American Journal of Applied Sciences

Original Research Paper

On the Mean Estimation using Stratified Double Median Ranked Set Sampling

[1]Mahmoud Ibrahim Syam, [2]Amer Ibrahim Al-Omari and [3]Kamarulzaman Ibrahim

[1,3]*Department of Mathematics, Al Al-Bayt University, Jordan*
[2]*School of Mathematical Sciences, University Kebangsaan Malaysia, Malaysia*

Article history
Received: 16-03-2014
Revised: 19-10-2015
Accepted: 24-03-2017

Corresponding Author:
Amer Ibrahim Al-Omari
Department of Mathematics,
Al al-Bayt University, Jordan
Email: alomari_amer@yahoo.com

Abstract: Stratified Double Median Ranked Set Sampling (SDMRSS) method is suggested for estimating the population mean. The SDMRSS is compared with the Simple Random Sampling (SRS), Stratified Simple Random Sampling (SSRS) and Stratified Ranked Set Sampling (SRSS) methods. It is shown that SDMRSS estimator is an unbiased of the population mean and is more efficient than the SRS, SSRS and SRSS counterparts. Also, SDMRSS increase the efficiency of mean estimation for specific value of the sample size. The SDMRSS is applied on real data set.

Keywords: Ranked Set Sampling, Double Ranked Set Sampling, Median Ranked Set Sampling, Mean, Efficiency

Introduction

During the last years, the Ranked Set Sampling method which was proposed by McIntyre (1952) for estimating the population mean of pasture yields was developed and modified by many authors. Dell and Clutter (1972) showed that the mean of the RSS is an unbiased estimator of the population mean even if there are errors in ranking. Muttlak (1997) suggested using Median Ranked Set Sampling (MRSS) to estimate the population mean. Al-Saleh and Al-Kadiri (2000) introduced Double Ranked Set Sampling for estimating the population mean. Al-Saleh and Al-Omari (2002) suggested Multistage Ranked Set Sampling that increases the efficiency of estimating the population mean for specific value of the sample size. Jemain and Al-Omari (2006) suggested Multistage Median Ranked Set Sampling (MMRSS) to estimate a population mean. Syam *et al*. (2013) considered the problem of estimating the population mean using stratified Double Percentile Ranked Set Sample. For more details about RSS see for example Bouza (2002; Samawi, 1996; Ohyama *et al*., 2008; Al-Omari and Jaber, 2008; Jemain *et al*., 2007; Takahasi and Wakimoto, 1968).

In this study, stratified double median ranked set sampling method is suggested for estimating the population mean of symmetric and asymmetric distributions. The paper is organized as follows: In section 2, some of sampling methods considered in this study are presented. Estimation of the population mean is given in section 3. A simulation study is considered in section 4. A real life example using SDMRSS is discussed in section 5. Finally, conclusions are provided in section 6.

Sampling Methods

Stratified Simple Random Sampling

In stratified sampling, the population of N units is first divided into L subpopulations of sizes $N_1, N_2,\ldots, N_L$ units, respectively. These subpopulations are non overlapping and together they comprise the whole population, i.e., $N_1+N_2+\ldots+N_L=N$. The subpopulations are called strata and the proportional allocation is considered. When the strata have been determined, a sample is chosen from each stratum, the drawing being made in different strata. The sample sizes within the strata are denoted by $n_1, n_2,\ldots, n_L$, respectively. If a simple random sample is taken from each stratum, the whole procedure is described as Stratified Simple Random Sampling (SSRS).

Ranked Set Sampling

The Ranked Set Sampling (RSS) is suggested by McIntyre (1952) can be described as follows: randomly select n samples of size n units each from the population of interest and then rank the units within each set with respect to the variable of interest. Select the smallest ranked unit from the first sample. The second smallest ranked unit from the second sample and the procedure is continued until the unit with the largest rank is selected for actual measurement from the n-th sample. Thus, a total of n measured units is obtained, one from each ordered sample of size n and this completed one cycle. The cycle may can be repeated m times if needed until nm units have been measured.

Median Ranked Set Sampling

The MRSS procedure as proposed by Muttlak (1997) depends on selecting n random samples of size n units from the population and the ranking of the units within each sample with respect to a variable of interest. If the sample size n is odd, then from each sample select for the measurement the $\left(\frac{n+1}{2}\right)^{th}$ smallest ranked unit, i.e., the median of the sample. If the sample size n is even, then select for the measurement from the first $\frac{n}{2}$ samples the $\left(\frac{n}{2}\right)^{th}$ smallest ranked unit of each set and from the second $\frac{n}{2}$ samples the $\left(\frac{n}{2}+1\right)^{th}$ smallest ranked unit. The cycle can be repeated m times if needed to get a sample of size nm units.

