MINIMAX RANKED SET SAMPLING

Amjad D. Al-Nasser[1] and Amer I. Al-Omari[**]
[*]Department of Statistics, Science Faculty, Yarmouk University, Irbid, Jordan
[**]Department of Mathematics, Faculty of Science, Al al-Bayt University, Al Mafraq, Jordan

ABSTRACT
In this article, a new sampling scheme based on ranked data is proposed. The main idea of the new sampling procedure is to produce a more flexible, reliable, cost-efficient design than the classical ranked set sampling and simple random sampling techniques. Accordingly, the new sampling scheme minimizes the number of wasted measurement units with high efficiency performances in estimating the population mean. Moreover, as different set sizes are used then the sample mean expected to be biased, to solve this problem an information theoretic weighted mean estimator is proposed. It is found that the weighted mean is more accurate and more efficient than the standard one. and the sample mean based a simple random sampling technique. Both estimators, weighted and un-weighted outperform the simple random sampling scheme in estimating the population mean. A real data set for estimating the average cohort percentage for survivals to age 65 in Jordan from 1960 to 2015 is used to illustrate the proposed sampling scheme.

KEYWORDS: Shannon Entropy, Ranked Set Sampling, Mean Estimation, Relative Efficiency, Survivals to age 65.

MSC: 62D05

RESUMEN
En este trabajo se propone un nuevo esquema de muestreo basado en datos ordenados. La idea principal del nuevo procedimiento de muestreo es producir una técnica que sea más flexible, fiable y barata que el clásico “ranked set sampling” y que el muestreo simple aleatorio. Así que el nuevo esquema de muestreo minimiza el número de medidas desechadas con una alta eficiencia en su desempeño al estimar la media de la población. Más aun, como se utilizan tamaños diferentes de los conjuntos se espera que el estimador de la media tenga sesgo, para resolver este problema se propone un estimador ponderado basado en teoría de la información. Se halla que la media ponderada es más acurada y más eficiente que el típico y que la media muestral basado en el muestreo simple aleatorio. Ambos estimadores, el ponderado y el no ponderado funcionan mejor que el del esquema simple aleatorio en la estimación de la media poblacional. Se utiliza, para ilustrar el comportamiento del propuesto esquema de muestreo, un conjunto de dato reales para estimar el promedio del porcentaje de las cohortes de la sobrevivencia a la edad de 65 en Jordania entre 1960 y 2015

PALABRAS CLAVE: Entropía de Shannon, Ranked Set Sampling, Estimación de la media, eficiencia relativa, Sobrevivencia a la edad de 65.

1. INTRODUCTION

McIntyre (1952) is the first advocating the use of ranked sets for unbiased selective sampling called ranked set sampling (RSS). The main idea of the RSS is to increase the efficiency of the population estimators comparing to the classical simple random sampling (SRS) technique. In RSS, m SRS sets are selected each of size m. The units within each SRS are ranked based on a free-cost method. Then the i^{th} measurement unit is selected from the i^{th} SRS, where i= 1, 2, …, m. This procedure may repeated r times to obtain a RSS of size $n = rm$. If the set size (i.e., m) equal to the number of selected sets (i.e., m) then the sampling scheme is known as balanced RSS. However, if the set size is fixed (i.e., n) but not equal to number of random sets (i.e., m);then the sample scheme is called unbalanced RSS. There are several statistical inference procedures use balanced or unbalanced RSS, in all of these procedures an important property is the sample mean based on ranked data should be unbiased with minimum variance comparing to SRS (Zhang et al., 2014; Jemain et al., 2008, Al-Talib and Al-Nasser, 2008; Al-Omari, 2012; Zamanzade and Vock, 2015; Zamanzade and Mahdizadeh, 2017; Zamanzade and Al-Omari, 2016, Al-Omari and Haq, 2016; Chen et al., 2004 and the references therein).
In the literature, there are many authors proposed improvements to the RSS scheme. Muttlak (1997) suggested using the median ranked set sampling (MRSS). Muttlak (2003), introduced percentiles ranked set sampling (PRSS). For symmetric distributions, and in order to estimate an unbiased population mean with high efficiency, Samawi et al. (1996) investigated the extreme ranked set samples (ERSS); While, Jemain and

Al-Omari (2006) suggested double quartile ranked set sampling. Moreover, a generalization sampling schemes are proposed by Al-Nasser (2007) called L ranked set sampling (LRSS) to outperform the RSS, MRSS and PRSS; and Al-Nasser and Bani-Mustafa (2009) suggested RERSS as an alternative sampling scheme.

In terms of using unequal set sizes, Al Odat and Al-Saleh (2001) introduced the moving extreme ranked set sampling. In this sampling scheme, m random samples of size i =1,2, …, m are drawn; and then, the maximum order statistics is selected from each set. To complete the scheme, another cycle of m random samples are drawn; then the minimum value is selected from each set. Then the MRSS size is $2m$, the process could be repeated r times to obtain a MERSS of size $n = 2rm$. Even when r =1, the cost of using MERSS is very high. In this article, a two cycles of MERSS are merged into one cycle to minimize the sampling cost and to be more flexible and reliable in using the measurement units in our applied researches.

