On the improvement of paired ranked set sampling to estimate population mean

Syed Abdul Rehman[1], Javid Shabbir[2]

ABSTRACT

In ecological and environmental sampling the quantification of units is either difficult or overly demanding in terms of the time, money, workload, it requires. For this reason efficient and cost-effective sampling methods need to be devised for data collecting. The most commonly used method for this purpose is the Ranked Set Sampling (RSS). In this paper, a sampling scheme called Improved Paired Ranked Set Sampling (IPRSS) is proposed to estimate the population mean. The performance of the proposed IPRSS is evaluated under perfect and imperfect rankings. A simulation study based on selected hypothetical distributions and a real-life data set showed that IPRSS is more precise than RSS, Paired RSS (PRSS) or Extreme RSS (ERSS).

Key words: order statistics, ranked set sampling, relative efficiency, unbiased estimator, imperfect ranking.

1. Introduction

There are several methods of sampling that can be used to survey the natural resources of agriculture, biology, ecology, environmental management and forestry, etc. Whatever method of sampling is used, the main objective is to obtain precise estimates of population parameters at lowest cost of time, money and labour. One efficient sampling method is RSS, which provides more efficient estimates than simple random sampling (SRS). RSS is used when the exact quantification of selected units is expensive yet ranking a small set of selected units is inexpensive. For example, if interest lies in estimating the average weight of abalone, then it is easy to rank a small set of abalone with respect to their visual size. Similarly, the fuel consumption of vehicles can be ranked by a visual inspection of the vehicle size. McIntyre (1952) was the first to suggest the RSS method and estimated the average pasture and forage yields. The theory of RSS procedure was developed by Takahasi and Wakimoto (1968) assuming perfect ranking. Dell and Clutter (1972) showed that the mean estimator under imperfect RSS remains an unbiased estimator of population mean. Lynne Stokes (1977) showed the possibility of the ranking of the study variable based on an inexpensive concomitant variable. More applications of RSS can be seen from Johnson et al. (1993), Patil (1995), Mode et al. (1999), Al-Saleh and Al-Shrafat (2001), Yu and Tam (2002), Al-Saleh and Al-Hadrami (2003), Chen et al. (2003), Buchanan et al. (2005), Haq

---
[1]Balochistan University of Information Technology, Engineering and Management Sciences, Pakistan. E-mail: rehmanbukhari725@gmail.com. ORCID: https://orcid.org/0000-0003-0992-9310.

[2]Quaid-i-Azam University, Islamabad, Pakistan. E-mail: javidshabbir@gmail.com. ORCID: https://orcid.org/0000-0002-0035-7072.

et al. (2013) and references cited therein.

Muttlak (1996) introduced the paired RSS (PRSS) and Median RSS (MRSS) schemes for estimation of population mean. Muttlak (1998) extended the work of Patil (1995) and showed that the mean estimator under MRSS is more efficient than the mean estimators under SRS and RSS. Samawi et al. (1996) suggested Extreme RSS (ERSS) for estimation of population mean. Muttlak (2003) suggested quartile RSS (QRSS) to get more representative units as a sample. Biradar and Santosha (2015) proposed Independent Extreme RSS (IERSS) by modifying original RSS. It is well understood that population parameters like mean and variance might not be estimated more efficiently when the population under study is suspected of containing some extreme values or outliers. The selection of any of these extreme values in a sample will affect both the precision and accuracy of estimates. All the above discussed RSS schemes, especially usual RSS, MRSS and ERSS, might suffer the consequences of extreme values being selected in a sample. In order to overcome this deficiency, we propose a new RSS scheme by taking the advantage of ranking made in RSS. Our proposed RSS scheme called Improved Paired RSS (IPRSS) is more efficient as compared to other existing RSS schemes to estimate the population mean. Another advantage of using the IPRSS is that it requires less number of units to identify or observe as compared to RSS and MRSS, while requiring more units to identify as compared to PRSS.

Let Y and X respectively be the study variable and concomitant variable respectively. Let $Y_{i(i)j}$ be the ith order statistic in the ith set of the jth cycle of the RSS scheme. Let $\bar{y}_{RSS}$ be the sample mean of study variable based on the RSS sample. Let $\Delta_i = \mu_{(i)} - \mu$ be the deviation of the ith order statistic from true mean. Similarly, other deviations can be defined in the same manner.

