

PREDICTION INTERVALS FOR CHARACTERISTICS OF FUTURE  
NORMAL SAMPLE UNDER MOVING RANKED SET SAMPLING

M. Al-Rawwash, M.T. Alodat, K.M. Aludaat, N. Odat, R. Muhaidat

## 1. INTRODUCTION

In a distinguished article aiming to estimate the population mean, (McIntyre, 1952) introduced a very clever idea for selecting a sample from the population of interest that appeared to be more representative than the Simple Random Sample (SRS) in carrying out the estimation process of the population mean. As a cornerstone in the literature of sampling theory, McIntyre's work has been referred to as the Ranked Set Sampling (RSS). The motivation of RSS in applied fields arises in certain situations where ranking is considered to be easy compared to the high cost of the traditional sampling techniques. On the other hand, the handiness, flexibility and robustness of the RSS compared to its counterparts are considered additional advantages of the RSS scheme. Numerous articles and several books discussed the idea of RSS and its modifications over the last five decades (Takahasi and Wakimoto, 1968; Dell and Clutter, 1972; Martin *et al.*, 1980; Kaur *et al.*, 1995; Muttlak, 1997; Alodat and Al-Saleh, 2001; Alodat *et al.*, 2009, Alodat *et al.*, 2010). To shed more light on the RSS approach, we describe the steps leading to an RSS of size  $m$  as follows. First we draw  $m$  independent random samples of size  $m$  from the target population. Accordingly, using a visual inspection or any other cheap method, we detect the  $i^{th}$  order statistic of the  $i^{th}$  sample for actual quantification. The resulting quantified ordered statistics constitute an RSS of size  $m$ . Recently, the RSS technique has taken the attention of a large number of researchers in ecology, wildlife and agriculture where ranking units, using a visually inspection, has a negligible cost relative to their quantification. Despite the fact that perfect ranking is considered in this article, yet we conduct a pair wise comparison between perfect and imperfect ranked set sampling and we elaborate on its importance during the course of this article.

Although the idea of RSS was introduced by (McIntyre, 1952), the theoretical and rigorous statistical framework of ranked set sampling was outlined for the first time by (Takahasi and Wakimoto, 1968). Compared to a simple random sample with the same size, the ranked set sample seems to be more representative for the population of interest and it produces estimators that are more efficient

(Chen, 2000). In fact, two factors may affect the efficiencies of estimators obtained via the RSS scheme namely, the set size ( $m$ ) and the ranking error. (Al-Saleh and Alomari, 2002) showed that the larger the set size, the larger the efficiency of the estimator. On the other hand, increasing the set size, leads to tangible difficulties in visual ranking of the sampled units. Carrying out the RSS procedure may result in some kind of ranking errors. (Takahasi and Wakimoto, 1968) studied the performance of RSS when the ranking is perfect while ranking errors and its implications have been studied later by (Dell and Clutter, 1972). It has been shown that ranking errors do not affect the superiority of RSS over SRS while estimating the population mean as well as other vital parameters (Presnell and Bohn, 1999). On the other hand, the advantage of RSS will faint in the case of imperfect sampling and eventually if the ranking process is assumed to be completely random; it will yield equal variances using RSS and SRS techniques (Dell and Clutter, 1972; Cobby *et al.*, 1985). Moreover, (Evans, 1967) noted that there is no practical difference between visual-type ranking and the actual detailed quantification of the units. For these reasons, modifications of RSS were presented in the literature to study the effect of ranking errors that the experimenter makes (Alodat and Al-Saleh, 2001; Al-Saleh and Alomari, 2002; Samawi *et al.*, 1996). Moreover, (Presnell and Bohn, 1999) pointed out some of the common errors presented in the RSS literature regarding perfect and imperfect ranking. They provided examples and counterexamples to show the possibility of obtaining more efficient results using imperfect ranking. They provide some theoretical detailed work to support their claims.

(Alodat and Al-Saleh, 2001) introduced the moving ranked set sampling (MRSS) which can be described as follows:

1. Select  $m$  simple random samples of sizes 1, 2, ...,  $m$ , respectively.
2. From the  $j^{\text{th}}$  sample,  $j = 1, 2, \dots, m$  quantify the  $j^{\text{th}}$  order statistic after detecting it via a visual inspection or any other crude method.
3. Repeat steps 1 and 2, but by quantifying the minimum i.e., the first order statistic from each sample.
4. The steps (1)-(3) could be repeated several times to increase the sample size for fixed  $m$ .

