On the population median estimation using quartile double ranked set sampling

Amer Ibrahim Al-Omari
Faculty of Science, Department of Mathematics
Al al-Bayt University, Mafraq, Jordan
alomari_amer@yahoo.com

Loai M. Al-Zubi
Faculty of Science, Department of Mathematics
Al al-Bayt University, Mafraq, Jordan

Ahmad Khazaleh
Faculty of Science, Department of Mathematics
Al al-Bayt University, Mafraq, Jordan

Abstract

In this article, quartile double ranked set sampling (QDRSS) method is considered for estimating the population median. The sample median based on QDRSS is suggested as an estimator of the population median. The QDRSS is compared with the simple random sampling (SRS), ranked set sampling (RSS) and quartile ranked set sampling (QRSS) methods for estimating the population median. To verify this method a real data example is applied. It turns out that for the symmetric distributions considered in this study, the QDRSS estimators are unbiased estimators of the population median and are larger than their counterparts using SRS, RSS and QRSS based on the same sample size of measured units. For asymmetric distributions, QDRSS is biased. It is more efficient than the SRS and the QRSS for all samples of size *m* while it is more efficient than RSS if $m > 4$.

***Keywords*:** Simple random sampling; Quartile ranked set sampling; Ranked set sampling; Quartile double ranked set sampling; Median.

1. Introduction

Ranked set sampling was first suggested by McIntyre (1952) as a cost efficient sampling procedure when compared to the commonly used simple random sampling in situations where visual ordering of set units can be done easily, but the exact measurement of the units is difficult and expensive. McIntyre (1952) found that the RSS is more efficient than SRS for estimating the population mean.

Let *X* be a random variable with a probability density function (pdf) $f(x)$, and a cumulative distribution function (cdf) $F(x)$ with mean μ and variance σ^2. Also, let $f_{(i:m)}(x)$ be the pdf of the *i*th order statistic of a random sample of size *m*, $X_{i1}, X_{i2}, ..., X_{im}$ for $i = 1, 2, ..., m$. Then, the pdf of $X_{(i:m)}$ is given by

$$f_{(i:m)}(x) = \frac{1}{B(i, m-i+1)} \frac{d}{dx} \int_0^{F(x)} u^{i-1} \left[1-u\right]^{m-i} \, f(x),$$

where $B(\theta,\tau)=\int_0^1 u^{\theta-1}(1-u)^{\tau-1}du, \theta>0, \tau>0$, with mean $\mu_{(i:m)}=\int_{-\infty}^{\infty} x f_{(i:m)}(x)dx$ and variance $\sigma^2_{(i:m)}=\int_{-\infty}^{\infty}\left(x-\mu_{(i:m)}\right)^2 f_{(i:m)}(x)dx$, David and Nagaraja (2003).

Takahasi and Wakimoto (1968) provided the necessary mathematical theory of RSS. They showed that

$$f(x)=\frac{1}{m}\sum_{i=1}^{m} f_{(i:m)}(x),\ \mu=\frac{1}{m}\sum_{i=1}^{m}\mu_{(i:m)} \text{ and } \sigma^2=\frac{1}{m}\sum_{i=1}^{m}\sigma^2_{(i:m)}+\frac{1}{m}\sum_{i=1}^{m}\left(\mu_{(i:m)}-\mu\right)^2.$$

Muttlak (1997) suggested median ranked set sampling for estimating the population mean. Al-Saleh and Al-Kadiri (2000) considered double ranked set sampling (DRSS) method for estimating the population mean, and they showed that the ranking at the second stage is easier than the ranking at the first stage.

The double ranked set sampling method can be described as follows: Randomly identify m^3 units from the target population and divide them randomly into *m* sets each of size m^2. The procedure of ranked set sampling is applied to these sets to obtain *m* ranked set samples each of size *m*, again reapply the ranked set sampling procedure on the *m* ranked set samples to obtain a DRSS of size *m*.

Al-Saleh and Al-Omari (2002) generalized the DRSS to multistage ranked set sampling to increase the efficiency of the estimators for specific value of the sample size. Muttlak (2003) proposed quartile ranked set sampling (QRSS) for estimating the population mean. Al-Omari and Al-Saleh (2009) suggested quartile double ranked set sampling (QDRSS) for estimating the population mean. Al-Omari (2010) suggested an estimator of the population median using double robust extreme ranked set sampling. Entropy estimation and goodness-of-fit tests for the inverse Gaussian and Laplace distributions using paired ranked set sampling method is suggested by Al-Omari and Haq (2015). Biradar and Santosha (2015) proposed estimation of the population mean using paired ranked set sampling. Santos and Barrios (2015) considered predictive accuracy of logistic regression model using ranked set samples. Confidence intervals and hypothesis tests for a population mean using ranked set sampling are considered by Stella et al. (2015). For more about RSS and its modifications see Sinha et al. (2006), Ozturk and Jozani (2014), Hatefi et al. (2014), Samawi and Al-Saleh (2014), Bouza (2002), and Tiwari and Pandey (2013).

2. Estimation of the population median

2.1. *Using SRS*

Let $X_1, X_2, ..., X_m$ be a random sample of size *m* from a distribution with pdf $f(x)$, cdf $F(x)$, mean μ, median η and variance σ^2.

