Hacettepe Journal of Mathematics and Statistics
Volume 44 (6) (2015), 1513 – 1525

Improved estimation from ranked set sampling

Raghunath Arnab *† and J. O. Olaomi‡

Abstract

Ranked set sampling is used when the measurement or quantification of units of the variable under study is difficult but the ranking of units of sets of small sizes can be done easily by an inexpensive method. Dell and Clutter (1972) showed that the sample mean based on ranked set sample is more efficient than the sample mean based on simple random sample with replacement sampling procedure for estimation of the population mean. In this paper Dell and Clutter estimator has been improved further by using the ranking variable x as an auxiliary variable when μ_x, the population mean of x is unknown. An empirical investigation based on life data shows all proposed estimators are approximately unbiased and bring gain in efficiency of up to 50 percent.

Keywords: Order statistics, ranked set sampling, relative efficiency, sampling with replacement.

2000 AMS Classification: 62D05

Received : 30.07.2014 *Accepted* : 19.11.2014 *Doi* : 10.15672/HJMS.2014438213

1. Introduction

Ranked set sampling (RSS) was introduced by McIntyre (1952) to estimate the mean pasture and forage yield. The RSS is used when precise measurement of the variable of interest is difficult or expensive but the variable can be ranked easily without measuring the actual variable by an inexpensive method such as visual perception, judgment and auxiliary information. For example, in estimating the mean height of trees in a forest, the heights of a small sample of two or three trees standing nearby can be ranked easily by visual inspection without measuring them. In estimating the number of bacterial cells per unit volume, we can rearrange two or three test tubes easily in order of concentration using optical instruments without measuring exact values. In a ranked set sampling, instead of selecting a single sample of size m, we select m-sets of samples each of size m. In

*Department of Statistics, University of Botswana, Department of Statistics, University of South Africa and University of Kwa-Zulu Natal, Email: arnabr@mopipi.ub.bw
†Corresponding Author.
‡Department of Statistics, University of South Africa, Email: olaomjo@unisa.ac.za

each of the sets all the elements are ranked but only one is measured. Finally, an average of the m-measured units is taken as an estimate of the population mean. Dell and Clutter (1972) proved that the sample mean based on the RSS is unbiased for the population mean regardless of the errors of ranking. The RSS mean is at least as precise as the sample mean of the simple random sampling with replacement (SRSWR) sampling scheme of the same size. Stokes (1980, 1980a, 1988) showed that RSS provides precise estimators for cumulative distribution function, population variance and correlation coefficient.

1.1. Rank set sampling by SRSWR method. First we choose a small number m (set-size) such that one can easily rank the m elements of the population with sufficient accuracy. Then the selection of RSS is as follows: Select a sample of m^2 units from a population U by SRSWR method. Allocate these m^2 units at random into m sets each of size m. Rank all the units in a set with respect to the values of the variable of interest y from 1 (minimum) to m (maximum) by a very inexpensive method such as eye inspection. No actual measurement is done at this stage. After the ranking has been completed, the unit holding rank 1 of the set-1, unit holding rank-2 of the set $2, \dots,$ and finally the unit holding rank m of the set m is measured accurately by using a suitable instrument. This completes a cycle of the sampling. The process is repeated for r cycles to obtain the desired sample of size $n = mr$ units. Thus in a RSS a total of m^2r units have been drawn from the population but only mr of them are measured and the rest $mr\,(m-1)$ are discarded. These measured mr observations are called "ranked set sample". Since the ordering of a large number of observations is difficult, increase of sample size $n = mr$ is done by increasing the number of cycles r.

It is obvious that the variable used for ranking x (say) e.g. eye estimation, judgment or auxiliary information is expected to have high correlation with the variable of interest y. Stokes (1977) considered ranking as an auxiliary variable. Prasad (1989), Kadilar et al. (2009) and Singh et al. (2014) used the estimation of the population mean μ_y assuming the population mean μ_x is known. In our present paper we have proposed improved methods of estimation of the population mean using the ranking variable as an auxiliary variable when the population mean μ_x is unknown. The proposed estimators fare better than the traditional estimator-sample mean. We also compared the performances of the proposed estimators through simulation studies based on live data collected by Platt et al. (1988), given by Chen et al. (2003). The simulation revealed that all the proposed estimators are approximately unbiased and bring gain in efficiency of up to 50%.