This paper considers nonparametric estimation for population mean based on a ranked data scheme of unequal set size under prefect ranking. This paper proposed un-weighted estimator as well as a weighted estimator which is motivated by the fact that the observations from unequal set sizes are not identically distributed. In other words, observations selected by different ranks should contribute differently because they follow different distributions.

The article organized as follows: Section (2) describes the new sampling scheme, and give an illustrative investigations for estimating the population mean in case of the standard uniform distribution. Also, a comparative study based on the relative efficiency criterion is given. Section (3) introduce a weighted mean estimation based on Shannon entropy to increase the accuracy and the efficiency of the proposed estimator. Section (4) we discuss the merits of the proposed sampling schemes. In Section (5) an illustration of the proposed sampling scheme is discussed by analyzing a real data set to estimate the average percentage cohort for the survival to age 65 in Jordan between 1960 and 2015. The article ends with some concluding remarks.

2. THE MINIMAX RSS SCHEME

In RSS scheme and its extensions, to draw a random sample of size m, we use m^2 sampling units, one of the previous schemes is folded ranked set sampling (FRSS) (Bani Mustafa et al., 2011) is used to reduce number of wasted sampling units and to improve the estimator efficiency in case of asymmetric distributions. The new sampling scheme has the same aims of FRSS by reducing number of wastes and increase the estimator efficiency but in symmetric distributions. The main idea is to use different set sizes and apply the same procedure of the ERSS. This can be clarify by applying the following steps:

Step (1): Draw m SRS of size i = 1,2,3, … , m

$$\begin{array}{lcllll}
SRS.1 & \rightarrow & x_1 & & & \\
SRS.2 & \rightarrow & x_1 & x_2 & & \\
SRS.3 & \rightarrow & x_1 & x_2 & x_3 & \\
\vdots & & \vdots & \vdots & \vdots & \vdots \\
SRS.m & \rightarrow & x_1 & x_2 & \cdots & x_m
\end{array}$$

Step (2): Arrange the sampling units within each SRS from smallest to largest

$$\begin{array}{lcllll}
SRS.1 & \rightarrow & x_{(1:1)} & & & \\
SRS.2 & \rightarrow & x_{(1:2)} & x_{(2:2)} & & \\
SRS.3 & \rightarrow & x_{(1:3)} & x_{(2:3)} & x_{(3:3)} & \\
\vdots & & \vdots & \vdots & \vdots & \vdots \\
SRS.m & \rightarrow & x_{(1:m)} & x_{(2:m)} & \cdots & x_{(m:m)}
\end{array}$$

Step (3): From the first SRS of size i=1, measure $x_{(1:1)}$; the minimum
Step (4): From the second SRS of size i=2; measure $x_{(2:2)}$; the maximum
Step (5): From the third SRS of size i=3; measure $x_{(1:3)}$; the minimum
Step (6): Continue in the same selection procedure till in the last sample you select the minimum if m is odd, or select the maximum if m is even.

$$\begin{array}{lcllllcl}
 & & \text{Ordered samples} & & & & & \text{MiniMaxRSS} \\
SRS.1 & \rightarrow & x_{(1:1)} & & & & \rightarrow & x_{[1:1]} \\
SRS.2 & \rightarrow & x_{(1:2)} & x_{(2:2)} & & & \rightarrow & x_{[2:2]} \\
SRS.3 & \rightarrow & x_{(1:3)} & x_{(2:3)} & x_{(3:3)} & & \rightarrow & x_{[1:3]} \\
\vdots & & \vdots & \vdots & \vdots & \vdots & &
\end{array}$$

$$SRS.m \quad \rightarrow \quad x_{(1:m)} \quad x_{(2:m)} \quad \cdots \quad x_{(m:m)} \quad \rightarrow \quad \begin{cases} x_{[1:m];if\ m\ is\ odd} \\ x_{[m:m];if\ m\ is\ even} \end{cases}$$

Step (7): Repeat the process *r* times to chose a MiniMax RSS of size *n* = *r.m*

Then the MiniMax RSS sample will be of the form:

$$\begin{cases} \left\{x_{[1:2i-1]k};\ x_{[2j:2j]k};\ i = 1,2,\dots,\dfrac{m+1}{2};\ j = 1,2,\dots,\dfrac{m-1}{2};\ k = 1,2,\dots,r\right\};\ \ if\ m\ is\ odd \\ \left\{x_{[1:2i-1]k};\ x_{[2i:2i]k};\ i = 1,2,\dots,\dfrac{m}{2};\ k = 1,2,\dots,r\right\} \qquad ;if\ m\ is\ even. \end{cases}$$

