

Percentile Double Ranked Set Sampling

¹Amer Ibrahim Al-Omari and ²Khalifa H. Jaber

¹Department of Mathematics, Jerash Private University, Jerash, Jordan

²Al Ain University of Science and Technology, UAE

Abstract: In this paper, percentile double ranked set sampling (PDRSS) method is suggested for estimating the population mean. The PDRSS method is compared with the simple random sampling (SRS), ranked set sampling (RSS), median ranked set sampling (MRSS) and the extreme ranked set sampling (ERSS) methods. When the underlying distribution is symmetric, it turns out that PDRSS produce unbiased estimators of the population mean and it is more efficient than SRS, RSS, MRSS and ERSS based on the same sample size. For asymmetric distribution considered in this study, it is shown that PDRSS has a small bias and it is more efficient than RSS, MRSS and ERSS for most cases considered in this study.

Key words: Ranked set sampling, simple random sampling, percentile, unbiased estimator.

INTRODUCTION

Ranked set sampling (RSS) is a cost effective sampling procedure when compared to the commonly used SRS in the situations where visual ranking of units can be done easily but the exact measurement of the units is difficult and expensive. McIntyre ^[4] proposed the sample mean based on RSS as an estimator of the population mean. He found that the estimator based on RSS is more efficient than SRS. Many modifications on RSS have been done since McIntyre ^[4]. Takahasi and Wakimoto ^[7] provided the necessary mathematical theory of RSS. Samawi and Muttlak ^[6] suggested RSS method for estimating the population ratio. Al-Saleh and Al-Kadiri ^[1] suggested double ranked set sampling method (DRSS) for estimating the population mean and they showed that the ranking at the second stage is easier than the ranking at the first stage. Muttlak ^[5] proposed percentile ranked set sampling (PRSS) for estimating the population mean. Jemain *et al.* ^[3] suggested multistage median ranked set sampling (MMRSS) for estimating the population median.

SAMPLING METHODS

Ranked set sampling: The RSS consist of selecting m random samples each of size m units from the population. The m units in each sample are ranked visually or by any inexpensive method with respect to the variable of interest. From the first set of m units, the smallest ranked unit is measured. From the second set

of m units, the second smallest ranked unit is measured. The process is continued until from the m th set of m units the largest ranked unit is measured. Repeat the process n times if needed to obtain a set of size mn from initial m^2n units.

Percentile ranked set sampling: In the percentile ranked set sampling (PRSS) procedure, select m random samples of size m units from the population and rank the units within each sample with respect to a variable of interest. If the sample size is even, select for measurement from the first $m/2$ samples the $(p(m+1))$ th smallest ranked unit and from the second $m/2$ samples the $(q(m+1))$ th smallest ranked unit, where $q=1-p$ and $0 \leq p \leq 1$. If the sample size is odd, select from the first $(m-1)/2$ samples the $(p(m+1))$ th smallest ranked unit and from the other $(m-1)/2$ samples the $(q(m+1))$ th smallest ranked unit and select from the remaining sample the median for that sample for actual measurement. The cycle may be repeated n times if needed to get mn units. These mn units are form the PRSS data.

Double ranked set sampling: The double ranked set sampling (DRSS) procedure can be described as follows: Identify m^3 units from the target population and divide these units randomly into m sets each of size m^2 . The procedure of ranked set sampling is applied on these sets to obtain m ranked set sampling each of size m , again apply the ranked set sampling procedure on the m ranked set sampling sets to obtain a DRSS of size m .

Corresponding Author: Amer Ibrahim Al-Omari, Department of Mathematics, Jerash Private University, Jerash Jordan

Percentile double ranked set sampling: To obtain a sample of size m based on PDRSS method, the following steps are required to be carried out:

Step 1: Randomly select m^3 units from the target population and divide them into m samples each of size m^2 .

Step 2: Apply the RSS method on the m sets; this step yields m ranked set samples each of size m .

Step 3: Without doing any actual quantifications on the m sets obtained in Step 2, apply the PRSS method described above. Repeat the process n times to obtain a set of size mn from initial m^3n units.

