Median Double Ranked Set Sampling Method

Nuran Medhat Al-Mawan[1, 3], El-Houssainy A. Rady[1] and Nasr I. Rashwan[2]

Department of Applied Statistics and Econometric, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt - Department of Basic Science, Faculty of Engineering, Modern Academy, Cairo, Egypt

Nuran_medhat@yahoo.com

Department of Applied Statistics and Econometric, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt;

Al-houssainy@yahoo.com

Department of Applied Statistics and Econometric, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt;

nasrrashwan@yahoo.com

Abstract:

In environmental monitoring and assessment, the main focus is to achieve observational economy and to collect data with unbiased, efficient and cost-effective sampling methods. Ranked set sampling (RSS) is one traditional method that is mostly used for accomplishing observational economy. In this article, we suggested new sampling method called median double ranked set sampling (MDRSS). The newly suggested sampling method MDRSS is compare to the simple random sampling (SRS), RSS, double ranked set sampling (DRSS), median ranked set sampling (MRSS). When the underlying distributions are symmetric and asymmetric, it is shown that, the variance of the mean estimator under MDRSS is always less than the variance of the mean estimator based on SRS and the other methods.

Indexing Terms/Keywords: Ranked set sampling; Median ranked set sampling; Double ranked set sampling; Relative precision.

1. Introduction

There are many sampling methods that can be used in surveys of natural resources in Agriculture, Biology, Environmental Management, Forestry, etc. The main objective of whatever sampling method is used is to obtain precise estimates of population parameters with minimum cost and expenditure. One method is ranked set sampling (RSS). RSS becomes an efficient alternative to simple random sampling (SRS) when taking exact measurements of selected units is very costly whereas ranking a small set of selected units is cheap. Ranking may be visually with respect to the study variable or by any inexpensive method. McIntyre (1952) was the first to suggest the RSS method for estimation of pasture and forage yields. McIntyre's goal was to maintain the unbiased of SRS while effectively incorporating into the estimates the outside information provided by visual inspection. Takahasi and Wakimoto (1968) developed the theory of RSS procedure under the assumption of perfect ranking. Dell and Clutter (1972) showed that the mean estimator under imperfect RSS remains an unbiased estimator of population mean. Muttlak (1997) suggested using median ranked set sampling method (MRSS) to estimate the population mean and he showed that it is more efficiently than the usual RSS method. Al-Saleh and Al-Kadiri (2000) suggested double ranked set sampling method (DRSS) for estimating the population mean and they showed that the sample mean based on DRSS is more efficient than the sample mean with RSS.

In this paper, modification for the RSS method is considered, namely median double ranked sampling. The newly suggested sampling method (MDRSS) is compared with RSS, MRSS and DRSS. It is shown that for the probability distributions considered in this study, MDRSS unbiased estimator has less variance than all the other methods. Computer simulation results are given to compare the efficiency for the estimators based on SRS with its counterparts RSS, MRSS, DRSS and MDRSS.

2. Ranked Set Sampling Methods

This section is concern with methods of ranked set sampling given a definition of each method and defined the mean and the variance of each case in the methods. Let X_1, X_2, X_n be a random sample with probability density function $f(x)$ with mean μ and variance $σ^2$. Let $X_{11}, X_{12}, X_{1n}; X_{21}, X_{22}, ..., X_{2n}; ..., X_{n1}, X_{n2}, ..., X_{nn}$ be independent random variables all with the same cumulative distribution function $F(x)$. Let $X_{(i:n)}$ denotes the ith order statistic from the ith sample of size n (i =1, 2, ..., n). The cycle may be repeated m times to get nm units. Let $X_{(i:n)j}$ denotes the jth cycle of size m (j= 1, 2, ..., m). The variance or the mean square error of the sample means for the RSS, DRSS, MRSS and MDRSS were calculated using the moments of the order statistics for the distributions considered in the study (see Harter and Balakrishnan (1996) and Balakrishnan and Chen (1997)). The relative efficiency (RE) of estimating the population mean using any of the RSS based methods with respect to the usual estimator SRS is defined as

$$RE(\bar{X}_{SRS}, \bar{X}_{RSS's}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{RSS's})}$$

