.0671	0.0166	0.0719	1.1472	1.177	1.0139	1.1538

Table 8 *AB and RE under Gamma(4, 3) with imperfect ranking*

		Absolute Bias (AB)			Relative Efficiency (RE)			
r	m	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$	$\bar{y}_{rss}$	$\bar{y}_m$	$\bar{y}_e$	$\bar{y}_{em}$
	3	0.0808	0.1101	0.0738	1.4542	1.5857	1.3107	2.274
	4	0.0583	0.0991	0.1031	1.4321	1.5754	1.2646	1.5612
5	5	0.102	0.1778	0.0464	1.5836	1.7534	1.3494	2.4664
	6	0.0712	0.1396	0.1024	1.4846	1.5437	1.2207	1.4824
	7	0.1112	0.2223	0.0275	1.621	1.8196	1.3001	2.4442
	8	0.079	0.1717	0.1053	1.5617	1.6147	1.2551	1.4729
	3	0.0832	0.1108	0.0724	1.4363	1.5592	1.2901	2.2832
	4	0.0608	0.0983	0.1061	1.4696	1.544	1.2795	1.5441
10	5	0.1017	0.1768	0.0476	1.5935	1.7237	1.2789	2.4111
	6	0.0706	0.1397	0.1036	1.5013	1.6188	1.2741	1.5401
	7	0.1111	0.2215	0.0288	1.5991	1.7385	1.2857	2.3308
	8	0.0785	0.1713	0.1041	1.5579	1.6064	1.2471	1.4395

Absolute Bias (AB) and Relative Efficiency (RE) of mean estimators of all considered sampling schemes are provided for several combinations of r and m in Tables 1-8. Tables 1-4 present AB and RE for perfect ranking while Tables 5-8 give AB and RE for imperfect ranking. AB and RE for even choices of set sizes (m) are shown in shaded rows as behavior of estimators under EMRSS may differ for even and odd set sizes. This can be noticed from Example 1 and Example 2 of EMRSS procedure given in Section 2. With respect to AB, EMRSS perform better than MRSS and ERSS when data is generated from Weibull distribution for both perfect and imperfect ranking. Among all other distributions, EMRSS perform better than ERSS only for certain choices of m. On the other hand EMRSS performs better than ERSS for all choices of r and m in term of efficiency for all distributions used for generating data. In context of efficiency EMRSS performs better than MRSS only for smaller choices of m. RE of all estimators increase with increase in set size m. Furthermore, relative efficiency of mean estimator under EMRSS is higher for perfect ranking as compared to imperfect ranking (see Tables 1-8) because imperfect ranking involve judgmental ranked units.

5 Application

Iris flower data set is used to illustrate the application of EMRSS and compared it with three well-known existing ranked set sampling schemes in terms of relative efficiency. The data consist of N=150 flowers with 5 variables.

Petal width is taken as the study variable and suppose that we are interested in estimating average width of petals. For $m = 4$, we take 8 independent samples each consisting 8 flowers and ranked them according to petal width. The ranked sets are presented in Table 9.

Table 9 *Ranked sets*

Samples	Ranks							
	1	2	3	4	5	6	7	8
1	**1.2**	1.3	1.3	1.8	2	2.1	2.2	2.4
2	**0.1**	0.2	0.2	1	1.2	1.4	1.9	2.3
3	0.1	0.2	1.1	1.3	1.4	2.1	2.4	**2.4**
4	0.1	0.3	1.0	1.0	1.4	1.5	2.3	**2.4**
5	0.2	0.2	0.2	**1.0**	1.6	1.7	1.9	2.3
6	0.1	0.2	1.0	**1.3**	1.4	1.4	1.8	1.9
7	0.1	0.2	0.3	1.2	**1.3**	1.5	1.6	1.9
8	0.2	0.2	0.2	1.3	**1.3**	1.5	2.2	2.3

The observations in bold form are taken from units that are selected under EMRSS. We obtain RSS, ERSS and MRSS from the above ranked sets in similar way. Table 10 provides ultimate ranked set samples obtained from Table 9, respective sample means and a 95 percent confidence limits for population mean in each case. In this example, we use $r = 1$ for simplicity.