2. Ranked Set Sampling (RSS)

By means of the RSS procedure, initially m^2 units are identified from the population and randomly allocated to m sets each of size m. Now within each set, units are ranked visually with respect to the variable under study or by any other economical method. The lowest ranked unit from the first set is selected and the second lowest ranked unit is selected from the second set. The procedure is carried out until the highest ranked unit is selected from the last set. A sample of m units is collected which also completes one cycle of a RSS. Repeating this procedure times results in a sample of size $n = mr$.

Let $Y_{1(1)j}, Y_{1(2)j}, \cdots, Y_{1(m)j}; Y_{2(1)j}, Y_{2(2)j}, \cdots, Y_{2(m)j}; \cdots, Y_{m(1)j}, Y_{m(2)j}, \cdots, Y_{m(m)j}$ be m independent random sets (samples) in the jth cycle, each of size m, such that $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, r$. Hence, the collected units in a sample by means of the RSS procedure are: $Y_{1(1)j}, Y_{2(2)j}, \cdots, Y_{m(m)j}$. Takahasi and Wakimoto (1968) showed that under perfect ranking, the mean of RSS is $\bar{y}_{RSS} = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m} Y_{i(i)j}$, which is an unbiased estimator of population mean μ and is more precise than $\bar{y}_{SRS} = \frac{1}{n} \sum_{i=1}^{n} Y_i$. The variance of $\bar{y}_{RSS}$ is given by

$$Var(\bar{y}_{RSS}) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \sum_{i=1}^{m} \Delta_{(i)}^2 \qquad (1)$$

where $\Delta_{(i)}$ is defined earlier.

2.1. Mathematics of RSS procedure

Let Y be the study variable with pdf $f(y)$ and CDF $F(y)$ with mean μ and variance σ^2. Let $Y_1, Y_2, \cdots, Y_m$ be a simple random sample of size m drawn from $f(y)$ for which the order statistics is $Y_{(1:m)}, Y_{(2:m)}, \cdots, Y_{(m:m)}$. The pdf and CDF of the ith order statistics $Y_{(i:m)}$ for $i = 1, 2, \cdots, m$, respectively, are given by

$$f_{(i:m)}(y) = \frac{m!}{(m-i)!(i-1)!}\{F(y)\}^{i-1}\{1-F(y)\}^{m-i}f(y) \qquad -\infty < y < \infty \qquad (2)$$

and

$$F_{(i:m)}(y) = \sum_{i=1}^{m}\binom{m}{r}\{F(y)\}^r\{1-F(y)\}^{m-r} \qquad (3)$$

The expression for mean and variance of $Y_{(i:m)}$, respectively, are given by

$$\mu_{(i:m)}(y) = \int y f_{(i:m)}(y) dy \qquad \text{and} \qquad (4)$$

$$\sigma^2_{(i:m)}(y) = \int (y - \mu_{(i:m)}(y))^2 f_{(i:m)}(y) dy, \qquad (5)$$

for detail see David and Nagaraja (2003).

3. RSS with errors in ranking (imperfect)

In practical situations, we come across problems in which visual ranking of the study variable is difficult or impossible. Lynne Stokes (1977) introduced a model to rank the study variable (Y) with respect to ranks of the auxiliary variable (X) correlated with (Y). This procedure was named imperfect RSS (IRSS). Following Lynne Stokes (1977), it is assumed that (Y, X) follows a bivariate normal distribution and the regression of Y on X is linear, i.e.

$$Y_{i[i]j} = \mu_y + \rho \frac{\sigma_y}{\sigma_x}\left(X_{i(i)j} - \mu_x\right) + \varepsilon_{ij} \qquad (6)$$

where ρ is the population correlation coefficient between Y and X, and ε_{ij} is the error term with zero mean and a constant variance, i.e.,

$$E(\varepsilon_{ij}) = 0 \quad and \quad Var(\varepsilon_{ij}) = \sigma^2_\varepsilon = \sigma^2_y(1 - \rho^2) \qquad (7)$$

