Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 68(1) 2014 55-65

A Generalized Approach on Ranked Set Sampling Theory for Large Set Size

Girish Chandra[1], Neeraj Tiwari[2] and Sawal Kishore Singh[3]
[1]*Tropical Forest Research Institute, Jabalpur*
[2]*Kumaun University, SSJ Campus, Almora*
[3]*National Institute of Technology, Patna*

Received 12 April 2010; Revised 18 October 2013; Accepted 11 December 2013

SUMMARY

In situations where the experimental or sampling units in a study can be more easily ranked than quantified, McIntyre (1952) proposed that the ranked set sampling (RSS) provides the unbiased estimator of population mean with smaller variance compared to the simple random sampling (SRS) of the same sample size. McIntyre's concept of RSS is completely non-parametric in nature and assumed that the set size used in the experiment is small. When set size is very large, there arises a case in which it is very difficult to assign proper individual rank to all the units by visual inspection or other rough-gauging methods. This means that for large set size, there may be more than one order statistics corresponding to each rank orders. In this paper we have generalized the ranked set sampling theory for large set size using the idea of Probability Proportion to Rank Size Matrix (PPRSM). An estimator of population mean has been proposed. Properties of the proposed estimator in the lines of Yanagawa and Shirahata (1976) have been discussed.

Keywords: Probability proportion to rank size matrix, Order statistic, Ranked set sample, Rank size matrix, Relative precision, Unbiased estimators.

1. INTRODUCTION

Ranked Set Sampling (RSS) is useful especially when visual inspection, or other rough gauging methods can order elements of each set readily, whereas the exact measurement of an element is costly in time or effort. Ranked set sampling was actually applied in the pastoral Research Laboratory, CSIRO, at Armidale, N.S.W., Australia; when a plate with four holes was randomly set on a field, pasture in each hole was visually compared, a hole was selected and its pasture was repeated in, and its dried weight was measured. The method has also been applied to estimate rice crop yields in Okinawa, Japan, where squares of rice field were visually compared before selecting for measurement. Takahasi and Wakimoto (1968) derived the theory of RSS and proposed independently the same estimator $\overline{X}$ of population mean μ, as suggested by McIntyre (1952). Yanagawa and Shirahata (1976) proposed ranked set sampling theory with a selective probability matrix to estimate the population mean. Their estimator was a generalization of McIntyre and Takahasi and Wakimoto's estimator. Yanagawa and Chen (1980) considered the ranked set sampling theory with half-selective probability matrix. The estimator, termed as MG estimator, was constructed with half of the selective probability matrix. Chen (1983) introduced an estimator of the population mean by using the idea of selective probability vector and used the optimization algorithm of linear programming to find the optimal solution of the selective probability vector under the condition of unbiasedness.

The basic concept behind RSS with equal allocation can be briefly described as follows: Suppose $(X_1, X_2, \ldots X_n)$ is a simple random sample from $F(x)$

Corresponding author: Girish Chandra
E-mail address: gchandra23@yahoo.com

with a mean μ and a finite variance σ^2. Then a standard unbiased estimator of μ is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ with var $\bar{X}$ = $\sigma^2|n$. In contrast to simple random sampling (SRS), RSS starts with taking a simple random sample of size k from the population and the k sampling units are ranked on the basis of personal judgment or a concomitant variable, say X, without actual measurement. Then the unit with rank 1 is identified and taken for the measurement and the remaining units of the sample are discarded. Next, another simple random sample of size k is drawn and the units of the sample are ranked by judgment, the unit with rank 2 is taken for the measurement and the remaining units are discarded. This process is continued until a simple random sample of size k is taken and ranked and the unit with rank k is taken for the measurement. This whole process is referred to as a cycle. The cycle then repeats m times to get a ranked set sample of size $n = km$ from the population of size $N = k^2m$.

When set size is very large, there may arise the situation where it is very difficult to assign proper individual rank to all the units by visual inspection or other rough-gauging methods. This means that for large set size, there may be more than one order statistics corresponding to each rank orders. Following the idea of Yanagawa and Shirahata (1976), this paper generalizes the ranked set sampling theory for large set size using the Probability Proportion to Rank Size Matrix (PPRSM).

In Section 2, we describe the procedures given by Yanagawa and Shirahata (1976), Yanagawa and Chen (1980) and Chen (1983) in brief. In Section 3, we introduce the idea and construction of Probability Proportion to Rank Size Matrix (PPRSM) and proposed an estimator of population mean μ. In Section 4, the properties of the proposed estimator have been discussed. Section 5 demonstrates the proposed design with the help of empirical study on two real data sets. The findings of the paper are concluded in Section 6.

2. A BRIEF DESCRIPTION OF VARIOUS PROCEDURES USING SELECTIVE PROBABILITY MATRIX/ VECTOR UNDER RSS

In this Section we shall describe in brief the procedures given by Yanagawa and Shirahata (YS-procedure (1976), Yanagawa and Chen (MG-procedure (1980) and Chen (Chen-procedure (1983)) which are as follows:

2.1 YS-Procedure

Select randomly mn elements from the population and split them into n sets each of which consists of m elements; rank the elements in each set according to the order of magnitude of the characteristic to be estimated; choose the j^{th} smallest element from the i^{th} set with probability p_{ij}, and measure the selected element for $i = 1, 2, ..., n$, and $j = 1, 2, ..., m$, where $\sum_{j=1}^{m} p_{ij} = 1$ for all i; estimate the population mean μ by means of the arithmetic mean of the measured values.