Note that the observed data are judgmental order statistics, where these values are independent but are not identically distributed. We will refer to this random sample as partial judgment order statistics. Consequently, the MiniMax RSS mean can be identify as follows:

$$\bar{X}_{MiniMaxRss} = \begin{cases} \dfrac{1}{mr}\left\{\displaystyle\sum_{k=1}^{r}\left(\sum_{i=1}^{(m+1)/2} X_{[1:2i-1]k} + \sum_{i=1}^{(m-1)/2} X_{[2i:2i]k}\right)\right\};\quad m\ is\ odd \\ \dfrac{1}{mr}\left\{\displaystyle\sum_{k=1}^{r}\left(\sum_{i=1}^{m/2} X_{[1:2i-1]k} + \sum_{i=1}^{m/2} X_{[2i:2i]k}\right)\right\};\qquad m\ is\ even. \end{cases}$$

Without lose of generality, let *r* =1; then expected value of the proposed estimator is given by:

$$E(\bar{X}_{MiniMaxRss}) = \begin{cases} \dfrac{1}{m}\left\{\displaystyle\sum_{i=1}^{(m+1)/2}\int x_{[1:2i-1]}dF(x_{[1:2i-1]}) + \sum_{i=1}^{(m-1)/2}\int x_{[2i:2i]}dF(x_{[2i:2i]})\right\};\quad m\ is\ odd, \\ \dfrac{1}{m}\left\{\displaystyle\sum_{i=1}^{m/2}\int x_{[1:2i]}dF(x_{[1:2i]} + \sum_{i=1}^{m/2}\int x_{[2i:2i]}dF(x_{[2i:2i]})\right\};\qquad m\ is\ even, \end{cases}$$

where the probability density functions of $x_{[i:m]}$ is given by:

$$dF\left(x_{[i:m]}\right) = f\left(x_{[i:m]}\right) = \frac{m!}{(i-1)!\,(m-i)!}\ x^{i-1}(1-x)^{m-i}.$$

The associated variance of this estimator is:

$$\sigma^2_{MiniMaxRSS} = \begin{cases} \dfrac{1}{m^2}\left\{\displaystyle\sum_{i=1}^{(m+1)/2}\sigma^2{}_{[1:2i-1]} + \sum_{i=1}^{(m-1)/2}\sigma^2{}_{[2i:2i]}\right\};\quad m\ is\ odd, \\ \dfrac{1}{m^2}\left\{\displaystyle\sum_{i=1}^{m/2}\sigma^2{}_{[1:2i-1]} + \sum_{i=1}^{m/2}\sigma^2{}_{[2i:2i]}\right\};\qquad m\ is\ even. \end{cases}$$

2.1 An Example of MiniMax RSS: Standard Uniform Distribution

In general, the samples based on ranked data have shown several evidences to conclude the superiority of such methods over the SRS while estimating the population mean. The relative efficiency (RE) is considered as the statistical criterion for investigating the new estimator performances. The RE can be computed as:

$\mathrm{RE} = \frac{\mathrm{MSE}(\bar{\mathrm{X}}_{\mathrm{SRS}})}{\mathrm{MSE}(\bar{\mathrm{X}}_{\mathrm{MiniMaxRss}})}$,

where *MSE* is the mean squared error.

To illustrate the RE of the proposed sampling scheme in estimating the population mean from the standard Uniform distribution i.e., U(0,1). The expected value and the variance of the ith order statistics from U(0,1) is:

$\mu_{(i:m)} = \frac{i}{m+1}, \sigma^2_{(i:m)} = \frac{i(m-i+1)}{(m+1)^2(m+2)}.$

Consequently, using the MMRSS, the expected value and the variance of the sample mean from a MMRSS is equal to:

$$\mu_{MiniMaxRSS} = \begin{cases} \frac{1}{m}\left\{\sum_{i=1}^{\frac{m+1}{2}} \frac{1}{2i} + \sum_{i=1}^{\frac{m-1}{2}} \frac{2i}{2i+1}\right\}; m \text{ is odd} \\ \frac{1}{m}\left\{\sum_{i=1}^{\frac{m}{2}} \frac{1}{2i} + \frac{2i}{2i+1}\right\}; m \text{ is even.} \end{cases}$$

Noting that, $H_m = \sum_{i=1}^{m} \frac{1}{i}$ is denoted by the m^th harmonic number and it is equal to

$\gamma + \psi_0(m+1)$,

where γ is the Euler-Mascheroni constant and ψ_0 is the digamma function. Therefore, the mean from the standard Uniform distribution based on MiniMaxRSS can be simplify to:

$$\mu_{MiniMaxRSS} = \begin{cases} \frac{1 + m - \log(4) + H_{\frac{m+1}{2}} - H_{\frac{m}{2}}}{2n}; m \text{ is odd} \\ \frac{m + \mathrm{H}_{\frac{m}{2}} + \psi[\frac{3}{2}] - \psi[\frac{3+m}{2}]}{2m}; m \text{ is even.} \end{cases}$$