Here X_{SRS} and $X_{RSS's}$ is both unbiased estimators. In case of biased estimator, the variance is replaced by the MSE of the estimator where

$$MSE(\bar{X}_{RSS's}) = Var(\bar{X}_{RSS's}) + [E(\bar{X}_{RSS's} - μ)]^2.$$

2.1. Ranked Set Sampling

The RSS scheme can be described as follows: In order to draw a sample of size n, identify n^2 units from the target population. The n units in each sample are ranked visually or by any inexpensive method with respect to the variable of interest. From the first set of n units, the smallest ranked unit is measured. From the second set of n units, the second smallest ranked unit is measured. The process is continued until from the nth set of n units the largest ranked until is measured. The estimator of the population mean for RSS of size n is given by

$$\bar{X}_{RSS} = \frac{1}{nm} \sum_{j=1}^{m} \sum_{i=1}^{n} X_{(i:n)j},$$

with variance

$$Var(\bar{X}_{RSS}) = \frac{1}{n^2 m} \sum_{i=1}^{n} \sigma^2_{(i:n)}, \text{ where } \sigma^2_{(i:n)} = E(X_{(i:n)} - \mu_{(i:n)})^2 = \sigma^2 - \sum_{i=1}^{n} (\mu_{(i:n)} - \mu)^2.$$

The RE for the symmetric and asymmetric distributions are

$$RE(\bar{X}_{SRS}, \bar{X}_{RSS}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{RSS})}.$$

2.2. Median Ranked Set Sampling

Muttlak (1997) introduced Median ranked set sampling (MRSS) procedure for estimating the population mean. MRSS procedure is explained as follows: for even sample size n, identify n^2 units from the target population. Each sample is ranked in itself as in ranked set sampling design. Select the $(\frac{n}{2})^{th}$ smallest rank from the first $\frac{n}{2}$ sets and select the $(\frac{n+2}{2})^{th}$ smallest rank from the other $\frac{n}{2}$ sets. Similarly, for odd sample size n, select the $(\frac{n+1}{2})^{th}$ smallest rank from all sets. The sample means under MRSS for even and odd n are given below:

$$\bar{X}_{MRSSE} = \frac{1}{nm} \sum_{j=1}^{m} (\sum_{i=1}^{\frac{n}{2}} X_{i(\frac{n}{2}:n)j} + \sum_{i=\frac{n}{2}+1}^{n} X_{i(\frac{n+2}{2}:n)j})$$

$$\text{and } \bar{X}_{MRSSO} = \frac{1}{nm} \sum_{j=1}^{m} (\sum_{i=1}^{n} X_{i(\frac{n+1}{2}:n)j}),$$

Both $\bar{X}_{MRSSE}$ and $\bar{X}_{MRSSO}$ are unbiased estimators. The variances of $Var(\bar{X}_{MRSSE})$ and $Var(\bar{X}_{MRSSO})$ respectively, are

$$Var(\bar{X}_{MRSSE}) = \frac{1}{n^2 m} (\sum_{i=1}^{\frac{n}{2}} Var(X_{i(\frac{n}{2}:n)}) + \sum_{i=\frac{n}{2}+1}^{n} Var(X_{i(\frac{n+2}{2}:n)})),$$

$$Var(\bar{X}_{MRSSO}) = \frac{1}{n^2 m} (\sum_{i=1}^{n} Var(X_{i(\frac{n+1}{2}:n)})).$$

The RE for the symmetric distribution is

$$RE(\bar{X}_{SRS}, \bar{X}_{MRSS}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{MRSS})},$$

$$\text{and } RE(\bar{X}_{SRS}, \bar{X}_{MRSS}) = \frac{Var(\bar{X}_{SRS})}{MSE(\bar{X}_{MRSS})} \text{ for asymmetric distribution.}$$