Note that if the sample size $m \leq 3$, the percentile double ranked set sampling will be reduced to the usual ranked set sampling procedure. However, we will always take the nearest integer of the $(p(m+1))$ th and $(q(m+1))$ th, where $q = 1 - p$ and $0 \leq p \leq 1$.

RESULTS AND DISCUSSION

Estimation of the population mean: Let X_1, X_2, \dots, X_m be a random sample with probability density function $f(x)$ with mean μ and variance σ^2 . Let $X_{11}, X_{12}, \dots, X_{1m}; X_{21}, X_{22}, \dots, X_{2m}; \dots, X_{m1}, X_{m2}, \dots, X_{mm}$ be independent random variables all with the same cumulative distribution function $F(x)$. The SRS estimator of the population mean from a sample of size m is given by

$$\bar{X}_{SRS} = \frac{1}{m} \sum_{i=1}^m X_i, \tag{1}$$

with variance

$$\text{Var}(\bar{X}_{SRS}) = \frac{\sigma^2}{m}. \tag{2}$$

The estimator of the population mean for a RSS of size m (McIntyre 1952) is given by

$$\bar{X}_{RSS} = \frac{1}{m} \sum_{i=1}^m X_{i(i:m)}, \tag{3}$$

with variance

$$\text{Var}(\bar{X}_{RSS}) = \frac{\sigma^2}{m} - \frac{1}{m^2} \sum_{i=1}^m (\mu_{(i:m)} - \mu)^2. \tag{4}$$

For the h th cycle, $h = 1, 2, \dots, n$, if the sample size is even, let $X_{i(p(m+1):m)h}^*$ be the $(p(m+1))$ th smallest ranked unit of the i th sample ($i = 1, 2, \dots, l$), where $l = m/2$ and let $X_{i(q(m+1):m)h}^*$ be the $(q(m+1))$ th smallest ranked unit of the i th sample ($i = l+1, l+2, \dots, m$). In this case, $X_{1(p(m+1):m)h}^*, X_{2(p(m+1):m)h}^*, \dots, X_{\frac{m}{2}(p(m+1):m)h}^*, X_{\frac{m+2}{2}(q(m+1):m)h}^*, \dots, X_{m(q(m+1):m)h}^*$ denote the measured PDRSSE.

For odd sample size, let $X_{i(p(m+1):m)h}^*$ be the $(p(m+1))$ th smallest ranked unit of the i th sample, ($i = 1, 2, \dots, k$) where $k = \frac{m-1}{2}$ and $X_{i(\frac{m+1}{2}:m)h}^*$ be the median of the i th sample of the rank $i = \frac{m+1}{2}$ and let $X_{i(q(m+1):m)h}^*$ be the $(q(m+1))$ th smallest ranked unit of the i th sample ($i = k+2, k+3, \dots, m$). In this case, $X_{1(p(m+1):m)h}^*, X_{2(p(m+1):m)h}^*, \dots, X_{\frac{m-1}{2}(p(m+1):m)h}^*, X_{\frac{m+1}{2}(\frac{m+1}{2}:m)h}^*, X_{\frac{m+3}{2}(q(m+1):m)h}^*, X_{\frac{m+5}{2}(q(m+1):m)h}^*, \dots, X_{m(q(m+1):m)h}^*$ denote the measured PDRSSO.

If m is even, the PDRSSE estimator of the population mean is defined as

$$\bar{X}_{PDRSSE}^* = \frac{1}{mn} \sum_{h=1}^n \left(\sum_{i=1}^{\frac{m}{2}} X_{i(p(m+1):m)h}^* + \sum_{i=\frac{m+2}{2}}^m X_{i(q(m+1):m)h}^* \right) \tag{5}$$

with variance

$$\text{Var}(\bar{X}_{PDRSSE}^*) = \frac{1}{nm^2} \sum_{h=1}^n \left(\sum_{i=1}^{\frac{m}{2}} \text{Var}(X_{i(p(m+1):m)h}^*) + \sum_{i=\frac{m+2}{2}}^m \text{Var}(X_{i(q(m+1):m)h}^*) \right). \tag{6}$$