In this sampling scheme, we select  $m(m+1)$  sampling units but we quantify only  $2m$  units that will be used to obtain prediction intervals for future characteristics of a future sample from normal distribution. Moreover, this modification has an advantage over the ordinary RSS, since visual detection of the extremes order statistics of samples is easier than detecting other order statistics. (Alodat and Al-Saleh, 2001) showed that the probability of making errors while ranking the samples via MRSS is smaller than its counterpart in RSS. This procedure was also studied by (Al-Saleh and Al-Hadrami, 2003) and (Alodat *et al.*, 2010). The rest of the article is structured as follows. In section 2, we present the notation and the theoretical setup of the moving ranked set sampling scheme. The prediction intervals for the sample mean as well as the extreme order statistics of a fu-

ture sample is constructed in section 3 and 4 based on MRSS and SRS, respectively. In section 5, we conduct numerical comparisons between MRSS and RSS using the length of the prediction interval as our main comparison criterion. The ranking error effect is discussed in section 6. Finally, we provide an overview application to grassland biodiversity data set in section 7 and conclude our work in section 8.

## 2. NOTATION AND THEORETICAL SETUP

Let  $\{X_{j1}^1, X_{j2}^1, \dots, X_{jj}^1\}$  and  $\{X_{j1}^2, X_{j2}^2, \dots, X_{jj}^2\}$ ,  $j = 1, 2, \dots, m$  be a collection of  $2m$  random samples from  $N(\mu, \sigma^2)$  and define  $Y_{1j}$  and  $Y_{2j}$  to be an MRSS sample such that  $Y_{1j} = \min\{X_{j1}^1, X_{j2}^1, \dots, X_{jj}^1\}$  and  $Y_{2j} = \min\{X_{j1}^2, X_{j2}^2, \dots, X_{jj}^2\}$ . Also, let  $\phi(x)$  and  $\Phi(x)$  denote the density and the cumulative distribution functions of the standard normal distribution. The fundamentals of the order statistic theory indicate that the random variables  $Y_{1j}$  and  $Y_{2j}$  have the following distribution functions (Arnold *et al.*, 1992)

$$\begin{aligned} F_{1j}(y; \mu, \sigma) &= 1 - \left[ 1 - \Phi\left(\frac{y - \mu}{\sigma}\right) \right]^j \\ &= 1 - \Phi^j\left(\frac{\mu - y}{\sigma}\right) \end{aligned}$$

and

$$F_{2j}(y; \mu, \sigma) = \Phi^j\left(\frac{y - \mu}{\sigma}\right),$$

respectively, for  $j = 1, 2, \dots, m$ . The corresponding probability density functions are

$$f_{1j}(y; \mu, \sigma) = \frac{j}{\sigma} \Phi^{j-1}\left(\frac{\mu - y}{\sigma}\right) \phi\left(\frac{y - \mu}{\sigma}\right)$$

and

$$f_{2j}(y; \mu, \sigma) = \frac{j}{\sigma} \Phi^{j-1}\left(\frac{y - \mu}{\sigma}\right) \phi\left(\frac{y - \mu}{\sigma}\right),$$

respectively. For simplicity, we may use the transformation  $y = \sigma v + \mu$  to show that

$$E(Y_{1j}) = \int_{-\infty}^{\infty} y f_{1j}(y; \mu, \sigma) dy = \mu + \sigma A_j,$$

where

$$A_j = \int_{-\infty}^{\infty} y f_{1j}(y; 0, 1) dy.$$

Similarly, we may obtain the expected value of  $Y_{2j}$  as follows

$$E(Y_{2j}) = \int_{-\infty}^{\infty} y f_{2j}(y; \mu, \sigma) dy = \mu + \sigma B_j,$$

where

$$B_j = \int_{-\infty}^{\infty} y f_{2j}(y; 0, 1) dy.$$

One may notice that the random variables  $-Y_{1j}$  and  $Y_{2j}$  have the same distribution when  $\mu = 0$  implying that  $B_j = -A_j$  for all  $j = 1, 2, \dots, m$ , (Arnold *et al.*, 1992).

### 3. PREDICTION INTERVALS USING MRSS

In this section, we focus our attention on developing prediction intervals for a new future observation as well as the sample mean and the extreme order statistics of a future sample under the normality assumption. To accomplish this mission, we define  $S_1(Y_1, Y_2, \dots, Y_m)$  to be a statistic based on a future sample obtained from  $N(\mu, \sigma^2)$  while  $X_1, X_2, \dots, X_n$  represents an observed random sample from  $N(\mu, \sigma^2)$ . To derive the prediction interval of the statistic  $S_1(Y_1, Y_2, \dots, Y_m)$ , we define another ancillary statistic  $S_3(S_1, S_2)$ , where  $S_2$  is only a function of  $X_1, X_2, \dots, X_n$ . Since the distribution of  $S_3$  does not depend on  $\mu$  and  $\sigma^2$ , then for a confidence level  $1 - \alpha_1 - \alpha_2$ , we have

$$P(a_{\alpha_1} < S_3(S_1, S_2) < a_{1-\alpha_2}) = 1 - \alpha_1 - \alpha_2$$

where  $a_{\alpha_1}$  and  $a_{1-\alpha_2}$  denote the  $100\alpha_1$  and  $100(1-\alpha_2)$  quantiles of the distribution of  $S_3(S_1, S_2)$ , respectively. A  $100(1-\alpha_1-\alpha_2)\%$  prediction interval of  $S_1$  can be obtained by solving the following inequality for  $S_1$ :