Now the variance can be written as:

$$\sigma^2_{MiniMaxRSS} = \begin{cases} \frac{1}{m^2}\left\{\sum_{i=1}^{\frac{m+1}{2}} \frac{2i-1}{(2i)^2(2i+1)} + \sum_{i=1}^{\frac{m-1}{2}} \frac{2i}{(2i+1)^2(2i+2)}\right\}; m \text{ is odd} \\ \frac{1}{m^2}\left\{\sum_{i=1}^{\frac{n}{2}} \frac{2i-1}{(2i)^2(2i+1)} + \frac{2i}{(2i+1)^2(2i+2)}\right\}; m \text{ is even,} \end{cases}$$

which can be simplify to:

$$\sigma^2_{MiniMaxRSS} = \begin{cases} \frac{\frac{24(1+m)}{2+m} - 2\pi^2 + 12\psi^{(1)}[2+m]}{12m^2}; m \text{ is odd} \\ \frac{\frac{24(1+m)}{2+m} - 2\pi^2 + 12\psi^{(1)}[2+m]}{12m^2}; m \text{ is even,} \end{cases}$$

where $\psi^{(1)}$ is the first derivative of the digamma (psi) function. Therefore, the relative efficiency with respect to SRS is given by:

$$RE = \frac{\sigma^2_{SRS}}{MSE_{MiniMaxRSS}},$$

where

$$\sigma^2_{SRS} = \frac{1}{12\,m},$$

And $MSE_{MiniMaxRSS} = \sigma^2_{MiniMaxRSS} + bias^2$. Given that $bias = \mu_{MiniMaxRSS} - \frac{1}{2}$.

Then, the exact RE were computed for different set sizes i.e., m = 3, 4, 5 and 6, using Mathematica.6 software is illustrated in Figure (1).

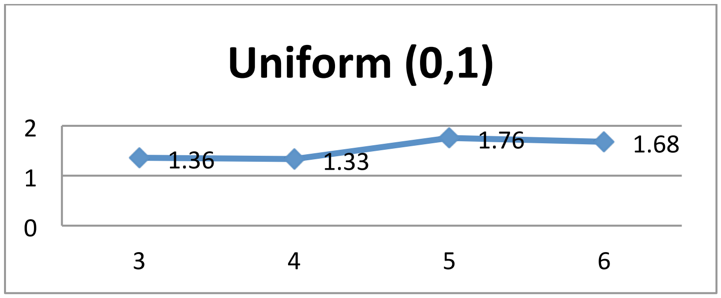

Figure (1): RE of the Standard Uniform Distribution

2.3 Efficiency Results: Comparing with SRS

To illustrate the RE of the proposed sampling scheme in estimating the population mean from a symmetric distribution, we considered a set of size *m= 3, 4, 5* and *6.* Then the exact RE were computed using Mathematica 6 software for several symmetric distributions. The results are presented in Table (1).
The computational results in Table (1) suggest that under the RE measure, the MiniMax RSS mean estimators are superior to the traditional SRS estimators when the underlying distribution is symmetric. It could be noted the following results:

- The MiniMax RSS estimators is more efficient than the SRS estimator with efficiency value more than 1 in all symmetric distributions and some of asymmetric distributions, (Beta, Rayleigh and Half Normal).
- The SRS estimator is an unbiased and more efficient in case of some asymmetric distributions, (Exponential, Gamma and Chi Square), when the set size is even.
- By increasing the set size (in both cases: odd set size and even set size separately), the relative efficiency value also improved in case of symmetric distributions.
- The bias is negative in case of odd set size, because number of minimum observations are more than the maximum observations within odd set sizes. Also, it is positive in case of even set size, because the expected value of the maximums is dominated the expected values of the minimums as we are using unequal set sizes.
- The MSE decreases as the set size increases in symmetric distributions.

Table .1: The RE of MiniMax RSS with respect to SRS for different distributions

	m	**3**			**4**			**5**			**6**		
	Distribution	**Bias**	**MSE**	**RE**	**Bias**	**MSE**	**RE**	**Bias**	**MSE**	**RE**	**Bias**	**MSE**	**RE**
Symmetric Distribution	**U(0,1)**	-0.027	0.020	1.360	0.054	0.016	1.330	-0.023	0.009	1.760	0.040	0.008	1.680
	N(0,1)	-0.094	0.258	1.290	0.186	0.207	1.210	-0.083	0.134	1.490	0.142	0.119	1.390
	Beta(3,3)	-0.018	0.009	1.330	0.035	0.007	1.260	-0.015	0.004	1.610	0.026	0.003	1.510
	Logistic(5,2)	-0.330	3.509	1.250	0.670	2.836	1.160	-0.300	1.921	1.370	0.510	1.741	1.260
	Student t(4)	-0.123	0.562	1.190	0.247	0.464	1.080	-0.113	0.335	1.190	0.191	0.307	1.090
	ArcSin (0,1)	-0.034	0.030	1.380	0.065	0.023	1.390	-0.027	0.013	1.880	0.047	0.011	1.840
Asymmetric Distribution	**Beta(5,2)**	-0.013	0.006	1.334	0.034	0.006	1.114	-0.007	0.003	1.615	0.031	0.004	1.199
	Rayleigh(1)	-0.054	0.108	1.323	0.137	0.097	1.102	-0.029	0.055	1.573	0.125	0.061	1.178
	HalfNormal(2)	-0.028	0.036	1.321	0.084	0.035	1.025	-0.006	0.018	1.552	0.084	0.023	1.029
	Exponential(1)	-0.056	0.265	1.256	0.229	0.289	**0.865**	-0.023	0.154	1.303	0.261	0.216	**0.772**
	Gamma(2,3)	-0.287	4.651	1.290	0.945	4.729	**0.952**	-0.023	2.523	1.427	0.992	3.299	**0.909**
	ChiSquare(3)	-0.155	1.564	1.279	0.553	1.636	**0.917**	-0.014	0.868	1.383	0.600	1.173	**0.853**