Double Median Ranked Set Sampling

The DMRSS (Samawi and Tawalbeh, 2002) is described as follows:

- Identify n^3 elements from the target population and divide these elements randomly into n sets, each of size n^2 elements
- If the sample size is even, chose from the first $\frac{n^2}{2}$ sets the $\left(\frac{n}{2}\right)^{th}$ smallest ranked unit of each set and from the second $\frac{n^2}{2}$ sets the $\left(\frac{n}{2}+1\right)^{th}$ smallest ranked unit. If the sample size is odd, select from all sets the $\left(\frac{n+1}{2}\right)^{th}$ smallest ranked unit. This step yields n sets each of size n
- Apply the MRSS procedure again on the sets obtained from Step (2) to obtain a DMRSS of size n. The cycle can be repeated m times if needed to get a sample of size nm units

Stratified Double Median Ranked Set Sampling

Referring to section 2.1, if the double median ranked set sampling is used in each stratum, the whole method is described as Stratified Double Median Ranked Set Sampling (SDMRSS). To explain the method, the following example with two strata is considered, the first stratum with even sample size and the second stratum with odd sample size.

Example 1

Consider a population with two strata, $L = 2$ ($h = 1,2$) and in the first stratum there are 27 elements divided into 3 sets with 9 elements in each set and in the second stratum there are 64 elements divided into 4 sets with 16 elements in each set as the following.

Stratum (1): Assume that the 27 elements are:

$$X_{11}^{(1)}, X_{12}^{(1)}, \cdots, X_{33}^{(1)}, X_{11}^{(2)}, X_{12}^{(2)}, \cdots, X_{33}^{(2)}, X_{11}^{(3)}, X_{12}^{(3)}, \cdots, X_{33}^{(3)}.$$

After ranking the elements in each set, the following sets are obtained:

$$\begin{bmatrix} X_{(11)}^{(1)} & X_{(12)}^{(1)} & X_{(13)}^{(1)} \\ X_{(21)}^{(1)} & X_{(22)}^{(1)} & X_{(23)}^{(1)} \\ X_{(31)}^{(1)} & X_{(32)}^{(1)} & X_{(33)}^{(1)} \end{bmatrix}, \begin{bmatrix} X_{(11)}^{(2)} & X_{(12)}^{(2)} & X_{(13)}^{(2)} \\ X_{(21)}^{(2)} & X_{(22)}^{(2)} & X_{(23)}^{(2)} \\ X_{(31)}^{(2)} & X_{(32)}^{(2)} & X_{(33)}^{(2)} \end{bmatrix} \text{ and } \begin{bmatrix} X_{(11)}^{(3)} & X_{(12)}^{(3)} & X_{(13)}^{(3)} \\ X_{(21)}^{(3)} & X_{(22)}^{(3)} & X_{(23)}^{(3)} \\ X_{(31)}^{(3)} & X_{(32)}^{(3)} & X_{(33)}^{(3)} \end{bmatrix}.$$

Apply the MRSS method on each of the above sets to get three sets as the following:

$$\text{Set }(1): X_{(12)}^{(1)}, X_{(22)}^{(1)}, X_{(32)}^{(1)},$$

$$\text{Set }(2): X_{(12)}^{(2)}, X_{(22)}^{(2)}, X_{(32)}^{(2)},$$

$$\text{Set }(3): X_{(12)}^{(3)}, X_{(22)}^{(3)}, X_{(32)}^{(3)}.$$

Now, apply the MRSS on these sets to get the double median ranked set sample units from the first stratum as:

$$X_{(22)}^{*(1)}, X_{(22)}^{*(2)}, X_{(22)}^{*(3)}.$$

Stratum (2): Assume that the 64 elements are:

$$Y_{11}^{(1)}, Y_{12}^{(1)}, \cdots, Y_{44}^{(1)}, Y_{11}^{(2)}, Y_{12}^{(2)}, \cdots, Y_{44}^{(2)},$$
$$Y_{11}^{(3)}, Y_{12}^{(3)}, \cdots, Y_{44}^{(3)}, Y_{11}^{(4)}, Y_{12}^{(4)}, \cdots, Y_{44}^{(4)}.$$

