Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal. 1(2008), 1 – 15
ISSN 2070-5948, DOI 10.1285/i20705948v1n1p1
http://siba2.unile.it/ese/ejasa
http://faculty.yu.edu.jo/alnasser/EJASA.htm

SOME VARIATIONS OF RANKED SET SAMPLING

Abdul Aziz Jemain[(1)] , **Amer Al-Omari**[(2)*] and **Kamarulzaman Ibrahim**[(1)]

[(1)] *School of Mathematical Sciences, Faculty of Science and Technology University Kebangsaan Malaysia, Malaysia.*
[(2)] *Department of Mathematics, Jerash Private University, Jordan.*

Received 3 March 2008; accepted 6 July 2008
Available online 18 August 2008

Abstract. *Balanced groups ranked set samples method (BGRSS) is suggested for estimating the population mean with samples of size* $m = 3k$ *where* $(k = 1,2,...)$. *The BGRSS sample mean is considered as an estimator of the population mean. It is found that the BGRSS produces unbiased estimators with smaller variance than the commonly used simple random sampling (SRS) for symmetric distributions considered in this study. For asymmetric distributions that we considered, the BGRSS estimators have a small bias. A real data set is used to illustrate the BGRSS method.*

Keywords. *Simple random sampling; ranked set sampling; balanced groups ranked set sampling.*

1. Introduction

The RSS suggested by McIntyre [5] for estimating mean pasture yields was found to have greater efficiency than SRS. He also suggested that this method is particularly suitable where the experimental or sampling units in a study can be more easily ranked than quantified. To obtain a sample of size m using RSS, randomly select m simple random samples each of size m from the target population and rank the units within each sample with respect to a variable of interest. The i th smallest rank united of the i th sample $(i = 1,2,...,m)$ is drawn and measured. This method is repeated n times if needed to obtain a sample of size mn out

* Corresponding author. Email: alomari_amer@yahoo.com; Tel: +962777906433.

of m^2n units. Takahasi and Wakimoto [11] proposed the same method with the mathematical theory of ranked set sampling. Dell and Clutter [2] showed that the mean of the RSS is an unbiased estimator of the population mean, whether there are errors in ranking or not. Samawi et al. [11] investigated the extreme ranked set samples (ERSS) for estimating a population mean. Muttlak [8] suggested using median ranked set sampling (MRSS) to estimate the population mean. Muttlak [6, 7] suggested quartile ranked set sampling (QRSS) and percentile ranked set sampling (PRSS) for estimating the population mean and showed that PRSS and QRSS produced unbiased estimators of the population mean when the underlying distribution is symmetric. Jemain and Al-Omari [4] suggested double quartile ranked set sampling (DQRSS) for estimating the population mean and showed that the mean based on DQRSS is an unbiased estimator and more efficient than those based on SRS, RSS and QRSS if the underlying distribution is symmetric. Details about RSS can be found in several works (see Al-Saleh and Al-Omari [1], Jemain and Al-Omari [3] and Ozturk and Deshpande [9]).This paper is presented as follows: in Section 2, we describe the BGRSS and illustrate two cases as examples. In Section 3, we derive the BGRSS estimators for the population mean for two cases when the sample size is odd or even. In addition, we study the properties of these estimators. In Section 4, results based on the uniform, normal and logistic distributions are provided. Simulation study using BGRSS for several distributions is presented in Section 5. This is followed by a real data set to illustrate the BGRSS, as given in Section 6. Finally, we summarize our results in Section 7.

2. Descriptions of BGRSS

The balanced groups ranked set sampling (BGRSS) can be described as follows:
Step 1: Randomly select $m = 3k$ $(k = 1,2,...)$ sets each of size m from the target population, and rank the units within each set with respect to the variable of interest.
Step 2:Allocate the $3k$ selected sets randomly into three groups, each of size k sets.
Step 3: For each group in step (2), select for measurement the lowest ranked unit from each set in the first group, and the median unit from each set in the second group, and the largest ranked unit from each set in the third group.
By this way we have a measured sample of size $m = 3k$ units in one cycle. The Steps 1-3 can be repeated n times to increase the sample size to $3kn$ out of $9k^2n$ units.
The BGRSS method differs from the usual RSS and ERSS methods. In the usual RSS we identify and measure the ith smallest ranked unit of the ith sample $(i = 1,2,...,m)$. In the case when m is odd, for ERSS we select the smallest ranked unit from the first $\frac{m-1}{2}$ sets and the largest ranked unit from the other $\frac{m-1}{2}$ sets. In the case when m is even we select the smallest ranked unit from the first $\frac{m}{2}$ sets and the largest ranked unit from the other $\frac{m}{2}$ sets. But in the BGRSS method, the measured units consist of $\frac{m}{3}$ minima, $\frac{m}{3}$ medians and $\frac{m}{3}$ maxima.
Indeed, the BGRSS method is easy to be applied since we only need to identify and measure the lowest rank units of the first k sets, and the medians of the second k sets, and the largest rank units from the last k sets. Here, k is any positive integer. However, for practical

purposes, k should be small in order to have a small sample size, so that the ranking is easy and errors in ranking is reduced. Let us consider the following example to illustrate BGRSS for estimating the population mean.

