RANKED SET SAMPLING

By

Munir Ahmad
M. Hanif
Hassen A. Muttlak

FOREWORD

The Editorial Board of Pakistan Journal of Statistics, in its meeting held in September, 2009 to collect papers on one topic and published in the form of Book. Each paper has been refereed by at least three experts actively engaged in “Ranked Set Sampling”. This book is the first in the series and we hope that in future, we shall be collecting papers and publishing in the form of books.

MUNIR AHMAD
Editor-in-Chief PJS and
Rector, NCBA&E, Lahore, Pakistan

PREFACE

Recently attention is being paid to the basic concepts of “Ranked Set Sampling” and there are a number of papers available in the literature. New techniques and approaches are being studied recently but there is no collection of papers that provide recent developments in the area. The motivation of this book is the amount of recent papers published by various authors on the topic of “Ranked Set Sampling” in Pakistan Journal of Statistics.

Our main objective is to present before a wider audience on the work done on “Ranked Set Sampling” during the last decade and to motivate statisticians in this part of the world to work on some latest statistical technologies developed in various aspects of sampling. This book does not show any overlap with the current developments in the area, instead it has added new approaches to the area, instead it has added new approaches to the area.

We are indebted to all the authors of the papers for their enormous hard work in preparation of the papers and their referees for the quality work they have done and to Mr. M. Imtiaz and Mr. M. Iftikhar for excellent job of reproduction / composition of papers and setting in the proper format.

Munir Ahmad
M. Hanif
Hassen A. Muttlak

CONTENTS

CHAPTER ONE

Stratified Ranked Set Sample

Hani M. Samawi
Jiann-Ping Hsu College of Public Health, Biostatistics Center, PO Box 8015
Georgia Southern University, Statesboro, GA 30460, USA
E-mail: hsamawi@georgiasouthern.edu

ABSTRACT

Stratified simple random sampling (SSRS) is used in certain types of surveys because it combines the conceptual simplicity of simple random sample (SRS) with potentially significant gains in efficiency. It is a convenient technique to use whenever we wish to ensure that our sample is representative of the, population and also to obtain separate estimates for parameters of each subdomain of the population. If sampling units in a study can be easily ranked compared to quantification, McIntyre (1952) proposed to use the mean of n units based on a ranked set sample (RSS) to estimate the population mean, and observed that it provides an unbiased estimator with a smaller variance compared to SRS of the same size n.

In this paper we introduce the concept of stratified ranked set sample (SRSS) for estimating the population mean. SRSS combines the advantages of stratification and RSS to obtain an unbiased estimator for the population mean, with potentially significant gains in efficiency. 'The conclusion of this study is that by using SRSS the efficiency of the estimator relative to SSRS and SRS has strictly increased. Results from uniform distribution are given. Computer simulated results on other distributions are also given. An example using real data is presented to illustrate the computations.

KEY WORDS

Simple random sample, stratified random sample, ranked set sample, stratified ranked set sample, order statistics.

1. INTRODUCTION

Ranked set sampling (RSS) was introduced by McIntyre (1952) to estimate the pasture yield. RSS procedure involves randomly drawing *n* sets of *n* units each from the population for which the mean is to be estimated. It is assumed that the units in each set can be ranked visually. From the first set of *n* units, the unit ranked lowest is measured. From the second set of *n* units, the unit ranked second lowest is measured. The process is continued until from the *n – th* set of n units the *n – th* ranked unit is measured. Talcahasi and Wakimoto (1968) warned that in practice the number of units which are easily ranked cannot be more than four.

A stratified simple random sample (SSRS), (for example see Hansen et al. 1953) is a sampling plan in which a population is divided into *L* mutually exclusive and exhaustive strata, and a simple random sample (SRS) of n_h elements is taken and quantified within each stratum h. The sampling is performed independently across the strata. In essence, we can think of a SSRS scheme as one consisting of L separate simple random samples.

A stratified ranked set sample (SRSS) is a sampling plan in which a population is divided into L mutually exclusive and exhaustive strata, and a ranked set sample (RSS) of n_h elements is quantified within each stratum, h =1, 2, …, L. The sampling is performed independently across the strata. Therefore, we can think of a SRSS scheme as a collection of L separate ranked set samples.

In this paper, we introduce the concept of SRSS to estimate the population mean. This study showed that the estimator using SRSS is at least more efficient than the one using SSRS. In Section 2, we describe some sampling plans, discuss estimation of population mean using these plans, and give some useful definitions and general results and results for the uniform distribution. Simulation results from non-uniform distributions are given in Section 3. In Section 4, we illustrate the method using real data. The discussion is given in Section 5.

2. SAMPLES AND ESTIMATION OF POPULATION MEAN

Suppose that the population is divided into L mutually exclusive and exhaustive strata. Let $X^*_{h11}, X^*_{h12}, \ldots, X^*_{h1n_h}; X^*_{h21}, X^*_{h22}, \ldots, X^*_{h2n_h}; \ldots; X^*_{hn_h1}, X^*_{hn_h2}, \ldots, X^*_{hn_hn_h}$ be n_h independent random samples of size n_h each one is taken from each stratum $(h = 1, 2, \ldots, L)$. Assume that each element X^*_{hij} in the sample has the same distribution function $F_h(x)$ and density function $f_h(x)$ with mean μ_h and variance σ^2_h. For simplicity of notation, we will assume that X_{hij} denotes the quantitative measure of the unit X^*_{hij}. Then, according to our description $X_{h11}, X_{h21}, \ldots, X_{hn_h1}$ could be considered as the SRS from the $h - th$ stratum. Let $X^*_{hi(1)}, X^*_{hi(2)}, \ldots, X^*_{hi(n_h)}$ be the ordered statistics of the $i - th$ sample

$X^*_{hi1}, X^*_{hi2}, ..., X^*_{hin_k}$ $(i = 1, 2, ...n_k)$ taken from the $h-th$ stratum. Then, $X_{h1(1)}, X_{h1(2)}, ..., X_{hn_h(n_h)}$ denotes the RSS for the $h-th$ stratum. If $N_1, N_2, ..., N_L$ represent the number of sampling units within respective strata, and $n_1, n_2, ..., n_L$ represent the number of sampling units measured within each stratum, then $N = \sum_{h=1}^{L} N_h$ will be the total population size, and $n = \sum_{h=1}^{L} n_h$ will be the total sample size.

2.1 Definitions, notations and some useful results

The following notations and results will be used throughout this paper. For all $i, i = 1, 2, ..., n_h$ and $h = 1, 2, ..., L$, let $\mu_h = E\left(X_{hij}\right), \sigma_h^2 = Var\left(X_{hij}\right)$, $\mu_{h(i)} = E\left(X_{hi(i)}\right)$, $\sigma^2_{h(i)} = Var\left(X_{hi(i)}\right)$, for all $j = 1, 2, ..., n_h$ and let $T_{h(i)} = \mu_{h(i)} - \mu_h$.

As in Dell and Clutter (1972), one can show easily that for a particular stratum $h, (1 = 1, 2, ..., L)$,

$$f_h(x) = \frac{1}{n_h} \sum_{i=1}^{n_h} f_{h(i)}(x),$$

and hence $\sum_{i=1}^{n_h} \mu_{h(i)} = n_h \mu_h$, $\sum_{i=1}^{n_h} T_{h(i)} = 0$ and $\sum_{i=1}^{n_h} \sigma^2_{h(i)} = n_h \sigma_h^2 - \sum_{i=1}^{n_h} T^2_{h(i)}$

The mean μ of the variable X for the entire population is given by

$$\mu = \frac{1}{N} \sum_{h=1}^{L} N_h \mu_h = \sum_{h=1}^{L} W_h \mu_h \qquad (2.1.1)$$

where $W_h = \dfrac{N_h}{N}$.

If within a particular stratum, h, we suppose to have selected SRS of n_h elements from N_h elements in the stratum and each sample element is measured with respect to some variable X, then the estimate of the mean ph using SRS of size n_h is given by

$$\bar{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} X_{hi1}. \qquad (2.1.2)$$

The mean and variance of $\bar{X}_h$ are known to be $E\left(\bar{X}_h\right)=\mu_h$ and $Var\left(\bar{X}_h\right)=\frac{\sigma_h^2}{n_h}$ respectively, assuming N_h's are large enough. The estimate of the population mean μ using SSRS of size n is defined by

$$\bar{X}_{SSRS}=\frac{1}{N}\sum_{h=1}^{L} N_h\bar{X}_h=\sum_{h=1}^{L} W_h\bar{X}_h \tag{2.1.3}$$

The mean and the variance of $\bar{X}_{SSRS}$ are known to be $E\left(\bar{X}_{SSRS}\right)=\mu$ and

$$Var\left(\bar{X}_{SSRS}\right)=\sum_{h=1}^{L} W_h^2\left(\frac{\sigma_h^2}{n_h}\right) \tag{2.1.4}$$

respectively, assuming N_h's are large enough.

If within a particular stratum h, we suppose to have selected RSS of n_h elements from N_h elements in the stratum and each sample element is measured with respect to some variable X, then the estimate of the mean μ_h using RSS of size μ_h is given by

$$\bar{X}_h=\frac{1}{n_h}\sum_{i=1}^{n_h} X_{hi(i)} \tag{2.1.5}$$

It can be shown that the mean and variance of $\bar{X}_{h(n_h)}$ are $E\left(\bar{X}_{h(n_h)}\right)=\mu_h$ and

$$Var\left(\bar{X}_{h(n_h)}\right)=\frac{\sigma_h^2}{n_h}-\frac{1}{n_h^2}\sum_{i=1}^{n_h} T_{h(i)}^2, \tag{2.1.6}$$

respectively, assuming N_h's are large enough. Therefore, the estimate of the population mean μ using SRSS of size n is defined by

$$\bar{X}_{SRSS}=\frac{1}{N}\sum_{h=1}^{L} N_h\bar{X}_{h(n_h)}=\sum_{h=1}^{L} W_h\bar{X}_{h(n_h)}. \tag{2.1.7}$$

It can be shown straightforward algebra that the mean and the variance of $\bar{X}_{SRSS}$ are $E\left(\bar{X}_{SRSS}\right)=\mu$ (i.e., and unbiased estimator) and

$$Var\left(\bar{X}_{SRSS}\right)=\sum_{h=1}^{L} W_h^2\left(\frac{\sigma_h^2}{n_h}-\frac{1}{n_h^2}\sum_{i=1}^{n_h} T_{h(i)}^2\right), \tag{2.1.8}$$

respectively, assuming N_h 's are large enough.

Therefore, the relative efficiency of the estimator of the population mean μ using SRSS with respect to the one using SRSS can be defined by

$$RE=\frac{Var\left(\bar{X}_{SSRS}\right)}{Var\left(\bar{X}_{SRSS}\right)}=\frac{1}{\left\{1-\frac{1}{Var\left(\bar{X}_{SSRS}\right)}\left(\sum_{h=1}^{L}\frac{W_h^2}{n_h^2}\sum_{i=1}^{n_h} T_{h(i)}^2\right)\right\}} \tag{2.1.9}$$

2.2 Results for the uniform distribution

Assume that a population of size N, with a variable X has a uniform distribution $U\left(U,\theta\right)$. Suppose we can divide this population into L strata with respect to some characteristics in the population. If we let $N_1, N_2, ..., N_L$ represent the number of sampling units within respective strata, and $n_1, n_2, ..., n_L$ represent the number of selected sampling units from respective strata, then $N=\sum_{h=1}^{L} N_h$ and $n=\sum_{h=1}^{L} n_h$. Assume that the random variable X_h has distribution $U\left(0,\theta_h\right)$. Thus, $\mu_h=E\left(X_h\right)=\frac{\theta_h}{2}$ and $\sigma_h^2=Var\left(X_h\right)=\frac{\theta_h^2}{12}$. Also, $\theta=\frac{1}{N}\sum_{h=1}^{L} N_h\theta_h=\sum_{h=1}^{L} W_h\theta_h$.

The mean and variance of the estimate $\bar{X}_{SSRS}$ of the population mean μ using SSRS of size n ware $E\left(\bar{X}_{SSRS}\right)=\frac{\theta}{2}=\mu$ and

$$Var\left(\bar{X}_{SSRS}\right)=\sum_{h=1}^{L} W_h^2\left(\frac{\theta_h^2}{12n_h}\right) \tag{2.2.1}$$

respectively.

The mean and variance of the estimate $\bar{X}_{SRSS}$ of the population mean μ using SRSS of size n are $E\left(\bar{X}_{SRSS}\right)=\frac{\theta}{2}=\mu$ and

$$Var\left(\bar{X}_{SRSS}\right) = \sum_{h=1}^{L} W_h^2 \left(\frac{\theta_h^2}{6n_h\left(n_h+1\right)} \right) \tag{2.2.2}$$

Also, if $n_h \geq 2$ then,

$$RE = \frac{Var\left(\bar{X}_{SSRS}\right)}{Var\left(\bar{X}_{SRSS}\right)} = \frac{\sum_{h=1}^{L} W_h^2 \left(\frac{\theta_h^2}{n_h} \right)}{2\sum_{h=1}^{L} W_h^2 \left(\frac{\theta_h^2}{n_h\left(n_h+1\right)} \right)} > 1\,, \tag{2.2.3}$$

which implies that SRSS gives a more efficient unbiased estimator for the uniform population mean compared to SSRS.

3. SIMULATION STUDY

The normal and exponential distributions are used in the simulation. Sample sizes $N = 10, 20$ and 30 and number of strata $L = 3$ are considered. For each of the possible combination of distribution, sample size and different choice of parameter, 2000 data sets were generated. The relative efficiencies of the estimate of the population mean using SRSS with respect to SSRS, SRS and RSS are obtained. All computer programs were written in Borland TURBO BASIC.

3.1 Result of the Simulation Study

The values obtained by simulation are given in Table 1. Our simulation indicates that estimating the population means using. SRSS is more efficient than using SSRS or SRS. In some cases, when the underlying distribution is normal with ($\mu_1 = 1.0, \quad \mu_2 = 3.0, \quad \mu_3 = 5.0$), the simulation indicates that estimating the population mean using SRSS is even more efficient than RSS.

Table 1
The relative efficiency of the simulation results

Distribution function	n	$RE\left(\bar{X}_{SRSS},\bar{X}_{SSRS}\right)$	$RE\left(\bar{X}_{SRSS},\bar{X}_{SRS}\right)$	$RE\left(\bar{X}_{SRSS},\bar{X}_{RSS}\right)$
Normal $W_1=0.3, W_2=0.3, W_3=0.4,$ $\mu_1=1.0,\mu_2=3.0,\mu_3=5.0$ $\sigma_1=1.0,\sigma_2=1.0,\sigma_3=1.0$	10 20 30	2.04 3.19 4.45	7.59 11.63 16.30	1.50 1.29 1.25
Normal $W_1=0.3, W_2=0.3, W_3=0.4,$ $\mu_1=1.0,\mu_2=2.0,\mu_3=3.0$ $\sigma_1=1.0,\sigma_2=1.0,\sigma_3=1.0$	10 20 30	2.08 3.19 4.42	3.48 5.70 7.57	0.72 0.65 0.67
Normal $W_1=0.3, W_2=0.3, W_3=0.4,$ $\mu_1=1.0,\mu_2=3.0,\mu_3=5.0$ $\sigma_1=1.0,\sigma_2=1.1,\sigma_3=1.2$	10 20 30	2.08 3.38 4.32	7.15 10.50 13.94	1.28 1.19 1.10
Exponential $W_1=0.3, W_2=0.3, W_3=0.4,$ $\mu_1=1.0,\mu_2=3.0,\mu_3=5.0$	10 20 30	2.82 3.04 3.50	3.55 3.78 4.17	1.28 0.96 0.86
Exponential $W_1=0.3, W_2=0.3, W_3=0.4,$ $\mu_1=5.0,\mu_2=10.0,\mu_3=15.0$	10 20 30	1.95 2.85 3.53	2.15 3.25 4.15	0.73 0.71 0.74

4. EXAMPLE: Body Mass Index Data

In Table 2 we present three sample of size 7 each, from baseline interview data for the Iowa 65+ Rural Health Study (RHS), which is a longitudinal cohort study of 3,673 individuals (1,420 men arid 2,253 women) ages 65 or older living in Washington and Iowa countries of the State of Iowa in 1982. This study is one of four supported by the National Institute on Aging and collectively referred to as EPESE, (Established Populations for Epidemiologic Studies of the Elderly), National Institute on Aging, 1986.

In the Iowa 65+ RHS there were 33 diabetic women aged 80 to 85, of whom 14 reported urinary incontinence. The question of interest is to estimate the mean body mass index (BMI) of diabetic women. The BMI is the ratio of the subject's weight (kilograms) divided by height (meters) squared. Note that, the BMI may be different for women with or without urinary incontinence. Thus, here is a situation where stratification might work well. The 33 women were divided into two strata, the first consists of those women without urinary incontinence and the second consists of those 14 women with urinary incontinence. Four samples of size $(n = 7)$ each were drawn from those women using SSRS, SRSS, RSS and SRS. Note that in case of SRSS and RSS the selecting samples are drawn with replacement. The calculated values of BMI are given in Table 2. These calculations indicate the same pattern of conclusions that were obtained earlier, and illustrate the method described in Section 2.

Table 2
Body Mass Index Samples of Diabetic Women Aged 80 to 85 Years with and without Urinary Incontinence

	SRS	RSS		SSRS	SRSS
	18.88	18.88		23.45	23.45
	19.76	22.88	Stratum 1	28.95	23.46
	20.57	23.45		30.17	30.10
	25.66	24.38		19.61	19.61
	26.01	26.30	Stratum 2	24.07	24.38
	28.95	27.31		27.49	31.31
	33.52	36.65		33.52	31.95
Estimated Mean	24.77	25.69		26.95	26.15
Standard Error	2.03	2.06		1.72	1.67

5. DISCUSSION

The BMI data is a good example where we need stratification to find an unbiased estimator for the population mean of those diabetic women aged 80 to 85 years. Since the 33 women were divided into two strata, the first consists of those women without urinary incontinence and the second consists of those women with urinary incontinence. It is clear that the mean of the BMI in each stratum will be different. Also, women can be ranked visually according to their BMI. In this situation we recommend using SRSS to estimate the mean BMI of those women. SRSS will give an unbiased and more efficient estimate of the BMI mean. Moreover, SRSS can provide an unbiased and more efficient estimate for the mean of each stratum.

Remark: We could not find a closed form for optimal allocation of units and also for optimal allocation of resources for n_h using SRSS. However, the near optimal allocation can be obtained from the formulae obtained by using SSRS, for example see, Hansen et al. 1953.

REFERENCES

1. Brock, D.B., Wineland T. Freeman, D.H., Lemke, J.H., Scherr, P.A. (1986), Demographic characteristics. In: Established Population for Epidemiologic Studies of the Elderly. Resource Data Book, Cornoni Huntley, J. Brock D.B., Ostfeld, A.M., Taylor, J.O. and Wallace, R.B. (eds). *National Institute on Aging*, NM Publication No. 86-2443. U.S. Government Printing Office, Washington, D.C.
2. Dell T.R. and Clutter J.L. (1972). Ranked Set Sampling Theory with Order Statistics Background. *Biometrics* 28, 545-555.
3. Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sampling Survey Methods and Theory*, Vols. 1 and 2, Wiley, New York.
4. McIntyre, G.A. (1952): A method of unbiased selective sampling, using ranked sets. *Australian J. Agri. Research* 8, 385-390.
5. Takahasi, K. and Wakimoto, K. (1968): On unbiased estimates of the population mean based on the stratified sampling by means of ordering. *Ann. Inst. Statist.* Math., 20, 1-31.

CHAPTER TWO

Using Ranked Set Sampling for Hypothesis Tests on the Scale Parameter of the Exponential and Uniform Distributions

Walid Abu-Dayyeh[1] and **Hassan A. Muttlak**[2]
[1] Department of Statistics, Faculty of Economics, University of UAE, P.O. Box 17555, Al-Ain, UAE.
[2] School of Computing & Mathematics, Deakin University, Geelong Campus, Geelong, Vectoria 3217, Australia.

ABSTRACT

The concept of ranked set sampling (RSS) was suggested by McIntyre (1952). Many authors including Takahasi and Wakimoto (1968). Stokes (1980) and Muttlak and McDonald (1990) have used RSS in estimation.

In this paper we will obtain the uniformly most powerful test (UMPT) and the likelihood ratio test (LRT) in case of exponential distribution and the UMPT in case of uniform distribution, using simple random sample (SRS) and then we will adapt the statistics of these tests to construct new tests using RSS. It turns out that the use of RSS gives much better results in terms of the power function compared to SRS.

KEY WORDS

Ranked set sampling; simple random sample; power of the test, UMPT and LRT.

1. INTRODUCTION

In many applications it is very difficult or expensive to measure the sampling units, but the units can be ranked with out any cost. It turns out that in such cases the use of RSS gives better estimate of the population mean compared to the SRS. In agricultural and environmental studies, it is possible to rank the sampling units without actually measuring them. For some such applications see Cobby et al. (1985), Muttlak and McDonald (1992), Johnson et al. (1993) and Patil and Taillie (1993). For the sampling method of RSS see Stokes (1986).

Many other uses of RSS have been studied in the literature. Takahasi and Wakimoto (1968) independently suggested the same method that was considered by McIntyre (1952). They proved that the mean of RSS is an unbiased estimator of the population mean with smaller variance than the variance of the sample mean of a SRS with the same sample size. Stokes (1980) discussed the estimation of the variance based on RSS. Muttlak and McDonald (1990a, 1990b) developed RSS theory when the sampling units are selected with size based probability of selection.

The object of this paper is to obtain the UMPT for the one sided alternative and the LRT for the two sided alternative in case of the exponential distribution and the UMPT for the two sided alternative in case of uniform distribution using SRS and will adapt these tests to RSS data. It turns out that the tests based on RSS have higher power than the corresponding tests based on the SRS.

2. EXPONENTIAL DISTRIBUTION

Let $X_1, X_2, ..., X_n$ be a random sample from the exponential distribution with pdf

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{if } x \geq 0 \\ 0 & \text{Otherwise} \end{cases}$$

We are interested in testing the hypotheses

$$H_0 : \theta = \theta_0 \text{ vs. } H_\alpha : \theta > \theta_0 \tag{1}$$

It is well known that the UMPT of size α for testing (1) is given by

$$\phi UMPT = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i > C_\alpha \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Without lost of generality we may take $\theta_0 = 1$. Then $c_\alpha = \frac{\chi^2_{2n,1-\alpha}}{2}$, where χ^2_m is the chi-square distribution with m degrees of freedom. The power of the test (2) is given by

$$\beta_{\phi UMPT}(\theta) = P_\theta\left(\sum_{i=1}^{n} X_i > \frac{1}{2}\chi^2_{2n,1-\alpha}\right) = P_\theta\left(W > \frac{1}{\theta}\chi^2_{2n,1-\alpha}\right),$$

where W is distributed χ^2_{2n}

To obtain the test using RSS let $X_{11}, X_{12}, ..., X_{1n}; X_{21}, X_{22}, ..., X_{2n}; \ldots\ldots; X_{n1}, X_{n2}, ..., X_{nn}$ be the n groups of n independent random variables all with the same cumulative distribution function $F(x)$. Let $X_{i(1)}, X_{i(2)}, ..., X_{i(n)}$ be the order statistics of the variables $X_{i1}, X_{i2}, ..., X_{in}$ in the i-th group $(i = 1, 2, ..., n)$ Then

$X_{1(1)}, X_{2(2)}, ..., X_{i(i)}, ..., X_{n(n)}$ denotes the ranked set sample, where $X_{i(i)}$ is the i-th order statistic in the i-th group. To simplify the notation, $X_{i(i)}$ will be denoted by Y_i through out this paper.

To test the same hypothesis (1) using the RSS we propose the following test

$$\phi_1 = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} Y_i > d \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where d is determined so that the test ϕ_1 has size α. To obtain the value of d, we need the distribution of $\sum_{i=1}^{n} Y_i$ under H_0. For this purpose we consider the following transformation: $Z_1 = Y_1, Z_2 = Y_1 + Y_2, Z_3 = Y_1 + Y_2 + Y_3, ..., Z_n = \sum_{i=1}^{n} Y_i$. We know that $Y_1, Y_2, ..., Y_n$, are independent random variables with joint pdf:

$$g_\theta(y_1, y_2, ..., y_n) = \begin{cases} \left\{ \prod_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} \left[1 - e^{-y_i/\theta}\right]^{i-1} \right\} \frac{1}{\theta^n} e^{-\sum_{i=1}^{n} (n-i+1) y_i / \theta}, y_i > 0, i = 1, ..., n \\ 0 \quad \text{Otherwise} \end{cases} \tag{4}$$

Then the joint pdf of $Z_1, Z_2, ..., Z_n$ is given by

$$h_\theta(z_n, z_2, ..., z_n) = g_\theta(z_1, z_2 - z_1, z_3 - z_2, ..., z_n - z_{n-1}),$$

which implies that the pdf of Z_n is:

$$k_\theta(z_n) = \int_0^{z_n} \int_0^{z_{n-1}} ... \int_0^{z_n} g_\theta(z_1, z_2 - z_1, z_3 - z_2, ..., z_{n-1}) d_{z_1} d_{z_2 ... d_{z_{n-1}}} \tag{5}$$

Therefore the power function of the test (3) is given by

$$\beta_{\phi_1}(\theta) = P_\theta\left(\sum_{i=1}^{n} Y_i > d\right) \int_d^\infty k_\theta(z_n) d_{z_n},$$

To find d, we need to solve

$$\beta_{\phi_1}(1) = \alpha = \int_d^\infty k_{\theta=1}(z_n) d_{z_n} \tag{6}$$

It is not easy to find d for general n and α. Therefore we will find d for $n = 3,4,$ and 5 and $\alpha = 0.05$. For $n = 3$, the pdf of z_3 is

$$k_\theta(z_3) = \frac{27}{2\theta} e^{-3z_3/\theta} \left\{ e^{2z_3/\theta} + 16e^{2z_3/\theta} - 4 \right\} - \frac{27}{2\theta^3} e^{-3z_3/\theta} \left\{ 13\theta^2 + 10\theta e z_3 + 8\theta_{z_3} e^{z_3/\theta} + 2z_3^2 \right\} \quad (7)$$

For $\alpha = 0.05$ and $n = 3$ we found that $d = 5.532$, using a computer mathematical program. Therefore, the power of this test is given by

$$\beta_{\phi_1}(\theta) = P_\theta\left(\sum_{i=1}^{3} Y_i > d \right) = \int_{5.532}^{\infty} k_\theta(z_3)\, dz_3, \quad (8)$$

Similarly, the power function can be written for $n = 4$ and 5. Table (1) shows the results for $\alpha = 0.05$ and $n = 4$ and 5 and different values of θ.

It appears from Table (1) that the power of the tests $\phi UMPT$ and ϕ_1 increases as θ increases and also as n increases and that the power of ϕ_1 is larger than the power of $\phi UMPT$ i.e. using RSS gives higher power of the test compared to SRS.

Table (1)

Values of $\beta_{UMPT}(\theta)$ and $\beta_{\phi_1}(\theta)$ for different values of θ and sample sizes $n = 3,4$ and 5 and $\alpha = 0.05$

θ	$\beta_{\phi UMPT}(\theta)$			$\beta_{\phi_1}(\theta)$		
	n = 3	$n = 4$	$n = 5$	$n = 3$	$n = 4$	$n = 5$
1.10	0.076	0.079	0.083	0.080	0.087	0.094
1.25	0.122	0.134	0.146	0.138	0.164	0.191
1.50	0.211	0.242	0.258	0.327	0.402	0.402
2.00	0.391	0.458	0.518	0.503	0.642	0.762
3.00	0.650	0.739	0.807	0.807	0.924	0.977
4.00	0.790	0.868	0.918	0.924	0.984	0.998
5.00	0.866	0.928	0.961	0.967	0.996	0.999
10.0	0.974	0.992	0.998	0.999	0.999	0.999

Next we will consider the LRT for testing the hypothesis

$$H_0 : \theta = 1 \text{ vs. } H_\alpha : \theta \neq 1. \quad (9)$$

It is well known that the LRT of size a is given by

$$\phi_{LRT} = \begin{cases} 0 & \text{if } \frac{\chi^2_{2n,\alpha/2}}{2} < \sum_{i=1}^{n} X_i < \frac{\chi^2_{2n,1-\alpha/2}}{2} \\ 1 & \text{Otherwise} \end{cases}$$

which implies that its power function is given by

$$\beta_{\phi UMPT}(\theta) = \quad 1 - P_\theta\left(\frac{\chi^2_{2n,\alpha/2}}{2} < \sum_{i=1}^{n} X_i < \frac{\chi^2_{2n,1-\alpha/2}}{2}\right)$$

$$1 - P_\theta\left(\frac{\chi^2_{2n,\alpha/2}}{\theta} < W < \frac{\chi^2_{2n,1-\alpha/2}}{\theta}\right)$$

where $W = \sum_{i=1}^{n} X_i$ is distributed as χ^2_{2n}.

To test the same hypothesis using the RSS, the following test is proposed:

$$\phi_2 = \begin{cases} 0 & \text{if } k_1 < \sum_{i=1}^{n} Y_i < k_2 \\ 1 & \text{Otherwise} \end{cases}$$

The power function of the test ϕ_2 is

$$\beta_{\phi_2}(\theta) = 1 - P_\theta\left(k_1 < \sum_{i=1}^{n} Y_i < K_2\right) = 1 - \int_{K_1}^{k_2} k_\theta(z_n)\,dz_n,$$

where $k_\theta(z_n)$ is defined in (5). To obtain the test of size α we need to find k_1 and k_2 to satisfy

$$\beta_{\phi_2}(1) = \alpha = 1 - \int_{k_1}^{k_2} k_{\theta=1}(z_n)\,d_{z_n}.$$

We will take $1 - \int_0^{k_1} k_{\theta=1}(z_n)\,dz_n = \alpha/2$ and $1 - \int_0^{k_2} k_{\theta=1}(z_n)\,dz_n = 1 - \alpha/2$. To compare the two tests ϕ_{LRT} and ϕ_2, we take $\sigma = 0.05$ and $n = 3,4$ and 5. Table (2) shows the power for both tests for n = 3,4 and 5 and a = 0.05.

Considering Table (2) we conclude that the power of the tests ϕ_{LRT} and ϕ_2 increases as θ moves away from 1 in both directions and as n increases and the power of ϕ_2 is higher than ϕ_{LRT} i.e. using RSS will increase the power of the test. Also, we notice that ϕ_{LRT} appears to be unbiased test while ϕ_2 is an unbiased test.

Table (2)
Values of $\beta_{LRT}(\theta)$ and $\beta_{\phi_2}(\theta)$ for different values of θ and sample sizes $n = 3, 4$ and 5 and $\alpha = 0.05$

θ	$\beta_{\phi LRT}(\theta)$			$\beta_{\phi_2}(\theta)$		
	$n=3$	$n=4$	$n=5$	$n=3$	$n=4$	$n=5$
0.05	0.999	0.999	0.999	0.999	0.999	0.999
0.10	0.946	0.995	0.999	0.999	0.999	0.999
0.25	0.450	0.633	0.776	0.822	0.969	0.996
0.50	0.129	0.1768	0.228	0.263	0.437	0.612
0.75	0.055	0.063	0.071	0.079	0.109	0.146
0.90	0.046	0.047	0.048	0.051	0.056	0.061
1.00	0.050	0.050	0.050	0.050	0.050	0.050
1.10	0.061	0.062	0.063	0.070	0.062	0.064
1.25	0.087	0.093	0.100	0.092	0.105	0.122
1.50	0.150	0.172	0.194	0.177	0.229	0.289
2.00	0.305	0.365	0.421	0.395	0.530	0.661
3.00	0.569	0.665	0.742	0.731	0.878	0.960
4.00	0.730	0.821	0.883	0.884	0.970	0.995
5.00	0.823	0.899	0.943	0.947	0.992	0.999
10.0	0.963	0.988	0.996	0.997	0.999	0.999

3. UNIFORM DISTRIBUTION

Let $X_1, X_2, ..., X_n$ be a random sample from the uniform distribution **with** probability density function

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

We want to test

$$H_0 : \theta = \theta_0 \text{ vs. } H_\alpha : \theta \neq \theta_0 . \quad (11)$$

As was done in case of the exponential distribution we assume that $\theta_0 = 1$ w.l.o.g. Since the UMPT test for (11) exists there is no need to consider the LRT. The UMPT of side α is given by

$$\phi_\mu = \begin{cases} 1 & \text{if } X_{(n)} > 1 \text{ or } X_{(n)} \le n\sqrt{\alpha} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where $X_{(n)}$ is the largest ordered statistic. Then the power function of this test is

$$\beta_{\phi_\mu}(\theta)=\begin{cases}1 & \text{if } \theta \le n\sqrt{\alpha}\\ \frac{\alpha}{\theta_n} & \text{if } n\sqrt{\alpha} < \theta \le 1\\ 1+\frac{\alpha-1}{\theta_n} & \text{if } \theta > 1\end{cases}$$

To test the same hypothesis using the RSS we propose the following test

$$\phi_3=\begin{cases}1 & \text{if } \max\{Y_i\}<c \text{ or } \max\{Y_i\}>1\\ 0 & \text{otherwise}\end{cases}$$

To find the value of c we must solve the equation

$$a=P_{\theta=1}\left(\max\{Y_i\}<c\right)=\prod_{i=1}^{n}\left(P_{\theta=1}\left(Y_i\right)<c\right)$$

which can be written as

$$\alpha=\prod_{i=1}^{n}\left(\int_0^c \frac{n!}{(i-1)!(n-i)!}\left(\frac{y_i}{\theta}\right)^{i-1}\left(1-\frac{y_i}{\theta}\right)^{n-1}\frac{1}{\theta}dy_i\right)$$

Then the power of this test can be written as

$$\beta_{\phi_2}(\theta)=\begin{cases}1 & \text{if } \theta\le c\\ \prod_{i=1}^{n}P_\theta\left(Y_i\le c\right) & \text{if } c<\theta\le 1\\ \prod_{i=1}^{n}P_\theta\left(Y_i\le c\right)+1-\prod_{i=1}^{n}P_\theta\left(Y_i\le 1\right) & \text{if } \theta>1\end{cases}$$

To compare the two tests ϕ_μ and ϕ_3 we take $\alpha=0.05$ and $n=3,4$ and 5. Table (3) shows the results for $n=3,4$ and 5 and $\alpha=0.05$ with different values of θ.

Considering Table (3) we see that the power of the tests ϕ_μ and ϕ_3 increases as θ moves away from 1 in both directions and as n increases and the power of ϕ_3 is larger than ϕ_μ, i.e. using RSS will increase the power of the test.

Table (3)
Values of $\beta_{\phi_\mu}(\theta)$ and $\beta_{\phi_3}(\theta)$ for different values of θ and sample sizes $n = 3, 4$ and 5 and $\alpha = 0.05$

θ	$\beta_{\phi\mu}(\theta)$			$\beta_{\phi_2}(\theta)$		
	$n = 3$	$n = 4$	$n = 5$	$n = 3$	$n = 4$	$n = 5$
0.25	0.999	0.999	0.999	0.999	0.999	0.999
0.50	0.400	0.800	0.999	0.946	0.999	0.999
0.60	0.232	0.386	0.643	0.497	0.999	0.999
0.75	0.119	0.158	0.211	0.194	0.330	0.528
0.90	0.069	0.076	0.085	0.084	0.104	0.129
1.00	0.050	0.050	0.053	0.050	0.053	0.050
1.10	0.286	0.351	0.410	0.298	0.374	0.445
1.25	0.514	0.611	0.689	0.561	0.683	0.779
1.50	0.719	0.812	0.875	0.795	0.899	0.955
2.00	0.881	0.941	0.970	0.947	0.988	0.998
3.00	0.965	0.988	0.996	0.993	0.999	0.999
4.00	0.985	0.997	0.999	0.999	0.999	0.999
5.00	0.992	0.999	0.999	0.999	0.999	0.999
10.0	0.999	0.999	0.999	0.999	0.999	0.999
10.0	0.963	0.988	0.996	0.997	0.999	0.999

REFERENCES

1. Cobby, J.M., Ridout, M.S., Bassett, P.J. and Large, R.V. (1985). An investigation to the use of ranked set sampling on grass and grass-cover swards. *Grass and Forage Science* 40, 257-263.
2. Johnson, G.D., Patil, G.P. and Sinha, A.K. (1993). Ranked set sampling for vegetation research. *Abstracta Botanica* 17, 87-102.
3. McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Austr. J. Agri. Res.* 3, 385-390.
4. Muttlak, H.A. and McDonald, L.L. (1990). Ranked set sampling with respect to concomitant variables and with size based probability of selection. *Commun. Statist. Theory Math.,* 19(1), 205-219.
5. Muttlak, H.A. and McDonald, L.L. (1990). Ranked set sampling with size based probability of selection. *Biometrics* 46, 435-445.
6. Muttlak, H.A. and McDonald, L.L. (1992). Ranked set sampling and line intercept method: A more efficient procedure. *Biom. Journal* 46, 435-445.
7. Patil, G.P. and Taillie, C. (1993). Environmental sampling, observational economy and statistical inference with emphasis on ranked set sampling,

encounter sampling and composite sampling. *Bulletin ISI Proceedings of 49 Session*, Firenze, 295-312.

8. Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics* 36, 35-42.
9. Stokes, S.L. (1986). Ranked set sampling. *Encyclopedia of Statistical Sciences* 1986 ed.
10. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.*, 20, 1-31.

CHAPTER THREE

A Note on Bayesian Estimation using Ranked Set Sample

M. Fraiwan Al-Saleh[1] and **Hassen A. Muttlak**[2]
[1] Department of Statistics, Yarmouk University, Irbid, Jordan.
[2] Faculty of Science and Technology, Deakin University
Geelong Campus, Geelong, Vectoria 3217, Australia.

ABSTRACT

Ranked set sampling (RSS) as suggested by McIntyre (1952) and Takahasi and Wakimoto (1968) mar be used in Bayesian estimation to reduce the Bayes risk. Bayesian estimation based on a ranked set sample (RSS) is investigated for exponential and normal distributions. We examine the Bayes risk of the Bayes estimator using RSS. It appears that as expected, the Bayes risk of the Bayes estimator using RSS is smaller than that of the corresponding Bayes estimator using simple random sample (SRS) of the same sample size.

KEYW ORDS

Order Statistics; Ranked Set Sampling; Bayes Estimators, Bayes Risk

1. INTRDOUCTION

Ranked set sampling (RSS) was first suggested by McIntyre (1952) who noted that RSS is highly beneficial and is much superior to the standard simple random sampling (SRS) for the estimation of the population mean. In many studies, it is possible to rank the sampling units rather cheaply without actually measuring them. See Halls and Dell (1966) and Muttlak and McDonald (1992) for some examples.

McIntyre (1952) gave no mathematical theory to support his suggestion. Takahasi and Wakimoto (1968) supplied the necessary mathematical theory. They proved that the sample mean of the RSS is an unbiased estimator of the population mean with smaller variance than the sample mean of a simple random sample (SRS) with the same sample size.

McIntyre (1952) and Takahasi and Wakimoto (1968) assumed perfect ranking of the elements. Dell and Clutter (1972) studied the case in which the ranking may not be perfect. Stokes (1980) proved that the estimator of the variance based on RSS data is an asymptotically unbiased estimator of the population variance and for large sample size, it is more efficient than the usual estimator based on SRS data with the same sample size. Muttlak and McDonald (1990a and 1990b) developed the RSS theory when the experimental units are selected with size-biased probability of selection. Sinha et al. (1992) considered estimating the mean and variance of the normal distribution and the mean of the exponential distribution. Lam et at. (1993) studied the parameters estimation of a two-parameter exponential distribution using RSS. Stokes (1995) considered estimation of and a for a family of random variables with cdfs of the form $F\left(\frac{x-\mu}{\sigma}\right)$.

In this this paper, Bayes estimation of the normal and exponential means using RSS is compared to that using SRS and it is shown that the former has smaller Bayes risk than the latter.

2. BAYESIAN ESTIMATION

Let $X_{11},X_{12},...,X_{1n};X_{21},X_{22},...,X_{2n},...,X_{2n},...,X_{n1}X_{n2},...,X_{nn}$ be n sets of n independent random variable all having the cdf $F(x\mid\theta)$. Let $X_{i(1)},X_{i(2)},...,X_{i(n)}$ denote the order statistics of $X_{i1},X_{i2},...,X_{in}\ (i=1,2,...,n)$. To simplify the notation we will use Y_i to denote the i^{th} order statistics of $X_{i1},X_{i2},...,X_{in}\ (i=1,2,...,n)$. Then $Y_1,Y_2,...,Y_n$, known as to ranked set sample, are independent random variables with densities

$$g_i(y_i\mid\theta)=n\binom{n-1}{i-1}\left[F(y_i\mid\theta)\right]^{i-1}\left[1-F(y_i\mid\theta)\right]^{n-i}f(y_i\mid\theta),i=1,2,...,n \quad (1)$$

If θ has a prior density $\pi(\theta)$, then the posterior distribution of θ given $Y_1=y_1,Y_2=y_2,....,Y_n=y_n$ is given by

$$\pi(\theta\mid y_1,y_2,...,y_n)=\frac{\prod_{i=1}^{n} g_i(y_i\mid\theta)\pi(\theta)}{\int\left[\prod_{i=1}^{n} g_i(y_i\mid\theta)\right]\pi(\theta)d\theta}$$

Thus

$$\pi(\theta\mid y_1,y_2,...,y_n)=\frac{\left\{\prod_{i=1}^{n}\left[F(y_i\mid\theta)\right]^{i-1}\left[1-F(y_i\mid\theta)\right]^{n-i}\right\}\left\{\prod_{i=1}^{n}f(y_i\mid\theta)\right\}\pi(\theta)}{\int\left\{\prod_{i=1}^{n}\left[F(y_i\mid\theta)\right]^{i-1}\left[1-F(y_i\mid\theta)\right]^{n-i}\right\}\left\{\prod_{i=1}^{n}f(y_i\mid\theta)\right\}\pi(\theta)d\theta}$$

(2)

Assume that $X_1, X_2, \ldots, X_n$ are independent random variables all having the same cdf $F(x \mid \theta)$. Then the posterior distribution of θ given $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ can be written as

$$\pi^*(\theta \mid x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} f(x_i \mid \theta)\pi(\theta)}{\int \left[\prod_{i=1}^{n} f(x_i \mid \theta)\right]\pi(\theta)\, d\theta} \tag{3}$$

Let δ be an estimator of θ and suppose that the loss function is $L(\theta, \delta)$. Then the risk function of δ is

$$R(\theta, \delta) = E_f\left(L(\theta, \delta)\right) \tag{4}$$

where the expectation is taken with respect of f. The Bayes risk is defined as $E_\pi\left(R(\theta, \delta)\right)$ where the expectation is taken with respect of π. Denote the bayes risk $r(\theta, \pi)$. If the Bayed risk is finite for some δ, then the Bayed estmator of θ is the estimator(s) that minimize the Bayes risk $r(\delta, \pi)$. See Berger (1985).

3. EXAMPLES

3.1 Exponential Distribution

Let $X_1, X_2, \ldots, X_n$ be iid with pdf $f(x \mid \theta)\frac{1}{\theta}e^{-x/\theta}; x > 0$. Let θ has the prior pdf $\pi(\theta)1\frac{1}{\Gamma(r)\theta^{r+1}}e^{-1/\theta}; \theta > 0$, i.e. θ has an inverse gamma distribution with $a = r$ and $\beta = 1$. The density of this distribution will be denoted by $IG(\alpha, \beta)$. Then the posterior distribution of θ given the SRS $x_1, x_2, \ldots, x_n$ is $IG\left(n + r, \frac{1}{\sum x_i + 1}\right)$.

The SRS Bayes estimator with respect to squared error loss is

$$\hat{\theta}_{SRS} = E(\theta \mid x_1, x_2, \ldots, x_n) = \frac{\sum x_i + 1}{n + r - 1}; n + r - 1 > 0.$$

The risk of $\hat{\theta}_{SRS}$ is

$$R\left(\hat{\theta}_{SRS},\theta\right)=Var\left(\hat{\theta}_{SRS}\right)+bias^2\left(\hat{\theta}_{SRS}\right)=\frac{\theta^2\left(n+1-1\right)^2-2\theta\left(r-1\right)+1}{\left(n+r-1\right)}$$

The Bayes risk of $\hat{\theta}_{SRS}$ can be shown to be

$$r\left(\hat{\theta}_{SRS},\theta\right)=\frac{1}{\left(n+r-1\right)\left(r-1\right)\left(r-2\right)};r>2 \tag{5}$$

Let $Y_1, Y_2, ..., Y_n$ be the RSS. The posterior distribution of θ given $y_1, y_2, ..., y_n$ is

$$\pi\left(\theta \mid y_1, y_2, ..., y_n\right)\propto\frac{1}{\theta^{n+r+1}}\prod_{i=1}^{n}\left\{\left(1-e^{-y_i/\theta}\right)^{i-1}e^{-\frac{1}{\theta}\left[1+\sum_{i=1}^{n}\left(n+1-i\right)y_i\right]}\right\} \tag{6}$$

For $n=2$,

$$\pi\left(\theta \mid y_1, y_2\right)\propto\frac{1}{\theta^{r+3}}\left[e^{-\frac{1}{\theta}\left(1+2_{y_1+y_2}\right)}-e^{-\frac{1}{\theta}\left(1+2_{y_1+y_2}\right)}\right]$$

Thus

$$\pi\left(\theta \mid y_1, y_2\right)=\frac{C_1}{C_1-C_2}IG\left(r+2,\frac{1}{1+2_{y_1+y_2}}\right)-\frac{C_2}{C_1-C_2}IG\left(r+2,\frac{1}{1+2_{y_1+y_2}}\right)$$

where

$$C_1=\frac{\Gamma\left(r+2\right)}{\left(1+2_{y_1+y_2}\right)^{r+2}} \text{ and } C_2=\frac{\Gamma\left(r+2\right)}{\left(1+2_{y_1+y_2}\right)^{r+2}}$$

The Bayed estimator of θ with respect to squared error loss based on RSS is

$$\hat{\theta}_{SRS}=\frac{C_1}{C_1-C_2}\frac{2_{y_1+y_2}+1}{r+1}-\frac{C_2}{C_1-C_2}\frac{2_{y_1+y_2}+1}{r+1}$$

which can be simplified to

$$\hat{\theta}_{SRS} = \frac{1}{r+1}\frac{\left(2_{y_1+y_2}+1\right)^{-(r+1)}-\left(2_{y_1+y_2}+1\right)^{-(r+1)}}{\left(2_{y_1+y_2}+1\right)^{-(r+2)}-\left(2_{y_1+y_2}+1\right)^{-(r+2)}}$$

For $n=3$ and using (6)

$$\pi\left(\theta \mid y_1,y_2,y_3\right) \propto \frac{1}{\theta^{r+4}}\left[e^{-\frac{1}{\theta}(1+3y_1+2y_2+y_3)} - 2e^{-\frac{1}{\theta}(1+3y_1+2y_2+2y_3)} + 2e^{-\frac{1}{\theta}(1+3y_1+3y_2+2y_3)} \right.$$
$$\left. -e^{-\frac{1}{\theta}(1+3y_1+2y_2+3y_3)} - e^{-\frac{1}{\theta}(1+3y_1+3y_2+y_3)} - e^{-\frac{1}{\theta}(1+3y_1+3y_2+3y_3)}\right]$$

Thus

$$\pi\left(\theta \mid y_1,y_2,y_3\right) = \left[C_1 IG\left(r+3,\frac{1}{1+3y_1+2y_2+2y_3}\right)\right.$$
$$-2C_2 IG\left(r+3,\frac{1}{1+3y_1+2y_2+2y_3}\right)+2C_3 IG\left(r+3,\frac{1}{1+3y_1+3y_2+2y_3}\right)$$
$$+C_4 IG\left(r+3,\frac{1}{1+3y_1+2y_2+3y_3}\right)-C_5 IG\left(r+3,\frac{1}{1+3y_1+3y_2+y_3}\right)$$
$$\left.-C_6 IG\left(r+3,\frac{1}{1+3y_1+3y_2+3y_3}\right)\right] \Big/ \left[C_1-2C_2+2C_3+C_4-C_5-C_6\right]$$

where

$$C_1 = \frac{\Gamma(r+3)}{\left(1+3y_1+2y_2+y_3\right)^{r+3}}, C_2 = \frac{\Gamma(r+3)}{\left(1+3y_1+2y_2+2y_3\right)^{r+3}},$$

$$C_3 = \frac{\Gamma(r+3)}{\left(1+3y_1+3y_2+2y_3\right)^{r+3}}, C_4 = \frac{\Gamma(r+3)}{\left(1+3y_1+2y_2+3y_3\right)^{r+3}},$$

$$C_5 = \frac{\Gamma(r+3)}{\left(1+3y_1+3y_2+y_3\right)^{r+3}}, C_6 = \frac{\Gamma(r+3)}{\left(1+3y_1+3y_2+3y_3\right)^{r+3}}$$

Therefore, the Bayes estimator based on RSS is given by

$$\hat{\theta}_{RSS} = \left[C_1 \left(\frac{1+3y_1+2y_2+y_3}{r+2} \right) \right.$$
$$-2C_2 \left(\frac{1+3y_1+2y_2+2y_3}{r+2} \right) + 2C_3 \left(\frac{1+3y_1+3y_2+2y_3}{r+2} \right)$$
$$+C_4 \left(\frac{1+3y_1+2y_2+3y_3}{r+2} \right) - C_5 \left(\frac{1+3y_1+3y_2+y_3}{r+2} \right)$$
$$\left. -C_6 \left(\frac{1+3y_1+3y_2+3y_3}{r+2} \right) \right] \Big/ \left[C_1 - 2C_2 + 2C_3 + C_4 - C_5 - C_6 \right]$$

To compare the SRS with the RSS for this example we need to find the Bayes risk $r\left[\hat{\theta}_{RSS}, \theta\right]$ for the RSS which can be wrtten for $n = 2$ as

$$r\left(\hat{\theta}_{RSS}, \theta\right) = \int_0^\infty \int_0^\infty \int_0^\infty \left(\hat{\theta}_{RSS}, \theta\right)^2 g\left(y_1\right) g\left(y_2\right) \pi\left(\theta\right) dy_1 dy_2 d\theta$$

where

$$g\left(y_i\right) = n \binom{n-1}{i-1} \left[F\left(y_i\right)\right]^{n-i} f\left(y_i\right);\ f\left(y_i\right) = \frac{1}{\theta} e^{-y_i/\theta},\ F\left(y_i\right) = 1 - e^{-y_i/\theta}, i = 1,2$$

and $\pi\left(\theta\right) = \dfrac{1}{\theta^{r+1}\Gamma\left(r\right)} e^{-1/\theta}$.

Similarly, the Bayes risk can be written for larger n. To evaluate the above integral, we use the IMSL computer program. Table 1 shows the efficiency of RSS with respect to SRS for n = 2, 3, 4 and 5 with different values of r. Considering Table (1), it can be seen that the use of RSS reduces the Bayes risk by about 6 times when n = r = 4. Relative precision is also reported for estimating the population mean using $SS\left(\hat{\mu}^*_{RSS}\right)$ and $SRS\left(\bar{X}_{SRS}\right)$. The relative precision for the MLE estimators using RSS and SRS is also reported.

3.2 Normal Distribution

Let $X_1, X_2, ..., X_n$ be iid with pdf, $f\left(x \mid \theta\right) = \dfrac{1}{\sqrt{2\pi}} e^{-\left(x-\theta\right)^2/2};\ -\infty < x < \infty$.

Let θ has the prior pdf, $\pi\left(\theta\right) = \dfrac{1}{\sqrt{2\pi}} e^{-\theta^2/2};\ -\infty < \theta < \infty$, i.e. has a standard normal distribution.

Assume that the loss function is the squared error loss function then the Bayes estimator is

$$\hat{\theta} = \int_{-\infty}^{\infty} \theta\pi\left(\theta \mid \underline{x}\right) d\theta .$$

The following method will be used to find the Bayes risk with respect to SRS and RSS

1. Pick a random value of θ from $\pi(\theta)$ call it θ_1, where $\pi(\theta)$ is our prior pdf.
2. Select $z_1, z_2, ..., z_n$ from $f\left(z \mid \theta_1\right)$.
3. Calculate the integral $\int \theta\pi\left(\theta \mid \underline{z}\right) d\theta$ and call it $\hat{\theta}_1$.
4. Repeat steps 2 and 3, L times (L is relatively large) and find $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_L$.
5. Approximate the risk of $\hat{\theta}_1$ as $R\left(\hat{\theta}, \theta_1\right) = \frac{1}{L} \sum_{i=I}^{L} \left(\hat{\theta}_1 - \theta_1\right)^2$
6. Repeat steps 1 to 5 in times for different values of θ say $\theta_2, \theta_3, ..., \theta_m$.
7. The Bayes risk of θ can be 'approximate as $r\left(\hat{\theta}, \pi\right) = \frac{1}{m} \sum_{i=1}^{m} R\left(\hat{\theta}, \theta_i\right)$.

Table (2) shows the efficiency of RSS with respect to SRS for n = 2, 3, 4 and 5 and it is also compared to other methods of estimation. We can see that the RSS reduces the Bayes risk by about 3 times when n = 4.

REFERENCES

1. Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer-Verlag, New York.
2. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background, *Biometrics* 28, 545-555.
3. Halls, L.S. and Dell, T.R. (1966). Trial of ranked set sampling for forage yields, *Forest Science* 12, (No. 1), 22-226.
4. Lam, Kin, Sinha, Bimal K., and Zhong Wu (1993). *Estimation of parameters in a two parameters exponential distribution using ranked set sample*, preprint.
5. McIntyre, G.A. (1982). A method of unbiased selective sampling, using ranked scts, *Aust. J. Agri. Res.*, 3, 385-390.
6. Muttlak, H.A. and McDonald, L.L. (1990a), Ranked set sampling with size biased probability of selection, *Biometrics*, 46, 435-445.

7. Muttlak, H.A. and McDonald, L.L. (1990b). Ranked set sampling with respect to a concomitant variables and with size biased probability of selection, *Commun. Statist. Theory Meth*. 19(1), 205-219.
8. Muttlak, H.A. and McDonald, L.L. (1992). Ranked set sampling and line intercept method: A more efficient procedure, *Biom. Journal* 46, 435-445.
9. Sinha, Bimal K, Sinha Bikas K. and Purkayastha, S. (1992). *On some aspects of ranked set sampling for estimation of normal and exponential parameters*, preprint.
10. Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics* 36, 35-42.
11. Stokes, S.L. (1995). Parametric ranked set sampling. *Statistics and Decisions*, In Press.
12. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Ann. Inst. of Statist. Math.,* 20, 1-31.

Table 1

Bayes relative efficiency (BRE) of $\hat{\theta}_{RSS}$ with respect of $\hat{\theta}_{SRS}$ and relative precision (PR) of RSS to SRS using other type of estimations for the exponential distribution1,2

	n			
	2	3	4	5
$BRE\left(\hat{\theta}_{RSS},\hat{\theta}_{SRS}\right);r=3$	1.51	1.60	1.72	2.27
$BRE\left(\hat{\theta}_{RSS},\hat{\theta}_{SRS}\right);r=3.5$	2.52	2.97	3.76	4.33
$BRE\left(\hat{\theta}_{RSS},\hat{\theta}_{SRS}\right);r=4$	4.19	4.96	5397	6.81
$RP\left(\hat{\mu}^*_{RSS},\bar{X}_{SRS}\right)$	1.33	1.64	1.92	2.14
$\lim_{m\to\infty}\left(\hat{\theta}^*_{ML},\hat{\theta}_{ML}\right)$	1.40	1.81	2.21	2.62

1) Values of RP in the fourth line are from Dell andClutter (1972)
2) The asymptotic RP in the last line represent the limiting RP of MLE using RSS w.r.t the MLE using SRS as reported by Stokes (1995).

Table 2

Bayes relative efficiency (BRE) of $\hat{\theta}_{RSS}$ with respect of $\hat{\theta}_{SRS}$ and relative precision (PR) of RSS to SRS using other type of estimations for the normal distribution1,2

	n			
	2	3	4	5
$BRE\left(\hat{\theta}_{RSS},\hat{\theta}_{SRS}\right)$	1.73	2.35	2.81	3.29
$RP\left(\hat{\mu}^*_{RSS},\bar{X}_{SRS}\right)$	1.47	1.91	2.35	2.77
$\lim_{m\to\infty}\left(\hat{\theta}^*_{ML},\hat{\theta}_{ML}\right)$	1.48	1.96	2.44	2.92

1) Values of RP in the fourth line are from Dell andClutter (1972)
2) The asymptotic RP in the last line represent the limiting RP of MLE using RSS w.r.t the MLE using SRS as reported by Stokes (1995).