After ranking the elements in each set, the following sets are obtained:

$$\begin{bmatrix} Y_{(11)}^{(1)} & Y_{(12)}^{(1)} & Y_{(13)}^{(1)} & Y_{(14)}^{(1)} \\ Y_{(21)}^{(1)} & Y_{(22)}^{(1)} & Y_{(23)}^{(1)} & Y_{(24)}^{(1)} \\ Y_{(31)}^{(1)} & Y_{(32)}^{(1)} & Y_{(33)}^{(1)} & Y_{(34)}^{(1)} \\ Y_{(41)}^{(1)} & Y_{(42)}^{(1)} & Y_{(43)}^{(1)} & Y_{(44)}^{(1)} \end{bmatrix}, \begin{bmatrix} Y_{(11)}^{(2)} & Y_{(12)}^{(2)} & Y_{(13)}^{(2)} & Y_{(14)}^{(2)} \\ Y_{(21)}^{(2)} & Y_{(22)}^{(2)} & Y_{(23)}^{(2)} & Y_{(24)}^{(2)} \\ Y_{(31)}^{(2)} & Y_{(32)}^{(2)} & Y_{(33)}^{(2)} & Y_{(34)}^{(2)} \\ Y_{(41)}^{(2)} & Y_{(42)}^{(2)} & Y_{(43)}^{(2)} & Y_{(44)}^{(2)} \end{bmatrix}$$
$$\begin{bmatrix} Y_{(11)}^{(3)} & Y_{(12)}^{(3)} & Y_{(13)}^{(3)} & Y_{(14)}^{(3)} \\ Y_{(21)}^{(3)} & Y_{(22)}^{(3)} & Y_{(23)}^{(3)} & Y_{(24)}^{(3)} \\ Y_{(31)}^{(3)} & Y_{(32)}^{(3)} & Y_{(33)}^{(3)} & Y_{(34)}^{(3)} \\ Y_{(41)}^{(3)} & Y_{(42)}^{(3)} & Y_{(43)}^{(3)} & Y_{(44)}^{(3)} \end{bmatrix} \text{ and } \begin{bmatrix} Y_{(11)}^{(4)} & Y_{(12)}^{(4)} & Y_{(13)}^{(4)} & Y_{(14)}^{(4)} \\ Y_{(21)}^{(4)} & Y_{(22)}^{(4)} & Y_{(23)}^{(4)} & Y_{(24)}^{(4)} \\ Y_{(31)}^{(4)} & Y_{(32)}^{(4)} & Y_{(33)}^{(4)} & Y_{(34)}^{(4)} \\ Y_{(41)}^{(4)} & Y_{(42)}^{(4)} & Y_{(43)}^{(4)} & Y_{(44)}^{(4)} \end{bmatrix}$$

Apply the MRSS on each of the 16 elements to get four sets as the following:

$$\text{Set (1)}: Y_{(12)}^{(1)}, Y_{(22)}^{(1)}, Y_{(32)}^{(1)}, Y_{(42)}^{(1)},$$

$$\text{Set (2)}: Y_{(12)}^{(2)}, Y_{(22)}^{(2)}, Y_{(32)}^{(2)}, Y_{(42)}^{(2)},$$

$$\text{Set (3)}: Y_{(13)}^{(3)}, Y_{(23)}^{(3)}, Y_{(33)}^{(3)}, Y_{(43)}^{(3)},$$

$$\text{Set (4)}: Y_{(13)}^{(4)}, Y_{(23)}^{(4)}, Y_{(33)}^{(4)}, Y_{(43)}^{(4)}.$$

The elements of the double median ranked set sample in the second stratum are:

$$Y_{(22)}^{*(1)}, Y_{(22)}^{*(2)}, Y_{(33)}^{*(3)}, Y_{(33)}^{*(4)}.$$

Therefore, the SDMRSS units from the two strata are:

$$X_{(22)}^{(1)}, X_{(22)}^{(2)}, X_{(22)}^{(3)}, Y_{(22)}^{(1)}, Y_{(22)}^{(2)}, Y_{(33)}^{(3)}, Y_{(33)}^{(4)}.$$

Estimation of the Population Mean

Assume that the variable of interest X has a density $f(x)$ and a cumulative distribution function $F(x)$, with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a SRS from $f(x)$. Let $X_{11}, X_{12}, \ldots, X_{1n}$; $X_{21}, X_{22}, \ldots, X_{2n}$; $X_{n1}, X_{n2}, \ldots, X_{nn}$ be n simple random samples each of size n. Let $X_{i(1:n)}, X_{i(2:n)}, \ldots, X_{i(n:n)}$ be the order statistics of the ith sample $X_{i1}, X_{i2}, \ldots, X_{in}$, $i = 1,2,\ldots,n$. Therefore, the measured units $X_{i(1:n)}, X_{i(2:n)}, \ldots, X_{i(n:n)}$ constitute the ranked set sample.