The SRS estimator of the population median from a sample of size *m* at the *h*th cycle $(h=1,2,...,n)$ is defined as

$$\hat{\eta}_{SRS}=\begin{cases} X_{\left(\frac{m+1}{2}\right)h} & \text{, if } m \text{ is odd} \\ \frac{1}{2}\left(X_{\left(\frac{m}{2}\right)h}+X_{\left(\frac{m+2}{2}\right)h}\right) & \text{, if } m \text{ is even.} \end{cases} \tag{1}$$

Based on $f_{(i:m)}(x)$ if *m* is odd, the pdf of $X_{\left(\frac{m+1}{2}\right)}$ is given by

$$f_{\left(\frac{m+1}{2}\right)}(x)=\frac{m!}{\left[\left(\frac{m+1}{2}\right)!\right]^{2}}\left[F(x)\left(1-F(x)\right)\right]^{\frac{m-1}{2}} f(x), \tag{2}$$

and if *m* is even

$$f_{\left(\frac{m}{2}\right)}(x)=\frac{m!}{\left(\frac{m-2}{2}\right)!\left(\frac{m}{2}\right)!}\left[F(x)\right]^{\frac{m-2}{2}}\left[1-F(x)\right]^{\frac{m}{2}} f(x), \tag{3}$$

and

$$f_{\left(\frac{m+2}{2}\right)}(x)=\frac{m!}{\left(\frac{m}{2}\right)!\left(\frac{m-2}{2}\right)!}\left[F(x)\right]^{\frac{m}{2}}\left[1-F(x)\right]^{\frac{m-2}{2}} f(x). \tag{4}$$

2.2. *Using RSS*

The RSS (McIntyre, 1952) involves randomly selecting m^2 units from the population. These units are randomly allocated into *m* sets, each of size *m*. The *m* units of each sample are ranked visually or by any inexpensive method with respect to a 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. The process can be repeated *n* times to get a sample of size *mn* from the initial m^2n units.

Let $X_{11h}, X_{12h}, ..., X_{1mh}$; $X_{21h}, X_{22h}, ..., X_{2mh}$; …; $X_{m1h}, X_{m2h}, ..., X_{mmh}$ be *m* independent simple random samples each of size *m* in the *h*th cycle $\left(h=1,2,...,n\right)$. Let $X_{i(1)h}$, $X_{i(2)h}$, …, $X_{i(m)h}$ be the order statistics of the *i*th sample $X_{i1h}, X_{i2h}, ..., X_{imh}$ for $i=1,2,...,m$. Therefore, $X_{1(1)h}, X_{2(2)h}, \ldots, X_{m(m)h}$ denote the measured RSS units.

The RSS estimator of the population median η from a sample of size *m* at the *h*th cycle $\left(h=1,2,...,n\right)$ is given by

$$\hat{\eta}_{RSS}=\text{Median}\left\{X_{1(1)h}, X_{2(2)h}, ..., X_{m(m)h}\right\}. \tag{5}$$

2.3. *Using QRSS*

The QRSS procedure, suggested by Muttlak (2003) is described as follows: select *m* random samples each of size *m* units from the target population and rank the units within each sample with respect to the variable of interest. If the sample size is even, select for measurement from the first $m/2$ samples the $q_1(m+1)$th smallest ranked unit and from the second $m/2$ samples the $q_3(m+1)\text{th}$ smallest ranked unit, where $q_1 = 0.25$ and $q_3 = 0.75$, where the nearest integers of $q_1(m+1)\text{th}$ and $q_3(m+1)\text{th}$ will always be taken. If the sample size is odd, select from the first $(m-1)/2$ samples the $q_1(m+1)\text{th}$ smallest ranked unit and from the other $(m-1)/2$ samples the $q_3(m+1)\text{th}$ smallest ranked unit, and from one sample the median of that sample for actual measurement. The procedure can be repeated *n* times if needed to increase the sample size to *nm* units.

If the sample size is even, at the *h*th cycle $(h = 1, 2, ..., n)$, let $X^*_{i(q_1(m+1))h}$ be the first quartile of the *i*th sample $\left(i = 1, 2, ..., \frac{m}{2}\right)$, and $X^*_{i(q_3(m+1))h}$ be the third quartile of the *i*th sample $\left(i = \frac{m+2}{2}, \frac{m+4}{2}, ..., m\right)$. Therefore, the measured QRSSE units are $X^*_{1(q_1(m+1))h}, ...,$ $X^*_{\frac{m}{2}(q_1(m+1))h}$, $X^*_{\frac{m+2}{2}(q_3(m+1))h}$, ..., $X^*_{m(q_3(m+1))h}$. The QRSSE estimator of the population median is given by

$$\hat{\eta}^*_{QRSSE} = \text{Median}\left\{X^*_{1(q_1(m+1))h}, ..., X^*_{\frac{m}{2}(q_1(m+1))h}, X^*_{\frac{m+2}{2}(q_3(m+1))h}, ..., X^*_{m(q_3(m+1))h}\right\}, \tag{6}$$