3. WEIGHTED ESTIMATION OF THE MEAN USING MINIMAX RSS

For a finite population X, let $F(x)$ and $f(x)$ denote, respectively, the cumulative distribution function (cdf) and the probability density function (pdf) with mean μ and variance σ^2. We wish to estimate the population mean using a random sample from MiniMax RSS scheme. Then, the judgmental order statistics can be written as:

$$\begin{cases} \left\{x_{[1:2i-1]};\ x_{[2j:2j]};\ i = 1,2,\dots,\dfrac{m+1}{2};\ j = 1,2,\dots,\dfrac{m-1}{2};\right\}; & if\ m\ is\ odd \\ \left\{x_{[1:2i-1]};\ x_{[2i:2i]};\ i = 1,2,\dots,\dfrac{m}{2}\right\} & ;if\ m\ is\ even \end{cases}$$

Note that judgmental order statistics are independent but not identically distributed. Accordingly, a weighted population mean estimator can be formulated as:

$$\bar{X}_w = \begin{cases} \left\{\displaystyle\sum_{i=1}^{(m+1)/2} w_i X_{[1:2i-1]} + \sum_{i=1}^{(m-1)/2} w_{i+\frac{m+1}{2}} X_{[2i:2i]}\right\}; & m\ is\ odd \\ \left\{\displaystyle\sum_{i=1}^{m/2} w_i X_{[1:2i-1]} + \sum_{i=1}^{m/2} w_{i+\frac{m}{2}} X_{[2i:2i]}\right\}; & m\ is\ even \end{cases}$$

where $w_1,\ w_2 \dots,\ w_m$ are nonnegative weights. Following Al-Nasser (2007), the optimal weights, which provide a measure of uncertainty, can be found by using an entropy measure from information theory under two constraints: (1) The moment condition: the weighted mean of MiniMax RSS is unbiased estimator of the population mean; and (2) Unity constraints: the sum of all weights is equal to one. Entropy generally is taken as a measure of expected information. A simple choice of this measure is Shannon (1948) entropy:

$$H(w) = -\sum_{i=1}^{m} w_i \ln(w_i),$$

where, $w_i \ln(w_i) = 0$ for $w_i = 0$, and $H(w)$ reaches a maximum when $w_1 = w_2 = \dots = w_m = \dfrac{1}{m}$. Then, by maximizing shannon entropy subject to the moment condition and the add up constraint $\sum_{j=1}^{m} w_j = 1$, leads to the optimal solution. This problem can be expressed by a nonlinear programming system (if m is odd):

$$Maximize\ -\sum_{j=1}^{m} w_j Ln(w_j),$$

Subject to

$$(1)\ \textstyle\sum_{i=1}^{(m+1)/2} w_i X_{[1:2i-1]} + \sum_{i=1}^{(m-1)/2} w_{i+\frac{m+1}{2}} X_{[2i:2i]} = \mu.$$

$$(2)\ \sum_{j=1}^{m} w_j = 1.$$

In case if m is even, then the first constrain is replaced by

$$\sum_{i=1}^{m/2} w_i X_{[1:2i-1]} + \sum_{i=1}^{m/2} w_{i+\frac{m}{2}} X_{[2i:2i]} = \mu$$

To recover the weights, one can form the Lagrangian function as

$$L = -\textstyle\sum_{i=1}^{m} w_i \ln(w_i) + \lambda_1\left(\mu - \left(\sum_{i=1}^{(m+1)/2} w_i X_{[1:2i-1]} + \sum_{i=1}^{(m-1)/2} w_{i+\frac{m+1}{2}} X_{[2i:2i]}\right)\right) + \lambda_2\left(1 - \sum_{i=1}^{m} w_i\right),$$

where λ_1 and λ_2 are a Lagrangian multiplier. Then, the unbiased estimator can be recovered through the weighted estimates, after finding the first order conditions and solve the nonlinear programming system. Accordingly, the optimal weighted mean can be expressed as:

$$\bar{X}_w = \begin{cases} \left\{ \sum_{i=1}^{(m+1)/2} \hat{w}_i X_{[1:2i-1]} + \sum_{i=1}^{(m-1)/2} \hat{w}_{i+\frac{m+1}{2}} X_{[2i:2i]} \right\}; & m \text{ is odd} \\ \left\{ \sum_{i=1}^{m/2} \hat{w}_i X_{[1:2i-1]} + \sum_{i=1}^{m/2} \hat{w}_{i+\frac{m}{2}} X_{[2i:2i]} \right\}; & m \text{ is even} \end{cases}$$

and its associated weighted variance will be defined as:

$$\sigma^2_{MiniMaxRSS} = \begin{cases} \frac{1}{m}\left\{ \sum_{i=1}^{(m+1)/2} \hat{w}_i\, \sigma^2_{[1:2i-1]} + \sum_{i=1}^{(m-1)/2} \hat{w}_{i+\frac{m+1}{2}}\, \sigma^2_{[2i:2i]} \right\}; & m \text{ is odd} \\ \frac{1}{m}\left\{ \sum_{i=1}^{m/2} \hat{w}_i\, \sigma^2_{[1:2i-1]} + \sum_{i=1}^{m/2} \hat{w}_{i+\frac{m}{2}}\, \sigma^2_{[2i:2i]} \right\}; & m \text{ is even.} \end{cases}$$

The results of the nonlinear programming system are given in Table (2) for symmetric distributions and Table (3) for asymmetric distributions.