The $n \times m$ matrix **P** with elements p_{ij} such that $\sum_{j=1}^{m} p_{ij} = 1$, $i = 1, 2, ..., n$ is called the Selective Probability Matrix (SPM). The estimator which is given by specifying that the SPM be equal to an identically matrix when $m = n$ is called McIntyre, Takahasi and Wakimoto's (MTW) estimator.

Let X_i denote the measured value of the j^{th} smallest element from the i^{th} set with probability p_{ij}, for $i = 1$, 2, ... n, and $j = 1, 2, ..., m$, where $\sum_{j=1}^{m} p_{ij} = 1$ for all i. Then the estimate $(\bar{X}_{YS})$ of the population mean under YS-procedure is given by

$$\bar{X}_{YS} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

2.2 MG-Procedure

Select $n = 2l$ sets of elements each of size m from the population randomly, and arrange them in order of magnitude among the m elements in each set by visual inspection or other rough-gauging methods. Let I_1, I_2, ..., I_l be independent random variables such that I_i takes value j with probability p_{ij} for $i = 1, 2, ..., l$, and $j = 1$, 2, ..., m, where $\sum_{j=1}^{m} p_{ij} = 1$ for all i. If $I_1 = i_1$, then select the i_1^{th} smallest element from the first set and the $(m - i_1 + 1)^{\text{th}}$ smallest element from the n^{th} set, and measure them. If $I_2 = i_2$, then select the i_2^{th} smallest element and the $(m - i_2 + 1)^{th}$ smallest element respectively from the second and the $(n - 1)^{th}$ set, and measure them, and so on. Estimate the population mean μ by the arithmetic mean of the measured values. This estimator is equivalent to the MTW-estimator when $m = n$, $p_{ij} = 1$ and $p_{ij} = 0$ for $i \neq j$, for $i = 1, 2, ..., l$, and $j = 1, 2, ..., m$. Note that the SPM in the MG-procedure is given by $l \times m$ matrix $P = (p_{ij})$ such that $\sum_{j=1}^{m} p_{ij} = 1$ for $i = 1, 2, ..., l$.

Let $X_{i:j}$ denote the j^{th} smallest order statistic of a random sample of size m selected from the i^{th} set, and if $I_1, I_2, ..., I_l = (i_1, i_2, ..., i_l)$ is selected. Therefore the estimator $(\bar{X}_{MG})$ of population mean under this procedure is given by

$$\bar{X}_{MG} = \frac{1}{n}\sum_{j=1}^{l}\left(X_{j:i_j} + X_{n-j+1:m-i_j+1}\right).$$

2.3 Chen's-Procedure

In this procedure, we select elements from the population and split them into n sets each of size m; rank the elements in each set according to the order of magnitude of the characteristic to be estimated. Choose one element from the i^{th} ordered set, $(O_{i1}, O_{i2}, ... O_{im})$, $i = 1, 2, ..., n$, say $(O_{1k_1}, O_{2k_2}, ... O_{nk_n})$. This means that from the first set choose O_{1k_1}, from the second set choose O_{2k_2} and finally choose O_{nk_n} from the n^{th} set. Suppose $(O_{1k_1}, O_{2k_2}, ... O_{nk_n})$ is drawn with probability $p_{k_1k_2...k_n}$, such that $\sum_{k_1k_2....k_n=1}^{m} p_{k_1k_2....k_n} = P_{ij}$ the probability of drawing O_{ij}. Measure the value of the characteristic of element O_{ik_i}, of the i^{th} set for $i = 1, 2, ..., n$ and $k = 1, 2 ...., m$. Estimate the population mean μ by means of the arithmetic mean of the measured values. The probability vector $\mathsf{P} = \left\{p_{k_1k_2.......k_n} \middle| 1 \le k_1 \le m, ...1 \le k_n \le m\right\}$ with $\sum_{k_1k_2.......k_n=1}^{m} p_{k_1k_2...k_n}$ is called the Selective Probability Vector (SPV). This estimator is equivalent to the MTW-estimator if $m = n$ and $p_{k_1k_2...k_n} = 1$ if $k_i = i$ for all i, $p_{k_1k_2...k_n} = 0$ if $k_i \neq i$ for all i.

Let $X_{i:j_i}$ denote the measured value of the characteristic of element $O_{i:j_i}$ of the i^{th} group, then the estimator $(\bar{X}_C)$ of the population mean under this procedure is given by $\bar{X}_C = \frac{1}{n}\sum_{j=1}^{n} X_{i:k_i}$.

3. PROBABILITY PROPORTION TO RANK SIZE MATRIX (PPRSM)

Suppose a sample of size nm is drawn from the infinite population having mean μ and finite variance. Split them into n sets, each of size m. Rank the elements in each set according to the order of magnitude of the characteristic to be estimated by visual inspection or some other rough-gauging methods. For large m, many units in each set may get the same rank because it is quite difficult to distinguish them by the rough-gauging methods, even the measure values are same. For example, suppose we want to estimate the average height of the trees in an area which were planted under a scheme of Government/ other Agencies over a period of two years. It is clear that all the trees which were planted at the commencement of the scheme will have more or less same height. Moreover, those which were planted at the end of the second year will also have similar heights but different from the earlier ones.