CHAPTER FOUR

Recent Developments in Ranked Set Sampling

Hassen A. Muttlak[1] and **M. Fraiwan Al-Saleh**[2]
[1] Department of Mathematical Sciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
[2] Department of Mathematics and Statistics, Sultan Qaboos University, Sultanate Oman

ABSTRACT

Considering situations where units are expensive to measure but can be ordered relatively cheaply or at no cost without actual measurements of the units under investigation. McIntyre (1952) pioneered the study of the method ranked set sampling (RSS). Considerable attention has been paid to this sampling method in the statistics literature for the last ten to fifteen years. In this paper we will review the recent developments in area of RSS, concentrating mainly on the last five to six years.

1. INTRODUCTION

McIntyre (1952) was the first to propose the method of ranked set sampling (RSS) to estimate population mean. Takahasi and Wakimoto (1968) independently described the same sampling method and presented the mathematical theory, which supports McIntyre's intuitive assertion. Dell and Clutter (1972) showed that errors in ranking reduce the efficiency of the RSS mean relative to the SRS mean. However, the RSS mean remains unbiased and more efficient than the SRS mean unless the ranking is so poor as to yield a random sample, in this case the RSS estimator performs just as well as the SRS mean.

The RSS method can be summarized as follows: From a population of interest, *n* random sets each of size *n* are selected. The members of each random set are ranked with respect to the variable of interest by a cost free method e.g. eyeballs. From the first ordered set, the smallest unit is selected for measurement. From the second ordered set the second smallest unit is selected for measurement. This continues until the largest element from the last ordered set is measured. This process may be repeated *r* times (i.e. *r* cycles or replications) to yield a sample of size *rn*. These *rn* units form the RSS data.

Let $X_1, X_2, \ldots, X_n$ be a random set with probability density function $f(x)$ with a finite mean μ and variance σ^2. Let $X_{11}, X_{12}, \ldots, X_{1n}$; $X_{21}, X_{22}, \ldots, X_{2n}$; ...;

X_{n1}, X_{n2}, ..., X_{nn} be independent random variables all with the same cumulative distribution function F(x). Let $X_{i(1)}$, $X_{i(2)}$, ..., $X_{i(n)}$ be the order statistics of X_{i1}, X_{i2}, ..., X_{in} (i = 1, 2, ..., n). Let $X_{(i:n)}$ denote the i^{th} order statistic from the i^{th} set of size n. If the cycle is repeated *r* times, let $X_{(i:n)j}$ denotes the i^{th} order statistic from the i^{th} set of size *n* in the j^{th} cycle. We will refer to this sampling method, which is due to McIntyre (1952), and Takahasi and Wakimoto (1968) as MTW RSS. The unbiased estimator of the population mean (see Takahasi and Wakimoto, 1968) is defined as

$$\overline{X}_{RSS} = \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} X_{(i:n)j} \ . \tag{1}$$

In this paper we review the recent developments in the area of ranked set sampling in the last five to six years. In Section 2, we review the recent developments in the nonparametric statistics methods using RSS data. In Section 3, we consider the latest developments in the area of parametric statistics including parameters estimation and testing hypotheses. The modifications of the MTW RSS methods are discussed in Section 4. The use of RSS in the regression estimation and Bayesian statistics are considered in Sections 5 and 6 respectively. The other works, which cannot fit under any of the previous sections, are discussed in the last section. For classified and extensively reviewed work in the area of RSS from 1952 to 1994 see Patil et al. (1994) and Kaur et al. (1995). Patil et al. (1999) presented a bibliographic list in most of the work published in the area of RSS up to the end of the twentieth century.

2. NONPARAMETRIC STATISTICS USING RANKED SET SAMPLING

Stokes and Sager (1988) were the first to consider a nonparametric setting using RSS data. They developed the properties of the empirical distribution function based on RSS and compared these properties to the usual empirical distribution function using simple random sample (SRS) data. Bohn and Wolfe (1992, 1994) developed the Mann-Whitney-Wilcoxon statistic using RSS for both perfect and imperfect ranking. Kvam and Samaniego (1993, 1994) developed the estimation of the population distribution function and population mean using unbalanced RSS data, i.e. the size of the i^{th} set need not to be the same for all sets and the various order statistics need not to be represented an equal number of time. Bohn (1994) and Hettmansperger (1995) considered the one ranked-set sample problem. They considered procedures called sign-rank and sign statistics respectively for RSS data. For a review for the early nonparametric work in the area of RSS see Bohn (1996).

1) Koti and Babu (1996) derived the exact distribution of the RSS sign test under the null hypothesis i.e. the exact null distribution. They compared

the power of the sign test based on RSS with the usual SRS sign test for the double exponential, Cauchy and contaminated normal distributions. They showed that the RSS sign test is superior to the SRS sign test. Finally they discussed the problem of imperfect ranking.

2) Huang (1997) considered the asymptotic properties of the nonparametric maximum likelihood estimator (NPMLE) of a distribution function using RSS. He showed that the NPMLE of a distribution function based on RSS is consistent and converges weakly to normal process. He also developed the covariance function of the limiting process. Finally he showed that the NPMLE of a distribution function based on RSS is asymptotically efficient compared to the usual NPMLE based on SRS.

3) The Neyman's optimal allocation requires the sample size corresponding to each rank to be proportional to its standard deviation, but in most applications the standard deviation is unknown. The performances of RSS methods are affected by the allocation of order statistics in the sample. Kaur, Patil, and Taillie (1997) considered the effects of unequal allocation for RSS with skew distributions on the estimation of the population mean. They considered two models of unequal allocation of skew distributions. The first model is the t-model where the largest order statistics is quantified t (≥ 1) times more than the rest of order statistics. The second model is the (s, t)-mode where the two largest order statistics are quantified more than the rest by factors of (s, t), 1≤ s≤ t, respectively. The Neyman's optimal allocation is performs better than the (s, t)-model, while the (s, t)-model perform better than the t-model. Finally the t-model performs better than the equal allocation model.

4) The RSS procedure that we described in Section 1 is called the balanced RSS procedure (i.e. in the ith set we observe $X_{(i:n)}$ the ith order statistics in the ith set of size n). But under the generalized version of RSS we observe $X_{(r_i;n_i)}$, so the data set is $(X_{(r_1:n_1)}, X_{(r_2:n_2)}, ..., X_{(r_k:n_k)})$. Kim and Arnold (1997) considered estimating the distribution function F under both balanced and unbalanced RSS. They start with a Dirichlet process as a prior for F. The estimate of distribution function F is updated (the posterior distribution function is again a Dirichlet process) based on a completed data. These two steps are repeated until the estimate of F is stabilized.

5) Barabesi (1998) developed a simple and fast method to calculate the exact distribution of the RSS sign test statistic based on the probability

generating function. He developed his method using the Mathematica package.

6) Chen (1999) was the first to use RSS data to estimate the density function using the kernel method of density estimation. He studied the properties of the RSS density estimation and found that the bias of the RSS density estimate is the same as the SRS estimate of the same size and the variance of the RSS estimate as a function of the set size decreases as the set size increases. The mean integrated square error of the pdf $f(x)$ is defined by $MISE(\hat{f}) = E\int[\hat{f}(x) - f(x)]^2$, where $\hat{f}(x)$ is the density estimation of $f(x)$. The MISE of the RSS estimate of $f(x)$ is found to be smaller than the SRS estimate whether or not there are errors in ranking. Finally some simulation studies were carried to find out how much MISE can be reduced by using RSS using the normal, gamma and extreme value distributions.

7) Aragon, Patil and Taillie (1999) reviewed the work of Stokes and Sager (1988) on the empirical distribution function using RSS and Bohn (1992) on the Mann-Whitney-Wilcoxon test based on RSS. They proposed a model for ranking error probability matrix, which can be use for evaluating RSS-based statistical methods.

8) Hartlaub and Wolfe (1999) generalized the one- and two-sample location problems considered in the previous nonparametric work in the area of RSS to m-sample location problem. They developed the RSS procedures for the m-sample location setting under the restriction that the treatment effect parameters follow a restricted umbrella pattern. They developed distribution-free test statistics for both cases where the peak of the umbrella is known and when it is unknown. They studied the properties of the null-distribution in the case of known peak of the umbrella. Finally they discussed finding the critical values for the test statistics for both cases of known and unknown umbrella peak.

9) The problem of estimating $\theta = p(X > c)$ using RSS is addressed by Li, Sinha and Chuiv (1999). They showed that the use of RSS instead of SRS improves the estimation of θ when the distribution is unspecified i.e. distribution free. They compared the performances of different estimators of θ using SRS and RSS methods and concluded that if the underlying distribution is normal, RSS estimators will outperform the SRS estimators.

10) Statistical inference using RSS depends on the location of measured observations. Özturk (1999a) suggested and used a selective design that determines the location of the measured observations in RSS. The sampling procedure may be summarized as follows: Select a set of k elements from a population with a cdf F(x), rank them with respect to the variable of interest and measure only $X_{(d_1;k)}$, where $X_{(d_1;k)}$ is the d1-st order statistic in the set. Return the remaining observations to the population. Then select another set of size k and measure $X_{(d_2;k)}$ and again return the remaining observations to the population. Repeat this process *r* times to get $X_{(d_1;k)i}, {}_{(d_2;k)i}, \ldots, {}_{(d_r;k)i}$, for i = 1, 2, …, n. This is called a selective RSS. The measured $X_{(d_1;k)i}, {}_{(d_2;k)i}, \ldots, {}_{(d_r;k)i}$ are a cycle in the selective RSS. If $r = k$ then a cycle in a selective RSS is the same as a cycle in usual RSS. The set $D = \{d_1,\ d_2,\ \ldots,\ d_r; k\}$ is called s design, where $d_1,\ d_2,\ \ldots,\ d_r$ determine the locations of the measured observation in a set of size k. He considered the one and two-sample sign test based on selective RSS and compared it to that based on MTW RSS. He showed that the selective RSS has higher Pitman efficiency than the MTW RSS of the same size when the design measured only the middle observations.

11) Özturk (1999b) developed a two-sample test using RSS. He showed that the test is a distribution free test and there is substantial increase in the efficiency of the test even with error in ranking. But the probability of type I error is inflated in the presence of errors in ranking. Finally he showed that the newly developed test is superior comparing to the two-sample Mann-Whitney-Wilcoxon RSS test if the underlying distribution has a heavy and long tail distribution and the number of observations in each cycle is small.

12) Presnell and Bohn (1999) developed the U-statistics using RSS data for one and two sample cases. They showed that their statistics are asymptotically efficient compared to their counterpart the SRS U-statistics, whether or not there are errors in ranking. They pointed out some errors in ranked set sampling literatures. Finally they came up with counterexamples to show that it is not necessary that perfect ranking will lead to more efficient estimators than imperfect ranking.

13) Chen (2000) investigated the properties of the sample quantiles of RSS. He showed that the RSS quantiles are strongly consistent estimator for any set size and obtained the asymptotic normality for a large sample size.

He also obtained the Bahadur representation of the RSS quantiles. The confidence interval and testing of hypotheses for the population quantiles were developed using RSS data. The newly developed RSS quantiles were compared with the usual SRS quantiles via their relative efficiency. The gain in efficiency by using RSS was shown to be very large, the largest being when the inference is on the median. Finally it was noticed that the gain in efficiency is decreases if we move away from the median to the extreme quantiles.

14) Özturk and Wolfe (2000) investigated the effect of different RSS protocols on sign test statistic. These RSS protocols are sequential, mid-range and fixed RSS designs. He showed that the sequential and mid-range RSS are optimal if only one observation from each set is measured. The fixed RSS design is optimal if only the middle-order statistic is measured. The sequential RSS protocol can be summarized as follows: Randomly selects cn units from an infinite population and partitions these units into c set each of size n units. In each set, rank the units with respect to a variable of interest. From the first set measure the observations $X_{(1)},\ ...,\ X_{(t)}$, where t = n/c is an integer. From the second set, the observations $X_{(t+1)},\ ...,\ X_{(2t)}$ are measured. This process continues until the observations $X_{(n-t+1)},\ ...,\ X_{(n)}$ are measured from the last set. The cycle may be repeated r times. For the mid-range RSS protocol, the observations at equal distance from the middle rank are measured in each set. Finally in fixed RSS protocol the same order statistics in each set are measured. For an integer $t \in \{1,\ ...,\ n\}$, let $D_t = \{d_1,\ ...,\ d_t\}$ be the set of order statistics to be measured from each set of size n units.

3. PARAMETRIC STATISTICS BASED ON RANKED SET SAMPLING

3.1 Parameter Estimation using RSS

Even though RSS method is nonparametric in nature several authors considered using RSS data to estimate the parameters of well-known distributions. Lam et al. (1994) considered the estimation of the parameters of two-parameter exponential distribution. Ni Chuiv et al. (1994) studied the estimation of the location parameter of the Cauchy Distribution. Fei et al. (1994) explored the estimation of the parameters for the two-parameter weibull and extreme-value distributions using RSS. Sinha et al. (1994) considered the estimation of a gamma mean based on RSS Lam et al. (1995) considered the estimation of the location and scale parameter of the logistic distribution using RSS. See Ni Chuiv and Sinha (1998) for a review of some development in the parametric estimation using RSS.

1) Stokes (1995) studied the maximum likelihood estimators under RSS of the parameters of the location-scale family having cumulative distribution function (cdf) of the form $F\left((x-\mu)/\sigma\right)$ with F known. Stokes (1995) considered several examples and demonstrated the dominance of the mle's under RSS over other estimators. The best linear unbiased estimators (BLUEs) of the location and scale parameters were proposed in the same study and shown to do nearly as well as their maximum likelihood counterparts in most cases.

2) Bhoj and Ahsanullah (1996) considered estimation of the parameters of the generalized geometric distribution using RSS. They obtained the minimum variance linear unbiased estimators (MVUE) of the parameters μ and σ for the generalized geometric distribution of the form $f(x) = p\sigma^{-1}b^{-p}\left((x-\mu)/\sigma + a\right)^{p-1}$, $\mu - a\sigma \le x \le \mu + (b-a)\sigma$ based on RSS. The MVUE of μ and σ based on RSS were shown to be more efficient than the MVUE based on n ordered statistics given by Downton (1954). Further the MVUE of the population mean μ based on RSS is more efficient than the usual RSS estimator of the population mean.

3) Sinha, B., Sinha, A. and Purkayastha (1996) proposed some best linear unbiased estimators (BLUEs) of the parameters of the normal and exponential distributions under RSS and some modifications of it. They first addressed the issue of how best to use the RSS, namely $X_{(i;n)}$, $i = 1, \ldots, n$ to estimate μ and came up with a BLUE of μ which is given by $\hat{\mu}_{blue} = \left[\sum_{i=1}^{n} X_{(i;n)}/\nu_i\right] \Big/ \sum_{i=1}^{n} 1/\nu_i$, where ν_i is the variance of the ith order statistics in a sample of size n from a standard normal population. They derived the BLUE of μ based on a partial RSS, namely $X_{(i;n)}$, $i = 1, \ldots, l$, where $l < n$. Also, they considered selecting the median of the ith sample in estimating μ. To estimate the variance of the normal distribution they proposed several estimators using RSS and some of its modifications. Most of the newly proposed estimators are more efficient than SRS estimators. But $\hat{\mu}_{blue}$ and the estimator based on the median of the ith sample are more efficient than the mean based on the MTW RSS. They proposed several estimators for the exponential parameter θ, the first being BLUE for θ based on the RSS, which is more

efficient than MTW RSS estimator for θ. The other estimators are based on partial RSS.

4) Barnett and Moore (1997) extended the work of Stokes (1995) and Sinha et al. (1996) to come up with an optimal ranked-set of the location and scale parameters when the nuisance parameter is unknown and the distribution need not to be symmetric. They showed that their estimators are more efficient than the estimators suggested by Stokes (1995) and Sinha et al. (1996) in the case of perfect ranking for the normal and exponential distributions. Unlike Stokes (1995) and Sinha et al. (1996) they considered the case of imperfect ranking and came up with BLUEs for the parameters.

5) Bhoj (1997a) obtained the minimum variance unbiased estimator (MVUE) of the location and scale parameters of the extreme value distribution using RSS. He compared these estimators with the ordered least squared estimates given by Lieblein and Zelen (1956). The MVUE turned out to dominate their order least squared counterparts. He also introduced an unbiased estimator for the population mean and showed it be more efficient than the MTW RSS. Finally he used a modification of RSS to come up with a more efficient estimator for the scale parameter in case of small sample size.

6) Li and Ni (1997) discussed the efficiency of using RSS to estimate the parameters of the normal, exponential two-parameter exponential and Cauchy distributions compared to the usual estimators using SRS. They established that using the right modifications of RSS would often result in a more efficient estimator for some parameters of those distributions.

7) As we saw many authors used the BLUE based on RSS to estimate the population mean and other parameters of interest instead of the MTW RSS estimator. However, the underlying distribution should be known to use the BLUE method. Tam, Yu, and Fung (1998) investigated the sensitivity of the BLUE to the misspecification of the underling distribution. They considered several distributions and compared their performance under both the usual RSS and the BLUE based on RSS. It truned out that in general the sensitivity of the BLUE depend on the kurtosis of the underling distribution.

8) The exponential distribution is very widely used in modeling real life problem. Bhoj (1999a) discussed estimating the exponential distribution parameter using three RSS methods. These methods are MTW RSS, NRSS suggested by Bhoj (1997c) and MRSS introduced by Bhoj (1999b). He compared these estimators among themselves and with the estimator

based on the ordered least squares. He found that the estimator based on NRSS is the most efficient one.

9) Bhoj (1999a) considers the estimation of the scale parameter of the Rayleigh distribution. He used the MTW RSS, NRSS suggested by Bhoj (1997b) and other modification of RSS method to estimate the parameter of interest. The newly suggested estimators were compared to the least squares estimator based on order observations. And found to be more efficient.

10) Chen (2000) generalized Stokes's (1995) results for multi-parameter families. He proved that the Fisher information matrix under RSS is equal to the Fisher information matrix under SRS minus an additional positive definite matrix. This showed that the maximum likelihood estimators under RSS are always more efficient than their counterparts based on SRS. He studied the effect of the errors in ranking and considered different models. Finally the effect of the underlying distribution is considered.

Many aspects of RSS have been considered in the literature. Most of the work has been devoted to estimation of the population mean μ, very little has been done in the estimation of the population variance σ^2. Stokes (1980) treated this problem and suggested an estimator for σ^2 which asymptotically unbiased and more efficient than the usual SRS unbiased estimator for σ^2.

Yu, Lam, and Sinha (1999) addressed the issue of variance estimation using RSS for the normal distribution. They suggested several unbiased estimators for σ^2 based on the balanced and unbalanced RSS data. They proposed several estimators for σ^2 based on single cycle and compared these estimators to Stokes's (1980) estimator. The new estimators truned out to be more efficient. In the case of more than one cycle they combined information from different cycles in different ways and proposed several unbiased estimators for σ^2. Finally they considered the problem of unequal replications (unbalanced RSS) and came up with suitable unbiased estimator for σ^2.

3.2 Testing Hypotheses using RSS

We now consider the work that has been done in the area of hypothesis testing using RSS.

1) In Abu-Dayyeh and Muttlak (1996) some hypothesis testing for the exponential and uniform distributions are considered. For the exponential distribution with pdf $f_\theta(x) = \theta^{-1} \; e^{-x/\theta}$, they considered testing the

hypothesis $H_o:\ \theta=\theta_o$ vs. $H_a:\theta>\theta_o$. They proposed a test based on RSS and compared it with the uniformly most powerful test (UMPT) of size α based on SRS. It truned out that their suggested test is more powerful than the UMPT. Also, they derived the likelihood ratio test (LRT) for the hypothesis $H_o:\ \theta=1$ vs. $H_a:\theta\neq 1$ based on RSS data and found it to be more powerful than SRS test. For the uniform distribution with the pdf $f_\theta(x)=\theta^{-1},\ 0\leq x\leq\theta$, they considered the following hypothesis is $H_o:\ \theta=\theta_o$ vs. $H_a:\theta\neq\theta_o$. In comparison with the UMPT based on SRS, they found their proposed test to be more powerful.

2) Muttlak and Abu-Dayyeh (1998) considered testing some hypotheses about the population mean μ and variance σ^2 of the normal distribution. For the population mean, they considered the case of σ^2 is known and unknown, in testing the hypothesis $H_o:\mu=\mu_o$ vs. $H_o:\mu\neq\mu_o$. They proposed test statistics based on RSS for both cases. They showed their proposed test statistics are more powerful than the SRS UMPT. A similar conclusion was reached when they compared their proposed test of the hypothesis $H_o:\sigma^2=\sigma_o^2$ vs. $H_a:\sigma^2>\sigma_o^2$ with the SRS UMPT he test based on RSS UMPT of the same hypothesis.

3) Shen and Yuan (1996) proposed a test for the normal mean based on modified partial ranked set sample when the variance is known. They showed that this test has more power as compared to the traditional normal test. This test is similar to the test derived by Shen (1994). Both tests are found to be more powerful than the usual normal test using SRS.

4. MODIFIED RANKED SET SAMPLING METHODS

Several modifications of the RSS method have been proposed. These modifications were found to be necessary because RSS is often difficult to apply in the field. It is also subject t errors in ranking, which reduce its efficiency. Thus the proposed modifications are aimed at remedying these problems. And some times only to increase the efficiency of the estimators of the parameters under consideration.

1) Muttlak (1996c) suggested a modification for RSS called pair RSS, which may be summarized as follows: In the case of even set size select k = n/2 sets of size *n* units and rank the units within each set with respect to a variable of interest. From the first set select the smallest and the largest for measurement. From the second set select the second smallest and the second largest for measurement. Continue until the units with the k^{th}

smallest and $(k+1)^{th}$ largest from the k^{th} set are chosen for measurement. For the case of odd set size select L = (n+1)/2 random sets of size *n* units. From L-1 sets and repeat the above procedure. From one set select for measurement the median of the set. He used the proposed pair RSS to estimate the population mean, showed it to be unbiased, and has smaller variance the usual SRS mean.

2) Samawi, Ahmed, and Abu-Dayyeh (1996) studied the properties of the extreme ranked set sampling (ERSS) in estimating the population mean. The ERSS procedure can be summarized as follows: Select *n* random sets of size *n* units from the population and rank the units within each set with respect to a variable of interest by visual inspection. If the set size *n* is even, select from n/2 sets the smallest unit and from the other n/2 sets the largest unit for actual measurement. If the set size is odd, select from (n-1)/2 sets the smallest unit, from the other (n-1)/2, the largest unit, and from one set the median of the sample for actual measurement. The cycle may be repeated *r* times to get nr units. These nr units form the ERSS data. Let $X_{i\,(1)}$ and $X_{i\,(n)}$ be the smallest and the largest of the i^{th} set respectively (i = 1, 2,..., n). If the cycle is repeated *r* times, the estimator of the population mean using ERSS is defined in the case of an even set size as

$$\overline{X}_{erss1} = \frac{1}{nr}\sum_{j=1}^{r}\left(\sum_{i=1}^{L} X_{i(1)j} + \sum_{i=L+1}^{n} X_{i(n)j}\right),$$

where L = n/2. In the case of an odd set size, the estimator of the population mean is defined as

$$\overline{X}_{erss2} = \frac{1}{nr}\sum_{j=1}^{r}\left(\sum_{i=1}^{L_1} X_{i(1)j} + \sum_{i=L_1+2}^{n} X_{i(n)j} + X_{i((n+1)/2)j}\right),$$

where L_1= (n-1)/2 and $X_{i\,((n+1)/2)}$ is the median of set i = (n+1)/2. They showed that the ERSS estimators are more efficient than usual SRS mean and unbiased if the underlying distribution is symmetric. Also, it is more efficient than RSS estimator for some probability distribution functions, e. g. uniform distribution.

3) Bhoj (1997b) proposed a modification to the MTW RSS called it new ranked set sampling (NRSS). The method is simply select *n* sets of size 2m units where n=2m. The 2m units of each set are ranked among themselves by visual inspection or any cost free method. Select from each ordered set two units for actual measurement. The choice of the two units from each set depended on the underlying distribution and the parameter(s) to be estimated. He used this method estimate the location

and scale parameters of the rectangular and logistic distributions. The new method improved the efficiency of the estimators considered with respect to the MTW RSS method.

4) Bhoj (1997c) proposed another modification to the MTW RSS, called modified ranked set sampling (MRSS). In this sampling method we select only two order statistics instead of *n* as we do in the MTW RSS. We select n/2 j^{th} order statistics and n/2 k^{th} order statistics. The choice of the j^{th} and k^{th} order statistics depend on the underlying distribution and the parameter(s) to be estimated. He used this method estimate the parameter of the uniform and logistic distributions. Again using the MRSS method improved the efficiency of the estimators considered in this study.

5) Muttlak (1997) suggested using the median ranked set sampling (MRSS) to estimate the population mean. The MRSS method can be summarized as follows: Select *n* random sets of size *n* units from the population under study and rank the units within each set with respect to a variable of interest. If the set size *n* is odd, from each sample select for measurement the $((n + 1)/2)^{th}$ smallest rank (the median of the sample). If the set size is even, select for measurement from each of the first n/2 sets, the $(n/2)^{th}$ smallest unit and from the second n/2 sets the $((n +2)/2)^{th}$ smallest unit. The cycle may be repeated *r* times to get nr units. These nr units form the MRSS data. If the set size is odd, let $X_{i((n+1)/2)}$ be the median of the i^{th} sample (i = 1, 2, ..., n) i.e. the $((n+1)/2)^{th}$ order statistic of the i^{th} set. If the set size is even let $X_{i(n/2)}$ be the $(n/2)^{th}$ order statistic of the i^{th} set (i = 1, 2, ..., L= n/2) and $X_{i((n+2)/2)}$ be the $((n+2)/2)^{th}$ order statistic of the i^{th} set (i = L+1, L+2, ..., n). The estimator of the population mean based on MRSS is defined in the case of an odd set size by

$$\overline{X}_{\text{mrss1}} = \frac{1}{nr} \sum_{j=1}^{r} \sum_{i=1}^{n} X_{i((n+1)/2)j} \ ,$$

and in the case of an even set size, it is defined as

$$\overline{X}_{\text{mrss2}} = \frac{1}{nr} \sum_{j=1}^{r} \left(\sum_{i=1}^{L} X_{i(n/2)j} + \sum_{i=L+1}^{n} X_{i((n+2)/2)j} \right)$$

where L = n/2. He showed MRSS estimators are unbiased for the population mean and more efficient than the RSS estimator is if the underlying distribution is symmetric.

6) The MTW RSS is an equal allocation scheme in the sense that all of the *n* order statistics are replicated equal number of times, namely, *r* number of

times. But it is quite possible that the ith order statistic is replicated ri number of times, i = 1, 2,…, n. Yu, Lam and Sinha (1997) considered estimation of the population mean under the above sampling procedure. They proposed an estimator for the population mean as:

$$\hat{\mu}_{RSS} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} X_{(i:n)j} / r_i \text{ with its variance } \operatorname{var}(\hat{\mu}_{RSS}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2_{(i:n)} / r_i \ .$$

It is not difficult to see that $\operatorname{var}(\hat{\mu}_{RSS}) < \operatorname{var}(\bar{X}_{SRS})$ for the possible location of ri.

7) Muttlak (1999a and 1999b) considered two phase sampling, in the first phase, units are selected with probability proportional to their size. In the second phase unites are selected using MRSS or ERSS to increase the efficiency of the estimators relative to SRS. In both papers he considered estimating the population mean and the population size using MRSS and ERSS.

8) Al-Saleh and Al-kadiri (1999) investigated a double ranked set sampling (DRSS) approach to RSS. This approach simply consists of applying the RSS technique to the resulting RSS samples. For example if the set size n = 2, then to obtain a DRSS of size 2, 8 randomly selected elements are divided into 4 sets of size 2 each. Applying the RSS procedure to this gives obtain 2 RSS of size 2 each. Applying the procedure again to the two sets gives one set of size 2, which is called Double RSS sample. It turns out that all identities that are valid for RSS continue to be valid in the case of DRSS. The efficiency of the method in estimating the population mean is higher than that of RSS. For example, for the uniform distribution with n =3, the efficiency of DRSS is 3.03 while it is 2.0 for RSS; for n = 4 it is 4.71 while it is 2.5 for RSS. The difficulty in ranking in the second stage is also investigated. It was shown using the concept of "Level of Distinguishability" that the ranking in the second stage is much easier than in the second stage; thus no extra effort is needed. That makes the method of practical use.

9) Al-Saleh and Al-Omari (1999) considered the multistage version of RSS and introduced the concept of steady state RSS and steady state efficiency, which is simply the limit of the efficiency as the number of stages gets large. The usefulness of this concept in Monte Carlo simulation was explored. The new technique was illustrated here by considering the yield of olive trees. The data, which was collected by the second Author, represents a useful application of RSS since ranking of the variable of interest is much easier than the actual measurement of this

variable. The chosen ranker judged visually, the smallest and the largest olive yields of two contiguous randomly selected trees in a field of trees in West Jordan. The process was repeated by the same ranker until he got multi stage ranked set sample (MSRSS) with 3 stages (i.e. s = 3) of size $n = 2$. The whole cycle was repeated $r = 4$ times, to give 4 MSRSS of size 2 each. The exact olive yield of each of the 64 trees used for the study was then exactly quantified. Hence it was possible to obtain the accuracy of the judgment ranking. This method turned out to be very useful in this application.

10) Hossain and Muttlak (1999) proposed another modification for RSS called paired ranked set sampling (PRSS). In the PRSS procedure, select two sets each of size n units and rank the units within each set by visual inspection. Select for actual measurements the kth smallest units from the first set, where $1 \le k \le n$ is pre-determined. From the second set select the (n-k+1)th smallest unit for actual measurement. The cycle may be repeated r times to obtain a sample of size 2n units. They used PRSS procedure to come up with MVUE for the population mean and standard deviation. They compared the PRSS estimators to the SRS and the usual RSS estimators for population mean and variance using various distributions namely the rectangular, normal, logistic, double exponential and exponential distributions. They showed that the PRSS estimators are more efficient than the SRS estimators and most of the RSS estimators for the distributions considered.

11) In the MTW RSS procedure we select n sets of size n units each, order the units within each set by visual inspection, and select the ith smallest unit from the ith set for actual measurement. Li, Sinha and Perron (1999) proposed to select $r < n$ order statistics, independently from r sets with distinct indices $k_1, \ldots, k_r$ denoted by $\tilde{k}$ where $\tilde{k}$ is a subset of (1, …, n). For example for a given r first select r indices ($k_1, \ldots, k_r$) from (1, …, n) at random without replacement, and then select $X_{(1k_1)}$ from random set of size n (as if from the first set) and select $X_{(2k_2)}$ form another random set of size n (as if from the second set) and so on. Select $X_{(rk_r)}$ from the rth random set (as if from the rth set). They study the properties of the estimators of μ and σ2 based on the new random selection sample ($X_{(1k_1)}, \ldots, X_{(rk_r)}$). They study both the parametric and nonparametric properties of the newly developed estimators for μ and σ2 for the normal, exponential and logistic distributions.

5. REGRESSION ESTIMATION USING RAKED SET SAMPLING

RSS utilizes cost free auxiliary information to rank randomly selected units with respect to a variable of interest before measuring a subset of these units. These measured units are chosen, on the basis of the ranking information, this makes the RSS estimator of the population mean more efficient than the SRS estimator with the same sample size. Patil et al. (1993) were the first to compare RSS estimator to the SRS regression estimator of the population mean and showed that it is more efficient unless the correlation coefficient between the variable of interest and the auxiliary variable is high i.e. $|\rho| > 0.85$. They assumed that variable of interest and the auxiliary variable jointly following the bivariate normal distribution.

1) Muttlak (1995) presents a RSS method to simple linear regression line. In that method, a sample of n pairs of observations $(x_1, y_{1(1)}), (x_2, y_{2(2)})$, $(x_n, y_{n(n)})$ are obtained where the $y_{i(i)}$ is the ith smallest value measured of the dependent variable in a potential sample size n and x_i the corresponding observed values of the independent variable, i = 1, …, n. He proposed estimators of the slope and intercept. The estimation of the parameters of the one-way design of experiment lay-out and multiple regression models using RSS are considered by Muttlak (1996a, 1996b) respectively. He showed that estimators of the parameters of interest based on RSS are more efficient than their counterpart SRS estimators. He used real data to illustrate the computations.

2) Samawi and Muttlak (1996) used the RSS to estimate the ratio of two population means. They showed that using RSS would increase the efficiency of the ratio estimator of RSS data as compared to the SRS ratio estimator.

3) Yu and Lam (1997) considered the case when the variable of interest Y is difficult to rank and measure, but there is a concomitant variable X that can be used to estimate the rank of Y. They proposed regression-type RSS estimators of the mean of the population for variable Y when the population mean of X is known and when it is unknown. Under the assumption that Y and X are jointly following the bivariate normal distribution, they showed that their estimators are more efficient than the RSS and SRS estimators unless $|\rho| < 0.40$. Finally they considered the case when the normality assumption does not hold and pointed out that

the results still hold if the shape of variable X only slightly departs from symmetry.

4) Following the footsteps of Patil et al. (1993), Muttlak (1998) compared the median RSS estimator to the SRS regression and RSS estimators of the population mean. He showed that the median RSS estimator is more efficient than SRS regression estimator is unless is $|\rho| > 0.90$. Also, the median RSS estimator dominates the RSS estimator under the assumption that the main and auxiliary variables are jointly following the bivariate normal distribution.

5) Barreto and Barnett (1999) considered a different approach from the one considered by Muttlak (1995) to estimate the slope and the intercept of the simple linear regression line. They considered m_i samples of size m_i at each value of the independent variable x_i, from which they choose and measure the RSS $y_{i,\,1(1)}$, $y_{i,\,2(2)}$, ..., $y_{i,\,m_i(m_i)}$, i =1, …, n. Clearly they finish up with *n* RSS samples one at each value of the independent variable. The best linear unbiased estimators for the slope and intercept are proposed based on the RSS data and shown to be more efficient for normal data than the usual simple linear regression estimators.

6. BAYESIAN STATISTICS WITH RANKED SET SAMPLING

1) Al-Saleh and Muttlak (1998) were the first to consider the Bayesian estimation using RSS. In this study the Bayesian estimation based on RSS was investigated for exponential and normal distribution. Given a RSS from the exponential distribution and using the inverse gamma prior for the mean of the distribution, the Bayes estimator was derived. The Bayes risk of this estimator was compared to the Bayes risk of the corresponding estimator using SRS. It turned out that the Bayes estimator with ranked set sampling is more efficient than that with SRS. Depending on the parameters of the prior and the set size, the efficiency can be as large as 6.8 for set size 5 and as low as 1.5 for set size 2. Similar results are obtained for the normal distribution. The Authors realized the complication of the close form of the Bayes estimate. For example, for the estimation of the exponential mean θ, when the prior is inverse gamma with $\alpha = r,\ \beta = 1$, the closed form of the Bayes estimator for a RSS sample of size 2 takes the form

$$\hat{\theta} = \frac{1}{r+1}\frac{(2y_1 + y_2 + 1)^{-(r+1)} - (2y_1 + 2y_2 + 1)^{-(r+1)}}{(2y_1 + y_2 + 1)^{-(r+2)} - (2y_1 + 2y_2 + 1)^{-(r+2)}}\ .$$

2) Kim and Arnold (1999) considered a generalized version of RSS, Bayesian parameter estimation of some specified parameters under both balanced and generalized RSS was accomplished using the Gibbs sampler. An algorithm was provided. The Authors considered the case of the exponential mean with gamma prior. Their numerical results show that the RSS is superior to the SRS.

3) Kvam and Tiwari (1999) considered the Bayes estimation of the distribution function from RSS data. They used the Singular ordered Dirichlet distribution as a prior. They derived the Bayes estimator of the distribution function as well as the generalized MLE, using the mean squared error loss function. No close form was provided in either case. The Authors used the Gibbs sampling technique to approximate the estimators. The methods were illustrated with data from the Natural environmental Research council of Great Britain, representing water discharge of food on the Nidd River in Yorkshire, England.

4) Lavine (1999) examined the RSS procedure from a Bayesian point of view. He determined whether RSS provides advantages over simple random sampling, and explored some optimality questions. The following is a summary of the main results of the paper:

Let X be the data from an experiment with density f and let $p(\theta)$ be the prior density of θ then the quantity

$$E(I(X \mid p)) = \iint p(\theta) f(x \mid \theta) \log \frac{f(x \mid \theta)}{f(x)} dx$$

is used to measure the expected utility of a sample. The author stated and proved the following theorem:

For any prior p and sample size $n > 0$, there exists a collection of ranks $r_1, r_2, \cdots, r_n$ such that the expected information in a RSS $Y_{r_1}, Y_{r_2}, \cdots, Y_{r_n}$ is greater than or equal to the expected information in a simple random sample of size n. The author provided an example to show that it is not necessarily true for arbitrarily selected ranks, that RSS sampling is more informative than SRS.

5) Al-Saleh, Al-Shrafat, and Muttlak (2000) considered the Bayesian estimation based on RSS. They showed that the Bayes risk of the Bayes estimator based on SRS is the average Bayes risk of all possible n^n RSS plans plus a positive quantity. So, it was concluded that there exists at least one RSS plan, for which the Bayes estimator dominates the Bayes estimator based on SRS. i.e., the Bayes risk of the Bayes estimator with respect to RSS is smaller than the Bayes risk of the Bayes estimator with respect to SRS for at least one RSS

plan. If f belongs to the exponential family then a dominant RSS plan is the balanced one. Milk yields of 403 sheep was collected and used to evaluate RSS and Bayesian RSS.

6) It was noted by Al-Saleh and AbuHawwas (2000) that the RSS Bayes estimators have very complicated closed forms even for small sample sizes. Their computations in the simple exponential case for sample sizes of 2 and 3 demonstrated this fact. This is due to the fact that the likelihood equation becomes very complicated, thus making the posterior also complicated no matter how simple the prior may be. These complications are due to the fact that there is no minimal sufficient statistic of lower dimension than the dimension of the data itself. This lower dimension minimal sufficient statistic usually exists in the case of SRS. In this paper the Authors used the concept of multiple imputation proposed by Rubin (1987), to obtain a formula that relates the posterior of the parameter using RSS to that using the full data. This formula facilitates the study of some of the theoretical properties of Bayes estimators and also provides clues for approximating the complicated exact forms.

7. OTHER WORKS BASED ON RANKED SET SAMPLING

Almost all the RSS works reviewed so far assume that we are sampling from an infinite population.

1) Patil, Sinha and Taillie (1995) were first to consider RSS and sampling from a finite population. They showed that RSS sampling mean $\hat{\mu}_{RSS} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{(i:n)j}$ from a finite population is an unbiased estimator for the population μ. They derived more than one expression for the variance of $\hat{\mu}_{RSS}$. The final expression is given by

$$\mathrm{var}(\hat{\mu}_{RSS}) = \frac{1}{nr}\left\{\frac{N-1-nr}{N-1}\quad \sigma^2 - \overline{\gamma}\right\},$$

where, N is the population size, $\overline{\gamma} = \frac{n!(n-1)!}{N(N-1)\ldots(N-2n+1)}\quad \gamma$, and $\gamma = (\mathrm{x} - \mu)'\,\Gamma(\mathrm{x} - \mu)$. The entries of the matrix $\Gamma_{N\mathrm{x}N}$ are functions of the population size N and the size of the set n, but do not depend on the population values x. When the population follows a linear or quadratic trend, they derived an explicit form for $\mathrm{var}(\hat{\mu}_{RSS})$. Finally they noticed that the efficiency of the RSS depends on the number cycles r unlike the infinite population case.

2) Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be a SRS without replacement from a finite population. Takahasi and Futatsuya (1998) showed that the joint distribution of $X_{(i)}$ and $X_{(j)}$ is positively likelihood ratio dependent and $Y_{(j)}$ is negatively regression dependent on $X_{(i)}$, where $X_{(i)}$, $X_{(j)}$ and $Y_{(j)}$ are the ith and jth order statistics. They used these results to show that when sampling without replacement from a finite population the efficiency of RSS estimator of the population mean with respect to SRS estimator is bounded below by one.

3) Barnett (1999) investigated the use of RSS method in estimating parameters of environmental variables. He considered various RSS estimators for the population mean. These estimators are the sample mean of SRS ($\overline{X}$), the mean of MTW RSS ($\overline{X}_{RSS}$), the estimator proposed by Kaur et al. (1997) $\overline{\overline{X}}_r$ and the estimator suggested by Barnett and Moore (1997) μ_X^*. He explored the various properties of the above estimators using the lognormal and extreme value distributions. In all case he showed that μ_X^* achieved highest efficiency. The amount of gain over $\overline{X}$ and $\overline{\overline{X}}_r$ depend on the underlying distribution.

4) Mode, Conquest and Marker (1999) investigated the conditions under which the RSS method will be a cost-efficient sampling method for environmental and ecological filed studies compare to SRS. They assumed that the ranking of the units is not cost free, but it will cost money. They present the ratios of measuring to ranking cost for the normal and exponential distributions with and without errors in ranking. They also presented the ratio of measuring to ranking cost for real life problem.

5) Samawi (1999) is the first to use RSS in the area of simulation. He showed that the efficiency of Monte Carlo methods of integral estimation could be substantially improved by using ranked simulated samples (RSIS) in place of uniform simulated samples (USIS). The author considered the integral of the form $\theta = \int_0^1 k(u)du$. Note that θ is simply $E(k(U))$, where $U \sim U(0,\ 1)$. Usual Monte Carlo methods simulate n values $u_1, u_2, \cdots, u_n$ from $U(0,\ 1)$ and approximate θ by the average of these values. In this paper instead, a RSS sample $u_{(1)}, u_{(2)}, \cdots u_{(n)}$ is simulated from the uniform distribution. This can be done by taking $U_{(i)}$ from f_i, where f_i is the density of the ith order statistics of a sample of size n from the uniform distribution. f_i is actually the density of a beta random variable with parameters i & $n-i+1$. It turns that the

average of these RSS values is a more efficient estimator than the usual estimator. Different Monte Carlo methods such as Crude, Antithetic, Control Variate and Importance sampling methods are investigated and found to benefit from this new sampling scheme.

6) Al-Saleh and Samawi (2000) used the idea of MSRSS and Steady state RSS introduced by Al-Saleh and Al-Omari (1999) to approximate integrals via Monte Carlo methods. This method provides substantial improvement over the usual Monte Carlo methods. The suggested procedure for estimating $\theta = \int_0^1 k(u)du$ consists of generating an independent sample, say

$$U_i^{\infty}, U_2^{\infty}, \cdots U_m^{\infty} \text{ from } f_i^{(\infty)}(x) = \begin{cases} m & , \quad \frac{i-1}{m} < x < \frac{i}{m} \\ 0 & , \quad o.w. \end{cases}$$

This sample is called a steady state simulated sample (SRSIS). The average of this sample is used to estimate the above integral instead of a sample from the standard uniform distribution (USIS). The generation of SRSIS costs no extra computer time. It is shown that the estimator of θ using SRSIS is unbiased with variance strictly less than that using USIS, and hence is more efficient. The efficiency for evaluating the quantity $\int_0^1 \exp(u^2)du$ ranges from 1348 for n=50 to 33719 for n=300. This actually shows how powerful and time saving this method is in approximating complicated integrals. Different Monte Carlo methods such as Crude, Antithetic, Control Variate, and Importance sampling methods were investigated and found to benefit from this new sampling scheme.

REFERENCES

1) Abu-Dayyeh, W. and Muttlak, H.A. (1996). Using ranked set sampling for hypothesis tests on the scale parameter for the exponential and uniform Distributions. *Pak. J. Statist.*, 12(2), 131-138.
2) Al-Saleh, M.F. and Muttlak, H.A. (1998). A note on Bayesian estimation using ranked set sample. *Pak. J. Statist.*, 14(1), 49-56.
3) Al-Saleh, M.F. and Al-Kadiri, M.A. (1999). Double ranked set sampling. *Statistics and Probability Letters*, 48, 205-212.
4) Al-Saleh, M.F. and Al-Omari, A. (1999). Multistage ranked set sampling. Submitted to *Statist. Plann. and Infer*.
5) Al-Saleh, F.M., Al-Shrafat, K. and Muttlak, H.A. (2000). Bayesian estimation using ranked set sampling. *Biometrical Journal*, 42, 3, 1-12.

6) Al-Saleh, F.M. and Samawi, H.M. (2000). On the efficiency of Monte Carlo methods using steady state ranked simulated samples. *Commun. Statist. – Simula*, 29 (3), 941-954.
7) Al-Saleh, M.F. and Abu Hawwas, J. (2000). Characterization of ranked set sampling Bayes estimators with application to the normal distribution, submitted to *Statistics and Probability Letters*.
8) Aragon, M.E.D., Patil, G.P., and Taillie, C. (1999). A performance indicator for ranked set sampling using raking error probability matrix. *Environmental and Ecological Statistics*, 6, 75-89.
9) Barabesi, L. (1998). The computation of the distribution of the sign test statistic for ranked-set sampling. *Commun. Statist. Simula.* 27(3), 833-842.
10) Barnett, V. and Moore, K. (1997). Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering. *J. App. Statist.*, 24(6), 697-710.
11) Barnett, V. (1999). Ranked set design for environmental investigations. *Environmental and Ecological Statistics*, 6, 59-74.
12) Barreto, M.C.M. and Barnett, V. (1999). Best linear unbiased estimators for the simple linear regression model using ranked set sampling. *Environmental and Ecological Statistics*, 6, 119-133.
13) Bhoj, D.S. and Ahsanullah, M. (1996). Estimation of parameters of the generalized geometric distribution using ranked set sampling. *Biometrics*, 52, 685-694.
14) Bhoj, D.S. (1997a). Estimation of parameters of the extreme value distribution using ranked set sampling. *Commun. Statist. Theory Meth.* 26(3), 653-667.
15) Bhoj, D.S. (1997b). New parametric ranked set sampling. *App. Statist. Scien.,* 6(4), 275-289.
16) Bhoj, D.S. (1997c). Estimation of parameters using modified ranked set sampling. *App. Statist. Scien.,* 6(2), 145-163.
17) Bhoj, D.S. (1999a). Estimation of parameters of the exponential distribution using three ranked set sampling procedures. *App. Statist. Scien.,* 8(3), 155-172.
18) Bhoj, D.S. (1999b). Minimum variance linear unbiased estimators of the Rayleigh Parameter based on ranked set sampling procedures. *App. Statist. Scien.,* 8, 269-277.
19) Bohn, L.L. and Wolfe, D.A. (1992). Nonparametric two-sample procedures for ranked set samples data. *J. Amer. Statist. Assoc.*, 87, 552-561.
20) Bohn, L.L. and Wolfe, D.A. (1994). The effect of imperfect judgment rankings on properties of procedures based on ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic. *J. Amer. Statist. Assoc.*, 89, 168-176.
21) Bohn, L.L. (1994). A ranked-set sample signed-rank statistic. Department of Statistics technical report number 426. The University of Florida, Gainesville, FL.

22) Bohn, L.L. (1996). A review of nonparametric ranked-set sampling methodology. *Commun. Statist. Theory Meth.* 25(11), 2675-2685.
23) Chen, Z. (1999). Density estimation using ranked-set sampling data. *Environmental and Ecological Statistics*, 6, 135-146.
24) Chen, Z. (2000). On ranked-set sample quantiles and their applications. *Journal of Statistical Planning and Inference*, 83, 125-135.
25) Chen, Z. (2000). The efficiency of ranked-set sampling relative to simple random sampling under multi-parameter families. *Statistica Sinica*. 10, 247-263.
26) Dell, D.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-55.
27) Downton, F. (1954). Lest-squares estimates using ordered observations. *Ann. Math. Statist.*, 25, 303-316.
28) Fei, H. and Sinha, B.K., Wu, Z. (1994). Estimation of parameters in two-parameter weibull and extreme-value distributions using ranked set sampling. *J. Statist. Res.*, 28, 149-161.
29) Hartlaub, B.A. and Wolfe, D.A. (1999). Distribution-free ranked set sample procedures for umbrella alternatives in the m-sample setting. *Environmental and Ecological Statistics*, 6, 105-118.
30) Hettmansperger, T. (1995). The ranked-set sample sign test. *Nonparametric Statistics,* 4, 263-270.
31) Hossain, A.S. and Muttlak, H.A. (1999). Paired ranked set sampling: A more efficient procedure. *Environmetrics*, 10, 195-212.
32) Huang, J. (1997). Asymptotic properties of the NPMLE of a distribution function based on ranked set samples. *The Annals of Statistics*, 25(3), 1036-1049.
33) Kaur, A., Patil, G.P., Sinha, B.K. and Taillie, C. (1995). Ranked set sampling: an annotated bibliography. *Environmental and Ecological Statistics*, 2, 25-54.
34) Kaur, A, Patil, G.P. and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions. *Biometrics*, 53, 123-130.
35) Kim, Y.H. and Arnold, B.C. (1999). Estimation of a distribution function under generalized ranked set sampling. *Commun. Statist. –Simula,* 28(3), 657-666.
36) Kim, Y.H. and Arnold, B.C. (1999). Parameter estimation under generalized ranked set sampling. *Statistics & Probability Letters*, 42, 353-360.
37) Koti, K.M. and Babu, G.J. (1996). Sign test for ranked-set sampling. *Commun. Statist. Theory Meth.* 25(7), 1617-1630.
38) Kvam, P.H. and Samaniego, F.J. (1993). On the inadmissibility of empirical averages as estimators in ranked set sampling. *Journal of Statistical Planning and Inference*, 36, 39-55.
39) Kvam, P.H. and Samaniego, F.J. (1994). Nonparametric maximum likelihood Estimation based on ranked set samples. *Journal of American Statistical Association*, 89, 526-537.

40) Kvam, P.H. and Tiwari, R.C. (1999). Bayes estimation of a distribution function using ranked set sampling. *Environmental and Ecological Statistics*, 6, 11-22
41) Lam, K., Sinha, B.K. and Wu, Z. (1994). Estimation of parameters in two-parameter exponential distribution using ranked set sample. *Ann. Inst. Statist. Math.,* 46, 723-736.
42) Lam, K., Sinha, B.K. and Wu, Z. (1995). Estimation of location and scale parameters of a logistic distribution using ranked set sample. In: Nagaraja, Sen and Morrison, ed., *Papers in Honor of Herbert A. David*, 187-197.
43) Lavine, M. (1999). The bayesics of ranked set sampling. *Environmental and Ecological Statistics*, 6, 47-57.
44) Li, D. Sinha, B.K. and Chuiv, N.N (1999). On estimation of P (X> c) based on ranked set sample. Statistical Inference and Design, Dixit and Satam (Eds), 47-54.
45) Li, D. Sinha, B.K. and Perron, F. (1999). Random selection in ranked set sampling and its applications. *J. Statist. Plann. and Infer.*, 76, 185-201.
46) Li, D. and Chuiv, N.N. (1997). On the efficiency of ranked set sampling strategies in parametric estimation. *Cal. Statist. Assoc. Bull.*, 47, 23-42.
47) Lieblein, J. and Zelen, M. (1956). Statistical investigations of the fatigue life of deep-grove ball bearings. *J. Res. Nat. Bur. Stand.*, 57, 273-316.
48) McIntyre, G. A. (1952). A method of unbiased selective sampling, using ranked sets. *Austr. J. Agri. Res.*, 3, 385-390.
49) Mode, N.A., Conquest, L.L. and Marker, D.A. (1999). Ranked set sampling for ecological research: Accounting for the total costs of sampling. *Environmetrics*, 10, 179-194.
50) Muttlak, H.A. (1995). Parameters estimation in a simple linear regression using rank set sampling. *Biometrical Journal*, 37(7), 799-810.
51) Muttlak, H.A. (1996a). Estimation of parameters for on-way layout with rank set sampling. *Biometrical Journal*, 38(4), 507-515.
52) Muttlak, H.A. (1996b). Estimation of parameters in a multiple regression model using rank set sampling. *Information & Optimization Sciences*, 17(3), 521-533.
53) Muttlak, H.A. (1996c). Pair ranked set sampling. *Biometrical Journal*, 38(7), 897-885.
54) Muttlak, H.A. (1997). Median ranked set sampling. *J. App. Statist. Scien.*, 6(4), 245-55.
55) Muttlak, H.A. (1998). Median ranked set sampling with concomitant variables and comparison with ranked set sampling and regression estimators. *Environmetrics*, 9, 255-267.

56) Muttlak, H.A. and Abu-Dayyeh, W. (1998). Testing some hypotheses about the normal distribution using ranked set sampling: a more powerful test. *Journal of Information & Optimization Sciences*, 19(1), 1-11.
57) Muttlak, H.A. (1999a). Median ranked set sampling with size biased probability of selection. *Biometrical Journal*, 40(4), 455-465.
58) Muttlak, H.A. (1999b). On extreme ranked set sampling with size biased probability of selection. *Far East Journal Theory of Statistics*, 3(2), 319-329.
59) Ni Chuiv, N., Sinha, B.K. and Wu, Z. (1994). Estimation of location parameter of a Cauchy distribution using ranked set sample. Technical Report, University of Maryland Baltimore County.
60) Ni Chuiv, N.N. and Sinha, B.K. (1998). On some aspects of ranked set sampling in parametric estimation. Handbook of statistics, vol. 17, 337-377.
61) Özturk, Ö. (1999a). One-and two-sample sign tests for ranked set sample selective designs. *Commun. Statist. Theory Meth.* 28(5), 1231-1245.
62) Özturk, Ö. (1999b). Two-sample inference based on one-sample ranked set sample sign statistics. *Nonparametric Statistics*. 10, 197-212.
63) Özturk, Ö. and Wolfe, D.A. (2000). Alternative ranked set sampling protocols for sign test. *Statistics & Probability Letters*, 47, 15-23.
64) Patil, G.P., Sinha, A.K. and Taillie, C. (1993). Relative precision of ranked set sampling: A comparison with the regression estimator. *Environmetrics*, 4, 399-412.
65) Patil, G.P., Sinha, A.K. and Taillie, C. (1994). Ranked set sampling. *A Handbook of Statistics*, Vol. 12, pp167-200.
66) Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Finite population corrections for ranked set sampling. *Ann. Inst. Statist. Math.,* 47(4), 621-636.
67) Patil, G.P., Sinha, A.K. and Taillie, C. (1999). Ranked set sampling: a bibliography. *Environmental and Ecological Statistics*, 6, 91-98.
68) Presnell, B. and Bohn, L.L. (1999). U-Statistics and imperfect ranking in ranked set sampling. *Nonparametric Statistics,* 10, 111-126.
69) Rubin, D. (1987). Multiple imputation for nonresponse in sample surveys and censuses, Wiley, New York.
70) Samawi, H.M., Ahmed, M.S. and Abu-Dayyeh. Estimating the population mean using extreme raked set sampling. *The Biometrical Journal*, 38(5), 577-86.
71) Samawi, H.M. and Muttlak, H.A. (1996). Estimation of ratio using rank set sampling. *Biometrical Journal*, 38(6), 753-764.
72) Samawi, H.M. (1999). More efficient Monte Carlo methods obtained by using ranked set simulated sample. *Commun. Statist. –Simula,* 28(3),
73) Shen, W.H. and Yuan, W. (1996). A test for a normal mean based on a modified partial ranked set sampling. *Pak. J. Statist.*, 11(3), 227-233.
74) Shen, W.H. (1994). Use the ranked set sampling for testing a normal mean. *Cal. Statist. Assoc. Bull.*, 44, 183-193.

75) Sinha, B.K., Wu, Z. and Fei, H. (1994). Estimation of a gamma mean based on ranked set sample. *Pak. J. Statist.*, 10(1), 235-249.
76) Sinha, B.K., A.K. Sinha, and Purkayastha, S. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics and Decisions,* 14, 223-240.
77) Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set Samples. *Biometrics,* 36, 35-42.
78) Stokes, S.L. and Sager, T.W. (1988). Characterization of a ranked-set sample with application to estimating distribution function. *J. Amer. Statist. Assoc.*, 83, 374-381.
79) Stokes, S.L. (1995). Parametric ranked set sampling. *Ann. Inst. Statist. Math.,* 47(3), 465-482.
80) Takahasi K. and Wakimoto K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.,* 21, 249-55.
81) Takahasi K. and Futatsuya, M. (1998). Dependence between order statistics in samples from finite population and its application to rank set sampling. *Ann. Inst. Statist. Math.,* 50, 49-70.
82) Tam, C.Y.C., Yu, P.L.H. and Fung, T.W.K. (1998). Sensitivity analysis of BLUE for the population mean based on a ranked set sample. *Commun. Statist. Theory Meth.* 27(4), 1075-1091.
83) Yu, P.L.H. and Lam, K. (1997). Regression estimator in ranked set sampling. *Biometrics*, 53, 1070-1080.
84) Yu, P.L.H., Lam, K. and Sinha, B.K. (1997). Estimation of mean based on unbalanced ranked set sample. Applied Statistical Science II, 87-97.
85) Yu, P.L.H., Lam, K. and Sinha, B.K. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. *Environmental and Ecological Statistics*, 6, 23-45.