If the sample size m is odd, let $X^*_{i(q_1(m+1))}$ be the first quartile of the *i*th sample $\left(i = 1, 2, ..., \frac{m-1}{2}\right)$, $X^*_{i\left(\frac{m+1}{2}\right)h}$ be the median of the *i*th sample of the rank $i = \frac{m+1}{2}$, and let $X^*_{i(q_3(m+1))h}$ be the third quartile of the *i*th sample $\left(i = \frac{m+3}{2}, \frac{m+5}{2}, ..., m\right)$. Therefore, the measured QDRSSO units are $X^*_{1(q_1(m+1))h}, ...,$ $X^*_{\frac{m-1}{2}(q_1(m+1))h}$, $X^*_{\frac{m+1}{2}\left(\frac{m+1}{2}\right)h}$, $X^*_{\frac{m+3}{2}(q_3(m+1))h}, ...,$ $X^*_{m(q_3(m+1))h}$. The estimator of the population median using QRSSO is defined as

$$\hat{\eta}^*_{QRSSO} = \text{Median}\left\{X^*_{1(q_1(m+1))h}, ..., X^*_{\frac{m-1}{2}(q_1(m+1))h}, X^*_{\frac{m+1}{2}\left(\frac{m+1}{2}\right)h}, X^*_{\frac{m+3}{2}(q_3(m+1))h}, ..., X^*_{m(q_3(m+1))h}\right\} \tag{7}$$

2.4 . *Using QDRSS*

The quartile double ranked set sampling method (Al-Omari and Al-Saleh, 2009) can be carried out as follows:

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

Step 2: Rank the units within each set with respect to the variable of interest, and then 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, apply the QRSS method on the *m* DRSS sets obtained in Step 2. The whole process can be repeated n times if needed to get a sample of size mn from the QDRSS data. For even and odd sample sizes we denote the measured QDRSS units as QDRSSE and QDRSSO, respectively.

Let us consider the following example. Select a random sample of size $m = 8$, so we will select $m^3 = 512$ units. Allocate them into 8 sets each of 64 units. Rank the units within each set with respect to the variable of interest. Let X_{jik} be the *i*th unit $(i = 1, 2, ..., 8)$ in the *j*th set $(j = 1, 2, ..., 8)$ in the *k*th subset $(k = 1, 2, ..., 8)$. Select the $X_{j(i)k}$ from each subset, the processes appears as shown below:

$$\underbrace{\begin{bmatrix} \left(X_{1(1)1}\right), X_{1(2)1}, ..., X_{1(7)1}, X_{1(8)1} \\ X_{1(1)2}, \left(X_{1(2)2}\right), ..., X_{1(7)2}, X_{1(8)2} \\ \vdots \qquad\qquad\qquad \vdots \\ X_{1(1)8}, X_{1(2)8}, ..., \left(X_{1(7)8}\right), X_{1(8)8} \\ X_{1(1)1}, X_{1(2)1}, ..., X_{1(7)1}, \left(X_{1(8)1}\right) \end{bmatrix}}_{\text{The 1st set of size 64 units}}, \ldots, \underbrace{\begin{bmatrix} \left(X_{8(1)1}\right), X_{8(2)1}, ..., X_{8(7)1}, X_{8(8)1} \\ X_{8(1)2}, \left(X_{8(2)2}\right), ..., X_{8(7)2}, X_{8(8)2} \\ \vdots \qquad\qquad\qquad \vdots \\ X_{8(1)7}, X_{8(2)7}, ..., \left(X_{8(7)7}\right), X_{8(8)7} \\ X_{8(1)8}, X_{8(2)8}, ..., X_{8(7)8}, \left(X_{8(8)8}\right) \end{bmatrix}}_{\text{The 8th set of size 64 units}}$$

Select the *i*th smallest ranked unit from the *i*th subset $(i = 1, 2, .., 8)$ in each set. This step yields 64 units, i.e., 8 RSS sets each of size 8 as follows:

$$\begin{array}{l} \left\{\left(X_{1(1)1}\right),\left(X_{1(2)2}\right),\left(X_{1(3)3}\right),\left(X_{1(4)4}\right),\left(X_{1(5)5}\right),\left(X_{1(6)6}\right),\left(X_{1(7)7}\right),\left(X_{1(8)8}\right)\right\}, \\ \left\{\left(X_{2(1)1}\right),\left(X_{2(2)2}\right),\left(X_{2(3)3}\right),\left(X_{2(4)4}\right),\left(X_{2(5)5}\right),\left(X_{2(6)6}\right),\left(X_{2(7)7}\right),\left(X_{2(8)8}\right)\right\}, \\ \left\{\left(X_{3(1)1}\right),\left(X_{3(2)2}\right),\left(X_{3(3)3}\right),\left(X_{3(4)4}\right),\left(X_{3(5)5}\right),\left(X_{3(6)6}\right),\left(X_{3(7)7}\right),\left(X_{3(8)8}\right)\right\}, \\ \left\{\left(X_{4(1)1}\right),\left(X_{4(2)2}\right),\left(X_{4(3)3}\right),\left(X_{4(4)4}\right),\left(X_{4(5)5}\right),\left(X_{4(6)6}\right),\left(X_{4(7)7}\right),\left(X_{4(8)8}\right)\right\}, \\ \left\{\left(X_{5(1)1}\right),\left(X_{5(2)2}\right),\left(X_{5(3)3}\right),\left(X_{5(4)4}\right),\left(X_{5(5)5}\right),\left(X_{5(6)6}\right),\left(X_{5(7)7}\right),\left(X_{5(8)8}\right)\right\}, \\ \left\{\left(X_{6(1)1}\right),\left(X_{6(2)2}\right),\left(X_{6(3)3}\right),\left(X_{6(4)4}\right),\left(X_{6(5)5}\right),\left(X_{6(6)6}\right),\left(X_{6(7)7}\right),\left(X_{6(8)8}\right)\right\}, \\ \left\{\left(X_{7(1)1}\right),\left(X_{7(2)2}\right),\left(X_{7(3)3}\right),\left(X_{7(4)4}\right),\left(X_{7(5)5}\right),\left(X_{7(6)6}\right),\left(X_{7(7)7}\right),\left(X_{7(8)8}\right)\right\}, \\ \left\{\left(X_{8(1)1}\right),\left(X_{8(2)2}\right),\left(X_{8(3)3}\right),\left(X_{8(4)4}\right),\left(X_{8(5)5}\right),\left(X_{8(6)6}\right),\left(X_{8(7)7}\right),\left(X_{8(8)8}\right)\right\}. \end{array}$$