Suppose m_i $(i = 1, 2, ..., n, 1 \le m_i \le m)$ is the number of order statistics in the i^{th} set and n_{ij} (i = 1, 2, ... n, $j = 1, 2, ... m_i$, $0 \le n_{ij} \le m$) is the number of units having the j^{th} order statistic in the i^{th} set. We can easily arrange these n_{ij}'s in the form of a matrix of order $n \times m$ we call it as Rank Size Matrix (RSM). Thus RSM, $\mathbf{R} = (n_{ij})_{n \times m}$ is

$$\mathbf{R} = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1j} & \dots & n_{1m} \\ n_{21} & n_{22} & \dots & n_{2j} & \dots & n_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ n_{i1} & n_{i2} & \dots & n_{ij} & \dots & n_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ n_{n1} & n_{n2} & \dots & n_{nj} & \dots & n_{nm} \end{bmatrix}_{n\times m} \qquad (1)$$

$$\sum_{j=1}^{m} n_{ij} = m, \; 0 \le n_{ij} \le m, \; i = 1, 2, ... n, \; j = 1, 2, ... m$$

Let p_{ij} denote the probability proportional to the number of elements n_{ij} in the $(i, j)^{th}$ cell of $\mathbf{R}$, *i.e.*

$p_{ij} = \dfrac{n_{ij}}{m}$. The matrix of p_{ij}'s becomes the Probability Proportion to Rank Size Matrix (PPRSM), $\mathbf{P} = (p_{ij})_{n\times m}$. This can be written as

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nj} & \cdots & p_{nm} \end{bmatrix}_{n\times m} \tag{2}$$

$p_{ij} = \dfrac{n_{ij}}{m}$, $\sum_{j=1}^{m} p_{ij} = 1$, 0 £ p_{ij} £ 1, i = 1, 2, ... n and j = 1, 2, ... m

The procedure for the construction of the estimator to be proposed is as follows:

(i) Select an order statistics from the i^{th} row with probability set $(p_{i1}\ p_{i2}\ \ldots\ p_{im_i})$, $1 \le m_i \le m$, i = 1, 2, ... n.

(ii) Measure the characteristic of the selected unit and denote it by X_i. Here X_i can be considered as a random variable and denote its cumulative distribution function (cdf) by $F(X)$.

The proposed estimator is,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3}$$

when n_{ij} = 1, for all i & j, and n = km for an integer k, this unbalanced ranked set sampling becomes balanced ranked set sampling. Further if $n = km$ for an integer k and X_i is selected for measurement with probability p_{ii} for each i, the estimator $\bar{X}$ equals the McIntyre and Takahasi and Wakimoto's (MTW) estimator.

4. PROPERTIES OF THE PROPOSED ESTIMATOR

It is clear from the above construction that the CDF of X_i is given by

$$P(X_i \le x) = \sum_{j=1}^{m_i} p_{ij} F_{m_i.j}(x) \tag{4}$$

where

$$dF_{m_i.j} = m_i \binom{m_i - 1}{j-1} F^{j-1}(x)\left[1 - F(x)\right]^{m_i - j} dF(x)$$

is the pdf of the j^{th} order statistic in the i^{th} set, j = 1, 2, ... m_i, i = 1, 2, ... n and $1 \le m_i \le m$. The following Theorem 1 gives the unbiasedness condition of the proposed estimator. In Theorem 2, the variance of the proposed estimator has been obtained. Theorem 3 shows the utility of our proposed estimator with respect to estimator under simple random sampling. In Theorem 4, we find the optimal RSM or corresponding PPRSM.

Theorem 1. The estimator of the population mean μ given in (3) is an unbiased estimator if and only if the Rank Size Matrix (RSM) **R** satisfies,

$$\sum_{i=1}^{n} n_{ij} = \frac{n^2 m}{\sum_{i=1}^{n} m_i}, \quad j = 1,\ 2,\ \ldots\ m_i \tag{5}$$

Proof: Let us denote by μ and $\mu_{m_i.j}$, the means of F and $F_{m_i.j}$ respectively. In general we have

$$\mu_{m_i.1} \le \mu_{m_i.2} \le \ldots \le \mu_{m_1.m_i}, \quad i = 1,\ 2,\ ..n \tag{6}$$

and

$$\mu = \frac{1}{\sum_{i=1}^{n} m_i}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \mu_{m_i.j} \tag{7}$$

Thus

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} p_{ij}\mu_{m_i.j}, \text{ therefore}$$

$$E(\bar{X}) - \mu = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} p_{ij}\mu_{m_i.j} - \frac{1}{\sum_{i=1}^{n} m_i}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mu_{m_i.j}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\frac{p_{ij}}{n} - \frac{1}{\sum_{i=1}^{n} m_i}\right)\mu_{m_i.j}$$

The necessary and sufficient condition for $E(\bar{X}) - \mu = 0$ for any CDF F is

$$\frac{p_{ij}}{n} = \frac{1}{\sum_{i=1}^{n} m_i}$$

Or

$$\sum_{i=1}^{n} p_{ij} = \frac{n^2}{\sum_{i=1}^{n} m_i}, \quad j = 1,\ 2,\ \dots m_i$$

Or

$$\sum_{i=1}^{n} n_{ij} = \frac{n^2 m}{\sum_{i=1}^{n} m_i}, \quad j = 1,\ 2,\ \dots m_i$$

which is the same as (5).