$$a_{\alpha_1} < S_3(S_1, S_2) < a_{1-\alpha_2} \tag{1}$$

Let  $Y^*$  be a new observation distributed as  $N(\mu, \sigma^2)$  and consider the statistic

$$F = \frac{Y^* - 0.5(\bar{Y}_1 + \bar{Y}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2}} \quad (2)$$

where  $\bar{Y}_1 = \frac{1}{m} \sum_{j=1}^m Y_{1j}$ ,  $\bar{Y}_2 = \frac{1}{m} \sum_{j=1}^m Y_{2j}$  and  $\bar{Y} = \frac{1}{2m} \sum_{i=1}^2 \sum_{j=1}^m Y_{ij}$ . Since the original data belongs to a location-scale family, we may write the random variable  $Y_{ij}$  as  $Y_{ij} = \mu + \sigma Z_{ij}$ , where  $Z_{ij}$  has the pdf  $f_{ij}(z; 0, 1)$  for  $i = 1, 2$  and  $j = 1, 2, \dots, m$ . Also, we may notice that the probability density function  $f_{ij}(z; 0, 1)$  does not depend on the parameters  $\mu$  and  $\sigma$  which implies that the distribution of the statistic (2) does not depend on  $\mu$  and  $\sigma$  too. Hence, equation (2) may be written as

$$\begin{aligned} \frac{Y^* - 0.5(\bar{Y}_1 + \bar{Y}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2}} &= \frac{\mu + \sigma Z^* - 0.5(\mu + \sigma \bar{Z}_1 + \mu + \sigma \bar{Z}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (\mu + \sigma Z_{ij} - \mu - \sigma \bar{Z})^2}} \\ &= \frac{Z^* - 0.5(\bar{Z}_1 + \bar{Z}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Z_{ij} - \bar{Z})^2}} \end{aligned}$$

where  $\bar{Z}_1 = \frac{1}{m} \sum_{j=1}^m Z_{1j}$ ,  $\bar{Z}_2 = \frac{1}{m} \sum_{j=1}^m Z_{2j}$ ,  $\bar{Z} = \frac{1}{2m} \sum_{i=1}^2 \sum_{j=1}^m Z_{ij}$  and  $Z^*$  is distributed as  $N(0, 1)$ . Eventually, the distribution of the statistic  $F$  could be found via Monte Carlo simulation method.

Similarly and based on a given future sample say  $W_1, W_2, \dots, W_n$ , we may define new statistics that correspond to the average ( $\bar{W}$ ), the minimum ( $W_{(1)}$ ) and the maximum ( $W_{(n)}$ ). To elaborate more on this idea, we define the random variable  $W_k = \mu + \sigma V_k$ , for  $k = 1, 2, \dots, n$  which allows us to construct the following parameter-free distribution statistics:

$$G = \frac{\bar{V} - 0.5(\bar{Z}_1 + \bar{Z}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Z_{ij} - \bar{Z})^2}}$$

$$H = \frac{V_{(1)} - 0.5(\bar{Z}_1 + \bar{Z}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Z_{ij} - \bar{Z})^2}},$$

and

$$E = \frac{V_{(n)} - 0.5(\bar{Z}_1 + \bar{Z}_2)}{\sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Z_{ij} - \bar{Z})^2}}$$

*Theorem 1.* Suppose that we have an MRSS, say  $Y_{ij}$ ,  $i = 1, 2$  and  $j = 1, 2, \dots, m$ , then based on the previously listed variables and assumptions we have

1. A  $100(1-\alpha)$  prediction interval of  $Y^*$  is

$$\bar{Y} + F_{\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2} < Y^* < \bar{Y} + F_{1-\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2}.$$

2. A  $100(1-\alpha)$  prediction interval of  $\bar{W}$  is

$$\bar{Y} + G_{\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2} < \bar{W} < \bar{Y} + G_{1-\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2}.$$

3. A  $100(1-\alpha)$  prediction interval of  $W_{(1)}$  is

$$\bar{Y} + H_{\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2} < W_{(1)} < \bar{Y} + H_{1-\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2}.$$

4. A  $100(1-\alpha)$  prediction interval of  $W_{(n)}$  is

$$\bar{Y} + E_{\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2} < W_{(n)} < \bar{Y} + E_{1-\alpha/2} \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2},$$

Where  $F_\alpha, G_\alpha, H_\alpha$  and  $E_\alpha$  are the  $\alpha^{th}$  quantiles of the random variables  $F, G, H$  and  $E$ , respectively.

*Proof.* The  $100(1-\alpha)$  prediction interval of  $Y^*$  will be derived using equations (1) and (2) and assuming that that  $S_3 = F$ ,  $S_1 = Y^*$ ,  $S_2 = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$  and  $\alpha_1 = \alpha_2 = \alpha/2$ . Similar approaches allow us to conclude the results presented in parts 2, 3 and 4.