If m is odd, the PDRSSO estimator of the population mean is given by

$$\bar{X}_{PDRSSO}^* = \frac{1}{mn} \sum_{h=1}^n \left(\sum_{i=1}^{\frac{m-1}{2}} X_{i(p(m+1):m)h}^* + X_{\frac{m+1}{2}(\frac{m+1}{2}:m)h}^* + \sum_{i=\frac{m+3}{2}}^m X_{i(q(m+1):m)h}^* \right) \tag{7}$$

with variance

$$\text{Var}(\bar{X}_{PDRSSO}^*) = \frac{1}{nm^2} \sum_{h=1}^n \left(\begin{aligned} &\sum_{i=1}^{\frac{m-1}{2}} \text{Var}(X_{i(p(m+1);m)h}^*) \\ &+ \text{Var}\left(X_{\frac{m+1}{2}(p(m+1);m)}^*\right) \\ &+ \sum_{i=\frac{m+3}{2}}^m \text{Var}(X_{i(q(m+1);m)h}^*) \end{aligned} \right) \cdot (8)$$

$$= \frac{1}{m} \left(\sum_{i=1}^{\frac{m-1}{2}} \mu_{(p(m+1))}^* + \mu_{\left(\frac{m+1}{2}\right)}^* + \sum_{i=\frac{m+3}{2}}^m \mu_{(q(m+1))}^* \right)$$

$$= \frac{1}{m} \left(\frac{m-1}{2} (\mu_{(p(m+1))}^*) + \mu_{\left(\frac{m+1}{2}\right)}^* + \frac{m-1}{2} (\mu_{(q(m+1))}^*) \right)$$

$$= \frac{1}{m} \left(\frac{m-1}{2} (\mu_{(p(m+1))}^* + \mu_{(q(m+1))}^*) + \mu \right)$$

$$= \frac{1}{m} \left(\frac{m-1}{2} (2\mu) + \mu \right) = \mu$$

It is important to note that, using PDRSSE method when m is even, the measured units $X_{1(p(m+1);m)h}^*$, $X_{2(p(m+1);m)h}^*$, ..., $X_{\frac{m}{2}(p(m+1);m)h}^*$ are iid, also $X_{\frac{m+2}{2}(q(m+1);m)h}^*$, $X_{\frac{m+4}{2}(q(m+1);m)h}^*$, ..., $X_{m(q(m+1);m)h}^*$ are iid and using PDRSSO when m is odd, the measured units $X_{1(p(m+1);m)h}^*$, $X_{2(p(m+1);m)h}^*$, ..., $X_{\frac{m-1}{2}(p(m+1);m)h}^*$ are iid and $X_{\frac{m+3}{2}(q(m+1);m)h}^*$, $X_{\frac{m+5}{2}(q(m+1);m)h}^*$, ..., $X_{m(q(m+1);m)h}^*$ are iid.

Now, we will prove that PDRSSE and PDRSSO are unbiased estimators of the population mean. If m is even, we have

$$E(\bar{X}_{PDRSSE}^*) = \frac{1}{m} \left(\begin{aligned} &\sum_{i=1}^{\frac{m}{2}} E(X_{i(p(m+1);m)}^*) \\ &+ \sum_{i=\frac{m+2}{2}}^m E(X_{i(q(m+1);m)}^*) \end{aligned} \right)$$

$$= \frac{1}{m} \left(\sum_{i=1}^{\frac{m}{2}} \mu_{(p(m+1))}^* + \sum_{i=\frac{m+2}{2}}^m \mu_{(q(m+1))}^* \right)$$

since the distribution is symmetric about μ , then $\mu - \mu_{(p(m+1))}^* = \mu_{(q(m+1))}^* - \mu$ and $\mu_{(p(m+1))}^* + \mu_{(q(m+1))}^* = 2\mu$ (David and Nagaraja [2]), so it is easy to show that

$$E(\bar{X}_{PDRSSE}^*) = \frac{1}{m} \left(\frac{m}{2} \mu_{(p(m+1))}^* + \frac{m}{2} \mu_{(q(m+1))}^* \right)$$

$$= \frac{1}{m} \left(\frac{m}{2} (\mu_{(p(m+1))}^* + \mu_{(q(m+1))}^*) \right)$$

$$= \frac{1}{m} \left(\frac{m}{2} (2\mu) \right) = \mu.$$

If m is odd, we have

$$E(\bar{X}_{PDRSSO}^*) = \frac{1}{m} \left(\begin{aligned} &\sum_{i=1}^{\frac{m-1}{2}} E(X_{i(p(m+1);m)}^*) + E\left(X_{\frac{m+1}{2}(p(m+1);m)}^*\right) \\ &+ \sum_{i=\frac{m+3}{2}}^m E(X_{i(q(m+1);m)}^*) \end{aligned} \right)$$