2.3. Double Ranked Set Sampling

Al-Saleh and Al-Kadiri (2000) introduced double ranked set sampling (DRSS) procedure for estimating the population mean. In order to select a double ranked set sample of size n, the experimenter needs to identify n^3 units from the target population. This may be difficult when an epidemic breaks out in some area or in queuing problems when data arrive in batches of varying sizes. Moreover, there may be a shortage of experimental units or ranking is difficult, time consuming and costly, see Samawi (2011) and Haq et al. (2013). In all such situations, the balanced DRSS cannot be conducted with full confidence or it is costly. The DRSS scheme can be described as follows: In order to draw a sample of size n, identify n^3 units from the target population. Randomly allocate these units to n sets each of size n^2. Apply RSS procedure to n sets in order to obtain n ranked set samples each of size n. Again, apply the RSS procedure to these n ranked sets, each of size n, to obtain a double ranked set sample of size n. The estimator of the population mean for DRSS of size n is given by

$$\bar{X}_{DRSS} = \frac{1}{nm} \sum_{j=1}^{m} \sum_{i=1}^{n} X^*_{(i:n)j},$$

with variance

$$Var(\bar{X}_{DRSS}) = \frac{1}{n^2 m} \sum_{i=1}^{n} \sigma^{*2}_{(i:n)}, \text{ where } \sigma^{*2}_{(i:n)} = E(X_{(i:n)} - \mu_{(i:n)})^2.$$

The RE for the symmetric distribution is

$$RE(\bar{X}_{SRS}, \bar{X}_{DRSS}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{DRSS})},$$

and $RE(\bar{X}_{SRS}, \bar{X}_{DRSS}) = \frac{Var(\bar{X}_{SRS})}{MSE(\bar{X}_{DRSS})}$ for asymmetric distribution.

3. Median Double Ranked Set Sampling Method

The MDRSS scheme can be described as follows: In order to draw a sample of size n, we identify n^3 units from the target population. Randomly allocate these units to n sets each of size n^2. Apply RSS procedure to n sets in order to obtain n ranked set samples each of size n. Apply the MRSS procedure to these n ranked sets, each of size n, to obtain a median double ranked set sample of size n. Let $X^*_{1(\frac{n}{2}:n)}, X^*_{2(\frac{n}{2}:n)}, ..., X^*_{\frac{n}{2}(\frac{n}{2}:n)}, X^*_{\frac{n+2}{2}(\frac{n+2}{2}:n)}, ...,$ $X^*_{n(\frac{n+2}{2}:n)}$ denote the MDRSSE. For odd sample size, let $X^*_{i(\frac{n+1}{2}:n)}$ be the median of the ith sample of the rank $\frac{n+1}{2}$ ($i = 1, 2, ..., n$). And $X^*_{1(\frac{n+1}{2}:n)}, X^*_{2(\frac{n+1}{2}:n)}, ..., X^*_{n(\frac{n+1}{2}:n)}$ denote the measured MDRSSO. The sample means under MDRSS for even and odd n are given below:

$$\bar{X}_{MDRSSE} = \frac{1}{nm}\sum_{j=1}^{m}(\sum_{i=1}^{\frac{n}{2}} X^*_{i(\frac{n}{2}:n)j} + \sum_{i=\frac{n}{2}+1}^{n} X^*_{i(\frac{n+2}{2}:n)j})$$

and $\bar{X}_{MDRSSO} = \frac{1}{nm}\sum_{j=1}^{m}(\sum_{i=1}^{n} X^*_{i(\frac{n+1}{2}:n)j})$

The variances of Var(X_{MDRSSE}) and Var(X_{MDRSSO}) respectively, are

$$\text{Var}(\bar{X}_{MDRSSE}) = \frac{1}{n^2m}(\sum_{i=1}^{\frac{n}{2}} Var(X^*_{i(\frac{n}{2}:n)}) + \sum_{i=\frac{n}{2}+1}^{n} Var(X^*_{i(\frac{n+2}{2}:n)})),$$

$$\text{Var}(\bar{X}_{MDRSSO}) = \frac{1}{n^2m}(\sum_{i=1}^{n} Var(X^*_{i(\frac{n+1}{2}:n)})).$$