Without doing any actual quantifications on these sets, rank the units within each set with respect to the variable of interest and then select the first quartile $X^{\nabla}_{i(2)1}$ from the *i*th set $(i = 1, 2, 3, 4)$ and select the third quartile $X^{\nabla}_{i(7)1}$ from the *i*th set $(i = 5, 6, 7, 8)$ as shown below:

$$\begin{array}{l} \left\{\left(X^{\nabla}_{1(1)1}\right),\boxed{\left(X^{\nabla}_{1(2)1}\right)},\left(X^{\nabla}_{1(3)1}\right),\left(X^{\nabla}_{1(4)1}\right),\left(X^{\nabla}_{1(5)1}\right),\left(X^{\nabla}_{1(6)1}\right),\left(X^{\nabla}_{1(7)1}\right),\left(X^{\nabla}_{1(8)1}\right)\right\}, \\ \left\{\left(X^{\nabla}_{1(1)2}\right),\boxed{\left(X^{\nabla}_{1(2)2}\right)},\left(X^{\nabla}_{1(3)2}\right),\left(X^{\nabla}_{1(4)2}\right),\left(X^{\nabla}_{1(5)2}\right),\left(X^{\nabla}_{1(6)2}\right),\left(X^{\nabla}_{1(7)2}\right),\left(X^{\nabla}_{1(8)2}\right)\right\}, \\ \left\{\left(X^{\nabla}_{1(1)3}\right),\boxed{\left(X^{\nabla}_{1(2)3}\right)},\left(X^{\nabla}_{1(3)3}\right),\left(X^{\nabla}_{1(4)3}\right),\left(X^{\nabla}_{1(5)3}\right),\left(X^{\nabla}_{1(6)3}\right),\left(X^{\nabla}_{1(7)3}\right),\left(X^{\nabla}_{1(8)3}\right)\right\}, \end{array}$$

$$\left\{\left(X^{\nabla}_{1(1)4}\right),\boxed{\left(X^{\nabla}_{1(2)4}\right)},\left(X^{\nabla}_{1(3)4}\right),\left(X^{\nabla}_{1(4)4}\right),\left(X^{\nabla}_{1(5)4}\right),\left(X^{\nabla}_{1(6)4}\right),\left(X^{\nabla}_{1(7)4}\right),\left(X^{\nabla}_{1(8)4}\right)\right\},$$

$$\left\{\left(X^{\nabla}_{1(1)5}\right),\left(X^{\nabla}_{1(2)5}\right),\left(X^{\nabla}_{1(3)5}\right),\left(X^{\nabla}_{1(4)5}\right),\left(X^{\nabla}_{1(5)5}\right),\left(X^{\nabla}_{1(6)5}\right),\boxed{\left(X^{\nabla}_{1(7)5}\right)},\left(X^{\nabla}_{1(8)5}\right)\right\},$$

$$\left\{\left(X^{\nabla}_{1(1)6}\right),\left(X^{\nabla}_{1(2)6}\right),\left(X^{\nabla}_{1(3)6}\right),\left(X^{\nabla}_{1(4)6}\right),\left(X^{\nabla}_{1(5)6}\right),\left(X^{\nabla}_{1(6)6}\right),\boxed{\left(X^{\nabla}_{1(7)6}\right)},\left(X^{\nabla}_{1(8)6}\right)\right\},$$

$$\left\{\left(X^{\nabla}_{1(1)7}\right),\left(X^{\nabla}_{1(2)7}\right),\left(X^{\nabla}_{1(3)7}\right),\left(X^{\nabla}_{1(4)7}\right),\left(X^{\nabla}_{1(5)7}\right),\left(X^{\nabla}_{1(6)7}\right),\boxed{\left(X^{\nabla}_{1(7)7}\right)},\left(X^{\nabla}_{1(8)7}\right)\right\},$$

$$\left\{\left(X^{\nabla}_{1(1)8}\right),\left(X^{\nabla}_{1(2)8}\right),\left(X^{\nabla}_{1(3)8}\right),\left(X^{\nabla}_{1(4)8}\right),\left(X^{\nabla}_{1(5)8}\right),\left(X^{\nabla}_{1(6)8}\right),\boxed{\left(X^{\nabla}_{1(7)8}\right)},\left(X^{\nabla}_{1(8)8}\right)\right\}.$$