Here, $X_{i(i)j}$ is the ith order statistics selected from the ith sample in the jth cycle of the auxiliary variable whose $X_{i[i])j}$ corresponding is the ith judgment order statistics of the study variable. The sample mean $\bar{y}_{IRSS}$ and its variance under IRSS respectively, are given by

$$\bar{y}_{IRSS} = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{m}Y_{i[i]j} \qquad (8)$$

and

$$Var(\bar{y}_{IRSS}) = \frac{1}{rm^2} \left\{ m\sigma_y^2(1-\rho^2) + \rho^2 \frac{\sigma_y^2}{\sigma_x^2} \sum_{i=1}^{m} \sigma_{x(i)}^2 \right\} \tag{9}$$

4. Paired Ranked Set Sampling (PRSS)

The PRSS procedure was suggested by Muttlak (1996) to estimate the mean of finite population. The procedure is as follows: For even sample size m, identify $m/2$ sets each of size m and rank the units in each set. The lowest and highest ranked units are collected from the first set, and the second lowest and second highest ranked units are collected from the second set. The process continues until $(m/2)^{th}$ and $((m+2)/2)^{th}$ ranked units are collected from the last set. For odd sample size m, identify $((m+1)/2)$ sets, each of size m, and select units as previously mentioned in the case of even size until $(m/2)^{th}$ set. Also $((m+1)/2)^{th}$ ranked unit is collected from the last set. This concludes in one cycle of a PRSS of size m. By repeating this process r times a sample of size m can be obtained. The sample mean under PRSS depending on even (E) and odd (O) sample sizes are calculated as, respectively

$$\bar{y}_{PRSS}^E = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m} \left(Y_{i(i)j} + Y_{i(m+1-i)j} \right) \tag{10}$$

and

$$\bar{y}_{PRSS}^O = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m+1)/2} Y_{i(i)j} + \sum_{i=1}^{(m-1)/2} Y_{i(m+1-i)j} \right) \tag{11}$$

The bias and variances of $\bar{y}_{PRSS}^E$ and $\bar{y}_{PRSS}^O$ respectively, are given by

$$Bias(\bar{y}_{PRSS}^E) = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m} \left(\Delta_{i(i)j} + \Delta_{i(m+1-i)j} \right) \tag{12}$$

$$Var(\bar{y}_{PRSS}^E) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \sum_{j=1}^{r} \sum_{i=1}^{m} \left(\Delta_{i(i)j}^2 + \Delta_{i(m+1-i)j}^2 \right)$$
$$+ \frac{2}{(rm)^2} \sum_{j=1}^{r} \sum_{i=1}^{m} \Delta_{i(i)j} \Delta_{i(m+1-i)j} \tag{13}$$

and

$$Bias(\bar{y}_{PRSS}^O) = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m+1)/2} \Delta_{i(i)j} + \sum_{i=1}^{(m-1)/2} \Delta_{i(m+1-i)j} \right) \tag{14}$$

$$Var(\bar{y}_{PRSS}^O) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \left(\sum_{i=1}^{(m+1)/2} \Delta_{(i)}^2 + \sum_{i=1}^{(m-1)/2} \Delta_{(m+1-i)}^2 \right)$$
$$+ \frac{2}{r^2m^2} \sum_{j=1}^{r} \sum_{i=1}^{(m-1)/2} \Delta_{(i)} \Delta_{(m+1-i)} \tag{15}$$

In the case of symmetrical distribution both $\bar{y}_{PRSS}^{E}$ and $\bar{y}_{PRSS}^{O}$ are unbiased estimators of population mean.

5. Extreme Ranked Set Sampling (ERSS)

The ERSS procedure was introduced by Samawi et al. (1996) to estimate the population mean. It is an essential sampling scheme in the case where population is fat-tailed, e.g. students t-distribution, Uniform distribution, Cauchy distribution, etc., units deviates from their mean beyond 5-sigma limits. The ERSS technique is executed as:

From the population, identify m^2 units and allocate them randomly to m sets, each of size . Now rank the units within each set. For $m = even$, select the lowest ranked units from the first $m/2$ sets and highest ranked units from the last $m/2$ sets. For $m = odd$, select the lowest ranked unit from the first $(m-1)/2$ sets, and the highest ranked units from set $(m+1)/2$ to set $(m-1)$. The median, i.e. $((m+1)/2))^{th}$ unit is selected from the last set. This conclude one cycle of ERSS of size m. A sample of size $n = mr$ can be obtained by repeating the ERSS process r times.