The notation for the DMRSS will be used by replacing the sets of ordered statistics $X_{i(1:n)}, X_{i(2:n)}, \ldots, X_{i(n:n)}$, $i = 1,2,\ldots,n$ which obtained from the sample $X_{i1}, X_{i2}, \ldots, X_{in}$, $i = 1,2,\ldots,n$. If the sample size n is even, the DMRSS units will be $X^*_{1\left(\frac{n}{2}:n\right)}, \ldots, X^*_{\frac{n}{2}\left(\frac{n}{2}:n\right)}$, $X^*_{\frac{n}{2}+1\left(\frac{n+2}{2}:n\right)}$, $\ldots, X^*_{n\left(\frac{n+2}{2}:n\right)}$ and if the sample size n is odd, the MDRSS units will be $X^*_{1\left(\frac{n+1}{2}:n\right)}$, ..., $X^*_{\frac{n+1}{2}-1\left(\frac{n+1}{2}:n\right)}$, $X^*_{\frac{n+1}{2}\left(\frac{n+1}{2}:n\right)}$, $X^*_{\frac{n+1}{2}+1\left(\frac{n+1}{2}:n\right)}$.

The notation for the SDMRSS units will be $X^*_{h1\left(\frac{n_h}{2}:n_h\right)}, \ldots, X^*_{h\frac{n_h}{2}\left(\frac{n_h}{2}:n_h\right)}$, $X^*_{h\frac{n_h}{2}+1\left(\frac{n_h+2}{2}:n_h\right)}$, $\ldots, X^*_{hn_h\left(\frac{n_h+2}{2}:n_h\right)}$ if n_h is even and if the sample size n_h is odd, the SDMRSS units will be $X^*_{h1\left(\frac{n_h+1}{2}:n_h\right)}, \ldots,$ $X^*_{h\frac{n_h+1}{2}-1\left(\frac{n_h+1}{2}:n_h\right)}$, $X^*_{h\frac{n_h+1}{2}\left(\frac{n_h+1}{2}:n_h\right)}$, $X^*_{h\frac{n_h+1}{2}+1\left(\frac{n_h+1}{2}:n_h\right)}$, $\ldots$, $X^*_{hn_h\left(\frac{n_h+1}{2}:n_h\right)}$ where $X^*_{hi\left(\frac{n_h+1}{2}:n_h\right)}$ is the median of the ith sample in hth stratum.

The Stratified Double Median Ranked Set Sampling estimator of the population mean when n_h is even is defined as:

$$\overline{X}^*_{SDMRSS1} = \sum_{h=1}^{L} W_h \cdot \overline{X}_{DMRSS1h}$$
$$= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\left(\sum_{i=1}^{\frac{n_h}{2}} X^*_{hi\left(\frac{n_h}{2}:n_h\right)} + \sum_{i=\frac{n_h}{2}+1}^{n_h} X^*_{hi\left(\frac{n_h+2}{2}:n_h\right)}\right). \quad (1)$$

But when n_h is odd, the stratified double median ranked set sampling estimator of the population mean is:

$$\overline{X}^*_{SDMRSS2} = \sum_{h=1}^{L} W_h \cdot \overline{X}_{DMRSS2h} = \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\sum_{i=1}^{n_h} X^*_{hi\left(\frac{n_h+1}{2}:n_h\right)}, \quad (2)$$

where $W_h = \frac{N_h}{N}, N_h$ is the stratum size and N is the total population size.