This process produces $\left\{X^{\nabla}_{1(2)1},X^{\nabla}_{1(2)2},X^{\nabla}_{1(2)3},X^{\nabla}_{1(2)4},X^{\nabla}_{1(7)5},X^{\nabla}_{1(7)6},X^{\nabla}_{1(7)7},X^{v}_{1(7)8}\right\}$ as a QDRSSE of size 8. The median of these units can be considered as an estimator of the population median. It is defined as

$$\hat{\eta}^{\nabla}_{QDRSSE}=\text{Median}\left\{X^{\nabla}_{1(2)1},X^{\nabla}_{1(2)2},X^{\nabla}_{1(2)3},X^{\nabla}_{1(2)4},X^{\nabla}_{1(7)5},X^{\nabla}_{1(7)6},X^{\nabla}_{1(7)7},X^{\nabla}_{1(7)8}\right\}. \tag{8}$$

The most interesting thing here is that the number of quantified units using QDRSS is 8 which will be compared with a SRS of size 8 is a small relative to to the number of sampled units 512. Hence, the information contained in the QDRSS sample is more than the information in the 8 units of the SRS.

In the hth cycle $(h=1,2,...,n)$ if the sample size is even, let $X^{\nabla}_{i(q_1(m+1))h}$ be the first quartile of the ith sample $\left(i=1,2,...,\frac{m}{2}\right)$, and $X^{\nabla}_{i(q_3(m+1))h}$ be the third quartile of the ith sample $\left(i=\frac{m+2}{2},\frac{m+4}{2},...,m\right)$. Hence, the measured QDRSSE units are $X^{\nabla}_{1(q_1(m+1))h},...,$ $X^{\nabla}_{\frac{m}{2}(q_1(m+1))h}$, $X^{\nabla}_{\frac{m+2}{2}(q_3(m+1))h}$,..., $X^{\nabla}_{m(q_3(m+1))h}$. The suggested QDRSSE estimator of the population median is given by

$$\hat{\eta}^{\nabla}_{QDRSSE}=\text{Median}\left\{X^{\nabla}_{1(q_1(m+1))h},...,X^{\nabla}_{\frac{m}{2}(q_1(m+1))h},X^{\nabla}_{\frac{m+2}{2}(q_3(m+1))h},...,X^{\nabla}_{m(q_3(m+1))h}\right\}, \tag{9}$$

If the sample size m is odd, let $X^{\nabla}_{i(q_1(m+1))}$ be the first quartile of the ith sample $\left(i=1,2,...,\frac{m-1}{2}\right)$, and $X^{\nabla}_{i\left(\frac{m+1}{2}\right)h}$ be the median of the ith sample of the rank $i=\frac{m+1}{2}$, and $X^{\nabla}_{i(q_3(m+1))h}$ be the third quartile of the ith sample $\left(i=\frac{m+3}{2},\frac{m+5}{2},...,m\right)$. Therefore, the QDRSSO measured units are $X^{\nabla}_{1(q_1(m+1))h},...,X^{\nabla}_{\frac{m-1}{2}(q_1(m+1))h}$, $X^{\nabla}_{\frac{m+1}{2}\left(\frac{m+1}{2}\right)h}$, $X^{\nabla}_{\frac{m+3}{2}(q_3(m+1))h}$, $...,X^{\nabla}_{m(q_3(m+1))h}$. The suggested estimator of the population median using QDRSSO is defined as

$$\hat{\eta}^{\nabla}_{QDRSSO} = \text{Median}\left\{ X^{\nabla}_{1(q_1(m+1))h}, \ldots, X^{\nabla}_{\frac{m-1}{2}(q_1(m+1))h}, X^{\nabla}_{\frac{m+1}{2}\left(\frac{m+1}{2}\right)h}, X^{\nabla}_{\frac{m+3}{2}(q_3(m+1))h}, \ldots, X^{\nabla}_{m(q_3(m+1))h} \right\} \quad (10)$$

4. Simulation Study

In this section, a simulation study is considered to compare the proposed estimators for the population median using QDRSS, QRSS, RSS relative to SRS. Six probability distribution functions were considered for the populations: Uniform, Normal, Logistic, Exponential, Gamma and Weibull. 60,000 samples were generated and the averages of these samples were compared.