CHAPTER FIVE

Random Ranked Set Samples

[1]**I. Rahimov** and **H.A. Muttlak**
Department of Mathematical Sciences,
King Fahd University of Petroleum & Minerals
Dhahran 31261, Saudi Arabia.

ABSTRACT

Ranked set sampling (RSS) as suggested by McIntyre (1952) uses fixed set size and number of cycles (or replications). In real life however, we may encounter problems that requiring random set size or number of cycles or both. In dealing with such problems we suggest several unbiased estimators of the population mean using random ranked set sampling (RRSS) method. The efficiencies of the estimators of the population mean under RRSS and RSS are compared. The results show, under certain conditions, the efficiency of estimators is improved by using RRSS. The asymptotic properties of the newly suggested estimators are also considered.

KEY WORDS

Asymptotic properties; discrete uniform distribution; efficiency; random number of replications; random set size; ranked set sampling.

1. INTRODUCTION

The ranked set sampling (RSS) has attracted number of authors as an efficient sampling method, particularly in the area of environmental and ecological investigations. The RSS proposed by McIntyre (1952) is a sampling method proven to be more efficient when units are difficult and costly to measure, but are easy and cheap to rank with respect to a variable of interest without actual measurement. One can often tell which tree is the tallest without measuring all the trees. The RSS method can be summarized as follows: From a population of interest, k random sets each of size *k* are selected. The members of each random set are ranked with respect to the variable of interest by a cost free method e.g. eye

[1] I. Rahimov is Professor of Probability and Statistics and H.A. Muttlak is Associate Professor of Statistics, Department of Mathematical Sciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia. The authors are indebted to King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia, for excellent research facilities.

inspection. From the first ordered set, the smallest unit is selected for measurement. From the second ordered set, the second smallest unit is selected for measurement. This continues until the largest element from the last ordered set is measured. This process may be repeated r times (i.e. r cycles or replications) to yield a sample of size rk.

The RSS is based on fixed set sizes and number of replications. But in some applications we might be faced with problems where k, r, or both cannot be fixed. The example considered by Muttlak and McDonald (1992) demonstrates the need for at least r to be random. The line intercept (transect) is a widely used sampling method in ecological and environmental studies. Units are plants, animal species and etc. The RSS method will apply after the units are sampled using the line transect in the first phase. We know that the number of units n_1 (say) that are intercepted by the line can not be fixed. Muttlak and McDonald (1992) assumed that $n_1 \geq k^2 r$. Obviously, this assumption will be violated in most applications and hence we cannot use RSS. As we can see, this situation can be handled easily using the random ranked set sampling (RRSS) method by letting either r or k to be random.

Takahasi and Wakimoto (1968) supplied mathematical theory to support McIntyre's (1952) suggestion. Stokes and Sager (1988) developed the properties of the empirical distribution function based on RSS and compared these properties to the usual empirical distribution function using simple random sample (SRS) data. Bohn and Wolfe (1992, 1994) developed the Mann-Whitney-Wilcoxon statistic using RSS for both perfect ranking and ranking with errors. Kvam and Samaniego (1993, 1994) developed the estimation of the population distribution function and population mean using unbalanced RSS data i.e. the size of the i^{th} set need not be the same for all sets and the various order statistics need not be represented an equal number of time. Koti and Babu (1996) derived the exact null distribution of the RSS sign test. Huang (1997) considers the asymptotic properties of the nonparametric maximum likelihood estimator (NPMLE) of a distribution function using RSS. Kim and Arnold (1999a, 1999b) considered estimating the distribution function F and Bayesian parameter estimation for specified parameter under both balanced and unbalanced RSS. Hartlaub and Wolfe (1999) generalized the one- and two-sample location problems considered in the previous nonparametric work in the area of RSS to m-sample location problem. Presnell and Bohn (1999) developed the U-statistics using RSS data for one and two sample cases. Özturk (2000) investigated the effect of different RSS protocols on sign test statistic.

Several authors considered some modifications of the RSS method either to improve the efficiency of the estimators or (and) to make the RSS method easier to implement in the field. Samawi et al. (1996) studied the properties of the extreme ranked set sampling (ERSS) in estimating the population mean. Muttlak (1997) suggested the use of median ranked ret sampling (MRSS) to estimate the population mean.

Li et al. (1999) introduced the notion of random selection of m sets out of k sets, m < k, where k is the set size in the usual RSS method. They studied the properties of the estimators of the population mean and variance based on the new randomly selected sample.

For classified and extensively reviewed work in the area of RSS from 1952 to 1994 see Patil et al. (1994) and Kaur et al. (1995). Finally for bibliography in the area of RSS see Patil et al. (1999).

In this paper we provided a new direction of RSS via the notion of random ranked set sampling (RRSS). In Section 2 the idea of RRSS is introduced in the case of one replication i.e. single cycle. The cases of r replications with random set size and fixed set size with random number of replications are discussed in Sections 3 and 4 respectively. The general case of RRSS with random set sizes and replications is considered in Section 5. The asymptotic properties of the estimator of the population mean suggested for the general case of RRSS is established in Section 6. In Section 7 we calculate the efficiency of the newly suggested estimators for specific probability distributions and compare these to the RSS. Some concluding remarks are given in the last Section.

2. SINGLE CYCLE WITH RANDOM SET SIZE

We consider the following family of random variables

$X_{11}, X_{12}, ..., X_{1\ell}; X_{21}, X_{22}, ..., X_{2\ell}; ....; X_{i1}, X_{i2}, ..., X_{i\ell}; ...; X_{\ell 1}, X_{\ell 2}, ..., X_{\ell\ell}$ where $X_{ij}, i, j = 1, 2, ...; \ell \in \Lambda = \{2, 3, ...\}$ are independent and identically distributed random variables with cdf F(x), pdf f(x), mean μ and variance σ^2. Let ν be a random variable taking values from $\Lambda = \{2, 3, ...\}$. Let $X_{i(1)}^{(\nu)}$, $X_{i(2)}^{(\nu)}$, ..., $X_{i(\nu)}^{(\nu)}$ be the order statistics of $X_{i1}, X_{i2}, ..., X_{i\nu}, i = 1, 2, ..., \ell$. To simplify the notations for any $\ell \in \Lambda$, we will use $y_i^{(\ell)} = X_{i(i)}^{(\ell)}$, $i = 1, 2, ..., \ell$. It is easy to see that $y_i^{(\ell)}, i = 1, 2, ..., \ell$ are independent but not identically distributed random variables. We propose

$$\bar{y}_{(\nu)} = \frac{1}{\nu} \sum_{i=1}^{\nu} y_i^{(\nu)} \tag{1}$$

as an estimator of the population mean μ. Assume from now on that the random variable ν and the family of random variables $X_{ij}^{(\ell)}$ are independent. We denote also the cdf, pdf, mean and variance of $y_i^{(\ell)}$ by $F_{\ell i}(x)$, $f_{\ell i}(x)$, $\mu_{\ell i}$, and $\sigma_{\ell i}^2$ respectively. It follows from the definition of $y_i^{(\ell)}$ for any $\ell \in \Lambda$ that

$$f(x) = \frac{1}{\ell}\sum_{i=1}^{\ell} f_{\ell i}(x). \tag{2}$$

The properties of the estimator $\overline{y}_{(\nu)}$ are:

(a) $\overline{y}_{(\nu)}$ is an unbiased estimator of population mean μ with a variance

(b) $V(\overline{y}_{(\nu)}) = E\left[\frac{1}{\nu}\sigma_{(\nu)}^2\right]$, where $\sigma_{(\ell)}^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}\sigma_{\ell i}^2$.

We can easily proof (i) by using the total probability formula and equation (2). For any ℓ we have

$$E[\frac{1}{\ell}\sum_{i=1}^{\ell} y_i^{(\ell)}] = \frac{1}{\ell}\sum_{i=1}^{\ell}\int_{-\infty}^{\infty} x f_{\ell i}(x)dx = \int_{-\infty}^{\infty} x \frac{1}{\ell}\sum_{i=1}^{\ell} f_{\ell i}(x)dx = \mu,$$

then

$$E(\overline{y}_{(\nu)}) = E\{E[\frac{1}{\nu}\sum_{i=1}^{\nu} y_i^{(\nu)} | \nu]\} = \sum_{\ell=1}^{\infty} E[\frac{1}{\ell}\sum_{i=1}^{\ell} y_i^{(\ell)}]P(\nu = \ell) = \mu. \tag{3}$$

This shows that $\overline{y}_{(\nu)}$ is an unbiased estimator of population mean μ for any random variable ν. Now we consider the proof of (ii). Again by using the total probability formula we can write

$$V(\overline{y}_{(\nu)}) = \sum_{\ell=1}^{\infty} E\{[\frac{1}{\nu}\sum_{i=1}^{\nu} y_i^{(\nu)} - \mu]^2 | \nu = \ell\} P(\nu = \ell)$$

$$= \sum_{\ell=1}^{\infty} E\{\frac{1}{\ell}\sum_{i=1}^{\ell}[y_i^{(\ell)} - E(y_i^{(\ell)})]\}^2 P(\nu = \ell),$$

then $V(\overline{y}_{(\nu)}) = E[\frac{1}{\nu^2}\sum_{i=1}^{\nu}\sigma_{\ell i}^2]$. If denote $\sigma_{(\ell)}^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}\sigma_{\ell i}^2$, then we get

$$V(\overline{y}_{(\nu)}) = E\left[\frac{1}{\nu}\sigma_{(\nu)}^2\right]. \tag{4}$$

Let us consider particular cases of the formula (4) for given distributions of ν.

Example 1: Assume that ν has geometric distribution truncated at one with a parameter p. In this case we obtain $V(\overline{y}_{(\nu)})$ in equation (4) as

$$V(\overline{y}_{(\nu)}) = p \sum_{\ell=2}^{\infty} \frac{1}{\ell} q^{\ell-2} \sigma^2_{(\ell)} , \tag{5}$$

where q = 1– p.

Example 2: Let ν have a binomial distribution truncated at zero and one with parameters n and p i.e. $P(\nu = k) = P(\xi = k \mid \xi > 1)$, $k = 2,3,...,n$ where ξ is a binomial random variable of the same parameters. In this case the variance of $V(\overline{y}_{(\nu)})$ is

$$V(\overline{y}_{(\nu)}) = \frac{q^n n!}{1 - q^n - npq^{n-1}} \sum_{\ell=2}^{n} \frac{1}{\ell!(n-\ell)!\ \ell} \left(\frac{p}{q}\right)^{\ell} \sigma^2_{(\ell)} . \tag{6}$$

Example 3: Let us assume that ν has a uniform distribution on the set {2, 3,…, N}. In this case the variance $V(\overline{y}_{(\nu)})$ is given by

$$V(\overline{y}_{(\nu)}) = \frac{1}{N-1} \sum_{\ell=2}^{N} \frac{1}{\ell} \sigma^2_{(\ell)} . \tag{7}$$

To compare the proposed estimator $\overline{y}_{(\nu)}$ to the RSS estimator, $\overline{y}_{(k)} = \frac{1}{k}\sum_{i=1}^{k} y_{(i)}$, where $y_{(i)}$ is the i[th] order statistic from the i[th] set of fixed size k, k =2, 3,…, N, it is easy to see that $V(\overline{y}_{(k)}) = \frac{1}{k}\sigma^2_{(k)}$, where $\sigma^2_{(k)} = \frac{1}{k}\sum_{i=1}^{k} \sigma^2_{ki}$. As shown by Takahasi and Wakimoto (1968), $\sigma^2_{(k)} > \sigma^2_{(k+1)}$ and consequently $\frac{1}{k}\sigma^2_{(k)}$ is also decreasing on k. Using these results and equation (7) we may state the following proposition.

Proposition 1: There exist $2 \le \tau \le N$ such that $V(\overline{y}_{(k)}) > V(\overline{y}_{(\nu)})$ for $k \le \tau$ and $V(\overline{y}_{(k)}) < V(\overline{y}_{(\nu)})$ for $\tau < k \le N$.

It is clear that the number τ depends on the form of the initial density function f(x). In Section 7 we will consider different concrete distributions to obtain the value τ.

3. FIXED NUMBER OF CYCLES WITH RANDOM SET SIZE

Let $\nu_1, \nu_2, ..., \nu_m$ are independent and identically distributed random variables taking values from Λ. If the process of Section 2 is repeated m times, i.e. we replicate the cycle m times with set sizes ν_i, where i = 1, 2,…, m, we will obtain a sequence of estimators $\bar{y}_{\nu_1 1}, \bar{y}_{\nu_2 2}, ..., \bar{y}_{\nu_m m}$. It is clear that $\bar{y}_{\nu_i i}, i \geq 1$ are independent and identically distributed with the following mean and variance respectively

$$E(\bar{y}_{\nu_i i}) = \mu \,,\; V(\bar{y}_{\nu_i i}) = E\left[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}\right],\; i = 1, 2, ..., m \tag{8}$$

Also, they have the common characteristic function

$$\varphi(t) \equiv E\left[e^{it\bar{y}_{\nu_i i}}\right] = E\left[\prod_{j=1}^{\nu_i}\varphi_j^{(\nu_i)}(t)\right], \tag{9}$$

for $i \geq 1$, where $\varphi_j^{(\nu_i)}(t)$ is the characteristic function of $y_j^{(\nu_i)}$, the j[th] order statistic with set size ν_i. We proposed as an estimator for the population mean μ as

$$\bar{\bar{y}}_{(m)} = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_{\nu_i i} \,. \tag{10}$$

It is not difficult to show that $\bar{\bar{y}}_{(m)}$ is an unbiased estimator for the population mean μ with a variance

$$V(\bar{\bar{y}}_{(m)}) = \frac{1}{m}E\left[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}\right]. \tag{11}$$

To compare the proposed estimator $\bar{\bar{y}}_{(m)}$ with a similar estimator in the usual RSS case where $\nu_1 = \nu_2 = ... = \nu_m = k$, we have again to make comparison between $E\left[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}\right]$ and $\sigma^2_{(k)}$. For example, this comparison may use Proposition 1 in the case when the random variables $\nu_i, i \geq 1$ have common discrete uniform distribution.

4. RANDOM NUMBER OF CYCLES WITH FIXED SET SIZE

Assume now that the set size is fixed and equal to k. From the usual (fixed set size) RSS, we have that $\overline{y}_{(k)}$ is an unbiased estimator of the population mean μ. Let the number of replications θ be a random variable taking values from Λ and independent of $\overline{y}_{(k)}$. We consider

$$\overline{\overline{y}}_{(k)} = \frac{1}{\theta} \sum_{i=1}^{\theta} \overline{y}_{(k)i} \tag{12}$$

as an estimator for μ. Since θ and $\overline{y}_{(k)i}$ are independent it is easy to show that $\overline{\overline{y}}_{(k)}$ is an unbiased estimator for μ with variance

$$V(\overline{\overline{y}}_{(k)}) = \frac{\sigma^2_{(k)}}{k} E\left[\frac{1}{\theta}\right]. \tag{13}$$

Example 4: Let us assume that θ has a uniform distribution on the set {2, 3,…, m}. Then the variance of $V(\overline{\overline{y}}_{(k)})$ is given by

$$V(\overline{\overline{y}}_{(k)}) = \frac{\sigma^2_{(k)}}{k(m-1)} \sum_{j=2}^{m} \frac{1}{j}. \tag{14}$$

Let $\overline{\overline{y}}_{(r)}$ denote the estimator of the RSS method with r replications. Then the variance of $\overline{\overline{y}}_{(r)}$ is given $V(\overline{\overline{y}}_{(r)}) = \frac{\sigma^2_{(k)}}{kr}$. We can compare the variance of $\overline{\overline{y}}_{(k)}$, which is given in equation (14) with the variance of $\overline{\overline{y}}_{(r)}$. We can see that the proposed estimator has smaller variance if $\frac{1}{m-1} \sum_{j=2}^{m} \frac{1}{j} < \frac{1}{r}$.

5. RANDOM NUMBER OF CYCLES WITH RANDOM SET SIZE

Let θ be a random variable independent of $\overline{y}_{\nu_i i}, i \geq 1$ and taking values from Λ. Then we propose

$$\overline{\overline{y}}_{(\theta)} = \frac{1}{\theta} \sum_{i=1}^{\theta} \overline{y}_{\nu_i i}, \tag{15}$$

as an estimator for μ. Since θ and $\overline{y}_{\nu_i i}$ are independent we can show that

$$E\left[\bar{\bar{y}}_{(\theta)}\right] = E\left\{E\left[\frac{1}{\theta}\sum_{i=1}^{\theta}\bar{y}_{\nu_i i}\,|\theta\right]\right\} = E\left\{\frac{1}{\theta}\sum_{i=1}^{\theta}E\left[\bar{y}_{\nu_i i}\right]\right\} = \mu ,$$

i.e. $\bar{\bar{y}}_{(\theta)}$ is an unbiased estimator for μ. To find the variance of $\bar{\bar{y}}_{(\theta)}$ we again use the total probability formula and obtain

$$V(\bar{\bar{y}}_{(\theta)}) = E\left\{E\left[\left[\frac{1}{\theta}\sum_{i=1}^{\theta}(\bar{y}_{\nu_i i}-\mu)\right]^2\middle|\theta\right]\right\} = E\left\{\frac{1}{\theta^2}\sum_{i,\ j=1}^{\theta}E\left[(\bar{y}_{\nu_i i}-\mu)(\bar{y}_{\nu_j j}-\mu)\right]\right\}.$$

Since $E[\bar{y}_{\nu_i i}] = E[\bar{y}_{\nu_j j}] = \mu$, we have

$$V(\bar{\bar{y}}_{(\theta)}) = E\left[\frac{1}{\theta}\right]E\left[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}\right] \tag{16}$$

Example 5: Let θ be a random variable having a Poisson distribution with parameter λ>0 truncated at zero and one, i.e. $P(\theta = m) = P(\xi = m|\xi > 1),\ m \geq 2$, where ξ is a Poisson random variable of the parameter λ. In this case we obtain that

$$E\left[\frac{1}{\theta}\right] = \frac{1}{e^{\lambda}-1-\lambda}T(\lambda), T(\lambda) = \sum_{j=2}^{\infty}\frac{\lambda^j}{j!\,j}.$$

Using the fact that $T'(\lambda) = \frac{1}{\lambda}(e^{\lambda}-1-\lambda)$, we obtain the variance as

$$V(\bar{\bar{y}}_{(\theta)}) = E\left[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}\right]\frac{1}{e^{\lambda}-1-\lambda}\int_0^{\lambda}\frac{e^u-1-u}{u}du .$$

Example 6: Let θ have a uniform distribution on the set {2, 3,…, m}. In this case we have

$$V(\bar{\bar{y}}_{(\theta)}) = \frac{1}{m-1}\sum_{j=2}^{m}\frac{1}{j}E\left[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}\right].$$

If in addition, the random variables $\nu_i, i \geq 1$ also have a common uniform distribution on the set {2, 3, …, N} as in example 3, we obtain

$$V(\bar{\bar{y}}_{(\theta)}) = A_m B_N , \tag{17}$$

where $A_m = \frac{1}{m-1}\sum_{j=2}^{m}\frac{1}{j}$ and $B_N = \frac{1}{N-1}\sum_{\ell=2}^{N}\frac{1}{\ell}\ \sigma_{(\ell)}^2$.

For comparison we consider the RSS unbiased estimator $\overline{\overline{y}}_{(r)}$ with fixed set size k, $2 \le k \le N$ and number of replication r, $2 \le r \le m$. The variance of $\overline{\overline{y}}_{(r)}$ is given by $V(\overline{\overline{y}}_{(r)}) = \frac{\sigma_{(k)}^2}{kr}$. Thus we have to compare A_m with 1/r and B_N with $\frac{\sigma_{(k)}^2}{k}$. In the latter case the comparison is based on Proposition 1. The following proposition is helpful in comparing the RRSS with the usual RSS method

Proposition 2: Let $\varepsilon_r = A_m - r^{-1}, r = 2,3,\dots,m$, then

(a) $\varepsilon_r > 0$ for $r > \frac{m^2+m}{2m+1}$;

(b) $\varepsilon_r < 0$ for $r < \frac{\sqrt{2m^3+2m^2+1}-2m-1}{m-2}$.

To prove the proposition we consider the sum

$$(m-1)\varepsilon_r = \sum_{j=2}^{r}\frac{r-j}{jr} + \sum_{j=r+1}^{m}\frac{r-j}{jr} = I_1 + I_2 .$$

It is not difficult to see that

$$I_1 > \frac{1}{r^2}\sum_{j=1}^{r}(r-j) = \frac{1}{2}, \text{ and } I_2 > \frac{r-m-1}{2r(r+1)} .$$

If we use the opposite inequalities, then we obtain

$$I_1 < \frac{r}{4}, \text{ and } I_2 < \frac{r-m-1}{2mr}(m-r) .$$

Using the bound for I_1 and I_2 in the previous equality we can obtain bounds for r.

The efficiency of the random ranked set sampling (RRSS) with random set size and number of replications with respect to RSS with set size k and r replications may be defined as

$$\tau(k,r) = \frac{\sigma_{(k)}^2/rk - A_m B_N}{\sigma_{(k)}^2/rk} = \frac{\sigma_{(k)}^2 - rkA_m B_N}{\sigma_{(k)}^2} .$$

Evaluation of function $\tau(k, r)$ for different probability distributions will be considered in Section 7.

6. ASYMPTOTIC PROPERTIES

In this section we will prove that under the very natural assumptions the estimator $\overline{\overline{y}}_{(\theta)}$ is asymptotically normal. Recall that $\overline{\overline{y}}_{(\theta)} = \frac{1}{\theta}\sum_{i=1}^{\theta} \overline{y}_{\nu_i i}$ and $\overline{y}_{\nu_i i} = \frac{1}{\nu_i}\sum_{j=1}^{\nu_i} y_j^{(\nu_i)}$. As mentioned before $\overline{y}_{\nu_i i}$, $i \geq 1$ are independent and identically distributed random variables such that $E[\overline{y}_{\nu_i i}] = \mu$, $B^2 \equiv V(\overline{y}_{\nu_i i}) = E[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}]$ and have characteristic function given in equation (9). Since $\sigma^2_{(n)} > \sigma^2_{(n+1)}$, n ≥1 and ν_i are random variables taking values from Λ, we find that $B^2 \leq E\left[\sigma^2_{(\nu_i)}\right] < \sigma^2_{(1)} = \sigma^2$. Thus, if the initial distribution of $X_{ij}^{(\ell)}$ has a finite variance, then $B^2 < \infty$.

Theorem: If $\sigma^2 < \infty$ and $\theta \to \infty$ in probability then

$$P\left(B^{-1}\sqrt{\theta}\left(\overline{\overline{y}}_{(\theta)} - \mu\right) \leq x\right) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2} dt$$

Proof. Let $W_\theta = B^{-1}\sqrt{\theta}\ \ (\overline{\overline{y}}_{(\theta)} - \mu)$, then by total probability formula we have

$$E\left[e^{itw_\theta}\right] = \sum_{\ell=1}^{\infty} E\left\{\exp\left[\frac{it}{B\sqrt{\ell}}\sum_{i=1}^{\ell}(\overline{y}_{\nu_i i} - \mu)\right]\right\} P(\theta = \ell) = E\left\{\left[\overline{\varphi}\left(\frac{t}{B\sqrt{\theta}}\right)\right]^{\theta}\right\},$$

where $\overline{\varphi}(t) = E\left\{\exp\left[it\left(\overline{y}_{\nu_i i} - \mu\right)\right]\right\}$. Since $B^2 < \infty$, we can write following representation for $\overline{\varphi}(t)$:

$$\overline{\varphi}(t) = 1 + itE\left(\overline{y}_{\nu_i i} - \mu\right) - \frac{t^2 B^2}{2} + t^2\varepsilon(t), \tag{18}$$

where $\varepsilon(t) \to 0$ as $t \to 0$. Then if $\theta \to \infty$ in probability, then $\varepsilon(\theta^{-1/2}) \to 0$ in probability. In fact for any δ >0 there is a t_o > 0 such that $|\varepsilon(t)| < \delta$ when $|t| < t_o$. Thus

$$P\left\{\left|\varepsilon\left(1/\sqrt{\theta}\right)\right|>\delta\right\}=P\left\{\left|\varepsilon\left(1/\sqrt{\theta}\right)\right|>\delta,1/\sqrt{\theta}\le t_o\right\}+P\left\{\left|\varepsilon\left(1/\sqrt{\theta}\right)\right|>\delta,1/\sqrt{\theta}>t_o\right\}. \quad (19)$$

It is easy to see that the first probability on the right side of equation (19) is equal to zero and the second is less than $p\left(\sqrt{\theta}<1/t_o\right)$ which tends to zero when $\theta\xrightarrow{p}\infty$. Using equation (18) and the simple formula $\ln(\alpha)=\alpha-1+o(\alpha-1)$, $\alpha\to 1$, we obtain that

$$\theta\ln\overline{\varphi}\left(\frac{t}{B\sqrt{\theta}}\right)=-\frac{t^2}{2}+\gamma(\theta)$$

where $\gamma(\theta)\xrightarrow{P}0$ as $\theta\xrightarrow{p}\infty$. Consequently

$$\left[\overline{\varphi}\left(\frac{t}{B\sqrt{\theta}}\right)\right]^{\theta}\xrightarrow{p}e^{-t^2/2}, \quad (20)$$

as $\theta\xrightarrow{p}\infty$. Since $\left|\overline{\varphi}\left(\frac{t}{B\sqrt{\theta}}\right)\right|\le 1$, by the dominated convergence theorem (see Shiryaev, (1996), Theorem 3, p 187 and remark on p 258). We conclude from (20) that

$$E\left[\overline{\varphi}\left(\frac{t}{B\sqrt{\theta}}\right)\right]^{\theta}\to e^{-t^2/2}$$

i.e. $B^{-1}\sqrt{\theta}\ \ (\overline{\overline{y}}_{\theta}-\mu)$ is asymptotically normal as $\theta\xrightarrow{p}\infty$.

7. EXAMPLES

In this section we will consider comparing the RRSS with the RSS for estimating the population mean μ if the parent distribution is known to be normal, exponential, double exponential or logistic. Also, we are assuming that the set size ν is a uniform random variable defined on the set [2, 3, …, N] and the number of replications θ is following a uniform distribution on the set [2, 3, …, m].

We calculate the value of τ, $2\le\tau\le N$ as suggested by Proposition 1, which will give $V(\overline{y}_{(k)})>V(\overline{y}_{(\nu)})$ for different parent distributions. Table 1 shows the values of τ with the corresponding $V(\overline{y}_{(\nu)})$ as if τ = N with the values of

$V(\bar{y}_{(k)})$ for set size k = 2, 3, 4, 5 for the above probability distributions. It is clear that the RRSS will do better than the RSS with set size k=3, for example if $\tau = 5$ for most of the distributions considered in this study.

Table 1
The value of the random set size τ along the corresponding variance $V(\bar{y}_{(\nu)})$ of RRSS as if N = τ and the RSS variance $V(\bar{y}_{(k)})$ for different set size k and different probability distributions

Distribution		k			
		2	**3**	**4**	**5**
Normal	$V(\bar{y}_{(k)})$	0.3408	0.1742	0.1065	0.0722
	τ	3	5	9	14
	$V(\bar{y}_{(\nu)})$	0.2575	0.1734	0.1053	0.0708
Exponential	$V(\bar{y}_{(k)})$	0.3750	0.2037	0.1303	0.0913
	τ	3	5	9	13
	$V(\bar{y}_{(\nu)})$	0.2894	0.2001	0.1248	0.0911
Logistic	$V(\bar{y}_{(k)})$	0.3480	0.1814	0.1128	0.0776
	τ	3	4	9	14
	$V(\bar{y}_{(\nu)})$	0.2647	0.2141	0.1104	0.0748
Double Exponential	$V(\bar{y}_{(k)})$	0.7368	0.3854	0.2453	0.1719
	τ	3	5	9	14
	$V(\bar{y}_{(\nu)})$	0.5611	0.3848	0.2385	0.1630

The value of the efficiency $\tau(k,\ r)$ of RRSS with respect to RSS is evaluated for k = 3, 5, r =3, 5, N =10, 15 and m = 10, 20. Table 2 shows different values $\tau(k,\ r)$ for the normal, exponential, double exponential and logistic distributions. We observe that the RRSS improves the efficiency of estimating the population mean if the values of N and m are moderately large. For example, if N = m =10 and r = k = 3, the RRSS is about 66% more efficient than the RSS.

Table 2
The efficiency of the RRSS $\tau(k, r)$ with respect to RSS for different probability distributions

N	m	k			
		3		5	
		r			
		3	5	3	5
Normal					
10	10	0.655	0.410	0.146	- 0.423
	20	0.774	0.624	0.455	0.092
15	10	0.755	0.591	0.408	0.014
	20	0.844	0.739	0.623	0.371
Exponential					
10	10	0.640	0.399	0.196	- 0.340
	20	0.770	0.618	0.487	0.145
15	10	0.746	0.577	0.435	0.058
	20	0.838	0.730	0.639	0.399
Logistic					
10	10	0.643	0.404	0.165	- 0.392
	20	0.772	0.620	0.467	0.112
15	10	0.751	0.585	0.418	0.029
	20	0.841	0.735	0.628	0.381
Double Exponential					
10	10	0.636	0.394	0.184	- 0.361
	20	0.768	0.613	0.479	0.132
15	10	0.744	0.574	0.426	0.044
	20	0.837	0.728	0.634	0.390

8. CONCLUDING COMMENTS

In this paper we have considered the case of random set size and/or random number of replications. The reason for considering such a method is to resolve the problem of unfixed number of units that we might come cross in real life problems like the line intercept (transect) example given in Section 1. It has been shown that under certain conditions the efficiency of the estimator of the population mean may be improved by using RRSS instead of RSS. The following conclusions are drawn:

1. In the case of single cycle with random set size we might be able to improve the efficiency of the estimator of the population mean by using RRSS instead of RSS by choosing the suitable distribution for the set size. The result of Table 1 confirms this fact in the case of choosing the discrete uniform distribution.
2. If the set size is fixed and the number of replications is random we can easily show that the RRSS is more efficient than RSS, if the number of replications are following the discrete uniform distribution.
3. The results of Table 2 show that in the case of random set size ν and number of cycles θ, the efficiency is substantially increased if the underlying for both ν and θ is discrete uniform.

The final recommendation is to use RRSS to handle some practical problems and/or to increase the efficiency of the estimator of the population mean.

REFERENCES

1. Bohn, L.L. and Wolfe, D.A. (1992). Nonparametric Two-Sample Procedures for Ranked Set Samples Data, *J. Amer. Statist. Assoc.*, 87, 552-561.
2. ------- (1994). The Effect of Imperfect Judgment Rankings on Properties of Procedures Based on Ranked-Set Samples analog of the Mann-Whitney-Wilcoxon Statistic, *J. Amer. Statist. Assoc.*, 89, 168-176.
3. Hartlaub, B.A. and Wolfe, D.A. (1999). Distribution-Free Ranked Set Sample Procedures for Umbrella Alternatives in the m-Sample Setting, *Environmental and Ecological Statistics*, 6, 105-118.
4. Huang, J. (1997). Asymptotic Properties of the NPMLE of a Distribution Function Based on Ranked Set Samples, *The Annals of Statistics*, 25(3), 1036-1049.
5. Kaur, A., Patil, G.P., Sinha, B.K. and Taillie, C. (1995). Ranked Set Sampling: An Annotated Bibliography, *Environmental and Ecological Statistics*, 2, 25-54.
6. Kim, Y.H. and Arnold, B.C. (1999a). Estimation of a Distribution Function under Generalized Ranked Set Sampling, *Commun. Statist. – Simula,* 28(3), 657-666.
7. ------ (1999b). Parameter Estimation under Generalized Ranked Set Sampling, *Statistics & Probability Letters*, 42, 353-360.
8. Koti, K.M. and Babu, G.J. (1996). Sign Test for Ranked-Set Sampling, *Commun. Statist. Theory Meth.* 25(7), 1617-1630.
9. Kvam, P.H. and Samaniego, F.J. (1993). On the Inadmissibility of Empirical Averages as Estimators in Ranked Set Sampling, *J. Statist. Plann. and Infer.*, 36, 39-55.
10. ------- (1994). Nonparametric Maximum Likelihood Estimation based on Ranked Set Samples, *J. Amer. Statist. Assoc.*, 89, 526-537.

11. Li, D. Sinha, B.K. and Perron, F. (1999). Random Selection in Ranked Set Sampling and its Applications, *J. Statist. Plann. and Infer.*, 76, 185-201.
12. McIntyre, G.A. (1952). A Method of Unbiased Selective Sampling, using Ranked Sets, *Aust. J. Agri. Res.*, 3, 385-390.
13. Muttlak, H.A. and McDonald, L.L. (1992). Ranked Set Sampling and the Line Intercept Method: A more efficient procedure, *The Biometrical Journal*, 34, 329-46.
14. Muttlak, H.A. (1997). Median Ranked Set Sampling, J. *App. Statist. Sci.*, 6(4), 245-55.
15. Özturk, Ö. (2000). Alternative Ranked Set Sampling Protocols for Sign Test, *Statistics & Probability Letters*, 47, 15-23.
16. Patil, G.P., Sinha, A.K. and Taillie, C. (1994). Ranked Set Sampling, *A Handbook of Statistics*, 12, 167-200.
17. ----- (1999). Ranked Set Sampling: A Bibliography, *Environmental and Ecological Statistics*, 6, 91-98.
18. Presnell, B. and Bohn, L.L. (1999). U-Statistics and Imperfect Ranking in Ranked Set Sampling, *Nonparametric Statistics,* 10, 111-126.
19. Samawi, H.M., Ahmed, M.S. and Abu-Dayyeh (1996). Estimating the Population Mean using Extreme Raked Set Sampling, *The Biometrical Journal*, 38(5), 577-86.
20. Shiryaev, A.N. (1996). *Probability, 2end*: Springer, New York.
21. Stokes, S.L. and Sager, T.W. (1988). Characterization of a Ranked-Set Sample with Application to Estimating Distribution Function, *J. Amer. Statist. Assoc.*, 83, 374-381.
22. Takahasi K. and Wakimoto K. (1968). On Unbiased Estimates of the Population Mean based on the Sample Stratified by Means of Ordering, *Ann. Inst. of Statist. Math.,* 21, 249-55.

CHAPTER SIX

Extreme ranked set sampling: A comparison with regression and ranked set sampling estimators

Hassen A. Muttlak
Department of Mathematical Sciences,
King Fahd University of Petroleum & Minerals,
Dhahran 31261, Saudi Arabia

ABSTRACT

Ranked set sampling (RSS), as suggested by McIntyre (1952), assumed perfect ranking i.e. there will be no errors in ranking the units with respect to the variable of interest. In fact, for most practical applications, it is not easy to rank the units without errors in ranking. As pointed out by Dell and Clutter (1972) there will be a loss in efficiency, i.e. RSS will give a larger variance due to the errors in ranking the units. To reduce the errors in ranking in estimating the population mean, the extreme ranked set sampling (ERSS) procedure is used and compared to its counterpart RSS and regression estimators. The regression estimator uses an auxiliary variable to estimate the population mean. Turns out that the ERSS estimator is more efficient than the regression estimator for most cases considered in this study unless the correlation between the variable of interest and the auxiliary is more than 0.80. In addition, the ERSS and RSS estimators are comparable if the ranking of the variable of interest is done using a concomitant variable. ERSS is used to estimate the population mean of a variable of interest when the ranking of this variable is acquired through a concomitant variable. Finally, ERSS is used to estimate the population mean of the variable of interest in the presence of errors in ranking and compared with RSS and simple random sampling (SRS) estimators.

KEYWORDS

Auxiliary variable; concomitant variable; efficiency; errors in ranking; simple random sampling.

1. INTRODUCTION

Ranked set sampling (RSS) was suggested by McIntyre (1952) without the mathematical theory to support his suggestion. Takahasi and Wakimoto (1968) supplied the mathematical theory. They proved that the sample mean of the ranked set sample (RSS) is an unbiased estimator of the population mean. In addition, it is more efficient than the sample mean of a simple random sample (SRS) with the same sample size.

Dell and Clutter (1972) studied the case in which the ranking may not be perfect i.e. there are errors in ranking the units with respect to the variable of interest. They showed that the mean of the RSS is an unbiased estimator of the population mean, whether or not there are errors in ranking, but there will be a loss in efficiency due to these errors.

Stokes (1977) considered the case when the variable of interest X is difficult to measure and order, but there is a concomitant variable Y which is correlated with X that can be used to judge the order of the variable Y.

Patil, Sinha, and Taillie (1993) showed the estimator of the population mean using RSS is considerably more efficient than the SRS regression estimator unless the correlation between the variable of interest and the concomitant variable is more than 0.85.

Samawi, Abu-Dayyeh and Ahmed (1996) suggested using extreme ranked set sampling (ERSS) to estimate the population mean of the variable of interest. They showed that the ERSS estimator is an unbiased estimator of the population mean if the underlying distribution is symmetric and that it is more efficient than the SRS estimator.

In this paper, the performance of the ERSS procedure for estimating the population mean is compared to the regression and the RSS estimators. If the variable of interest and the auxiliary variable follow a bivariate normal distribution, it has been found that the ERSS estimator is more efficient than the regression estimator unless the correlation between the two variables $|\rho| > 0.8$. In addition, if the units are ranked using a concomitant variable and the cycle is repeated once the RSS and the ERSS estimators are more efficient than the regression estimator provided that $|\rho| < 0.80$. Nevertheless, if the cycle is repeated more than once and $|\rho| < 0.80$, the three methods are comparable. ERSS is used to estimate the population mean of a variable of interest when it is difficult to measure or rank, but there is a concomitant variable available which can be used to rank the variable of interest. We assumed that the variable of interest and the auxiliary follow a bivariate normal distribution since the regression and ERSS estimators are unbiased under bivariate normality and to simplify the calculations. Computer simulation results are given to compare the efficiency of ERSS estimator of the population mean with its counterparts SRS and RSS in the presence of errors in raking the variable of interest for some probability distributions.

2. SAMPLING METHODS

2.1 Ranked set sampling

The ranked set sampling (RSS) method can be summarized as follows: Select n random samples of size n units and rank the units within each sample with respect to a variable of interest by a visual inspection. Then select for actual measurement the smallest unit from the first sample. From the second sample, select for actual measurement the second smallest unit. The procedure is continued until the largest from the n^{th} sample is selected for measurement. In this way, we obtain a total of n measured units, one from each sample. The cycle may be repeated r times until nr units have been measured. These nr units form the RSS data.

Let $X_1, X_2, \ldots, X_n$ be a random sample with probability density function f (x) with a finite mean μ and variance σ^2. Let $X_{11}, X_{12}, \ldots, X_{1n}; X_{21}, X_{22}, \ldots, X_{2n}; \ldots; X_{n1}, X_{n2}, \ldots, X_{nn}$ be independent random variables all with the same cumulative distribution function F(x). Let $X_{(i:n)}$ denote the i^{th} order statistic from the i^{th} sample of size n (i =1, 2…, n) and let $X_{(i:n)j}$ denote the i^{th} order statistic from the i^{th} sample of size n in the j^{th} cycle (j = 1, 2, …, r). The unbiased estimator of the population mean is defined as

$$\overline{X}_{rss} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{(i:n)j} \,.$$

The variance of $\overline{X}rss$ is given by

$$\operatorname{var}(\overline{X}_{rss}) = \frac{1}{n^2 r}\sum_{i=1}^{n}\sigma^2_{(i:n)} \,,$$

where $\sigma^2_{(i:n)} = E\ [X_{(i:n)} - E(X_{(i:n)})]^2$.

2.2 Ranked set sampling with concomitant variable

Suppose that the variable of interest X is difficult to measure and to order, but there is available a concomitant variable Y, which is correlated with X. We may use the variable Y to estimate the rank of X variable as follows: Select n samples of size n bivariate units from the population and rank each sample with respect to the variable Y by visual inspection. Select for actual measurement, from the first sample of size n the X associated with the smallest Y. From the second sample of size n, select for actual measurement the X associated with the second smallest Y. We continue this way until the X associated with the largest Y from the n^{th} sample is selected for actual measurement. The cycle may be repeated r times until nr X's are selected for actual measurement. Note that ranking of the variable X will be with errors in ranking i.e. $X_{[i:n]j}$ is the i^{th} judgment order statistic from the i^{th} sample of size n in the j^{th} cycle.

Assume that (X, Y) has a bivariate normal and the regression of X on Y is linear. Following Stokes (1977), we can write

$$X_{[i:n]j} = \mu_x + \frac{\rho\sigma_x}{\sigma_y}\left(Y_{(i:n)j} - \mu_y\right) + \varepsilon_{ij}, \tag{1}$$

where $Y_{(i:\,n)\,j}$ is the i^{th} order statistic of the i^{th} sample of size n units assuming the ranking of the units with respect to the variable Y is perfect in the j^{th} cycle and ε_{ij} is the error term with mean equal to zero. The mean of the variable of interest X with ranking based on the concomitant variable Y can be written as

$$\overline{X}_{rssc} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{[i:n]j}\ . \tag{2}$$

The variance of $\overline{X}rssc$ is given by

$$\operatorname{var}\left(\overline{X}_{rssc}\right) = \frac{1}{n^2 r}\left[n\,\sigma_x^2\left(1-\rho^2\right) + \frac{\rho^2\sigma_x^2}{\sigma_y^2}\sum_{i=1}^{n}\sigma_{y(i:n)}^2\right]. \tag{3}$$

2.3 Extreme ranked set sampling

RSS as suggested by McIntyre (1952) and Takahasi and Wakimoto (1968) can be modified to introduce a new sampling method called extreme ranked set sampling **(**ERSS). In the ERSS procedure, select n random samples of size n units from the population and rank the units within each sample with respect to a variable of interest by visual inspection. If the sample size n is even, select from n/2 samples the smallest unit and from the other n/2 samples the largest unit for actual measurement. If the sample size is odd, select from (n-1)/2 samples the smallest unit, from the other (n-1)/2 the largest unit and from one sample the median of the sample for actual measurement. The cycle may be repeated r times to get nr units. These nr units form the ERSS data.

We can note that the ERSS can be performed in practical applications with less errors in ranking the units since all we have to do is to find the largest or the smallest of the sample and measure it, comparing to its counterpart RSS. The ERSS method is very easy to apply in the field and will save time in performing the ranking of the units with respect to the variable of interest. In addition, this method will reduce the errors in ranking and hence increase the efficiency of the ERSS method.

Let $X_{i\,(1)}$, $X_{i\,(2)}$, …, $X_{i\,(n)}$ be the order statistics of X_{i1}, X_{i2},..., X_{in} (i = 1, 2, ..., n). If the cycle is repeated once, let $X_{i\,(1)}$ and $X_{i\,(n)}$ be the smallest and the largest of the i^{th} sample respectively (i = 1, 2, ..., n). The estimator of the population mean using ERSS with r cycles can be defined in the case of an even sample size as

$$\overline{X}_{erss1} = \frac{1}{nr}\left(\sum_{i=1}^{L_1}\sum_{j=1}^{r} X_{i(1)j} + \sum_{i=L_1+1}^{n}\sum_{j=1}^{r} X_{i(n)j}\right),$$

where $L_1 = n/2$. In the case of an odd sample size, the estimator of the population mean can be defined as

$$\overline{X}_{erss2} = \frac{1}{nr}\left(\sum_{i=1}^{L_2}\sum_{j=1}^{r} X_{i(1)j} + \sum_{i=L_2+2}^{n}\sum_{j=1}^{r} X_{i(n)j} + \sum_{j=1}^{r} X_{i((n+1)/2)j}\right),$$

where $L_2 = (n-1)/2$ and $X_{i\,((n+1)/2)}$ is the median of sample i = (n+1)/2 if the sample size is odd. To simplify the notation, let $X_{(i:\,e),\,j}$ denote the smallest of the i^{th} sample (i = 1, 2,…, L_1) and the largest of the i^{th} sample (i = L_1+1, L_1+2, …, n) in the j^{th} cycle (j = 1, 2, …, r) if the sample size n is even. Also denote the smallest of the i^{th} sample (i = 1, 2, …, L_2), the median of the i^{th} sample (i = (n+1)/2) and the largest of the i^{th} sample (i = L_2+2, L_2+3, …, n) in the j^{th} cycle if the sample size n is odd. The estimator of the population mean then can be written as

$$\overline{X}_{erss} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{(i:e)j}\,.$$

The variance of $\overline{X}_{erss}$ can be written as

$$\operatorname{var}\left(\overline{X}_{erss}\right) = \frac{1}{n^2 r}\sum_{i=1}^{n}\sigma^2_{(i:e)}\ ,$$

where $\sigma^2_{(i:n)} = E\left[X_{(i:e)} - E\left(X_{(i:e)}\right)\right]^2$

It can be shown that $\overline{X}_{erss}$ is an unbiased estimator of the population mean if the underline distribution is symmetric. In addition, it is more efficient than the sample mean $\left(\overline{X}_{srs}\right)$ of simple random sample (SRS) with the same sample size. The analogues of equation (1), (2) and (3) are

$$X_{[i:e]j} = \mu_x + \frac{\rho\sigma_x}{\sigma_y}\left(Y_{(i:e)j} - \mu_y\right) + \varepsilon_{ij}\,,$$

$$\overline{X}_{erssc} = \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} X_{[i:e],j}$$

and

$$\text{var}\left(\overline{X}_{erssc}\right) = \frac{1}{n^2 r}\left[n\sigma_x^2 \left(1-\rho^2\right) + \frac{\rho^2 \sigma_x^2}{\sigma_y^2} \sum_{i=1}^{n} \sigma_{y(i:e)}^2 \right].$$

Finally, the ERSS estimator for the population mean will not necessarily be unbiased if the judgment ranking is not perfect.

3. COMPARISON OF THE ERSS ESTIMATOR WITH THE REGRESSION AND RSS ESTIMATORS

Suppose that (X, Y) follow a bivariate normal distribution and if the mean of an auxiliary variable Y is known. The linear regression of the variable X on the variable Y is defined

$$X_i = \beta_0 + \beta_1 Y_i + \varepsilon_i, \; i = 1, \; 2, \; ..., \; nr,$$

where β_0 and β_1 are the intercept and the slope of the regression line respectively and ε_i is the random error component with expected value of zero. The linear regression estimator of the population mean (μ_x) of the variable of interest X is

$$\overline{X}_{lr} = \overline{X} + \hat{\beta}_1 (\mu_y - \overline{Y}),$$

where $\overline{X}$ and $\overline{Y}$ are the sample means of the variable of interest X and the auxiliary variable Y respectively, based on sample size nr. In addition, μ_y is the population mean of the auxiliary Y and $\hat{\beta}_1$ is the least square estimator of β_1.

In most applications, the population mean of the auxiliary variable is unknown and we usually estimate it using the method of double sampling. To estimate the population mean μ_y of the auxiliary variable Y, we need to select a large random sample of size n^2r (say). In addition, we need a random sample of size nr units to study the variable of interest X. The regression estimator of the population mean (μ_x) of the variable of interest X can be defined using the double sampling method to estimate the mean (μ_y) of the auxiliary Y as

$$\overline{X}_{lrd} = \overline{X} + \hat{\beta}_1 \left(\overline{Y}_d - \overline{Y}\right),$$

where $\overline{Y}_d$ is an unbiased estimator of the population mean of the auxiliary variable Y (μ_y). Sukhatme and Sukhatme (1970, p. 221) showed that $\overline{X}_{lrd}$ is an unbiased estimator of the population mean (μ_x) of the variable of interest X with variance

$$\operatorname{var}\left(\overline{X}_{lrd}\right) = \sigma_x^2\left(1-\rho^2\right)\left(\frac{1}{nr}-\frac{1}{n^2r}\right)\left(1+\frac{1}{nr-3}\right)+\frac{\sigma_x^2}{n^2r},$$

if (X, Y) follows a bivariate normal distribution.

Let us assume that the ranking on the variable of interest is perfect. The relative efficiency for $\overline{X}_{rss}$ with respect to $\overline{X}_{lrd}$ can be found as

$$eff_1 = eff\left(\overline{X}_{lrd}, \overline{X}_{rss}\right) = \frac{\operatorname{var}(\overline{X}_{lrd})}{\operatorname{var}(\overline{X}_{rss})} = \frac{\left(1-\rho^2\right)\left(\frac{n-1}{n}\right)\left(1+\frac{1}{nr-3}\right)+\frac{1}{n}}{\frac{1}{n}\sum_{i=1}^{n}\sigma_{z(i:n)}^2}.$$

The relative efficiency for $\overline{X}_{errs}$ with respect to $\overline{X}_{lrd}$ can easily be shown to be

$$eff_2 = eff\left(\overline{X}_{lrd}, \overline{X}_{erss}\right) = \frac{\operatorname{var}(\overline{X}_{lrd})}{\operatorname{var}(\overline{X}_{erss})} = \frac{\left(1-\rho^2\right)\left(\frac{n-1}{n}\right)\left(1+\frac{1}{nr-3}\right)+\frac{1}{n}}{\frac{1}{n}\sum_{i=1}^{n}\sigma_{z(i:e)}^2}.$$

The values of eff_1 and eff_2 are summarized in Table 1 for both RSS and ERSS. Calculations were done using sample sizes of n = 4, 5, 6 and 7, cycles r = 1, 2, 4, 5, ∞ and with values of $\rho = \pm 0.25,\ \pm 0.5,\ \pm 0.75,\ \pm 0.90$.

Table 1
Summary of the relative efficiency values, for estimating the population mean using RSS and ERSS methods with respect to the regression estimator if the population mean of the auxiliary is unknown

	RSS					**ERSS**				
	1	**2**	**3**	**5**	**∞**	**1**	**2**	**3**	**5**	**∞**
	eff$_1$					**eff$_2$**				
	ρ = ± 0.25									
4	3.89	2.57	2.42	2.34	2.23	3.37	2.22	2.09	2.02	1.94
5	3.37	2.93	2.80	2.72	2.63	3.19	2.55	2.44	2.37	2.29
6	3.68	3.30	3.19	3.11	3.02	2.91	2.49	2.40	2.33	2.28
7	4.12	3.65	3.56	3.49	3.40	3.13	2.79	2.71	2.65	2.59
	ρ = ± 0.50									
4	3.23	2.17	2.05	1.98	1.91	2.79	1.88	1.78	1.72	1.65
5	3.05	2.45	2.35	2.29	2.22	2.65	2.13	2.05	1.99	1.93
6	3.19	2.74	2.65	2.60	2.52	2.40	2.07	2.00	1.95	1.90
7	3.40	3.03	2.95	2.89	2.82	2.59	2.31	2.24	2.20	2.15

Table 1 (continued)

	RSS					**ERSS**				
	1	**2**	**3**	**5**	∞	**1**	**2**	**3**	**5**	∞
			eff_1					**eff_2**		
					$\rho = \pm 0.75$					
4	2.13	1.51	1.44	1.40	1.36	1.84	1.31	1.25	1.21	1.18
5	2.01	1.66	1.60	1.57	1.52	1.75	1.44	1.39	1.36	1.32
6	2.08	1.82	1.77	1.74	1.69	1.57	1.38	1.34	1.31	1.28
7	2.20	1.98	1.94	1.90	1.86	1.67	1.51	1.47	1.45	1.41
					$\rho = \pm 0.90$					
4	1.26	0.99	0.96	0.94	0.93	1.09	0.86	0.83	0.81	0.80
5	1.19	1.04	1.01	0.99	0.98	1.03	0.90	0.88	0.86	0.85
6	1.20	1.09	1.07	1.05	1.04	0.91	0.82	0.81	0.79	0.78
7	1.24	1.15	1.13	1.12	1.10	0.95	0.88	0.86	0.85	0.84

Considering Table 1 we can see that a gain in efficiency is obtained by using ERSS instead of the regression estimator for different values of n and r. For example, for n = 6, r =1 and $\rho = \pm 0.25$, the relative efficiency is 2.91.

Finally, if the ranking of the variable of interest X is done using the concomitant variable Y, the relative efficiency of $\overline{X}_{rssc}$ with respect to $\overline{X}_{lrd}$ can be found as

$$eff_3 = \frac{\operatorname{var}(\overline{X}_{lrd})}{\operatorname{var}(\overline{X}_{rssc})} = \frac{\left(1-\rho^2\right)\left(\frac{n-1}{n}\right)\left(1+\frac{1}{nr-3}\right)+\frac{1}{n}}{\left(1-\rho^2\right)+\frac{\rho^2}{n}\sum_{i=1}^{n}\sigma^2_{z(i:n)}}.$$

The relative efficiency of $\overline{X}_{errsc}$ with respect to $\overline{X}_{lrd}$ can be found as

$$eff_4 = \frac{\operatorname{var}(\overline{X}_{lrd})}{\operatorname{var}(\overline{X}_{erssc})} = \frac{\left(1-\rho^2\right)\left(\frac{n-1}{n}\right)\left(1+\frac{1}{nr-3}\right)+\frac{1}{n}}{\left(1-\rho^2\right)+\frac{\rho^2}{n}\sum_{i=1}^{n}\sigma^2_{z(i:e)}}.$$

Results of eff_3 and eff_4 are summarised in Table 2 for both ERSS and RSS. Again, calculations were done using sample sizes of n = 4, 5, 6 and 7, cycles r = 1, 2, 3, 4, 5, ∞ and with values of $\rho = \pm 0.25,\ \pm 0.5,\ \pm 0.75,\ \pm 0.80$. A gain in efficiency is obtained by using ERSS with ranking done using a concomitant variable for different values of n, r = 1, and for $|\rho| < 0.80$. For example, if n = 6, r =1 and $\rho = \pm 0.25$, the relative efficiency is 1.25.

Table 2
Summary of the relative efficiency values, for estimating the population mean using RSS and ERSS methods with respect to the regression estimator if the population mean of the auxiliary is unknown.

	RSS					ERSS				
	1	**2**	**3**	**5**	∞	**1**	**2**	**3**	**5**	∞
	eff$_1$					**eff$_2$**				
$\rho = \pm 0.25$										
4	1.72	1.13	1.07	1.03	0.99	1.71	1.13	1.06	1.03	0.98
5	1.38	1.10	1.05	1.03	0.99	1.37	1.10	1.05	1.03	0.98
6	1.26	1.08	1.04	1.02	0.99	1.25	1.07	1.04	1.01	0.98
7	1.20	1.07	1.04	1.02	0.99	1.19	1.06	1.03	1.02	0.98
$\rho = \pm 0.50$										
4	1.61	1.08	1.02	0.99	0.95	1.58	1.06	1.01	0.97	0.93
5	1.31	1.05	1.01	0.98	0.95	1.29	1.04	1.02	0.98	0.94
6	1.21	1.05	1.01	0.98	0.96	1.17	1.02	0.99	0.95	0.93
7	1.15	1.03	1.00	0.98	0.96	1.12	1.00	0.98	0.96	0.93
$\rho = \pm 0.75$										
4	1.34	0.95	0.91	0.88	0.85	1.27	0.90	0.86	0.84	0.81
5	1.13	0.94	0.90	0.88	0.86	1.08	0.89	0.86	0.84	0.82
6	1.06	0.93	0.91	0.89	0.87	0.97	0.85	0.83	0.81	0.79
7	1.03	0.93	0.91	0.89	0.87	0.95	0.86	0.84	0.82	0.80
$\rho = \pm 0.80$										
4	1.25	0.91	0.87	0.85	0.83	1.17	0.85	0.82	0.79	0.77
5	1.07	0.90	0.87	0.85	0.83	1.01	0.84	0.82	0.80	0.78
6	1.01	0.89	0.87	0.85	0.83	0.91	0.80	0.78	0.76	0.75
7	0.98	0.89	0.87	0.85	0.83	0.89	0.81	0.79	0.78	0.76

4. ERSS ESTIMATOR WITH CONCOMITANT VARIABLES

To compare the two estimators $\overline{X}_{rssc}$ and $\overline{X}_{erssc}$ with respect to $\overline{X}_{srs}$, it is assumed that the two variables X and Y follow the bivariate normal distribution. The benefit of using the concomitant variables will depend upon the correlation between the variable of interest X and the concomitant variable Y. If X and Y are independent the estimators $\overline{X}_{rssc}$ and $\overline{X}_{erssc}$ will have the same variance as $\overline{X}_{srs}$. The relative efficiency of $\overline{X}_{rssc}$ with respect to $\overline{X}_{srs}$ can be defined as

$$eff_{rssc} = \frac{\text{var}(\overline{X}_{srs})}{\text{var}(\overline{X}_{rssc})} = \frac{\sigma_x^2 / nr}{\frac{1}{n^2 r}\left[n\sigma_x^2\left(1-\rho^2\right) + \frac{\rho^2\sigma_x^2}{\sigma_y^2}\sum_{i=1}^{n}\sigma_{y(i:n)}^2 \right]}.$$

This can be simplified to

$$eff_{rssc} = \frac{1}{(1-\rho^2) + \frac{\rho^2}{n}\sum_{i=1}^{n}\sigma_{z(i:n)}^2}.$$

The relative efficiency of $\overline{X}_{erssc}$ with respect to $\overline{X}_{srs}$ can be found as

$$eff_{erssc} = \frac{\text{var}(\overline{X}_{srs})}{\text{var}(\overline{X}_{erssc})} = \frac{1}{\left(1-\rho^2\right) + \frac{\rho^2}{n}\sum_{i=1}^{n}\sigma_{z(i:e)}^2}.$$

Results of eff_{rssc} and eff_{errsc} are summarized in Table 3 for both RSS and ERSS. Calculations were done with sample sizes n = 3, 4, 5, 6 and 7 and with values of $\rho = \pm 0.25, \pm 0.5, \pm 0.75, \pm 0.90$.