Theorem 2: If $\Re$ is the class of all RSM's **R** defined by

$$\Re = \left\{ \mathbf{R} : \mathbf{R} \text{ is a RSM such that } \sum_{i=1}^{n} n_{ij} = \frac{n^2 m}{\sum_{i=1}^{n} m_i} \text{ for each j.} \right\} \tag{8}$$

Then the variance of $\bar{X}$ when $\mathbf{R} \in \Re$ is given by

$$V\left(\bar{X}\middle|\mathbf{R}\right) = \frac{1}{n}\int x^2 dF(x) - \frac{1}{n^2 m^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)^2 \tag{9}$$

Proof:

$$Var\left(\bar{X}\middle|\mathbf{R}\right) = E(\bar{X})^2 - \left(E(\bar{X})\right)^2$$

$$= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right)^2$$

Since all X_i's are independent, therefore

$$Var\left(\bar{X}\middle|\mathbf{R}\right) = \frac{1}{n^2}\sum_{i=1}^{n} E\left(X_i^{\,2}\right) - \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} p_{ij}\mu_{m_i.j}\right)^2$$

$$= \frac{1}{n}\int x^2 dF(x) - \frac{1}{n^2 m^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)^2$$

which is the same as (9).

Theorem 3: If $Var(\bar{X}\,|\,\mathbf{R_R})$ denote the variance of the estimator $\bar{X}$ under simple random sampling, then for any $\mathbf{R} \in \Re$, $Var\left(\bar{X}\middle|\mathbf{R_R}\right) \ge V\left(\bar{X}\middle|\mathbf{R}\right)$, for any fixed F with finite variance, and

$$Var\left(\bar{X}\middle|\mathbf{R_R}\right) = V\left(\bar{X}\middle|\mathbf{R}\right) \text{iff } \mathbf{R} = \mathbf{R_R}\,.$$

Proof: For simple random sampling the RSM $\mathbf{R_R} = (n_{ij})_{n\times m}$ is

$$n_{ij} = \begin{cases} \dfrac{m}{m_i}; & i = 1,\ 2,\dots n,\ j = 1,\ 2,\dots m_i \\ 0\quad; & \text{otherwise} \end{cases} \tag{10}$$

We know for $\mathbf{R} \in \Re$

$$E_{\mathbf{R}}\left(X_i\right) = \sum_{j=1}^{m_i} p_{ij}\mu_{m_i.j} = \frac{1}{m}\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}$$

and $E_{\mathbf{R_R}}(X_i) = \mu$

Therefore we have from (9) and (10)

$$Var\left(\bar{X}\middle|\mathbf{R_R}\right) - Var\left(\bar{X}\middle|\mathbf{R}\right)$$

$$= \frac{1}{n}\int x^2 dF(x) - \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{m_i}\sum_{j=1}^{m_i}\mu_{m_i.j}\right)^2$$

$$-\left(\frac{1}{n}\int x^2 dF(x) - \frac{1}{n^2}\sum_{i=1}^{n} E_{\mathbf{R}}^{\,2}\left(X_i\right)\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E_{\mathbf{R}}^2\left(X_i\right) - \frac{1}{n^2}\sum_{i=1}^{n}\mu^2$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} E_{\mathbf{R}}^2\left(X_i\right) - n\mu^2\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left(E_{\mathbf{R}}\left(X_i\right) - \mu\right)^2 \ \ge 0$$

with equality if $E_{\mathbf{R}}(X_i) = \mu = E_{\mathbf{R_R}}(X_i)$,

i.e. $\mathbf{R} = \mathbf{R_R}$ or $n_{ij} = \dfrac{m}{m_i}$, $i = 1, 2, \dots n$

Theorem 4: The RSM that maximizes

$$\sum_{i=1}^{n}\sum_{\alpha=j}^{m_i}\sum_{\alpha'=j'}^{m_i} n_{i\alpha}n_{i\alpha'}, \tag{11}$$

for all j, $j' = 1, 2, \ldots m_{i}$, $i = 1, 2, \ldots n$ minimizes $Var\left(\bar{X}\middle|\mathbf{R}\right)$, $\mathbf{R} \in \Re$ for all cdf F.

The maximum value of (11) is given by

$$\frac{[d_j] + (d_j - [d_j])(d_{j'} - [d_{j'}])}{m^2}, \text{ if } [d_j] = [d_{j'}] \quad (12)$$

and

$$\frac{d_j}{m^2}, \text{ if } [d_j] < [d_{j'}] \quad (13)$$

where

$$d_j = \frac{n}{m_j}(m_j - j + 1) \quad (14)$$

and $[d_j]$ is the maximum integer not larger than d_j

The proof of Theorem 4 is based on the following Lemma 1.