#### 4. PREDICTION INTERVAL USING SRS

To better illustrate the performance of the MRSS in obtaining the prediction intervals and for the sake of comparing these results to those using SRS, we plan to derive the aforementioned prediction intervals using the SRS scheme. To carry out this mission, we assume that  $X_1, X_2, \dots, X_{2m}$  is a random sample of size  $2m$  from  $N(\mu, \sigma^2)$  and  $U_1, U_2, \dots, U_n$  is a future sample distributed as  $N(\mu, \sigma^2)$ . Once again we make use of the standardization of the random variable  $X_j$  as  $X_j = \mu + \sigma R_j$ , where  $R_j$  has a standard normal distribution. Accordingly, we may introduce the following variables that represent a parameter-free distribution statistics

- a. To predict a new observation ( $U_1$ ), we define the random variable  $I$  as follows

$$I = \frac{U_1 - \bar{X}}{\sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}} = \frac{V - \bar{R}}{\sqrt{\sum_{j=1}^{2m} (R_j - \bar{R})^2}},$$

where  $\bar{X} = \frac{1}{2m} \sum_{j=1}^{2m} X_j$  and  $\bar{R} = \frac{1}{2m} \sum_{j=1}^{2m} R_j$ .

- b. To predict the sample mean ( $\bar{U}$ ) of a future sample of size  $n$ , we define the random variable  $J$  as follows

$$J = \frac{\bar{U} - \bar{X}}{\sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}} = \frac{\bar{V} - \bar{R}}{\sqrt{\sum_{j=1}^{2m} (R_j - \bar{R})^2}}$$

- c. Similarly, to predict the minimum ( $U_{(1)}$ ) of a future sample, we define the following random variable

$$K = \frac{U_{(1)} - \bar{X}}{\sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}} = \frac{V_{(1)} - \bar{R}}{\sqrt{\sum_{j=1}^{2m} (R_j - \bar{R})^2}}$$

- d. Finally, to predict the maximum ( $U_{(n)}$ ) of a future sample, we need to define the following random variable  $L$

$$L = \frac{U_{(n)} - \bar{X}}{\sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}} = \frac{V_{(n)} - \bar{R}}{\sqrt{\sum_{j=1}^{2m} (R_j - \bar{R})^2}}$$

*Theorem 2.* Based on the SRS scheme and considering the random variables  $I, J, K$  and  $L$ , we may construct the prediction intervals for  $U_1, \bar{U}, U_{(1)}$  and  $U_{(n)}$  as follows:

1. A  $100(1-\alpha)$  prediction interval for  $U_1$  is

$$\bar{X} + I_{\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2} < U_1 < \bar{X} + I_{1-\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}.$$

2. A  $100(1-\alpha)$  prediction interval for  $\bar{U}$  is

$$\bar{X} + J_{\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2} < \bar{U} < \bar{X} + J_{1-\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}$$

3. A  $100(1-\alpha)$  prediction interval for  $U_{(1)}$  is

$$\bar{X} + K_{\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2} < U_{(1)} < \bar{X} + K_{1-\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}$$

4. A  $100(1-\alpha)$  prediction interval for  $U_{(n)}$  is

$$\bar{X} + L_{\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2} < U_{(n)} < \bar{X} + L_{1-\alpha/2} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2},$$

where  $I_\alpha, J_\alpha, K_\alpha$  and  $L_\alpha$  are the  $\alpha^{th}$  quantiles of the random variables  $I, J, K$  and  $L$ , respectively. Moreover, the distributions of these random variables can be obtained using Monte-Carlo simulation setup.



TABLE 1  
 The upper and lower quantiles of F, G, H, E, I, J, K and L when  $n = 5$  and  $\alpha = 0.05$

$m$	F	G	H	E	I	J	K	L
2	-1.9702	-1.1186	-3.2073	-0.2292	-2.0723	-1.2467	-3.3517	-0.2939
	1.9671	1.1296	0.2517	3.1208	2.1067	1.2218	0.3091	3.3427
5	-0.6259	-0.3162	-0.8699	-0.0469	-0.8089	-0.4144	-1.1126	-0.0813
	0.6279	0.3133	0.0382	0.8557	0.8043	0.4128	0.0716	1.1237
8	-0.4087	-0.1945	-0.5505	-0.0222	-0.5695	-0.2813	-0.7809	-0.0438
	0.4044	0.1971	0.0194	0.5429	0.5658	0.2831	0.0418	0.7701
12	-0.2931	-0.1339	-0.3845	-0.0118	-0.4351	-0.2140	-0.5958	-0.0289
	0.2879	0.1338	0.0128	0.3851	0.4459	0.2076	0.0217	0.6022

5. SIMULATION STUDY

To compare the MRSS prediction intervals with the SRS prediction interval using the aforementioned procedures, we use the expected lengths of the prediction intervals as a comparison criterion. To do so, we let  $T$  denote a random variable in the set  $\{F, G, H, E, I, J, K, L\}$  where  $T_\alpha$  denotes the  $\alpha^{th}$  quantile of the random variable  $T$ . The results concerning the previously mentioned random variables are presented in Tables 1 and 2 for different set size ( $m$ ) and assuming the significance level  $\alpha = 0.05$ . More precisely, Table 1 gives the lower and upper quantiles for  $n = 5$  future samples while Table 2 presents these quantiles for  $n = 10$  future samples. The results of Tables 1 and 2 will be used to obtain the length of the prediction intervals using the two schemes, namely SRS and MRSS.