It is important to note that, using MDRSSE method when n is even, the measured units $X^*_{1(\frac{n}{2}:n)}, X^*_{2(\frac{n}{2}:n)}, ..., X^*_{\frac{n}{2}(\frac{n}{2}:n)}$ are *iid*, also, $X^*_{\frac{n}{2}(\frac{n}{2}:n)}, X^*_{\frac{n+2}{2}(\frac{n+2}{2}:n)}, ..., X^*_{n(\frac{n+2}{2}:n)}$ are *iid* and using MDRSSO when n is odd, the measured units $X^*_{1(\frac{n+1}{2}:n)}, X^*_{2(\frac{n+1}{2}:n)}, ..., X^*_{n(\frac{n+1}{2}:n)}$ are *iid*.

The RE for the symmetric distribution is

$$RE(\bar{X}_{SRS}, \bar{X}_{MDRSS}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{MDRSS})},$$

and $RE(\bar{X}_{SRS}, \bar{X}_{MDRSS}) = \frac{Var(\bar{X}_{SRS})}{MSE(\bar{X}_{MDRSS})}$ for asymmetric distribution.

<u>Lemma</u> (i) $\bar{X}_{MDRSSE}$ and $\bar{X}_{MDRSSO}$ are unbiased estimators.

(ii) Var($\bar{X}_{MDRSSE}$), Var($\bar{X}_{MDRSSO}$) $\prec$ Var ($\bar{X}_{SRS}$).

(iii) If the distribution is not symmetric about μ, then the mean square error (MSE) of $\bar{X}_{MDRSSO}$ is less than the variance of Var($\bar{X}_{SRS}$)**.**

Proof. To prove **(i)**, if n is odd, we have

$$\mathsf{E}(\bar{X}_{MDRSSO}) = \frac{1}{nm}\sum_{j=1}^{m}(\sum_{i=1}^{n} E(X^*_{i(\frac{n+1}{2}:n)j})),$$

$$= \frac{1}{nm}\sum_{j=1}^{m}(\sum_{i=1}^{n} \mu^*_{(\frac{n+1}{2}:n)}) = \frac{1}{nm}\left(nm\,\mu^*_{(\frac{n+1}{2})}\right) = \mu^*_{(\frac{n+1}{2})},$$

Since the distribution is symmetric about μ, then median = mean, which implies $\mu^*_{\left(\frac{n+1}{2}\right)} = \mu_{\left(\frac{n+1}{2}\right)} = \mu$.

Therefore:

$$E(\bar{X}_{MDRSSO}) = \mu.$$

If n is even, then

$$E(\bar{X}_{MDRSSE}) = \frac{1}{nm}\sum_{j=1}^{m}\left(\sum_{i=1}^{\frac{n}{2}} E\left(X^*_{i\left(\frac{n}{2}:n\right)j}\right) + \sum_{i=\frac{n}{2}+1}^{n} E\left(X^*_{i\left(\frac{n+2}{2}:n\right)j}\right)\right),$$

$$= \frac{1}{nm}\sum_{j=1}^{m}\left(\sum_{i=1}^{\frac{n}{2}} \mu^*_{\left(\frac{n}{2}:n\right)} + \sum_{i=\frac{n}{2}+1}^{n} \mu^*_{\left(\frac{n+2}{2}:n\right)}\right)$$

$$= \frac{1}{nm}\left(\frac{nm}{2}\mu^*_{\left(\frac{n}{2}\right)} + \frac{nm}{2}\mu^*_{\left(\frac{n+2}{2}\right)}\right) = \frac{1}{2}\left(\mu^*_{\left(\frac{n}{2}\right)} + \mu^*_{\left(\frac{n+2}{2}\right)}\right)$$