Table 3
Summary of the relative efficiency values, for estimating the population mean using RSS and ERSS methods with concomitant variables.

n	ρ = ± 0.25		ρ = ± 0.50		ρ = ± 0.75		ρ = ± 0.90	
	eff_{rssc}	eff_{erssc}	eff_{rssc}	eff_{erssc}	eff_{rssc}	eff_{erssc}	eff_{rssc}	eff_{erssc}
3	1.025	1.025	1.109	1.109	1.287	1.287	1.631	1.631
4	1.037	1.033	1.168	1.146	1.477	1.400	1.873	1.699
5	1.042	1.038	1.190	1.171	1.561	1.489	2.073	1.899
6	1.045	1.038	1.207	1.171	1.628	1.489	2.251	1.897
7	1.047	1.041	1.220	1.188	1.684	1.554	2.410	2.056

A gain in efficiency is obtained by using ERSS with a concomitant variable to estimate the population mean for different values of n. For example, if n = 6 and ρ = ± 0.90, the relative efficiency of the ERSS estimator is 1.897.

5. ERSS WITH ERRORS IN RANKING

Ranked set sampling with errors in ranking (RSSE) is considered by Dell and Clutter (1972); that is the quantified observation from the ith sample in the jth cycle may not be the i^th^ order statistic but rather the i^th^ "judgment order statistic". Let $X_{[i:n]j}$ be the i^th^ judgment order statistic from the i^th^ sample of size n in the j^th^ cycle (to distinguish it from the actual order statistic $X_{(i:n)j}$). Assuming the cycle is repeated once, the unbiased estimator of the population mean using RSSE is defined as follows

$$\tilde{X}_{rsse} = \frac{1}{n}\sum_{i=1}^{n} X_{[i:n]} ,$$

Dell and Clutter (1972) showed that $\tilde{X}_{rsse}$ is an unbiased estimator of the population mean with smaller variance than $\overline{X}_{srs}$ the sample mean of SRS with the same size.

Let

$$\tilde{X}_{ersse} = \frac{1}{n}\sum_{i=1}^{n} X_{[i:e]} ,$$

be an estimator of the population mean based on ERSS with errors in ranking. The properties of $\tilde{X}_{ersse}$ are

(a) $\tilde{X}_{ersse}$ is unbiased estimator of the population mean if the distribution is symmetric about the population mean μ and the error in ranking following a normal with mean 0 and variance σ_e^2 and

(b) $Var\left(\tilde{X}_{ersse}\right)$ is less than $Var\left(\tilde{X}_{srs}\right)$.

It is easy to prove (i) and (ii) using results by and Takahasi and Wakimoto (1968), Dell and Clutter (1972) and Samawi et al. (1996).

The ERSS and RSS with errors in ranking were simulated in computer trials. Four probability distributions were considered for the population: normal, gamma, uniform, and weibull. Five Thousand random numbers were generated. For each computer simulation trials were run with n = 4, 6 and 8. The model for these simulations was the same as the model considered by Dell and Clutter (1972); the elements are ranked on the basis of elements that are equal to true values generated from the above probability distributions plus random error components assumed to be distributed normally with mean 0 and variance σ_e^2.

A sample of size n elements was generated from a normal distribution with mean 0 and standard deviation σ_e, and added to the n elements generated from the parent distribution. When this had been accomplished the ERSS and RSS methods were used to rank the units. Estimates for the mean, variance and mean square errors (MSE) were found for the ERSS and RSS data after 5,000 elements had been compute.

The efficiency of estimating the population mean using ERSS with errors in ranking with respect to SRS is defined as follows

$$eff_{erss} = eff\left(\tilde{X}_{srs},\ \tilde{X}_{ersse}\right) = \frac{Var\left(\tilde{X}_{srs}\right)}{Var\left(\tilde{X}_{ersse}\right)},$$

if the distribution is symmetric, otherwise i.e. if the distribution is not symmetric

$$eff_{erss} = eff\left(\tilde{X}_{srs},\ \tilde{X}_{ersse}\right) = \frac{Var\left(\tilde{X}_{srs}\right)}{MSE\left(\tilde{X}_{ersse}\right)}.$$

The efficiency of RSS with errors in ranking with respect to SRS is defined as

$$eff_{rss} = eff\left(\tilde{X}_{srs},\ \tilde{X}_{rsse}\right) = \frac{Var\left(\tilde{X}_{srs}\right)}{Var\left(\tilde{X}_{rsse}\right)},$$

were $\tilde{X}_{rsse}$ is the sample of RSS with errors in ranking data with only one cycle.

Results of these simulations are summarised in Table 4 for both ERSS and RSS. For each population simulations were run with sample of sizes n = 4, 6 and 8 and with values of $\sigma_e = 0.0$, 0.25, 0.5, 1 and 1.5.

Table 4
Summary of efficiency values from computer simulation for four distributions, for estimating the mean of the population using ERSS and RSS with errors in ranking

σ_e	ERSS			RSS		
	Sample Size			Sample Size		
	4	6	8	4	6	8
Uniform (0, 4)						
0.00	3.13	5.45	8.44	2.50	3.50	4.50
0.25	2.98	5.27	7.67	2.40	3.20	3.65
0.50	1.90	2.34	2.89	1.53	1.66	1.78
0.75	1.23	1.31	1.37	1.17	1.18	1.19
1.00	1.06	1.08	1.09	1.02	1.06	1.04
1.50	1.00	1.02	1.03	1.00	1.01	1.02
Normal (2, 1)						
0.00	2.10	2.34	2.66	2.35	3.18	3.98
0.25	2.01	2.32	2.52	2.27	3.06	3.87
0.50	1.98	2.26	2.38	2.12	2.79	3.33
0.75	1.68	1.73	1.92	1.75	2.04	2.23
1.00	1.40	1.47	1.51	1.42	1.46	1.57
1.50	1.12	1.18	1.21	1.14	1.08	1.13
Gamma (3)						
0.00	1.62	1.30	0.88	2.19	3.08	3.42
0.25	1.49	1.22	0.86	2.12	2.82	3.50
0.50	1.55	1.20	0.82	2.09	2.80	3.33
0.75	1.41	1.18	0.77	1.88	2.58	2.77
1.00	1.26	1.03	0.75	1.62	1.98	2.15
1.50	1.07	0.99	0.71	1.25	1.31	1.35
Weibull (2, 3)						
0.00	2.05	2.17	1.82	2.38	3.31	3.93
0.25	2.04	2.15	1.80	2.27	3.17	3.81
0.50	1.92	2.00	1.77	2.20	2.82	3.56
0.75	1.72	1.76	1.49	1.93	2.40	2.67
1.00	1.46	1.47	1.28	1.57	1.72	1.83
1.50	1.15	1.22	1.16	1.16	1.25	1.22

6. RESULTS AND DISCUSSION

In this paper, the bivariate normal distribution is considered, because it is the most widely use in the regression models. Additionally this assumption makes the calculations of the relative efficiency much easier.

Considering the results in Tables 1 and 2 the following conclusions are made: The relative efficiency obtained using ERSS to estimate the population mean of the variable of interest depends upon the sample size n, the number of cycles r and the value of the correlation coefficient ρ between the variable of interest and the auxiliary variable. If the ranking of the variable of interest is perfect (Table 1) the ERSS estimator is superior to its counterpart, the regression estimator, unless the value of $|\rho| > 0.80$. If the ranking of the variable of interest is done using a concomitant variable (Tables 2) and the number of cycles is 1, the ERSS estimator is more efficient (unless the value of $|\rho| > 0.80$ then the regression estimator is more efficient). But if $r > 1$ and the value of $|\rho| < 0.80$ there is not much difference between the two estimators. Finally, if $|\rho| > 0.80$ and $r > 1$ the regression estimator is superior.

In the basis of Table 3 we could conclude that: The relative efficiency obtained using ERSS increases as $|\rho|$ and /or n increases. The relative efficiency obtained using ERSS with a concomitant variable is very close to that of RSS if $|\rho| < 0.8$ and /or $n < 6$.

Considering Table 4 the following conclusions can be made: A gain efficiency obtained using ERSS with errors in ranking if the underline distribution is symmetric around μ. For example, for $n = 6$ and $\sigma_e = 0.5$, the efficiency of the ERRS 2.34 for estimating the population mean of a uniform distribution. The efficiency of ERSS decreases as the value of σ_e is increases. If the parent distribution is uniform the gain in efficiency using ERSS with or without errors in raking is much higher than its counterpart RSS.

As we know, in real life applications it is not easy or sometimes impossible to rank the units without errors in ranking, especially if the sample size is more than 5 or 6 units. The recommendation is to use the ERSS with an even sample size if it is difficult to rank the units using RSS. Since it is very easy to find the largest or the smallest of the i^{th} sample with minimum errors in ranking.

ACKNOWLEDGMENT

The author is indebted to King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia for the excellent research facilities.

REFERENCES

1. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics,* 28, 545-555.
2. McIntyre, G.A. (1952). A method of unbiased selective sampling, using ranked sets. *Aust. J. Agri. Res.,* 3, 385-390.
3. Patil, G.P., Sinha, A.K. and Taillie, C. (1993). Relative efficiency of ranked set sampling: A comparison with the regression estimator. *Environmetrics,* 4, 399-412.
4. Samawi, H., Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling. *The Biometrical Journal*, 30, 577-586.
5. Stokes, S.L. (1977). Ranked set sampling with concomitant variables. *Commun. in Statist.,* A6, 1207-1211.
6. Sukhatme, P.V. and Sukatme, B.V. (1970). *Sampling Theory of Surveys with Applications*. Iowa State University Press, Ames, Iowa.
7. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.,* 20, 1-31.
8. Tietjen, G.L., Kahaner, D.K and Beckman, R.J. (1977). Variances and covariances of the normal order statistics for sample sizes 2 to 50. *Selected Tables in Mathematical Statistics,* 5, 1-75.

CHAPTER SEVEN

Variance Estimation for the Location-scale Family Distributions using Ranked Set Sampling

Walid Abu-Dayyeh[1] and **Hassen A. Muttlak**[2]
[1] Department of Statistics, Yarmouk University, Irbid, Jordan
[2] Department Mathematical Sciences, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia.
Email: hamstat@kfupm.edu.sa

ABSTRACT

The ranked set sampling (RSS) has been shown as an efficient method in estimating the population mean and other parameters of interest of certain probability distributions. Very little, has however been done in estimating the population variance using RSS. In this paper, several unbiased estimators of the population variance are proposed for the location-scale family distributions using RSS and some its modifications. The proposed estimators are compared to the usual simple random sampling (SRS) estimators for several probability distributions. Most of these proposed estimators are proven to be more efficient then the usual SRS estimators.

KEY WORDS

Median ranked set sampling; order statistic; relative precision and simple random sampling.

1. INTRODUCTION

The ranked sample sampling (RSS) method as suggested by McIntyre (1952) and used to estimate mean pasture yield can be summarized as follows: Select m random samples of size m units and rank the units within each sample with respect to a variable of interest by a visual inspection or any other cost free method. Then select for actual measurement the smallest unit from the first sample. From the second sample, select for actual measurement the second smallest unit. The procedure is continued until the largest unit from the m^{th} sample is selected for measurement. In this way, we obtain a total of m measured units, one from each sample. The cycle may be repeated r times until mr units have been measured. These mr units form the RSS data.

Various authors have studied many aspects of RSS. For more details see Patil et al. (1995) and Sinha, et al. (1996). Most of the work done on the RSS is devoted to estimate the population mean μ. Little has been done in estimating the population variance σ^2 using RSS.

Stoke (1980) showed that the estimator of σ^2 based on RSS data is asymptotically unbiased estimator of σ^2. Further, if mr is large enough it is more efficient than the usual estimator based on simple random sample (SRS).

Stokes (1995) considered estimation of μ and σ for the family of random variables with probability distribution function of the form $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ using the maximum likelihood method.

Finally, Yu et al. (1999) considered estimating the variance of a normal population using RSS. They proposed several estimators for σ^2 and compared the performances of these estimators.

In this paper, we consider estimating the population variance σ^2 of the location-scale family having a probability distribution function of the form $f(x,\theta,\lambda) = \mathrm{g}\left((\mathrm{x}\text{-}\theta)/\lambda\right)/\lambda$. Estimating σ^2 is considered when the location parameter θ known and when it is unknown. Under both cases, we proposed several unbiased estimators using RSS and some of its modifications. These estimators are then compared with the usual SRS estimators via their variances when the underlying distributions are normal, logistic, and student t and double exponential to the usual estimators using SRS. It turns out that most of the newly suggested estimators are more efficient than the usual estimators based on SRS.

2. VARIANCE ESTIMATION IN CASE OF KNOWN LOCATION PARAMETER

Let $X_1, X_2, \ldots, X_n$, where n = mr be a random sample with distribution function

$$f(x,\theta,\lambda) = \mathrm{g}((\mathrm{x}\text{-}\theta)/\lambda)/\lambda\,, \tag{2.1}$$

where θ is a location parameter, λ is a scale parameter and g is a probability distribution. Suppose without lost of generality that the value of the location parameter θ is zero. The well-known method of moment estimator for the population variance is

$$\hat{\sigma}_{01}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2\,, \tag{2.2}$$

with variance

$$V\left(\hat{\sigma}_{01}^2\right)=\frac{1}{n}\left[E\left(X_i^{\,4}\right)-E\left(X_i^{\,2}\right)^2\right].$$

Let X_1, X_2,..., X_m be a SRS of size m with probability distribution function f(x) with finite mean and variance. Let X_{11}, X_{12},..., X_{1m}; X_{21}, X_{22}, ..., X_{2m}; ...; X_{m1}, X_{m2},..., X_{mm} be independent random variables all with the same cumulative distribution function F(x). Let $X_{(i)j}$ denotes the ith order statistic from the ith sample of size m (i = 1,2,..., m) in the jth cycle (j = 1,2,…, r). Takahasi and Wakimoto (1968) defined the unbiased estimator of the population mean using RSS to be

$$\hat{\mu}=\frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m}X_{(i)j}\ ,$$

with variance

$$V(\hat{\mu})=\frac{1}{m^2 r}\sum_{i=1}^{m}\sigma_{(i)}^2\ ,$$

where $\sigma_{(i)}^2=E\left[X_{(i)}-E\left(X_{(i)}\right)\right]^2$.

We propose the following unbiased estimator of σ^2 which is based on RSS data from a population with probability distribution as defined in equation (2.1) with $\theta = 0$

$$\hat{\sigma}_{02}^2=\frac{1}{rm}\sum_{j=1}^{r}\sum_{i=1}^{m}X_{(i)j}^2 \tag{2.3}$$

with variance

$$V(\hat{\sigma}_{02}^2)=\frac{1}{(rm)^2}\sum_{j=1}^{r}\sum_{i=1}^{m}V\left(X_{(i)j}^2\right)=\frac{1}{(rm)^2}\sum_{j=1}^{r}\sum_{i=1}^{m}\left[E\left(X_{(i)j}^4\right)-\left(E\left(X_{(i)j}^2\right)\right)^2\right],$$

where $X_{(i)j}$ is the ith order statistic of a SRS of size m of the jth cycle from the above population.

We now consider some modifications of RSS. The smallest, largest or the median of the ith sample i = 1,2,…, m, will be selected for measurement instead of selecting the ith smallest of the ith sample as we have done for the usual RSS. Again we replicate the selection procedure r times to get a sample of size n = rm units.

We propose the following unbiased estimator of σ^2, which is based on the smallest order statistic ($X_{(1)k}$, k = 1,2, …, n =mr) of a SRS of size m with r cycles

$$\hat{\sigma}_{03}^2 = \frac{{\sigma_0}^2}{n\,E(M_{1:m}^2)} \sum_{k=1}^{n} X_{(1)k}^2 \ , \qquad (2.4)$$

with variance

$$V\left(\hat{\sigma}_{03}^2\right) = \frac{\sigma_0^4}{n\left(E\left(M_{1:m}^2\right)\right)^2} \left[E\left(X_{(1)k}^4\right) - \left(E\left(X_{(1)k}^2\right)\right)^2\right],$$

where $\sigma^2 = \mathrm{V(X)} = \lambda^2\sigma_0^2$, X~ $f(x,\theta,\lambda)$, $\sigma_0^2 = V(M)$, $M \sim f(x,0,1)$ and $M_{i:m}$ is the i$^{\text{th}}$ order statistic of a SRS of size m from $f(x,0,1)$. Finally, note that $\sigma^4 = \left(\mathrm{V(X)}\right)^2$ and $\sigma_0^4 = \left(V(M)\right)^2$.

Now using the largest order statistic of a SRS of size m with r cycles the following unbiased estimator for the population variance σ^2 is proposed

$$\hat{\sigma}_{04}^2 = \frac{\sigma_0^2}{nE(M_{m:m}^2)} \sum_{k=1}^{n} X_{(m)k}^2 \ , \qquad (2.5)$$

with variance

$$V\left(\hat{\sigma}_{04}^2\right) = \frac{\sigma_0^4}{n\left(E\left(M_{m:m}^2\right)\right)^2} \left[E\left(X_{(m)k}^4\right) - \left(E\left(X_{(m)k}^2\right)\right)^2\right].$$

The median ranked set sampling MRSS can be summarized as follows: Select m random samples of size m units from the population and rank the units within each sample with respect to a variable of interest. If the sample size m is odd, from each sample select for measurement the ((m + 1)/2)$^{\text{th}}$ smallest rank (the median of the sample). If the sample size is even, select for measurement from the first m/2 sample the (m/2)$^{\text{th}}$ smallest rank and from the second m/2 sample the ((m +2)/2)$^{\text{th}}$ smallest rank. The cycle may be repeated r times to get mr units. Let $X_{(med)k} = X_{\left(\frac{m+1}{2}\right)k}, k = 1, 2, ..., n = mr$ denote the median of the ith sample from the jth cycle if the sample size m is odd. If the sample size is even let $X_{(med)k} = X_{\left(\frac{m}{2}\right)\mathrm{k}}$ m/2 times and $X_{(med)k} = X_{\left(\frac{m}{2}+1\right)\mathrm{k}}$ m/2 times for the j$^{\text{th}}$ cycle. The unbiased estimator for σ^2 using MRSS if the probability distribution is symmetric in the case of even sample size can be defined as following:

$$\hat{\sigma}_{05}^{2} = \frac{{\sigma_o}^2}{n\, E\left(M_{med:m}^{2}\right)} \sum_{k=1}^{n} X_{(med)k}^{2} \tag{2.6}$$

with variance

$$V(\hat{\sigma}_{05}^{2}) = \frac{\sigma_0^4}{n\left(E\left(M_{med:m}^{2}\right)\right)^2} \left[E\left(X_{(med)k}^{4}\right) - \left(E\left(X_{(med)k}^{2}\right)\right)^2\right].$$

We will use the variances to compare the estimators given in equations (2.2–2.6) by the relative precision of δ_2 to δ_1, which is defined as follows:

$$RP\left(\delta_2, \delta_1\right) = \frac{V(\delta_1)}{V(\delta_2)}.$$

The relative precision of the estimators using RSS and some modification of RSS equations (2.3–2.6) with respect to SRS estimator equation (2.2) is calculated for the distributions: logistic, normal, student t and double exponential. Table 1 shows the relative precision for sample size m = 2, 3, 4, 5, 6 and replication r = 5 and 10.

From Table 1, we can see that the estimators of the population variance using RSS or some of its modification will improve the relative precision for most of the cases considered in this study. For example for m = 4 and r = 5, the $RP(\hat{\sigma}_{03}^2, \hat{\sigma}_{01}^2)$ values for the logistic, normal, student t and double exponential distributions are 1.572, 1.614, 1.632 and 1.477, respectively. Considering Table 1 the following remarks can be made:

1. Using RSS or some of its modification will improved the efficiency of the newly suggested estimators over the SRS estimator for the distribution considered with the exception of using the median for the normal distribution.
2. Increasing the sample size m will increase the relative precision for almost all the cases considered.
3. Increasing the number of cycle r more than 5 will not increase the relative precision.

3. VARIANCE ESTIMATION IN CASE OF UNKNOWN LOCATION PARAMETER

Let X_1, X_2, ..., X_n, where n = rm be a random sample with probability distribution function as defined in equation (2.1). The sample variance is used to estimate the population variance σ^2 using SRS, which is

$$\hat{\sigma}_{11}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \tag{3.1}$$

with a variance (Stuart and Ord, 1994, P. 370)

$$V\left(\hat{\sigma}_{11}^2\right) = \frac{1}{n}\left[\mu_4 - \left(\tfrac{n-3}{n-1}\right)\sigma^4\right],$$

where μ_4 is the population 4th central moment.

Now we will consider unbiased estimators based on RSS. The first estimator we call the McIntyre modified estimator is defined as follows:

$$\hat{\sigma}_{12}^2 = \frac{1}{(mr-1)\left[1 + \frac{r}{m(rm-1)}\sum_{i=1}^{m} M_{i:m}^2\right]}\sum_{j=1}^{r}\sum_{i=1}^{m}(X_{(i)j} - \hat{\mu})^2 \tag{3.2}$$

with variance

$$V(\hat{\sigma}_{12}^2) = \frac{m}{(mr-1)\left(m(mr-1) + r\sum_{i=1}^{m} M_{i:m}\right)}\left\{\left(\frac{mr-1}{mr}\right)^2\sum_{j=1}^{r}\sum_{i=1}^{m}\mu_{4(i)j}\right.$$

$$+4\sum_{j=1}^{r}\sum_{i=1}^{m}\mu_{(i)j}^2\sigma_{(i)j}^2 + 4\left(\frac{mr-1}{mr}\right)\sum_{j=1}^{r}\sum_{i=1}^{m}\mu_{(i)j}\,\mu_{3(i)j}$$

$$\left. + \frac{4m}{(mr)^2}\sum_{j=1}^{r}\sum_{i=1}^{m}\sum_{s=i+1}^{m}\sigma_{(i)j}^2\sigma_{(s)j}^2 + \frac{2(m-1)-(mr-1)^2}{(mr)^2}\sum_{j=1}^{r}\sum_{i=1}^{m}\sigma_{(i)j}^4\right\},$$

where $\mu_{k(i)j} = E\left(X_{(i)j} - \mu_{(i)j}\right)^k$, $\mu_{(i)j} = \mathrm{E}\left(X_{(i)j}\right)$, $\sigma^2{}_{(i)j} = V\left(X_{(i)j}\right)$.

The following two unbiased estimators are based on the smallest and the largest order statistic of sample of size m with r cycles. First, let consider the estimator based on the smallest order statistic

$$\hat{\sigma}_{13}^2 = \frac{\sigma_0^2}{(\text{n-1})\text{V}_{1:\text{m}}}\sum_{k=1}^{n}\left(X_{(1)k} - \overline{X}_{(1)}\right)^2, \tag{3.3}$$

with variance

$$V(\hat{\sigma}_{13}^2) = \frac{\sigma_0^4}{nV_{1:m}^2}\left[\mu_4 - \left(\tfrac{m-3}{m-1}\right)V_{1:m}^2\right],$$

where $\overline{X}_{(1)} = \frac{1}{n} \sum_{k=1}^{n} X_{(1)k}$, $V_{1:m} = V\left(M_{(1)k}\right)$, μ_4 is the population 4th central moment and the other terms are as defined is Section 2. Now for the largest order statistic, we propose the estimator

$$\hat{\sigma}_{14}^2 = \frac{\sigma_0^2}{(\text{n-1})\text{V}_{\text{m:m}}} \sum_{k=1}^{n} \left(X_{(m)k} - \overline{X}_{(m)} \right)^2 , \tag{3.4}$$

with variance

$$V(\hat{\sigma}_{14}^2) = \frac{\sigma_0^4}{nV_{m:m}^2} \sum_{\text{k}=1}^{\text{n}} \left[\mu_4 - \left(\tfrac{m-3}{m-1}\right) V_{m:m}^2 \right] ,$$

where $\overline{X}_{(m)} = \frac{1}{n} \sum_{k=1}^{n} X_{(m)k}$, $V_{m:m} = V\left(M_{(m)k}\right)$ and the other terms as defined before.

Finally, using MRSS and assuming that the underlining distribution is symmetric when the sample size is even we propose the following unbiased estimator:

$$\hat{\sigma}_{15}^2 = \frac{\sigma_0^2}{(n-1)\text{V}_{\text{med:m}}} \sum_{k=1}^{n} \left(X_{(med)k} - \overline{X}_{(med)} \right)^2 , \tag{3.5}$$

with variance

$$V(\hat{\sigma}_{15}^2) = \frac{\sigma_0^4}{nV_{med:m}^2} \left[\mu_4 - \left(\tfrac{m-3}{m-1}\right) V_{med:m}^2 \right],$$

where $\overline{X}_{(med)} = \frac{1}{n} \sum_{k=1}^{n} X_{(med)k}$, $V_{med:m} = V\left(M_{(med)k}\right)$ and other terms as defined before.

As was done in Section 2, we use the variances to compare the estimators given in equations (3.1–3.5). The relative precision of the estimators using RSS and some its modifications equations (3.2–3.5) with respect to SRS estimator equation (3.1) is calculated for the distributions: logistic, normal, student t and double exponential. Table 2 shows the relative precision for sample size m = 2, 3, 4, 5, 6 and replication r = 5 and 10.

From Table 2, we can see that the estimators of the population variance using RSS will improve the relative precision for all the cases considered in this study. For example for m =5 and r = 5, the $RP(\hat{\sigma}_{12}^2, \hat{\sigma}_{11}^2)$ values for the logistic, normal, double exponential and student t distributions are 1.367, 1.311, 8.313 and 1.120, respectively. In addition, the median estimator is more efficient than the SRS estimator for all the distributions considered except for the normal distribution. However, the other estimators based on the smallest and largest order statistics are not doing well as compared to the estimator based on SRS.

4. CONCLUSIONS

RSS method is very efficient and widely used method in estimating the population mean. However, little has been done for estimating the population variance. The question is if we collect our data using RSS method to estimate the population mean or some other parameters of interest can we use this data to estimate the population variance. In this paper, we have proposed several unbiased estimators of the population variance, using the method of RSS and some of its modifications for the location-scale parameter family of distributions. Our results clearly indicate the superiority of RSS estimators over the usual SRS estimators. Except for the normal distribution, MRSS estimator dominates the SRS estimator for all other distribution considered.

The estimators based on respectively the first and the last order statistics (i.e. the smallest and the largest) haven been shown to do quite well for the case of known location parameter.

The estimators proposed in this study are all unbiased for estimating the population variance of the location-scale family distributions, but the amount of improvement in the relative precision depends on the underlining distribution.

Finally, we note that the relative precision of the proposed estimators under each of RSS, MRSS, smallest and largest methods increase with the increasing sample size m. However, the RSS is more difficult to perform than the other methods for large sample sizes, makes the proposed estimators under MRSS, smallest and largest are more appealing than RSS in case where they perform better.

REFERENCES

1. McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Aust. J. Agri. Res.,* 3, 385-390.
2. Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Ranked set sampling: An annotated bibliography. *Environmental and Ecological Statistics*, 2, 25-54.
3. Sinha, B.K., Sinha, B.K. and Purkayastha, S. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics and Decisions,* 14, 223-240.
4. Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics,* 36, 35-42.

5. Stokes, S.L. (1995). Parametric ranked set sampling. *Ann. Inst. Statist. Math.,* 47(3), 465-482.
6. Stuart, A. and Ord, J.K. (1994). Kendall's Advanced Theory of Statistics, Sixth Edition, Volume 1: *Distribution Theory*. Edward Arnold.
7. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.*, 20, 1-31.
8. Yu, P.L.H, Lam, K. and Sinha, B. K. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. *Environmental and Ecological Statistics* 6, 23-45.

Table 1
The relative precision of the estimators based on RSS with respect to SRS estimator for the population variance for the logistic, normal, double exponential and student t distributions when the location parameter is known

m	$RP(\hat{\sigma}_{02}^2,\hat{\sigma}_{01}^2)$		$RP(\hat{\sigma}_{03}^2,\hat{\sigma}_{01}^2)$		$RP(\hat{\sigma}_{04}^2,\hat{\sigma}_{01}^2)$		$RP(\hat{\sigma}_{05}^2,\hat{\sigma}_{01}^2)$	
	r		R		r		R	
	5	10	5	10	5	10	5	10
Logistic distribution								
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.061	1.061	1.259	1.259	1.259	1.259	1.234	1.234
4	1.131	1.131	1.572	1.572	1.572	1.572	1.234	1.234
5	1.199	1.199	1.882	1.882	1.882	1.882	1.344	1.344
6	1.267	1.267	2.175	2.175	2.175	2.175	1.344	1.344
Normal distribution								
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.082	1.082	1.268	1.268	1.268	1.268	0.983	0.983
4	1.179	1.179	1.614	1.614	1.614	1.614	0.983	0.983
5	1.279	1.279	1.974	1.974	1.974	1.974	0.984	0.984
6	1.379	1.379	2.329	2.329	2.329	2.329	0.984	0.984
Double exponential distribution								
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.049	1.049	1.294	1.294	1.294	1.294	1.205	1.205
4	1.102	1.102	1.632	1.632	1.632	1.632	1.205	1.205
5	1.155	1.155	1.962	1.962	1.962	1.962	1.352	1.352
6	1.256	1.256	2.275	2.275	2.275	2.275	1.352	1.352
Student t distribution								
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.031	1.031	1.225	1.225	1.225	1.225	2.671	2.671
4	1.060	1.060	1.477	1.477	1.477	1.477	2.671	2.671
5	1.087	1.087	1.708	1.708	1.708	1.708	3.150	3.150
6	1.112	1.112	1.912	1.912	1.912	1.912	3.150	3.150

Table 2
The relative precision of the estimators based on RSS with respect to SRS estimator for the population variance for the logistic, normal, double exponential and student t distributions when the location parameter is unknown

m	$RP\left(\hat{\sigma}_{12}^2,\hat{\sigma}_{11}^2\right)$		$RP\left(\hat{\sigma}_{13}^2,\hat{\sigma}_{11}^2\right)$		$RP\left(\hat{\sigma}_{14}^2,\hat{\sigma}_{11}^2\right)$		$RP\left(\hat{\sigma}_{15}^2,\hat{\sigma}_{11}^2\right)$	
	r		r		r		r	
	5	10	5	10	5	10	5	10
Logistic distribution								
2	1.177	1.083	0.963	0.961	0.963	0.961	0.963	0.961
3	1.248	1.150	0.896	0.894	0.896	0.894	1.225	1.228
4	1.308	1.216	0.856	0.854	0.856	0.854	1.237	1.242
5	1.367	1.281	0.830	0.829	0.830	0.829	1.332	1.338
6	1.424	1.344	0.813	0.812	0.813	0.812	1.348	1.353
Normal distribution								
2	1.024	1.011	0.976	0.973	0.976	0.973	0.976	0.973
3	1.113	1.097	0.955	0.950	0.955	0.950	0.986	0.985
4	1.211	1.195	0.939	0.932	0.939	0.932	0.983	0.981
5	1.311	1.295	0.926	0.917	0.926	0.917	0.987	0.986
6	1.410	1.397	0.914	0.905	0.914	0.905	0.987	0.985
Double exponential distribution								
2	2.628	2.595	1.456	1.492	1.456	1.492	1.456	1.492
3	4.414	4.215	1.905	2.001	1.905	2.001	1.183	1.195
4	6.313	6.151	2.280	2.447	2.280	2.447	1.243	1.259
5	8.313	8.062	2.592	2.831	2.592	2.831	1.310	1.332
6	10.41	10.05	2.854	3.166	2.854	3.166	1.321	1.344
Student t distribution								
2	1.040	1.079	0.495	0.487	0.495	0.487	0.495	0.487
3	1.069	1.109	0.301	0.294	0.301	0.294	2.432	2.556
4	1.095	1.132	0.227	0.221	0.227	0.221	1.511	1.538
5	1.120	1.172	0.188	0.183	0.188	0.183	2.796	2.977
6	1.143	1.154	0.164	0.159	0.164	0.159	1.982	2.050

CHAPTER EIGHT

New Median Ranked Set Sampling

Dinesh S. Bhoj
Department of Mathematical Sciences
Rutgers University, USA

ABSTRACT

A new median ranked set sampling (NMRSS) procedure is proposed to estimate the mean of a population. The estimator based on NMRSS is compared with estimators based on other ranked set sampling procedures. It is shown that the relative precisions of the estimator based on NMRSS are higher than those based on other ranked set sampling methods for unimodal symmetric distributions around the mean. The NMRSS method also works quite well for moderate skew distributions.

KEY WORDS

Mean square error; modified ranked set sampling; ordered observations; ranked set sampling; relative precision; unbiased estimator; variance.

1. INTRODUCTION

Ranked set sampling (RSS) was first introduced by McIntyre (1952) in relation to estimating pasture yields. This is a cost-efficient alternative to simple random sampling (SRS) if observations are rather cheaply or cost free ranked without actually measuring them. Dell and Clutter (1972) and Takahasi and Wakimoto (1968) provided mathematical foundation for RSS. They showed that the sample mean of the RSS is an unbiased estimator for the population mean with smaller variance than that of the sample mean of SRS with the same sample size. Dell and Clutter (1972) also showed that the estimator for the population mean based on RSS is at least as efficient as the SRS estimator even when there are ranking errors.

The procedure of selection of RSS involves drawing of n random samples with n units in each sample. The n units in each sample are ranked with respect to a variable of interest. Then the unit with the lowest rank is measured from the first sample, the unit with the second lowest rank is measured from the second sample, and this procedure is continued until the unit with the highest rank is quantified from the n^{th} sample. The n^2 ordered observations in the n samples can be displayed as:

$$x_{(11)} x_{(12)} \cdots x_{(1n)}$$
$$x_{(21)} x_{(22)} \cdots x_{(2n)}$$
$$\vdots \quad \vdots \quad \vdots$$
$$x_{(n1)} x_{(n2)} \cdots x_{(nn)}$$

only $x_{(11)}$, $x_{(22)}$, ... $x_{(nn)}$ are accurately measured and they constitute the RSS data. If n is small, the cycle may be repeated r times to increase the sample size. For convenience, we assume that r = 1. McIntyre (1952) proposed the estimator $\bar{\mu} = \sum x_{(ii)}/n$ for the population mean μ.

2. MODIFIED RANKED SET SAMPLING

The RSS procedure has been modified by some authors (Bhoj (1997) and Muttlak (1997)) which has further reduced the variance of the estimator for the mean. Bhoj (1997) proposed a general modified ranked set sampling in the parametric setting. In this scheme he proposed to select only two order statistics for even n=2m. He suggests to select the jth order statistic from the first m samples and kth order statistic from the last m samples. The choices of the jth and kth order statistics depend on the distribution under consideration and the parameter(s) to be estimated. If one is interested in estimating the mean of a symmetrical distribution around mean, the modified ranked set sampling procedure becomes the median ranked set sampling (MRSS). Muttlak (1997) proposed to use MRSS procedure for symmetrical and also for skew distributions.

In MRSS procedure, we draw n random samples of size n from the population and rank n observations in each sample. If n is odd, we measure the observation with rank (n+1)/2 from each sample. If n is even, we measure mth order statistic from the first m samples and (m+1)th order statistic from the last m samples. Then our estimator for μ is

$$\hat{\mu} = \frac{1}{n}\left[\sum_{i=1}^{m} x_{(im)} + \sum_{i=m+1}^{n} x_{(im+1)}\right], \text{ for even n,}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{(ik)} \quad , \quad \text{for odd n,}$$

where $k = (n+1)/2$.

Let v_i denote the variance of $y_{(ii)} = (x_{(ii)} - \mu)/\sigma$, where σ^2 is the variance of the population. Then the variance of $\hat{\mu}$ is given by

$$Var(\hat{\mu}) = (v_m + v_{m+1})\sigma^2 / 2n \text{, for even n,}$$
$$= v_k \sigma^2 / n \text{, for odd n.}$$

The following results hold for $\hat{\mu}$.

a) $\hat{\mu}$ is an unbiased estimator for μ if the distribution is symmetric about μ.
b) $Var(\hat{\mu}) < Var(\overline{\mu})$ for n>2 if the symmetric distribution about μ is unimodal. For n=2, RSS and MRSS are identical and hence $Var(\hat{\mu}) = Var(\overline{\mu})$.
c) For any distribution, $Var(\hat{\mu}) < Var(\overline{x})$, where $\overline{x}$ is the sample mean based on SRS.
d) If the distribution is asymmetric about μ, $\hat{\mu}$ is a biased estimator for μ. In this case, Muttlak (1997) demonstrated that for most distributions, the mean square error (MSE) of $\hat{\mu}$ is less than $Var(\overline{x})$ for small sample sizes.

3. NEW MEDIAN RANKED SET SAMPLING

The accuracy of $\hat{\mu}$ over $\overline{\mu}$ is compared by computing the relative precision, $RP_1 = Var(\overline{\mu}) / Var(\hat{\mu})$, where $\hat{\mu}$ is an unbiased estimator for μ. The MRSS procedure does not perform well for even n as compared to odd n. To exemplify this characteristic of MRSS, we present in Table 1 the values of RP_1 for the logistic distribution along with the relative percentage increases (RPI) in RP_1. The RPI in RP_1 is defined as

$$\frac{RP_1 \text{ for } n - RP_1 \text{ for } (n-1)}{RP_1 \text{ for } (n-1)} .100 \text{ , } n > 2 .$$

Table 1
RP_1 and RPI for the logistic distribution

N	**2**	**3**	**4**	**5**	**6**	**7**
RP_1	1.0000	1.3876	1.4276	1.6154	1.6556	1.7742
RPI	-	38.76	2.87	13.17	2.49	7.16

It is clear that the values of RPI are higher when we move from even to odd values of n, and they are lower when we switch from odd to even values of n. Therefore, in this section we propose a new median ranked set sampling (NMRSS) for even n=2m. In this procedure, we draw first m samples of size (n-1) and last m samples of size (n+1). As in RSS or MRSS, we order the observations in each sample with visual inspection or methods not requiring actual measurements. Then we quantify the median from each sample for estimating the population mean. Thus our NMRSS data are $x_{(im)}, i = 1, 2, ...m$ from samples of size (n-1) and $x_{(i\,m+1)}, i = m+1, ..., n$ from samples of size (n+1). Now we propose an estimator

$$\mu^* = w\sum_{i=1}^{m} x_{(im)} \Big/ m + (1-w) \sum_{i=m+1}^{n} x_{(im+1)} \Big/ m \;,$$

where $w = (n-1)/2n$. The variance of μ^* is given by $Var(\mu^*) = \left\{ w^2 v_{m(n-1)} \big/ m + (1-w)^2 v_{m+1(n+1))} \big/ m \right\} \sigma^2$, where $V_{m(n_i)}$ denotes the variance of the m^{th} order statistic from a sample of size n_i. In the next section we compare the three estimators of μ based on three ranked set sampling procedures.

4. COMPARISON OF ESTIMATOR

In this section we compare the estimators $\hat{\mu}$ and μ^* with the estimator $\overline{\mu}$, an unbiased estimator for μ. $\hat{\mu}$ and μ^* are unbiased estimators for μ if the distribution is symmetric about μ. For asymmetric distributions, $\hat{\mu}$ and μ^* are biased estimators. In this case, for comparison, we use mean square error (MSE) of the estimator, where MSE = Variance + (bias)^2. We have computed the two relative precisions

$$RP_1 = \frac{Var(\overline{\mu})}{Var(\hat{\mu})} \text{, for a symmetric distribution,}$$

$$= \frac{Var(\overline{\mu})}{MSE(\hat{\mu})} \text{, for a skew distribution,}$$

and $$RP_2 = \frac{Var(\overline{\mu})}{Var(\mu^*)} \text{, for a symmetric distribution,}$$

$$= \frac{Var(\overline{\mu})}{MSE(\mu^*)} \text{, for a skew distribution.}$$

The values of RP_1 and RP_2 are presented in Table 2 for some distributions and three sample sizes. The moments of order statistics for these distributions are readily available; see Harter and Balakrishnan (1996). We can draw the following conclusions:

Table 2
Relative precisions of the estimators $\hat{\mu}$ and μ^*

Distribution	RP_1			RP_2		
	n = 2	n = 4	n = 6	n = 2	n = 4	n = 6
Normal	1.000	1.182	1.275	1.083	1.216	1.293
Logistic	1.000	1.427	1.656	1.230	1.515	1.701
Laplace	1.000	1.884	2.533	1.484	2.161	2.722
Gamma (3)	1.000	1.209	1.049	1.165	1.209	1.023
Gamma (5)	1.000	1.198	1.120	1.131	1.208	1.106
Weibull (2)	1.000	1.104	1.061	1.054	1.110	1.053
Weibull (40	1.000	1.122	1.184	1.045	1.146	1.196
Extreme Value	1.000	1.297	1.192	1.208	1.318	1.173

a) For unimodal symmetric distributions around μ, the new median ranked set sampling is better than the median ranked set sampling procedure.
b) For moderate skew distributions, NMRSS is better than MRSS for small n.
c) For extremely skew distributions, NMRSS is better than MRSS for n=2. We checked this for Exponential, Pareto(5) and Weibull (.5) distributions. The values of RP_2 for these distributions are, respectively, 1.334, 1.875 and 2.671.
d) Both MRSS and NMRSS are not suitable for a rectangular distribution (not given in Table 2). This is expected since the optimal choices in this case are the extreme order statistics; see Bhoj (1997) and Bhoj and Ahsanullah (1996).
e) The values of RP_1 and RP_2 increase with n for symmetrical distributions around μ. The values of RP_1 and RP_2 are higher for n=4 when compared to n=2 for moderate skew distributions. However, for $n \geq 6$, the values of RP_1 and RP_2 decrease as n increases. Although *Var(*μ^**)* is smaller than *Var(*$\hat{\mu}$*)*, the bias in μ^* increases faster than bias in $\hat{\mu}$ as n increases. Therefore MSE*(*μ^**)* may be higher than MSE*(*$\hat{\mu}$*)* for some skew distributions for $n \geq 6$.

We revise Table 1 where we have computed RPI for the logistic distribution with MRSS procedure. The values of RPI in RP_2 with NMRSS for n=3,4,5,6 and 7 are, respectively, 12.80, 9.17, 6.64, 5.31 and 4.29. This clearly shows that NMRSS procedure works well for even n.

Based on the computations of relative precisions, we recommend the NMRSS procedure for even values of n when the samples are drawn from unimodal symmetric distributions. For moderate skew distributions, NMRSS may be recommended for $n = 2$ and 4. Most importantly for most skew distributions we can recommend the NMRSS procedure with n=2. In order to increase the sample size, the cycle may be repeated $r \geq 2$ times.

REFERENCES

1. Bhoj, D.S. and Ahsanullah, M. (1996). Estimation of parameters of the generalized geometric distribution using ranked set sampling. *Biometrics*, 52, 685-694.
2. Bhoj, D.S. (1997). Estimation of parameters using modified ranked set sampling. *App. Statist. Scien.*, II, Nova Science Publishers, Inc., New York. M. Ahsanullah, Editor, 154-163.
3. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-555.
4. Harter, H.L. and Balakrishnan, N. (1996). *CRC Handbook of Tables for the Use of Order Statistics in Estimation*. CRC Press, Boca Raton.
5. McIntyre, G.A. (1952). A method of unbiased selective sampling, using ranked sets. *Aust. J. of Agri. Res.*, 3, 385-390.
6. Muttlak, H.A. (1997). Median ranked set sampling. *J. App. Statist. Scien.*, 6, 245-255.
7. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.*, 20, 1-31.

CHAPTER NINE

Statistical Quality Control based on Pair and Selected Ranked Set Sampling

Hassen A. Muttlak and **Walid S. Al-Sabah**
Department of Mathematical Sciences
King Fahd University of Petroleum & Minerals,
Dhahran 31261 Saudi Arabia

ABSTRACT

Different quality control charts for the sample mean are developed using pair ranked set sampling (PRSS), and selected ranked set sampling (SRSS). These new charts are compared to the usual control charts based on simple random sampling (SRS) data. The charts based on PRSS and SRSS are shown to have smaller average run length (ARL) than the classical chart especially if the process starts to get out of control. Through this study we are assuming that the underlying distribution is normal.

KEY WORDS

Average run length, ranked set sampling, lower confidence limit, simple random sampling and upper confidence limit.

1. INTRODUCTION

This study is concerned with the idea of developing quality control charts using pair ranked set sampling (RSS) and selected ranked set sampling (SRSS) data. These newly developed control charts are considered as alternatives, and more efficient methods than the usual control charts based on the simple random sampling (SRS) method.

The RSS method was first suggested by McIntyre (1952) who noted that it is highly beneficial and superior to the standard simple random sampling (SRS) for estimation of the population mean. Takahasi and Wakimoto (1968) supplied the necessary mathematical theory. Dell and Clutter (1972) studied the case in which the ranking may not be perfect, i.e. there are errors in ranking the unit with respect to the variable of interest.

As pointed out by Dell and Clutter (1972) there will be loss in efficiency depending on the amount of errors in ranking the units. To overcome this problem, or at least to reduce the errors in ranking the units selected from the population, Samawi et al. (1996) studied the properties of estimating the population mean using the extreme ranked set sampling method, Muttlak (1997) suggested the median ranked set sampling method and Hossain and Muttlak (1999) and (2001) respectively suggested using pair ranked set sampling (PRSS) and selected ranked set sampling (SRSS).

Muttlak and Al-Sabah (2001) developed quality control charts for the sample mean using ranked set sampling, median ranked set sampling and extreme ranked set sampling. They compared the newly developed charts using average run length (ARL) to the Shewhart control chart. They showed that the newly charts are more efficient than the Shewhart charts i.e. they have smaller average run length. Finally they collected a real life data set and used to construct quality control charts for the newly suggested control charts.

In this paper the pair ranked set sampling (PRSS), and selected ranked set sampling (SRSS) will be used to develop control charts for the population mean. These charts are compared to the well-known quality control charts for variables using the usual simple random sampling (SRS) data; see for example Montgomery (1995). The control charts for PRSS and SRSS data are shown to have smaller average run length (ARL) then the usual control charts based on SRS data.

2. SAMPLING METHODS

2.1 Ranked Set Sampling (RSS)

The RSS procedure can be summarized as follows: Select *n* random sets, each of size *n* units from the population, and rank the units within each set with respect to a variable of interest. Then an actual measurement is taken from the unit with the smallest rank from the first set. From the second set, an actual measurement is taken from the unit with the second smallest rank, and the procedure is continued until the unit with the largest rank is chosen for actual measurement from the *nth* set. In this way, we obtain a sample of *n* measured units, one from each set. The cycle may be repeated *r* times until *nr* units have been measured. These nr units are forming the RSS data.

Let $X_{(i:n)j}$ denote the *ith* order statistic from the *ith* set of size *n* in the *jth* cycle. Then the unbiased estimator of the population mean, see Takahasi and Wakimoto (1968) using RSS data based on the *jth* cycle is defined as:

$$\bar{X}_{rss,\,j} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:n)\,j}\,;\, j = 1, 2, ..., r\,. \tag{1}$$

The variance of $\bar{X}_{rss,\,j}$ is given by

$$\operatorname{var}\left(\bar{X}_{rss,\,j}\right)=\frac{1}{n^{2}}\sum_{i=1}^{n}\sigma_{(i:n)}^{2}\,, \tag{2}$$

where $\sigma_{(i:n)}^{2}=E\left[X_{(i:n)}-E\left(X_{(i:n)}\right)\right]^{2}$.

2.2 Ranked set sampling with concomitant variable

Suppose that the variable of interest X is difficult to measure and to order, but there is a concomitant variable Y, which is correlated with X. The variable Y may be used to acquire the rank of X as follows: Select n^2 bivariate units from the population and group them into *n* sets of size *n* each. From the first set of size *n*, the X associated with the smallest Y is measured. From the second set of size *n* the X associated with the second smallest Y is measured. We continue this way until the X associated with the largest Y from the *nth* set is measured. The cycle is repeated *r* times until *nr* Xs have been measured. Note that ranking of the variable X will be with errors in ranking i.e. $X_{[i:n],j}$ is the *ith* judgment order statistic from the *ith* set of size *n* in the *jth* cycle of size *r*. This method is called imperfect ranked set sampling (IRSS).

Assume that (X, Y) has a bivariate normal distribution and the regression of X on Y is linear. Then following Stokes (1977) we can write

$$X=\mu_{x}+\ \rho\frac{\sigma_{x}}{\sigma_{y}}\left(Y-\mu_{y}\right)+\varepsilon \tag{3}$$

where Y and ε are independent and ε has mean 0 and variance $\sigma_{x}^{2}(1-\rho^{2})$, ρ is the correlation between X and Y and $\mu_{x},\mu_{y},\sigma_{x},\sigma_{y}$ are the means and standard deviations of the variables X and Y.

Let $Y_{(i:n),j}$ and $X_{[i:n],j}$ be the *ith* smallest value of Y and the corresponding value of X obtained from the *ith* set in the *jth* cycle respectively. We can write the above equation

$$X_{[i:n],\,j}=\mu_{x}+\frac{\rho\sigma_{x}}{\sigma_{y}}\ (Y_{(i:n),j}-\mu_{y})+\varepsilon_{ij}\,,\quad i=1,\,2,\ldots,\,n,\,j=1,\,2,\ldots,\,r. \tag{4}$$

The unbiased estimator of the mean of the variable of interest X with ranking based on the concomitant variable Y, i.e. using IRSS method, can be written for the *jth* cycle as

$$\bar{X}_{irss,\ j} = \frac{1}{n}\sum_{i=1}^{n} X_{[i:n],j}\,;\ j = 1,2,...,r\ . \tag{5}$$

The variance of $\bar{X}_{irss,j}$ (see Stokes, 1977) is given by

$$\text{var}\left(\bar{X}_{irss,j}\right) = \frac{\sigma_x^2}{n}\left[\ (1-\rho^2) + \frac{\rho^2}{n\sigma_y^2}\sum_{i=1}^{n}\sigma_{y(i:n)}^2\ \right]. \tag{6}$$

2.3 Pair Ranked Set Sampling (PRSS)

In the paired ranked set sampling (PRSS) method, two sets of n random elements are required to obtain a sample of size two. At first n elements are selected randomly and ordered, the k-th smallest element of the set is considered for measurement, where $1 \le k \le n$ is pre-determined, see Hossain and Muttlak (1999). Similarly, second set of size n elements is again selected randomly and ordered, and the (n-k+1)-th smallest of the set is measured. The procedure can be repeated r times to obtain a sample of size 2r. Note that in the usual RSS method the sample size is required to be a multiple of n and in the PRSS method it is required to be a multiple of 2 and does not depend on the choice of the set size n.

Once the value of k is determined, an estimator of the population mean μ for the jth cycle can be written as

$$\bar{X}_{prssj} = \frac{1}{2}(X_{(k:n)1j} + X_{(k':n)2j});\ j = 1,2,...,r \tag{7}$$

where $k' = n - k + 1$. Clearly for a symmetric distribution $\bar{X}_{prssj}$ is an unbiased estimator for μ with variance

$$\text{var}(\bar{X}_{prssj}) = \frac{\sigma^2}{2t^2}\sigma_{(k:n)}^2\ . \tag{8}$$

where $\sigma_{(k:n)}^2 = E\ \ [X_{(k:n)} - \ E(X_{(k:n)})]^2$ and t is known constant depending on the underlying distribution, t = 1 for normal distribution. For more details see Hossain and Muttlak (1999).

2.4. Selected Ranked Set Sampling (SRSS)

Consider the situation where, instead of selecting n random sets of size n elements each as in the RSS, only $k < n$ random set of size n elements are selected, and instead of measuring the ith smallest order statistic of the ith set, n_{ith} smallest order statistic of the ith set is considered for measurement. The values of

$$n_1, n_2, ..., n_k \ (1 \le n_1 < n_2 < ... < n_k \le n)$$

are required to be determined beforehand, see Hossain and Muttlak (2001).

The procedure of selected ranked set sampling (SRSS) can be described as follows: At first, a set of $n > k$ elements is randomly selected and they are ordered by visual inspection and the n_1-th smallest is selected for measurements. Another set of n elements is randomly selected and they are ordered and the n_2-th smallest element is measured, and the procedure is continued until the n_k-th smallest is measured.

Once the values of $n_1, n_2, ..., n_k$ and $c_i, i = 1, 2, ..., k$ are determined (see Hossain and Muttlak, 2001), the SRSS method will be use to collect the data. Let $X_{(n_i:n)j\,;\,j=1,2,...,r}$ is the n_{ith} order statistics of the n_{ith} set of size n in the jth cycle. If the underlining distribution is normal an unbiased estimator of the population mean μ for the jth cycle is

$$\bar{X}_{srssj} = \sum_{i=1}^{k} c_i X_{(n_i:n)j} \tag{9}$$

with

$$c_i = \frac{S_1 - S_2 \alpha_{(n_i:n)}}{D_s \sigma^2_{(n_i:n)}}$$

where $\alpha_{(n_i:n)}$ and $\sigma^2_{(n_i:n)}$ are the expected value and the variance of the n_i -th order statistics for standard normal respectively. Also,

$$S_1 = \sum_{i=1}^{k} \frac{\alpha^2_{(n_i:n)}}{\sigma^2_{(n_i:n)}}, \ S_2 = \sum_{i=1}^{k} \frac{\alpha_{(n_i:n)}}{\sigma^2_{(n_i:n)}}, \ S_3 = \sum_{i=1}^{k} \frac{1}{\sigma^2_{(n_i:n)}} \ \text{and} \ D_s = S_1 S_3 - S_2^2$$

The variance of $\bar{X}_{srssj}$ can be shown to be

$$\text{var}(\bar{X}_{srssj}) = \theta^2 \frac{S_1}{D_s} \tag{10}$$

where θ is the value of the scale parameter for the underlying distribution. For more details see Hossain and Muttlak (2001).

3. QUALITY CONTROL CHARTS

3.1. Quality Control Chart using SRS

Let X_{ij} for $i=1,2,...,n$ and $j=1,2,...,r$ denote the *ith* unit in the *jth* SRS of size *n* and $X_{ij} \sim N(\mu,\sigma^2)$. If the population mean μ and variance σ^2 are known then the Shewhart control chart for

$$\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij};\, j=1,2,...,r \tag{11}$$

is given by

$$UCL = \mu + 3\frac{\sigma}{\sqrt{n}}$$
$$CL = \mu$$
$$LCL = \mu - 3\frac{\sigma}{\sqrt{n}} \tag{12}$$

where UCL, CL and LCL denote the upper central limit, central limit and lower central limit respectively. The sample means $\bar{X}_j, j=1,2,...,r$ can be plotted in the above charts.

For this chart the average run length (ARL) is equal to 1/α, where α is the probability of type I error if the process is under control. But if the process starts to get out of control then ARL= 1/β, where β is the probability of type II error, see Montgomery (1995).

In most real life problems μ and variance σ^2 are unknown. We use the collected data to estimate μ and σ, obviously the unbiased estimator for μ is

$$\bar{\bar{X}} = \frac{1}{r}\sum_{j=1}^{r} \bar{X}_j\ . \tag{13}$$

But

$$\bar{S} = \frac{1}{r}\sum_{j=1}^{r} S_j \tag{14}$$

where $S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)^2}$ is a biased estimate for σ. We can use $\bar{S}/c_4$ as an unbiased estimator for σ, where c_4 can be shown to equal

$\left(2/n-1\right)^2 \dfrac{\Gamma(n-2)}{\Gamma[(n-1)/2]}$. We may now define the control limits for the sample mean $\bar{X}_j$ to be

$$UCL = \bar{\bar{X}} + \frac{3\bar{S}}{c_4\sqrt{n}}$$
$$CL = \bar{\bar{X}}$$
$$LCL = \bar{\bar{X}} - \frac{3\bar{S}}{c_4\sqrt{n}} \qquad (15)$$

After establishing the above chart, the sample means $\bar{X}_j ; j = 1,2,...,r$ are plotted in the chart. For more details see Montgomery (1995).