Table 10 *Ranked Sets Samples*

Sample number	RSS	ERSS	MRSS	EMRSS
1	1.2	1.2	1.8	1.2
2	0.2	0.1	1	0.1
3	1.1	0.1	1.3	2.4
4	1	0.1	1	2.4
5	1.6	2.3	1.6	0.2
6	1.4	1.9	1.4	0.1
7	1.6	1.9	1.3	1.9
8	2.3	2.3	1.3	1.9
$\bar{y}_{(\bullet)}$	1.3	1.2375	1.3375	1.275
$\bar{y}_{(\bullet)} - 1.96 \times \sqrt{V(\bar{y}_{(\bullet)})}$	1.086468	1.03043	1.08111	1.063593
$\bar{y}_{(\bullet)} + 1.96 \times \sqrt{V(\bar{y}_{(\bullet)})}$	1.513531	1.44456	1.59388	1.486406

The last three rows of Table 10 provides ($\bar{y}_{(\bullet)}$) sample mean obtained using proposed and existing ranking mechanisms, lower and upper confidence limits for true population mean of petal width. To compute these limits variances of the mean estimators are obtained through repeated sampling from the population of 90 units (see detail later). In Table 11, variance and relative efficiency of mean estimator under different ranked set sampling are presented.

Table 11 *Variance and Relative Efficiency*

	Variance			
m	$\bar{y}_{rss}$	$\bar{y}_e$	$\bar{y}_m$	$\bar{y}_{em}$
4	0.01805953	0.01791547	0.02620678	0.01646436
5	0.01186896	0.01116143	0.01711133	0.0116339
6	0.008625955	0.007666784	0.01222065	0.004757325
7	0.006512832	0.00544884	0.00886409	0.004189423
	Relative Efficiency			
m	$\bar{y}_{rss}$	$\bar{y}_e$	$\bar{y}_m$	$\bar{y}_{em}$
4	4.021466	4.053802	2.771259	4.41109
5	4.895174	5.205482	3.395449	4.99408
6	5.612965	6.315189	3.961915	10.1774
7	6.372104	7.616382	4.681862	9.906006

The first four rows of Table 11 are obtained by applying formulas given in Equations (2.2), (2.5), (2.6) and (3.8), where $\sigma_i^2 = \sigma^2 - \wp_{(i)}$ for $i = 1, 2, 3...2m$ is obtained using following algorithm.

1. Select $2m$ samples each of size $2m$ for $m = 4, 5, 6, 7$ from population of 150 Iris flowers.
2. Ranked each sample within itself from smallest to largest according to petal width and compute $\hat{\sigma}^2_{(i)k} = \frac{1}{2m}\sum_{j=1}^{2m}\left(y_{j(i)k} - \mu_{(i)k}\right)^2$.
3. Repeat Steps (1) and (2) 10,000 times.
4. Compute $\sigma^2_{(i)} = \sum_{k=1}^{10000}\hat{\sigma}^2_{(i)k}/10000$ for $i = 1, 2, ..., 2m$.

The variance and relative efficiency of estimators under RSS, ERSS, MRSS and EMRSS with respect to variance of mean estimator under SRSWOR are given in Table 11. The numerical results suggest that EMRSS performs better than well known existing ranked set sampling schemes for all choices of m except of $m = 5$. For $m = 5$, EMRSS is slightly less efficient than ERSS. The petal width of Iris flowers has a fat tailed distribution as shown in Figure 1.