Since the distribution is symmetric about μ, then $\mu^*_{\left(\frac{n}{2}\right)} + \mu^*_{\left(\frac{n+2}{2}\right)} = 2\mu^* = 2\mu$. Therefore:

$$E(\bar{X}_{MDRSSE}) = \frac{1}{2}(2\mu) = \mu.$$

To prove **(ii)**, If n is even, then

$$Var(\bar{X}_{MDRSSE}) = \frac{1}{n^2 m}\left(\sum_{i=1}^{\frac{n}{2}} Var\left(X^*_{i\left(\frac{n}{2}:n\right)}\right) + \sum_{i=\frac{n}{2}+1}^{n} Var\left(X^*_{i\left(\frac{n+2}{2}:n\right)}\right)\right)$$

$$= \frac{1}{n^2 m}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^{*2}_{i\left(\frac{n}{2}:n\right)} + \sum_{i=\frac{n}{2}+1}^{n} \sigma^{*2}_{i\left(\frac{n+2}{2}:n\right)}\right)$$

$$= \frac{1}{n^2 m}\left(\frac{n}{2}\sigma^{*2}_{\left(\frac{n}{2}:n\right)} + \frac{n}{2}\sigma^{*2}_{\left(\frac{n+2}{2}:n\right)}\right)$$

Nevertheless, $\sigma^{*2}_{\frac{n}{2}} < \sigma^2$ for each i, this implies

$$Var(\bar{X}_{MDRSSE}) = \frac{1}{2nm}\left(\sigma^{*2}_{\frac{n}{2}} + \frac{n}{2}\sigma^{*2}_{\frac{n+2}{2}}\right) < \frac{\sigma^2}{nm} = Var(\bar{X}_{SRS})$$

Therefore: $Var(\bar{X}_{MDRSSO}) < Var(\bar{X}_{SRS})$, similarly when n is odd and similarly prove **(iii)**.

4. Simulation Study

Based on 30000 replications, a computer simulation is conduced to study the behavior of the efficiency of the sample mean using RSS, MRSS, DRSS and MDRSS with respect to SRS. The sample size n should be kept very small so that visual ranking errors will not destroy the gain in efficiency. Random observations are generated from **(1)** Standard normal distribution with $(\mu = 0, \sigma^2 = 1)$ **(2)** Beta distribution with $(\alpha = 3, \beta = 3)$ **(3)** Logistic distribution with $(\alpha = -1, \beta = 1)$

(4) Exponential distribution with $(\lambda = 1)$ **(5)** Gamma distribution with $(\alpha = 1, \beta = \frac{1}{2})$ **(6)** Weibull distribution with $(\lambda = 1, k = 2)$, were generated by the (MATHEMTICA) package. Samawi and Muttlak (1996) used 5000 replications in their paper because of that the replications were extended to be more than 5000.

The simulation results are summarized in a table and list of figures for the performance of the samples means with sample sizes $n = 6, 9, 11$ and one cycle $(m = 1)$. The results of the table and these figures show the relative efficiency (the ratio of the variances) of the estimators of the mean relative to SRS.

Table 1   Exact relative efficiency relative to SRS, RSS, MRSS, DRSS and MDRSS under symmetric and asymmetric distributions with different sample sizes of *n*.

Distribution	n	RSS	MRSS	DRSS	MDRSS
Normal (0,1)	6 9 11	3.208 4.431 5.123	4.081 6.074 7.197	11.887 21.609 29.397	23.415 55.473 83.102
Beta (3,3)	6 9 11	3.350 4.840 5.755	3.383 5.106 5.979	22.827 52.326 77.794	23.754 56.301 84.300
Logistic (-1,1)	6 9 11	2.888 3.958 4.563	4.855 7.399 8.995	6.093 9.537 11.963	22.428 55.852 82.459
Exponential (1)	6 9 11	2.398 3.147 3.679	4.500 7.097 5.772	1.062 1.098 1.017	16.745 51.066 73.891
Gamma (1,2)	6 9 11	2.425 3.242 3.654	5.341 5.276 10.280	1.712 1.088 1.155	14.539 51.579 72.128
Weibull (1,2)	6 9 11	2.424 3.160 3.715	4.984 8.561 10.745	1.825 1.087 1.168	12.190 49.839 71.555

- Table 1 shows that estimating the population mean using MDRSS is more efficient than RSS, MRSS and DRSS.
- Regarding the sample size n, the efficiency of the mean estimators using MDRSS relative to the other methods increases for almost results.
- Underlying the symmetric distributions, the population mean estimator using DRSS is more efficient than RSS and MRSS.
- Underlying the asymmetric distributions, the population mean estimator using MRSS is more efficient than DRSS and RSS.