Lemma 1: Assume that the real numbers A_i, B_i, $i = 1, 2, \ldots m$ satisfy

$$B_1 \le B_2 \le \ldots \le B_m \quad (15)$$

$$\sum_{i=1}^{n} A_i = 0 \quad (16)$$

and

$$\sum_{i=j}^{m} A_i \ge 0 \quad \text{for } j = 1, 2, \ldots m \quad (17)$$

Then $\sum_{i=1}^{m} A_i B_i \ge 0$

Proof of Lemma 1: From (17) we have $A_m \ge 0$ and $A_m + A_{m-1} \ge 0$. Hence from (15) and (16)

$$\sum_{i=1}^{m} A_i B_i = A_1 B_1 + \sum_{i=2}^{m} A_i B_i + \sum_{i=2}^{m} A_i B_1 - \sum_{i=2}^{m} A_i B_1$$

$$= \sum_{i=2}^{m} A_i(B_i - B_1) + B_1 \sum_{i=1}^{m} A_i = \sum_{i=2}^{m} A_i(B_i - B_1)$$

$$= \sum_{i=2}^{m-2} A_i(B_i - B_1) + A_{m-1}(B_{m-1} - B_1) + A_m(B_m - B_1)$$

$$\ge \sum_{i=2}^{m-2} A_i(B_i - B_1) + A_{m-1}(B_{m-1} - B_1) + A_m(B_{m-1} - B_1),$$

$$= \sum_{i=2}^{m-2} A_i(B_i - B_1) + (A_{m-1} + A_m)(B_{m-1} - B_1)$$

$$= \sum_{i=2}^{m-3} A_i(B_i - B_1) + A_{m-2}(B_{m-2} - B_1) + (A_{m-1} + A_m)(B_{m-1} - B_1)$$

$$\ge \sum_{i-2}^{m-.3} A_i(B_i - B_1) + A_{m-2}(B_{m-2} - B_1) + (A_{m-1} + A_m)(B_{m-2} - B_1)$$

$$= \sum_{i=2}^{m-3} A_i(B_i - B_1) + (A_{m-2} + A_{m-1} + A_m)(B_{m-2} - B_1)$$

A repetition of this process shows that

$$\sum_{i=1}^{m} A_i B_i \ge (A_2 + A_3 + \ldots . A_m)(B_2 - B_1) \ge 0$$

This completes the proof.

Proof of Theorem 4: From (9) it is clear that the problem of minimizing $Var(\bar{X} \mid \mathbf{R})$ implies that the problem of maximizing $\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)^2$. In order to find the optimal RSM which maximizes $\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)^2$ let us take two RSM **R** and **R***
defined as $\mathbf{R} = (n_{ij})_{n\times m}$ and $\mathbf{R}^* = (n_{ij}^*)_{n\times m}$. Let **R*** is the required optimal RSM, then in order to show

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}^*\mu_{m_i.j}\right)^2 - \sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)^2$$

$$= \sum_{i=1}^{n}\left[\left(\sum_{j=1}^{m_i} n_{ij}^*\mu_{m_i.j}\right)^2 - \left(\sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)^2\right]$$

$$= \sum_{i=1}^{n}\left(\sum_{j=1}^{m_i} n_{ij}^*\mu_{m_i.j} + \sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)\left(\sum_{j=1}^{m_i} n_{ij}^*\mu_{m_i.j} + \sum_{j=1}^{m_i} n_{ij}\mu_{m_i.j}\right)$$

$$= \sum_{i=1}^{m}\left(\sum_{j=1}^{m_i}\left(n_{ij}^* + n_{ij}\right)\mu_{m_i.j}\right)\left(\sum_{j'=1}^{m_i}\left(n_{ij'}^* - n_{ij'}\right)\mu_{m_i.j'}\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(n_{ij}^* + n_{ij}\right) \mu_{m_i . j} \sum_{j'=1}^{m_i} \left(n_{ij'}^* - n_{ij'}\right) \mu_{m_i . j'}$$

$$= \sum_{j=1}^{m_i} \mu_{m_i . j} \sum_{i=1}^{n} \sum_{j'=1}^{m_i} \left(n_{ij}^* n_{ij'}^* - n_{ij} n_{ij'}\right) \mu_{m_i . j'} \geq 0 \qquad (18)$$

In view of Lemma 1 we may show that

$$\sum_{j'=1}^{m_i} \sum_{\alpha=j}^{m_i} \sum_{i=1}^{n} \left(n_{ij}^* n_{ij'}^* - n_{i\alpha} n_{ij'}\right) \mu_{m_i . j'} \geq 0$$

For all, j = 1, 2, ... m_i, $\mu_{m_i . j'}$ satisfies (15) and $\sum_{i=1}^{n} \sum_{j'=1}^{m_i} \left(n_{ij}^* n_{ij'}^* - n_{ij} n_{ij'}\right) \mu_{m_i . j'}$ satisfies (16) as

$$\sum_{i=1}^{n} \sum_{j'=1}^{m_i} \left(n_{ij}^* n_{ij'}^* - n_{ij} n_{ij'}\right) \mu_{m_i . j'}$$

$$= \sum_{i=1}^{n} \left(n_{ij}^* \sum_{j'=1}^{m_i} n_{ij'} \mu_{m_i . j'} - n_{ij} \sum_{j'=1}^{m_i} n_{ij'} \mu_{m_i . j'} \right)$$

$$= \sum_{i=1}^{n} \left(n_{ij}^* m E_{\mathbf{R}} \left(X_i\right) - n_{ij} m E_{\mathbf{R}} \left(X_i\right) \right) = 0$$

To show (18), the same Lemma implies that we may show

$$\sum_{i=1}^{n} \sum_{\alpha=j}^{m_i} \sum_{\alpha'=j'}^{m_i} n_{i\alpha}^* n_{i\alpha'}^* - n_{i\alpha} n_{i\alpha'} \geq 0 \text{ for all } j, j' = 1, 2, \ldots m_i.$$

This fact implies that the RSM, **R*** maximizes (11) and hence is optimal RSM. The maximum value of this RSM can be found by the softwares on Operation Research, as suggested by Yanagawa and Shirahata (1976). Details of these methods are omitted for brevity.