3.2. Quality Control Charts using PRSS

The PRSS mean $\bar{X}_{prssj}$ of the jth cycle defined on Section 2.3 can be plotted on the control chart based on PRSS

$$UCL = \mu + 3\sigma_{\bar{X}_{prss}}$$
$$CL = \mu$$
$$LCL = \mu - 3\sigma_{\bar{X}_{prss}} \qquad (16)$$

where $\sigma_{\bar{X}_{prss}} = \sqrt{\frac{\sigma^2}{2}\sigma^2_{(k:n)}}$ and $\sigma^2_{(k:n)}$ as defined in Section 2.3.

We use the average run length (ARL) to compare the PRSS control charts to the Shewhart control chart. The ARL assumes that the process is under control with mean μ_o and standard deviation σ_o, and at some point in time the process may start to get out of control i.e. the mean is shifted from μ_0 to $\mu_o + \delta\sigma_o/\sqrt{n} = \mu$. We are assuming that the process is following the normal distribution with mean μ_0 and variance σ_o^2 if the process is under control, and the shift on the process mean is $\delta = \frac{\sqrt{n}}{\sigma_o}\left|\mu - \mu_0\right|$. If $\delta = 0$ the process is under control and in this case if the point is outside the control limits it is a false alarm.

The simulation is done only for the rule: a point out of control limits, see Champ and Woodall (1987). For each value of δ we simulate 1,000,000 replications. We calculate the values of the limits in equation (16) using the results of the order statistics for the standard normal distribution, see for example Harter and Balakrishnan (1996).

Ranking the variable of interest without errors in ranking the units is called perfect ranking. But if the units cannot be raked perfectly or the ranking is done on a concomitant variable we call that imperfect ranking, see Section 2.2. Since the perfect ranking and SRS are special cases of the imperfect ranking with $\rho = 1$ and $\rho = 0$ respectively, we will consider the case of imperfect ranked set sampling (IPRSS) with different values of ρ. Following the same procedure that we used in Section 1.2, we only need equation (6) to perform our simulation, which can be written as

$$\mathrm{var}(\bar{X}_{iprssj}) = \frac{\sigma_x^2}{2}\left[(1-\rho^2)+\rho^2\sigma_{z(k:n)}^2\right] \tag{17}$$

where σ_x^2 is the variance of the variable of interest X and $\sigma_{z(k:n)}^2$ is the variance of the k-th order statistic of a standard normal distribution.

The control chart given in equation (16) is based on the perfect PRSS, we need to modify it to the case of imperfect ranking by substituting for the variance of $\bar{X}_{iprssj}$ given in equation (17) to get

$$\begin{aligned} UCL &= \mu + 3\sigma_{\bar{X}_{iprss}} \\ CL &= \mu \\ LCL &= \mu - 3\sigma_{\bar{X}_{iprss}} \end{aligned} \tag{18}$$

where $\sigma_{\bar{X}_{iprss}} = \sqrt{\frac{\sigma_x^2}{2}\left[(1-\rho^2)+\rho^2\sigma_{z(k:n)}^2\right]}$. Note that if the ranking of the units is done perfectly, i.e. there are no errors in ranking then we set $\rho = 1$ in equation (18).

In our simulation we considered both X and Y as standard normal random variable, this implies $\sigma_x^2 = 1$. The computer simulations are run for $\rho = 0, 0.25, 0.5, 0.75, 0.9, 1.0$, n = 3, 4, 5, 6 and for δ = 0, 0.1, 0.2, 0.3, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 3.2. Results are shown in Tables 1- 4. Considering these results the following remarks are made.

1. If the process is under control, i.e. $\delta = 0$, PRSS reduces the false alarm to ARL= 355.89. But in the case PRSS we are only measuring two units, one form each set of size n, i.e. a sample size of 2. For example if n=4 for all other cases we have to measure 4 units, one from each set of size 4, but for the case of PRSS we are only measuring 2 units, one unit from two sets of size 4 units.
2. The PRSS method is dominated SRS if the process starts to get out of control i.e. $\delta>0$, for example if $\delta=0.4$ and n = 4 the ARL using PRSS is 99 as compare to 200.01, for SRS.
3. If the sample size increases there will not be much of a change in the ARL if $\delta=0$. But the ARL will keep decreasing if $\delta>0$, for example if the sample size is 5 and $\delta=0.4$ the ARL is 81 as compared to 99 in the case of n=4.
4. Imperfect ranking decreases the efficiency of PRSS and the ARL will be larger which depend on the values of ρ. But PRSS is still doing better than SRS method by having smaller ARL for the same value of ρ.
5. The ARL for the PRSS will decrease much faster than SRS if δ increases.

3.3 Quality Control Charts using SRSS

The SRSS mean $\bar{X}_{srssj}$ of the jth cycle defined on Section 2.4 can be plotted on the control chart based on SRSS, which can be defined as follows:

$$\begin{aligned} UCL &= \mu + 3\sigma_{\bar{X}_{srss}} \\ CL &= \mu \\ LCL &= \mu - 3\sigma_{\bar{X}_{srss}} \end{aligned} \tag{19}$$

where $\sigma_{\bar{X}_{srss}} = \sqrt{\sigma^2 \sum_{i=1}^{k} c_i^2 \sigma_{(n_i:n)}^2}$.

As we did in the pervious sections, we used the ARL to compare the selected ranked set sampling (SRSS) control charts to the other control charts. We used the same values for δ, ρ and n that we used in the previous sections, and we run our simulation for 1,000, 000 replications. Following the same procedure that we used in Section 3.2, but here we use SRSS instead of PRSS i.e. imperfect selected ranked set sampling (ISRSS). The analogues of equations (17) and (18) are

$$\operatorname{var}(\bar{X}_{isrssj}) = \sigma_x^2 \left[(1-\rho^2) + \rho^2 \sum_{i=1}^{k} c_i^2 \sigma_{z(n_i:n)}^2 \right] \tag{20}$$

and

$$UCL = \mu + 3\,\sigma_{\overline{X}_{isrss}}$$
$$CL = \mu$$
$$LCL = \mu - 3\sigma_{\overline{X}_{isrss}} \qquad (21)$$

where $\sigma_{\overline{X}_{isrss}} = \sqrt{\sigma_x^2\left[(1-\rho^2)+\rho^2\sum_{i=1}^{k} c_i^2\sigma_{z(n_i:n)}^2\right]}$ and $\sigma_{z(n_i:n)}^2$ is the variance of the n_{ith} order statistic of a standard normal distribution.

We considered in our simulation both X and Y as standard normal random variable. The computer simulation is run for the same values of ρ, δ and n that considered before and for different values of k. Results are shown in Tables 5- 14. Considering the results of Tables 5-14, the following remarks can be made:

1. If the process under control i.e. $\delta = 0$ SRSS is dominated SRS and PRSS methods in reducing the number of false alarm, i.e. reducing ARL. Please note that in SRSS we are only measuring k units out of the n units in each set, where k < n.
2. If the number of measured units k remains constant, but the sample size n increases then the ARL will be decrease as n increases. For example if δ=0.4, k=2 and if n=3, 4, 5, or 6 then the corresponding ARL is 133.51, 120.51, 110.79, 103.62 respectively. In the other we do not see this pattern if n remains constant and k increases.
3. The SRSS reduces the ARL over SRS and PRSS for must cases considered in this study if the process starts to get out of control.
4. Imperfect ranking decreases the efficiency of SRSS, as it is the case for other methods.

4. CONCLUSIONS AND RECOMMENDATIONS

The ranked set sampling has attracted a number of authors as an efficient sampling method. The RSS method that proved to be more efficient when units are difficult and costly to measure, but are easy and cheap to rank with respect to a variable of interest without actual measurement. In this study we used two modifications of RSS to develop several quality control charts for the variables of interest using the sample mean. These charts are compared with the classical control charts using simple random sampling data. It is clear that all the newly developed charts are more efficient than the classical control chart. The following are some specific conclusions.

1. All the newly developed control charts dominate the classical charts. If the process starts to get out of control by reducing the number of average run

length (ARL) substantially. But number of false alarms is not reduced by the same amount if the process is under control.

2. Errors in ranking will reduce the ARL for both cases considered. The amount of reduction in the ARL will depend on the amount of errors committed in ranking the units of the variable of interest. For example if we are using a concomitant variable to rank our variable, then the ARL will depend on the correlation between the two variables.

Finally we recommend using the PRSS and/ or SRSS to build the quality control charts. Since they are reducing the ARL comparing to SRS method.

ACKNOWLEDGMENTS

This work was supported by King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia under the fast track project # FT /2000-15.

REFERENCES

1. Champ, C.W. and Woodall, W.H. (1987). Exact results for Shewhart charts with supplementary run rules. *Technometrics*, 29, 393-399.
2. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics* 28, 545-553.
3. Harter, H.L. and Balakrishnan, N. (1996). *CRC Handbook of tables for the use of order statistics in estimation*, CRC press Boca Raton.
4. Hossain, S.S. and Muttlak, H.A. (1999). Paired Ranked Set Sampling: A More Efficient Procedure. *Environmetrics*. Vol. 10. 195-212.
5. Hossain, S.S. and Muttlak, H.A. (2001). Selected Ranked Set Sampling. *Aust. & New Zealand J. Statist.,* 43. 311-325.
6. McIntyre, G.A. (1952). A method for unbiased selective sampling, using, ranked sets. *Aust. J. Agri. Res.,* 3, 385-390.
7. Montgomery, D.C. (1995). *Statistical Quality Control*. New York: Wiley.
8. Muttlak, H.A. (1997). Median ranked set sampling. *J. App. Statist. Scien.,* 6, 245-255.
9. Muttlak, H.A. and AL-Sabah (2001) *Statistical quality control using ranked set sampling*. Technical Report No. FT/2000-15, Department of Mathematical Sciences, KFUPM.
10. Samawi, H., Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling. *The Biometrical Journal*, 30**,** 577-586.
11. Stokes, S.L. (1977). Ranked set sampling with concomitant variables, *Commun. in Statist. - Theo. and Meth.,* A6, 1207-1211.
12. Takahasi, K. and Wakimoto, K. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.* 20, 1-31.

Table 1
ARL values when n=3 using PRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.24	369.89	368.25	367.08	362.18	355.89
0.1	347.63	350.53	354.31	346.62	336.38	321.85
0.2	310.71	309.44	298.88	288.37	270.55	250.92
0.3	251.51	247.35	241.02	225.41	204.66	184.51
0.4	200.36	197.06	185.41	162.02	139.47	118.60
0.8	71.51	69.04	61.35	47.91	36.78	28.12
1.2	27.84	26.53	22.79	16.70	12.09	8.83
1.6	12.39	11.75	9.92	7.07	5.06	3.70
2.0	6.30	5.98	5.03	3.61	2.65	2.02
2.4	3.64	3.47	2.95	2.19	1.69	1.39
3.2	1.73	1.66	1.48	1.25	1.11	1.04

Table 2
ARL values when n=4 using PRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.12	369.52	369.79	366.84	364.41	351.63
0.1	345.54	349.04	347.94	339.99	330.49	308.13
0.2	311.08	300.48	300.53	279.65	258.41	230.08
0.3	251.76	251.72	238.22	212.24	182.88	151.45
0.4	201.12	195.97	182.83	155.20	126.15	98.99
0.8	71.56	68.60	59.71	43.78	30.88	20.97
1.2	27.82	26.38	22.02	14.98	9.83	6.33
1.6	12.38	11.65	9.53	6.30	4.11	2.71
2.0	6.30	5.92	4.84	3.24	2.21	1.59
2.4	3.65	3.44	2.85	2.00	1.48	1.19
3.2	1.73	1.65	1.45	1.19	1.06	1.01

Table 3
ARL values when n=5 using PRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	369.80	368.46	370.50	367.84	367.92	355.40
0.1	352.84	348.85	348.22	339.38	328.12	304.61
0.2	308.58	305.24	298.86	274.41	247.83	208.77
0.3	252.14	250.15	235.47	207.47	170.65	130.62
0.4	199.99	195.60	180.46	147.37	114.38	81.09
0.8	71.55	68.49	58.15	40.54	26.02	15.18
1.2	27.82	26.24	21.41	13.60	8.05	4.47
1.6	12.39	11.59	9.23	5.70	3.39	2.02
2.0	6.30	5.88	4.68	2.95	1.88	1.30
2.4	3.65	3.41	2.76	1.85	1.32	1.07
3.2	1.73	1.64	1.42	1.15	1.03	1.00

Table 4
ARL values when n=6 using PRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.04	371.35	369.86	367.38	365.68	354.12
0.1	352.76	352.92	348.23	339.98	323.76	294.75
0.2	307.94	307.06	295.46	272.13	239.73	193.56
0.3	253.67	248.76	234.94	201.68	161.39	116.45
0.4	200.51	196.20	179.64	144.66	106.93	69.67
0.8	71.71	68.22	57.43	38.64	23.29	12.12
1.2	27.84	26.12	20.95	12.83	7.11	3.57
1.6	12.38	11.53	9.05	5.37	3.02	1.70
2.0	6.31	5.85	4.59	2.80	1.72	1.18
2.4	3.64	3.40	2.72	1.77	1.25	1.03
3.2	1.73	1.64	1.41	1.13	1.02	1.00

Table 5
ARL values when n =3 and k=2 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.6	370.22	369.13	362.63	330.69	324.73
0.1	352.87	352.83	350.06	339.51	321.79	299.90
0.2	307.90	307.07	301.88	286.28	266.81	245.59
0.3	252.54	250.54	242.22	225.58	204.78	184.75
0.4	200.18	197.29	188.62	168.77	151.01	133.51
0.8	71.63	69.54	63.27	52.56	44.29	36.96
1.2	27.83	26.79	23.83	18.89	15.11	12.36
1.6	12.38	11.88	10.40	8.07	6.37	5.17
2.0	6.31	6.04	5.28	4.10	3.27	2.69
2.4	3.65	3.56	3.08	2.45	2.01	1.71
3.2	1.73	1.68	1.53	1.32	1.19	1.11

Table 6
ARL values when n = 4 and k= 2 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	373.01	370.01	369.62	359.51	334.67	311.88
0.1	354.50	353.99	350.55	335.67	314.99	284.88
0.2	309.60	308.89	302.21	284.37	257.01	229.46
0.3	252.04	250.26	241.70	218.66	193.21	168.30
0.4	201.30	197.72	186.29	163.99	144.67	120.51
0.8	71.38	64.43	62.30	49.81	39.29	31.41
1.2	27.78	26.60	23.19	17.54	13.22	10.23
1.6	12.39	11.79	10.10	7.46	5.58	4.27
2.0	6.30	5.99	5.12	3.80	2.88	2.27
2.4	3.64	3.48	3.00	2.29	1.810	1.50
3.2	1.73	1.67	1.50	1.27	1.14	1.06

Table 7
ARL values when 4 = and k= 3 using SRSS

δ	ρ					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.22	369.97	368.89	365.15	355.64	334.57
0.1	352.85	352.13	348.52	338.31	317.94	291.43
0.2	311.52	308.78	299.70	282.13	253.45	221.51
0.3	256.31	253.35	243.42	218.19	187.41	157.68
0.4	203.84	200.06	186.94	160.30	133.80	108.81
0.8	74.17	71.54	63.19	48.44	36.50	27.30
1.2	29.08	27.77	23.71	17.14	12.25	8.81
1.6	13.04	12.35	10.36	7.31	5.15	3.72
2.0	6.64	6.28	5.26	3.73	2.70	2.03
2.4	3.83	3.63	3.08	2.26	1.72	1.39
3.2	1.79	1.72	1.53	1.26	1.11	1.04

Table 8
ARL values when n =5 and k=2 using SRSS

δ	ρ					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.05	369.94	365.95	358.34	335.26	298.28
0.1	353.24	352.20	348.81	336.12	308.75	272.75
0.2	308.85	304.66	299.13	277.84	249.34	216.69
0.3	253.81	250.91	240.36	214.86	187.16	157.44
0.4	199.58	196.35	185.39	160.21	134.48	110.79
0.8	71.31	69.01	61.29	47.62	36.53	27.85
1.2	27.81	26.54	22.82	16.65	12.11	8.90
1.6	12.38	11.76	9.91	7.07	5.06	3.72
2.0	6.30	5.97	5.03	3.61	2.65	2.02
2.4	3.65	3.46	2.95	2.19	1.69	1.38
3.2	1.73	1.66	1.49	1.24	1.11	1.04

Table 9
ARL values when n=5 and k=3 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	372.87	369.86	367.7	363.01	348.57	323.29
0.1	354.60	352.99	347.10	335.60	309.47	271.09
0.2	311.36	307.58	300.10	275.48	241.55	199.83
0.3	256.38	253.42	241.39	213.75	176.78	139.21
0.4	205.00	200.51	188.55	158.24	126.23	96.11
0.8	75.54	72.46	63.45	47.20	33.93	23.87
1.2	29.82	28.33	23.86	16.68	11.36	7.72
1.6	13.42	12.66	10.47	7.11	4.80	3.294
2.0	6.84	6.44	5.31	3.64	2.53	1.83
2.4	3.94	3.72	3.10	2.21	1.63	1.29
3.2	1.83	1.75	1.54	1.25	1.09	1.02

Table 10
ARL values when n = 5 and k=4 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	372.47	368.64	366.99	364.26	362.16	348.20
0.1	351.33	349.21	348.03	342.50	329.94	308.25
0.2	309.05	308.45	300.75	284.07	258.36	227.41
0.3	255.13	253.04	240.86	215.86	185.43	152.76
0.4	202.01	198.94	184.96	158.21	128.18	100.14
0.8	73.70	70.93	61.74	45.38	31.92	21.52
1.2	29.01	27.46	22.94	15.59	10.21	6.52
1.6	12.96	12.22	9.98	6.58	4.26	2.78
2.0	6.62	6.21	5.06	3.37	2.28	1.62
2.4	3.82	3.59	2.97	2.07	1.51	1.20
3.2	1.79	1.71	1.49	1.21	1.06	1.01

Table 11
ARL values when n =6 and k=2 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	372.52	371.78	369.95	358.49	332.95	289.92
0.1	353.23	351.98	347.21	332.45	307.07	265.01
0.2	307.35	305.87	296.31	278.10	246.38	207.39
0.3	253.90	249.65	241.71	212.74	181.76	149.27
0.4	200.14	196.75	183.30	157.925	129.91	103.62
0.8	71.42	68.96	60.83	46.29	34.38	25.34
1.2	27.84	26.48	22.50	16.05	11.29	7.98
1.6	12.39	11.72	9.77	6.79	4.72	3.34
2.0	6.31	5.96	4.96	3.47	2.46	1.85
2.4	3.64	3.45	2.91	2.12	1.61	1.30
3.2	1.73	1.66	1.47	1.22	1.09	1.03

Table 12
ARL values when n =6 and k=3 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.04	369.76	367.89	362.41	346.92	310.48
0.1	352.78	354.77	349.78	332.40	302.54	254.29
0.2	311.22	308.40	300.11	273.01	233.97	185.00
0.3	258.84	254.20	243.29	210.07	170.27	128.22
0.4	206.11	203.18	187.42	155.27	120.69	87.55
0.8	76.33	73.30	63.72	46.44	32.18	21.68
1.2	30.32	28.76	24.06	16.36	10.78	7.01
1.6	13.66	12.85	10.53	6.98	4.55	3.01
2.0	6.97	6.55	5.35	3.57	2.42	1.71
2.4	4.02	3.78	3.12	2.17	1.58	1.24
3.2	1.86	1.78	1.545	1.24	1.08	1.01

Table 13
ARL values when n = 6 and k= 4 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	369.57	363.87	359.21	354.24	350.95	342.11
0.1	331.08	329.61	330.43	325.41	316.20	300.93
0.2	290.38	288.79	282.81	268.14	245.72	216.43
0.3	240.96	237.41	227.43	202.43	174.16	143.19
0.4	202.58	187.45	175.75	148.52	120.01	91.92
0.8	73.87	67.60	58.56	42.56	29.17	18.91
1.2	29.01	26.42	21.90	14.64	9.28	5.66
1.6	12.99	11.83	9.60	6.20	3.90	2.46
2.0	6.61	6.06	4.90	3.21	2.12	1.47
2.4	3.82	3.53	2.89	1.99	1.43	1.14
3.2	1.79	1.69	1.47	1.19	1.05	1.01

Table 14
ARL values when n = 6 and k= 5 using SRSS

δ	q					
	0.0	0.25	0.50	0.75	0.90	1.00
0.0	370.04	369.05	368.40	366.27	364.19	352.12
0.1	352.54	355.13	353.29	340.82	325.04	299.51
0.2	310.23	309.37	302.17	278.18	250.93	209.47
0.3	256.30	253.28	242.37	209.18	175.66	135.92
0.4	203.48	200.58	186.31	153.12	119.64	86.12
0.8	75.01	71.83	61.48	43.32	28.36	17.12
1.2	29.57	27.83	22.84	14.75	8.93	5.12
1.6	13.25	12.39	9.93	6.21	3.75	1.39
2.0	6.75	6.30	5.04	3.20	2.04	1.11
2.4	3.89	3.65	2.95	1.98	1.47	1.02
3.2	1.81	1.73	1.49	1.43	1.04	1.00

CHAPTER TEN

Modified Ranked Set Sampling Methods

Hassen A. Muttlak
Department Mathematical Sciences
KFUPM, Dhahran 31261, Saudi Arabia

ABSTRACT

The ranked set sampling method (RSS) as suggested by McIntyre (1952) may be modified to yield new sampling methods with improved. Several modifications for the RSS are introduced by several authors such as extreme ranked set sampling (ERSS), suggested by Samawi et al. (1996), median ranked set sampling (MRSS), suggested by Muttlak (1997), etc. In this study a few other modifications for the RSS are introduced and compared to the RSS, ERSS and MRSS. It turns out that for probability distributions considered in this study, we can always improve upon the efficiency of RSS by using some sort of modification for the RSS method.

KEY WORDS

Extreme ranked set sampling, median ranked set sampling, simple random sampling, percentile ranked sampling and relative precision.

1. INTRODUCTION

Ranked set sampling (RSS) was first suggested by McIntyre (1952) without the mathematical theory to support his suggestion. Takahasi and Wakimoto (1968) supplied the necessary mathematical theory. They proved that the sample mean of the ranked set sample (RSS) is an unbiased estimator of the population mean with smaller variance than the sample mean of a simple random sample (SRS) with the same sample size. Dell and Clutter (1972) studied the case in which the ranking may not be perfect i.e., there are errors in ranking the units. Muttlak (1996) suggested using pair ranked set sampling instead of RSS. This can be used when it is difficult to select a large number of units from the population of interest. Samawi et al. (1996) suggested using extreme ranked set sampling (ERSS) to estimate the population mean. They showed that the ERSS estimator is an unbiased estimator of the population mean if the underlying distribution is symmetric and it is more efficient than the SRS estimator. Muttlak (1997) suggested using median ranked set sampling (MRSS) to estimate the population mean more efficiently than the usual RSS method. For review and more bibliography on the RSS see Patil et al. (1999).

In this paper, a further modification of the RSS method is considered, namely, percentile ranked sampling (PRSS) with different values of $0 \leq p \leq 1$. The newly suggested sampling method is compared with RSS, ERSS and MRSS. It is shown that for the probability distributions considered in this study, we can always improve the relative precision and reduce the errors in ranking by using the modified sampling method instead of the usual RSS method.

2. NOTIONS AND SOME USEFUL RESULTS

Let $X_1, X_2, ..., X_n$ be a random sample with probability density function f(x) with a finite mean μ and variance σ^2. Let $X_{11}, X_{12}, ..., X_{1n}; X_{21}, X_{22}, ..., X_{2n}; ...; X_{n1}, X_{n2}, ..., X_{nn}$ be independent random variables all with the same cumulative distribution function F(x). Let $X_{(i:n)}$ denotes the i^{th} order statistic from the i^{th} sample of size n (i = 1, 2, ..., n). The unbiased estimator of the population mean using RSS is defined as

$$\overline{X}_{rss} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:n)} .$$

The variance of $\overline{X}rss$ is given by

$$\text{var}(\overline{X}_{rss}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:n)} ,$$

where $\sigma^2_{(i:n)} = E \ [X_{(i:n)} - \ E(X_{(i:n)})]^2$.

Let $X_{(i:\ e)}$, denote the smallest of the i^{th} sample (i = 1, 2, ..., L= n/2) and the largest of the i^{th} sample (i = L+1, L+2, ..., n) if the sample size n is even. Also denote the smallest of the i^{th} sample (i = 1, 2, ..., L_1= (n-1)/2), the median of the i^{th} sample (i = (n+1)/2) and the largest of the i^{th} sample (i = L_1+2, L_1+3, ..., n) if the sample size n is odd. The estimator of the population mean based on ERSS with one cycle can be written as

$$\overline{X}_{errs} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:e)} .$$

The variance of $\overline{X}_{errs}$ can be written as

$$\text{var}(\overline{X}_{errs}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:e)} ,$$

where $\sigma^2_{(i:n)} = E \ [X_{(i:e)} - \ E(X_{(i:e)})]^2$. For more details, see Samawi et al (1996).

Let $X_{(i:\ m)}$, denote the median of the i^{th} sample if the sample size is odd, and the $(n/2)^{th}$ order statistic of the i^{th} sample (i = 1, 2, ..., L= n/2) and the $((n+2)/2)^{th}$ order statistic of the i^{th} sample (i = L+1, L+2, ..., n) if the sample size is even. The estimator of the population mean using MRSS then can be written as

$$\overline{X}_{mrss} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:\ m)} \ .$$

The variance of $\overline{X}_{mrss}$ can be written as

$$\text{var}(\overline{X}_{mrss}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:m)}$$

where $\sigma^2_{(i:m)} = E\ \ [X_{(i:m)} - \ \ E(X_{(i:m)})]^2$. For more details, see Muttlak (1997).

3. PERCENTILE RANKED SET SAMPLING

In the percentile ranked set sampling (PRSS) procedure, select n random samples of size n units from the population and rank the units within each sample with respect to a variable of interest. If the sample size is even, select for measurement from the first n/2 samples the $(p(n+1))^{th}$ smallest rank and from the second n/2 samples the $(q(n+1))^{th}$ smallest rank, where $0 \le p \le 1$ and q = 1-p. If the sample size is odd, select from the first (n-1)/2 samples the $(p(n+1))^{th}$ smallest rank and from the other (n-1)/2 samples the $(q(n+1))^{th}$ smallest rank, and from one sample the median for that sample for actual measurement. The cycle may be repeated r times to get nr units. These nr units form the PRSS data.

Let X_{11}, X_{12}, ..., X_{1n}; X_{21}, X_{22}, ..., X_{2n}; ...; X_{n1}, X_{n2}, ..., X_{nn} be independent random variables all with the same cumulative distribution function F(x). Let $X_{i\ (p(n+1))}$ and $X_{i\ (q(n+1))}$ denote the $(p(n+1))^{th}$ order statistic $(q(n+1))^{th}$ order statistic of the i^{th} sample respectively (i = 1, 2, ..., n), where $0 \le p \le 1$ and q = 1-p. The estimator of the population mean using percentile ranked set sample (PRSS) with one cycle can be defined in the case of an even sample size as

$$\overline{X}_{prss1} = \frac{1}{n}\left(\sum_{i=1}^{L_1} X_{i(p(n+1))} + \sum_{i=L_1+1}^{n} X_{i(q(n+1))}\right),$$

where L_1 = n/2. In the case of an odd sample size, the estimator of the population mean can be defined as

$$\overline{X}_{prss2} = \frac{1}{n}\left(\sum_{i=1}^{L_2} X_{i(p(n+1))} + \sum_{i=L_2+2}^{n} X_{i(q(n+1))} + X_{i((n+1)/2)}\right),$$

where $L_2 = (n-1)/2$ and $X_{i\,((n+1)/2)}$ is the median of sample i = (n+1)/2

The variance of $\overline{X}_{prss}$ can be written as

$$\text{var}(\overline{X}_{prss}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2_{(i:p)}$$

where $\sigma^2_{(i:p)} = E\ [X_{(i:p)} - E(X_{(i:p)})]^2$. Here $X_{(i:p)}$ is the $p_{(n+1)}{}^{th}$ order statistic of the i^{th} sample.

Let $\overline{X}_{srs}$ denote the sample mean of simple random sample (SRS) of size n. The properties of $\overline{X}_{prss}$ are

1. $\overline{X}_{prss}$ is an unbiased estimator of the population mean μ if the underling distribution is symmetric about the population μ and
2. Var ($\overline{X}_{prss}$) is less than Var ($\overline{X}_{srs}$).
3. If the distribution is not symmetric about μ than the mean square error (MSE) of $\overline{X}_{prss}$ is less than the variance of $\overline{X}_{srs}$.

It is not difficult to prove (1)-(3) using the results by Takahasi and Wakimoto (1968), Samawi et al (1996) and Muttlak (1997).

To compare the proposed estimators for the population mean using PRSS with RSS, ERSS, MRSS and SRS methods, eight probability distribution functions were considered: rectangular, normal, exponential, gamma, weibull, double exponential, inverse Gaussian and lognormal. The variance or the mean square error of the sample means for the RSS, ERSS, MRSS and PRSS with different values of p were calculated for the above distributions using the moments of the order statistics, see Harter and Balakrishnan (1996) and Balakrishnan and Chen (1997). The relative precision (RP) of estimating the population mean using any of the RSS based methods with respect to the usual estimator using SRS is defined as following

$$RP(\overline{X}_{srs}, \overline{X}_{rss}) = \frac{Var(\overline{X}_{srs})}{Var(\overline{X}_{rss})},$$

if the distribution is symmetric and

$$RP(\overline{X}_{srs}, \overline{X}_{rss}) = \frac{Var(\overline{X}_{srs})}{MSE(\overline{X}_{rss})},$$

if the distribution is not symmetric.

Results are summarized by the relative precision (RP) and the bias in Tables I-III for RSS ERSS, MRSS and PRSS with p = 20%, 30% and 40%. For each population calculations were done with sample size n = 8 in Table I, n =9 in Table II and n=10 inn Table III. Considering the results in Tables I-III, a gain in efficiency is obtained by using PRSS for different values of n and for all the distributions considered in this study. For example, for n = 10 in Table III and p = 0.3 the relative precision (RP) of the PRSS is 5.329 for estimating the population mean of a weibull distribution with shape parameter 2.5.

4. PERCENTILE RANKED SET SAMPLING WITH ERRORS IN RANKING

Dell and Clutter (1972) considered the case in which there are errors in ranking; that is, the quantified observation from the i^{th} sample in the j^{th} cycle may be not the i^{th} order statistic but rather the i^{th} judgment order statistic. They showed that the sample mean of RSS with errors in ranking is an unbiased estimator of the population mean μ, regardless of the errors in ranking and has a smaller variance than the usual estimator based on SRS with the same sample size.

Let $X_{i\,[p(n+1)]}$ and $X_{i\,[q(n+1)]}$ denote the $[p(n+1)]^{th}$ and $[q(n+1)]^{th}$ judgment order statistics respectively, of the i^{th} sample (i = 1, 2, ..., n), where $0 \le p \le 1$ and q = 1-p. If the cycle is repeated once, the estimator of the population mean using percentile ranked set sample (PRSS) with errors in ranking can be defined in the case of an even sample size as

$$\tilde{X}_{prsse1} = \frac{1}{n}\left(\sum_{i=1}^{L_1} X_{i[p(n+1)]} + \sum_{i=L_1+1}^{n} X_{i[q(n+1)]} \right),$$

where $L_1 = n/2$. In the case of an odd sample size, the estimator of the population mean can be defined as

$$\tilde{X}_{prsse2} = \frac{1}{n}\left(\sum_{i=1}^{L_2} X_{i[p(n+1)]} + \sum_{i=L_2+2}^{n} X_{i[q(n+1)]} + X_{i[(n+1)/2]} \right),$$

where $L_2 = (n-1)/2$ and $X_{i\,[(n+1)/2]}$ is the judgment median of sample i = (n+1)/2.

The variance of $\tilde{X}_{prsse}$ can be written as

$$\text{var}(\tilde{X}_{prsse}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{[i:p]}$$

where $\sigma^2_{[i:p]} = E\left[X_{[i:p]} - E(X_{[i:p]})\right]^2$. Here $X_{[i:p]}$ is the $p_{(n+1)}{}^{th}$ judgment order statistic of the i^{th} sample.

Let $\overline{X}_{srs}$ denote the sample means of simple random sample (SRS) of size n. The properties of $\tilde{X}_{prsse}$ are

1. $\tilde{X}_{prsse}$ is an unbiased estimator of the population mean μ if the underling distribution is symmetric about μ and
2. Var ($\tilde{X}_{prsse}$) is less than Var ($\overline{X}_{srs}$).
3. If the distribution is not symmetric about μ than the mean square error (MSE) of $\tilde{X}_{prsse}$ is less than the variance of $\overline{X}_{srs}$.

It is not difficult to prove a-c, using the results by Takahasi and Wakimoto (1968), Dell and Clutter (1972) Samawi et al (1996) and Muttlak (1997).

5. ACKNOWLEDGMENTS

The author is indebted to King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia for excellent research facilities. I am thankful for the referee for healthy comments.

REFERENCES

1. Balakrishnan, N., and Chen, W.W.S. (1997). *CRC Handbook of Tables for Order Statistics from Inverse Gussian Distributions with Aplications*, CRC Press, Boca Raton.
2. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics* 28, 545-555.
3. Harter, H.L. and Balakrishnan, N. (1996). *CRC Handbook of Tables for the use of Order Statistics in Estimation*. CRC Press, Boca Raton.
4. McIntyre, G.A., (1952). A method of unbiased selective sampling, using ranked sets. *Aust. J. Agri. Res.,* 3, 385-390.
5. Muttlak, H.A. (1996). Pair ranked set sampling. *The Biometrical Journal,* 38, 879-885.
6. Muttlak, H.A. (1997). Median ranked set sampling, *J. App. Statist. Scien.*, 6.
7. Patil, G.P., Sinha, A.K., and Taillie, C. (1999). Ranked set sampling: A bibliography. *Environmental and Ecological Statistics*, 6, 91-98.
8. Samawi, H., Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling. *The Biometrical Journal*, 30, 577-586.
9. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.*, 20, 1-31.

Table I
Summary of the relative precision (RP) values for estimating the population mean using RSS, ERSS, MRSS, and PRSS with values of p = 0.2, 0.3 and 0.4, with sample size n = 8.

Distribution	RSS	ERSS	MRSS	PRSS	20%	30%	40%
Uniform (0, 1)	RP Bias	4.50	8.348	3.375	4.821	3.750	3.375
Normal (0, 1)	RP Bias	3.999	2.682	5.342	4.177	4.981	5.342
Exponential (1)	RP Bias	2.943	0.457 0.421	1.673 0.241	3.545 0.007	2.426 0.174	1.673 0.241
Gamma (2)	RP Bias	3.354	0.725 0.345	2.404 0.253	3.812 0.009	3.126 0.184	2.404 0.253
Gamma (3)	RP Bias	3.535	0.939 0.453	2.903 0.257	3.921 0.009	3.533 0.182	2.903 0.257
Gamma (5)	RP Bias	3.702	1.253 0.459	3.524 0.259	4.018 0.011	4.044 0.187	3.524 0.259
Lognormal (0, 1)	RP Bias	1.891	0.279 1.037	1.814 0.538	4.068 0.083	2.586 0.416	1.814 0.538
Double Exponential (0,1)	RP Bias	3.124	1.309	9.509	3.768	6.863	9.509
Inverse Gaussian (0.5)	RP Bias	3.885	1.951 0.115	4.633 0.065	4.124 0.003	4.641 0.047	4.633 0.065
Inverse Gaussian (1)	RP Bias	3.603	1.127 0.219	3.432 0.124	3.989 0.006	3.932 0.090	3.432 0.124
Inverse Gaussian (1.5)	RP Bias	3.262	0.715 0.308	2.563 0.173	3.821 0.008	3.276 0.127	2.563 0.173
Inverse Gaussian (2.5)	RP Bias	2.657	0.404 0.434	1.747 0.241	3.546 0.015	2.495 0.178	1.747 0.241
Weibull (0.5)	RP Bias	1.665	0.222 2.472	1.478 1.272	3.750 0.207	2.164 0.993	1.478 1.272
Weibull (1.5)	RP Bias	3.647	0.972 0.163	2.718 0.094	3.934 0.002	3.370 0.067	2.718 0.094
Weibull (2)	RP Bias	3.962	1.750 0.075	3.788 0.042	4.122 0.002	4.092 0.031	3.788 0.042
Weibull(2.5)	RP Bias	4.088	2.534 0.035	4.524 .0190	4.187 0.001	4.502 0.014	4.524 .0190

Table II
Summary of the relative precision (RP) values for estimating the population mean using RSS, ERSS, MRSS, and PRSS with values of p = 0.2, 0.3 and 0.4, with sample size n = 9.

Distribution	RSS	ERSS	MRSS	PRSS	20%	30%	40%
Uniform (0, 1)	RP	5.0	10.19	3.667	5.729	4.365	3.819
	Bias						
Normal (0, 1)	RP	4.394	2.798	6.020	4.431	5.365	5.863
	Bias						
Exponential (1)	RP	3.181	0.484	1.432	3.770	2.577	1.636
	Bias		0.389	0.254	0.000	0.158	0.232
Gamma (2)	RP	3.650	0.771	2.166	4.102	3.344	2.409
	Bias		0.412	0.267	0.001	0.167	0.244
Gamma(3)	RP	3.858	1.003	2.708	4.241	3.795	2.949
	Bias		0.419	0.271	0.001	0.170	0.248
Gamma (5)	RP	4.052	1.347	3.439	4.362	4.292	3.665
	Bias		0.425	0.274	0.002	0.172	0.251
Lognormal (0, 1)	RP	1.980	0.288	1.528	4.199	2.635	1.719
	Bias		0.975	0.562	0.064	0.387	0.523
Double Exponential (0,1)	RP	3.374	1.299	11.42	3.695	6.735	9.900
	Bias						
Inverse Gaussian (0.5)	RP	4.265	2.139	4.914	4.495	5.036	4.993
	Bias		0.106	0.068	0.001	0.043	0.062
Inverse Gaussian (1)	RP	3.936	1.211	3.062	4.311	4.227	3.547
	Bias		0.203	0.141	0.006	0.082	0.119
Inverse Gaussian (1.5)	RP	3.542	0.760	2.320	4.100	3.487	2.569
	Bias		0.286	0.183	0.003	0.116	0.167
Inverse Gaussian (2.5)	RP	2.850	0.736	1.494	3.739	2.615	1.692
	Bias		0.228	0.253	0.007	0.163	0.233
Weibull (0.5)	RP	1.736	0.228	1.228	3.807	2.182	1.387
	Bias		2.327	1.327	0.162	0.926	1.239
Weibull (1.5)	RP	3.992	1.039	2.530	4.244	3.642	2.770
	Bias		0.151	0.098	0.008	0.061	0.090
Weibull (2)	RP	4.406	1.891	3.855	4.504	4.457	4.039
	Bias		0.069	0.045	0.003	0.028	0.041
Weibull (2.5)	RP	4.507	2.779	4.899	4.596	4.927	4.930
	Bias		0.032	0.020	0.001	0.013	0.018

Table III
Summary of the relative precision (RP) values for estimating the population mean using RSS, ERSS, MRSS, and PRSS with values of p = 0.2, 0.3 and 0.4, with sample size n =10.

Distribution	RSS	ERSS	MRSS	PRSS	20%	30%	40%
Uniform (0, 1)	RP	5.50	12.10	4.033	6.722	5.042	4.321
	Bias						
Normal (0, 1)	RP	4.795	2.904	6.620	4.662	5.714	6.332
	Bias						
Exponential (1)	RP	3.414	0.292	1.328	2.985	3.259	1.736
	Bias		0.514	0.254	0.007	0.117	0.213
Gamma (2)	RP	3.940	0.486	2.061	3.591	4.002	2.556
	Bias		0.544	0.267	0.072	0.125	0.224
Gamma (3)	RP	4.177	0.656	2.628	3.880	4.407	3.140
	Bias		0.554	0.271	0.072	0.127	0.228
Gamma (5)	RP	4.397	0.934	3.428	4.155	4.829	3.898
	Bias		0.561	0.274	0.072	0.129	0.230
Lognormal (0,1)	RP	2.064	0.182	1.392	3.522	3.205	1.759
	Bias		1.288	0.562	0.076	0.313	0.489
Double Exponential (0,1)	RP	3.617	1.291	12.63	3.633	6.593	9.940
	Bias						
Inverse Gaussian (0.5)	RP	4.641	1.727	5.165	4.479	5.422	5.330
	Bias		0.140	0.068	0.018	0.032	0.057
Inverse Gaussian (1)	RP	4.263	0.828	3.267	4.055	4.774	3.758
	Bias		0.269	0.131	0.034	0.062	0.110
Inverse Gaussian (1.5)	RP	3.820	0.485	2.203	3.602	4.121	2.707
	Bias		0.377	0.183	0.047	0.087	0.154
Inverse Gaussian (2.5)	RP	3.038	0.259	1.374	2.997	3.244	1.772
	Bias		0.533	0.254	0.060	0.125	0.215
Weibull (0.5)	RP	1.807	0.142	1.110	3.099	2.672	1.408
	Bias		3.074	1.328	0.169	0.754	1.162
Weibull (1.5)	RP	4.333	0.669	2.457	3.915	4.287	2.965
	Bias		0.199	0.099	0.003	0.045	0.083
Weibull (2)	RP	4.752	1.381	3.953	4.511	4.969	4.320
	Bias		0.092	0.045	0.012	0.021	0.038
Weibull (2.5)	RP	4.922	2.379	5.245	4.754	5.329	5.313
	Bias		0.042	0.020	0.004	0.010	0.017

CHAPTER ELEVEN

Estimation of Reliability based on Exponential Distribution and Ranked Set Sample

Sukuman Sarikavanij[1*], **Montip Tiensuwan**[2**] and **Bimal Sinha**[1***]
[1] Department of Mathematics and Statistics,
University of Maryland, Baltimore County, USA
[2] Department of Mathematics, Mahidol University, Bangkok,
Thailand. Email: g4536132@student.mahidol.ac.th

ABSTRACT

Assume X (strength) $\sim (1/\theta_1)e^{-x/\theta_1}$, $x > 0$, $\theta_1 > 0$, independent of Y (stress) $\sim (1/\theta_2)e^{-y/\theta_2}$, $y > 0$, $\theta_2 > 0$. In this paper we consider the problem of estimation of the reliability $R(\theta_1, \theta_2) = P(X > Y)$. We consider both simple random sample (SRS) and ranked set sample (RSS), and provide several estimates of R along with their comparisons.

1. INTRODUCTION

In this paper we consider the problem of estimation of the reliability $R(\theta_1, \theta_2) = P(X > Y)$, based on $X_1, \ldots, X_N \sim iid \sim X$ where X is the strength with pdf, $f(x) = (1/\theta_1)e^{-x/\theta_1}$, and $Y_1, \ldots, Y_M \sim iid \sim Y$ where Y is the stress with pdf, $f(y) = (1/\theta_2)e^{-y/\theta_2}$, and X and Y are independent. We consider both simple random sample (SRS) and ranked set sample (RSS), and provide several estimates of R. Under RSS, we have used three estimates of R. The comparisons of the estimates of R are conducted for large sample sizes as well as small sample sizes.

For details about RSS, we refer to Stokes (1980), McIntyre (1952), Takahasi and Wakimoto (1968), Dell and Clutter (1972) and Sinha, Sinha and Purkayastha (1995).

* Supported by a fellowship from the Institute for the Promotion of Teaching Science and Technology (IPST).

** Supported by a fellowship from Thailand Research Fund (TRF).

*** Supported by UMBC's Presidential Research Professorship grant.

2. MAIN RESULT

Since $X \sim Exp(\theta_1)$ and $Y \sim Exp(\theta_2)$ $R(\theta_1,\theta_2) = \theta_1/(\theta_1+\theta_2)$. For estimation of $R(\theta_1,\theta_2)$ based on SRS, let $X_1,\ldots,X_N$ ~ *iid* ~ $Exp(\theta_1)$, $Y_1,\ldots,Y_M$ ~ *iid* ~ $Exp(\theta_2)$. Obviously standard estimates of θ_1 and θ_2 are $\bar{X}$ and $\bar{Y}$, respectively. So we use

$$\hat{R}_{SRS}(\theta_1,\theta_2) = \frac{\bar{X}}{\bar{X}+\bar{Y}}. \tag{1}$$

By the Central Limit Theorem, for large *N* and *M*, $\bar{X} \sim N\left(\theta_1, \frac{\theta_1^2}{N}\right)$, $\bar{Y} \sim N\left(\theta_2, \frac{\theta_2^2}{M}\right)$.

Therefore, by using standard Taylor expansion, we get for large *N* and *M*

$$\hat{R}_{SRS}(\theta_1,\theta_2) \sim N\left[R(\theta_1,\theta_2), \frac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^4}\left(\frac{1}{N}+\frac{1}{M}\right)\right]. \tag{2}$$

For using RSS, we write *N=kn* and *M=sm*, and draw RSS from the *X*-population as $\left\{X_{(ii)}^{(j)}\right\}$, $i=1,\ldots,k$; $j=1,\ldots,n$ and from the *Y*-population as $\left\{Y_{(ii)}^{(j)}\right\}$, $i=1,\ldots,s$; $j=1,\ldots,m$, (see McIntyre (1952)). From McIntyre (1952) and Sinha, Sinha and Purkayastha (1995), the estimates of θ_1 and θ_2 based on RSS are obtained as

$$\hat{\theta}_{1Mc} = \sum_{j=1}^{n}\sum_{i=1}^{k}\frac{X_{(ii)}^{(j)}}{kn}, \quad \hat{\theta}_{2Mc} = \sum_{j=1}^{m}\sum_{i=1}^{s}\frac{Y_{(ii)}^{(j)}}{ms} \tag{3}$$

$$\hat{\theta}_{1Blue} = \left[\sum_{j=1}^{n}\sum_{i=1}^{k}\frac{X_{(ii)}^{(j)}}{c_{i:k}a_{i:k}}\right]\Big/\left[\sum_{j=1}^{n}\sum_{i=1}^{k}\frac{1}{a_{i:k}}\right], \quad \hat{\theta}_{2Blue} = \left[\sum_{j=1}^{m}\sum_{i=1}^{s}\frac{Y_{(ii)}^{(j)}}{c_{i:s}a_{i:s}}\right]\Big/\left[\sum_{j=1}^{m}\sum_{i=1}^{s}\frac{1}{a_{i:s}}\right] \tag{4}$$

where $a_{i:k} = \frac{d_{i:k}}{c_{i:k}^2}$, $d_{i:k} = \sum_{l=1}^{i}\left(\frac{1}{k-l+1}\right)^2$, $c_{i:k} = \sum_{l=1}^{i}\left(\frac{1}{k-l+1}\right)$, $a_{i:s} = \frac{d_{i:s}}{c_{i:s}^2}$, $d_{i:s} = \sum_{l=1}^{i}\left(\frac{1}{s-l+1}\right)^2$, $c_{i:s} = \sum_{l=1}^{i}\left(\frac{1}{s-l+1}\right)$. Here $\hat{\theta}_{Blue}$ is the best linear unbiased estimate of θ based on RSS-data. For $\hat{\theta}_{Opt}$, our strategy is a variation of the usual RSS sample, which is based on always drawing the r^{th} order statistic from each row of all the cycles, r depending on the set size, resulting in $X_{(ir_k)}^{(j)}$, $i=1,\ldots,k;\ j=1,\ldots,n$ and $Y_{(ir_s)}^{(j)}$, $i=1,\ldots,s;\ j=1,\ldots,m$. Following Sinha et al. (1995), we use

$$\hat{\theta}_{1Opt} = \left[\sum_{j=1}^{n}\sum_{i=1}^{k}\frac{X_{(ir_k)}^{(j)}}{c_{i:k}}\right]\Big/kn\,,\ \hat{\theta}_{2Opt} = \left[\sum_{j=1}^{m}\sum_{i=1}^{s}\frac{X_{(ir_s)}^{(j)}}{c_{i:s}}\right]\Big/sm\,. \tag{5}$$

Here r_k is such that $a_{r:k}$ is the smallest among $a_{1:k},\ldots,a_{k:k}$ and r_s is such that $a_{r:s}$ is the smallest among $a_{1:s},\ldots,a_{s:s}$.

Once θ_1 and θ_2 are estimated as above, an estimate of $R(\theta_1,\theta_2)$ is obtained by $\hat{R}_{RSS}(\theta_1,\theta_2)=\dfrac{\hat{\theta}_{1RSS}}{\hat{\theta}_{1RSS}+\hat{\theta}_{2RSS}}$. To study the large sample properties of $\hat{R}_{RSS}(\theta_1,\theta_2)$, we first state the following theorem where proof follows from the CLT.

Theorem 2.1 For large *n* and *m*, the distributions of the estimates of θ_1 and θ_2 based on RSS are given by

a) $$\hat{\theta}_{1Mc}\sim N\left[\theta_1,\frac{\theta_1^2}{k^2 n}\sum_{i=1}^{k}d_{i:k}\right],\ \hat{\theta}_{2Mc}\sim N\left[\theta_2,\frac{\theta_2^2}{s^2 m}\sum_{i=1}^{s}d_{i:s}\right] \tag{6}$$

b) $$\hat{\theta}_{1Blue}\sim N\left[\theta_1,\frac{\theta_1^2}{n}\left(\sum_{i=1}^{k}\frac{1}{a_{i:k}}\right)^{-1}\right],\ \hat{\theta}_{2Blue}\sim N\left[\theta_2,\frac{\theta_2^2}{m}\left(\sum_{i=1}^{s}\frac{1}{a_{i:s}}\right)^{-1}\right] \tag{7}$$

c) $$\hat{\theta}_{1Opt}\sim N\left[\theta_1,\frac{\theta_1^2 a_{r:k}}{kn}\right],\ \hat{\theta}_{2Opt}\sim N\left[\theta_2,\frac{\theta_2^2 a_{r:s}}{sm}\right] \tag{8}$$

The large sample distributions of $\hat{R}_{RSS}(\theta_1,\theta_2)$ are stated below.

Theorem 2.2 For large *n* and *m*, the distributions of the estimates of $R(\theta_1,\theta_2)$ based on RSS are given by the following:

a) $$\hat{R}_{Mc}(\theta_1,\theta_2)\ \sim N\left[R(\theta_1,\theta_2),\frac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^4}\left(\sum_{i=1}^{k}d_{i:k}/k^2 n+\sum_{i=1}^{s}d_{i:s}/s^2 m\right)\right] \tag{9}$$

b) $\hat{R}_{Blue}(\theta_1,\theta_2) \sim N\left[R(\theta_1,\theta_2), \dfrac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^4}\left(\dfrac{1}{n}\left(\sum_{i=1}^{k}\dfrac{1}{a_{i:k}}\right)^{-1} + \dfrac{1}{m}\left(\sum_{i=1}^{s}\dfrac{1}{a_{i:s}}\right)^{-1}\right)\right]$ (10)

c) $\hat{R}_{Opt}(\theta_1,\theta_2) \sim N\left[R(\theta_1,\theta_2), \dfrac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^4}\left(\dfrac{a_{r:k}}{kn} + \dfrac{a_{r:s}}{sm}\right)\right]$ (11)

Proof. Follows from Theorem 2.1 and Taylor expansion.

3. COMPARISON OF ESTIMATES

In this section we provide a comparison of the above estimates of $R(\theta_1,\theta_2)$. We first mention about the large sample result.

Theorem 3.1 For large *n* and *m*, $Var(\hat{R}_{Opt}) < Var(\hat{R}_{Blue}) < Var(\hat{R}_{Mc}) < Var(\hat{R}_{SRS})$.

Proof:

1. To compare $Var(\hat{R}_{SRS})$ with $Var(\hat{R}_{Mc})$ is equivalent to comparing $1/N$ with $\sum_{i=1}^{k} d_{i:k}/k^2 n$ and $1/M$ with $\sum_{i=1}^{s} d_{i:s}/s^2 m$.

   Since $N=kn$ and $\sum_{i=1}^{k} d_{i:k} \le k$, so $\sum_{i=1}^{k} d_{i:k}/k^2 n = \sum_{i=1}^{k} d_{i:k}/Nk \le 1/N$. Similarly we get $\sum_{i=1}^{s} d_{i:s}/sM \le 1/M$. So $Var(\hat{R}_{Mc}) < Var(\hat{R}_{SRS})$.

2. To compare $Var(\hat{R}_{Mc})$ with $Var(\hat{R}_{Blue})$ is equivalent to comparing $\sum_{i=1}^{k} d_{i:k}/k^2 n$ with $\dfrac{1}{n}\left(\sum_{i=1}^{k}\dfrac{1}{a_{i:k}}\right)^{-1}$ and $\sum_{i=1}^{s} d_{i:s}/s^2 m$ with $\dfrac{1}{m}\left(\sum_{i=1}^{s}\dfrac{1}{a_{i:s}}\right)^{-1}$.

   From the numerical computations, we verified that $\left(\sum_{i=1}^{k}\dfrac{1}{a_{i:k}}\right)^{-1} < \sum_{i=1}^{k} d_{i:k}/k^2$ for all *k*. Of course the same is true for comparing $\sum_{i=1}^{s} d_{i:s}/s^2$ with $\left(\sum_{i=1}^{s}\dfrac{1}{a_{i:s}}\right)^{-1}$.

   So $Var(\hat{R}_{Blue}) < Var(\hat{R}_{Mc})$.

3. To compare $Var(\hat{R}_{Blue})$ with $Var(\hat{R}_{Opt})$ is equivalent to comparing $\frac{1}{n}\left(\sum_{i=1}^{k}\frac{1}{a_{i:k}}\right)^{-1}$ with $\frac{a_{r:k}}{kn}$ and $\frac{1}{m}\left(\sum_{i=1}^{s}\frac{1}{a_{i:s}}\right)^{-1}$ with $\frac{a_{r:s}}{sm}$.

Since $a_{r:k} \le a_{i:k}, \forall i$, then $\frac{1}{a_{r:k}} \ge \frac{1}{a_{i:k}}$ and $\frac{k}{a_{r:k}} \ge \sum_{i=1}^{k}\left(\frac{1}{a_{i:k}}\right)$, so $\frac{a_{r:k}}{k} \le \left(\sum_{i=1}^{k}\frac{1}{a_{i:k}}\right)^{-1}$.

Similarly, we get $\frac{a_{r:s}}{s} \le \left(\sum_{i=1}^{s}\frac{1}{a_{i:s}}\right)^{-1}$. So $Var(\hat{R}_{Opt}) < Var(\hat{R}_{Blue})$

This completes the proof.

We conclude this paper with a small sample comparison of the above estimates of $R(\theta_1,\theta_2)$ based on 1000 simulations using SAS. We have taken *N*=*M*=10, and *n*=*m*=5, *k*=*s*=2. The table below shows the bias and the variance of the proposed estimates of $R(\theta_1,\theta_2)$.

Table 3.1 Comparison of estimates of $R(\theta_1,\theta_2)$ in small samples

	$\theta_1=1$, $\theta_2=1$ *R*=0.5		$\theta_1=1$, $\theta_2=2$ *R*=0.33		$\theta_1=1$, $\theta_2=3$ *R*=0.25		$\theta_1=1$, $\theta_2=4$ *R*=0.2	
	bias	var	bias	var	bias	var	bias	var
SRS	0.00179	0.01041	0.00781	0.00842	0.00933	0.00625	0.00943	0.00472
McIntyre	0.00618	0.00797	0.01031	0.00648	0.01082	0.00477	0.01036	0.00357
Blue	-0.00358	0.00884	0.00205	0.00705	0.00399	0.00515	0.00459	0.00383
Optimum	-0.00292	0.00807	0.00744	0.00656	0.00841	0.00484	0.00831	0.00362

Table 3.1 (continued)

	$\theta_1=2$, $\theta_2=1$ *R*=0.67		$\theta_1=3$, $\theta_2=1$ *R*=0.75		$\theta_1=4$, $\theta_2=1$ *R*=0.8	
	bias	var	bias	var	bias	var
SRS	-0.00469	0.00863	-0.00677	0.00648	-0.00729	0.00492
McIntyre	0.00073	0.00645	-0.00146	0.00475	-0.00234	0.00355
Blue	-0.00854	0.00732	-0.00961	0.00546	-0.00949	0.00413
Optimum	-0.00221	0.00651	-0.00396	0.00478	-0.00448	0.00357

It follows from the above table that, even in small samples, the estimates of $R(\theta_1, \theta_2)$ based on RSS have both smaller bias and smaller variance compared to the SRS-based estimate. It also happens that the estimate of $R(\theta_1, \theta_2)$ based on McIntyre procedure is marginally better than the two other RSS-based estimates.