Example

Case 1: Let $k=1$, so $m=3$. Then we may have 3 sets of SRS each of size 3, as follows:

$$\{X_{11}, X_{12}, X_{13}\}, \{X_{21}, X_{22}, X_{23}\}, \{X_{31}, X_{32}, X_{33}\}.$$

After ranking the units with respect to a variable of interest allocate them into three groups where each contains one set of size three units as shown below:

First group,

$$A_1 = \{X_{1(1:3)}, X_{1(2:3)}, X_{1(3:3)}\},$$

Second group,

$$A_2 = \{X_{2(1:3)}, X_{2(2:3)}, X_{2(3:3)}\},$$

Third group,

$$A_3 = \{X_{3(1:3)}, X_{3(2:3)}, X_{3(3:3)}\}.$$

Now, select the smallest rank unit form the first group, the median from the second group, and the largest rank unit from the third group as:

$$X_{1(1:3)} = min(A_1),\ X_{2(2:3)} = median(A_2),\ X_{3(3:3)} = max(A_3).$$

The final set $\{X_{1(1:3)}, X_{2(2:3)}, X_{3(3:3)}\}$ is the BGRSS of size 3. These units are used for estimating the mean μ of the variable of interest as:

$$\hat{\mu}_{BGRSSO} = \frac{X_{1(1:3)} + X_{2(2:3)} + X_{3(3:3)}}{3}.$$

It is of interest to note that if $k=1$, the BGRSSO is the same as the usual RSS method in the case of estimating the population mean.

Case 2: If $k=2$, then $m=6$. So, we have six SRS sets each of size six. We first rank the unit within each set with respect to a variable of interest, and then allocate them into 3 groups where each contains two sets each of size six. After ranking and applying the BGRSS method, the sets appear as shown below:

$$\text{Firstgroup}\begin{cases} A_1 = \{(X_{1(1:6)}), X_{1(2:6)}, X_{1(3:6)}, X_{1(4:6)}, X_{1(5:6)}, X_{1(6:6)}\} \\ A_2 = \{(X_{2(1:6)}), X_{2(2:6)}, X_{2(3:6)}, X_{2(4:6)}, X_{2(5:6)}, X_{2(6:6)}\} \\ \end{cases}$$

$$\text{Secondgroup}\begin{cases} A_3 = \{X_{3(1:6)}, X_{3(2:6)}, (X_{3(3:6)}), (X_{3(4:6)}), X_{3(5:6)}, X_{3(6:6)}\} \\ A_4 = \{X_{4(1:6)}, X_{4(2:6)}, (X_{4(3:6)}), (X_{4(4:6)}), X_{4(5:6)}, X_{4(6:6)}\} \end{cases}$$

$$\text{Thirdgroup}\begin{cases} A_5 = \{X_{5(1:6)}, X_{5(2:6)}, X_{5(3:6)}, X_{5(4:6)}, X_{5(5:6)}, (X_{5(6:6)})\} \\ A_6 = \{X_{6(1:6)}, X_{6(2:6)}, X_{6(3:6)}, X_{6(4:6)}, X_{6(5:6)}, (X_{6(6:6)})\} \end{cases}$$

Finally, the set $\left\{X_{1(1:6)}, X_{2(1:6)}, \frac{1}{2}(X_{3(3:6)} + X_{3(4:6)}), \frac{1}{2}(X_{4(3:6)} + X_{4(4:6)}), X_{5(6:6)}, X_{6(6:6)}\right\}$ is a BGRSS of size 6, which can be used for estimating the population mean as:

$$\hat{\mu}_{BGRSSE} = \frac{1}{6}\left(X_{1(1:6)} + X_{2(1:6)} + \frac{1}{2}(X_{3(3:6)} + X_{3(4:6)}) + \frac{1}{2}(X_{4(3:6)} + X_{4(4:6)}) + X_{5(6:6)} + X_{6(6:6)} \right)$$

In the following section, we shall give some notations and introduce an estimator of the population mean using BGRSS.