In the jth cycle , ERSS is termed as: $Y_{1(1)j}, Y_{2(1)j}, \cdots, Y_{m/2(1)j}, Y_{m+2/2(m)j}, \cdots, Y_{m(m)j}$ for even m. While, for odd m, $Y_{1(1)j}, Y_{2(1)j}, \cdots, Y_{m-1/2(1)j}, Y_{m+1/2(m)j}, \cdots, Y_{m(m)j}, Y_{m(m-1/2)j}$ units are selected. The estimators based on ERSS, are given by:

$$\bar{y}_{ERSS}^{E} = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{m/2} Y_{i(1)j} + \sum_{i=(m+2)/2}^{m} Y_{i(m)j} \right) \tag{16}$$

and

$$\bar{y}_{ERSS}^{O} = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m-1)/2} Y_{i(1)j} + \sum_{i=(m+1)/2}^{m-1} Y_{i(m)j} + Y_{m(m+1)/2} \right) \tag{17}$$

The bias and variance of each estimator is obtained as

$$Bias(\bar{y}_{ERSS}^{E}) = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{m/2} \Delta_{(1)} + \sum_{i=(m+2)/2}^{m} \Delta_{(m)} \right) \tag{18}$$

$$Var(\bar{y}_{ERSS}^{E}) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \left(\sum_{i=1}^{m/2} \Delta_{(1)}^{2} + \sum_{i=(m+2)/2}^{m} \Delta_{(m)}^{2} \right) \tag{19}$$

and

$$Bias(\bar{y}_{ERSS}^{O}) = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m-1)/2} \Delta_{(1)} + \sum_{i=(m+1)/2}^{m-1} \Delta_{(m)} + \Delta_{((m+1)/2)} \right) \tag{20}$$

$$Var(\bar{y}_{ERSS}^{O}) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \left(\sum_{i=1}^{(m-1)/2} \Delta_{(1)}^{2} + \sum_{i=(m+1)/2}^{m-1} \Delta_{(m)}^{2} + \Delta_{((m+1)/2)}^{2} \right) \tag{21}$$

6. Improved Paired Ranked Set Sampling (IPRSS)

The proposed RSS procedure called IPRSS is an improvement in PRSS and is an effort to overcome the drawback of ERSS. There is a great chance that ERSS will select the extreme values contained by population under study by considering only the first and last order statistic of sets, which not only affects the representativeness of a sample but the accuracy and precisions both are also highly affected. The proposed IPRSS scheme is also an effort to counter this drawback.

The IPRSS procedure is as follows:

For an even sample size m , identify $m/2$ sets each of size $m+2$ from a population and within each set rank the units. Now, choose the second lowest and second highest ranked units from the first set. Similarly select the third lowest and third highest ranked units from the second set. The procedure continues until $(m+4)^{th}$ and $(m+4/2)^{th}$ ranked units are selected from the last set.

For an illustration, a selection of units by IPRSS for $m = 4$ is given by

$$
\begin{array}{cccccc}
Y_{1(1)j} & \underline{Y_{1(2)j}} & Y_{1(3)j} & Y_{1(4)j} & \underline{Y_{1(5)j}} & Y_{1(6)j} \\
Y_{2(1)j} & Y_{2(2)j} & \underline{Y_{2(3)j}} & \underline{Y_{2(4)j}} & Y_{2(5)j} & Y_{2(6)j}
\end{array}
$$

In case of odd sample size m, identify $m+1/2$ sets each of size $m+2$ and select units as aforesaid in case of even until $m-1/2$ set. Also $(m+3/2)^{th}$ ranked unit is choosen from the last set. For example $m = 5$, then the IPRSS procedure can be illustrated as

$$
\begin{array}{ccccccc}
Y_{1(1)j} & \underline{Y_{1(2)j}} & Y_{1(3)j} & Y_{1(4)j} & Y_{1(5)j} & \underline{Y_{1(6)j}} & Y_{1(7)j} \\
Y_{2(1)j} & Y_{2(2)j} & \underline{Y_{2(3)j}} & Y_{2(4)j} & \underline{Y_{2(5)j}} & Y_{2(6)j} & Y_{2(7)j} \\
Y_{3(1)j} & Y_{3(2)j} & Y_{3(3)j} & \underline{Y_{3(4)j}} & Y_{3(5)j} & Y_{3(6)j} & Y_{3(7)j}
\end{array}
$$