The variance of the estimator SDMRSS1 is given by:

$$Var\left(\overline{X}^*_{SDMRSS1}\right) = \sum_{h=1}^{L} \frac{W_h^2}{n_h^2}\left(\sum_{i=1}^{\frac{n_h}{2}} \sigma^{2*}_{hi\left(\frac{n_h}{2}:n_h\right)} + \sum_{i=\frac{n_h}{2}+1}^{n_h} \sigma^{2*}_{hi\left(\frac{n_h+2}{2}:n_h\right)}\right). \quad (3)$$

The variance of SDMRSS2 (when n_h is odd) is given by:

$$Var\left(\overline{X}^*_{SDMRSS2}\right) = \sum_{h=1}^{L} \frac{W_h^2}{n_h^2}\sum_{i=1}^{n_h} \sigma^{2*}_{hi\left(\frac{n_h+1}{2}:n_h\right)}. \quad (4)$$

Lemma 1. $\overline{X}_{SDMRSS}$ is an unbiased estimator of the mean of symmetric distribution.

Proof. Two cases are considered:

First: If the sample sizes in the strata n_h, $h = 1,2,\ldots, L$ are even:

$$\overline{X}^*_{SDMRSS1} = \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\left(\sum_{i=1}^{\frac{n_h}{2}} X^*_{hi\left(\frac{n_h}{2}:n_h\right)} + \sum_{i=\frac{n_h}{2}+1}^{n_h} X^*_{hi\left(\frac{n_h+2}{2}:n_h\right)}\right)$$

$$E\left(\overline{X}^*_{SDMRSS1}\right) = E\left[\sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\left(\sum_{i=1}^{\frac{n_h}{2}} X^*_{hi\left(\frac{n_h}{2}:n_h\right)} + \sum_{i=\frac{n_h}{2}+1}^{n_h} X^*_{hi\left(\frac{n_h+2}{2}:n_h\right)}\right)\right]$$
$$= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\left(\sum_{i=1}^{\frac{n_h}{2}} E\left(X^*_{hi\left(\frac{n_h}{2}:n_h\right)}\right) + \sum_{i=\frac{n_h}{2}+1}^{n_h} E\left(X^*_{hi\left(\frac{n_h+2}{2}:n_h\right)}\right)\right)$$
$$= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\left(\sum_{i=1}^{\frac{n_h}{2}} \mu^*_{h\left(\frac{n_h}{2}\right)} + \sum_{i=\frac{n_h}{2}+1}^{n_h} \mu^*_{h\left(\frac{n_h+2}{2}\right)}\right).$$

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

$$E\left(\overline{X}^*_{SDMRSS1}\right) = \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\sum_{i=1}^{n_h} \mu^*_{hi}.$$

Based on Al-Saleh and Al-Kadiri (2000) we have $\mu_h = \frac{1}{n_h}\sum_{i=1}^{n_h}\mu_{hi}^*$, therefore:

$$E\left(\overline{X}^*_{SDMRSS1}\right) = \sum_{h=1}^{L} W_h \cdot \mu_h = \mu.$$

Second: If the sample sizes in the strata n_h, h = 1,2,…, L are odd:

$$\begin{aligned} E\left(\overline{X}^*_{SDMRSS2}\right) &= E\left(\sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\sum_{i=1}^{n_h} X^*_{hi\left(\frac{n_h+1}{2}:n_h\right)}\right) \\ &= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\sum_{i=1}^{n_h} E\left(X^*_{hi\left(\frac{n_h+1}{2}:n_h\right)}\right) \\ &= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\sum_{i=1}^{n_h} \mu^*_{h\left(\frac{n_h+1}{2}\right)}. \end{aligned}$$

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

$$\begin{aligned} E\left(\overline{X}^*_{SDMRSS2}\right) &= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}\sum_{i=1}^{n_h} \mu^*_{h\left(\frac{n_h+1}{2}\right)} \\ &= \sum_{h=1}^{L} W_h \cdot \frac{1}{n_h}(n_h)(\mu_h) \\ &= \sum_{h=1}^{L} W_h \cdot \mu_h = \mu. \end{aligned}$$

Lemma 2. If the parent distribution is symmetric about μ, then $Var\left(\overline{X}^*_{SDMRSS}\right) < Var\left(\overline{X}_{SRS}\right)$.

Proof. Two cases are considered.