3. Estimation of the population mean using BGRSS

Let $X_1, X_2, ..., X_m$ be a random sample with probability density function $f(x)$, with mean μ, and variance σ^2. Let $X_{11}, X_{12}, ..., X_{1m}; X_{21}, X_{22}, ..., X_{2m}; ...; X_{m1}, X_{m2}, ..., X_{mm}$ be independent random variables all with the same cumulative distribution function $F(x)$. If m is odd, let $X_{i(1:m)}$ be the lowest rank unit of the ith sample $(i = 1,2,...,k)$, and $X_{i\left(\frac{m+1}{2}:m\right)}$ be the median of the ith sample $(i = k+1, k+2, ..., 2k)$, and let $X_{i(m:m)}$ be the largest rank unit of the ith sample $(i = 2k+1, 2k+2, ..., 3k)$. Note that, the measured units, $X_{1(1:m)}, X_{2(1:m)}, ..., X_{k(1:m)}$ are iid, $X_{k+1\left(\frac{m+1}{2}:m\right)}, ..., X_{2k\left(\frac{m+1}{2}:m\right)}$ are iid and $X_{2k+1(m:m)}, ..., X_{3k(m:m)}$ are iid. However, all units are mutually independent but not identically distributed and will be denoted as the measured BGRSSO. The BGRSSO estimator of the population mean can be defined as:

$$\hat{\mu}_{BGRSSO} = \frac{1}{3k}\left(\sum_{i=1}^{k} X_{i(1:m)} + \sum_{i=k+1}^{2k} X_{i\left(\frac{m+1}{2}:m\right)} + \sum_{i=2k+1}^{3k} X_{i(m:m)} \right), \tag{1}$$

with variance

$$\mathrm{Var}(\hat{\mu}_{BGRSSO}) = \frac{1}{9k^2}\left(\sum_{i=1}^{k} \mathrm{Var}(X_{i(1:m)}) + \sum_{i=k+1}^{2k} \mathrm{Var}\left(X_{i\left(\frac{m+1}{2}:m\right)} \right) + \sum_{i=2k+1}^{3k} \mathrm{Var}(X_{i(m:m)}) \right). \tag{2}$$

In the case of even sample size, let $X_{i(1:m)}$ be the lowest rank unit of the ith sample $(i = 1,2,...,k)$, and $\frac{1}{2}\left(X_{i\left(\frac{m}{2}:m\right)} + X_{i\left(\frac{m+2}{2}:m\right)} \right)$ be the median of the ith sample $(i = k+1, k+2,...,2k)$, and let $X_{i(m:m)}$ be the largest rank unit of the ith sample $(i = 2k+1, 2k+2,...,3k)$. Note that, $X_{1(1:m)}, X_{2(1:m)},..., X_{k(1:m)}$ are iid, $\frac{1}{2}\left(X_{k+1\left(\frac{m}{2}:m\right)} + X_{k+1\left(\frac{m+2}{2}:m\right)} \right)$, $\frac{1}{2}\left(X_{k+2\left(\frac{m}{2}:m\right)} + X_{k+2\left(\frac{m+2}{2}:m\right)} \right),..., \frac{1}{2}\left(X_{2k\left(\frac{m}{2}:m\right)} + X_{2k\left(\frac{m+2}{2}:m\right)} \right)$ are iid and $X_{2k+1(m:m)}$, $X_{2k+2(m:m)}$,..., $X_{3k(m:m)}$ are iid. However, all measured units are mutually independent and not identically distributed and will be denoted as the measured BGRSSE. The BGRSSE estimator of the population mean is defined as:

$$\hat{\mu}_{BGRSSE} = \frac{1}{3k}\left(\sum_{i=1}^{k} X_{i(1:m)} + \sum_{i=k+1}^{2k}\left(\frac{1}{2}\left(X_{i\left(\frac{m}{2}:m\right)} + X_{i\left(\frac{m+2}{2}:m\right)} \right)\right) + \sum_{i=2k+1}^{3k} X_{i(m:m)} \right), \tag{3}$$

with variance

$$\operatorname{Var}\left(\hat{\mu}_{BGRSSE}\right) = \frac{1}{9k^2}\left(\sum_{i=1}^{k}\operatorname{Var}\left(X_{i(1:m)}\right) + \frac{1}{4}\sum_{i=k+1}^{2k}\left(\operatorname{Var}\left(X_{i\left(\frac{m}{2}:m\right)} \right) + \operatorname{Var}\left(X_{i\left(\frac{m+2}{2}:m\right)} \right) + 2\operatorname{Cov}\left(X_{i\left(\frac{m}{2}:m\right)}, X_{i\left(\frac{m+2}{2}:m\right)} \right) \right) + \sum_{i=2k+1}^{3k}\operatorname{Var}\left(X_{i(m:m)}\right) \right). \tag{4}$$