REFERENCES

1. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-555.
2. McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. *Aust. J. Agri. Res.*, 3, 385-390.
3. Sinha, Bimal K., Sinha, Bikas K. and Purkayastha, S. (1995). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics & Decisions*, 14, 223-240.
4. Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics*, 36, 35-42.
5. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math*, 20, 1-31.

CHAPTER TWELVE

Estimation of Quantiles of Uniform Distribution Using Generalized Ranked-Set Sampling

A. Adatia[1] and **A.K. Md. Ehsanes Saleh**[2]
[1] University of Regina, Regina, Saskatchewan, Canada
[2] School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada, Email: esaleh@math.carleton.ca

ABSTRACT

The ranked-set sampling technique has been generalized so that more efficient estimators may be obtained. The generalized ranked-set sampling technique is applied in the estimation of quantiles of the uniform distribution. Three estimators are proposed. These include generalized ranked-set minimum variance unbiased estimator, simple estimator and ranked-set sample estimator. Coefficients, variances and relative efficiencies are derived. The estimators are compared to the best linear unbiased estimator of the quantiles.

KEYWORDS

Uniform distribution, order statistics, linear estimation, generalized ranked-set sampling, ranked-set sampling

1. INTRODUCTION

In applied statistics, experimenters often encounter situations where the actual measurements of the sample observations are difficult to make due to constraints in cost, time and other factors. However, ranking of the potential sample data is relatively easy. In these situations, McIntyre (1952) advocated the use of ranked-set sampling. He applied the ranked-set sampling technique in assessing the yields of pasture plots without actually carrying out the time-consuming process of mowing and weighing the hay for a large number of plots. Since then, the technique has been studied and applied to several areas of applied research. Takahasi and Wakimato (1968) and Dell and Clutter (1972) studied theoretical aspects of this technique on the assumption of perfect judgment ranking and imperfect judgment ranking respectively. Patil, Sinha and Taillie (1993) studied the same technique when the sample is from a finite population. Patil, Sinha and Taillie (1994) have reviewed various aspects of the ranked-set sampling. Also, Bohn (1996) discussed the application of this technique in nonparametric procedures.

In this paper the ranked-set sampling technique has been generalized so that more efficient estimators may be obtained. The generalized ranked-set sampling technique is applied in the estimation of quantiles of the uniform distribution. Three estimators are proposed. These are generalized ranked-set minimum variance unbiased estimator (GR-MVUE), simple estimator (SE) and ranked-set sample estimator (RSS). Coefficients, variances, and relative efficiencies are derived. The estimators are compared to the best linear unbiased estimators (BLUE) of the quantiles.

In generalized ranked-set sampling, first a set of N elements is randomly selected from a given population. The sample is ordered without making actual measurements. The unit identified with the N_1 rank is accurately measured. Next, a second set of N elements is randomly selected from the population. Again the units are ordered and the unit with the N_2 rank is accurately measured. The process is continued until N set of N elements is selected. The units are again ordered and the unit with N_N rank is accurately measured. The ordered sample of the N sets can be represented as follows:

Set 1	$X_{(11)}$	$X_{(12)}$	...	$X_{(1N)}$
Set 2	$X_{(21)}$	$X_{(22)}$	...	$X_{(2N)}$
⋮	⋮	⋮	...	⋮
Set N	$X_{(N1)}$	$X_{(N2)}$	...	$X_{(NN)}$

The generalized ranked-set sample of size N consists of units which are accurately measured i.e. $(X_{(1N_1)}, X_{(2N_2)}, \ldots, X_{(NN_N)})$ where $1 \le N_i \le N$ and $1 \le i \le N$. The generalized ranked-set sample actually includes the usual ranked-set sample which is obtained when $N_1 = 1$, $N_2 = 2, \ldots, N_N = N$.

2. ESTIMATORS

2.1 Best Linear Unbiased Estimator of Quantiles

Let the random variable *X* have a uniform distribution with probability density function

$$f(x) = \frac{1}{2\sqrt{3}\sigma} \quad , \quad \mu - \sqrt{3}\sigma \le x \le \mu + \sqrt{3}\sigma \, , \; \sigma > 0$$

where μ and σ are the location and scale parameters respectively.

The quantile function of the distribution is defined as

$$Q(\xi) = \mu + \sqrt{3}\sigma(2\xi - 1), \quad 0 \le \xi \le 1.$$

The best linear unbiased estimator (BLUE) of the quantile function Q(ξ) is

$$\hat{Q}(\xi)_{BLUE} = \hat{\mu} + \sqrt{3}\hat{\sigma}(2\xi - 1)$$

where the location and scale parameters are estimated by their respective BLUEs. Downton (1954) and Sarhan and Greenberg (1962) have obtained the results for BLUEs for the location and the scale parameter.

The variance is given by

$$V(\hat{Q}(\xi)_{BLUE}) = 6\sigma^2\left(3(N-1)+(N+1)(2\xi-1)^2\right)\Big/\left((N^2-1)(N+2)\right)$$

2.2 Estimator of Quantiles based on generalized ranked-set sampling

Let $Z_{(iN_j)} = (X_{(iN_j)} - \mu)/\sigma$

$\alpha_{(iN_j)} \equiv E(Z_{(iN_j)})$

$\omega_{iN_jN_j} \equiv Var(Z_{(iN_j)})$, i = 1, 2,…, N, j = 1,2,…,N.

Therefore $E(X_{(iN_j)}) \equiv \mu + \sigma\alpha_{(iN_j)}$ and $Var(X_{(iN_j)}) \equiv \omega_{iN_jN_j}\sigma^2$.

Let $\alpha_S = \left(\alpha_{(1N_1)}, \alpha_{(2N_2)}, ..., \alpha_{(NN_N)}\right)^T$ where T implies the transpose

$\mathbf{1}^T = (1,\dots 1)$

$S = \{N_1, N_2, ..., N_N\}$

$X_S = \left(X_{(1N_1)}, X_{(2N_2)}, ..., X_{(NN_N)}\right)^T$

and $Var(X_S) = \Omega_S\sigma^2$

where Ω_S is a N × N diagonal matrix with $\omega_{iN_iN_i}$ as the (i,i)th element.

Then $E(X_S) = \mu 1 + \sigma\alpha_S = A_S\theta$

where $A_S^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \alpha_{(1N_1)} & \alpha_{(2N_2)} & \dots & \alpha_{(NN_N)} \end{pmatrix}$ and $\theta^T = (\mu, \sigma)$.

Least squares estimator of θ is obtained by applying the Gauss and Markov theorem (Sarhan and Greenberg (1962)). Then

$$\hat{\theta}_S = (A_S^T\Omega_S^{-1}A_S)^{-1}A_S^T\Omega_S^{-1}X_S \text{ and } Var\left(\hat{\theta}_S\right) = (A_S^T\Omega_S^{-1}A_S)^{-1}\sigma^2$$

where superscript -1 implies inverse.

Therefore, based on generalized ranked-set sample X_S with

$S = \{N_1, N_2, \ldots, N_N\}$ the estimator for the quantile function is

$$\hat{Q}(\xi)_S = \hat{\mu}_S + \sqrt{3}\hat{\sigma}_S(2\xi - 1)$$

where

$$\hat{\mu}_S = \sum_{i=1}^{N} \frac{(T_{1S} - \alpha_{(iN_i)}T_{3S})}{(T_{1S}T_{2S} - T_{3S}^2)}\left(X_{(iN_i)} / \omega_{iN_iN_i}\right)$$

$$\hat{\sigma}_S = \sum_{i=1}^{N} \frac{(\alpha_{(iN_i)}T_{2S} - T_{3S})}{(T_{1S}T_{2S} - T_{3S}^2)}\left(X_{(iN_i)} / \omega_{iN_iN_i}\right)$$

$$T_{1S} = \left(\alpha_s{}^T \cdot \Omega_S{}^{-1} \cdot \alpha_s\right) = \sum_{i=1}^{N} \alpha_{(iN_i)}^2 / \omega_{iN_iN_i}$$

$$T_{2S} = \left(1^T \cdot \Omega_S{}^{-1} \cdot 1\right) = \sum_{i=1}^{N} 1 / \omega_{iN_iN_i}$$

$$T_{3S} = \left(1^T \cdot \Omega_S{}^{-1} \cdot \alpha_S\right) = \sum_{i=1}^{N} \alpha_{(iN_i)} / \omega_{iN_iN_i}$$

The variances of the estimator is given by

$$V(\hat{Q}(\xi)_S = \frac{3\sigma^2}{(T_{1S}T_{2S} - T_{3S}^2)}\left[T_{1S} + T_{2S}(2\xi - 1)^2 - 2T_{3S}(2\xi - 1)\right]$$

where

$$\alpha_{(iN_j)} = \sqrt{3}\left(\frac{2N_j}{(N+1)} - 1\right)$$

$$\omega_{iN_jN_j} = \frac{12N_j(N - N_j + 1)}{(N+1)^2(N+2)}$$

$$Cov(X_{(iN_j)}, X_{(iN_k)}) = \omega_{iN_jN_k} = \frac{12N_j(N - N_k + 1)}{(N+1)^2(N+2)} \quad \text{for } j < k.$$

2.3 Generalized Ranked-Set Minimum Variance Unbiased Estimators

Generalized ranked-set minimum variance unbiased estimator (GR-MVUE) is obtained from the generalized ranked-set estimator when all possible choices of S are considered. The best choice of S is the one which gives the minimum variance of the estimator. This S is denoted by $S_{GR\text{-}MVUE}$. The estimator is denoted by

$\hat{Q}(\xi)_{GR-MVUE}$. Table 1 provides ranks $S_{GR\text{-}MVUE}$ and coefficients of the estimator for $N = 2(1)10$ and $\xi = 0.01, 0.05, 0.5, 0.9, 0.95$

2.4 Simple Estimator (SE)

The simple estimator is obtained from the generalized ranked-set estimators as follows:

When sample size is even, $S = \{N_1, N_2, ..., N_N\}$ where $N_1 = N_2 = \ldots = N_{N/2} = 1$ and $N_{N/2+1} = \cdots = N_N = N$ then $S = \{1,...,1,N,...,N\}$. Then from generalized ranked-set estimators, the simple estimator for the quantile function Q(ξ) when sample sizes are even is:

$$\hat{Q}(\xi)_{SE} = \hat{\mu}_{SE} + \sqrt{3}\hat{\sigma}_{SE}(2\xi - 1)$$
$$= \sum_{i=1}^{N/2} \mathrm{B_i}(N)_{SE} X_{(i1)} + \sum_{j=N/2+1}^{N/2} \mathrm{B_j}(N)_{SE} X_{(iN)}$$

where

$$B_i(N)_{SE} = 1/N - (1+N)(2\xi - 1)/(N(N-1)), \quad \mathrm{i} = 1,..,\mathrm{N}/2$$
$$B_j(N)_{SE} = 1/N + (1+N)(2\xi - 1)/(N(N-1)), \quad \mathrm{j} = \mathrm{N}/2+1,..,\mathrm{N}$$

SE = {1,...,1,N,...,N}

$$Var(\hat{Q}(\xi)_{SE} = 12\sigma^2 \left(3/(N+2)(N+1)^2\right) + (2\xi - 1)^2 / \left((N+2)(N-1)^2\right)$$

When sample size is odd $S = \{N_1, N_2, ..., N_N\}$ where $N_1 = N_2 = \ldots = N_{(N+1)/2} = 1$ and $N_{(N+1)/2+1} = \cdots = N_N = N$ then $S = \{1,...,1,1,N,...,N\}$. Then from generalized ranked-set estimator, the simple estimator for the quantile function Q(ξ) when sample sizes are odd is:

$$\hat{Q}(\xi)_{SE} = \hat{\mu}_{SE} + \sqrt{3}\hat{\sigma}_{SE}(2\xi - 1)$$
$$= \sum_{i=1}^{(N-1)/2+1} \mathrm{B_i}(N)_{SE} X_{(i1)} + \sum_{j=(N-1)/2+2}^{N} \mathrm{B_j}(N)_{SE} X_{(iN)}$$

where

$$B_i(N)_{SE} = 1/(N+1) - (2\xi - 1)^2/(N-1)$$
$$B_j(N)_{SE} = 1/(N-1) + (2\xi - 1)^2 (1+N)/(N-1)^2$$

SE = {1,1,...,1,1,N,N,...N}

$$Var(\hat{Q}(\xi))_{SE} = 12\sigma^2 (3N^2/((1+N)^3(N+2)(N-1)$$
$$+(2\xi - 1)^2 / \left((N-1)^3(N+1)(N+2)\right) + 2\sqrt{3}N / \left((N+2)(N^2-1)^2)\right)$$

2.5 Ranked-Set Sample Estimator (RSS)

The ranked-set sample estimator (RSS) for $Q(\xi)$ is obtained from the generalized ranked-set estimator when S = {1,2,..,N}. The estimator is

$$\hat{Q}(\xi)_{RSS} = \hat{\mu}_{RSS} + \sqrt{3}\hat{\sigma}_{RSS}(2\xi - 1)$$
$$= \sum_{i=1}^{N} \mathrm{B_i}(N)_{RSS} X_{(i1)} + \sum_{i=1}^{N} \mathrm{C_i}(N)_{RSS} X_{(i1)}$$

where

$$B_i(N)_{RSS} = (N+1)/(2i(N-i+1)S_N)$$
$$C_i(N)_{RSS} = \sqrt{3}\ (N+1)(2i-N-1)(2\xi-1)/\left(2\sqrt{3}i(N-i+1)\{(N+1)S_N - 2N\}\right)$$
$$S_N = \sum_{i=1}^{N}(1/i)$$

RSS = {1,2,...,N}

The variance of the estimator is

$$Var(\hat{Q}(\xi)_{RSS}) = 6\sigma^2\left(3/((N+1)(N+2)S_N)\right.$$
$$\left. + (2\xi-1)^2/\left((N+2)\{(N+1)S_N - 2N\}\right)\right)$$

3. COMPARISONS

In this section comparison has been made between generalized ranked-set minimum variance unbiased estimator (GR-MVUE), simple estimator (SE), ranked-set sample estimator (RSS) and best linear unbiased estimator (BLUE). Generalized ranked-set minimum variance unbiased estimator is more efficient than simple estimator, ranked-set sample estimator and best linear unbiased estimator. This is apparent from Table 2 and the properties of the simple estimator and ranked-set sample estimators. The simple estimator is more efficient than ranked-set sample estimators as well as the best linear unbiased estimators. The ranked-set sample estimator is more efficient than the best linear unbiased estimator.

Simple estimator is more efficient than ranked-set sample estimator for N = 3 and $0 \le \xi \le \frac{1}{44}\left(22 + 11\sqrt{3} - 6\sqrt{11}\right)$ and N ≥ 4 and $0 \le \xi \le 1$. This can be seen by considering even and odd sample cases.

When N is even, let N = 2m where m ≥ 1. When m ≥ 2, $S_{2m} < (2m+1)/2$ and therefore

$$
\begin{aligned}
&Var\left(\hat{Q}(\xi)_{RSS}\right)-Var\left(\hat{Q}(\xi)_{SE}\right)\\
&\quad=6\sigma^{2}\left(\left(3/((N+1)(N+2)S_{N}\right)+(2\xi-1)^{2}/\left((N+2)\left\{(N+1)S_{N}-2N\right\}\right)\right)\\
&\quad\quad-12\sigma^{2}\left(\left(3/(N+2)(N+1)^{2}\right)+(2\xi-1)^{2}/\left((N+2)(N-1)^{2}\right)\right)\\
&\quad=\sigma^{2}\left(\frac{18}{(2m+1)(2m+2)S_{2m}}-\frac{36}{(2m+2)(2m+1)^{2}}\right.\\
&\quad\quad\left.+\frac{6\,(2\xi-1)^{2}}{(2m+2)\left\{(2m+1)S_{2m}-2m\right\}}-\frac{12(2\xi-1)^{2}}{(2m+2)(2m-1)^{2}}\right)\\
&\quad>\sigma^{2}\left((\frac{18}{(2m+1)(2m+2)(2m+1)/2}-\frac{36}{(2m+2)(2m+1)^{2}}\right.\\
&\quad\quad\left.+\frac{6\,(2\xi-1)^{2}}{(2m+2)\left\{(2m+1)(2m+1)/2-2m\right\}}-\frac{12(2\xi-1)^{2}}{(2m+2)(2m-1)^{2}}\right)\\
&\quad=0.
\end{aligned}
$$

Therefore when $N \geq 4$ and is even, Simple Estimator (SE) is more efficient than ranked-set sample estimators (RSS)

When N is odd, let $N = 2m+1$. When $m \geq 2$, $S_{2m+1} < 3/2 + 4m^3/(2m+1)^2$, then

$$
\begin{aligned}
&Var\left(\hat{Q}(\xi)_{RSS}\right)-Var\left(\hat{Q}(\xi)_{SE}\right)\\
&\quad=6\sigma^{2}\left(\left(3/((N+1)(N+2)S_{N}\right)+(2\xi-1)^{2}/\left((N+2)\left\{(N+1)S_{N}-2N\right\}\right)\right)\\
&\quad\quad-12\sigma^{2}\left(3N^{2}/(1+N)^{3}(N+2)(N-1)\right.\\
&\quad\quad+(2\xi-1)^{2}N^{2}/\left((N-1)^{3}(N+1)(N+2)\right)\\
&\quad\quad\left.+2\sqrt{3}N/\left((N+2)(N^{2}-1)^{2}\right)\right)
\end{aligned}
$$

$$= \sigma^2 \left(\frac{18}{(2m+2)(2m+3)S_{2m+1}} + \frac{6(2\xi-1)^2}{(2m+3)\left\{(2m+2)S_{2m+1} - 2(2m+1)\right\}} - \frac{36(2m+1)^2}{(2m+2)^3(2m+3)(2m)} - \frac{12(2\xi-1)^2(2m+1)^2}{(2m)^3(2m+2)(2m+3)} - \frac{24\sqrt{3}(2m+1)(2\xi-1)}{(2m+3)((2m+1)^2 - 1)^2} \right)$$

$$= \sigma^2 \left(\frac{18}{(2m+2)(2m+3)S_{2m+1}} + \frac{6(2\xi-1)^2}{(2m+3)\left\{(2m+2)S_{2m+1} - 2(2m+1)\right\}} - \frac{36(2m+1)^2}{(2m+2)^3(2m+3)(2m)} - \frac{12(2\xi-1)^2(2m+1)^2}{(2m)^3(2m+2)(2m+3)} - \frac{24\sqrt{3}(2m+1)(2\xi-1)}{(2m+3)((2m+1)^2 - 1)^2} \right)$$

$$> \sigma^2 \left(\frac{18}{(2m+2)(2m+3)\left(3/2 + 4m^3/(2m+1)^2\right)} + \frac{6(2\xi-1)^2}{(2m+3)\left\{(2m+2)\left(3/2 + 4m^3/(2m+1)^2\right) - 2(2m+1)\right\}} - \frac{36(2m+1)^2}{(2m+2)^3(2m+3)(2m)} - \frac{12(2\xi-1)^2(2m+1)^2}{(2m)^3(2m+2)(2m+3)} - \frac{24\sqrt{3}(2m+1)(2\xi-1)}{(2m+3)\left((2m+1)^2 - 1\right)^2} \right)$$

After simplifying, the expression is

$$\begin{aligned}
&\mathrm{Var}\left(\hat{Q}(\xi)_{RSS}\right) - \mathrm{Var}\left(\hat{Q}(\xi)_{SE}\right)\\
&> (3(1+2m)(-3 + 12\xi - 12\xi^2\\
&\quad +m(-33 + 6\sqrt{3} + 132\xi - 12\sqrt{3}\xi - 132\xi^2)\\
&\quad +m^2(-156 + 48\sqrt{3} + 588\xi - 96\sqrt{3}\xi - 588\xi^2)\\
&\quad +m^3(-392 + 138\sqrt{3} + 1340\xi - 276\sqrt{3}\xi - 1340\xi^2)\\
&\quad +m^4(-488 + 208\sqrt{3} + 1544\xi - 416\sqrt{3}\xi - 1544\xi^2)\\
&\quad +m^5(-160 + 304\sqrt{3} + 448\xi - 608\sqrt{3}\xi - 448\xi^2)\\
&\quad +m^6(160 + 480\sqrt{3} - 1120\xi - 960\sqrt{3}\xi + 1120\xi^2)\\
&\quad +m^7(128 + 544\sqrt{3} - 1664\xi - 1088\sqrt{3}\xi + 1664\xi^2)\\
&\quad +m^8(256 + 384\sqrt{3} - 1024\xi - 768\sqrt{3}\xi + 1024\xi^2)\\
&\quad +m^9(256 + 128\sqrt{3} - 256\xi - 256\sqrt{3}\xi + 256\xi^2)))/\\
&\quad (4m^3(1+m)^3(3+2m)(3+12m+12m^2\\
&\quad +8m^3)(1+3m+4m^3+8m^4))
\end{aligned}$$

The bracket containing ξ terms in the numerator of the right hand side of the above equation, is rewritten as $(-3 + 12\xi - 12\xi^2 + mf_1(\xi) + m^2f_2(\xi) + m^3f_3(\xi) + m^4f_4(\xi) + m^5f_5(\xi) + m^6f_6(\xi) + m^7f_7(\xi) + m^8f_8(\xi) + m^9f_9(\xi))$ where $f_i(\xi)$ are the coefficients of m^i ($1 \le i \le 9$). The function $f_9(\xi)$ is positive for $0 \le \xi \le 1$. The function $f_i(\xi)$ ($1 \le i \le 8$) is either positive or negative depending on the value of ξ. For $m \ge 15$ and $0 \le \xi \le 1$,

$$\begin{aligned}
&m^8(12f_9(\xi)+f_8(\xi)) + m^7(24f_9(\xi)+f_7(\xi)) + m^6(20f_9(\xi)+f_6(\xi)) + m^5(21f_9(\xi)+f_5(\xi))\\
&\quad + m^4(25f_9(\xi)+f_4(\xi)) + m^3(19f_9(\xi)+f_3(\xi)) + m^2(7f_9(\xi)+f_2(\xi)) + m(2f_9(\xi)+f_1(\xi))\\
&\quad +((((((((m-12)m-24)m-20)m-21)m-25)m-19)m-7)m-2)f_9(\xi) - 3 + 12\xi - 12\xi^2 > 0
\end{aligned}$$

Therefore, when N is odd and $N \ge 31$ and $0 \le \xi \le 1$

$$Var(\hat{Q}(\xi)_{RSS}) - Var(\hat{Q}(\xi)_{SE}) > 0$$

It can also be shown that when N = 3 and $0 \le \xi \le \frac{1}{44}\left(22 + 11\sqrt{3} - 6\sqrt{11}\right)$ or when $5 \le N \le 30$ and is odd and $0 \le \xi \le 1$ $\quad Var(\hat{Q}(\xi)_{RSS}) > Var(\hat{Q}(\xi)_{SE})$.

Simple estimator is also more efficient than best linear unbiased estimator when $N \ge 4$ and $0 \le \xi \le 1$. The relative efficiencies of simple estimator compared to ranked-set sample estimator and best linear unbiased estimator are presented in Table 3 for $2 \le N \le 10$ and $\xi = 0.01, 0.05, 0.5, 0.9, 0.95$

Ranked-set sample estimator is also more efficient than best linear unbiased estimator when $N \ge 4$ and $0 \le \xi \le 1$. The relative efficiencies of ranked-set sample estimator compared to best linear unbiased estimator are presented in Table 4 for $2 \le N \le 10$ and $\xi = 0.01, 0.05, 0.5, 0.9, 0.95$

4. CONCLUSION

It is evident that the generalized ranked-set minimum variance unbiased estimator, the simple estimator and the ranked-set sample estimator are all more efficient than the best linear unbiased estimator.

The simple estimators are more efficient than ranked-set sample estimators as well as the best linear unbiased estimators. The simple estimator has a closed form and the expression for the variances, has been derived. The simple estimators are more useful than the ranked-set sample estimators or the best linear estimators.

REFERENCES

1. Bohn, Lora L. (1996), A review of nonparametric ranked-set sampling methodology, *Commun. Statist.- Theory. Meth.*, 25, 2675-2685.
2. Dell, T.R. and Cutter J.L. (1972), Ranked-set sampling theory with order statistics background, *Biometrics*, 28, 545-555.
3. Downton, F. (1954), Least-squares estimates using ordered observation, *Ann. Math. Statist.*, 25, 303-316.
4. McIntyre, G.A. (1952), A method of unbiased selective sampling, using ranked sets, *Australian Journal of Agricultural Research*, 3, 385-390.
5. Patil, G.P., Sinha, A.K. and Taillie C. (1993), Ranked set sampling from a finite population in the presence of a trend on a site, *Journal of Applied Statistical Science*, 1, 51-65.
6. Patil, G.P., Sinha, A.K. and Taillie C. (1994), Ranked set sampling, In *Handbook of Statistics Volume 12: Environmental Statistics*, G.P. Patil, and C.R. Rao (eds). North Holland/Elsevier Science Publishers, New York and Amsterdam, 167-200.
7. Sarhan, A.E. and Greenberg B.G. (Eds). (1962), *Contributions to order statistics*, Wiley, New York.
8. 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.

Table 1: Coefficients for computing $\hat{Q}(\xi)_{GR-MVUE}$

N	$S_{GR\text{-}MVUE}$	ξ	1	2	3	4	5	6	7	8	9	10
2	{1,2}	0.95	-0.2794	1.2794								
3	{1,3,3}	0.95	-0.0196	0.5098	0.5098							
4	{3,4,4,4}	0.95	0.2010	0.2663	0.2663	0.2663						
5	{1,5,5,5,5}	0.95	0.1103	0.2224	0.2224	0.2224	0.2224					
6	{1,6,6,6,6,6}	0.95	0.1363	0.1727	0.1727	0.1727	0.1727	0.1727				
7	{1,7,7,7,7,7,7}	0.95	0.1536	0.1411	0.1411	0.1411	0.1411	0.1411	0.1411			
8	{1,8,8,8,8,8,8,8}	0.95	0.1660	0.1191	0.1191	0.1191	0.1191	0.1191	0.1191	0.1191		
9	{1,1,9,9,9,9,9,9,9}	0.95	0.0876	0.0876	0.1178	0.1178	0.1178	0.1178	0.1178	0.1178	0.1178	
10	{1,1,10,10,10,10,10,10,10,10}	0.95	0.0912	0.0912	0.1022	0.1022	0.1022	0.1022	0.1022	0.1022	0.1022	0.1022
2	{1,2}	0.9	-0.1928	1.1928								
3	{2,3,3}	0.9	0.0762	0.4619	0.4619							
4	{2,4,4,4}	0.9	0.1726	0.2758	0.2758	0.2758						
5	{1,5,5,5,5}	0.9	0.1536	0.2116	0.2116	0.2116	0.2116					
6	{1,6,6,6,6,6}	0.9	0.1767	0.1647	0.1647	0.1647	0.1647	0.1647				
7	{1,7,7,7,7,7,7}	0.9	0.1921	0.1347	0.1347	0.1347	0.1347	0.1347	0.1347			
8	{1,1,8,8,8,8,8,8}	0.9	0.1015	0.1015	0.1328	0.1328	0.1328	0.1328	0.1328	0.1328		
9	{1,1,9,9,9,9,9,9,9}	0.9	0.1057	0.1057	0.1127	0.1127	0.1127	0.1127	0.1127	0.1127	0.1127	
10	{1,1,10,10,10,10,10,10,10,10}	0.9	0.1089	0.1089	0.0978	0.0978	0.0978	0.0978	0.0978	0.0978	0.0978	0.0978
2	{1,2}	0.5	0.5000	0.5000								
3	{1,2,3}	0.5	0.3636	0.2727	0.3636							
4	{1,1,4,4}	0.5	0.2500	0.2500	0.2500	0.2500						
5	{1,1,1,5,5}	0.5	0.1667	0.1667	0.1667	0.2500	0.2500					
6	{1,1,1,6,6,6}	0.5	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667				
7	{1,1,1,1,7,7,7}	0.5	0.1250	0.1250	0.1250	0.1250	0.1667	0.1667	0.1667			
8	{1,1,1,1,8,8,8,8}	0.5	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250		
9	{1,1,1,1,1,9,9,9,9}	0.5	0.1000	0.1000	0.1000	0.1000	0.1000	0.1250	0.1250	0.1250	0.1250	
10	{1,1,1,1,1,10,10,10,10,10}	0.5	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000

Table 1 (continued)

N	$S_{GR\text{-}MVUE}$	ξ	1	2	3	4	5	6	7	8	9	10
2	{1,2}	0.1	1.1928	-0.1928								
3	{1,1,2}	0.1	0.4619	0.4619	0.0762							
4	{1,1,1,3}	0.1	0.2758	0.2758	0.2758	0.1726						
5	{1,1,1,1,5}	0.1	0.2116	0.2116	0.2116	0.2116	0.1536					
6	{1,1,1,1,1,6}	0.1	0.1647	0.1647	0.1647	0.1647	0.1647	0.1767				
7	{1,1,1,1,1,1,7}	0.1	0.1347	0.1347	0.1347	0.1347	0.1347	0.1347	0.1921			
8	{1,1,1,1,1,1,8,8}	0.1	0.1328	0.1328	0.1328	0.1328	0.1328	0.1328	0.1015	0.1015		
9	{1,1,1,1,1,1,1,9,9}	0.1	0.1127	0.1127	0.1127	0.1127	0.1127	0.1127	0.1127	0.1057	0.1057	
10	{1,1,1,1,1,1,1,1,10,10}	0.1	0.0978	0.0978	0.0978	0.0978	0.0978	0.0978	0.0978	0.0978	0.1089	0.1089
2	{1,2}	0.05	1.2794	-0.2794								
3	{1,1,3}	0.05	0.5098	0.5098	-0.0196							
4	{1,1,1,2}	0.05	0.2663	0.2663	0.2663	0.2010						
5	{1,1,1,1,5}	0.05	0.2224	0.2224	0.2224	0.2224	0.1103					
6	{1,1,1,1,1,6}	0.05	0.1727	0.1727	0.1727	0.1727	0.1727	0.1363				
7	{1,1,1,1,1,1,7}	0.05	0.1411	0.1411	0.1411	0.1411	0.1411	0.1411	0.1536			
8	{1,1,1,1,1,1,1,8}	0.05	0.1191	0.1191	0.1191	0.1191	0.1191	0.1191	0.1191	0.1660		
9	{1,1,1,1,1,1,1,9,9}	0.05	0.1178	0.1178	0.1178	0.1178	0.1178	0.1178	0.1178	0.0876	0.0876	
10	{1,1,1,1,1,1,1,1,10,10}	0.05	0.1022	0.1022	0.1022	0.1022	0.1022	0.1022	0.1022	0.1022	0.0912	0.0912

Table 2: Variances and relative efficiencies for generalized ranked-set minimum variance unbiased estimator

N	ξ	$\frac{Var\left(\hat{Q}(\xi)_{GR-MVUE}\right)}{\sigma^2}$	$\frac{Var\left(\hat{Q}(\xi)_{BLUE}\right)}{Var\left(\hat{Q}(\xi)_{GR-MVUE}\right)}$	$\frac{Var\left(\hat{Q}(\xi)_{RSS}\right)}{Var\left(\hat{Q}(\xi)_{GR-MVUE}\right)}$	$\frac{Var\left(\hat{Q}(\xi)_{SE}\right)}{Var\left(\hat{Q}(\xi)_{MVUE}\right)}$
2	0.95	3.4300	0.7915	1.0000	1
3	0.95	0.7023	1.9736	1.7371	1.9989
4	0.95	0.2625	3.3148	2.3743	1.6002
5	0.95	0.1500	4.0131	2.5016	1.9035
6	0.95	0.0924	4.7911	2.6951	1.5191
7	0.95	0.0626	5.4349	2.8214	1.7113
8	0.95	0.0451	5.9706	2.9036	1.4245
9	0.95	0.0331	6.6032	3.0406	1.5658
10	0.95	0.0248	7.3010	3.2085	1.4006
2	0.9	2.9200	0.8425	1.0000	1.0000
3	0.9	0.5865	2.1894	1.8192	2.1315
4	0.9	0.2620	3.1047	2.1103	1.4591
5	0.9	0.1448	3.9074	2.3205	1.8060
6	0.9	0.0919	4.5421	2.4417	1.4171
7	0.9	0.0637	5.0381	2.5056	1.5568
8	0.9	0.0450	5.6677	2.6458	1.3370
9	0.9	0.0327	6.3290	2.8021	1.4789
10	0.9	0.0248	6.9205	2.9283	1.3161
2	0.5	1.0000	1.5000	1.0000	1.0000
3	0.5	0.4909	1.8333	1.0000	1.0313
4	0.5	0.2400	2.5000	1.2000	1.0000
5	0.5	0.1488	2.8800	1.2613	1.0000
6	0.5	0.0918	3.5000	1.4286	1.0000
7	0.5	0.0638	3.9184	1.5112	1.0000
8	0.5	0.0444	4.5000	1.6557	1.0000
9	0.5	0.0331	4.9383	1.7456	1.0000
10	0.5	0.0248	5.5000	1.8778	1.0000
2	0.1	2.9200	0.8425	1.0000	1.0000
3	0.1	0.5865	2.1894	1.8192	1.0682
4	0.1	0.2620	3.1047	2.1103	1.4591
5	0.1	0.1448	3.9074	2.3205	1.2363
6	0.1	0.0919	4.5421	2.4417	1.4171
7	0.1	0.0637	5.0381	2.5056	1.2045
8	0.1	0.0450	5.6677	2.6458	1.3370
9	0.1	0.0327	6.3290	2.8021	1.2193
10	0.1	0.0248	6.9205	2.9283	1.3161
2	0.05	3.4300	0.7915	1.0000	1.0000
3	0.05	0.7023	1.9736	1.7371	1.0000
4	0.05	0.2625	3.3148	2.3743	1.6002
5	0.05	0.1500	4.0131	2.5016	1.2851
6	0.05	0.0924	4.7911	2.6951	1.5191
7	0.05	0.0626	5.4349	2.8214	1.3075
8	0.05	0.0451	5.9706	2.9036	1.4245
9	0.05	0.0331	6.6032	3.0406	1.2772
10	0.05	0.0248	7.3010	3.2085	1.4006

Table 3: Variances and relative efficiencies for simple estimator

N	ξ	$\frac{Var\left(\hat{Q}(\xi)_{SE}\right)}{\sigma^2}$	$\frac{Var\left(\hat{Q}(\xi)_{RSS}\right)}{Var\left(\hat{Q}(\xi)_{SE}\right)}$	$\frac{Var\left(\hat{Q}(\xi)_{BLUE}\right)}{Var\left(\hat{Q}(\xi)_{SE}\right)}$
2	**0.95**	**3.4300**	**1.0000**	**0.7915**
3	**0.95**	**1.4037**	**0.8690**	**0.9874**
4	**0.95**	**0.4200**	**1.4837**	**2.0714**
5	**0.95**	**0.2856**	**1.3142**	**2.1083**
6	**0.95**	**0.1404**	**1.7742**	**3.1539**
7	**0.95**	**0.1071**	**1.6487**	**3.1759**
8	**0.95**	**0.0643**	**2.0384**	**4.1914**
9	**0.95**	**0.0519**	**1.9419**	**4.2171**
10	**0.95**	**0.0348**	**2.2907**	**5.2126**
2	**0.9**	**2.9200**	**1.0000**	**0.8425**
3	**0.9**	**1.2500**	**0.8535**	**1.0272**
4	**0.9**	**0.3822**	**1.4464**	**2.1279**
5	**0.9**	**0.2615**	**1.2848**	**2.1635**
6	**0.9**	**0.1302**	**1.7230**	**3.2051**
7	**0.9**	**0.0992**	**1.6094**	**3.2362**
8	**0.9**	**0.0601**	**1.9790**	**4.2393**
9	**0.9**	**0.0484**	**1.8948**	**4.2796**
10	**0.9**	**0.0327**	**2.2250**	**5.2583**
2	**0.5**	**1.0000**	**1.0000**	**1.5000**
3	**0.5**	**0.5063**	**0.9697**	**1.7778**
4	**0.5**	**0.2400**	**1.2000**	**2.5000**
5	**0.5**	**0.1488**	**1.2613**	**2.8800**
6	**0.5**	**0.0918**	**1.4286**	**3.5000**
7	**0.5**	**0.0638**	**1.5112**	**3.9184**
8	**0.5**	**0.0444**	**1.6557**	**4.5000**
9	**0.5**	**0.0331**	**1.7456**	**4.9383**
10	**0.5**	**0.0248**	**1.8778**	**5.5000**
2	**0.1**	**2.9200**	**1.0000**	**0.8425**
3	**0.1**	**0.6265**	**1.7030**	**2.0495**
4	**0.1**	**0.3822**	**1.4464**	**2.1279**
5	**0.1**	**0.1790**	**1.8769**	**3.1604**
6	**0.1**	**0.1302**	**1.7230**	**3.2051**
7	**0.1**	**0.0768**	**2.0801**	**4.1826**
8	**0.1**	**0.0601**	**1.9790**	**4.2393**
9	**0.1**	**0.0399**	**2.2982**	**5.1909**
10	**0.1**	**0.0327**	**2.2250**	**5.2583**
2	**0.05**	**3.4300**	**1.0000**	**0.7915**
3	**0.05**	**0.7023**	**1.7371**	**1.9736**
4	**0.05**	**0.4200**	**1.4837**	**2.0714**
5	**0.05**	**0.1928**	**1.9466**	**3.1229**
6	**0.05**	**0.1404**	**1.7742**	**3.1539**
7	**0.05**	**0.0818**	**2.1578**	**4.1566**
8	**0.05**	**0.0643**	**2.0384**	**4.1914**
9	**0.05**	**0.0423**	**2.3807**	**5.1701**
10	**0.05**	**0.0348**	**2.2907**	**5.2126**

Table 4: Variances and relative efficiencies of ranked-set sample estimator

N	ξ	$\frac{Var\left(\hat{Q}(\xi)_{RSS}\right)}{\sigma^2}$	$\frac{Var\left(\hat{Q}(\xi)_{BLUE}\right)}{Var\left(\hat{Q}(\xi)_{RSS}\right)}$
2	0.95	3.4300	0.7915
3	0.95	1.2199	1.1362
4	0.95	0.6232	1.3961
5	0.95	0.3753	1.6043
6	0.95	0.2492	1.7777
7	0.95	0.1765	1.9263
8	0.95	0.1310	2.0562
9	0.95	0.1008	2.1717
10	0.95	0.0797	2.2755
2	0.9	2.9200	0.8425
3	0.9	1.0669	1.2035
4	0.9	0.5528	1.4712
5	0.9	0.3360	1.6839
6	0.9	0.2244	1.8602
7	0.9	0.1597	2.0108
8	0.9	0.1190	2.1421
9	0.9	0.0918	2.2586
10	0.9	0.0727	2.3633
2	0.5	1.0000	1.5000
3	0.5	0.4909	1.8333
4	0.5	0.2880	2.0833
5	0.5	0.1877	2.2833
6	0.5	0.1312	2.4500
7	0.5	0.0964	2.5929
8	0.5	0.0736	2.7179
9	0.5	0.0578	2.8290
10	0.5	0.0466	2.9290
2	0.1	2.9200	0.8425
3	0.1	1.0669	1.2035
4	0.1	0.5528	1.4712
5	0.1	0.3360	1.6839
6	0.1	0.2244	1.8602
7	0.1	0.1597	2.0108
8	0.1	0.1190	2.1421
9	0.1	0.0918	2.2586
10	0.1	0.0727	2.3633
2	0.05	3.4300	0.7915
3	0.05	1.2199	1.1362
4	0.05	0.6232	1.3961
5	0.05	0.3753	1.6043
6	0.05	0.2492	1.7777
7	0.05	0.1765	1.9263
8	0.05	0.1310	2.0562
9	0.05	0.1008	2.1717
10	0.05	0.0797	2.2755

CHAPTER THIRTEEN

Multistage Median Ranked Set Samples for Estimating the Population Mean

Abdul Aziz Jemain[1] and **Amer Ibrahim Al-Omari**[2]
School of Mathematical Sciences, Faculty of Science and Technology
43600 UKM Bangi, Selangor, Malaysia
Email: [1]azizj@pkrisc.cc.ukm.my, [2]alomari_amer@yahoo.com

ABSTRACT

Multistage median ranked set samples method (MMRSS) is considered. The estimator of population mean using MMRSS method is compared with that using simple random sapling (SRS) and ranked set sampling (RSS) methods. It is noted that the estimator of population mean using MMRSS is unbiased and more efficient than its counter parts for almost all distribution considered if the underlying distribution is symmetric. For asymmetric distributions considered in this study, MMRSS estimator has a smaller bias, and it's preferable for even sample size.

KEYWORDS

Ranked set sampling; median ranked set sampling, multistage median ranked set sampling.

1. INTRODUCTION

Ranked set sampling (RSS) method was first proposed by McIntyre (1952) for estimating the mean of pasture yields. McIntyre showed that the mean of m units in the ranked set sampling was unbiased and had a smaller variance than the mean of the same number of observations selected by simple random sampling. Hence the ranked set sampling is more efficient than simple random sampling when estimating the population mean. Takahasi and Wakimoto (1968) provided the mathematical properties of RSS. Dell and Clutter (1972) showed that RSS estimator is an unbiased for the population mean regardless of error in ranking. Muttlak (1997) suggested using median ranked set sampling (MRSS) method, and showed that MRSS estimator is more efficient than the usual RSS estimator based on the same sample size. Al-Saleh and Al-Omari (2002) introduced multistage ranked set sampling, that increase the relative efficiency for estimating the population mean for fixed sample size.

In this paper we introduce a new modification of RSS, namely, multistage median ranked set sampling. The usual sample mean is suggested as an estimator of the population mean using the MMRSS procedure.

2. MULTISTAGE MEDIAN RANKED SET SAMPLING

Multistage median ranked set sampling procedure can be described as follows:

Step 1: Randomly selected m^{r+1} sample units from the target population, where r is the number of stages.

Step 2: Allocate the m^{r+1} sample units as randomly as possible into m^r sets each of size m.

Step 3: For each m^r sets in step 2, if the sample size *m* is odd, select for measurement from each m^r sets the $((m+1)/2)$th smallest rank, i.e. median of the sample. If the sample size *m* is even, select for measurement from the first each $m^r/2$ sets the $(m/2)$th smallest rank and from each other $m^r/2$ sets the $((m+2)/2)$th smallest rank. This step yield m^{r-1} sets each of size *m*.

Step 4: Without doing any actual quantification, repeat step 3 on the m^{r-1} ranked set to obtain m^{r-2} second stage ranked sets, each of size m. The process is continued using step 3 up to the rth stage to get a sample of size *m* from MMRSS.

Finally, the m units identified in step 4 are actually measured only for estimating the mean of the variable of interest. The whole process can be repeated k times to obtain an MMRSS of size $n = km$.

To clarify this method, let $X_i^{(r)j}$ be the *i*th sample unit of the jth set at stage *r*.

Example 1: Consider the case of $m = 3$ and $r = 2$, so that we have 27 units.

$$X_1^{(0)}, X_2^{(0)}, X_3^{(0)}, X_4^{(0)}, X_5^{(0)}, X_6^{(0)}, X_7^{(0)}, X_8^{(0)}, X_9^{(0)}, X_{10}^{(0)}, X_{11}^{(0)}, X_{12}^{(0)}, X_{13}^{(0)}, X_{14}^{(0)}$$
$$X_{15}^{(0)}, X_{16}^{(0)}, X_{17}^{(0)}, X_{18}^{(0)}, X_{19}^{(0)}, X_{20}^{(0)}, X_{21}^{(0)}, X_{22}^{(0)}, X_{23}^{(0)}, X_{24}^{(0)}, X_{25}^{(0)}, X_{26}^{(0)}, X_{27}^{(0)}.$$

Allocate them into 9 sets each of size 3 at zero stage (SRS), and then rank visually the units within each sample with respect to the variable of interest as following

Thus, the set $\left\{Y_{(m+1)/2}^{(2)1}, Y_{(m+1)/2}^{(2)2}, Y_{(m+1)/2}^{(2)3}\right\}$ is a second stage median ranked set samples. The actual quantified for estimating the variable of interest will achieved using only these three units. Thus, the number of quantified units, which is 3, is a small portion of the number of sampled units, which is 27, but all sampled units add to the information content of the quantified units. Hence, we compare this sample with a sample of size 3, not 27, of SRS.

Figure 1: Display of 27 units in 9 sets up to 2 stage using MMRSSO

Sets	First stage	Second stage
$A^{(0)1}=\left\{X_{(1)}^{(0)1}, X_{(2)}^{(0)1}, X_{(3)}^{(0)1}\right\}$	$Y_{(m+1)/2}^{(1)1}=\text{med}\left(A^{(0)1}\right)$	
$A^{(0)2}=\left\{X_{(1)}^{(0)2}, X_{(2)}^{(0)2}, X_{(3)}^{(0)2}\right\}$	$Y_{(m+1)/2}^{(1)2}=\text{med}\left(A^{(0)2}\right)$	$Y_{(m+1)/2}^{(2)1}=\text{med}\left\{Y_{(m+1)/2}^{(1)1}, Y_{(m+1)/2}^{(1)2}, Y_{(m+1)/2}^{(1)3}\right\}$
$A^{(0)3}=\left\{X_{(1)}^{(0)3}, X_{(2)}^{(0)3}, X_{(3)}^{(0)3}\right\}$	$Y_{(m+1)/2}^{(1)3}=\text{med}\left(A^{(0)3}\right)$	
$A^{(0)4}=\left\{X_{(1)}^{(0)4}, X_{(2)}^{(0)4}, X_{(3)}^{(0)4}\right\}$	$Y_{(m+1)/2}^{(1)4}=\text{med}\left(A^{(0)4}\right)$	
$A^{(0)5}=\left\{X_{(1)}^{(0)5}, X_{(2)}^{(0)5}, X_{(3)}^{(0)5}\right\}$	$Y_{(m+1)/2}^{(1)5}=\text{med}\left(A^{(0)5}\right)$	$Y_{(m+1)/2}^{(2)2}=\text{med}\left\{Y_{(m+1)/2}^{(1)4}, Y_{(m+1)/2}^{(1)5}, Y_{(m+1)/2}^{(1)6}\right\}$
$A^{(0)6}=\left\{X_{(1)}^{(0)6}, X_{(2)}^{(0)6}, X_{(3)}^{(0)6}\right\}$	$Y_{(m+1)/2}^{(1)6}=\text{med}\left(A^{(0)6}\right)$	
$A^{(0)7}=\left\{X_{(1)}^{(0)7}, X_{(2)}^{(0)7}, X_{(3)}^{(0)7}\right\}$	$Y_{(m+1)/2}^{(1)7}=\text{med}\left(A^{(0)7}\right)$	
$A^{(0)8}=\left\{X_{(1)}^{(0)8}, X_{(2)}^{(0)8}, X_{(3)}^{(0)8}\right\}$	$Y_{(m+1)/2}^{(1)8}=\text{med}\left(A^{(0)8}\right)$	$Y_{(m+1)/2}^{(2)3}=\text{med}\left\{Y_{(m+1)/2}^{(1)7}, Y_{(m+1)/2}^{(1)8}, Y_{(m+1)/2}^{(1)9}\right\}$
$A^{(0)9}=\left\{X_{(1)}^{(0)9}, X_{(2)}^{(0)9}, X_{(3)}^{(0)9}\right\}$	$Y_{(m+1)/2}^{(1)9}=\text{med}\left(A^{(0)9}\right)$	

3. GENERAL SETUP AND SOME BASIC RESULTS

Let $X_{11}, X_{12}, ..., X_{1m}$; $X_{21}, X_{22}, ..., X_{2m}$;…; $X_{m1}, X_{m2}, ..., X_{mm}$; be m independent random samples each of size m, assume that each variable has the same distribution function $F(x)$ with mean μ and variance σ^2. Let $X_{i(1)}, X_{i(2)}, ..., X_{i(m)}$ $(i=1,2,...,m)$ be the ordered statistics of the ith sample $X_{i1}, X_{i2}, ..., X_{im}$ $(i=1,2,...,m)$. Let $Y_1, Y_2, ..., Y_m$ be RSS, then $Y_i \overset{d}{=} X_{(i)}$.

At the kth cycle $(k=1,2,...,n)$, for odd sample size, let $Y^{(r)}_{i((m+1)/2)k}$ be the median of the ith sample $(i=1,2,...,m)$ at stage r.

Thus, the quantified sample $Y^{(r)}_{1\left(\frac{m+1}{2}\right)k}$, $Y^{(r)}_{2\left(\frac{m+1}{2}\right)k}$, ..., $Y^{(r)}_{m\left(\frac{m+1}{2}\right)k}$ will denote the MMRSSO.

For even sample size, let $Y^{(r)}_{i(m/2)k}$ be the $(m/2)$th order statistic of the ith sample $(i=1,2,...,l; l=m/2)$, and let $Y^{(r)}_{i((m+2)/2)k}$ be the $((m+2)/2)$th order statistic of the ith sample $(i=l+1,l+2,...,m)$ each at stage r. Thus, the quantified sample

$$Y^{(r)}_{1\left(\frac{m}{2}\right)k},\ Y^{(r)}_{2\left(\frac{m}{2}\right)k},\ \ldots,\ Y^{(r)}_{\frac{m}{2}\left(\frac{m}{2}\right)k},\ Y^{(r)}_{\frac{m}{2}+1\left(\frac{m}{2}+1\right)k},\ldots,Y^{(r)}_{m\left(\frac{m}{2}+1\right)k}$$

will denote the MMRSSE.

The estimator of the population mean μ using multistage median ranked set samples can be defined as

$$\hat{\mu}_{MMRSS}=\begin{cases}\overline{Y}^{(r)}_{MMRSSO}=\dfrac{1}{mn}\sum\limits_{k=1}^{n}\sum\limits_{i=1}^{m}Y^{(r)}_{i((m+1)/2)k}, & \text{if } m \text{ is odd}\\ \overline{Y}^{(r)}_{MMRSSE}=\dfrac{1}{mn}\sum\limits_{k=1}^{n}\left(\sum\limits_{i=1}^{l}Y^{(r)}_{i(m/2)k}+\sum\limits_{i=l+1}^{m}Y^{(r)}_{i((m+2)/2)k}\right), & \text{if } m \text{ is even}, l=m/2.\end{cases} \tag{3.1}$$

Assume that the cycle is repeated once; let us define the following notations:

Let $\mu=E(X_i)$, $\sigma^2=\text{Var}(X_i)$, $(i=1,2,...,m)$, $\mu^{(r)}_{(m+1)/2}=E\left(Y^{(r)}_{i(m+1)/2}\right)$, $\sigma^{2(r)}_{(m+1)/2}=\text{Var}\left(Y^{(r)}_{i(m+1)/2}\right)$, $\mu^{(r)}_{(m/2)}=E\left(Y^{(r)}_{i(m/2)}\right)$, $\sigma^{2(r)}_{(m/2)}=\text{Var}\left(Y^{(r)}_{i(m/2)}\right)$, $\mu^{(r)}_{(m+2)/2}=E\left(Y^{(r)}_{i(m+2)/2}\right)$, $\sigma^{2(r)}_{(m+2)/2}=\text{Var}\left(Y^{(r)}_{i(m+2)/2}\right)$.

Based on these notations we have

$$\overline{Y}^{(r)}_{MMRSSO}=\frac{1}{m}\overline{Y}^{(r)}_{(m+1)/2}, \tag{3.2}$$

$$E\left(\overline{Y}^{(r)}_{MMRSSO}\right)=\frac{1}{m}\mu^{(r)}_{(m+1)/2} \quad (3.3)$$

$$\text{Var}\left(\overline{Y}^{(r)}_{MMRSSO}\right)=\frac{1}{m}\sigma^{2(r)}_{(m+1)/2}. \quad (3.4)$$

$$\overline{Y}^{(r)}_{MMRSSE}=\frac{1}{2}\left(\overline{Y}^{(r)}_{(m/2)}+\overline{Y}^{(r)}_{(m+2)/2}\right), \quad (3.5)$$

$$E\left(\overline{Y}^{(r)}_{MMRSSE}\right)=\frac{1}{2}\left(\mu^{(r)}_{(m/2)}+\mu^{(r)}_{(m+2)/2}\right), \quad (3.6)$$

$$\text{Var}\left(\overline{Y}^{(r)}_{MMRSSE}\right)=\frac{1}{2m}\left(\sigma^{2(r)}_{(m/2)}+\sigma^{2(r)}_{(m+2)/2}\right). \quad (3.7)$$

For a random sample from a continuous population whose pdf is symmetrical about $x=\mu$, (H. A. David and H. N. Nagaraja (2003)), showed that $f_{(i)}(\mu+x)=f_{(m-i+1)}(\mu-x)$. Assume that the distribution is symmetric about $x=0$, then $X_{(i)}\overset{d}{=}-X_{(m-i+1)}$. Hence, $\mu_{(i)}=-\mu_{(m-i+1)}$ and $\sigma^2_{(i)}=\sigma^2_{(m-i+1)}$.

This implies to, $\mu^{(r)}_{(m/2)}=-\mu^{(r)}_{(m+2)/2}$ and $\sigma^{2(r)}_{(m/2)}=\sigma^{2(r)}_{(m+2)/2}$, and if m is odd, $\mu^{(r)}_{(m+1)/2}=\mu=0$. Therefore, $E\left(\overline{Y}^{(r)}_{MMRSSE}\right)=0$, $E\left(\overline{Y}^{(r)}_{MMRSSO}\right)=0$, and $Var\left(\overline{Y}^{(r)}_{MMRSSE}\right)=\frac{1}{m}\sigma^{2(r)}_{(m/2)}$.

The mean square error (MSE) of the estimator $\overline{Y}^{(r)}_{MMRSS}$ is given by

$$\text{MSE}\left(\overline{Y}^{(r)}_{MMRSS}\right)=\text{Var}\left(\overline{Y}^{(r)}_{MMRSS}\right)+\left(\text{Bias}\left(\overline{Y}^{(r)}_{MMRSS}\right)\right)^2 \quad (3.8)$$

Note that, for symmetric distributions

$\text{Bias}\left(\overline{Y}^{(r)}_{MMRSS}\right)=0$, and $\text{MSE}\left(\overline{Y}^{(r)}_{MMRSS}\right)=\text{Var}\left(\overline{Y}^{(r)}_{MMRSS}\right)$.

Lemma 1:

1) If the distribution is symmetric about the population mean, then $\overline{Y}^{(r)}_{MMRSSO}$ and $\overline{Y}^{(r)}_{MMRSSE}$ are unbiased estimators of a population mean, i.e. $E\left(\overline{Y}^{(r)}_{MMRSSO}\right)=\mu$ and $E\left(\overline{Y}^{(r)}_{MMRSSE}\right)=\mu$.

2) If the distribution is symmetric about the population mean, then the relative efficiency of $\bar{Y}^{(r)}_{MMRSSO}$ and $\bar{Y}^{(r)}_{MMRSSE}$ are increasing in r $(r \geq 1)$, except for the uniform distribution, the lemma is true if $r \geq 2$.

Proof:

To prove 1, assume m is odd, then we have

$$E\left(\bar{Y}^{(r)}_{MMRSSO}\right) = E\left(\frac{1}{m}\sum_{i=1}^{m} Y^{(r)}_{i((m+1)/2)}\right) = \frac{1}{m}\sum_{i=1}^{m} E\left(Y^{(r)}_{i((m+1)/2)}\right)$$

$$= \frac{1}{m}\left(\sum_{i=1}^{m} \mu^{(r)}_{i((m+1)/2)}\right) = \frac{1}{m}\left(\sum_{i=1}^{m} \mu\right) = \mu$$

In the case of m is even, we have

$$E\left(\bar{Y}^{(r)}_{MMRSSE}\right) = E\left(\frac{1}{m}\left(\sum_{i=1}^{l} Y^{(r)}_{i(m/2)} + \sum_{i=l+1}^{m} Y^{(r)}_{i((m+2)/2)}\right)\right)$$

$$= \frac{1}{m}\left(\sum_{i=1}^{l} E\left(Y^{(r)}_{(i(m/2))}\right) + \sum_{i=l+1}^{m} E\left(Y^{(r)}_{i((m+2)/2)}\right)\right)$$

$$= \frac{1}{m}\left(\sum_{i=1}^{l} \mu^{(r)}_{i(m/2)} + \sum_{i=l+1}^{m} \mu^{(r)}_{i((m+2)/2)}\right)$$

Since the distribution is symmetric about the population mean, and $\bar{Y}^{(r)}_{MMRSSO}$, $\bar{Y}^{(r)}_{MMRSSE}$ are unbiased estimators, then we have $\mu^{(r)}_{i(m/2)} = \mu - \varepsilon$ and $\mu^{(r)}_{i((m+2)/2)} = \mu + \varepsilon$, where ε is real number. Therefore,

$$E\left(\bar{Y}^{(r)}_{MMRSSE}\right) = \frac{1}{m}\left(\sum_{i=1}^{l} (\mu - \varepsilon) + \sum_{i=l+1}^{m} (\mu + \varepsilon)\right) = \mu .$$

To prove 2, when m is odd, Let $Y^{(r)}_{1\left(\frac{m+1}{2}\right)}, Y^{(r)}_{2\left(\frac{m+1}{2}\right)}, \ldots, Y^{(r)}_{\frac{m+1}{2}\left(\frac{m+1}{2}\right)}, \ldots, Y^{(r)}_{m\left(\frac{m+1}{2}\right)}$ be an MMRSSO at stage r. Let $\operatorname{Var}\left(\bar{Y}^{(r)}_{MMRSSO}\right) = \frac{1}{m^2}\sum_{i=1}^{m} \operatorname{Var}\left(Y^{(r)}_{i(m+1)/2}\right)$.