Table 2: Optimal weights for given *m* under MiniMax RSS: Symmetric distributions

Distribution	*w* \ *m*	3	4	5	6
U(0,1)	1	0.337	0.263	0.202	0.169
	2	0.398	0.212	0.215	0.156
	3	0.265	0.351	0.182	0.194
	4		0.175	0.226	0.142
	5			0.175	0.205
	6				0.135
N(0,1)	1	0.337	0.263	0.200	0.169
	2	0.398	0.211	0.215	0.156
	3	0.265	0.352	0.181	0.193
	4		0.174	0.229	0.142
	5			0.175	0.206
	6				0.134
Beta(3,3)	1	0.338	0.263	0.202	0.169
	2	0.397	0.211	0.216	0.156
	3	0.265	0.352	0.181	0.193
	4		0.174	0.227	0.142
	5			0.174	0.206
	6				0.134
Logistic(5,2)	1	0.337	0.263	0.199	0.169
	2	0.398	0.211	0.215	0.156
	3	0.265	0.353	0.181	0.193
	4		0.173	0.230	0.142
	5			0.175	0.207
	6				0.133
T(4)	1	0.337	0.263	0.202	0.169
	2	0.398	0.211	0.216	0.156
	3	0.265	0.353	0.181	0.193
	4		0.173	0.228	0.142
	5			0.173	0.207
	6				0.133
ArcSin(0,1)	1	0.337	0.262	0.202	0.169
	2	0.398	0.211	0.215	0.155
	3	0.265	0.350	0.182	0.194
	4		0.176	0.225	0.142
	5			0.176	0.204
	6				0.136

Then the relative efficiency (*RE*) of estimating the population mean using the weighted MiniMax RSS method with respect to the usual estimator using SRS is defined as:

$$RE_w = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_w)}.$$

The RE_w in Table (4) summarizes results for weighted MiniMax RSS. For each distribution, calculations were done when the set sizes *m* equal to 3, 4, 5, and 6.

Table 3: Optimal weights for given *m* under MiniMax RSS: Asymmetric distributions

Distribution	w \ m	3	4	5	6
Beta(5,2)	1	0.337	0.258	0.201	0.168
	2	0.391	0.211	0.206	0.148
	3	0.272	0.367	0.192	0.209
	4		0.164	0.216	0.135
	5			0.186	0.218
	6				0.122
Rayleigh(2)	1	0.337	0.261	0.199	0.169
	2	0.391	0.208	0.207	0.149
	3	0.271	0.366	0.188	0.207
	4		0.166	0.217	0.135
	5			0.188	0.218
	6				0.123
HalfNormal	1	0.337	0.261	0.200	0.171
	2	0.387	0.205	0.204	0.147
	3	0.276	0.375	0.195	0.214
	4		0.159	0.207	0.126
	5			0.194	0.227
	6				0.115
Exp(1)	1	0.336	0.267	0.200	0.175
	2	0.380	0.201	0.195	0.143
	3	0.285	0.387	0.207	0.226
	4		0.145	0.189	0.117
	5			0.209	0.239
	6				0.099
Gamma(2,3)	1	0.336	0.264	0.201	0.171
	2	0.385	0.203	0.200	0.148
	3	0.279	0.378	0.197	0.217
	4		0.155	0.203	0.124
	5			0.198	0.230
	6				0.109
ChiSquare(3)	1	0.336	0.263	0.200	0.171
	2	0.383	0.206	0.199	0.146
	3	0.281	0.381	0.202	0.223
	4		0.150	0.197	0.123
	5			0.202	0.232
	6				0.105

Table 4: Relative efficiency (RE) by using weighted MiniMax RSS *vs* SRS.

	Distribution \ m	3	4	5	6
Symmetric Distributions	U(0,1)	1.386	1.619	1.847	2.088
	N(0,1)	1.322	1.451	1.566	1.663
	Beta(3,3)	1.352	1.525	1.684	1.840
	Logistic(5,2)	1.285	1.370	1.439	1.489
	T(4)	1.217	1.238	1.242	1.234
	ArcSin(0,1)	1.401	1.682	1.965	2.275
Asymmetric Distributions	Beta(5,2)	1.309	1.490	1.607	1.767
	Rayleigh(1)	1.300	1.462	1.569	1.700
	HalfNormal(2)	1.274	1.447	1.535	1.681
	Exponential(1)	1.188	1.302	1.344	1.422
	Gamma(2,3)	1.234	1.366	1.419	1.518
	ChiSquare(3)	1.216	1.342	1.392	1.483

The results indicate that there is an improvement by using the weighted MiniMax RSS. In fact, the weighted MiniMax RSS estimators are unbiased and outperform the SRS estimators.