If the underlying distribution is symmetric about μ, then $X_{(i:m)} \overset{d}{=} -X_{(m-i+1:m)}$, so that $E\left(X_{(i:m)}\right) = -E\left(X_{(m-i+1:m)}\right)$ and $\operatorname{Var}\left(X_{(i:m)}\right) = \operatorname{Var}\left(X_{(m-i+1:m)}\right)$ for all i, $(i = 1,2,...,m)$, (see David and Nagaraja 2003). Based on these results and since the measured k units in each group are iid, we have:

$$E\left(\hat{\mu}_{BGRSSO}\right) = 0, E\left(\hat{\mu}_{BGRSSE}\right) = 0, \tag{5}$$

and Equations (3) and (4) will be respectively as

$$\operatorname{Var}\left(\hat{\mu}_{BGRSSO}\right)=\frac{1}{9k}\left(2\operatorname{Var}\left(X_{(1:m)}\right)+\operatorname{Var}\left(X_{\left(\frac{m+1}{2}:m\right)}\right)\right), \tag{6}$$

and

$$\operatorname{Var}\left(\hat{\mu}_{BGRSSE}\right)=\frac{1}{9k}\left(\begin{array}{l}2\operatorname{Var}\left(X_{(1:m)}\right)+\frac{1}{2}\operatorname{Var}\left(X_{\left(\frac{m}{2}:m\right)}\right)\\ +\frac{1}{2}\operatorname{Cov}\left(X_{\left(\frac{m}{2}:m\right)},X_{\left(\frac{m+2}{2}:m\right)}\right)\\ \\ \end{array}\right). \tag{7}$$

The properties of $\hat{\mu}_{BGRSS}$ are:

1. If the underlying distribution is symmetric about the population mean μ, then
    - a) $\hat{\mu}_{BGRSS}$ is an unbiased estimator of the population mean μ.
    - b) $\operatorname{Var}\left(\hat{\mu}_{BGRSS}\right)<\operatorname{Var}\left(\hat{\mu}_{SRS}\right)$.
2. If the underlying distribution is asymmetric about μ, then the mean square error of $\hat{\mu}_{BGRSS}$ is less than the variance of $\hat{\mu}_{SRS}$ for some distributions considered in this study, specially with small sample size.

In the following section we will illustrate the BGRSS method for estimating the population mean of the uniform, normal and logistic distributions.

4. Results for Some Selected Distributions

The efficiency of $\hat{\mu}_{BGRSS}$ with respect to $\hat{\mu}_{SRS}$ for estimating the population mean is defined as:

$$eff\left(\hat{\mu}_{BGRSS},\hat{\mu}_{SRS}\right)=\frac{\operatorname{Var}\left(\hat{\mu}_{SRS}\right)}{\operatorname{Var}\left(\hat{\mu}_{BGRSS}\right)}. \tag{8}$$

The SRS estimator of the population mean from a sample of size m has the variance given by

$$\operatorname{Var}\left(\hat{\mu}_{SRS}\right)=\frac{\sigma^{2}}{m}. \tag{9}$$

4.1 Uniform distribution

If $X_1, X_2, ..., X_m$ constitute a random sample from standard uniform distribution, then we have two cases.

First: When m is odd, the minimum $X_{(1:m)}$ has beta distribution with parameters $(1, m)$. So that

$$F_{(1:m)}(x) = B_{1:m}[F(x)], \tag{10}$$

with mean and variance, respectively, are

$$E(X_{(1:m)}) = \frac{1}{m+1} \text{ and } \mathrm{Var}(X_{(1:m)}) = \frac{m}{(m+1)^2(m+2)}. \tag{11}$$

The median $X_{\left(\frac{m+1}{2}:m\right)}$ has beta distribution with parameters $\left(\frac{m+1}{2}, \frac{m+1}{2}\right)$, so that

$$F_{\left(\frac{m+1}{2}:m\right)}(x) = B_{\frac{m+1}{2}:\frac{m+1}{2}}[F(x)], \tag{12}$$

and the mean and variance, respectively, are

$$E\left(X_{\left(\frac{m+1}{2}:m\right)}\right) = \frac{1}{2} \text{ and } \mathrm{Var}\left(X_{\left(\frac{m+1}{2}:m\right)}\right) = \frac{1}{4(m+2)}. \tag{13}$$

Finally, the maximum $X_{(m:m)}$ has beta distribution with parameters $(m,1)$. Therefore

$$F_{(m:m)}(x) = B_{m:m}[F(x)], \tag{14}$$

$$E(X_{(m:m)}) = \frac{m}{m+1} \text{ and } \mathrm{Var}(X_{(m:m)}) = \frac{m}{(m+1)^2(m+2)}. \tag{15}$$