TABLE 2  
 The upper and lower quantiles of G, H, E, J, K and L when  $n = 10$  and  $\alpha = 0.05$

$m$	G	H	E	J	K	L
2	-0.9659	-3.7409	0.0872	-1.0959	-3.9786	0.0655
	0.9641	-0.0789	3.7266	1.0962	-0.0604	3.9673
5	-0.2516	-0.9647	0.1035	-0.3355	-1.2662	0.1030
	0.2399	-0.1064	0.9692	0.3361	-0.1049	1.2548
8	-0.1485	-0.6026	0.0859	-0.2296	-0.8534	0.0957
	0.1458	-0.0859	0.6101	0.2129	-0.0963	0.8615
12	-0.0999	-0.4251	0.0667	-0.1674	-0.6479	0.0842
	0.0972	-0.0650	0.4252	0.1572	-0.0858	0.6550

To obtain the expected interval length, we need to define  $Q$  as follows

$$Q = \begin{cases} \sqrt{\sum_{j=1}^{2m} (X_j - \bar{X})^2}, & \text{if SRS scheme is used,} \\ \sqrt{\sum_{i=1}^2 \sum_{j=1}^m (Y_{ij} - \bar{Y})^2}, & \text{if MRSS scheme is used.} \end{cases}$$

Then the expected length of any interval has the form

$$E(\text{Interval length}) = (T_{1-\alpha/2} - T_{\alpha/2})E(Q).$$

It is possible to obtain  $E(Q)$  in a closed form in the case of SRS while the mission might not be feasible in MRSS case. As a result, we use simulation to obtain the values of the expected interval length. Since the normal distribution belongs to the location scale family, then without loss of generality, we may assume  $\mu=0$  and  $\sigma=1$ . Table 3 contains the expected length of the prediction intervals for SRS and MRSS for different values of  $m$  and  $n$  based on 5000 Monte-Carlo simulations. It can be seen from Table 3 that the MRSS intervals are shorter than the SRS intervals in terms of their expected length. Moreover, the larger the sample size, the smaller the expected length.

TABLE 3  
The expected interval length of prediction intervals when  $\alpha = 0.05$

$n$	$m$	$F$	$G$	$H$	$E$	$I$	$J$	$K$	$L$
5	2	6.4590	3.6893	5.6793	5.4940	6.6891	3.9512	5.8596	5.8209
	5	4.3635	2.1903	3.1603	3.1409	4.6997	2.4099	3.4501	3.5106
	8	4.1074	1.9779	2.8790	2.8548	4.3165	2.1460	3.1278	3.0944
	12	4.0220	1.8538	2.7504	2.7474	4.1770	1.9989	2.9246	2.9922
	$m$								
10	2	6.4590	3.1671	6.0069	5.9697	6.6891	3.5088	6.2716	6.2454
	5	4.3635	1.7104	2.9867	3.0126	4.6997	1.9566	3.3833	3.3556
	8	4.1074	1.4868	2.6099	2.6476	4.3165	1.6824	2.8796	2.9116
	12	4.0220	1.3635	2.4924	2.4816	4.1770	1.5390	2.6654	2.7063

## 6. EFFECT OF RANKING ERRORS

In ranked set sampling, the resulting sample is obtained under the assumption that the error in personal judgment is absent. However, we can not ignore the ranking errors in an RSS sample. As mentioned in section 1, many authors have studied the effect of ranking errors on the efficiency of the RSS estimation approach. To name a few, (Dell and Clutter, 1972; Stokes, 1976 and Nahhas *et al.*, 2004) proposed some models for visual ranking errors. In this section, we consider the model proposed by (Dell and Clutter, 1972) where they assumed that the  $i^{th}$  visual score for the  $i^{th}$  observation in RSS set is defined as  $V_i = X_i + \tau_i$ , where  $\tau_1, \tau_2, \dots, \tau_n$  are independent and identically distributed as  $N(0, \sigma_\tau^2)$  independent of the  $X_i$ 's. To obtain an RSS sample with ranking errors, according to the additive model, we adopt the following steps:

1. Obtain  $V_i = X_i + \tau_i$ , where  $\tau_1, \tau_2, \dots, \tau_n$  are independent and identically  $N(0, \sigma_\tau^2)$ .
2. Rank the  $V_i$ 's in an ascending order so that we may obtain  $V_{(1)} \leq V_{(2)} \leq \dots \leq V_{(n)}$ . Also, let  $X_{[i]}$  denote the value of  $X$  associated with the  $i^{th}$  value  $V_{(i)}$ .