Let $Z_i^{(r-1)}$ be the ith median of the sample $Y_{1\left(\frac{m+1}{2}\right)}^{(r-1)}$, $Y_{2\left(\frac{m+1}{2}\right)}^{(r-1)}, \ldots, Y_{\frac{m+1}{2}\left(\frac{m+1}{2}\right)}^{(r-1)}, \ldots,$ $Y_{m\left(\frac{m+1}{2}\right)}^{(r-1)}$, and let $\bar{Z}^{(r-1)} = \frac{1}{m}\sum_{i=1}^{m} Z_i^{(r-1)}$. Then, $\bar{Z}^{(r-1)} = \bar{Y}^{(r-1)}$, $\operatorname{Var}\left(\bar{Y}_{MMRSSO}^{(r-1)}\right) = \operatorname{Var}\left(\bar{Z}^{(r-1)}\right)$ and $Y_i^{(r)}$ has the same distribution as $Z_i^{(r-1)}$.

Hence,

$$\begin{aligned}\operatorname{Var}\left(\bar{Y}_{MMRSSO}^{(r-1)}\right) &= \frac{1}{m^2}\sum_{i=1}^{m}\operatorname{Var}\left(Z_i^{(r-1)}\right) + \frac{1}{m^2}\sum_{i\neq j}^{m}\operatorname{Cov}\left(Z_i^{(r-1)}, Z_j^{(r-1)}\right)\\ &= \frac{1}{m^2}\sum_{i=1}^{m}\operatorname{Var}(Y_i^{(r)}) + \frac{1}{m^2}\sum_{i\neq j}^{m}\operatorname{Cov}\left(Z_i^{(r-1)}, Z_j^{(r-1)}\right)\end{aligned}$$

Since, $\operatorname{Cov}\left(Z_i^{(r-1)}, Z_j^{(r-1)}\right) \geq 0$ (Lehman (1966), Essary et al. (1997)), and Yang (1982) showed that for the median of iid sample, $\operatorname{Var}\left(X_{\frac{m+1}{2}}\right) \leq \sigma^2$. Therefore, $\operatorname{Var}\left(\bar{Y}_{MMRSSO}^{(r)}\right) \leq \operatorname{Var}\left(\bar{Y}_{MMRSSO}^{(r-1)}\right)$ and hence the relative efficiency of $\bar{Y}_{MMRSSO}^{(r)}$ is increasing in r.

In the case of even sample size, let

$$Z_i^{(r-1)} = \begin{cases} \left(\frac{m}{2}\right)\text{th of}\left(Y_{i\left(\frac{m}{2}\right)}^{(r-1)}\right), & 1 \leq i \leq \frac{m}{2} \\ \left(\frac{m+2}{2}\right)\text{th of}\left(Y_{i\left(\frac{m+2}{2}\right)}^{(r-1)}\right), & \frac{m+2}{2} \leq i \leq m. \end{cases}$$

Then the proof is directly as in the case of odd sample size.

Corollary:

1) If the distribution is symmetric about the population mean, then $\operatorname{Var}\left(\bar{Y}_{MMRSSO}^{(r)}\right)$ and $\operatorname{Var}\left(\bar{Y}_{MMRSSE}^{(r)}\right)$ are less than $\operatorname{Var}(\bar{X}_{SRS})$ at any stage r.

2) For asymmetric distribution about a population mean, we have $\operatorname{MSE}\left(\bar{Y}_{MMRSSO}^{(r)}\right) < \operatorname{Var}\left(\bar{X}_{SRS}\right)$ and $\operatorname{MSE}\left(\bar{Y}_{MMRSSE}^{(r)}\right) < \operatorname{Var}\left(\bar{X}_{SRS}\right)$.

4. EFFICIENCY OF MMRSS METHOD

Let $X_{i1}, X_{i2}, ..., X_{im}$ be independent random sample with cdf *F(x)* and pdf *f(x)*, with mean μ and variance σ^2.

The SRS estimator of the population mean μ is given by

$$\bar{X}_{SRS} = \frac{1}{m}\sum_{i=1}^{m} X_i , \qquad (4.1)$$

with mean $E\left(\bar{X}_{SRS}\right) = \mu$ and variance

$$\text{Var}\left(\bar{X}_{SRS}\right) = \frac{\sigma^2}{m} . \qquad (4.2)$$

The estimator of the population mean μ using RSS is defined as

$$\bar{X}_{RSS} = \frac{1}{m}\sum_{i=1}^{m} X_{i(i)} , \qquad (4.3)$$

with mean $E\left(\bar{X}_{RSS}\right) = \mu$ and variance

$$\text{Var}\left(\bar{X}_{RSS}\right) = \frac{\sigma^2}{m} - \frac{1}{m^2}\sum_{i=1}^{m}\left(\mu_i - \mu\right)^2 . \qquad (4.4)$$

From (3.7) and (4.2), if the parent distribution is symmetric about its mean, the relative efficiency of $\bar{Y}^{(r)}_{MMRSS}$ with respect to $\bar{X}_{SRS}$ is given by

$$eff\left(\bar{X}_{SRS}, \bar{Y}_{RSS}\right) = \frac{\text{Var}\left(\bar{X}_{SRS}\right)}{\text{Var}\left(\bar{Y}_{RSS}\right)} \text{ and } eff\left(\bar{X}_{SRS}, \bar{Y}^{(r)}_{MMRSS}\right) = \frac{\text{Var}\left(\bar{X}_{SRS}\right)}{\text{Var}\left(\bar{Y}^{(r)}_{MMRSS}\right)} \qquad (4.5)$$

and if the distribution is non symmetric, using (3.8) and (4.2) the relative efficiency is given by

$$eff(\bar{X}_{SRS}, \bar{Y}^{(r)}_{MMRSS}) = \frac{\text{MSE}\left(\bar{X}_{SRS}\right)}{\text{MSE}\left(\bar{Y}^{(r)}_{MMRSS}\right)} . \qquad (4.6)$$

4.1. Results for Uniform Distribution

Assume that the variable of interest *X* has a uniform $(0, \theta)$, namely,

$$f(x;\theta)=\begin{cases}1/\theta & 0\le x\le\theta;\theta>0,\\ 0 & \text{elsewhere.}\end{cases}$$

So that $\mu=\frac{\theta}{2}$ and $\sigma^2=\frac{\theta^2}{12}$, and $X_{i:m}$ has a beta$(i,m-i+1)$ with pdf

$$f_{i:m}(x)=\frac{m!}{(i-1)!(m-i)!}\left(\frac{x}{\theta}\right)^{i-1}\left(1-\frac{x}{\theta}\right)^{m-i}\frac{1}{\theta},\ 0<x<\theta.$$

with meant $E\left(X_{i:m}\right)=\frac{i\theta}{m+1}$ and variance

$$\mathrm{Var}\left(X_{i:m}\right)=\frac{i(m-i+1)\theta^2}{(m+1)^2(m+2)}. \tag{4.1.1}$$

From (4.1) and (4.2), the SRS estimator of the population mean $\mu=\frac{\theta}{2}$ from a sample of size m, has mean $\theta/2$, and the variance is given by

$$\mathrm{Var}\left(\bar{X}_{SRS}\right)=\frac{1}{m^2}\sum_{i=1}^{m}\mathrm{Var}(X_{i:m})=\frac{1}{m^2}\left(\frac{m\theta^2}{12}\right)=\frac{\theta^2}{12m}. \tag{4.1.2}$$

From (4.4), the RSS estimator of the population mean $\mu=\frac{\theta}{2}$ has mean

$$E\left(\bar{Y}_{RSS}\right)=\frac{1}{m}\sum_{i=1}^{m}E(Y_i)=\frac{\theta}{2}$$

and variance given by

$$\mathrm{Var}\left(\bar{Y}_{RSS}\right)=\frac{1}{m^2}\sum_{i=1}^{m}Var\left(X_{(i:m)}\right)=\frac{\theta^2}{6m(m+1)} \tag{4.1.3}$$

In the case of even sample size, let $m\ge 4$ and $r=1$, from (3.3) and (4.1.1) the estimator of the population mean $\mu=\frac{\theta}{2}$ using MMRSSE is

$$\bar{Y}_{MMRSSE}^{(r)}=\frac{1}{m}\left(\sum_{i=1}^{l}Y_{i(m/2)}^{(r)}+\sum_{i=l+1}^{m}Y_{i((m+2)/2)}^{(r)}\right)$$

with variance given by $\mathrm{Var}\left(\bar{Y}_{MMRSSE}^{(r)}\right)=\frac{\theta^2}{4(m+1)^2}$. (4.1.4)

Using equations (4.1.2) and (4.1.4), the relative efficiency of $\bar{Y}^{(r)}_{MMRSSE}$ with respect to $\bar{X}_{SRS}$ for estimating the population mean is defined as

$$eff\left(\bar{X}_{SRS}, \bar{Y}^{(r)}_{MMRSSE}\right) = \frac{\operatorname{Var}\left(\bar{X}_{SRS}\right)}{\operatorname{Var}\left(\bar{Y}^{(r)}_{MMRSSE}\right)} = \frac{(m+1)^2}{3m} > 1 .$$

This implies that the variance of the sample mean using MMRSSE for estimating the population mean is less than the variance of the sample mean using SRS. Assume that the parent distribution is U(0,1), let $m = 4$ and $r = 1$, from (4.1.1) we have, $\sigma^{2(1)}_{2:4} = \dfrac{1}{25}$ and $\sigma^{2(1)}_{3:4} = \dfrac{1}{25}$, and

$$\begin{aligned}\bar{Y}^{(1)}_{MMRSSE} &= \frac{1}{4}\left(\sum_{i=1}^{2} Y^{(1)}_{i(2)} + \sum_{i=3}^{4} Y^{(1)}_{i(3)}\right) \\ &= \frac{1}{4}\left(Y^{(1)}_{1(2)} + Y^{(1)}_{2(2)} + Y^{(1)}_{3(3)} + Y^{(1)}_{4(3)}\right),\end{aligned}$$

$$\operatorname{Var}\left(\bar{Y}^{(1)}_{MMRSSE}\right) = \frac{1}{16}\left(2\cdot\frac{1}{25} + 2\cdot\frac{1}{25}\right) = \frac{1}{100} .$$

And the variance of a SRS of size $m = 4$ from U(0,1) is $\operatorname{Var}\left(\bar{X}_{SRS}\right) = 0.0208$. Thus, the relative efficiency of $\bar{Y}^{(1)}_{MMRSSE}$ with respect to $\bar{X}_{SRS}$ is $eff\left(\bar{X}_{SRS}, \bar{Y}^{(1)}_{MMRSSE}\right) = 2.083$ which agrees with simulation results.

For odd sample size, if $m \geq 3$ and $r = 1$, from (3.2), the estimator of the population mean $\mu = \dfrac{\theta}{2}$ using MMRSSO is $\bar{Y}^{(1)}_{MMRSSO} = \dfrac{1}{m}\sum_{i=1}^{m} Y^{(1)}_{i((m+1)/2)}$ with variance

$$\operatorname{Var}\left(\bar{Y}^{(r)}_{MMRSSO}\right) = \frac{\theta^2}{4m(m+2)} . \qquad (4.1.5)$$

From (4.1.2) and (4.1.5), the relative efficiency of $\bar{Y}^{(r)}_{MMRSSO}$ with respect to $\bar{X}_{SRS}$ for estimating the population mean is given by

$$eff\left(\bar{X}_{SRS}, \bar{Y}^{(r)}_{MMRSSO}\right) = \frac{\operatorname{Var}\left(\bar{X}_{SRS}\right)}{\operatorname{Var}\left(\bar{Y}^{(r)}_{MMRSSO}\right)} = \frac{(m+2)}{3} > 1 .$$

Clearly, that the variance of the sample means using MMRSSO method for estimating the population mean is less than the variance of the sample mean using SRS method.

For *m=3* and $r=1$, assume the parent distribution is U(0,1), from (4.1.1) we have

$$\sigma_{(2)}^{2(1)} = \frac{1}{20},$$

and $$\overline{Y}_{MMRSSO}^{(1)} = \frac{1}{3}\sum_{i=1}^{m} Y_{i(2)k}^{(1)} = Y_{1(2)}^{(1)} + Y_{2(2)}^{(1)} + Y_{3(2)}^{(1)},$$

$$\operatorname{Var}\left(\overline{Y}_{MMRSSO}^{(1)}\right) = \sigma_{\left(\frac{m+1}{2}\right)}^{2(1)} = \frac{1}{9}\cdot 3\cdot\frac{1}{20} = \frac{1}{60}.$$

The variance of a SRS of size, $m=3$, is 0.028. Hence, the relative efficiency of $\overline{Y}_{MMRSSO}^{(1)}$ with respect to $\overline{X}_{SRS}$ is given by $eff\left(\overline{X}_{SRS}, \overline{Y}_{MMRSSO}^{(1)}\right) = 1.666$. This agrees with simulation result. If *m=5*, we have $eff\left(\overline{X}_{SRS}, \overline{Y}_{MMRSSO}^{(1)}\right) = 2.333$, which indicate that the efficiency of $\overline{Y}_{MMRSSO}^{(r)}$ is increasing in the sample size.

5. SIMULATION STUDY

To compare the relative efficiency of the proposed estimators for the population mean using MMRSS against usual estimators using RSS and SRS, we compared the average of 70.000 sample estimates with sample sizes $m=3,4$ and 5 for $r=1,2,3$ and 4. Eight distributions, namely, uniform, normal, logistic, exponential, lognormal, weibull, beta and gamma are considered. The relative efficiency of the sampling methods considered in this study can be computed using (4.5) and (4.6).

Table 1
The relative efficiency of RSS and MMRSS estimators for estimating the population mean of symmetric distributions

Distribution	RSS		MMRSS			
	m		$r=1$	$r=2$	$r=3$	$r=4$
Uniform (0,1)	3	2.000	1.672	3.152	6.366	13.626
	4	2.500	2.085	5.530	20.973	45.022
	5	3.000	2.367	6.876	21.351	79.421
Normal (0,1)	3	1.917	2.235	4.969	11.219	25.030
	4	2.367	2.780	7.674	20.973	56.599
	5	2.734	3.441	11.906	39.515	147.782
Normal (1,2)	*3*	1.916	2.196	4.982	11.202	24.813
	4	2.346	2.768	7.697	20.669	56.745
	5	2.797	3.512	12.294	40.819	151.347
Logistic (-1,1)	*3*	1.843	2.544	6.168	14.059	31.518
	4	2.239	3.152	8.904	24.297	65.155
	5	2.550	4.131	15.397	50.784	189.983

Table 2
The relative efficiency of RSS and MMRSS estimators for estimating the population mean of non symmetric distributions

Distribution			RSS	MMRSS			
	m			$r=1$	$r=2$	$r=3$	$r=4$
Exponential (1)	3	*RP*	1.606	2.218	3.074	3.372	3.444
		Bias	0.000	0.166	0.244	0.279	0.295
	4	*RP*	1.948	2.473	3.890	5.589	7.482
		Bias	0.000	0.168	0.196	0.187	0.173
	5	*RP*	2.159	2.188	2.201	2.125	2.105
		Bias	0.000	0.218	0.281	0.300	0.305
LogNormal (0,1)	3	*RP*	1.292	3.485	4.100	3.955	3.863
		Bias	0.000	0.397	0.545	0.607	0.631
	4	*RP*	1.447	3.318	4.274	5.158	6.009
		Bias	0.000	0.398	0.460	0.450	0.430
	5	*RP*	1.533	2.703	2.322	2.190	2.145
		Bias	0.000	0.496	0.609	0.638	0.648
Weibull (1,3)	3	*RP*	1.629	2.257	3.131	3.418	3.519
		Bias	0.000	0.500	0.731	0.835	0.881
	4	*RP*	1.943	2.447	3.863	5.584	7.433
		Bias	0.000	0.498	0.586	0.561	0.520
	5	*RP*	2.191	2.239	2.240	2.174	2.147
		Bias	0.000	0.652	0.843	0.897	0.914

Distribution			**RSS**	**MMRSS**			
	m			$r=1$	$r=2$	$r=3$	$r=4$
Beta (7,4)	3	*RP*	2.000	2.157	4.564	9.406	18.695
		Bias	0.000	0.005	0.007	0.008	0.008
	4	*RP*	2.394	2.605	6.766	17.765	43.601
		Bias	0.000	0.004	0.005	0.005	0.005
	5	*RP*	2.765	3.034	9.130	22.232	37.638
		Bias	0.000	0.006	0.008	0.008	0.009
Gamma (3,1)	3	*RP*	1.827	2.268	4.020	6.036	7.583
		Bias	0.000	0.180	0.260	0.296	0.312
	4	*RP*	2.202	2.649	5.541	10.043	10.138
		Bias	0.000	0.178	0.208	0.202	0.203
	5	*RP*	2.519	2.859	4.547	5.280	5.548
		Bias	0.000	0.233	0.298	0.318	0.324

From Tables 3.1 and 3.2 we can conclude the following:

1. It can be observed that the estimator of the population mean obtained from MMRSS is more efficient than the usual SRS and RSS estimators of population mean. i.e.,
$$\mathrm{Var}\left(\bar{Y}^{(r)}_{MMRSSO}\right) < \mathrm{Var}(\bar{X}_{SRS}) \text{ and } \mathrm{Var}\left(\bar{Y}^{(r)}_{MMRSSE}\right) < \mathrm{Var}(\bar{X}_{SRS}).$$

2. For uniform (0,1) distribution, only at the first stage, the relative efficiency of MMRSS estimator is less than the relative efficiency of the usual RSS estimator of population mean.

3. The relative efficiency of the MMRSS estimators is increasing in r for fixed sample size. As an example if the underlying distribution is normal (1,2), the values of the relative efficiency using $r=1,2,3$ and 4 with $m=3$ are 2.196, 4.982, 11.202 and 24.813 respectively. This emphasized that
$$\mathrm{Var}\left(\bar{Y}^{(r)}_{MMRSS}\right) \le \mathrm{Var}\left(\bar{Y}^{(r-1)}_{MMRSS}\right) \le \cdots \le \mathrm{Var}\left(\bar{Y}^{(1)}_{MMRSS}\right) \le \mathrm{Var}\left(\bar{Y}^{(0)}_{MMRSS}\right).$$

4. For non symmetric distributions the MMRSS estimator for estimating the population mean has a smaller bias, as an example for beta distribution with parameters 7 and 4, the relative efficiency using MMRSS method with $m=3$ at the first stage is 2.157 with bias 0.005 while at the 4th stage with the same sample size, the relative efficiency is 18.695 with bias 0.008. i.e. that,
$$\mathrm{MSE}\left(\bar{Y}^{(4)}_{MMRSS}\right) \le \mathrm{MSE}\left(\bar{Y}^{(3)}_{MMRSS}\right) \le \cdots \le \mathrm{MSE}\left(\bar{Y}^{(0)}_{MMRSS}\right) = \mathrm{MSE}\left(\bar{X}_{SRS}\right).$$

6. CONCLUDING REMARKS

Gain in efficiency is attained for estimating the population mean using median multistage ranked set samples, specially if the underlying distribution is symmetric about its mean. For asymmetric (skewed) distributions, the gain in efficiency is substantial with even sample size. The use of multistage median ranked set samples method is feasible for estimating the population mean if odd sample size is considered, because we only identify the median of the sample.

7. ACKNOWLEDGMENTS

We are grateful to a referee and the editor of Pakistan Journal of Statistics who provided very constructive comments.

REFERENCES

1. Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. *J. Statist. Plann. and Infer.*, 102, 273-286.
2. Dell, T.R., and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrika,* 28, 545-555.
3. Essary, J., Proschan, F., Walkup, D., (1967). Association of random variables, with application. *Ann. Math. Statist,* 38, 1466-1474.
4. David, H.A. and Nagaraja, H.N. (2003), *Order Statistics*, Third Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.
5. McIntyre, G.A. (1952). A method for unbiased selective sampling using, ranked sets. *Aust. J. of Agri. Res.*, 3, 385-390.
6. Muttlak, H.A. (1997). Median ranked set sampling, *J. Appl. Statist. Scien.*, 6(4), 577-586.
7. Samawi, H., Abu-Dayyeh, W., and Ahmed, S. (1996). Extreme ranked set sampling, *The Biometrical Journal* 30, 577-586.
8. Takahasi, K. and Wakimoto, K. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.*, 20, 1-31.

CHAPTER FOURTEEN

Double Quartile Ranked Set Samples

Abdul Aziz Jemain[1] and **Amer Ibrahim Al-Omari**[2]
School of Mathematical Sciences, Faculty of Science and Technology
43600 UKM Bangi, Selangor, Malaysia
Email: [1]azizj@pkrisc.cc.ukm.my; [2]alomari_amer@yahoo.com

ABSTRACT

Double quartile ranked set sampling procedure (DQRSS) and its properties for estimating the population mean are introduced. The performance of DQRSS with respect to simple random sapling (SRS), ranked set sampling (RSS) and quartile ranked set samples (QRSS) for estimating the population mean, is considered. The DQRSS estimator is unbiased of the population mean for symmetric distributions about its mean. In addition, the DQRSS method is more efficient than the SRS, RSS, and QRSS for all symmetric and asymmetric distributions considered in this study. For asymmetric distributions considered in this study, DQRSS estimator has a smaller bias.

KEYWORDS

Ranked set sampling, quartile ranked set sampling, double quartile ranked set sampling.

1. INTRODUCTION

McIntyre (1952) introduced ranked set sampling method for estimating the mean of pasture yields. In situations where the experimental or sampling units in a study can be more easily ranked than quantified, McIntyre proposed that the mean of m sample units based on a RSS as an estimator of the population mean. This estimator is unbiased estimator with a smaller variance compared to the usual sample mean based on a SRS of the same size. Takahasi and Wakimoto (1968) provided the mathematical properties of RSS. Dell and Clutter (1972) showed that RSS estimator is an unbiased for the population mean regardless of error in ranking. Samawi et al. (1996) suggested using extreme ranked set sampling (ERSS) for estimating a population mean, and showed that for symmetric distributions, the ERSS estimator is unbiased and has a smaller variance than the SRS estimator. Muttlak (1997) suggested using median ranked set sampling (MRSS) to increase the efficiency of the estimator and to reduce errors in ranking. Al-Saleh and Al-Kadiri (2000) introduced double ranked set sampling for estimating the population mean, they showed that the ranking in the second stage is easier than the ranking in the first stage. Al-Saleh and Al-Omari (2002) introduced multistage ranked set sampling, that increase the relative efficiency for estimating the population mean for fixed sample size. Muttlak (2003) proposed QRSS for estimating the population mean and to reducing the errors in ranking comparing to RSS.

2. SAMPLING METHODS

2.1 Quartile ranked set samples

In QRSS method, select m units from the population and rank the units within each sample with respect to a variable of interest. If the sample size is even, select for measurement from the first $m/2$ samples the $(\mathrm{q}_1(m+1))\text{th}$ smallest rank and from the second $m/2$ samples the $(\mathrm{q}_\mathrm{u}(m+1))\text{th}$ smallest rank. If the sample size is odd, select from the first $(m-1)/2$ samples the $(\mathrm{q}_1(m+1))\text{th}$ smallest rank and from the other $(m-1)/2$ samples the $(\mathrm{q}_\mathrm{u}(m+1))\text{th}$ smallest rank, and from one sample the median for that sample for actual measurement.

2.2 Double ranked set sampling

DRSS can be described as follows:

1) Identify m^3 elements from the target population and divide these elements randomly into *m* sets each of size m^2 elements.
2) Use the usual RSS procedure on each set to obtain *m* ranked set samples of size *m* each.
3) Apply the RSS procedure again on step (2) to obtain a DRSS of size *m*.

In this article, we consider double quartile ranked set samples (DQRSS) as a modification of RSS for estimating the population mean. The performance of DQRSS with respect to SRS, RSS and QRSS for estimating the population mean, is considered. The results indicates that the use of DQRSS for estimating the population mean is more efficient than SRS, RSS and QRSS for all distributions considered in this study. For asymmetric distributions, the DQRSS estimator has smaller bias with variance smaller than that of the SRS estimator.

2.3 Double quartile ranked set samples

The DQRSS procedure can be described as follows:

Step 1: Select m^3 units from the population and divide them into m^2 samples each of size m .

Step 2: If the sample size is even, select from the first $m^2/2$ samples the $(\mathrm{q}_1(m+1))$th smallest rank, and from the second $m^2/2$ samples the $(\mathrm{q}_3(m+1))$th smallest rank. If the sample size is odd, select from the first $m(m-1)/2$ samples the $(\mathrm{q}_1(m+1))$th smallest rank, the median from the next m samples and the $(\mathrm{q}_3(m+1))$th smallest rank from the second $m(m-1)/2$ samples. This step yield m sets each of size m.

Step 3: Apply the QRSS procedure on the m sets obtained in step 2, to get a DQRSS sample of size m.

Step 4: The whole cycle may be repeated n times to obtain a sample of size mn from DQRSS.

Note that we will take the nearest integer of $(\mathrm{q}_1(m+1))$th and $(\mathrm{q}_3(m+1))$th, where $\mathrm{q}_1 = 0.25$ and $\mathrm{q}_3 = 0.75$.

3. ESTIMATING OF THE POPULATION MEAN

Let $X_{11}, X_{12}, ..., X_{1m}$; $X_{21}, X_{22}, ..., X_{2m}$;…; $X_{m1}, X_{m2}, ..., X_{mm}$; be m independent random samples of size m and assume that each variable X_{ij} has the same distribution function $F(x)$ with mean μ and variance σ^2. Let $X_{i(1)}, X_{i(2)}, ..., X_{i(m)}$ $(i=1,2,...,m)$ be the ordered statistics of the ith sample $X_{i1}, X_{i2}, ..., X_{im}$ $(i=1,2,...,m)$. Let $Y_1, Y_2, ..., Y_m$ be RSS, then $Y_i \overset{d}{=} X_{(i)}$. The estimator of the population mean μ using RSS is defined by $\bar{Y}_{RSS} = \frac{1}{m}\sum_{i=1}^{m} Y_i$, with variance given by

$$\mathrm{Var}(\bar{Y}_{RSS}) = \frac{\sigma^2}{m} - \frac{1}{m^2}\sum_{i=1}^{m}(\mu_{(i)} - \mu)^2 .$$

The estimator of the population mean μ using SRS is defined by $\bar{X}_{SRS} = \frac{1}{m}\sum_{i=1}^{m} X_i$, with variance σ^2/m.

At the kth cycle $(k = 1, 2, ..., n)$, for even sample size, let $Y^*_{i(q_1(m+1))k}$ be the first quartile of the ith sample $(i = 1, 2, ..., l; l = m/2)$, and let $Y^*_{i(q_3(m+1))k}$ be the third quartile of the ith sample $(i = l+1, ..., m)$. The quantified sample $Y^*_{1(q_1(m+1))k}$, $Y^*_{2(q_1(m+1))k}, \ldots, Y^*_{\frac{m}{2}(q_1(m+1))k}$, $Y^*_{\frac{m}{2}+1(q_3(m+1))k}, \ldots, Y^*_{m(q_3(m+1))k}$, will denote the DQRSSE.

If the sample size is odd, let $Y^*_{i(q_1(m+1))}$ be the first quartile of the ith sample $(i = 1, 2, ..., h)$, where $h = (m-1)/2$, $Y^*_{i((m+1)/2)k}$ is the median of the ith sample ($i = (m+1)/2$), and $Y^*_{i(q_3(m+1))k}$ the third quartile of the ith sample $(i = h+2, ..., m)$. The quantified sample $Y^*_{1(q_1(m+1))k}$, $Y^*_{2(q_1(m+1))k}, \ldots, Y^*_{\frac{m-1}{2}(q_1(m+1))k}$, $Y^*_{\frac{m-1}{2}+1(\frac{m+1}{2})k}$, $Y^*_{\frac{m-1}{2}+2(q_3(m+1))k}, \ldots, Y^*_{m(q_3(m+1))k}$ will denote the DQRSSO.

The estimator of the population mean using DQRSS can be defined as

$$\bar{Y}^*_{DQRSS} = \begin{cases} \bar{Y}^*_{DQRSSE} = \dfrac{1}{mn}\displaystyle\sum_{k=1}^{n}\left(\sum_{i=1}^{l} Y^*_{i(q_1(m+1))k} + \sum_{i=l+1}^{m} Y^*_{i(q_3(m+1))k}\right), l = m/2 \\ \bar{Y}^*_{DQRSSO} = \dfrac{1}{mn}\displaystyle\sum_{k=1}^{n}\left(\sum_{i=1}^{h} Y^*_{i(q_1(m+1))k} + Y^*_{(h+1)\left(\frac{m+1}{2}\right)k} + \sum_{i=h+2}^{m} Y^*_{i(q_3(m+1))k}\right), h = (m\text{-}1)/2 \end{cases}$$

The variance of $\bar{Y}^*_{DQRSS}$ for even and odd sample size can be given respectively by

$$\sigma^{*2}_{DQRSSE} = \frac{1}{nm^2}\sum_{k=1}^{n}\left(\sum_{i=1}^{l}\sigma^{*2}_{i(q_1(m+1))k} + \sum_{i=l+1}^{m}\sigma^{*2}_{i(q_3(m+1))k}\right), \ l = m/2.$$

$$\sigma^{*2}_{DQRSSO} = \frac{1}{nm^2}\sum_{k=1}^{n}\left(\sum_{i=1}^{h}\sigma^{*2}_{i(q_1(m+1))k} + \sigma^{*2}_{(h+1)\left(\frac{m+1}{2}\right)k} + \sum_{i=h+2}^{m}\sigma^{*2}_{i(q_3(m+1))k}\right),$$

$$h = (m\text{-}1)/2$$

Assume that Y^*_i has the mean μ^*_i and the variance $\sigma^{*2}_{(i)}$, Al-Saleh and Al-Kadiri (2000) showed that

$$\mu = \sum_{i=1}^{m} \mu^{*}_{(i)} \ , \ \sigma^2 = \frac{1}{m}\left[\sum_{i=1}^{m} \sigma^{*2}_{(i)} + \sum_{i=1}^{m} (\mu^{*}_{(i)} - \mu)^2 \right]$$

where μ and σ^2 are the mean and the variance of the population respectively.

Lemma 1:

Let X be a random variable of pdf $f(x)$ and cdf $F(x)$. Its mean and variance are μ and σ^2 respectively. A random sample of size m was selected and ranked, let $X_{r:m}$ be the rth smallest value of the sample, where $r = 1,...,m$. The pdf and cdf for $X_{r:m}$ are

$$f_{r:m}(x) = \frac{1}{B(r, m-r+1)} F^{r-1}(x)(1-F(x))^{m-r} f(x),$$

$$F_{r:m}(x) = FB(F(x); r, m-r+1),$$

respectively, where $FB(F(x); r, m-r+1)$ is a beta distribution function with parameters $(r, m-r+1)$. Let denote the mean and the variance of $X_{r:m}$ as $\mu_{r:m}$ and $\sigma^2_{r:m}$ respectively. Then

a. $\mu_{r:m} = F^{-1}[\alpha(r)]$

b. $\mu_{m-r+1:m} = F^{-1}[1-\alpha(r)]$

c. $\sigma^2_{r:m} + (\mu_{r:m} - \mu)^2 < \sigma^2$

where $\alpha(r) = QB(p_r; r, m-r+1))$ which is a quartile function for beta distribution and $p_r = r/(m+1)$.

If $f(x)$ is symmetry then

d. $\mu_{r:m} + \mu_{m-r+1:m} = 2\mu$

e. $\sigma^2_{r:m} = \sigma^2_{m-r+1:m}$

Proof:

The variance of $X_{r:m}$ is given by

$$\sigma^2_{r:m} = \int (x-\mu_{r:m})^2 f_{r:m}(x)dx = \int (x-\mu)^2 f_{r:m}(x)dx - (\mu_{r:m} - \mu)^2$$

Substituting $f_{r:m}(x)$ and rearranging the above equation produces

$$\sigma^2_{r:m} + (\mu_{r:m} - \mu)^2 = \int (x-\mu)^2 \left(\frac{1}{B(r, m-r+1)} F^{r-1}(x)(1-F(x))^{m-r} \right) f(x)dx$$

As $\dfrac{F^{r-1}(x)[1-F(x)]^{m-r}}{B(r, m-r+1)} < 1$, so

$$\sigma^2_{r:m} + (\mu_{r:m} - \mu)^2 < \sigma^2 = \int (x-\mu)^2 f(x)dx$$

Using Taylor series, as given in David & Nagarajah (2003), can be shown that

$$E(X_{r:m}) = \mu_{r:m} = \int x f_{r:m}(x)dx \simeq F^{-1}_{r:m}(p_r)$$

Let $F_{r:m}(x) = FB(F(x); r, m-r+1) = p_r$

Utilizing this relationship produces

$$\mu_{r:m} = F^{-1}_{r:m}(p_r) = F^{-1}[\alpha(r)] \text{ where } \alpha(r) = QB(p_r; r, m-r+1))$$

Let $F_{m-r+1:m}(x) = FB(F(x); m-r+1, r) = q_r, \quad q_r + p_r = 1$

$$\mu_{m-r+1:m} = F^{-1}_{m-r+1:m}(q_r) = F^{-1}[QB(q_r; m-r+1, r)]$$

Since $QB(1-p_r; m-r+1, r) = 1 - QB(p_r; r, m-r+1)$, then

$$\mu_{m-r+1:m} = F^{-1}[1-\alpha(r)]$$

If $f(x)$ is symmetry for any $0 \le \alpha(r) \le 1$

$$F^{-1}[1-\alpha(r)] - \mu = \mu - F^{-1}[\alpha(r)]$$

So $F^{-1}[1-\alpha(r)] + F^{-1}[\alpha(r)] = \mu_{r:m} + \mu_{m-r+1:m} = 2\mu$

The variance of $X_{r:m}$ is given by

$$\sigma^2_{r:m} = \int \frac{\left(F^{-1}(u) - F^{-1}[\alpha(r)]\right)^2}{B(r, m-r+1)} u^{r-1}(1-u)^{m-r} du$$

For symmetrical $f(x)$,

$$\sigma^2_{r:m} = \int \frac{\left(F^{-1}(1-u) - F^{-1}[1-\alpha(r)]\right)^2}{B(r, m-r+1)} u^{r-1}(1-u)^{m-r}\, du$$

$$= \int \frac{\left(F^{-1}(u) - F^{-1}[1-\alpha(r)]\right)^2}{B(m-r+1, r)} u^{m-r}(1-u)^{r-1}\, du = \sigma^2_{m-r+1:m}$$

Lemma 2:

Let $Y_{r:m}$ be the rth smallest value of a random sample of size m. The sample was selected from a population of pdf

$$f_{r:m}(x) = \frac{1}{B(r, m-r+1)} F^{r-1}(x)(1-F(x))^{m-r} f(x)$$

where the mean and variance correspond to pdf $f(x)$ are μ and σ^2 respectively. In addition, let $Y_{m-r+1:m}$ be the $(m-r+1)$th smallest value, $E(Y_{r:m}) = \mu^*_{r:m}$, $E(Y_{m-r+1:m}) = \mu^*_{m-r+1:m}$, $\text{Var}(Y_{r:m}) = \sigma^{*2}_{r:m}$, $\text{Var}(Y_{m-r+1:m}) = \sigma^{*2}_{m-r+1:m}$. Then,

a. $\mu^*_{r:m} = F^{-1}[\alpha \circ \alpha(r)]$

b. $\mu^*_{m-r+1:m} = F^{-1}[1 - \alpha \circ \alpha(r)]$

c. $\sigma^{*2}_{r:m} + (\mu^*_{r:m} - \mu_{r:m})^2 + (\mu_{r:m} - \mu)^2 < \sigma^2$.

If $f(x)$ is symmetry then

d. $\mu^*_{r:m} + \mu^*_{m-r+1:m} = 2\mu$

e. $\sigma^{*2}_{r:m} = \sigma^{*2}_{m-r+1:m}$

Proof:

Using the results of lemma 1

$$\mu^*_{r:m} = F^{-1}_{r:m}[\alpha(r)] = F^{-1}[\alpha \circ \alpha(r)]$$
$$\mu^*_{m-r+1:m} = F^{-1}_{r:m}[1-\alpha(r)] = F^{-1}[1 - \alpha \circ \alpha(r)]$$
$$\sigma^{*2}_{r:m} + (\mu^*_{r:m} - \mu_{r:m})^2 < \sigma^2_{r:m}$$

Since $\sigma^2_{r:m} + (\mu_{r:m} - \mu)^2 < \sigma^2$ so

$$\sigma^{*2}_{r:m} + (\mu^*_{r:m} - \mu_{r:m})^2 + (\mu_{r:m} - \mu)^2 < \sigma^2$$

For any symmetry distribution, and $\alpha \in [0,1]$

$$(\mu - F^{-1}(\alpha)) = (F^{-1}(1-\alpha) - \mu)$$

so, $\mu^*_{r:m} + \mu^*_{m-r+1:m} = 2\mu$

The variance of $Y_{r:m}$ is equal to

$$\sigma^{*2}_{r:m} = \int \frac{\left(F^{-1}_{r:m}(u) - F^{-1}_{r:m}[\alpha(r)]\right)^2}{B(r, m-r+1)} u^{r-1}(1-u)^{m-r}\, du$$

$$= \int \frac{\left(F^{-1}(u) - F^{-1}[\alpha \circ \alpha(r)]\right)^2}{B(r, m-r+1)} u^{r-1}(1-u)^{m-r}\, du$$

$$= \int \frac{\left(F^{-1}(u) - F^{-1}[1-\alpha \circ \alpha(r)]\right)^2}{B(m-r+1, r)} u^{m-r}(1-u)^{r-1}\, du = \sigma^{*2}_{m-r+1:m}$$

Lemma 3:

1. $\hat{\mu}^*_{DQRSS}$ is an unbiased estimator of the population mean, under the assumption that the population is symmetric about its mean**.**
2. $\text{Var}\left(\bar{Y}^*_{DQRSS}\right)$ is less than each of $\text{Var}(\bar{X}_{SRS})$, $\text{Var}(\bar{X}_{RSS})$ and $\text{Var}(\bar{Y}_{QRSS})$.
3. The mean square error of DQRSS estimator is less than the variance of the SRS estimator for asymmetric distributions i.e., $\text{MSE}\left(\bar{Y}^*_{DPRSS}\right) < \text{Var}(\bar{X}_{SRS})$.

Proof:

For *m* even

$$\hat{\mu}_{DQRSSE} = \frac{1}{m}\left(\sum_{i=1}^{l} Y_{i(r:m)} + \sum_{i=l+1}^{m} Y_{i(m-r+1:m)}\right)$$

$$E\left(\hat{\mu}_{DQRSSE}\right) = \frac{1}{m}\left(\sum_{i=1}^{\frac{m}{2}} E\left(Y_{i(r:m)}\right) + \sum_{i=\frac{m}{2}+1}^{m} E\left(Y_{i(m-r+1:m)}\right)\right)$$

$$= \frac{1}{m}\left(\frac{m}{2}\mu^*_{r:m} + \frac{m}{2}\mu^*_{m-r+1:m}\right) = \mu$$

and

$$\operatorname{Var}\left(\hat{\mu}_{DQRSSE}\right)=\frac{1}{m^2}\left(\sum_{i=1}^{\frac{m}{2}}\operatorname{Var}\left(Y_{i(r:m)}\right)+\sum_{i=\frac{m}{2}+1}^{m}\operatorname{Var}\left(Y_{i(m-r+1:m)}\right)\right)$$

$$=\frac{1}{2m}\left(\sigma^{*2}_{r:m}+\sigma^{*2}_{m-r+1:m}\right)<\frac{\sigma^2}{m}$$

For m odd

$$\hat{\mu}_{DQRSSO}=\frac{1}{m}\left(\sum_{i=1}^{\frac{m-1}{2}}Y_{i(r:m)}+\sum_{i=\frac{m+3}{2}}^{m}Y_{i(m-r+1:m)}+Y_{\frac{m+1}{2}:m}\right)$$

$$E(\hat{\mu}_{DQRSSO})=\frac{1}{m}\left(\sum_{i=1}^{\frac{m-1}{2}}E\left(Y_{i(r:m)}\right)+\sum_{i=\frac{m+3}{2}}^{m}E\left(Y_{i(m-r+1:m)}\right)+E\left(Y_{\frac{m+1}{2}:m}\right)\right)$$

$$=\frac{1}{m}\left(\frac{m-1}{2}\left(\mu^{*}_{r:m}+\mu^{*}_{m-r+1:m}\right)+\mu\right)=\mu$$

and

$$\operatorname{Var}\left(\hat{\mu}_{DQRSSO}\right)=\frac{1}{m^2}\left(\sum_{i=1}^{\frac{m-1}{2}}\operatorname{Var}\left(Y_{i(r:m)}\right)+\sum_{i=\frac{m+3}{2}}^{n}\operatorname{Var}\left(Y_{i(m-r+1:m)}\right)+\operatorname{Var}\left(Y_{\frac{m+1}{2}:m}\right)\right)$$

$$=\frac{1}{m^2}\left(\frac{m-1}{2}\left(\sigma^{*2}_{r:m}+\sigma^{*2}_{m-r+1:m}\right)+\sigma^{*2}_{\frac{m+1}{2}:m}\right)<\frac{\sigma^2}{m}$$

4. EFFICIENCY OF DQRSS

To compare the considered estimators for the population mean using DQRSS with respect to the SRS, RSS, and QRSS procedures. Three symmetric distributions, namely, uniform, normal and logistic and three asymmetric distributions, namely, exponential, gamma and weibull are considered. The relative efficiency of the unbiased estimators using ranked set samples procedures for estimating the population mean with respect to SRS is defined as $eff\left(\bar{X}_{SRS},\bar{Y}_{RSS}\right)=\frac{\operatorname{Var}(\bar{X}_{SRS})}{\operatorname{Var}\left(\bar{Y}_{RSS}\right)}$, and for biased estimators the relative efficiency is defined as $eff\left(\bar{X}_{SRS},\bar{Y}_{RSS}\right)=\frac{\operatorname{Var}(\bar{X}_{SRS})}{\operatorname{MSE}\left(\bar{Y}_{RSS}\right)}$.

Assume the cycle is repeated once, Tables 1 and 2 summarize the relative efficiency of the RSS, QRSS and DQRSS estimators with sample sizes $m=6,7,10,11$ and 12, for each simulation, 60,000 iterations were performed.

Table 1
The relative efficiency for estimating the population mean using RSS, QRSS, and DQRSS with respect to SRS with sample size $m = 6$ and 7.

Distribution		$m = 6$			$m = 7$		
		RSS	***QRSS***	***DQRSS***	***RSS***	***QRSS***	***DQRSS***
Uniform (0,1)	*eff*	3.500	3.214	16.966	4.000	3.809	23.445
	Bias						
Uniform (0,2)	*eff*	3.500	3.232	17.267	4.000	3.770	23.021
	Bias						
Normal (0,1)	*eff*	3.191	3.639	11.906	3.658	4.065	14.669
	Bias						
Normal (1,2)	*eff*	3.210	3.645	11.950	3.631	4.051	14.590
	Bias						
Logistic (-1,1)	*eff*	2.868	3.729	11.707	3.259	4.144	13.845
	Bias						
Exponential (1)	*eff*	2.430	3.009	9.549	2.746	3.321	8.692
	Bias		0.092	0.016		0.075	0.059
Exponential (2)	*eff*	2.407	3.016	9.569	2.735	3.327	8.598
	Bias		0.046	0.008		0.038	0.029
Exponential (3)	*eff*	2.467	3.051	9.764	2.693	3.293	8.513
	Bias		0.031	0.005		0.025	0.020
Gamma (1,2)	*eff*	2.391	3.022	9.395	2.715	3.333	8.575
	Bias		0.183	0.033		0.150	0.119
Gamma (1,3)	*eff*	2.416	3.025	9.572	2.669	3.282	8.496
	Bias		0.279	0.047		0.230	0.178
Weibull (1,3)	*eff*	2.459	3.029	9.660	2.755	3.334	8.503
	Bias		0.274	0.047		0.227	0.178

Table 2
The relative efficiency for estimating the population mean using RSS, QRSS and DQRSS with respect to SRS with sample size $m = 10, 11$ and 12

Distribution		m = 10			m = 11			m = 12		
		RSS	*QRSS*	*DQRSS*	*RSS*	*QRSS*	*DQRSS*	*RSS*	*QRSS*	*DQRSS*
Uniform	*eff*	**5.500**	**5.085**	**38.097**	**6.000**	**5.637**	**47.852**	**6.500**	**6.730**	**66.637**
(0,1)	*Bias*									
Uniform	*eff*	**5.500**	**5.128**	**38.463**	**6.000**	**5.680**	**47.627**	**6.500**	**6.667**	**66.234**
(0,2)	*Bias*									
Normal	*eff*	**4.827**	**5.736**	**31.288**	**5.197**	**6.067**	**35.034**	**5.673**	**6.338**	**37.261**
(0,1)	*Bias*									
Normal	*eff*	**4.844**	**5.850**	**31.721**	**5.195**	**6.240**	**35.046**	**5.652**	**6.412**	**36.958**
(1,2)	*Bias*									
Logistic	*eff*	**4.198**	**6.270**	**32.220**	**4.533**	**6.755**	**34.801**	**4.911**	**6.728**	**34.315**
(-1,1)	*Bias*									
Exponential	*eff*	**3.440**	**3.281**	**15.024**	**3.671**	**3.542**	**28.555**	**3.922**	**4.693**	**8.303**
(1)	*Bias*		**0.117**	**0.056**		**0.105**	**0.001**		**0.061**	**0.083**
Exponential	*eff*	**3.426**	**3.288**	**14.916**	**3.659**	**3.521**	**28.406**	**3.962**	**4.735**	**8.409**
(2)	*Bias*		**0.059**	**0.028**		**0.053**	**0.000**		**0.031**	**0.042**
Exponential	*eff*	**3.394**	**3.252**	**14.844**	**3.653**	**3.535**	**28.775**	**3.964**	**4.773**	**8.452**
(3)	*Bias*		**0.039**	**0.019**		**0.035**	**0.000**		**0.020**	**0.028**
Gamma	*eff*	**3.440**	**3.276**	**14.878**	**3.723**	**3.594**	**28.877**	**3.919**	**4.697**	**8.372**
(1,2)	*Bias*		**0.234**	**0.113**		**0.210**	**0.001**		**0.123**	**0.167**
Gamma	*eff*	**3.460**	**3.274**	**14.963**	**3.638**	**3.539**	**28.510**	**3.990**	**4.711**	**8.350**
(1,3)	*Bias*		**0.354**	**0.170**		**0.314**	**0.002**		**0.184**	**0.250**
Weibull	*eff*	**3.471**	**3.245**	**14.808**	**3.699**	**3.576**	**28.675**	**3.960**	**4.751**	**8.480**
(1,3)	*Bias*		**0.352**	**0.170**		**0.313**	**0.002**		**0.185**	**0.249**

From simulation results, we conclude the following:

1. A gain in efficiency is attained using DQRSS for estimating the population mean for all cases that considered in this study. As an example for normal (0,1), with $m = 11$, the relative efficiency of the DQRSS 53.034 for estimating the population mean comparing this value with its counterpart 5.197, 6.067 using RSS and QRSS respectively.
2. If the underlying distribution is asymmetric, again in efficiency is attained using DQRSS, regardless of a smaller bias. As an example, for $m = 11$ the relative efficiency of the DQRSS 28.877 with bias 0.001 for estimating the population mean of a gamma distribution with parameters 1 and 2, while for $m = 11$ the relative efficiency using RSS is 3.723 and by using QRSS its 3.594 with bias 0.210.

5. DOUBLE QUARTILE RANKED SET SAMPLING WITH ERRORS IN RANKING

Dell and Clutter (1972) showed that the sample mean using RSS is unbiased estimator of the population mean regardless of whatever the ranking is perfect or not, and has a smaller variance than its counterpart SRS with the same sample size.

Muttlak (2003) showed that QRSS with errors in ranking is unbiased estimator of the population mean when the underlying distribution is assumed to be symmetric about its mean.

Let $Y^*_{i[q_1(m+1)]}$ and $Y^*_{i[q_3(m+1)]}$ be the first and third judgment quartile of the ith sample $(i=1,2,\ldots,m)$ respectively with errors in ranking. The estimator of the population mean with error in ranking using DQRSS can be defined as

$$\hat{Y}^*_{DQRSS_e} = \begin{cases} \hat{Y}^*_{DQRSSE_e} = \dfrac{1}{mn}\sum_{k=1}^{n}\left(\sum_{i=1}^{l} Y^*_{i[\mathrm{q}_1(m+1)]k} + \sum_{i=l+1}^{m} Y^*_{i[\mathrm{q}_3(m+1)]k}\right), l = m/2 \\ \hat{Y}^*_{DQRSSO_e} = \dfrac{1}{mn}\sum_{k=1}^{n}\left(\sum_{i=1}^{h} Y^*_{i[\mathrm{q}_1(m+1)]k} + Y^*_{(h+1)\left[\left(\frac{m+1}{2}\right)\right]k} + \sum_{i=h+2}^{m} Y^*_{i[\mathrm{q}_3(m+1)]k}\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad h=(m\text{-}1)/2 \end{cases}$$

The estimator of the population mean μ with errors in ranking has the following properties:

1. $\hat{Y}^*_{DQRSS_e}$ is unbiased estimator of the population mean if the population is symmetric about its mean.
2. $\mathrm{Var}\left(\hat{Y}^*_{DQRSS_e}\right)$ is less than $\mathrm{Var}\left(\bar{X}_{SRS}\right)$.
3. For asymmetric distribution about its mean, $\mathrm{MSE}\left(\hat{Y}^*_{DQRSS_e}\right) < \mathrm{Var}\left(\bar{X}_{SRS}\right)$

The above properties can be proved based on Takahasi and Wakimoto (1968), Dell and Clutter (1972), Muttlak (2003) and AL-Saleh and AL-Kadiri (2000).

In this article, it is observed that the DQRSS estimator is unbiased of the population mean if the underlying distribution is symmetric, and more efficient than the SRS, RSS and QRSS. The authors suggest using the DQRSS for estimating the population mean of symmetric distribution and asymmetric distribution when the biased is small; also, we can use DQRSS to reduce the errors in ranking than RSS.

6. ACKNOWLEDGMENTS

We are grateful to a referee and the Editor, Pakistan Journal of Statistics who provided very constructive comments.

REFERENCES

1. Al-Saleh, M.F. and Al-Kadiri, M. (2000). Double ranked set sampling. *Statist. Probab. Lett.*, 48, 205 -212.
2. Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. *J. Statist. Plann. Infer.*, 102, 273-286.
3. Dell, T.R., and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrika* 28, 545-555.
4. David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, Third Edition, John Wiley & sons, Inc., Hoboken, New Jersey.
5. McIntyre, G.A. (1952). A method for unbiased selective sampling using, ranked sets. *Austr. J. Agri. Res.*, 3, 385-390.
6. Muttlak, H.A. (2003). Investigating the use of quantile ranked set samples for estimating the population mean. *J. Appl. Math. Comp.*, 146, 437- 443.
7. Muttlak, H.A. (1997). Median ranked set sampling, *J. Appl. Statist. Scien.*, 6(4) 577-586.
8. Samawi, H, Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling, *Bio. Journal,* 30, 577-586.
9. Takahasi, K. and Wakimoto, K. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.*, 20, 1-31.

CHAPTER FIFTEEN

Multistage Quartile Ranked Set Samples

Abdul Aziz Jemain[1] and **Amer Ibrahim Al-Omari**[2]
School of Mathematical Sciences, Faculty of Science and Technology
University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
Email [1]azizj@pkrisc.cc.ukm.my [2]alomari_amer@yahoo.com

ABSTRACT

Multistage quartile ranked set sampling (MQRSS) method is suggested for estimating the population mean. The MQRSS is compared with the simple random sampling (SRS), ranked set sampling (RSS) suggested by McIntyre (1952) and quartile ranked set sampling (QRSS) suggested by Muttlak (2003) based on the same sample size. We show that MQRSS estimator is an unbiased of the population mean and more efficient than SRS, QRSS and RSS ($r > 1$, r is the number of stage) when the underlying distribution is symmetric about its mean. Also, by MQRSS we can increase the efficiency of mean estimator for specific value of the sample size. For asymmetric distributions considered in this study, MQRSS estimator has a smaller bias. A collection of a real data is used to illustrate the method.

KEYWORDS

Simple random sampling; ranked set sampling; quartile ranked set sampling; multistage ranked set sampling; symmetric distribution; asymmetric distribution.

1. INTRODUCTION

The RSS was suggested by McIntyre (1952) for estimating mean pasture yields with greater efficiency than SRS. In situations where the experimental or sampling units in a study can be more easily ranked than quantified, McIntyre proposed that the mean of m sample units based on a RSS as an estimator of the population mean. Takahasi and Wakimoto (1968) independently introduced the same method. Dell and Clutter (1972) showed that the mean of the RSS is an unbiased of the population mean, whatever or not there are errors in ranking. Samawi et al. (1996) investigated variety of extreme ranked set samples (ERSS) for estimating a population means. Muttlak (1997) suggested using median ranked set sampling (MRSS) to estimate the population mean. Al-Saleh and Al-Kadiri (2000) introduced double ranked set sampling for estimating the population mean, they showed that the ranking in the second stage is easier than the ranking in the first stage. Al-Saleh and Al-Omari (2002) suggested multistage ranked set sampling (MSRSS) that increase the efficiency of estimating the population mean for specific value of the sample size. Muttlak (2003) suggested QRSS for estimating the population mean and showed using QRSS procedure will reduce the errors in ranking comparing to RSS since we only select and measure the first or the third quartile of the sample. Jemain and Al-Omari (2006) suggested double quartile ranked set sampling (DQRSS) for estimating the population mean and showed that the DQRSS mean is an unbiased estimator and more efficient than the SRS, RSS and the QRSS if the underlying distribution is symmetric.

In this paper, MQRSS is considered. The properties of MQRSS for estimating the population mean are discussed. Also, MQRSS is compared with SRS, RSS, QRSS and DQRSS methods. The method is illustrated by using real data set. However, by MQRSS we can increase the efficiency of the mean estimator for specific value of the sample size *m* by increase the number of stages. Also MQRSS we can use larger sample as compared to the usual RSS, since all we have to do is to find the first or the third quartile of the *i*th sample and measure it.

2. SAMPLING METHODS

2.1 Ranked set sampling

To obtain a sample of size *m* by the usual RSS as suggested by McIntyre (1952), select *m* random samples each of size *m* from the target population and rank the units within each sample with respect to a variable of interest. The *i*th smallest of the *i*th sample $(i = 1, 2, ..., m)$ is drawn and measured. The method is repeated *n* times if needed to increase sample size.

2.2 Quartile ranked set sampling

The QRSS procedure suggested by Muttlak (2003) involves selecting *m* random samples each of size *m* units from the target population and ranks the units within each sample with respect to the variable of interest. If the sample size *m* is even, then select and measure from the first $m/2$ samples the $(q_1(m+1)$th smallest rank unit and from the second $m/2$ samples the $(q_3(m+1)$th smallest rank unit. Note that we will take the nearest integer of $(q_1(m+1))$th and $(q_3(m+1))$th where $q_1 = 0.25$, and $q_3 = 0.75$. If the sample size *m* is odd, select and measure from the first $(m-1)/2$ samples the $(q_1(m+1)$th smallest rank unit and from the other $(m-1)/2$ samples the $(q_3(m+1)$th smallest rank unit, and from one sample the median for that sample. The cycle can be repeated *n* times if needed to get a sample of size *nm* units.

2.3 Multistage quartile ranked set samples

The MQRSS procedure is described as in the following steps:

Step 1: Randomly selected m^{r+1} sample units from the target population, where r is the number of stages and m is the sample size.