From (11), (13) and (15), we have $E(\hat{\mu}_{BGRSSO}) = \frac{1}{2}$. Hence, $\hat{\mu}_{BGRSSO}$ is an unbiased estimator with variance given by

$$\mathrm{Var}(\hat{\mu}_{BGRSSO}) = \frac{m^2 + 10m + 1}{12m(m+1)^2(m+2)}. \tag{16}$$

From (8), (9) and (16), the efficiency of $\hat{\mu}_{BGRSSO}$ with respect to $\hat{\mu}_{SRS}$ is given by

$$eff(, \hat{\mu}_{BGRSSO}, \hat{\mu}_{SRS}) = \frac{(m+1)^2(m+2)}{m^2 + 10m + 1}. \tag{17}$$

For example, if $m=3$, we have $eff\left(\hat{\mu}_{SRS},\hat{\mu}_{BGRSSO}\right)=\frac{80}{40}=2$. Note that when, $m=3$, the efficiency value is equal to that obtained using the usual RSS method. For $m=9$, we have $eff\left(\hat{\mu}_{BGRSSO},\hat{\mu}_{SRS}\right)=\frac{29700}{4644}=6.395$.

Second: If m is even, then $X_{\left(\frac{m}{2}:m\right)}$ has beta distribution with parameters $\left(\frac{m}{2},\frac{m+2}{2}\right)$. So that

$$F_{\left(\frac{m}{2}:m\right)}(x)=B_{\frac{m}{2}:m}\left[F(x)\right], \tag{18}$$

and

$$E\left(X_{\left(\frac{m}{2}:m\right)}\right)=\frac{m}{2(m+1)} \text{ and } \mathrm{Var}\left(X_{\left(\frac{m}{2}:m\right)}\right)=\frac{m}{4(m+1)^2}. \tag{19}$$

Also, $X_{\left(\frac{m+2}{2}:m\right)}$ has beta distribution with parameters $\left(\frac{m+2}{2},\frac{m}{2}\right)$. Therefore,

$$F_{\left(\frac{m+2}{2}:m\right)}(x)=B_{\frac{m+2}{2}:m}\left[F(x)\right]. \tag{20}$$

Thus,

$$E\left(X_{\left(\frac{m+2}{2}:m\right)}\right)=\frac{m+2}{2(m+1)}\, and\, \mathrm{Var}\left(X_{\left(\frac{m+2}{2}:m\right)}\right)=\frac{m}{4(m+1)^2}. \tag{21}$$

The variance of the median given by

$$\mathrm{Var}\left(\frac{1}{2}\left(X_{\left(\frac{m}{2}:m\right)}+X_{\left(\frac{m+2}{2}:m\right)}\right)\right)=\frac{m}{4(m+2)(m+1)}. \tag{22}$$

Based on Equations (19), (21) and (22), it can be shown that

$$\mathrm{Var}\left(\hat{\mu}_{BGRSSE}\right)=\frac{m+9}{12(m+1)^2(m+2)}. \tag{23}$$

From (8), (9) and (23) the efficiency of $\hat{\mu}_{BGRSSE}$ with respect to $\hat{\mu}_{SRS}$ can be given by

$$eff\left(\hat{\mu}_{BGRSSE},\hat{\mu}_{SRS}\right)=\frac{(m+1)^2(m+2)}{m(m+9)}. \tag{24}$$

For example, if $m = 6$, $eff(\hat{\mu}_{SRS}, \hat{\mu}_{BGRSSE}) = \frac{392}{90} = 4.356$.

Now we will compare the $\hat{\mu}_{BGRSS}$ estimators with $\hat{\mu}_{RSS}$ based on the same number of measured units. For uniform (0,1), we know that

$$\text{Var}(\hat{\mu}_{RSS}) = \frac{1}{6m(m+1)}. \tag{25}$$

From (16) and (25), the efficiency of $\hat{\mu}_{BGRSSO}$ with respect to $\hat{\mu}_{RSS}$ is given by

$$eff(\hat{\mu}_{BGRSSO}, \hat{\mu}_{RSS}) = \frac{\text{Var}(\hat{\mu}_{RSS})}{\text{Var}(\hat{\mu}_{BGRSSO})} = \frac{2(m+1)(m+2)}{m^2+10m+1}. \tag{26}$$

It is clear that $\frac{2(m+1)(m+2)}{m^2+10m+1} > 1$. For even sample size, from (23) and (25) the efficiency of $\hat{\mu}_{BGRSSE}$ with respect to $\hat{\mu}_{RSS}$ is given by

$$eff(\hat{\mu}_{BGRSSE}, \hat{\mu}_{RSS}) = \frac{\text{Var}(\hat{\mu}_{RSS})}{\text{Var}(\hat{\mu}_{BGRSSE})} = \frac{2(m+1)(m+2)}{m(m+9)}. \tag{27}$$

It is easy to show that $\frac{2(m+1)(m+2)}{m(m+9)} > 1$. From Equations (26) and (27) it is obvious that BGRSS is more efficient than RSS in estimating the mean of the standard uniform distribution.