This concludes one cycle of an IPRSS of size m. The process can be repeated r times to obtain IPRSS of size $n = mr$. The sample means under IPRSS depending on even (E) and odd (O) sample sizes are respectively given by

$$
\bar{y}_{IPRSS}^{E} = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m/2} \left(Y_{i(i+1)j} + Y_{i(m+2-i)j} \right) \tag{22}
$$

and

$$
\bar{y}_{IPRSS}^{O} = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m+1)/2} Y_{i(i+1)j} + \sum_{i=1}^{(m-1)/2} Y_{i(m+2-i)j} \right) \tag{23}
$$

In case of a symmetric distribution about zero, i.e. $Y_{(i+1)} \cong Y_{(m+2-i)}$ for $i = 1, 2, \cdots, m$. Arnold et al. (1992) showed that $\mu_{(i+1)} = -\mu_{(m+2-i)}$ and $\sigma_{(i+1)}^2 \cong -\sigma_{(m+2-i)}^2$. This indicates that if m is odd, then $\mu_{(m+1/2)} = \mu = 0$. Hence, the suggested IPRSS provides unbiased estimators, however in the case of asymmetrical distribution the expressions for bias along

with the variance of $\bar{y}_{IPRSS}^{E}$ and $\bar{y}_{IPRSS}^{O}$ respectively, are given as

$$Bias(\bar{y}_{IPRSS}^{E}) = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m/2} \left(\Delta_{(i+1)} + \Delta_{(m+2-i)} \right) \tag{24}$$

$$Var(\bar{y}_{IPRSS}^{E}) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \sum_{i=1}^{m/2} \left(\Delta_{(i+1)}^{2} + \Delta_{(m+2-i)}^{2} \right)$$
$$+ \frac{2}{(rm)^2} \sum_{j=1}^{r} \sum_{i=1}^{m/2} \Delta_{(i+1)} \Delta_{(m+1-i)} \tag{25}$$

and

$$Bias(\bar{y}_{IPRSS}^{O}) = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m+1)/2} \Delta_{(i+1)} + \sum_{i=1}^{(m-1)/2} \Delta_{(m+2-i)} \right) \tag{26}$$

$$Var(\bar{y}_{IPRSS}^{O}) = Var(\bar{y}_{SRS}) - \frac{1}{(rm)^2} \left(\sum_{i=1}^{(m+1)/2} \Delta_{(i+1)}^{2} + \sum_{i=1}^{(m-1)/2} \Delta_{(m+2-i)}^{2} \right)$$
$$+ \frac{2}{(rm)^2} \sum_{j=1}^{r} \sum_{i=1}^{(m-1)/2} \Delta_{(i+1)} \Delta_{(m+2-i)} \tag{27}$$

7. IPRSS with ranking errors (imperfect ranking)

In the case of ranking errors, the quantified observations of the study variable from the ith sample may not be the ith order statistics instead of the ith judgment order statistics. Dell and Clutter (1972) demonstrated that the sample mean of RSS with errors in ranking is an unbiased estimator of the population mean, and has smaller variance than the usual estimator based on SRS with the same sample size. However, the variance of the estimator with errors in ranking will be larger than the variance of the estimator with perfect ranking. Let $Y_{[i+1]}$ and $Y_{[m+2-i]}$ denote the $(i+1)^{th}$ and $(m+2-i)^{th}$ judgment order statistics of the sample for $i = 1, 2, \cdots, m+2$. Then, the IPRSS estimator of population mean with errors in ranking is defined as:

$$\bar{y}_{IPRSS}^{E} = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m/2} \left(Y_{i[i+1]j} + Y_{i[m+2-i]j} \right) \tag{28}$$

and

$$\bar{y}_{IPRSS}^{O} = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m+1)/2} Y_{i[i+1]} + \sum_{i=1}^{(m-1)/2} Y_{i[m+2-i]j} \right) \tag{29}$$