3. The values  $X_{[1]}, X_{[2]}, \dots, X_{[n]}$  represent an RSS sample with raking errors.

TABLE 4  
 Values of  $E(Q)$  for different values of  $\sigma_\tau^2$  and  $m$

$\sigma_\tau^2$	$m$	$E(Q)$	
		With ranking errors	Without ranking error
0.1	2	1.6217	1.6406
	3	2.3227	2.3017
	5	3.4766	3.4608
	8	5.0514	5.0560
	12	6.9109	6.91263
0.5	2	1.6229	1.6330
	3	2.2668	2.3016
	5	3.3814	3.4919
	8	4.8278	5.0492
	12	6.5332	6.9138
1.0	2	1.6281	1.6363
	3	2.2991	2.2992
	5	3.1270	3.4993
	8	4.4728	5.0554
	12	5.9191	6.9250

To obtain an MRSS with ranking errors, we adopt the above model and replace  $n$  by  $j$ . Accordingly, we use the  $X_{[j]}$ 's in stead of  $X_{(j)}$ 's. Since  $T_{1-\alpha/2}$  and  $T_{\alpha/2}$  are obtained from the exact distribution via simulation without ranking errors, then it is sufficient to make the comparison only based on the values of  $E(Q)$ . Table 4 contains the values  $E(Q)$  for different values of  $\sigma_\tau^2$  and  $m$ . It can be clearly seen that the larger the value of  $\sigma_\tau^2$ , the larger the difference between the two values of  $E(Q)$ . Also for small values of  $m$ , the difference between the two values of  $E(Q)$  gets smaller.

To elaborate more on this idea, we highlight some of the recent work related to imperfect ranking errors. (Frey, 2007) proposed a model for imperfect ranking that assumed to be valid and flexible to cover a wide range of judgment ranking errors. On the other hand, (Presnell and Bohn, 1999) have considered the model of (Dell and Clutter, 1972) in addition to an artificial model for judgment ranking error. As an illustration, they showed that the imperfect ranking has asymptotic efficiency of 8/3 relative to perfect ranking if their judgment ranking is considered such that

$$J_1 = \begin{cases} 1, & |X_2 - 0.5| > |X_1 - 0.5| \\ 2, & \text{otherwise} \end{cases}$$

where  $X_1$  and  $X_2$  are identical and independently distributed as  $U(0,1)$  and  $J_1 = 1$  if the first order statistic is ranked without error. This example led them to conclude that the efficiency induced based on perfect ranking versus imperfect ranking can not be decisive to either of the two sides. In our simulation study, we

see that the numerical results are comparable for small values of  $\sigma_\tau^2$  and they are not far away when  $\sigma_\tau^2$  is large. It is important to note that small values of  $\sigma_\tau^2$  imply more accurate ranking, while large values of  $\sigma_\tau^2$  imply more ranking errors.

According to these studies in addition to the findings of this article, we may conclude that the imperfect ranking may have positive or negative effect on estimation accuracy depending on the model that is considered for judgment ranking errors. For this, the numerical results presented in this section lie among the findings of (Presnell and Bohn, 1999) and therefore we expect to have an effect on the expected length of the prediction intervals due to ranking errors.

## 7. APPLICATION TO GRASSLAND BIODIVERSITY DATA

Since the introduction of the RSS concept by (McIntyre, 1952), the idea has received an extensive attention due to its tremendous valuable applications in applied fields including but not limited to engineering, communications and ecology (Patil, 1995). For example, (Halls and Dell, 1966) utilized the RSS technique to estimate the weights of browse and herbage in a pine-hardwood forest of east Texas. They concluded that RSS is more efficient than SRS. Similarly, RSS was found to be more robust when applied for estimating the shrub phytomass in forest stands (Martin *et al.*, 1980). Further applications of RSS can be found in (Evans, 1967 and Cobby *et al.*, 1985).

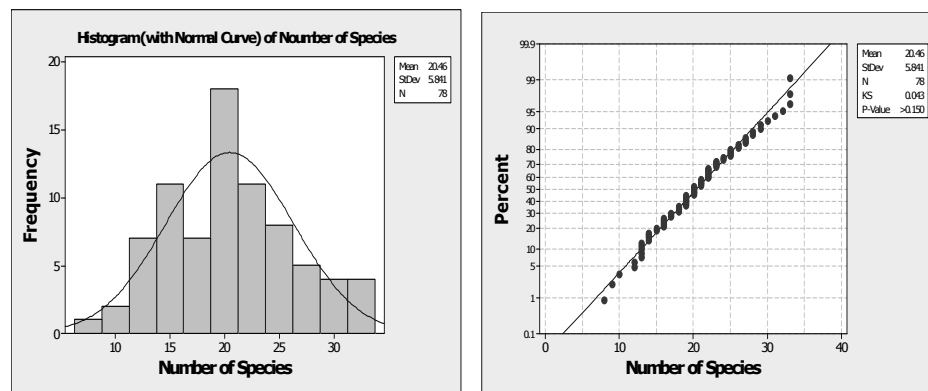


Figure 1 – Histogram with normal curve and normal probability plot for the number of species data.