If the distribution is symmetric the efficiency of RSS, QRSS and QDRSS relative to SRS, is defined as, respectively,

$$eff\left(\hat{\eta}_{RSS}, \hat{\eta}_{SRS}\right) = \frac{\text{Var}\left(\hat{\eta}_{SRS}\right)}{\text{Var}\left(\hat{\eta}_{RSS}\right)},\ eff\left(\hat{\eta}_{QRSS}, \hat{\eta}_{SRS}\right) = \frac{\text{Var}\left(\hat{\eta}_{SRS}\right)}{\text{Var}\left(\hat{\eta}_{QRSS}\right)},$$

and

$$eff\left(\hat{\eta}^{\nabla}_{QDRSS}, \hat{\eta}_{SRS}\right) = \frac{\text{Var}\left(\hat{\eta}_{SRS}\right)}{\text{Var}\left(\hat{\eta}^{\nabla}_{QDRSS}\right)}.$$

If the distribution is asymmetric, the efficiency is defined by

$$eff\left(\hat{\eta}_{RSS}, \hat{\eta}_{SRS}\right) = \frac{\text{MSE}\left(\hat{\eta}_{SRS}\right)}{\text{MSE}\left(\hat{\eta}_{RSS}\right)},\ eff\left(\hat{\eta}_{QRSS}, \hat{\eta}_{SRS}\right) = \frac{\text{MSE}\left(\hat{\eta}_{SRS}\right)}{\text{MSE}\left(\hat{\eta}_{QRSS}\right)},$$

and

$$eff\left(\hat{\eta}^{\nabla}_{QDRSS}, \hat{\eta}_{SRS}\right) = \frac{\text{MSE}\left(\hat{\eta}_{SRS}\right)}{\text{MSE}\left(\hat{\eta}^{\nabla}_{QDRSS}\right)}, \text{ where } \text{MSE}(\Psi) = \text{Var}(\Psi) + \left[E(\Psi) - \eta\right]^2.$$

Results of simulation in terms of the efficiency and bias values for RSS, QRSS and DQRSS are summarized for $m = 4, 5$ in Table 1, for $m = 6, 7$ in Table 2, for $m = 10, 11$ in Table 3 and for $m = 12$ in Table 4.

Table 1: The efficiency and bias values of RSS, QRSS, and QDRSS with respect to SRS in estimating the population mean with $m = 4$ and 5.

Distribution		$m = 4$			$m = 5$		
		RSS	**QRSS**	**QDRSS**	**RSS**	**QRSS**	**QDRSS**
Uniform (0,1)	*Eff*	1.988	2.400	4.517	1.885	2.358	3.696
Normal (0,1)	*Eff*	2.206	1.979	3.016	2.101	2.680	4.440
Logistic (-1,1)	*Eff*	2.268	1.872	2.631	2.177	2.813	4.713
Exponential (1)	*Eff*	2.299	1.313	1.347	2.296	3.049	5.105
	Bias	0.094	0.218	0.268	0.043	0.034	0.021
Gamma (1,2)	*Eff*	2.314	1.311	1.381	2.319	3.058	5.025
	Bias	0.192	0.437	0.531	0.080	0.065	0.039
Weibull (1,3)	*Eff*	2.275	1.271	1.369	2.281	3.054	5.187
	Bias	0.029	0.662	0.804	0.127	0.105	0.061

Table 2: The efficiency and bias values of RSS, QRSS, and QDRSS with respect to SRS in estimating the population mean with $m=6$ and 7.

Distribution		$m=6$			$m=7$		
		RSS	**QRSS**	**QDRSS**	**RSS**	**QRSS**	**QDRSS**
Uniform (0,1)	*Eff*	2.382	2.816	6.111	2.226	2.041	2.886
Normal (0,1)	*Eff*	2.750	3.106	6.568	2.522	2.351	3.334
Logistic (-1,1)	*Eff*	2.756	3.190	6.682	2.551	2.368	3.397
Exponential (1)	*Eff*	2.945	3.114	5.850	2.670	2.437	3.674
	Bias	0.048	0.056	0.064	0.026	0.027	0.019
Gamma (1,2)	*Eff*	2.857	3.183	5.851	2.700	2.480	3.708
	Bias	0.098	0.110	0.130	0.054	0.062	0.042
Weibull (1,3)	*Eff*	2.877	3.159	5.764	2.647	2.501	3.637
	Bias	0.149	0.165	0.198	0.080	0.083	0.062

Table 3: The efficiency and bias values of RSS, QRSS, and QDRSS with respect to SRS in estimating the population mean with $m=10$ and 11.

Distribution		$m=10$			$m=11$		
		RSS	**QRSS**	**QDRSS**	**RSS**	**QRSS**	**QDRSS**
Uniform (0,1)	*Eff*	3.143	3.803	11.336	2.838	2.301	3.165
Normal (0,1)	*Eff*	3.567	4.057	11.553	3.258	2.502	3.620
Logistic (-1,1)	*Eff*	3.483	4.131	11.146	3.131	2.479	3.605
Exponential (1)	*Eff*	3.671	4.016	8.295	3.247	2.604	3.769
	Bias	0.019	0.032	0.055	0.013	0.018	0.012
Gamma (1,2)	*Eff*	3.637	4.045	8.252	3.323	2.603	3.758
	Bias	0.043	0.063	0.113	0.026	0.035	0.024
Weibull (1,3)	*Eff*	3.666	4.089	8.170	3.329	2.560	3.669
	Bias	0.062	0.095	0.166	0.040	0.052	0.038

Table 4: The efficiency and bias values of RSS, QRSS, and QDRSS with respect to SRS in estimating the population mean with respect to SRS with $m=12$.