5. EMPIRICAL STUDIES

In what follows, we consider two empirical studies to demonstrate the utility of the proposed procedure.

1. For the purpose of comparing the proposed estimators, an empirical study on Davison (2005) was carried out wherein a part the data (Table 1a) of girth (inches) and height (feet) of 31 trees of an experiment on 'volume from measurements of girth and height for future trees' is taken for the estimation of population mean. Girth is the tree

Table 1a. Observations on girth and height

S.No.	Girth	Height
1	8.3	70
2	8.6	65
3	8.8	63
4	10.5	72
5	10.7	81
6	10.8	83
7	11.0	66
8	11.0	75
9	11.1	80
10	11.2	75
11	11.3	79
12	11.4	76
13	11.4	76
14	11.7	69
15	12.0	75
16	12.9	74
17	12.9	85
18	13.3	86
19	13.7	71
20	13.8	64
21	14.0	78
22	14.2	80
23	14.5	74
24	16.0	72
25	16.3	77
26	17.3	81
27	17.5	82
28	17.9	80
29	18.0	80
30	18.0	80
31	20.6	87

Table 1b. Ranked set sample of set size 5 on girth

8.3	11	11	17.3	18.0
8.6	10.7	11.1	11.7	14.2
10.5	11.3	11.7	16.0	16.3
8.8	11.4	14.0	14.2	18.0
10.7	11.3	11.4	11.7	20.6

Table 1c. Ranked set sample of set size 5 on height

66	71	72	80	80
65	74	80	80	82
63	72	75	75	81
65	74	80	80	84
65	76	76	80	87

Table 1d. RSMs **R**(*G*) & **R**(*H*) for girth and height respectively

$$\mathbf{R}(G) = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 3 & 1 & 0 & 0 \end{bmatrix}_{5\times5} \qquad \mathbf{R}(H) = \begin{bmatrix} 1 & 1 & 1 & 2 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 2 & 1 & 1 & 0 \end{bmatrix}_{5\times5}$$

$$\sum_{j=1}^{5} n_{ij} = 5,\ 0 \le n_{ij} \le 5,\ i = 1, 2, \ldots 5,\ j = 1, 2, \ldots 5$$

Table 1e. PPRSMs, **P**(*G*) and **P**(*H*) for girth and height respectively

$$\mathbf{P}(\mathrm{G}) = \begin{bmatrix} 1/5 & 2/5 & 1/5 & 1/5 & 0 \\ 1/5 & 1/5 & 2/5 & 1/5 & 0 \\ 1/5 & 2/5 & 2/5 & 1/5 & 0 \\ 1/5 & 1/5 & 2/5 & 1/5 & 0 \\ 1/5 & 3/5 & 1/5 & 0 & 0 \end{bmatrix}_{5\times5} \qquad \mathbf{P}(\mathrm{H}) = \begin{bmatrix} 1/5 & 2/5 & 1/5 & 1/5 & 0 \\ 1/5 & 1/5 & 2/5 & 2/5 & 0 \\ 1/5 & 1/5 & 2/5 & 1/5 & 0 \\ 1/5 & 1/5 & 2/5 & 1/5 & 0 \\ 1/5 & 2/5 & 1/5 & 1/5 & 0 \end{bmatrix}_{5\times5}$$

$$p_{ij} = \frac{n_{ij}}{5},\ \sum_{j=1}^{5} p_{ij} = 1,\ 0 \le p_{ij} \le 1,\ \iota = 1, 2, \ldots 5,\ j = 1, 2 \ldots 5$$

Table 1f. RSM and PPRSM under SRS

$$\mathbf{R}_R = \begin{bmatrix} 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \end{bmatrix}_{5\times5} \qquad \mathbf{P}_P \begin{bmatrix} 1/5\ 1/5\ 1/5\ 1/5\ 1/5 \\ 1/5\ 1/5\ 1/5\ 1/5\ 1/5 \\ 1/5\ 1/5\ 1/5\ 1/5\ 1/5 \\ 1/5\ 1/5\ 1/5\ 1/5\ 1/5 \\ 1/5\ 1/5\ 1/5\ 1/5\ 1/5 \end{bmatrix}_{5\times5}$$

Table 1g. Relative precision based on PPRSM over an estimator based on SRS

Parameters	Variance of estimators using PPRSM	Variance of estimators using SRS	Relative Precision
Girth	0.45	5.12	11.38
Height	2.63	42.42	16.13

diameter measured at 4 ft 6 inch above the ground. The Ranked Set Sample of set size 5 for girth and height parameters is arranged into 5 × 5 matrices (Table 1b and Table 1c). After ignoring the fractional values the RSM and PPRSM for both the parameters, the values are presented in Table 1d and Table 1e. Table 1f gives the same matrices for the case of SRS. The calculated value of relative efficiency of PPRSM based estimator over SRS based estimator for both the parameters are shown in Table 1g.