4. MERITS OF THE MINIMAX RSS SCHEME

The theoretical part of this article showed that the proposed sampling scheme is efficient and could be more accurate when we give the right weights for each measurement. Therefore, MiniMax RSS could be consider a robust extended form of the RSS scheme with some additional benefits. In general, there are several considerations, which make the MiniMax RSS better than many of the ranked sampling scheme. Starting with unequal set sizes, in many applications; specially, the medical and reliability applications, it is very hard to obtain all the time equal set samples. Then, cost considerations, it is known that in most of ranked data schemes such as RSS for a set of size m we usually discard m^2 - m sampling units. While by using the MiniMaxRSS we discard $\frac{m(m+1)}{2} - m = \frac{m(m-1)}{2}$. Table (5) shows the number of used and wasted sampling units in ranked data schemes and MiniMaxRSS and their sampling unit's saving ratio.

Table 5. Saving Percentages by using MiniMaxRSS scheme

	MiniMax RSS		RSS		
m	Measured	Wasted	Measured	Wasted	Saving percentage $\frac{MiniMaxRSS(Wasted)}{RSS\ (Wasted)}$
3	6	3	9	6	50%
4	10	6	16	12	50%
5	15	10	25	20	50%
6	21	15	36	30	50%

The results indicated that we need double number of sampling units by using RSS to observe the same set size by using MiniMax RSS. Moreover, by using a weighted MiniMax RSS the mean estimators became unbiased and more efficient than their counterpart estimators based on SRS.

5. ESTIMATING THE AVERAGE OF SURVIVAL TO AGES 65 IN JORDAN

We illustrate the MiniMax RSS scheme in data collection and estimating the population mean using a real data set obtained from the United Nation Division (2017) about the Jordan - Health Status - Survival to age 65. Survival to age 65 is an important indicator used by the united nation as well as each country to measure the percentage of a cohort of newborn infants that would survive to age 65. The important of this indicator is in its impact on the describing life quality and in describing the health status of the citizens, which effect on the happiness rate in that country. Noting that, age 65 reflects the number of years a person could be expected to receive unreduced Social Security retirement benefits (United Nation, 2015). The data file consists of information on female and male (% of cohort) from 1960 to 2015. The summary statistics on these data is given in Table (6) The results indicate that, the percent of female that survival to age 65 in Jordan increased from 44.87.13% in 1960 to 84.36% in 2015 at an average annual rate of 72.46%, while, the percent of male increased from 44.87% in 1960 to 78.36% in 2015 at an average annual rate of 66.47%.

Table 6: Summary Statistics to Survive to age 65 in Jordan

	N	Min	Q1	Q2	Q3	Max	μ	σ
Male	56	44.87	58.601	69.904	75.079	78.36	66.465	10.155
Female	56	49.19	64.929	76.158	81.028	84.36	72.459	10.421

In this article, a random sample of size 15 is selected by two sampling techniques; SRS and MiniMax RSS. In using the proposed MiniMax RSS scheme, three SRS were selected without replacement from the population each of size 1, 2 and 3; respectively. Then, the process is repeated five times (see Table 6). It is worth to say that, the random selection was on the auxiliary variable "year", and both associated values of the indicator for the male and female were included if the measurement "year" is selected for inclusion in the MiniMax RSS. Then the weighted and un weighted averages are computed for both genders. Noting that, in order to compute the weighted average, first we implemented the Kolmogrov Smirnov test on the actual data, the results indicated that, the population is Normal for both genders (Z=1.06, p=0.211) and (Z=1.113, p= 0.168) for male and female, respectively. Therefore, the weights in Table.2 are used when the population is Normal and the set size is 3 in computing the weighted average (see Table 7). To measure the accuracy of the estimator, the % Error is used:

$$\%Error = \frac{|estimator - \mu|}{|\mu|} \times 100$$

Table 6: MiniMax RSS for the survival to age 65 data using m =3.

Set #	Sample #	MiniMax RSS Male	MiniMax RSS Female	SRS Male	SRS Female
1	1	{60.38}	{66.84}	44.87	49.19
	2	{57.33, 73.93}	{63.57, 79.84}	48.74	53.81
	3	{50.81, 54.04, 55.13}	{56.24, 59.90, 61.12}	52.94	58.67
2	1	{61.39}	{69.02}	54.04	59.9
	2	{45.78, 76.55}	{50.29, 82.51}	55.13	61.12
	3	{69.67, 73.30, 77.0}	{75.94, 79.24, 83}	56.23	62.34
3	1	{77.45}	{83.48}	58.35	64.66
	2	{74.85, 78.13}	{80.47, 84.15}	63.93	70.56
	3	{56.23, 64.69, 77.22}	{62.34, 71.33, 83.24}	64.69	71.33
4	1	{48.74}	{53.81}	70.14	76.37
	2	{46.67, 70.60}	{51.39, 76.81}	71.07	77.24
	3	{72.01, 75.99, 76.77}	{78.10, 81.95, 82.76}	71.54	77.67
5	1	{71.54}	{77.67}	72.33	78.39
	2	{44.87,72.97}	{49.19, 78.95}	75.16	81.11
	3	{66.21, 72.65, 78.36}	{72.88, 78.67, 84.36}	77	83