the bias and variance of $\bar{y}_{IPRSS}^{E}$ and $\bar{y}_{IPRSS}^{O}$ are

$$Bias(\bar{y}_{IPRSS}^{E}) = \frac{1}{rm} \sum_{j=1}^{r} \sum_{i=1}^{m/2} \left(\Delta_{[i+1]} + \Delta_{[m+2-i]} \right) \tag{30}$$

$$Var(\bar{y}_{IPRSS}^{E}) = Var(\bar{y}_{SRS}) - \frac{1}{rm^2} \sum_{i=1}^{m/2} \left(\Delta_{[i+1]}^{2} + \Delta_{[m+2-i]}^{2} \right)$$

$$+ \frac{2}{(rm)^2} \sum_{j=1}^{r} \sum_{i=1}^{m/2} \Delta_{[i+1]} \Delta_{[m+1-i]} \tag{31}$$

and

$$Bias(\bar{y}_{IPRSS}^{O}) = \frac{1}{rm} \sum_{j=1}^{r} \left(\sum_{i=1}^{(m+1)/2} \Delta_{[i+1]} + \sum_{i=1}^{(m-1)/2} \Delta_{[m+2-i]} \right) \tag{32}$$

$$Var(\bar{y}_{IPRSS}^{O}) = Var(\bar{y}_{SRS}) - \frac{1}{(rm)^2} \left(\sum_{i=1}^{(m+1)/2} \Delta_{[i+1]}^{2} + \sum_{i=1}^{(m-1)/2} \Delta_{[m+2-i]}^{2} \right)$$

$$+ \frac{2}{(rm)^2} \sum_{j=1}^{r} \sum_{i=1}^{(m-1)/2} \Delta_{[i+1]} \Delta_{[m+2-i]} \tag{33}$$

The relative efficiency (RE) of the proposed $\bar{y}_{IPRSS}$ with respect to $\bar{y}_{RSS}$, $\bar{y}_{PRSS}$ and $\bar{y}_{ERSS}$ is defined as:

$$RE(\bar{y}_{IPRSS}, \bar{y}_{\bullet}) = \frac{MSE(\bar{y}_{\bullet})}{MSE(\bar{y}_{IPRSS})} \tag{34}$$

where

$$MSE(\bar{y}_{\bullet}) = Var(\bar{y}_{\bullet}) + (Bias(\bar{y}_{\bullet}))^2 \tag{35}$$

and

$$MSE(\bar{y}_{\bullet}) = Var(\bar{y}_{IPRSS}) + \left(Bias(\bar{y}_{IPRSS}) \right)^2 \tag{36}$$

8. Empirical study

To evaluate the efficiency comparison of IPRSS against other competing procedures, a simulation study is conducted based on the following steps:
Generate a hypothetical (Normal, Logistic, Exponential and Poisson) distribution, say "X" of size 1000 and also generate "$e \sim N(1000, 0, 1)$". Compute the study variable "Y" by using the relation $Y = \rho X + e\sqrt{1 - \rho^2}$, where ρ is the population coefficient of correlation between the study variable and the concomitant variable whose value is taken fixed as 0.50, 0.70 and 0.90. Draw a sample of size $n = mr$, using SRSWOR, and estimate mean for each sampling procedure, for all the pairs of $r = 1, 3$ and $m = 4, 6, 8$. This procedure is repeated 10,000 times. The results are given in Tables 1 and 2.

Table 1: RE of IPRSS over other sampling procedures based on perfect ranking

Distribution	m/r	RSS		ERSS		PRSS	
		1	3	1	3	1	3
Normal(1000,100,5)	4	1.05	1.09	1.24	1.28	1.47	1.48
	6	1.04	1.03	1.45	1.45	4.98	11.96
	8	1.01	1.04	1.58	1.66	12.33	35.00
Logistic(1000,5,2)	4	1.29	1.27	1.68	1.73	1.73	1.64
	6	1.24	1.21	2.05	2.05	5.17	12.29
	8	1.18	1.18	2.36	2.43	12.41	33.98
Exponential(1000,1)	4	1.16	1.09	1.48	1.44	1.56	1.48
	6	1.07	1.04	1.80	2.05	4.79	11.5
	8	1.08	1.03	2.43	3.04	12.11	30.58
Poisson(1000,3)	4	1.39	1.36	2.03	1.98	1.91	1.76
	6	1.44	1.39	3.11	3.27	5.00	11.97
	8	1.40	1.29	4.24	4.62	12.21	30.4