To represent more about the distributions, the simulation results are summarized by list of figures from Figure 1 to Figure 3. The figures show more about the methods with discuss the better and the worst methods in each distribution.

a) Standard Normal Distribution

b) Beta Distribution

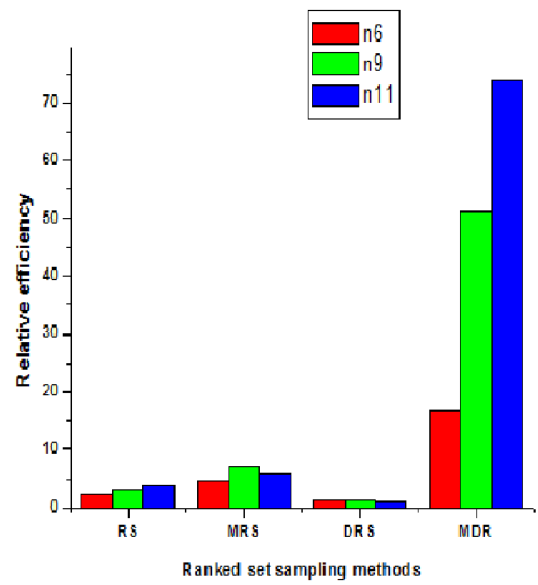**c) Logistic Distribution**
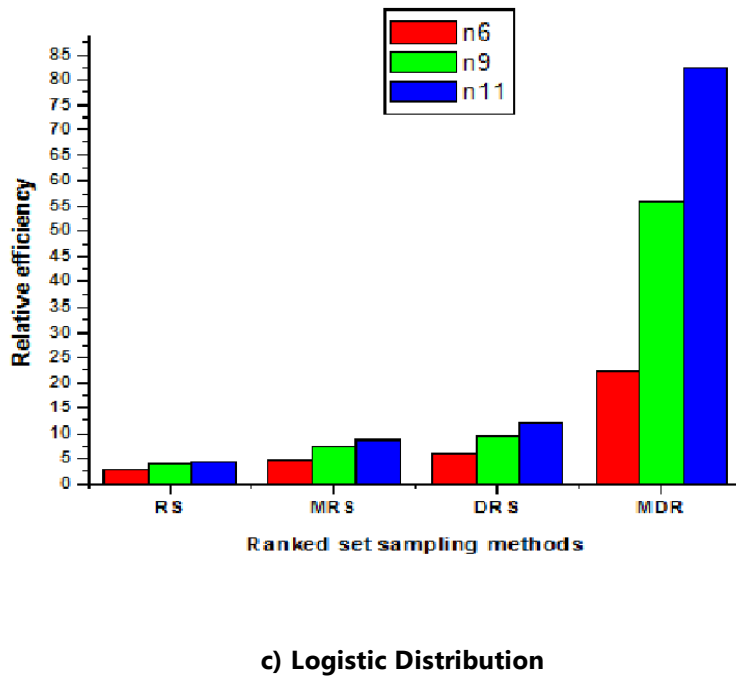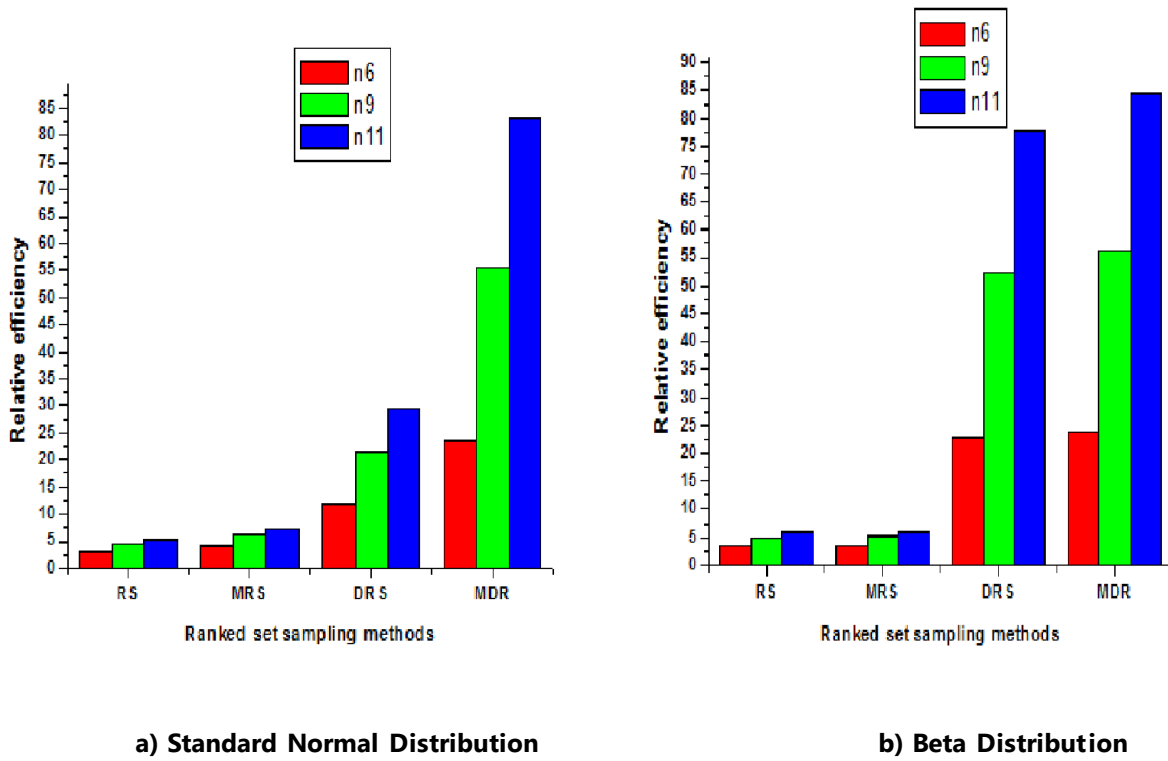
Figure 1   RE for symmetric distributions.