Table 7: Population Mean Estimation Based on Different Sampling Scheme

Method	MiniMax RSS				SRS	
Gender	Un weighted Average	*%Error*	Weighted Average	*%Error*	Sample Mean	*%Error*
Male	67.11	0.97%	67.85	2.08%	62.41	6.1%
Female	73.23	1.06%	73.97	2.08%	68.36	5.65%

The results of the MiniMax RSS scheme is more accurate in estimating the population mean than using the SRS of the same size. These results can be considered as an advantage of the proposed method to be a robust alternative to the SRS in estimating the population mean.

5. CONCLUDING REMARKS

A new ranked sampling scheme called MiniMax RSS is proposed in this article. There are two benefits by using the proposed MiniMax RSS, gain in efficiency when estimating the population mean from a given distribution, and reduce 50% of the number of wasted measurement units. Moreover, to improve the accuracy and the efficiency of the MiniMax RSS estimator a weighted mean is used. The relative efficiency results and the real data analysis in this article recommended of using the MiniMax RSS for estimating the population mean as an alternative sampling scheme.

ACKNOWLEDGMENTS: Both authors thank the suggestions of anonymous referees which allowed improving the quality of the present paper.

RECEIVED: NOVEMBER, 2017
REVISED: NOVEMBER, 2017

REFERENCES

[1] AL-NASSER, A.D. (2007): L Ranked set sampling, A generalization procedure for robust visual sampling. **Communications in Statistics, Simulation and Computation**, 36, 33-43.

[2] AL-NASSER, A.D., AND A. BANI MUSTAFA. (2009): Robust extreme ranked set sampling. **Journal of Statistical Computation and Simulation**, 79, 859–867.

[3] AL-OMARI, A.I. (2012): Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling. **Statistics and Probability Letters**, 82, 1883–1890.

[4] AL-OMARI, A.I. AND HAQ, A. (2016): 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**, 86, 2262–2272.
[5] AL-TALIB, M., AND AL-NASSER, A.D. (2008): Estimation of Gini-Index from Continuous Distribution Based on Ranked Set Sampling. **Electronic Journal of Applied Statistical Analysis**, 1, 37-48.
[6] BANI MUTAFA, A., AL-NASSER, A.D. AND ASLAM, M. (2011): Folded ranked set sampling. **Communications of the Korean Statistical Society**. 8, 147 - 153.
[7] CHEN, Z., BAI, Z. AND SINHA, B.K. (2004): **Ranked Set Sampling–Theory and Application**. Springer, New York.
[8] JEMAIN, A.A. AND AL-OMARI, A.I. (2006): Double quartile ranked set samples. **Pakistan Journal of Statistics**, 22, 217-228.
[9] JEMAIN, A.A., AL-OMARI, A.I. AND IBRAHIM, K. (2008): Some variations of ranked set sampling. **Electronic Journal of Applied Statistical Analysis**, 1, 1-15.
[10] MCINTYRE, G.A. (1952): A method for unbiased selective sampling, using ranked sets. **Australian Journal of Agricultural Research**, 3, 385-390.
[11] 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.
[12] MUTTLAK, H.A. (1997): Median ranked set sampling. **J. of Applied Statistical Sciences**, 6, 245-255.
[13] SAMAWI H, ABU-DAYYEH, W. AND AHMED, S. (1996): Extreme ranked set sampling. **The Biometrical Journal**, 30, 577-586.
[14] SHANNON, C.E. (1948): A Mathematical Theory of Communication. **Bell System Tech. J.**, 27, 379-423.
[15] UNITED NATIONS, Department of Economic and Social Affairs, Population Division (2015): **World Population Ageing**, 2015 (ST/ESA/SER.A/390):
[16] UNITED NATIONS, Population Division. (2017): **World Population Prospects. New York, United Nations, Department of Economic and Social Affairs**. Available at esa.un.org/wpp/unpp/panel_population.htm.
[17] ZAMANZADE, E. AND AL-OMARI, A.I. New ranked set sampling for estimating the population mean and variance. **Hacettepe Journal of Mathematics and Statistics**, 45, 1891-1905.
[18] ZAMANZADE, E., VOCK, M. (2015): Variance estimation in ranked set sampling using a concomitant variable, **Statistics and Probability Letters**, 105, 1-5.
[19] ZAMANZADE, E., AND MAHDIZADEH, M. (2017): A more efficient proportion estimator in ranked set sampling. **Statistics and Probability Letters**, 129, 28-33.
[20] ZHANG, L., DONG, X., XU, X., CUI, L. (2014): Weighted Estimation of Quantiles Using Unbalanced Ranked Set Sampling. **Quality Technology & Quantitative Management**, 11, 281-295.