From Table 1g, it is clear that the relative precision of the proposed procedure is quite high for both the parameters (girth and height) of the trees considered under the study.

2. We borrow another data set from Parsad *et* al. (2010) on biometrical character, average number of green leaves of hybrid Jowar crops to compare the performance of RSS and SRS based estimators for set sizes 10 and 15. Here we have taken the absolute values of the data while making RSM. Table 2a and Table 2b show the RSM and PPRSM for set sizes 10 and 15 respectively. In Table 2c we have shown the efficiency gain of PPRSM based estimator over SRS based estimator for set sizes 10 and 15.

Table 2a. Ranked set sample, RSM and PPRSM for set size 10

6.4	6.4	6.8	6.9	6.9	7.5	7.7	8.8	10.6	10.6
4.8	4.9	5.5	6.8	7.7	8.4	9.3	9.5	9.8	11.8
5.6	5.9	7.4	8.2	9.3	9.4	11.6	11.6	11.8	11.8
6.4	7.2	7.4	8.4	8.4	8.8	8.8	9.5	9.5	11.6
4.8	5.7	5.9	8.1	9.4	9.5	10.5	11.5	11.6	11.8
6.4	7.2	7.5	7.5	8.4	9.3	9.4	9.7	9.9	10.6
6.9	7.4	8.2	8.2	8.4	8.4	8.8	9.9	10.0	10.6
4.8	6.4	6.8	7.4	7.7	8.4	9.3	9.3	9.4	9.5
4.8	5.0	5.9	6.4	7.5	7.7	7.7	9.6	10.0	11.6
5.6	5.6	6.4	6.4	6.8	6.8	6.8	6.9	7.5	8.2
2	3	2	1	2	0	0	0	0	0
2	1	1	1	2	2	1	0	0	0
2	1	1	2	4	0	0	0	0	0
1	2	4	2	1	0	0	0	0	0
1	2	1	2	1	1	2	0	0	0
1	1	3	2	2	1	0	0	0	0
2	4	1	2	1	0	0	0	0	0
1	1	2	2	3	1	0	0	0	0
2	2	3	2	1	0	0	0	0	0
2	2	4	2	0	0	0	0	0	0
0.2	0.3	0.2	0.1	0.2	0	0	0	0	0
0.2	0.1	0.1	0.1	0.2	0.2	0.1	0	0	0
0.2	0.1	0.1	0.2	0.4	0	0	0	0	0
0.1	0.2	0.4	0.2	0.1	0	0	0	0	0
0.1	0.2	0.1	0.2	0.1	0.1	0.2	0	0	0
0.1	0.1	0.3	0.2	0.2	0.1	0	0	0	0
0.2	0.4	0.1	0.2	0.1	0	0	0	0	0
0.1	0.1	0.2	0.2	0.3	0.1	0	0	0	0
0.2	0.2	0.3	0.2	0.1	0	0	0	0	0
0.2	0.2	0.4	0.2	0	0	0	0	0	0