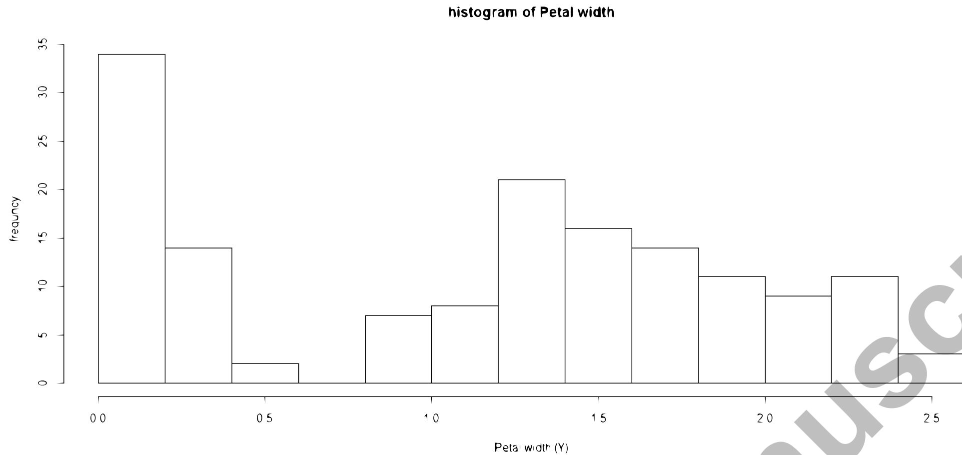

Figure 1 *Histogram of Petal width.*

The histogram in Figure 1 shows that the distribution of petal width is departed from normality. It also reveals that much data are densed at left tail of the distribution as it can be seen from 0 to 0.25 the bar is too high while it is low between 0.5 and 1.

6 Conclusion

A comparatively more representative, efficient and easy to handle ranked set sampling procedure is proposed and applied to estimate population mean of the study variable. The proposed sampling scheme is a mixture of ERSS and MRSS. It is shown both mathematically and through simulation study that under certain conditions EMRSS performs better than ERSS and MRSS for estimating population mean. Alternatively, we can conclude that it performs better than ERSS for all choices of m but it works better than MRSS only for relatively small m. The numerical application shows that EMRSS perform well when underlying distribution has fat tail. The suggested sampling scheme can be utilized in research areas where ranking all observations is tedious but obtain median and extreme values (five number summary statistics) are easy.

Acknowledgement

The authors are very grateful to the reviewers for their valuable comments for improvement of this paper.

References

[] Al-Omari, A. I. and Bouza, C. N. (2014). Review of ranked set sampling: modifications and applications.. *Revista Investigacin Operacional*, 3:215–240.

[] Balci, S., Akkaya, A., and Ulgen, B. (2013). Modified maximum likelihood estimators using ranked set sampling. *Journal of Computational and Applied Mathematics*, 238:171–179.

[] Biradar, B. and Santosha, C. D. (2015). Estimation of the population mean based on extremes ranked set sampling. *American Journal of Mathematics and Statistics*, 5(1):32–36.

[] Chen, Z., B. Z. and Sinha, B. (2002). Ranked set sampling: Theory and applications. *Springer Lecture Notes in Statistics*, 176.

[] Cobby J.M., Ridout M.S., B. P. and R.V., L. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards.. *Grass and Forage Science*, 40:257–263.

[] Cobby J.M., Ridout M.S., B. P. and R.V., L. (1987). Ranked set sampling with non-random selection of sets and errors in ranking. *Applied Statistics*, 36:145–152.

[] Dell, T. and Clutter, J. (1972). Ranked set sampling theorywith order statistics background. *Biometrics*, 28:545–555.

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

[] Muttlak, H. A. (1997). Median ranked set sampling. *Journal of Applied Statistical Sciences*, 6'.

[] Muttlak, H. A. (2003). Investigating the use of quartile ranked set samples for estimating the population mean. *Applied Mathematics and Computation.*, 146':437–443.

[] Patil, G. P., S. B. and Egemen, D. (2002). Ranked set sampling. *Encyclopedia of environmetrics*, 3:16841690.

[] Samawi, H., Abu-Dayyeh, W., and Ahmed, S. (1996). Extreme ranked set sampling. *The Biometrical Journal*, 30:577–586.

[] Stokes, S. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics*, 36:35–42.

[] Stokes, S. and Sager, T. (1988). Characterization of a ranked set sample with application to estimating distribution functions. *Journal of American Statistical Association*, 83:374–381.

[] 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*, 20:1–31.

Department of Statistics Quaid-i-Azam University Islamabad Pakistan

E-mail: shakeelatish05@gmail.com