Distribution		RSS	QRSS	QDRSSE
Uniform (0,1)	*Eff*	3.464	4.067	14.298
Normal (0,1)	*Eff*	3.953	4.139	12.492
Logistic (-1,1)	*Eff*	3.833	4.080	11.863
Exponential (1)	*Eff*	3.914	3.647	5.021
	Bias	0.017	0.047	0.094
Gamma (1,2)	*Eff*	3.997	3.697	4.985
	Bias	0.032	0.093	0.188
Weibull (1,3)	*Eff*	3.989	3.654	5.119
	Bias	0.044	0.140	0.283

According to these results, we conclude:

1) If the underlying distribution is symmetric about its mean, then

a) $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}^{\nabla}_{QDRSSO}$ are unbiased estimators of the population median with smaller variance than the $\hat{\eta}_{SRS}$ estimator based on the same sample size. As an example, for

$m = 7$, the efficiency of $\hat{\eta}^{\nabla}_{QDRSSO}$ is 3.334 for estimating the population median of the standard normal distribution.

b) $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}^{\nabla}_{QDRSSO}$ are more efficient than $\hat{\eta}_{RSS}$. For example, when $m = 11$ the efficiency of $\hat{\eta}^{\nabla}_{QDRSSO}$ and $\hat{\eta}_{RSS}$ are 3.165 and 2.838, respectively, for estimating the population median of the standard uniform distribution.

c) $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}^{\nabla}_{QDRSSO}$ are more efficient than $\hat{\eta}_{QRSS}$. For $m = 10$, the efficiency values of $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}_{QRSS}$ are 11.146 and 4.131, respectively for estimating the median of the Logistic distribution with parameters -1 and 1.

2) If the underlying distribution is asymmetric, we noted that

a) $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}^{\nabla}_{QDRSSO}$ have a small bias. As an example, for $m = 12$ the efficiency of $\hat{\eta}^{\nabla}_{QDRSSE}$ is 5.021 with bias 0.094 for estimating the median of the exponential distribution with parameter 1.

b) $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}^{\nabla}_{QDRSSO}$ are more efficient than $\hat{\eta}_{RSS}$ if $m > 4$ and they are more efficient than $\hat{\eta}_{QRSS}$ for all cases considered in this study based on the same number of measured units. For example, with $m = 10$ the efficiency values of $\hat{\eta}_{RSS}$, $\hat{\eta}_{QRSS}$ and $\hat{\eta}^{\nabla}_{QDRSSE}$ are, respectively, 3.666, 4.089 and 8.170 for estimating the median of Weibull distribution with parameters 1 and 3.

3) Comparing $\hat{\eta}^{\nabla}_{QDRSSE}$ to $\hat{\eta}^{\nabla}_{QDRSSO}$, it is found that $\hat{\eta}^{\nabla}_{QDRSSE}$ is more efficient. For example, for $m = 6$ and 7, the efficiency of $\hat{\eta}^{\nabla}_{QDRSSE}$ and $\hat{\eta}^{\nabla}_{QDRSSO}$ are, respectively, 6.568 and 3.334 for estimating the median of standard normal distribution. This may be due to that: in the case of odd sample size we select only the median of the set of the rank $i = \frac{m+1}{2}$, while with even sample size we identify the first or the third quartile of the *i*th sample.

5. Real Data Application

In this section, to evaluate the performance of QDRSS in estimating the population median of a real data, a study is conducted to estimate the median weight of 342 students. Balanced ranked set sampling is considered and all samples were done without replacement.

Let Z_i for $i = 1, 2, ..., 342$ be the weight of the ith student in the population. The mean μ, median η and the variance σ^2 of the population are, respectively

$$\mu = \frac{1}{342}\sum_{i=1}^{342} Z_i = 50.047\,\text{kg}, \quad \eta = \text{Median}\{Z_i, i = 1, 2, ..., 342\} = \frac{Z_{171} + Z_{172}}{2} = 48,$$

and

$$\sigma^2 = \frac{1}{342}\sum_{i=1}^{342}(Z_i - \mu)^2 = 258.93\,\text{kg}^2.$$

The skewness of the 342 observations is 1.244, which means that these data are asymmetrically distributed, and so the QDRSS estimators will be biased. Hence, the bias and mean squared error (MSE) of the estimators were computed. The efficiency of RSS, QRSS and QDRSS with respect to SRS are obtained using Equations (8), (9), and (10). The simulated median, bias, MSE and the efficiency values are summarized in Table 5.

Table 5: The efficiency and bias values of RSS, QRSS and QDRSS relative to SRS with sample size $m = 4, 5, 6, 7, 10, 11$ for estimating the median weight of 342 students.