Table 2b: Ranked set sample, RSM and PPRSM for set size 15

5.7	5.9	6.4	6.8	7.4	7.5	7.7	8.2	8.2	8.3	8.4	9.3	9.3	9.5	11.6
5.0	6.4	6.4	6.8	7.4	7.4	8.2	8.2	9.5	9.6	10.6	10.6	10.6	11.6	11.8
4.8	5.5	7.3	7.7	8.1	8.4	9.3	9.8	9.8	9.8	9.8	9.9	9.9	11.6	11.6
4.8	5.0	6.4	7.5	7.5	7.5	8.8	9.7	9.7	9.8	9.8	10.0	10.6	11.6	11.8
5.7	5.9	5.9	6.4	7.5	7.7	8.1	8.2	8.3	8.3	8.4	8.4	9.3	10.2	11.6
4.8	4.8	5.7	5.9	6.4	6.4	7.3	7.3	7.7	8.1	8.4	9.3	9.8	9.8	11.6
5.6	5.7	6.4	7.2	7.4	4.5	7.7	8.1	8.2	8.8	9.3	9.4	9.8	10.6	11.8
5.5	5.7	6.4	6.4	6.8	7.4	7.4	8.2	8.3	9.3	9.3	9.6	9.9	10.6	10.6
6.4	6.4	6.4	6.9	7.3	7.3	7.5	7.5	7.7	7.7	8.4	8.4	9.8	9.8	11.6
7.3	7.4	7.5	7.7	7.7	7.7	8.2	8.2	8.4	8.4	9.6	9.7	9.8	9.8	11.6
5.0	5.0	5.6	6.4	6.8	6.9	7.2	7.5	7.7	8.1	8.3	9.4	9.9	9.9	10.0
5.0	5.0	5.5	6.4	6.4	7.3	7.4	7.7	7.7	8.4	8.4	8.8	9.3	9.4	9.8
6.4	6.8	6.8	6.9	6.9	7.2	7.4	7.7	8.2	8.8	8.8	9.4	9.6	10.2	10.6
5.0	5.5	5.9	5.9	6.4	6.8	7.7	8.4	8.4	8.4	9.6	9.8	10.2	11.6	11.8
5.0	5.0	5.6	5.6	6.4	6.4	6.8	7.2	7.2	7.4	9.4	9.8	9.8	10.6	11.8
3	2	6	2	1	1	0	0	0	0	0	0	0	0	0
1	2	3	2	2	3	2	0	0	0	0	0	0	0	0
1	1	1	3	1	6	2	0	0	0	0	0	0	0	0
2	1	3	1	4	1	1	2	0	0	0	0	0	0	0
4	8	1	1	1	0	0	0	0	0	0	0	0	0	0
2	4	2	3	1	2	1	0	0	0	0	0	0	0	0
3	2	4	3	1	1	1	0	0	0	0	0	0	0	0
4	3	2	2	2	2	0	0	0	0	0	0	0	0	0
3	3	6	2	1	0	0	0	0	0	0	0	0	0	0
2	8	4	1	0	0	0	0	0	0	0	0	0	0	0
2	2	3	4	1	3	0	0	0	0	0	0	0	0	0
2	3	2	4	3	1	0	0	0	0	0	0	0	0	0
1	6	2	2	2	1	0	0	0	0	0	0	0	0	0
1	4	1	3	3	2	0	0	0	0	0	0	0	0	0
2	4	4	1	2	1	1	0	0	0	0	0	0	0	0
3/15	2/15	6/15	2/15	1/15	1/15	0	0	0	0	0	0	0	0	0
1/15	2/15	3/15	2/15	2/15	3/15	2/15	0	0	0	0	0	0	0	0
1/15	1/15	1/15	3/15	1/15	6/15	2/15	0	0	0	0	0	0	0	0
2/15	1/15	3/15	1/15	4/15	1/15	1/15	2/15	0	0	0	0	0	0	0
4/15	8/15	1/15	1/15	1/15	0	0	0	0	0	0	0	0	0	0
2/15	4/15	2/15	3/15	1/15	2/15	1/15	0	0	0	0	0	0	0	0
3/15	2/15	4/15	3/15	1/15	1/15	1/15	0	0	0	0	0	0	0	0
4/15	3/15	2/15	2/15	2/15	2/15	0	0	0	0	0	0	0	0	0
3/15	3/15	6/15	2/12	1/15	0	0	0	0	0	0	0	0	0	0
2/15	8/15	4/15	1/15	0	0	0	0	0	0	0	0	0	0	0
2/15	2/15	3/15	4/15	1/15	3/15	0	0	0	0	0	0	0	0	0
2/15	3/15	2/15	4/15	3/15	1/15	0	0	0	0	0	0	0	0	0
1/15	6/15	2/15	2/15	2/15	1/15	0	0	0	0	0	0	0	0	0
1/15	4/15	1/15	3/15	3/15	2/15	0	0	0	0	0	0	0	0	0
2/15	4/15	4/15	1/15	2/15	1/15	1/15	0	0	0	0	0	0	0	0

Table 2c: Efficiency gain based on PPRSM over an estimator based on SRS

Set size	Variance of estimators using PPRSM	Variance of estimators using SRS	Relative Precision
10	0.23	4.47	19.43
15	2.75	61.26	22.28

6. CONCLUSIONS

When the measurements of units are very costly and time consuming and there is heterogeneity between the units of the population, the SRS become useless. In such situations, RSS is a cost-effective and precise method of sample selection. Furthermore, if the set size used in the experiment is very large then the technique of construction of PPRSM is appropriate for estimating the population mean. In this discussion, the proposed design is more efficient than SRS for estimating mean of the population, under the assumption that ranking of sampling units are easier than actual measurements. It also contains the information about all order statistics. We have demonstrated theoretically as well as with the help of two empirical examples that the proposed procedure is more advantageous in comparison to existing procedures of RSS for large set size in the sense that it provides unbiased estimators as well as it give equal importance to all the rank orders.

ACKNOWLEDGEMENTS

The authors are thankful to the reviewer and Dr. Rajender Parsad, Coordinating Editor for their valuable and constructive comments, which led to considerable improvement in presentation of this work.

REFERENCES

Chen, S.H. (1983). Ranked set sampling theory with selective probability vector. *J. Statist. Plann. Inf.,* **8**, 161-174.

Davison, A.C. (2005). Practical Report on Cherry Tree Data. *Chair of Statistics Stat.h*ttp://stat.epfl.ch/webdav/site/stat/shared/Regression/trees.pdf

McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets. *Austr. J. Agric. Res.,* **3,** 385-390.

Parsad, R., Gupta, V.K., Umarao A. and Kole, B. (2010). Analysis of Data: Design Resources Server. Indian Agricultural Statistics Research Institute (ICAR), New Delhi 110 012, India. www.iasri.res.in/design (accessed lastly on <May 21, 2010>).

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.

Yanagawa, T. and Chen, S. (1980). The MG-procedure in ranked set sampling. *J. Statist. Plann. Inf.,* **4,** 33-44.

Yanagawa, T. and Shirahata, S. (1976). Ranked set sampling theory with selective probability matrix. *Austr. J. Agric. Res.,* **18,** 45-52.