Table 2: RE of IPRSS over other sampling procedures based on imperfect ranking

Distribution	r	m/ρ	RSS			ERSS			PRSS		
			0.5	0.7	0.9	0.5	0.7	0.9	0.5	0.7	0.9
Normal (1000,100,5)	1	4	1.04	1.05	1.05	1.18	1.21	1.22	1.35	1.4	1.43
		6	1.05	1.05	1.05	1.36	1.39	1.40	3.66	4.17	4.44
		8	1.04	1.03	1.02	1.42	1.47	1.50	7.98	9.59	10.49
	3	4	1.08	1.08	1.08	1.23	1.25	1.25	1.34	1.39	1.42
		6	1.03	1.03	1.03	1.34	1.38	1.40	8.13	9.61	10.41
		8	1.08	1.07	1.06	1.50	1.56	1.58	21.76	26.76	29.53
Logistic (1000,5,2)	1	4	1.20	1.24	1.26	1.47	1.57	1.62	1.44	1.54	1.61
		6	1.19	1.21	1.22	1.64	1.79	1.89	3.14	3.79	4.20
		8	1.13	1.16	1.18	1.76	1.96	2.08	6.11	7.94	9.17
	3	4	1.18	1.22	1.24	1.45	1.53	1.58	1.44	1.55	1.62
		6	1.14	1.17	1.19	1.63	1.79	1.88	6.59	8.41	9.58
		8	1.12	1.14	1.16	1.82	2.02	2.14	15.61	20.97	24.56
Exponential (1000,1)	1	4	1.02	1.03	1.04	1.11	1.17	1.25	1.06	1.09	1.13
		6	1.01	1.01	1.01	1.25	1.4	1.57	1.38	1.55	1.74
		8	1.01	1.00	1.00	1.46	1.78	2.13	1.8	2.2	2.66
	3	4	1.02	1.02	1.04	1.09	1.15	1.21	1.03	1.04	1.07
		6	1.01	1.01	1.02	1.52	1.78	2.05	1.88	2.36	2.84
		8	1.00	1.00	1.00	2.07	2.71	3.37	2.86	3.96	5.09
Poisson (1000,3)	1	4	1.16	1.22	1.27	1.43	1.6	1.74	1.31	1.44	1.55
		6	1.16	1.23	1.29	1.69	2.04	2.35	1.96	2.41	2.81
		8	1.09	1.14	1.19	1.89	2.4	2.87	3.12	4.25	5.3
	3	4	1.15	1.18	1.21	1.42	1.58	1.70	1.25	1.34	1.41
		6	1.11	1.15	1.18	1.79	2.18	2.51	3.43	4.68	5.74
		8	1.09	1.11	1.13	2.22	2.92	3.53	6.61	9.75	12.51

The simulation results show that the proposed IPRSS performs better than RSS, PRSS and ERSS. It is also observable from the simulation results that the efficiency of IPRSS significantly increases in comparison with other techniques when the number of cycles (r) is increased, i.e. in each cycle IPRSS avoids two suspected extreme values that are selected by other RSS schemes. Hence, with the increase in cycle (r), the number of suspected extreme values avoided by IPRSS also increases, which help us getting a sample free from outliers or extreme values. It is also notable that under imperfect ranking, all the RSS techniques including IPRSS perform poor relative to the case of perfect ranking. Since IPRSS collects a sample that avoids extreme values, so the results of IPRSS will be certainly better than other techniques. Since IPRSS is an improvement in the PRSS and it is designed to avoid the selection of extreme values in a sample so it is obvious that IPRSS will select more representative sample especially when population contains extreme values.

9. Application on real-life data set

For practical application, we consider the data provided by Singh and Mangat (2013). The net irrigated area is considered as a study variable while the number of tractors is considered as a concomitant variable. A simulation study is conducted by selecting 10,000 samples (the simulation procedure is discussed above). For simplicity we fix $r = 1$ and $m = 3,4,5,6,7$ using the procedure of RSS, ERSS, PRSS and IPRSS, by means of both perfect and imperfect rankings.