- For symmetric distributions in this study, the results for RSS and MRSS are close more than the results from DRSS and MDRSS.
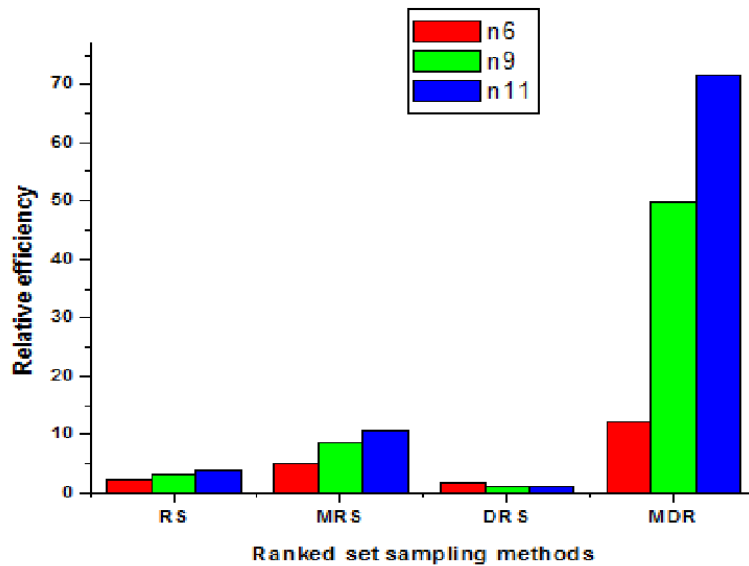
- For beta distribution, the results for RSS and MRSS are look constantly for all the sample sizes.
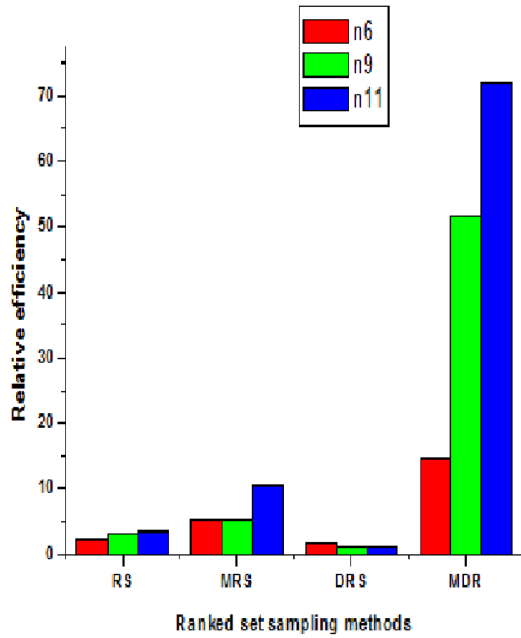
a) Exponential Distribution