Method		Sample size					
		$m=4$	$m=5$	$m=6$	$m=7$	$m=10$	$m=11$
SRS	Median	48.467	48.030	48.033	47.954	47.900	47.918
	Bias	0.467	0.030	0.033	-0.046	-0.101	-0.083
	MSE	64.180	56.268	41.996	40.156	26.608	26.084
RSS	Median	48.073	47.893	47.900	47.917	47.905	47.933
	Bias	0.073	-0.107	-0.100	-0.083	-0.094	-0.067
	MSE	27.461	26.494	15.026	17.016	8.120	9.382
	Efficiency	2.337	2.124	2.795	2.360	3.277	2.780
QRSS	Median	49.028	47.870	47.914	47.902	47.897	47.918
	Bias	1.028	-0.120	-0.087	-0.098	-0.103	-0.082
	MSE	39.385	20.756	13.201	18.095	6.666	11.873
	Efficiency	1.630	2.711	3.181	2.220	3.992	2.197
QDRSS	Median	49.286	47.916	47.856	47.897	47.798	47.939
	Bias	1.285	-0.084	-0.144	-0.103	-0.202	-0.061
	MSE	31.072	13.360	5.945	13.208	2.189	8.424
	Efficiency	2.066	4.212	7.064	3.040	12.155	3.096

Table 5 shows that there is a small difference between the true and the estimated median. This difference is due to skewness of the data used in this example. For $m = 4$, RSS is more efficient than QDRSS. While, QDRSS is more efficient than RSS. In addition, it can be noted that QDRSS is more efficient than QRSS for all sample sizes considered in Table 5. Furthermore, the results of real data example are agreed with the results of the simulation study conducted in Section 4.

6. Conclusion

In estimating the population median, a good achievement is gained in efficiency using QDRSS, QRSS, RSS regardless the underlying distribution whether it is symmetric or asymmetric. QDRSS estimators are unbiased estimators of the population median when distributions are symmetric. In addition, it is found that QDRSS is more efficient than RSS if $m > 4$ and more efficient than QRSS in all cases considered in this study. However, the QDRSS is recommended for estimating the population median of symmetric distributions.

Acknowledgement

The authors are thankful to the reviewer and the associate editor for valuable comments that significantly improved the current version of the article.

References

1. Al-Omari, A.I. (2010). Estimation of the population median of symmetric and asymmetric distributions using double robust extreme ranked set sampling. *Revista Investigación Operacional*, 31(3): 199-207.

2. Al-Omari, A.I. and Haq, A. (2015). Entropy estimation and goodness-of-fit tests for the inverse Gaussian and Laplace distributions using paired ranked set sampling. *Journal of Statistical Computation and Simulation*, accepted, DOI:10.1080/00949 655.2015. 1109097.

3. Al-Omari, A.I. and Jaber, K. (2008). Percentile double ranked set sampling. *Journal of Mathematics and Statistics*, 4(1): 60-64.

4. Al-Saleh, M.F. and Al-Kadiri, M.A. (2000). Double ranked set sampling. *Statistics and Probability Letters*, 48(2): 205-212.

5. Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. *Journal of Statistical Planning and Inference*, 102(2): 273-286.

6. Al-Omari, A.I. and Al-Saleh, M.F. (2009). Quartile double rankled set sampling for estimating the population mean. *Economic Quality Control*, 24(2): 243 – 253.

7. Balakrishnan, N. and Li, T. (2006). Confidence intervals for quantiles and tolerance intervals based on ordered ranked set samples. *Annals of the Institute of Statistical Mathematics*, 58: 757-777.

8. Biradar, B.S. and Santosha, C.D. (2015). Estimation of the population mean using paired ranked set sampling. *Open Journal of Statistics*, 5, 97-103.

9. Bouza, C.N. (2002). Ranked set subsampling the non-response strata for estimating the difference of means. *Biometrical Journal*, 44: 903–915.

10. David, H.A. and Nagaraja, H.N. (2003). Order Statistics, 3rd ed., John Wiley & Sons, Inc., Hoboken, NJ.

11. Hatefi, A., Jozani, M.J., and Ziou, D. (2014). Estimation and classification for finite mixture models under ranked set sampling. *Statistica Sinica*, 24: 675-698.

12. McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3: 385-390.

13. Muttlak, H.A. (1997). Median ranked set sampling. *Journal of Applied Statistical Sciences*, 6(4): 245-255.

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

15. Ozurk, O. and Jozani, M.J. (2014). Inclusion probabilities in partially rank ordered set sampling. *Computational Statistics and Data Analysis*, 69: 122-132.

16. Samawi, H.M. and Al-Saleh, M.F. (2013). Valid estimation of odds ratio using two types of moving extreme ranked set sampling. *Journal of the Korean Statistical Society*, 42: 17–24.

17. Santos, K.C.P. and Barrios, E.B. (2015). Improving predictive accuracy of logistic regression model using ranked set samples. *Communications in Statistics-Simulation and Computation*, accepted, doi.org/10.1080/03610918.2014.955113.

18. Sinha, B.K., Sengupta, S., Mukhuti, S. (2006). Unbiased estimation of the distribution function of an exponential population using order statistics with application in ranked set sampling. *Communications in Statistics-Theory and Methods*,35(9):1655–1670.

19. Stella M., Elizabeth, A. S., James, A.T., and Douglas, A.W. (2015). Confidence intervals and hypothesis tests for a population mean using ranked set sampling: an auditing application. *Journal of Contemporary Management*, 4(20): 1-17.

20. 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.

21. Tiwari, N. and Pandey, G.S. (2013). Application of ranked set sampling design in environmental investigations for real data set. *Thailand Statistician*, 11(2): 173-184.