First: If the sample sizes in the strata n_h, h = 1,2,…, L are even. The variance of $\overline{X}_{SDMRSS1}$ is:

$$Var\left(\overline{X}^*_{SDMRSS1}\right) = \sum_{h=1}^{L} \frac{W_h^2}{n_h^2}\left(\sum_{i=1}^{\frac{n_h}{2}} \sigma^{2*}_{hi\left(\frac{n_h}{2}:n_h\right)} + \sum_{i=\frac{n_h}{2}+1}^{n_h} \sigma^{2*}_{hi\left(\frac{n_h+2}{2}:n_h\right)}\right).$$

Nevertheless, $\sigma^{2*}_{hi} < \sigma^2_h$ for each stratum h = 1,2,…, L, this implies:

$$\begin{aligned} Var\left(\overline{X}^*_{SDMRSS1}\right) &= \sum_{h=1}^{L} W_h^2 \cdot \frac{1}{n_h}\sigma^{2*}_{hi} \\ &< \sum_{h=1}^{L} W_h^2 \cdot \frac{1}{n_h}\sigma^2_h \\ &= Var\left(\overline{X}_{SSRS}\right) \\ &< Var\left(\overline{X}_{SRS}\right). \end{aligned}$$

Second: The proof in case of odd sample sizes is similar.

Simulation Study Based on SDMRSS

In this section, a simulation study is conducted to investigate the performance of SDMRSS for estimating the population mean. Symmetric and asymmetric distributions are considered for samples of sizes n = 9, 12, 14, 15, 18. Assume that the population is partitioned into two or three strata and proportional allocation is considered. The simulation is performed for the SRSS, SSRS and SRS data sets from different distributions. The symmetric distributions are uniform and normal and the asymmetric distributions are exponential, gamma and Weibull. Using 100000 replications, estimates of the means, variances and Mean Squared Errors (MSE) are computed.

If the distribution is symmetric, then the efficiency of SDMRSS relative to T is defined by:

$$eff\left(\overline{X}^*_{SDMRSS}, \overline{X}_T\right) = \frac{Var\left(\overline{X}_T\right)}{Var\left(\overline{X}^*_{SDMRSS}\right)}, \quad (5)$$

where, T = SRS, $SSRS$, $SRSS$.

But if the distribution is asymmetric, the efficiency is defined as:

$$eff\left(\overline{X}^*_{SDMRSS}, \overline{X}_T\right) = \frac{MSE\left(\overline{X}_T\right)}{MSE\left(\overline{X}^*_{SDMRSS}\right)}. \quad (6)$$

The results are summarized in Table 1-5.

Based on the tables, we can conclude that gain in efficiency is attained using SDMRSS method as relative to other methods considered in this study to estimate the population mean of the variable of interest.

Table 1. The efficiency of SDMRSS relative to SRSS, SSRS and SRS for n = 9 and n = 12 with two strata

	n = 9, n_1 = 4 and n_2 = 5			n = 12, n_1 = 5 and n_2 = 7		
	SRSS	SSRS	SRS	SRSS	SSRS	SRS
Uniform (0,1)	24.9532	29.0126	29.3264	38.1847	39.3721	40.8392
Normal (0,1)	22.5374	24.3527	24.5821	34.6814	36.3885	35.6294
Exponential (1)	19.2431	17.6932	17.7357	23.3628	24.5624	22.9326
Gamma (1,2)	17.4071	17.9437	17.8426	23.4775	24.2247	23.4635
Weibull (1,2)	17.3286	17.0184	16.9452	22.7541	22.8352	22.3715

Table 2. The efficiency of SDMRSS relative to SRSS, SSRS and SRS for n = 14 and n = 18 with two strata

	$n = 14$, n_1, 8 and $n_2 = 6$			$n = 18$, $n_1 = 8$ and $n_2 = 10$		
	SRSS	SSRS	SRS	SRSS	SSRS	SRS
Uniform (0,1)	43.5647	47.8914	45.5718	60.8657	66.2043	63.2537
Normal (0,1)	39.4729	41.7583	38.7629	45.3256	49.0021	49.3312
Exponential (1)	21.8473	23.1638	21.5547	22.6374	21.3546	21.9903
Gamma (1,2)	23.2713	22.9246	22.3222	22.4617	22.3258	22.4366
Weibull (1,2)	21.2374	21.3621	21.2300	23.7332	22.9557	29.8471

Table 3. The efficiency of SDMRSS relative to SRSS, SSRS and SRS for n = 15 and n = 18 with three strata