4.2 Normal distribution

Let $X_1, X_2, ..., X_9$ constitute a random sample from normal population with mean 0 and variance 1. Let us define the error function $Erf(z)$ to be the integral of the Gaussian distribution as given by $Erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. The random variable $X_{(1:9)}$ has the cdf

$$F_{(1:9)}(x) = 1 + \frac{1}{512}\left(-1 + Erf\left[\frac{x}{\sqrt{2}}\right]\right)^9,$$

with mean $E(X_{(1:9)}) = -1.48501$ and variance $\text{Var}(X_{(1:9)}) = 0.357353$. Also, the median $X_{(5:9)}$ has cdf

$$F_{(5:9)}(x) = \frac{1}{256}\left(1+Erf\left[\frac{x}{\sqrt{2}}\right]\right)^5 \left(\begin{array}{l} 128-325Erf\left[\frac{x}{\sqrt{2}}\right]+345Erf\left[\frac{x}{\sqrt{2}}\right]^2 \\ -175Erf\left[\frac{x}{\sqrt{2}}\right]^3+35Erf\left[\frac{x}{\sqrt{2}}\right]^4 \end{array} \right),$$

with mean $E\left(X_{(5:9)}\right)=0$ and $\mathrm{Var}\left(X_{(5:9)}\right)=0.166101$. The maximum $X_{(9:9)}$ has the cdf

$$F_{(9:9)}(x) = \frac{1}{512}\left(1+Erf\left[\frac{x}{\sqrt{2}}\right]\right)^9,$$

with mean $E\left(X_{(1:9)}\right)=1.48501$ and $\mathrm{Var}\left(X_{(9:9)}\right)=0.357353$.

It is clear that $E\left(\hat{\mu}_{BGRSSO}\right)=0$ and the variance $\mathrm{Var}\left(\hat{\mu}_{BGRSSO}\right)=0.0326225$. By Equation (9) the *SRS* estimator of normal mean from a sample of size 9 has the variance $\mathrm{Var}\left(\hat{\mu}_{SRS}\right)=0.11111$. Therefore the efficiency of BGRSSO with respect to SRS is given by

$$eff\left(\hat{\mu}_{BGRSSO},\hat{\mu}_{SRS}\right) = \frac{0.11111}{0.0326225} = 3.406.$$

4.3 Logistic distribution

If $X_1, X_2, ..., X_9$ constitute a random sample from logistic distribution with parameters 0 and 1, then the random variable $X_{(1:9)}$ has the cdf

$$F_{(1:9)}(x) = 1-\left(1-\frac{1}{1+e^{-x}}\right)^9,$$

with mean $E\left(X_{(1:9)}\right)=-2.71786$ and $\mathrm{Var}\left(X_{(1:9)}\right)=1.76245$. Also, the median $X_{(5:9)}$ has the cdf given by

$$F_{(5:9)}(x) = \frac{1}{\left(1+e^{x}\right)^9}\left\{e^{5x}\left[126+e^{x}\left(84+e^{x}\left(36+e^{x}\left(9+e^{x}\right)\right)\right)\right]\right\},$$

mean $E\left(X_{(5:9)}\right)=0$, $\mathrm{Var}\left(X_{(5:9)}\right)=0.442646$, and the maximum $X_{(9:9)}$ has

$$F_{(9:9)}(x) = \frac{1}{\left(1+e^{-x}\right)^9},$$

with mean $E\left(X_{(1:9)}\right)=2.71786$ and $\mathrm{Var}\left(X_{(9:9)}\right)=1.76245$.

It is clear that $E\left(\hat{\mu}_{BGRSSO}\right)=0$ and variance $\mathrm{Var}\left(\hat{\mu}_{BGRSSO}\right)=0.146946$. From Equation (9), the *SRS* estimator of logistic mean from a sample of size 9 has the variance

$\mathrm{Var}(\hat{\mu}_{SRS}) = 0.36554$. Therefore, from (8) we have

$$eff(\hat{\mu}_{BGRSSO}, \hat{\mu}_{SRS}) = \frac{0.36554}{0.146946} = 2.487.$$

We can see that the $BGRSS$ method is more efficient than the SRS for estimating the mean of the uniform distribution, normal distribution and logistic distribution based on the same number of measured units.