Step 2: Allocate the m^{r+1} selected units as randomly as possible into m^r sets, each of size m.

Step 3: For each set in Step (2), if the sample size is even, select from the first $m^r/2$ samples, the $(q_1(m+1))$th smallest rank unit and from the second $m^r/2$ samples the $(q_3(m+1))$th smallest rank unit, where $q_1 = 0.25$ and $q_3 = 0.75$. The obtained sample will denoted by MQRSSE.

If the sample size is odd, select from the first $\left(m^r - m^{r-1}\right)/2$ samples the $(q_1(m+1))$th smallest rank unit and from the next m^{r-1} samples the median of each sample and from the other $\left(m^r - m^{r-1}\right)/2$ samples the $(q_3(m+1))$th smallest rank unit. Such sample will be denoted by MQRSSO.

Step 4: Repeat Step (3) on the m^{r-1} quartile ranked sets to obtain m^{r-2} second stage quartile ranked sets each of size m.

Step 5: The process continues until we end up with one rth stage of quartile ranked set samples of size m.

The whole process can be repeated n times if needed to get a sample of size nm from MQRSS data. Note that we always take the nearest integer of $(q_1(m+1))$th and $(q_3(m+1))$th. It is of interest to note that if $r = 1$ and $m \leq 3$ the MQRSS will be reduced to the usual RSS method. It is very important to emphasize here that the ranking at all stages are done by visually inspection or by any other cheap method and the actual quantification is only done on the last sample of size m that is obtained at the last stage.

To clarify this procedure, consider the cases in the following example:

Example:

Let $m=4$ and $r=2$, so that we may have a random sample of size 16, allocate them into 16 subsets each of size 4 units. Let $X_{i(j:m)}^{(r)}$ be the *j*th minimum $(j=1,2,3,4)$ of the *i*th set $(i=1,2,...,16)$ at stage *r*. After ranking the units within each subset appear as shown below:

$$A_1^{(0)}=\left\{X_{1(1:4)}^{(0)},X_{1(2:4)}^{(0)},X_{1(3:4)}^{(0)},X_{1(4:4)}^{(0)}\right\},\dots,\ A_{16}^{(0)}=\left\{X_{16(1:4)}^{(0)},X_{16(2:4)}^{(0)},X_{16(3:4)}^{(0)},X_{16(4:4)}^{(0)}\right\}.$$

Now, to apply the MQRSSE procedure on each of the 16 sets, the first quartile is the smallest rank and the third quartile is the largest rank. Thus for $r=1$, we will select the first quartile from the first 8 sets and the third quartile from the other 8 sets as:

$$X_{1(1:4)}^{(1)}=\min\left(A_1^{(0)}\right),\ X_{2(1:4)}^{(1)}=\min\left(A_2^{(0)}\right),$$

$$X_{3(1:4)}^{(1)}=\min\left(A_3^{(0)}\right),\ X_{4(1:4)}^{(1)}=\min\left(A_4^{(0)}\right),$$

$$X_{5(1:4)}^{(1)}=\min\left(A_5^{(0)}\right),\ X_{6(1:4)}^{(1)}=\min\left(A_6^{(0)}\right),$$

$$X_{7(1:4)}^{(1)}=\min\left(A_7^{(0)}\right),\ X_{8(1:4)}^{(1)}=\min\left(A_8^{(0)}\right),$$

and

$$X_{9(4:4)}^{(1)}=\max\left(A_9^{(0)}\right),X_{10(4:4)}^{(1)}=\max\left(A_{10}^{(0)}\right),$$

$$X_{11(4:4)}^{(1)}=\max\left(A_{11}^{(0)}\right),X_{12(4:4)}^{(1)}=\max\left(A_{12}^{(0)}\right),$$

$$X_{13(4:4)}^{(1)}=\max\left(A_{13}^{(0)}\right),X_{14(4:4)}^{(1)}=\max\left(A_{14}^{(0)}\right),$$

$$X_{15(4:4)}^{(1)}=\max\left(A_{15}^{(0)}\right),X_{16(4:4)}^{(1)}=\max\left(A_{16}^{(0)}\right).$$

This step yields 4 sets each of size 4 at the first stage. The obtained sets are:

$$A_1^{(1)}=\left\{X_{1(1:4)}^{(1)},X_{2(1:4)}^{(1)},X_{3(1:4)}^{(1)},X_{4(1:4)}^{(1)}\right\},$$

$$A_2^{(1)}=\left\{X_{5(1:4)}^{(1)},X_{6(1:4)}^{(1)},X_{7(1:4)}^{(1)},X_{8(1:4)}^{(1)}\right\},$$

$$A_3^{(1)}=\left\{X_{9(4:4)}^{(1)},X_{10(4:4)}^{(1)},X_{11(4:4)}^{(1)},X_{12(4:4)}^{(1)}\right\},$$

$$A_4^{(1)}=\left\{X_{13(4:4)}^{(1)},X_{14(4:4)}^{(1)},X_{15(4:4)}^{(1)},X_{16(4:4)}^{(1)}\right\}.$$

For $r=2$, reapply the MQRSSE method on these 4 sets, so we will select the smallest rank from the first 2 sets largest rank from the other 2 sets as:

$$X_{1(1:4)}^{(2)} = \min\left(A_1^{(1)}\right), X_{2(1:4)}^{(2)} = \min\left(A_2^{(1)}\right),$$
$$X_{3(4:4)}^{(2)} = \max\left(A_3^{(1)}\right), X_{4(4:4)}^{(2)} = \max\left(A_4^{(1)}\right).$$

The final set $\left\{X_{1(1:4)}^{(2)}, X_{2(1:4)}^{(2)}, X_{3(4:4)}^{(2)}, X_{4(4:4)}^{(2)}\right\}$ is a second stage quartile ranked set sample. It is of interest to note that $X_{1(1:4)}^{(2)}, X_{2(1:4)}^{(2)}$ are iid, also $X_{3(4:4)}^{(2)}, X_{4(4:4)}^{(2)}$ are iid. These 4 units exactly are measured for estimating the mean of the variable of interest as:

$$\bar{X}_{MQRSSE}^{(2)} = \frac{X_{1(1:4)}^{(2)} + X_{2(1:4)}^{(2)} + X_{3(4:4)}^{(2)} + X_{4(4:4)}^{(2)}}{4}.$$

Thus, the number of quantified units, which is 4, is small portion to the number of sampled units, which is 64, but all sampled units add to the information content of the quantified units. Hence, it makes sense to compare the information in this sample with that of a SRS of the size 4 and not 64.

3. ESTIMATION OF THE POPULATION MEAN

Let $X_1,\ X_2, \ldots,\ X_m$ be a random sample with probability density function $f(x)$ with mean μ and variance σ^2. Let $X_{11}, X_{12}, \ldots, X_{1m}$; $X_{21}, X_{22}, \ldots, X_{2m}$; …, $X_{m1}, X_{m2},\ \ldots, X_{mm}$ be independent random variables all with the same cumulative distribution function $F(x)$. The SRS estimator of the population mean from a sample of size *m* is given by

$$\bar{X}_{SRS} = \frac{1}{m}\sum_{i=1}^{m} X_i, \qquad (3.1)$$

with variance

$$\operatorname{Var}\left(\bar{X}_{SRS}\right) = \frac{\sigma^2}{m}. \qquad (3.2)$$

The estimator of the population for a RSS of size *m* (see McIntyre (1952)) is given by

$$\bar{X}_{RSS} = \frac{1}{m}\sum_{i=1}^{m} X_{i(i:m)}, \qquad (3.3)$$

with variance

$$\operatorname{Var}\left(\bar{X}_{RSS}\right)=\frac{1}{m^2}\sum_{i=1}^{m}\operatorname{Var}\left(X_{i(i:m)}\right)=\frac{\sigma^2}{m}-\frac{1}{m^2}\sum_{i=1}^{m}\left(\mu_{(i:m)}-\mu\right)^2 . \qquad (3.4)$$

Since, $\sum_{i=1}^{m}\left(\mu_{(i:m)}-\mu\right)^2 \geq 0$, then $\bar{X}_{RSS}$ is more efficient than $\bar{X}_{SRS}$ based on the same number of measured observations.

The MSRSS estimator of the population mean from a sample of size m (see Al-Saleh and Al-Omari (2002)) is given by

$$\bar{X}_{MSRSS}^{(r)}=\frac{1}{m}\sum_{i=1}^{m}X_i^{(r)} , \qquad (3.5)$$

with variance

$$\operatorname{Var}\left(\bar{X}_{MSRSS}^{(r)}\right)=\frac{1}{m}\left(\sigma^2-\frac{1}{m}\sum_{i=1}^{m}\left(\mu_i^{(r)}-\mu\right)^2\right), \qquad (3.6)$$

where μ and σ^2 are the mean and the variance of the population, respectively. It is of interest to note here that the MSRSS method suggested by AL-Saleh and AL-Omari (2002) constitute by apply the usual RSS method on m^{r-1} sets each of size m^2 up to rth stage, which is difference from our work based on MQRSS where we apply the QRSS method on m^r sets each of size m up to rth stage.

Now to estimate the population mean using MQRSS method, at the rth stage if the sample size is even, let $X_{i(q_1(m+1):m)}^{(r)}$ be the $(q_1(m+1))$th smallest rank unit of the ith sample $\left(i=1,2,\ldots,\frac{m}{2}\right)$ and $X_{i(q_3(m+1):m)}^{(r)}$ be the $(q_3(m+1))$th smallest rank unit of the ith sample $\left(i=\frac{m+2}{2},\frac{m+4}{2},\ldots,m\right)$. Note that the units $X_{1(q_1(m+1):m)}^{(r)}$, $X_{2(q_1(m+1):m)}^{(r)}$,…, $X_{\frac{m}{2}(q_1(m+1):m)}^{(r)}$ are iid and also $X_{\frac{m+2}{2}(q_3(m+1):m)}^{(r)}$,…, $X_{m(q_3(m+1):m)}^{(r)}$, are iid. However, all units are mutually independent but not identically distributed. These measured units denote the MQRSSE.

For odd sample size, let $X_{i(q_1(m+1):m)}^{(r)}$ be the $\left(q_1\left(m+1\right)\right)th$ smallest rank unit of the ith sample $\left(i=1,2,\ldots,\frac{m-1}{2}\right)$ and $X_{i\left(\frac{m+1}{2}:m\right)}^{(r)}$ be the median of the ith sample of the rank, $i = \frac{m+1}{2}$ and $X^{(r)}_{i(q_3(m+1):m)}$ be the $(q_3(m+1))$th smallest rank unit of the ith sample $\left(i = \frac{m+3}{2}, \frac{m+5}{2}, \dots, m\right)$. Note that $X^{(r)}_{1(q_1(m+1):m)}$, $X^{(r)}_{2(q_1(m+1):m)}, \dots,$ $X^{(r)}_{\frac{m-1}{2}(q_1(m+1):m)}$, are iid and $X^{(r)}_{\frac{m+1}{2}\left(\frac{m+1}{2}:m\right)}, X^{(r)}_{\frac{m+3}{2}(q_3(m+1):m)}, \dots, X^{(r)}_{m(q_3(m+1):m)}$ are iid. However, all units are mutually independent but not identically distributed. These measured units denote the MQRSSO.

The MQRSS estimators of the population mean in the case of an even and odd sample sizes respectively are given by

$$\bar{X}^{(r)}_{MQRSSE} = \frac{1}{m}\left(\sum_{i=1}^{\frac{m}{2}} X^{(r)}_{i(q_1(m+1):m)} + \sum_{i=\frac{m+2}{2}}^{m} X^{(r)}_{i(q_3(m+1):m)}\right), \tag{3.7}$$

$$\bar{X}^{(r)}_{MQRSSO} = \frac{1}{m}\left(\sum_{i=1}^{\frac{m-1}{2}} X^{(r)}_{i(q_1(m+1):m)} + X^{(r)}_{\frac{m+1}{2}\left(\frac{m+1}{2}:m\right)} + \sum_{i=\frac{m+3}{2}}^{m} X^{(r)}_{i(q_3(m+1):m)}\right). \tag{3.8}$$

The variances of $\bar{X}^{(r)}_{MQRSSE}$ and $\bar{X}^{(r)}_{MQRSSO}$ respectively are given by

$$\operatorname{Var}\left(\bar{X}^{(r)}_{MQRSSE}\right) = \frac{1}{m^2}\left(\sum_{i=1}^{\frac{m}{2}} \operatorname{Var}\left(X^{(r)}_{i(q_1(m+1):m)}\right) + \sum_{i=\frac{m+2}{2}}^{m} \operatorname{Var}\left(X^{(r)}_{i(q_3(m+1):m)}\right)\right), \tag{3.9}$$

$$\operatorname{Var}\left(\bar{X}^{(r)}_{MQRSSO}\right) = \frac{1}{m^2}\left(\sum_{i=1}^{\frac{m-1}{2}} \operatorname{Var}\left(X^{(r)}_{i(q_1(m+1):m)}\right) + \operatorname{Var}\left(X^{(r)}_{\left(\frac{m+1}{2}:m\right)}\right) + \sum_{i=\frac{m+3}{2}}^{m} \operatorname{Var}\left(X^{(r)}_{i(q_3(m+1):m)}\right)\right). \tag{3.10}$$

Equation (3.9) and (3.10), respectively can be written as

$$\operatorname{Var}\left(\bar{X}^{(r)}_{MQRSSE}\right) = \frac{1}{2m}\left(\operatorname{Var}\left(X^{(r)}_{(q_1(m+1):m)}\right) + \operatorname{Var}\left(X^{(r)}_{(q_3(m+1):m)}\right)\right). \tag{3.11}$$

$$\text{Var}\left(\bar{X}^{(r)}_{MQRSSO}\right)=\frac{(m-1)}{2m^2}\left(\text{Var}\left(X^{(r)}_{(q_1(m+1):m)}\right)+\text{Var}\left(X^{(r)}_{(q_3(m+1):m)}\right)\right).$$

$$+\frac{1}{m^2}\text{Var}\left(X^{(r)}_{\left(\frac{m+1}{2}:m\right)}\right) \tag{3.12}$$

The properties of the MQRSS estimators are:

(1) If the parent distribution is symmetric about the population mean μ, then
   (a) The MQRSS estimators are unbiased of the population mean.
   (b) The efficiency of $\bar{X}^{(r)}_{MQRSS}$ is increasing in *r*.
   (c) For $r \geq 2$, $\text{Var}\left(\bar{X}^{(r)}_{MQRSS}\right) < \text{Var}(\bar{X}_{RSS})$.

(2) If the underlying distribution is asymmetric μ, then for $m \geq 5$, it is found that $\text{MSE}\left(\bar{X}^{(r)}_{MQRSS}\right) < \text{Var}\left(\bar{X}_{SRS}\right)$, where the MSE is the mean square error of $\bar{X}^{(r)}_{MQRSS}$.

4. SIMULATION STUDY

To compare the proposed estimators for the population mean using MQRSS against the usual estimators using SRS and RSS methods. Six probability distribution functions were considered for the populations: uniform, normal, logistic, exponential, gamma and weibull. The efficiency of estimating the population mean using the RSS with respect to SRS estimator is defined by

$$eff\left(\bar{X}_{SRS}, \bar{X}_{RSS}\right)=\frac{\text{Var}\left(\bar{X}_{SRS}\right)}{\text{Var}\left(\bar{X}_{RSS}\right)}. \tag{4.1}$$

If the distribution is symmetric the efficiency of the MQRSS with respect to SRS is defined as

$$eff^{(r)}\left(\bar{X}_{SRS}, \bar{X}^{(r)}_{MQRSS}\right)=\frac{\text{Var}\left(\bar{X}_{SRS}\right)}{\text{Var}\left(\bar{X}^{(r)}_{MQRSS}\right)}, \tag{4.2}$$

but if the distribution is asymmetric the efficiency is defined follows

$$eff^{(r)}\left(\bar{X}_{SRS}, \bar{X}^{(r)}_{MQRSS}\right)=\frac{\text{Var}\left(\bar{X}_{SRS}\right)}{\text{MSE}\left(\bar{X}^{(r)}_{MQRSS}\right)}. \tag{4.3}$$

We compared the average of the 70,000 sample estimates. The simulation was done using the *Mathematica 5.2* program. The mean square error (MSE) of the $\bar{X}^{(r)}_{MQRSS}$ is given by

$$\text{MSE}\left(\bar{X}^{(r)}_{MQRSS}\right) = \text{Var}\left(\bar{X}^{(r)}_{MQRSS}\right) + \left(\text{Bias}\left(\bar{X}^{(r)}_{MQRSS}\right)\right)^2 . \quad (4.4)$$

It is necessary to emphasize here that to estimate the population mean by a sample of size m using RSS method we have to identify m^2 units and only measure m of them. And when we use MQRSS we must identify m^{r+1} units and only measure m of them. But the comparison between the RSS and MQRSS is done based on the same number of measured units, m, which obtained at the last stage.

Results are summarized by the efficiency values and bias in Table 1, 2 and 3 with sample sizes $m = 3, 4, 5$ and 10 for stages $r = 1,2,3$ using both RSS and MQRSS.

Table 1: The efficiency values for estimating the population mean using RSS and MQRSSO with sample size $m = 3$ for $r = 1,2$ and 3

Distribution		RSS	MQRSSO		
			$r=1$	$r=2$	$r=3$
Uniform (0,1)	*Eff*	2.000	2.000	5.713	16.501
Uniform (0,2)	*Eff*	2.000	2.000	5.773	16.041
Normal (0,1)	*Eff*	1.914	1.914	3.295	4.998
Normal (1,2)	*Eff*	1.910	1.910	3.296	5.143
Logistic (-1,1)	*Eff*	1.849	1.849	2.384	2.713
Exponential (1)	*Eff*	1.636	1.636	1.394	0.678
	Bias		0.000	0.230	0.551
Exponential (2)	*Eff*	1.687	1.641	1.359	0.691
	Bias		0.000	0.116	0.273
Exponential (3)	*Eff*	1.672	1.610	1.373	0.681
	Bias		0.000	0.078	0.184
Gamma (1,2)	*Eff*	1.655	1.615	1.374	0.680
	Bias		0.000	0.462	1.105
Gamma (1,3)	*Eff*	1.593	1.638	1.399	0.681
	Bias		0.000	0.701	1.657
Weibull (1,3)	*Eff*	1.633	1.633	1.357	0.683
	Bias		0.000	0.704	1.660

Table 2: The efficiency values for estimating the population mean using RSS and MQRSSE with sample size $m = 4$ for $r = 1,2$ and 3

Distribution		RSS	MQRSSE		
			$r = 1$	$r = 2$	$r = 3$
Uniform (0,1)	*Eff*	2.500	3.125	27.018	367.084
Uniform (0,2)	*Eff*	2.500	3.148	26.876	373.279
Normal (0,1)	*Eff*	2.347	2.034	3.432	4.901
Normal (1,2)	*Eff*	2.319	2.012	3.406	4.892
Logistic (-1,1)	*Eff*	2.229	1.706	1.904	2.017
Exponential (1)	*Eff*	1.922	1.162	0.352	0.119
	Bias	0.000	0.168	0.719	1.382
Exponential (2)	*Eff*	1.912	1.208	0.343	0.120
	Bias	0.000	0.082	0.363	0.690
Exponential (3)	*Eff*	1.900	1.175	0.352	0.117
	Bias	0.000	0.055	0.240	0.461
Gamma (1,2)	*Eff*	1.891	1.170	0.349	0.120
	Bias	0.000	0.331	1.443	2.767
Gamma (1,3)	*Eff*	1.940	1.160	0.348	0.119
	Bias	0.000	0.505	2.164	4.144
Weibull (1,3)	*Eff*	1.937	1.154	0.349	0.117
	Bias	0.000	0.505	2.160	4.152

Table 3: The efficiency values for estimating the population mean using RSS and MQRSSO with sample size $m = 5$ for $r = 1,2$ and 3

Distribution		RSS	MQRSSO		
			$r = 1$	$r = 2$	$r = 3$
Uniform (0,1)	*Eff*	3.000	2.562	9.401	33.264
Uniform (0,2)	*Eff*	3.000	2.548	9.248	33.543
Normal (0,1)	*Eff*	2.749	3.271	10.338	32.177
Normal (1,2)	*Eff*	2.812	3.327	10.276	31.935
Logistic (-1,1)	*Eff*	2.563	3.637	11.437	33.681
Exponential (1)	*Eff*	2.177	2.607	5.579	14.248
	Bias	0.000	0.151	0.130	0.085
Exponential (2)	*Eff*	2.206	2.610	5.753	14.115
	Bias	0.000	0.074	0.064	0.043
Exponential (3)	*Eff*	2.177	2.616	5.703	14.369
	Bias	0.000	0.050	0.043	0.028
Gamma (1,2)	*Eff*	2.240	2.622	5.675	14.230
	Bias	0.000	0.300	0.258	0.170
Gamma (1,3)	*Eff*	2.161	2.638	5.670	14.279
	Bias	0.000	0.449	0.389	0.256
Weibull (1,3)	*Eff*	2.236	2.649	5.640	14.227
	Bias	0.000	0.447	0.385	0.256

Table 4: The efficiency values for estimating the population mean using RSS and MQRSSE with sample size $m=10$ for $r=1,2$ and 3

Distribution		RSS	MQRSSE		
			$r=1$	$r=2$	$r=3$
Uniform (0,1)	*Eff*	5.500	5.085	38.097	250.384
Uniform (0,2)	*Eff*	5.500	5.128	38.463	250.729
Normal (0,1)	*Eff*	4.787	5.736	31.288	160.948
Normal (1,2)	*Eff*	4.779	5.850	31.721	162.267
Logistic (-1,1)	*Eff*	4.198	6.270	32.220	152.985
Exponential (1)	*Eff*	3.440	3.281	15.024	38.506
	Bias		0.117	0.056	0.042
Exponential (2)	*Eff*	3.426	3.288	14.916	38.924
	Bias		0.059	0.028	0.021
Exponential (3)	*Eff*	3.394	3.252	14.844	38.725
	Bias		0.039	0.019	0.014
Gamma (1,2)	*Eff*	3.440	3.276	14.878	38.583
	Bias		0.234	0.113	0.084
Gamma (1,3)	*Eff*	3.460	3.274	14.963	39.124
	Bias		0.354	0.170	0.125
Weibull (1,3)	*Eff*	3.471	3.245	14.808	38.509
	Bias		0.352	0.170	0.126

Considering the results Tables 1-4, we can conclude the following:

(1) A gain in efficiency is obtained using MQRSS for different values of m with $r=1,2$ and 3 for all symmetric distributions considered in this study and for asymmetric distributions if the sample size $m \geq 5$ and for $m \leq 4$ in some cases.

(2) For asymmetric distributions considered in this study, the MQRSS estimator has a smaller bias. As an example, for estimating the population mean of exponential distribution with parameter 3 for $m=10$ and $r=3$ the efficiency of MQRSS is 38.725 and the bias 0.014.

(3) For all symmetric distributions we considered, the efficiency of the $\overline{X}^{(r)}_{MQRSS}$ is increasing in r. For example, for $m=5$ and $r=1,2$ and 3 the efficiency values of MQRSS are 3.327, 10.334 and 30.796 respectively for estimating the population mean of a normal distribution with parameters 1 and 2. Also, for asymmetric distributions the efficiency of $\overline{X}^{(r)}_{MQRSS}$ is increasing in *r* on the converse of the bias which is decreasing in *r*.

(4) For $r > 1$, the $\bar{X}^{(r)}_{MQRSS}$ is more efficient than $\bar{X}_{RSS}$ with the same number of quantified units. As an example, for $m = 3$ and $r = 2$ the efficiency of the MQRSS is 5.713 for estimating the population mean of a standard uniform distribution.

(5) For $r = 1$ and 2, the MQRSS is same as the QRSS and DQRSS respectively and it is found that the MQRSS is more efficient than both of QRSS and DQRSS $r > 1$ and $r > 2$ respectively.

5. APPLICATION TO REAL DATA SET

We illustrate the performance of the multistage ranked set samples method for mean estimation using a collection of real data set which consists of the olive yield of each of 64 trees, for more details see Al-Saleh and Al-Omari (2002). In this study, balanced ranked set sampling is considered. All sampling was done without replacement using the statistical programming *Mathematica 5.2*. We obtained the mean and the variance of the sample mean using SRS, RSS and MQRSS methods with set sizes $m = 3, 4$ and 5. We compared the averages of the 70,000 sample estimate.

Let, u_i be the olive yield of the ith tree i = 1,2, ..., 64 . The mean μ, and the variance σ^2 of the population, respectively, are

$$\mu = \frac{1}{64}\sum_{i=1}^{64} u_i = 9.777 \text{ kg/tree}, \text{ and } \sigma^2 = \frac{1}{64}\sum_{i=1}^{64}(u_i - \mu)^2 = 26.112 \text{ kg}^2/\text{tree}.$$

The skewness, kurtosis, and the median of the population, respectively, are 0.484, 2.071 and 8.250. The skewness should be close to zero for symmetrically distributed data, while for our data that considered, the skewness is 0.484, which mean that these data are asymmetrically distributed. Hence, we compute the mean square error of $\bar{X}^{(r)}_{MQRSS}$ and the efficiency values of $\bar{X}_{RSS}$ and $\bar{X}^{(r)}_{MQRSS}$ relative to $\bar{X}_{SRS}$ can be computed using the relations 13 and 15 respectively. We calculate the efficiency of RSS and of MQRSS and $m = 3, 4, 5$. Results are summarized by the efficiency and the bias values in Tables 5 with $m = 3, 4, 5$ for $r = 1, 2, 3$ for RSS and MQRSS.

Table 5: The efficiency values of RSS and MQRSS relative to SRS with sample size $m = 3, 4, 5$ for stages $r = 1, 2, 3$

Methods			Sample size		
			$m = 3$	$m = 4$	$m = 5$
SRS		Mean	9.787	9.784	9.772
		Variance	8.344	6.159	4.843
RSS		Mean	9.784	9.773	9.773
		Variance	4.294	2.564	1.696
		Efficiency	1.954	2.383	2.870
MQRSS	Stage				
	$r = 1$	Mean	9.748	10.203	9.435
		Bias	0.029	0.426	0.341
		MSE	4.279	2.468	1.976
		Efficiency	1.922	2.482	2.430
	$r = 2$	Mean	10.183	11.058	9.666
		Bias	0.407	1.281	0.111
		MSE	1.760	2.070	0.598
		Efficiency	4.741	2.960	8.061
	$r = 3$	Mean	10.439	11.521	9.843
		Bias	0.663	1.744	0.067
		MSE	1.014	3.104	0.151
		Efficiency	8.170	1.984	32.124

Based on Table 5, the MQRSS mean at ant stage is closed to the population mean 9.777, and the bias that because our data are asymmetrically distributed. It can be noted that the MQRSS is much more efficient than SRS.

5. CONCLUDING REMARKS

It is recommended to use MQRSS for estimating the population mean if the underlying distribution is symmetric, and if the distribution is asymmetric with larger sample size when estimating the mean, since only we have to do is identify and measure the first or the third quartile of the ith sample.

ACKNOWLEDGEMENTS

The authors are indebted to the referee and the Editor for helpful suggestions.

REFERENCES

1. Al-Saleh, M.F. and Al-Kadiri, M. (2000). Double ranked set sampling. *Statist. Probab. Lett.*, 48, 205 - 212.
2. Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. *J. Statist. Plann. and Infer.*, 102, 273-286.
3. Dell, T.R., and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrika* 28, 545-555.
4. Jemain, A.A. and Al-Omari, A.I. (2006). Double quartile ranked set samples. *Pak. J. Statist.*, 22(3), 217-228.
5. David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, Third Edition, John Wiley & Sons, Inc., Hoboken, New Jersey.
6. McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets. *Aust. J. Agri. Res.*, 3, 385-390.
7. Muttlak, H.A. (2003). Investigating the use of quartile ranked set samples for estimating the population mean. *J. App. Math. and Comput.*, 146, 437- 443.
8. Muttlak, H.A. (1997). Median ranked set sampling, *J. App. Statist. Scien*., 6(4), 577-586.
9. Samawi, H., Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling, *The Biometrical Journal.* 30, 577-586.
10. Takahasi, K. and Wakimoto, K. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math*. 20, 1-31.

CHAPTER SIXTEEN

Modified Ratio Estimator for the Population Mean using Double Median Ranked Set Sampling

Abdul Aziz Jemain, Amer Al-Omari and **Kamarulzaman Ibrahim**
School of Mathematical Sciences, Faculty of Science and Technology,
University Kebangsaan Malaysia 43600 UKM Bangi, Selangor, Malaysia

SUMMARY

In this paper, ranked set sampling (RSS), median ranked set sampling (MRSS) and double median ranked set sampling (DMRSS) methods are used for estimating the population mean based on a modified ratio estimator. It is found that, RSS, MRSS and DMRSS produce approximately unbiased estimators of the population mean and these estimators are more efficient than those obtained using simple random sampling (SRS) based on the same sample size. Also, it is found that, DMRSS is more efficient than both RSS and MRSS methods.

1. INTRODUCTION

Ranked set sampling is introduced by McIntyre (1952) for estimating mean pasture and forage yields as a more efficient and cost effective method than the commonly used simple random sampling in the situations where visual ordering of sample units can be done easily, but the exact measurement of the units is difficult and expensive. Takahasi and Wakimoto (1968) provided the necessary mathematical theory of RSS. Samawi and Muttlak (1996) suggested the used of RSS to estimate the population ratio. Muttlak (1997) suggested using median ranked set sampling (MRSS) to estimate the population mean. Samawi and Muttlak (2001) used MRSS to estimate the population ratio. Al-Saleh and Al-Kadiri (2000) suggested double ranked set sampling method (DRSS) for estimating the population mean, and they showed that the ranking at the second stage is easier than the ranking at the first stage. Samawi and Tawalbeh (2002) suggested double median ranked set sampling method for estimating the population mean and ratio. Jemain and Al-Omari (2006) proposed multistage median ranked set sampling (MMRSS) method for estimating the population mean.

In this paper, RSS, MRSS and DMRSS methods are used for estimating the population mean of the variable of interest Y using information in the auxiliary variable X based on a modified ratio estimator. The modified ratio estimators for the population mean obtained using RSS, MRSS and DMRSS are compared with the counterparts using SRS.

2. SAMPLING METHODS

2.1 Ranked set sampling

The RSS 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 the variable of interest. From the first set of m units, the smallest unit is measured. From the second set of m units, the second smallest unit is measured. The process is continued until from the mth set of m units the largest unit is measured. Repeating the process n times yields a set of size mn from the initial nm^2 units.

2.2 Median ranked set sampling

In median ranked set sampling (MRSS) method select m random samples each of size m units from the population and rank the units within each sample with respect to the variable of interest. If the sample size m is odd, then from each sample select for measurement the $((m+1)/2)$th smallest rank (the median of the sample). If the sample size m is even, then select for measurement the $(m/2)$th smallest rank from the first $m/2$ samples, and the $((m+2)/2)$th smallest rank from the second $m/2$ samples. The cycle can be repeated n times if needed to obtain a sample of size nm (Muttlak 1997).

2.3 Double ranked set sampling

The double ranked set sampling (DRSS) procedure can be described as the followings: Identify m^3 units from the target population and divide these units randomly into m sets each of size m^2. The procedure of ranked set sampling is applied on each m^2 units to obtain m ranked set sampling each of size m, then again apply the ranked set sampling procedure on the m ranked set sampling sets obtained in the first stage to obtain a DRSS of size m (Al-Saleh and Al-Kadiri 2000).

3. ESTIMATORS FOR THE POPULATION MEAN

Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_m,Y_m)$ be a bivariate normal random sample with pdf $f(x,y)$, cdf $F(x,y)$, with means μ_X, μ_Y, variances σ_X, σ_Y and correlation coefficient ρ. Assume that the ranking is performed on the variable X to estimate the mean of the variable of interest Y. Let (X_{11},Y_{11}), $(X_{12},Y_{12}),\ldots,$ (X_{1m},Y_{1m}), (X_{21},Y_{21}), $(X_{22},Y_{22}),\ldots,$ $(X_{2m},Y_{2m}),\ldots,(X_{m1},Y_{m1}),(X_{m2},Y_{m2}),\ldots,(X_{mm},Y_{mm})$ be m independent bivariate normal random samples each of size m. Let

$\left(X_{i(1)},Y_{i[1]}\right),\left(X_{i(2)},Y_{i[2]}\right),\ldots,\left(X_{i(m)},Y_{i[m]}\right)$ be the order statistics of $X_{i1},X_{i2},...,X_{im}$ and the judgment order of Y_{i1}, $Y_{i2},\ldots,Y_{im}$, $\left(i=1,2,...,m\right)$.

3.1 Using SRS

The SRS estimator of the population mean μ_Y as suggested by Singh and Tailor (2003) is given by

$$\hat{\mu}_{YSRS}=\bar{Y}_{SRS}\left(\frac{\mu_X+\rho}{\bar{X}_{SRS}+\rho}\right), \tag{1}$$

with bias and MSE, respectively, given by

$$\text{Bias}\left(\hat{\mu}_{YSRS}\right)=\frac{1-f}{m}\mu_Y C_X^2\theta\left(\theta-K\right), \tag{2}$$

and

$$\text{MSE}\left(\hat{\mu}_{YSRS}\right)=\frac{1-f}{m}\mu_Y^2\left(C_Y^2+\theta C_X^2\left(\theta-2K\right)\right) \tag{3}$$

where, *M* and *m* are the population and sample size respectively and

$$f=\frac{m}{M},\quad C_Y^2=\frac{\sigma_Y^2}{\mu_Y^2},\quad \theta=\frac{\mu_x}{\mu_x+\rho},\quad C_X^2=\frac{\sigma_X^2}{\mu_X^2},\quad K=\rho\frac{C_Y}{C_X},\quad \rho=\frac{\sigma_{XY}}{\sigma_X\sigma_Y},$$

$$\sigma_X^2=\left(M-1\right)^{-1}\sum_{i=1}^{M}\left(X_i-\mu_X\right)^2,\ \sigma_Y^2=\left(M-1\right)^{-1}\sum_{i=1}^{M}\left(Y_i-\mu_Y\right)^2 \text{ and}$$

$$\sigma_{XY}^2=\left(M-1\right)^{-1}\sum_{i=1}^{M}\left(X_i-\mu_X\right)\left(Y_i-\mu_Y\right).$$

For more details about ratio estimation see Raj (1968) and Cochran (1977). The transformed ratio estimator suggested by Singh and Tailor (2003) can be exploited to estimate the population mean μ_Y using RSS, MRSS and DMRSS methods.

3.2 Using RSS

Assume that the ranking is performed on the auxiliary variable *X*, the only measured units, using RSS are denoted by $\left(X_{1(1)},Y_{1[1]}\right),\left(X_{2(2)},Y_{2[2]}\right),\ldots,$ $\left(X_{m(m)},Y_{m[m]}\right)$. The suggested RSS estimator of the population mean μ_Y from a bivariate normal sample of size *m* is defined as:

$$\hat{\mu}_{YRSS}=\bar{Y}_{RSS}\left(\frac{\mu_X+\rho}{\bar{X}_{RSS}+\rho}\right), \tag{4}$$

where $\bar{X}_{RSS} = \frac{1}{m}\sum_{i=1}^{m} X_{i(i)}$ and $\bar{Y}_{RSS} = \frac{1}{m}\sum_{i=1}^{m} Y_{i[i]}$. This estimator can be approximated using Taylor expansion as:

$$\hat{\mu}_{YRSS} \cong \bar{Y}_{RSS} - H\left(\bar{X}_{RSS} - \mu_X\right) + HG\left(\bar{X}_{RSS} - \mu_X\right)^2 - G\left(\bar{X}_{RSS} - \mu_X\right)\left(\bar{Y}_{RSS} - \mu_Y\right) \quad (5)$$

where, $H = \mu_Y / \left(\mu_X + \rho\right)$ and $G = 1/\left(\mu_X + \rho\right)$. Take the expectation of (5) yields

$$E\left(\hat{\mu}_{YRSS}\right) \cong \mu_Y + G\left(H - \beta\right)\text{Var}\left(\bar{X}_{RSS}\right), \quad (6)$$

since, $\text{Cov}\left(\bar{X}_{RSS}, \bar{Y}_{RSS}\right) = E\left(\left(\bar{X}_{RSS} - \mu_X\right)\left(\bar{Y}_{RSS} - \mu_Y\right)\right)$, $\text{Cov}\left(\bar{X}_{RSS}, \bar{Y}_{RSS}\right) = \beta\text{Var}\left(\bar{X}_{RSS}\right)$ and $\beta = \rho\frac{\sigma_Y}{\sigma_X}$. Therefore, the bias of $\hat{\mu}_{YRSS}$ is given by

$$\text{Bias}\left(\hat{\mu}_{YRSS}\right) \cong G\left(H - \beta\right)\text{Var}\left(\bar{X}_{RSS}\right). \quad (7)$$

To the first order of approximation, the estimator of the population mean $\hat{\mu}_{YRSS}$ is given by:

$$\hat{\mu}_{YRSS} \cong \bar{Y}_{RSS} - H\left(\bar{X}_{RSS} - \mu_X\right). \quad (8)$$

The expectation of (8) is $E\left(\hat{\mu}_{YRSS}\right) \cong \mu_Y$, implies that the estimator is approximately unbiased. Using $\text{Var}\left(\bar{Y}_{RSS}\right) \cong \beta^2\text{Var}\left(\bar{X}_{RSS}\right) + \frac{1}{m}\sigma_Y^2\left(1 - \rho^2\right)$, the variance and MSE of (8), respectively, can be found as

$$\text{Var}\left(\hat{\mu}_{YRSS}\right) \cong \left(H - \beta\right)^2 \text{Var}\left(\bar{X}_{RSS}\right) + \frac{1}{m}\sigma_Y^2\left(1 - \rho^2\right), \quad (9)$$

and

$$\text{MSE}\left(\hat{\mu}_{YRSS}\right) \cong \left(H - \beta\right)^2 \text{Var}\left(\bar{X}_{RSS}\right)\left[1 + G^2\text{Var}\left(\bar{X}_{RSS}\right)\right] + \frac{1}{m}\sigma_Y^2\left(1 - \rho^2\right). \quad (10)$$

3.3 Using MRSS

If the sample size m is odd, then $\left(X_{1\left(\frac{m+1}{2}\right)}, Y_{1\left[\frac{m+1}{2}\right]}\right), \left(X_{2\left(\frac{m+1}{2}\right)}, Y_{2\left[\frac{m+1}{2}\right]}\right), \ldots,$

$\left(X_{m\left(\frac{m+1}{2}\right)}, Y_{m\left[\frac{m+1}{2}\right]}\right)$ denote the measured MRSSO. If the sample size m is even, then $\left(X_{1\left(\frac{m}{2}\right)}, Y_{1\left[\frac{m}{2}\right]}\right)$, $\left(X_{2\left(\frac{m}{2}\right)}, Y_{2\left[\frac{m}{2}\right]}\right)$,..., $\left(X_{\frac{m}{2}\left(\frac{m}{2}\right)}, Y_{\frac{m}{2}\left[\frac{m}{2}\right]}\right)$, $\left(X_{\frac{m+2}{2}\left(\frac{m+2}{2}\right)}, Y_{\frac{m+2}{2}\left[\frac{m+2}{2}\right]}\right)$, ..., $\left(X_{m\left(\frac{m+2}{2}\right)}, Y_{m\left[\frac{m+2}{2}\right]}\right)$ denote the measured MRSSE. The estimator of the population mean μ_Y for a MRSS of size m is defined as:

$$\hat{\mu}_{YMRSS} = \bar{Y}_{MRSS}\left(\frac{\mu_X + \rho}{\bar{X}_{MRSS} + \rho}\right), \tag{11}$$

where if m is odd $\bar{X}_{MRSS}$ and $\bar{Y}_{MRSS}$, respectively, are defined as

$$\bar{X}_{MRSSO} = \frac{1}{m}\sum_{i=1}^{m} X_{i\left(\frac{m+1}{2}\right)} \text{ and } \bar{Y}_{MRSSO} = \frac{1}{m}\sum_{i=1}^{m} Y_{i\left[\frac{m+1}{2}\right]},$$

and if m is even, $\bar{X}_{MRSS}$ and $\bar{Y}_{MRSS}$, respectively, are given by:

$$\bar{X}_{MRSSE} = \frac{1}{m}\left(\sum_{i=1}^{\frac{m}{2}} X_{i\left(\frac{m}{2}\right)} + \sum_{i=\frac{m+2}{2}}^{m} X_{i\left(\frac{m+2}{2}\right)}\right) \text{ and } \bar{Y}_{MRSSE} = \frac{1}{m}\left(\sum_{i=1}^{\frac{m}{2}} Y_{i\left[\frac{m}{2}\right]} + \sum_{i=\frac{m+2}{2}}^{m} Y_{i\left[\frac{m+2}{2}\right]}\right).$$

The estimator $\hat{\mu}_{YMRSS}$ can be approximated as

$$\begin{aligned}\hat{\mu}_{YMRSS} \cong \bar{Y}_{MRSS} - H\left(\bar{X}_{MRSS} - \mu_X\right) + HG\left(\bar{X}_{MRSS} - \mu_X\right) \\ - G\left(\bar{X}_{MRSS} - \mu_X\right)\left(\bar{Y}_{MRSS} - \mu_Y\right)\end{aligned} \tag{12}$$

Take the expectation of (12) yields

$$E\left(\hat{\mu}_{YMRSS}\right) \cong \mu_Y + G\left(H - \beta\right)\text{Var}\left(\bar{X}_{MRSS}\right). \tag{13}$$

Therefore, the bias of $\hat{\mu}_{YMRSS}$ is given by

$$\text{Bias}\left(\hat{\mu}_{YMRSS}\right) \cong G\left(H - \beta\right)\text{Var}\left(\bar{X}_{MRSS}\right). \tag{14}$$

And the variance and MSE, respectively, are given by

$$\operatorname{Var}\left(\hat{\mu}_{YMRSS}\right) \cong \left(H-\beta\right)^2 \operatorname{Var}\left(\bar{X}_{MRSS}\right) + \frac{1}{m}\sigma_Y^2\left(1-\rho^2\right), \tag{15}$$

and

$$\operatorname{MSE}\left(\hat{\mu}_{YMRSS}\right) \cong \left(H-\beta\right)^2 \operatorname{Var}\left(\bar{X}_{MRSS}\right)\left[1+G^2\operatorname{Var}\left(\bar{X}_{MRSS}\right)\right] + \frac{1}{m}\sigma_Y^2\left(1-\rho^2\right). \tag{16}$$

3.4 Using DMRSS

If the sample size m is odd, then $\left(X^*_{1\left(\frac{m+1}{2}\right)}, Y^*_{1\left[\frac{m+1}{2}\right]}\right)$, $\left(X^*_{2\left(\frac{m+1}{2}\right)}, Y^*_{2\left[\frac{m+1}{2}\right]}\right)$,…, $\left(X^*_{m\left(\frac{m+1}{2}\right)}, Y^*_{m\left[\frac{m+1}{2}\right]}\right)$ denote the measured DMRSSO. And if m is even, then $\left(X^*_{1\left(\frac{m}{2}\right)}, Y^*_{1\left[\frac{m}{2}\right]}\right)$, $\left(X^*_{2\left(\frac{m}{2}\right)}, Y^*_{2\left[\frac{m}{2}\right]}\right)$,…, $\left(X^*_{\frac{m}{2}\left(\frac{m}{2}\right)}, Y^*_{\frac{m}{2}\left[\frac{m}{2}\right]}\right)$, $\left(X^*_{\frac{m+2}{2}\left(\frac{m+2}{2}\right)}, Y^*_{\frac{m+2}{2}\left[\frac{m+2}{2}\right]}\right)$, …, $\left(X^*_{m\left(\frac{m+2}{2}\right)}, Y^*_{m\left[\frac{m+2}{2}\right]}\right)$ denote the DMRSSE. The DMRSS estimator of the population mean μ_Y is given by

$$\hat{\mu}_{YDMRSS} = \bar{Y}^*_{DMRSS}\left(\frac{\mu_X+\rho}{\bar{X}^*_{DMRSS}+\rho}\right), \tag{17}$$

where if m is odd $\bar{X}^*_{DMRSS}$ and $\bar{Y}^*_{DMRSS}$, respectively, are defined as

$$\bar{X}^*_{DMRSSO} = \frac{1}{m}\sum_{i=1}^{m} X^*_{i\left(\frac{m+1}{2}\right)} \text{ and } \bar{Y}^*_{DMRSSO} = \frac{1}{m}\sum_{i=1}^{m} Y^*_{i\left[\frac{m+1}{2}\right]},$$

and if m is even, $\bar{X}^*_{DMRSS}$ and $\bar{Y}^*_{DMRSS}$, respectively, are defined as

$$\bar{X}^*_{DMRSSE} = \frac{1}{m}\left(\sum_{i=1}^{\frac{m}{2}} X^*_{i\left(\frac{m}{2}\right)} + \sum_{i=\frac{m+2}{2}}^{m} X^*_{i\left(\frac{m+2}{2}\right)}\right)$$

and

$$\bar{Y}^*_{DMRSSE} = \frac{1}{m}\left(\sum_{i=1}^{\frac{m}{2}} Y^*_{i\left[\frac{m}{2}\right]} + \sum_{i=\frac{m+2}{2}}^{m} Y^*_{i\left[\frac{m+2}{2}\right]}\right).$$

The estimator $\hat{\mu}_{YDMRSS}$ can be approximated as

$$\hat{\mu}_{YMRSS} \cong \bar{Y}^*_{DMRSS} - H\left(\bar{X}^*_{DMRSS} - \mu_X\right) + HG\left(\bar{X}^*_{DMRSS} - \mu_X\right) - G\left(\bar{X}^*_{DMRSS} - \mu_X\right)\left(\bar{Y}^*_{DMRSS} - \mu_Y\right) \tag{18}$$

Take the expectation of (18) yields

$$E\left(\hat{\mu}_{YDMRSS}\right) \cong \mu_Y + G\left(H - \beta\right)\mathrm{Var}\left(\bar{X}^*_{DMRSS}\right). \tag{19}$$

Therefore, the bias, variance and MSE of $\hat{\mu}_{YDMRSS}$, respectively are given by

$$\mathrm{Bias}\left(\hat{\mu}_{YDMRSS}\right) \cong G\left(H - \beta\right)\mathrm{Var}\left(\bar{X}^*_{DMRSS}\right) \tag{20}$$

$$\mathrm{Var}\left(\hat{\mu}_{YDMRSS}\right) \cong \left(H - \beta\right)^2 \mathrm{Var}\left(\bar{X}^*_{DMRSS}\right) + \frac{1}{m}\sigma_Y^2\left(1 - \rho^2\right), \tag{21}$$

and

$$\mathrm{MSE}\left(\hat{\mu}_{YDMRSS}\right) \cong \left(H - \beta\right)^2 \mathrm{Var}\left(\bar{X}^*_{DMRSS}\right)\left[1 + G^2\mathrm{Var}\left(\bar{X}^*_{DMRSS}\right)\right] + \frac{1}{m}\sigma_Y^2\left(1 - \rho^2\right) \tag{22}$$

4. SIMULATION STUDY

A simulation study was conducted to investigate the performance of SRS, RSS, MRSS and DMRSS methods for estimating the population mean where the ranking was performed on the variable *X*. The samples were generated from bivariate normal distribution with parameters $\mu_X = 6$, $\mu_Y = 3$, $\sigma_X = \sigma_Y = 1$ and $\rho = \pm 0.99$, ± 0.90, $\pm 0.80, \pm 0.70, \pm 0.50$. Based on 60,000 replications, the efficiency and the bias of $\hat{\mu}_{YSRS}$, $\hat{\mu}_{YRSS}$, $\hat{\mu}_{YMRSS}$ and $\hat{\mu}_{YDMRSS}$ are obtained and the results for $m = 3,4$ are presented in Table 1 and for $m = 5,6$ in Table 2. The efficiency of $\hat{\mu}_{YRSS}$, $\hat{\mu}_{YMRSS}$ and $\hat{\mu}_{YDMRSS}$ with respect to $\hat{\mu}_{YSRS}$, respectively, are defined as:

$$\mathit{eff}\left(\hat{\mu}_{YSRS}, \hat{\mu}_{YRSS}\right) = \frac{\mathrm{MSE}\left(\hat{\mu}_{YSRS}\right)}{\mathrm{MSE}\left(\hat{\mu}_{YRSS}\right)}, \tag{21}$$

$$eff\left(\hat{\mu}_{YSRS},\hat{\mu}_{YMRSS}\right)=\frac{\text{MSE}\left(\hat{\mu}_{YSRS}\right)}{\text{MSE}\left(\hat{\mu}_{YMRSS}\right)}, \tag{22}$$

$$eff\left(\hat{\mu}_{YSRS},\hat{\mu}_{YDMRSS}\right)=\frac{\text{MSE}\left(\hat{\mu}_{YSRS}\right)}{\text{MSE}\left(\hat{\mu}_{YDMRSS}\right)}. \tag{23}$$

From the results of simulation given in Tables 1 and 2, we can conclude the followings:

1. The estimators of population mean obtained by RSS, MRSS and DMRSS is more efficient compared to the usual SRS estimator based on the same number of measured units.
2. Based on the same sample size, $eff\left(\hat{\mu}_{YSRS},\hat{\mu}_{YDMRSS}\right) > eff\left(\hat{\mu}_{YSRS},\hat{\mu}_{YMRSS}\right) > eff\left(\hat{\mu}_{YSRS},\hat{\mu}_{YRSS}\right)$. This is particularly apparent when ρ is close to 1. For example, for $m=3$, the efficiency of RSS, MRSS and DMRSS, respectively are 1.846, 2.099 and 4.127 with $\rho=0.99$.
3. It is found that, for the same value of the correlation coefficient the absolute value of the bias satisfies the inequality, $\text{Bias}\left(\hat{\mu}_{YDMRSS}\right)\le \text{Bias}\left(\hat{\mu}_{YMRSS}\right)$. For example, for $\rho=-0.80$ and $m=6$, the absolute bias of the estimators using MRSS and DMRSS, respectively, are 0.008 and 0.004.
4. The efficiency of each the estimators, $\hat{\mu}_{YRSS}$, $\hat{\mu}_{YMRSS}$ and $\hat{\mu}_{YDMRSS}$ is increasing with the sample size for the same value of ρ. For example, for $m=3,4,5,6$, the efficiency of DMRSS estimator, respectively, are 1.100, 1.114, 1.118 and 1.125 for $\rho=0.70$.
5. For the estimators considered, the negative values of the correlation coefficient ρ give higher values of the efficiency than the positive values for a given sample size. For example, with $m=5$ using DMRSS for $\rho=0.90$ and -0.90, the efficiency values are 1.922 and 6.644 respectively.
6. The efficiency of estimators using any of RSS, MRSS and DMRSS methods is found to be increasing as the magnitude of the correlation coefficient increase. As an example, for $m=6$, the efficiency of using MRSS for $\rho=0.50,0.70,0.80,0.90$, 0.99 are 1.012, 1.080, 1.231, 1.647 and 3.351. Similarly, for $\rho=-0.50,-0.70,$ $-0.80,-0.90,-0.99$ the efficiency are 2.275, 3.474, 4.774, 7.480 and 15.089 respectively.

Table 1
The efficiency and bias values of estimating the population mean using RSS, MRSS and DMRSS with respect to SRS for $m=3$, 4.

ρ		$m=3$			$m=4$		
		RSS	**MRSS**	**DMRSS**	**RSS**	**MRSS**	**DMRSS**
0.99	*eff*	1.846	2.099	4.127	2.264	2.565	5.615
	Bias	-0.010	-0.008	-0.004	-0.006	-0.006	-0.002
0.90	*eff*	1.350	1.413	1.752	1.440	1.512	1.868
	Bias	-0.009	-0.007	-0.001	-0.006	-0.006	-0.001
0.80	*eff*	1.121	1.186	1.287	1.176	1.205	1.297
	Bias	-0.004	-0.006	-0.003	-0.004	-0.003	0.000
0.70	*eff*	1.039	1.080	1.100	1.081	1.082	1.114
	Bias	-0.006	-0.003	-0.003	-0.003	-0.003	-0.001
0.50	*eff*	1.006	1.015	1.018	1.014	1.017	1.026
	Bias	0.001	-0.001	0.000	0.000	-0.001	0.000
-0.99	*eff*	1.954	2.326	5.114	2.465	2.849	7.554
	Bias	0.037	0.032	0.011	0.026	0.021	0.005
-0.90	*eff*	1.835	2.098	3.963	2.213	2.512	5.173
	Bias	0.037	0.029	0.013	0.023	0.018	0.005
-0.80	*eff*	1.710	1.904	3.163	2.028	2.236	3.798
	Bias	0.035	0.028	0.011	0.014	0.019	0.006
-0.70	*eff*	1.599	1.782	2.598	1.837	1.991	2.981
	Bias	0.032	0.022	0.010	0.022	0.014	0.005
-0.50	*eff*	1.413	1.515	1.933	1.570	1.632	2.130
	Bias	0.023	0.017	0.007	0.013	0.011	0.003

Table 2
The efficiency and bias values of estimating the population mean using RSS, MRSS and DMRSS with respect to SRS for $m = 5$, 6.

ρ		$m=5$			$m=6$		
		RSS	**MRSS**	**DMRSS**	**RSS**	**MRSS**	**DMRSS**
0.99	*eff*	2.631	3.093	7.335	2.921	3.531	8.574
	Bias	-0.004	-0.004	0.001	-0.003	-0.002	0.001
0.90	*eff*	1.516	1.611	1.922	1.587	1.647	1.983
	Bias	-0.003	-0.003	0.001	-0.001	-0.002	0.002
0.80	*Eff*	1.220	1.221	1.314	1.211	1.231	1.321
	Bias	-0.004	-0.004	0.000	-0.002	-0.003	0.002
0.70	*eff*	1.078	1.085	1.118	1.085	1.080	1.125
	Bias	-0.003	-0.001	0.000	-0.001	0.000	0.000
0.50	*eff*	1.006	1.012	1.027	1.010	1.012	1.029
	Bias	-0.001	0.001	0.000	0.000	0.000	0.000
-0.99	*eff*	2.885	3.595	11.580	3.298	4.223	15.089
	Bias	0.013	0.013	-0.001	0.011	0.009	-0.002
-0.90	*eff*	2.581	2.994	6.644	2.830	3.350	7.480
	Bias	0.018	0.011	-0.001	0.009	0.010	-0.005
-0.80	*eff*	2.268	2.545	4.448	2.466	2.801	4.774
	Bias	0.013	0.009	0.002	0.006	0.008	-0.004
-0.70	*eff*	2.017	2.214	3.346	2.135	2.389	3.474
	Bias	0.013	0.011	-0.001	0.009	0.007	-0.003
-0.50	*eff*	1.669	1.794	2.189	1.708	1.848	2.275
	Bias	0.008	0.005	-0.001	0.010	0.004	-0.002

5. CONCLUSION

The suggested estimators for the population mean using RSS, MRSS and DMRSS methods are more efficient than the SRS estimator based on the same sample size. The efficiency of the suggested estimators is increasing in the sample size and also it is increasing as the magnitude of the correlation coefficient increase. When these three methods RSS, MRSS and DMRSS are compared, it is found that DMRSS are most efficient.

REFERENCES

1. Al-Saleh, M.F. and Al-Kadiri, M.A. (2000). Double ranked set sampling. *Statist. and Prob. Lett.*, 48(2), 205-212.
2. Cochran, W.G. (1977). Sampling Technique. 3rd edition, Wiley and Sons. New York.
3. Jemain, A.A. and Al-Omari, A.I. (2006). Multistage median ranked set samples for estimating the population mean, *Pak. J. Statist.*, 22(3), 195-207.
4. McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets, *Aust. J. Agri. Res.*, 3, 385-390.
5. Muttlak, H.A. (1997). Median ranked set sampling, *J. App. Statist. Scien.,* 6(4), 245-255.
6. Samawi. H.M. and Muttlak, H.A. (1996). Estimation of ratio using rank set sampling. *The Biometrical Journal,* 36(6), 753-764.
7. Samawi. H.M. and Muttlak, H.A. (2001). On ratio estimation using median ranked set sampling. *J. App. Statist. Scien.,* 10(2), 89-98.
8. Samawi. H.M. and Tawalbeh, E.M. (2002). Double median ranked set sample: Comparison to other double ranked samples for mean and ratio estimators. *J. Mod. App. Statist. Meth.,* 1(2), 428-442.
9. Singh, H.P. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6(4), 555-560.
10. Raj, D. (1968). *Sampling Theory*. TMH Edition, Tata McGraw-Hill Publishing Company LTD, Bombay-New Delhi.
11. Takahasi K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.,* 20, 1-31.