The $MSE(\bar{y}(\bullet))$ of each sampling procedure is calculated as,

$$MSE(\bar{y}(\bullet)) = \frac{1}{10,000} \sum_{i=1}^{10,000} (\bar{y}(\bullet) - \mu_y)^2 \tag{37}$$

The relative efficiencies are given in Table 3.

Table 3: RE of IPRSS over other sampling procedures

| Ranking | m | RE | | |
		RSS	ERSS	PRSS
	3	1.746	1.731	1.910
	4	1.358	2.436	1.613
Perfect	5	1.334	2.338	1.450
	6	1.157	4.270	1.379
	7	1.127	4.195	1.195
	3	1.617	1.613	1.715
	4	1.441	2.390	1.532
Imperfect	5	1.345	2.190	1.390
	6	1.207	3.582	1.239
	7	1.124	3.358	1.190

The results in Table 3 show that IPRSS performs better than RSS, ERSS and PRSS on real-life data set. Furthermore, IPRSS performs better than ERSS distinguishably when the sample size is increased.

10. Conclusions

In this article we proposed an efficient sampling scheme for the estimation of population mean. We showed numerically that the mean estimators under IPRSS are more precise than the mean estimators under RSS, ERSS and PRSS, for both perfect and imperfect rankings based on various distributions and a real-life data set. Generally, the proposed mean estimates under IPRSS are more efficient than the estimates under RSS, ERSS, and PRSS when more cycles of small samples are selected. Hence, the use of IPRSS is recommended over the existing RSS schemes. The current work can be extended to develop ratio and regression type estimators of population mean based on IPRSS to get more precise estimates.

Using the idea suggested in this paper, other RSS techniques can be modified to eliminate the effects of extreme values in samples.

Acknowledgements

References

Al-Saleh, M. F. and Al-Hadrami, S. A. (2003). Parametric estimation for the location parameter for symmetric distributions using moving extremes ranked set sampling with application to trees data. *Environmetrics: The official journal of the International Environmetrics Society*, 14(7):651–664.

Al-Saleh, M. F. and Al-Shrafat, K. (2001). Estimation of average milk yield using ranked set sampling. *Environmetrics: The official journal of the International Environmetrics Society*, 12(4):395–399.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). A first course in order statistics, vol. 54. *SIAM, Philadelphia*.

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

Buchanan, R. A., Conquest, L. L., and Courbois, J.-Y. (2005). A cost analysis of ranked set sampling to estimate a population mean. *Environmetrics: The official journal of the International Environmetrics Society*, 16(3):235–256.

Chen, Z., Bai, Z., and Sinha, B. (2003). *Ranked set sampling: theory and applications*, volume 176. Springer Science & Business Media.

David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*. John Wiley and Sons New York, 3 edition.

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

Haq, A., Brown, J., Moltchanova, E., and Al-Omari, A. I. (2013). Partial ranked set sampling design. *Environmetrics*, 24(3):201–207.

Johnson, G., Patil, G., and Sinha, A. (1993). Ranked set sampling for vegetation research. *Abstracta Botanica*, pages 87–102.

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

McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets. *Australian journal of agricultural research*, 3(4):385–390.

Mode, N. A., Conquest, L. L., and Marker, D. A. (1999). Ranked set sampling for ecological research: accounting for the total costs of sampling. *Environmetrics: The official journal of the International Environmetrics Society*, 10(2):179–194.

Muttlak, H. (1998). Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *Environmetrics: The official journal of the International Environmetrics Society*, 9(3):255–267.

Muttlak, H. A. (1996). Pair rank set sampling. *Biometrical journal*, 38(7):879–885.

Muttlak, H. A. (2003). Investigating the use of quartile ranked set samples for estimating the population mean. *Applied Mathematics and Computation*, 146(2-3):437–443.

Patil, G. (1995). ranked set sampling. *Environmental and Ecological Statistics*, 2(4):271–285.

Samawi, H. M., Ahmed, M. S., and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, 38(5):577–586.

Singh, R. and Mangat, N. S. (2013). *Elements of survey sampling*, volume 15. Springer Science & Business Media.

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.

Yu, P. L. and Tam, C. Y. (2002). Ranked set sampling in the presence of censored data. *Environmetrics: The official journal of the International Environmetrics Society*, 13(4):379–396.