5. Simulation study

In this section, we compare the efficiency of the proposed estimators of the population mean using $BGRSS$ method relative to SRS method. Three symmetric distributions, namely, uniform, normal and logistic, and also three asymmetric distributions, exponential, beta and gamma are considered. We compare the average of 70,000 sample estimates using $k = 1,2,...,7$ corresponding to the sample sizes $m = 3,6,...,21$ respectively. If the distribution is symmetric the efficiency of the $BGRSS$ relative to SRS can be obtained using Equation (8). But if the distribution is asymmetric the efficiency is defined as:

$$eff(\hat{\mu}_{BGRSS}, \hat{\mu}_{SRS}) = \frac{\mathrm{Var}(\hat{\mu}_{SRS})}{\mathrm{MSE}(\hat{\mu}_{BGRSS})}, \tag{28}$$

where the $\mathrm{MSE}(\hat{\mu}_{BGRSS})$ is the mean square error of the $\hat{\mu}_{BGRSS}$, and $\mathrm{MSE}(\hat{\mu}_{BGRSSO}) = \mathrm{Var}(\hat{\mu}_{BGRSSO}) + (E(\hat{\mu}_{BGRSSO}) - \mu)^2$. Results of the efficiency and bias values are given in Table 1. Based on Table 1, we may conclude the following:

(1) Gain in efficiency is obtained by using BGRSS compared to SRS for estimating the population mean for different values of m if the underlying distribution is symmetric about its mean. For example, for $m = 9$, the efficiency of the BGRSS is 6.420 for estimating the population mean of a uniform distribution $U(0,1)$.

(2) When comparing the efficiencies obtained for the symmetric distributions considered in this study, the BGRSS is most efficient for estimating the mean of the uniform distribution.

(3) For asymmetric distributions considered in this study, the BGRSS mean estimator has a small bias. For example, if the distribution is $B(7,4)$ with sample size $m = 9$, the efficiency of BGRSS is 3.684 with bias 0.006.

6. Example for real data set

In this section, a collection of a real data set is used to illustrate the BGRSS for estimating the population mean. These data consists of height (H), and the weight (W) of 348 students. Table 2 contains the summary statistics of the data. We will show how to use the RSS and BGRSS to estimate the mean of the students height based on their weight. It is known that the coefficient of skewness should be close to zero for symmetrically distributed data. But since the coefficient of skewness of the weight and height are 1.1954 and 0.0289, respectively, so that these data are asymmetrically distributed.

To illustrate the RSS and BGRSS methods, we first fix the values of height (H) and do the ranking based on the values of the weight (W). Let $k=2$, then $m=6$. For estimating the mean of height consider the following steps:

Step 1: Randomly select 6 independent simple random samples each of size 6 ordered pairs (W,H) as:

Set 1:{(47, 159),(35, 146),(60, 145),(37, 144),(61, 173),(34, 139)}
Set 2:{(33, 146), (36, 137), (52, 170), (48, 160), (73, 162), (62, 165)}
Set 3:{(54, 160), (56, 165), (39, 151), (37, 155), (55, 147), (46, 160)}
Set 4:{(40, 148), (68, 164), (51, 156), (36, 149), (60, 160), (95, 162)}
Set 5:{(40, 154), (54, 156), (33, 141), (34, 141), (37, 142), (41, 158)}
Set 6:{(50, 158), (56, 166), (87, 155), (37, 144), (40, 138), (49, 163)}.

Step 2: For each set in Step 1, rank the pairs within each set based on their weight (in bold) from the lowest to highest as shown below:

Set 1: (**34**,139),(35,146),(37,144),(47,159),(60,145),(61,173)}
Set 2: (**33**,146),(36,137),(48,160),(52,170),(62,165),(73,162)}
Set 3: (37,155),(39,151),(**46**,160),(**54**,160),(55,147),(56,165)}
Set 4: (36,149),(40,148),(**51**,156),(**60**,160),(68,164),(95,162)}
Set 5: (33,141),(34,141),(37,142),(40,154),(41,158),(**54**,156)}
Set 6: (37,144),(40,138),(49,163),(50,158),(56,166),(**87**,155)}

Step 3: Now, we will consider the SRS, RSS and BGRSS as:
1. Under SRS, we have 6 estimates of the mean. Let $\hat{\mu}_{H,SRSi}$ be the mean of the *ith* set $(i=1,2,...,6)$. So we have:

$$\hat{\mu}_{H,SRS1}=151,\ \hat{\mu}_{H,SRS2}=156.66,\ \hat{\mu}_{H,SRS3}=156.33,\ \hat{\mu}_{H,SRS4}=156.5,\ \hat{\mu}_{H,SRS5}=148.66,\ \hat{\mu}_{H,SRS6}=154,$$