In our study, we used a data set on grassland biodiversity in central Europe to illustrate the usefulness of our statistical method. This data was collected based on a biodiversity project carried out in the Thuringer Schiefergebirge/Frankenwald, a plateau-like mountain range at the Thuringian/Bavarian border in central Germany with a maximum height of 870 meters. The average annual temperature in this area varies between 68F and 78F and the average annual precipi-

tation varies between 950 and 1099 mm (Perner *et al.*, 2005). The studied plant communities located between 11.018° and 11.638° eastern longitudes and between 50.358° and 50.578° northern latitudes comprising a total of one hectare. We used grassland biodiversity in 78 sites. The data set was considered by (Alodat *et al.*, 2011) for estimating the mean and the standard deviation using MRSS. The histogram and the normal probability plot are given in figure 1. The histogram suggests a normal distribution while the normal probability plot shows a p-value larger than 0.15. So, the data provides us with no evidence to reject the normality assumption. In this section, we apply the MRSS procedure to the 78 observations and we divide the first 35 observations into 7 sets of sizes 2, 3, ..., 8, and the last 35 observations into 7 sets of sizes 2, 3, ..., 8. Our plan is carried out by selecting the minimum of each set in the first group while we selected the maximum of each set in the second group (see table 5).

TABLE 5  
*Illustration of MRSS for the Grassland Biodiversity*

Group	Set	Minimum	Maximum
1.	{9, 21}	9	
	{15, 22, 17}	15	
	{18, 17, 19, 19}	17	
	{8, 33, 22, 33, 23}	8	
	{23, 19, 20, 19, 17, 12}	12	
	{18, 23, 26, 24, 13, 10, 15}	10	
	{23, 13, 22, 13, 19, 16, 21, 24}	13	
2.	{22, 13}		22
	{16, 22, 31}		31
	{23, 14, 13, 19}		23
	{16, 22, 20, 18, 21}		22
	{20, 25, 30, 28, 21, 27}		30
	{25, 14, 28, 33, 20, 22, 21}		33
	{27, 25, 27, 29, 26, 14, 25, 32}		32

Accordingly, we obtain the set of minima {9, 15, 17, 8, 12, 10, 13} and the set of maxima {22, 31, 23, 22, 30, 33, 32}. On the other hand, a simple random sample of size 14 is also obtained and the elements of this SRS are {9, 21, 15, 22, 17, 18, 17, 19, 19, 8, 33, 22, 33, 23}. The data set obtained via MRSS provides the values  $\bar{Y} = 19.786$  and  $Q = 32.594$  while the SRS data set gives the following results:  $\bar{X} = 19.714$  and  $Q = 25.862$ . Also, we used simulation to obtain the quantiles of  $F, G, H, E, I, J, K$  and  $L$  for  $m = 7$  and the results are summarized in table 6. From this table, we see that the length of each prediction interval using MRSS is smaller than the length of each prediction interval using SRS.

TABLE 6  
*Prediction intervals for species data using SRS and MRSS*

Statistic	Prediction Interval	Length
<i>F</i>	(4.8092, 34.6155)	29.8063
<i>G</i>	(12.4632, 26.9784)	14.5153
<i>H</i>	(-0.33706, 20.2945)	20.6315
<i>E</i>	(18.6707, 40.2393)	21.5686
<i>I</i>	(3.60571, 35.655)	32.0493
<i>J</i>	(11.6388, 35.655)	24.0162
<i>K</i>	(-2.71866, 20.8626)	23.5812
<i>L</i>	(18.5213, 41.1684)	22.6471

## 8. CONCLUSIONS

In this paper, we considered the problem of constructing classical prediction intervals for new observation, mean, minimum and maximum of a future sample from the normal distribution under the MRSS scheme. In terms of the interval expected length, we produced prediction intervals that are shorter than those obtained via the SRS scheme. This kind of prediction can be easily extended to other distribution families. Moreover, prediction intervals concerning characteristics of a future sample obtained via different ranked set sampling schemes could be considered. So we leave this to future research.

## ACKNOWLEDGMENT

We would like to thank the referee for the valuable comments that helped us to improve the content of this article.