	$n = 15$, $n_1 = 3$, $n_2 = 5$, $n_3 = 7$			$n = 18$, $n_1 = 4$, $n_2 = 6$, $n_3 = 8$		
	SRSS	SSRS	SRS	SRSS	SSRS	SRS
Uniform (0,1)	50.5472	52.3169	51.7284	63.3704	67.3274	65.4118
Normal (0,1)	35.643	38.3769	37.54	51.7528	52.5461	52.3271
Exponential (1)	24.0326	23.9002	23.2035	24.7361	25.3261	23.1043
Gamma (1,2)	21.4375	22.2438	21.8793	25.0143	24.2637	23.9927
Weibull (1,2)	22.3258	22.1002	22.3726	24.8933	23.0143	23.8436

Table 4. The values of bias of SDMRSS for different distributions and different numbers of strata

Sample size	No. of strata	Exp (1)	Gamma (1, 2)	Weibull (1, 2)
$n = 9$, $n_1 = 4$, $n_2 = 5$	2	0.0565	0.1776	0.7564
$n = 12$, $n_1 = 5$, $n_2 = 7$	2	0.0271	0.1989	0.3143
$n = 14$, $n_1 = 8$, $n_2 = 6$	2	0.0168	0.1033	0.0276
$n = 18$, $n_1 = 10$, $n_2 = 8$	2	0.0372	0.1101	0.0335
$n = 15$, $n_1 = 3$, $n_2 = 5$, $n_3 = 7$	3	0.0106	0.0754	0.0021
$n = 18$, $n_1 = 4$, $n_2 = 6$, $n_3 = 8$	3	0.0323	0.1011	0.0235

Table 5. The efficiency of SSRS, SRSS and SDMRSS relative to SRS for marks

Sampling method		$n = 7$, $n_1 = 4$, $n_2 = 3$	$n = 12$, $n_1 = 4$, $n_2 = 7$	$n = 14$, $n_1 = 8$, $n_2 = 6$
SRS	Mean	61.741	61.625	61.503
	Variance	57.314	42.592	38.728
SSRS	Mean	61.452	61.784	62.018
	Variance	32.416	28.715	22.825
	Efficiency	3.761	3.829	3.648
SRSS	Mean	62.331	62.072	61.912
	Variance	19.463	16.873	14.625
	Efficiency	3.984	4.533	4.327
SDMRSS	Mean	61.802	61.745	61.623
	Variance	13.426	11.039	10.819
	Efficiency	4.813	4.967	4.878

When the performances of the suggested SDMRSS estimators are compared, the efficiency of the suggested estimator is found to be more superior when the underlying distributions are symmetric as compared to asymmetric distributions.

The relative efficiency of SDMRSS estimator with respect to those estimators based on SRS, SSRS and SRSS are increasing as the sample size increases.

Real Data Example using SDMRSS

The marks of 787 students from scientific majors in Foundation Program at Qatar University during academic semester Fall 2011 are collected as a population to calculate its mean and variance. 100,000 samples using each of stratified double median ranked set sample, simple random sample, stratified ranked set sample and stratified simple random sample methods with sample size n = 7, 12, 14 using Mat lab 7 are generated. The mean and variance are obtained for each method and compared to evaluate the performance of SDMRSS to estimate the population mean for real data. Stratification is done according to the gender (males and females) and proportional allocation is considered.

Let x_{1i}, i = 1,2,…531 be the mark of ith female student in the population. Let x_{2j}, j = 1, 2,…, 256 be the mark of jth male student in the population.

The mean μ and the variance σ^2 of the population are:

$$\mu = \frac{1}{787}\sum_{i=1}^{787} x_i = 61.98 \text{ and } \sigma^2 = \frac{1}{787}\sum_{i=1}^{787}\left(x_i - \mu\right)^2 = 1060.86$$

The mean μ_1 and the variance σ_1^2 of the female population are:

$$\mu_1 = \frac{1}{531}\sum_{i=1}^{531} x_{1i} = 65.08 \text{ and } \sigma_1^2 = \frac{1}{531}\sum_{i=1}^{531}\left(x_{1i} - \mu\right)^2 = 1025.85$$

The mean μ_2 and the variance σ_2^2 of the male population respectively are:

$$\mu_2 = \frac{1}{256}\sum_{j=1}^{256} x_{2j} = 55.55 \text{ and } \sigma_2^2 = \frac{1}{256}\sum_{j=1}^{256}\left(x_{2j} - \mu\right)^2 = 1076.29$$

The skewness and median of the 787 students are -0.64 and 71.00. Skewness and median of 531 female students are -0.74 and 75.00. Skewness and median of 256 male students are -0.48 and 65.00.