2. Under RSS method, from the ith set, select and measure the height corresponding to the ith ordered weight values. The six RSS height are: 139,137,160,160,158 and 155. Hence, the RSS estimator of the mean is given by:

$$\hat{\mu}_{H,RSS}=\frac{139+137+160+160+158+155}{6}=\frac{909}{6}=151.5$$

3. Under BGRSS, the lowest ranked units are measured from the first two sets, the median is measured from the third and fourth sets and largest ranked units are measured from the last two sets. Thus, height values considered are: 139, 146, 157.5, 158, 156, 155. The mean height can be estimated using BGRSS as:

$$\hat{\mu}_{H,BGRSS}=\frac{139+146+157.5+158+156+155}{6}=\frac{911.5}{6}=151.92.$$

Table 3 provide the means, variances and the efficiencies based on 50,000 simulated values of $\hat{\mu}_{H,SRS}$ and $\hat{\mu}_{H,BGRSS}$. From Table 3, the estimated mean is found to be close to the real value of the population mean. Also, we can see that the BGRSS method is more efficient than the SRS for estimating the population mean of the height.

7. Summary

A gain in efficiency is obtained using BGRSS for estimating the population mean. It is found that BGRSS is more appropriate for estimating the population mean of symmetric distributions than asymmetric distributions considered in this study. Thus, it is recommended to use BGRSS for estimating the population mean of symmetric distribution, also for estimating the mean of symmetric distributions when the sample size is small since the bias is very negligible.

Appendix: Tables

Table 1 - The efficiency values for estimating the population mean using BGRSS with respect to SRS with *m*= 3,6,…,21.

Distribution		$m=3$	$m=6$	$m=9$	$m=12$	$m=15$	$m=18$	$m=21$
Uniform (0,1)	*Eff*	2.000	4.265	6.420	9.439	11.567	14.977	17.155
Normal (0,1)	*Eff*	1.917	2.883	3.515	3.984	4.343	4.655	4.925
Logistic (0,1)	*Eff*	1.841	2.281	2.524	2.661	2.785	2.812	2.823
Exponential (1)	*Eff*	1.638	1.484	0.972	0.597	0.387	0.265	0.194
	Bias	0.000	0.135	0.229	0.309	0.369	0.424	0.470
Beta (7,4)	*Eff*	1.994	3.087	3.684	3.957	3.843	3.550	3.143
	Bias	0.000	0.004	0.006	0.008	0.010	0.011	0.012
Gamma (2,1)	*Eff*	1.759	1.926	1.456	0.935	0.636	0.451	0.332
	Bias	0.000	0.140	0.241	0.328	0.394	0.451	0.501

Table 2 - Summary statistics of 348 students data.

	Mean	*Variance*	*Skewness*
Weight (W) in kg	50.3017	275.7560	1.1954
Height (H) in cm	152.3250	131.0560	0.0289
Correlation coefficient	0.6775		

Table 3 - Summary results of estimating the population mean of the height of 348 students using BGRSS method with *m*= 3,6,…,18.

Size	**SRS**		**BGRSS**		**Efficiency**
m	*Mean*	*MSE*	*Mean*	*MSE*	
3	152.353	43.4037	152.359	31.3387	1.38496
6	152.332	21.4179	151.623	10.1066	2.11918
9	152.328	14.1717	151.278	6.59716	2.14815
12	152.323	10.6051	151.048	5.23777	2.02473
15	152.313	6.94952	150.702	4.57425	1.51924
18	152.339	5.82804	150.586	4.56858	1.27563

References

Al-Saleh M. F. and Al-Omari A. I. Multistage ranked set sampling. Journal of Statistical Planning and Inference. 2002;102:273--286.

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

Jemain A. A., and Al-Omari, A. I. Double percentile ranked set samples for estimating the population mean. Advances and Applications in Statistics. 2006a;6(3):261--276.

Jemain A. A. and Al-Omari A. I. Double quartile ranked set samples. Pakistan Journal of Statistics. 2006b;22(3):217--228.

McIntyre G. A. A method for unbiased selective sampling using ranked sets. Australian Journal of Agricultural Research. 1952;3:385--390.

Muttlak H. A. Investigating the use of quartile ranked set samples for estimating the population mean. Journal of Applied Mathematics and Computation. 2003a;146:437--443.

Muttlak H. A. Modified ranked set sampling methods. Pakistan Journal of Statistics. 2003b;19(3):315--323.

Muttlak H. A. Median ranked set sampling, Journal of Applied Statistical Sciences. 1997;6(4):577--586.

Ozturk O. and Deshpande J.V. Ranked Set Sample nonparametric quantile confidence intervals. Journal of Statistical Planning and Inference. 2006;136:570--577.

Samawi H, Abu-Dayyeh W and Ahmed S. Extreme ranked set sampling, The Biometrical Journal. 1996;30:577--586.

Takahasi K., and Wakimoto K. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. Annals of the Institute of Statistical Mathematics. 1968;20:1--31.