b) Gamma Distribution

c) Weibull Distribution

Figure 2   RE for asymmetric distributions.

- The figures a, b and c have the same behavior for all the methods. Weibull (1, k) random variable is an exponential random variable with mean k. Gamma distribution with shape parameter α =1 and scale parameter β is an exponential (β) distribution. Both the gamma and weibull distributions can be seen as generalizations of the exponential distribution.

5. Summary and Conclusions

From the simulation results given in the tables the following can be concluded:

1. The MDRSS method is asymptotically more efficient than all the methods in the study.
2. Regarding the sample size n, the efficiency of the mean estimators using the ranked set sampling methods increase when the sample size increases for almost results.
3. Underlying the symmetric distributions, the population mean estimator using DRSS is more efficient than RSS and MRSS.
4. Underlying the asymmetric distributions, the population mean estimator using MRSS is more efficient than DRSS and RSS.

References

1. Al-Saleh, M. F., and Al- Kadiri, M. A. 2000. Double ranked set sampling. Statistics and Probability Letters 48, 205 -212.
2. Balakrishnan, N., and Chen, W.W. 1997. CRC Handbook of Tables for Order Statistics from Inverse Gussian Distributions with Applications, CRC Press, Boca Raton.
3. Dell, T. R. and Clutter, J. L. 1972. Ranked set sampling theory with order statistics background. Biometrics 28, 545 -555.
4. Haq, A., Brown, J., Moltchanova, E., and Al-Omari, A. I. 2013. Partial ranked set sampling design. Environ metrics 24, 201-207.
5. Haq, A., Brown, J., Moltchanova, E., and Al-Omari, A. I. 2016. Paired double-ranked set sampling. Communications in Statistics - Theory and Methods 10, 2873-2889.
6. Harter, H. L. and Balakrishnan, N. 1996.CRC Handbook of Tables for the use of Order Statistics in Estimation. CRC Press, Boca Raton.
7. McIntyre, G. A. 1952. A method of unbiased selective sampling, using ranked sets. Australian J. Agricultural Research 3, 385-390.
8. Muttlak, H. A. 1997. Median ranked set sampling. Journal of Applied Statistical Sciences 6, 245-255.
9. Muttlak, H. A. 1998. Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimator. Environmetrics 9, 255-267.
10. Patil, G. P., Sinha, A. K., Taillie, C. 1993. Relative precision of ranked set sampling: Comparison with the regression estimator. Environmetrics 4, 399-412.
11. Richard, S. L., Dennis, W.,and William, M. 2008. Mathematical Statistics with Applications. 7th ed. Thomson Higher Education. USE.
12. Samawi, H. M. 2011. Varied set size ranked set sampling with applications to mean and ratio estimation. The International Journal of Simulation Modeling 31, 6–13.
13. Samawi, H. M., and Muttlak, H. A. 1996. Estimation of Ratio Using Rank Set Sampling. Biometrical Journal 38, 753 -764.
14. 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.