Since the skewness for all students, female students and male students are -0.64, -0.74, -0.48, then the marks data are asymmetrically distributed, which means that SDMRSS estimator is biased and hence the mean squared errors of the estimators will be calculated. The efficiency of SSRS, SRSS and SDMRSS with respect to SRS are computed and summarized in Table 5.

From Table 5, the following are noticed:

- The values of estimated mean using SDMRSS are very close to the population mean
- The variance of SDMRSS estimator is less than the variances of SRS, SSRS and SRSS. Therefore, the efficiency values using SDMRSS are greater than those obtained using SRS, SSRS and SRSS
- Results in this real life example agrees with the theoretical results

Conclusion

In this study, new estimators of the population mean are suggested using SDMRSS. The performance of the estimators based on SDMRSS are compared with those using SRSS, SSRS and SRS for the same number of measured units. It is shown that SDMRSS estimators are unbiased of the population mean and are more efficient than their counterparts using SRSS, SSRS and SRS. Therefore, the SDMRSS is recommended for estimating the mean of symmetric and asymmetric distributions.

Acknowledgement

The author thanks the editors and the anonymous referees for their helpful comments that substantially improved this paper.

Author's Contributions

The work is the result of a full collaboration of the authors. However:

Mahmoud Ibrahim Syam: Carried out the simulations and application.

Amer Ibrahim Al-Omari: Participated in the theoretical part and discussion of this paper.

Kamarulzaman Ibrahim: Participated in reviewing the manuscript and giving ideas.

Ethics

This paper is original and has not published elsewhere.

References

Al-Omari, A.I. and K. Jaber, 2008. Percentile double ranked set sampling. J. Math. Statist., 4: 60-64. DOI: 10.3844/jmssp.2008.60.64

Al-Saleh, M.F. and M. Al-Kadiri, 2000. Double-ranked set sampling. Statist. Probab. Lett., 48: 205-212. DOI: 10.1016/S0167-7152(99)00206-0

Al-Saleh, M.F. and A.I. Al-Omari, 2002. Multistage ranked set sampling. J. Statist. Plann. Inference, 102: 273-286. DOI: 10.1016/S0378-3758(01)00086-6

Bouza, C.N., 2002. Ranked set sub sampling the non response strata for estimating the difference of means. Biometrical. J., 44: 903-915. DOI: 10.1002 /1521-4036(200210)44:7<903::AID-BIMJ9 03>3.0.CO;2-A

Dell, T.R. and J.L. Clutter, 1972. Ranked set sampling theory with order statistics background. Biometrika, 28: 545-555. DOI: 10.2307/2556166

Jemain, A.A. and A.I. Al-Omari, 2006. Double quartile ranked set samples. Pak. J. Statist., 22: 217-228.

Jemain, A.A., A.I. Al-Omari and K. Ibrahim, 2007. Multistage median ranked set sampling for estimating the population median. J. Math. Statist., 3: 58-64. DOI: 10.3844/jmssp.2007.58.64

McIntyre, G.A., 1952. A method for unbiased selective sampling, using ranked sets. Aus. J. Agric. Res., 3: 385-390. DOI: 10.1071/AR9520385

Muttlak, H.A., 1997. Median ranked set sampling. J. Applied Statis. Sci., 6: 577-586.

Ohyama, T., J.A. Doi and T. Yanagawa, 2008. Estimating population characteristics by incorporating prior values in stratified random sampling/ranked set sampling. J. Statist. Plann. Inference, 138: 4021-4032. DOI: 10.1016/j.jspi.2008.02.003

Samawi, H.M., 1996. Stratified ranked set sample. Pak. J. Statist., 12: 9-16.

Samawi, H.M. and E.M. Tawalbeh, 2002. Double median ranked set sample: Comparing to other double ranked samples for mean and ratio estimators. J. Mod. Applied Stat. Meth., 1: 428-442. DOI: 10.22237/jmasm/1036109460

Syam, M.I., K. Ibrahim and A.I. Al-Omari, 2013. The efficiency of stratified double percentile ranked set sample for estimating the population mean. Far East J. Math. Sci., 73: 157-177.

Takahasi, K. and K. Wakimoto, 1968. On unbiased estimates of the population mean based on the sample stratified by means of ordering. Ann. Institute Statist. Math., 20: 1-31. DOI: 10.1007/ BF02911622