Simulation study: To investigate the performance of the suggested estimators of the population mean using SRS, RSS, MRSS, ERSS and PDRSS methods, a simulation study is conducted. Three symmetric distributions, namely, uniform, normal and logistic and three asymmetric distributions, exponential, gamma and weibull, are considered. If the parent distribution is symmetric about μ , the efficiency of RSS, MRSS, ERSS and PDRSS procedures, respectively is defined as

$$eff(\bar{X}_{RSS}, \bar{X}_{SRS}) = \frac{\text{Var}(\bar{X}_{SRS})}{\text{Var}(\bar{X}_{RSS})},$$

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

$$eff(\bar{X}_{ERSS}, \bar{X}_{SRS}) = \frac{\text{Var}(\bar{X}_{SRS})}{\text{Var}(\bar{X}_{ERSS})},$$

$$eff(\bar{X}_{PDRSS}^*, \bar{X}_{SRS}) = \frac{\text{Var}(\bar{X}_{SRS})}{\text{Var}(\bar{X}_{PDRSS}^*)}.$$

If the distribution is asymmetric, the efficiencies are:

$$eff(\bar{X}_{MRSS}, \bar{X}_{SRS}) = \frac{\text{Var}(\bar{X}_{SRS})}{\text{MSE}(\bar{X}_{MRSS})},$$

$$eff(\bar{X}_{ERSS}, \bar{X}_{SRS}) = \frac{\text{Var}(\bar{X}_{SRS})}{\text{MSE}(\bar{X}_{ERSS})},$$

$$eff(\bar{X}_{PDRSS}^*, \bar{X}_{SRS}) = \frac{\text{Var}(\bar{X}_{SRS})}{\text{MSE}(\bar{X}_{PDRSS}^*)}.$$

For simulations, we assume that the cycle is repeated once. In Tables 1, 2 and 3, we summarized the results for $m = 7, 10, 11$ respectively using RSS, MRSS, ERSS and PDRSS methods.

Based on Tables 1, 2 and 3, we can conclude the following:

Table 1: The efficiency of RSS, MRSS, ERSS and PDRSS relative to SRS for estimating the population mean with $m = 7$

Distribution		RSS	MRSS	ERSS	PDRSS $p20%$	$p30%$	$p40%$
Uniform (0,1)	efficiency	4.000	2.986	5.711	8.360	8.360	7.040
Normal (0,1)	efficiency	3.654	4.741	2.762	8.350	8.350	11.156
Logistic (-1,1)	efficiency	3.258	5.787	1.997	8.770	8.770	13.361
Exponential (1)	efficiency	2.700	1.791	0.812	5.151	5.151	2.104
	Bias		0.241	0.280	0.097	0.097	0.241
Gamma (1,2)	efficiency	2.718	1.808	0.812	5.292	5.292	2.111
	Bias		0.480	0.563	0.194	0.194	0.483
Weibull (1,3)	efficiency	2.717	1.791	0.814	5.189	5.189	2.096
	Bias		0.721	0.838	0.292	0.292	0.725

Table 2: The relative efficiency of RSS, MRSS, ERSS and PDRSS relative to SRS for estimating the population mean with $m = 10$

Distribution		RSS	MRSS	ERSS	PDRSS $p25%$	$p35%$	$p45%$
Uniform (0,1)	efficiency	5.500	4.023	12.282	13.196	11.449	11.006
Normal (0,1)	efficiency	4.827	6.606	2.938	15.253	18.546	20.255
Logistic (-1,1)	efficiency	4.198	8.247	1.899	16.464	21.919	24.412
Exponential (1)	efficiency	3.440	1.335	0.224	3.675	1.591	1.175
	Bias		0.254	0.516	0.145	0.243	0.286
Gamma (1,2)	efficiency	3.452	1.329	0.293	3.698	1.580	1.168
	Bias		0.508	1.029	0.290	0.486	0.572
Weibull (1,3)	efficiency	3.471	1.332	0.293	3.679	1.588	1.171
	Bias		0.762	1.547	0.435	0.730	0.859