*Department of Statistics,  
 Yarmouk University, Jordan  
 rawwash@yu.edu.jo*

MOHAMMAD AL-RAWWASH

*Department of Statistics,  
 Yarmouk University, Jordan  
 alodatmts@yahoo.com*

MOH'D ALODAT

*Department of Statistics,  
 Yarmouk University, Jordan  
 aludaatk@hotmai.com*

KHALED ALUDAAT

*Department of Biological Sciences,  
 Al-Hussein Bin Talal University, Jordan  
 odatnidal@hotmail.com*

NIDAL ALODAT

*Department of Biological Sciences,  
 Yarmouk University, Jordan,  
 riyadmuhaidat@yahoo.com*

RIYAD MUHAIDAT

## REFERENCES

- M. T. ALODAT, K. M. ALUDAAT, N. ODAT, R. MUHAIDAT (2011), *Parameters estimation of normal population via moving ranked set sampling with application to grassland biodiversity*. Submitted.
- M. T. ALODAT, M. Y. AL-RAWWASH, I. M. NAWAJAH (2010), *Analysis of simple linear regression model via ranked set sampling*, "Communications in Statistics-Theory and Methods", 39(14), 2604-2616.
- M. T. ALODAT, M. Y. AL-RAWWASH, I. M. NAWAJAH (2009), *Analysis of simple linear regression via median ranked set sampling*, "METRON - International Journal of Statistics", vol. LXVII, n. 1, pp. 1-18.
- M. T. ALODAT, M. F. AL-SALEH (2001), *Variation of ranked set sampling*, "Journal of Applied Statistical Sciences", 10, 137-146.
- M. F. AL-SALEH, S. A. AL-HADRAMI (2003), *Parametric estimation for the location parameter for symmetric distributions using moving extremes ranked set sampling with application to trees data*, "Environmetrics", 14(7), 651-664.
- M. F. AL-SALEH, A. I. ALOMARI (2002), *Multistage ranked set sampling*, "Journal of Statistical planning and Inference", 102, 273-286.
- C. B. ARNOLD, N. BALAKRISHNAN, H. N. NAGARAJA (1992), *A first course in order statistic*, John Wiley and Sons, Inc, New York.
- Z. CHEN (2000), *The efficiency of ranked-set sampling relative to simple random sampling under multi-parameter families*, "Statistics Sinica", 10, 247-263.
- J. M. COBBY, M. S. RIDOUT, P. J. BASSETT, R. V. LARGE (1985), *An investigation into the use of ranked set sampling on grass and grass-clover swards*, "Grass and Forage Science", 40, 257, 263.
- T. R. DELL, J. L. CLUTTER (1972), *Ranked set sampling theory with order statistics background*, "Biometrics", 28, 545-553.
- M. J. EVANS (1967), *Application of ranked set sampling to regeneration surveys in areas direct-seeded to longleaf pine*, Masters Thesis, School of Forestry and Wildlife Management, Louisiana State University, Baton Rouge.
- J. C. FREY (2007), *New important rankings models for ranked set sampling*, "Journal of Statistical Planning and Inference", 137, 1433-1445.
- L. K. HALLS, T. R. DELL (1966), *Trial of ranked set sampling for forage yields*, "Forest Science", 12, 22-26.
- A. KAUR, G. P. PATIL, A. K. SINHA, C. TAILLIE (1995), *Ranked set sampling: an annotated bibliography*, "Environmental Ecological Statistics", 2, 25-54.
- W. L. MARTIN, T. L. SHARK, R. G. ODERWALD, D. W. SMITH (1980), *Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian oak forest*, Publication Number FWS-4-80, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, VA.
- G. A. MCINTYRE (1952), *A method of unbiased selective sampling using ranked sets*, "Australian Journal of Agricultural Research", 3, 385-390.
- H. MUTTLAK (1997), *Median ranked set sampling*, "Journal of Applied Statistical Sciences", 6, 245-255.
- R. W. NAHHAS, D. A. WOLFE, H. CHEN (2004), *Ranked set sampling: ranking error models and estimation of visual judgment error variance*, "Biometrical Journal", 46(2), 255-263.
- G. P. PATIL (1995), *Statistical ecology and related ecological statistics-25 years*, "Environmental and Ecological Statistics", 2, 81-89.
- J. PERNER, C. WYTRYKUSH, A. KAHMEN, N. BUCHMANN, I. EGERER, S. CREUTZBURG, N. ODAT, V. AUDORFF, W. WEISSER (2005), *Effects of plant diversity, plant productivity and habitat parameters on arthropod abundance in mountain European grasslands*, "Ecography", 28, 429-442.
- B. PRESNELL, L. L. BOHN (1999), *U-Statistics and imperfect ranking in ranked set sampling*, "Journal of Nonparametric Statistics", 10, No. 2, 111-126.

- K. TAKAHASI, K. WAKIMOTO (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.
- H. M. SAMAWI, M. S. AHMAD, W. ABU-DAYYEH (1996), *Estimating the population mean using extreme ranked set sampling*, "The Biometrical Journal", 38 (5), 577-586.
- S. L. STOKES (1976), *An investigation of the consequences of ranked set sampling*, Ph.D. Thesis, Department of Statistics, University of North Carolina, Chapel Hill.

#### SUMMARY

##### *Prediction intervals for characteristics of future normal sample under moving ranked set sampling*

In this article, we derive prediction intervals for the characteristics of a future sample from normal population when the sample is selected via moving extreme ranked set sampling. We conduct a simulation study to compare these intervals with their counterparts using simple random sampling. Finally, we apply our findings on grassland biodiversity real data set in central Europe.