Table 3: The relative efficiency of RSS, MRSS, ERSS and PDRSS relative to SRS for estimating the population mean with $m = 11$

Distribution		RSS	MRSS	ERSS	PDRSS $p25%$	$p35%$	$p45%$
Uniform (0,1)	efficiency	6.000	4.380	12.063	15.339	13.354	12.595
Normal (0,1)	efficiency	5.197	7.290	3.234	16.588	20.588	22.711
Logistic (-1,1)	efficiency	4.533	9.163	2.078	17.377	23.783	27.468
Exponential (1)	efficiency	3.671	1.162	0.306	4.374	1.649	1.148
	Bias		0.263	0.481	0.125	0.227	0.278
Gamma (1,2)	efficiency	3.723	1.322	0.311	4.294	1.661	1.132
	Bias		0.526	0.960	0.250	0.452	0.556
Weibull (1,3)	efficiency	3.699	1.167	0.308	4.219	1.666	1.139
	Bias		0.791	1.446	0.3376	0.680	0.833

- A gain in efficiency is obtained using PDRSS for estimating the population mean. For example, for $p = 45%$ and $m = 11$ the efficiency of PDRSS is 12.595 for estimating the mean of the standard uniform distribution
- The efficiency of PDRSS is increasing in the sample size when the distribution is symmetric. For example, with $p = 25%$ for $m = 10$ and 11, the efficiency of PDRSS is 15.253 and 16.588 for estimating the mean of the standard normal distribution
- For the normal and logistic distributions, PDRSS is more efficient than RSS and MRSS for all cases considered in this study. For the uniform distribution, PDRSS is more efficient than ERSS for most cases considered in this study
- For asymmetric distributions considered in this study, PDRSS has a small bias. However, the MSE

of PDRSS is less than the variance of SRS based on the same number of measured units.

CONCLUSIONS

In this paper, we suggested PDRSS method for estimating the population mean. Based on PDRSS, the findings of this paper may be summarized as follows:

- If the distribution is symmetric about μ ,

$$E(\bar{X}_{PDRSS}^*) = \mu,$$

$$\text{Var}(\bar{X}_{PDRSS}^*) < \text{Var}(\bar{X}_{SRS}),$$

$$\text{Var}(\bar{X}_{PDRSS}^*) < \text{Var}(\bar{X}_{MRSS}),$$
 and $\text{Var}(\bar{X}_{PDRSS}^*) < \text{Var}(\bar{X}_{ERSS})$ for most cases considered in this study.

- If the underlying distribution is asymmetric about μ , the mean square error (MSE) of \bar{X}_{PDRSS}^* is less than $\text{Var}(\bar{X}_{SRS})$.

It is recommended to use PDRSS for estimating the mean of symmetric distribution.

REFERENCES

1. Al-Saleh, M.F. and M. Al-Kadiri, 2000. Double ranked set sampling. *Stat. Prob. Lett.*, 48, 205 - 212.
2. David, H.A. and Nagaraja, H.N., 2003. *Order Statistics*, 3rd Edn. John Wiley and Sons, Inc., Hoboken, New Jersey.
3. Jemain, A.A., A.I. Al-Omari and K. Ibrahim, 2007. Multistage median ranked set sampling for estimating the population median. *J. Math. Stat.*, 3 (2): 58-64.
4. McIntyre, G.A., 1952. A method for unbiased selective sampling using ranked sets. *Aust. J. Agric. Res.*, 3: 385-390.
5. Muttalak, H.A., 2003. Modified ranked set sampling methods. *Pak. J. Stat.*, 19 (3): 315-323.
6. Samawi. H.M. and H.A. Muttalak, 1996. Estimation of ratio using rank set sampling. *The Biometrical J.*, 36 (6): 753-764.
7. Takahasi K. and K. Wakimoto, 1968. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Stat. Math.*, 20: 1-31.