1.2. A fundamental equality. Let $y_{i1|k}, \dots, y_{ij|k}, \dots, y_{im|k}$ and $x_{i1|k}, \dots, x_{ij|k}, \dots, x_{im|k}$ be the value of the variable of interest y and x of the ith set of elements of the kth cycle, $i = 1, \dots, m;\quad k = 1, \dots, r$. Further, let $y_{i(j)|k}$ and $x_{i(j)|k}$ be the smallest jth observation (order statistic) of $y_{i1|k}, \dots, y_{ij|k}, \dots, y_{im|k}$ and $x_{i1|k}, \dots, x_{ij|k}, \dots, x_{im|k}$ respectively. Here we first assume that y increases with x i.e. $x_{ij|k} > x_{i'j'|k'}$ implies $y_{ij|k} > y_{i'j'|k'}$. Ranking of heights of two and three trees nearby through visual inspection, the eye estimates (x) is expected to provide perfect ranking. Obviously, perfect ranking is not always possible. So, the theory of judgement ranking has been introduced in section 2.6. Let $y_{i1|k}, \dots, y_{ij|k}, \dots, y_{im|k}$ be a random sample from a population with cumulative distribution function (cdf) $F(y)$ and probability density function (pdf) $f(y)$. Similarly $x_{i1|k}, \dots, x_{ij|k}, \dots, x_{im|k}$ are the random sample from a population with cdf $F(x)$and pdf $f(x)$ respectively. Let the mean and variance of x and y be $\mu_x,\ \mu_y$ and

$\sigma_x^2,\ \sigma_y^2$ respectively. Then we have the following equalities following Stokes (1980):

$$\sum_{j=1}^{m} y_{ij|k} = \sum_{j=1}^{m} y_{i(j)|k}, \qquad \sum_{j=1}^{m} x_{ij|k} = \sum_{j=1}^{m} x_{i(j)|k} \tag{1.1}$$

$$\begin{aligned} \sum_{j=1}^{m}(x_{ij|k}-\mu_x)^2 &= \sum_{j=1}^{m},(x_{i(j)|k}-\mu_x)^2, \\ \sum_{j=1}^{m}(y_{ij|k}-\mu_y)^2 &= \sum_{j=1}^{m}(y_{i(j)|k}-\mu_y)^2 \end{aligned} \tag{1.2}$$

and

$$\sum_{j=1}^{m}(x_{ij|k}-\mu_x)(y_{ij|k}-\mu_y) = \sum_{j=1}^{m}(x_{i(j)|k}-\mu_x)(y_{i(j)|k}-\mu_y) \tag{1.3}$$

Let $\mu_{x(j)|m} = E\{x_{i(j)|k}\}$ and $\mu_{y(j)|m} = E\{y_{i(j)|k}\}$ be the mean of the jth order-statistic of random samples of size m of the variables x and y for the cycle k. The order statistics $\mu_{x(j)|m}$ and $\mu_{y(j)|m}$ depend on m but is independent of the set i and the cycle k.

The equation (1.1) yields

$$\begin{aligned} E\left\{\frac{1}{m}\sum_{j=1}^{m} x_{ij|k}\right\} &= E\left\{\frac{1}{m}\sum_{j=1}^{m} x_{i(j)|k}\right\} \\ \text{i.e.} \quad \mu_x &= \frac{1}{m}\sum_{j=1}^{m}\mu_{x(j)|m} \end{aligned} \tag{1.4}$$

Similarly,

$$\mu_y = \frac{1}{m}\sum_{j=1}^{m}\mu_{y(j)|m} \tag{1.5}$$

the equation (1.2) yields

$$\begin{aligned} \sum_{j=1}^{m} E(x_{ij|k}-\mu_x)^2 &= \sum_{j=1}^{m} E(x_{i(j)|k}-\mu_x)^2 \\ \text{i.e.} \quad m\sigma_x^2 &= \sum_{j=1}^{m}\left\{\sigma_{x(j)|m}^2 + (\mu_{x(j)|m}-\mu_x)^2\right\} \\ (\text{where } \sigma_{x(j)|m}^2 &= \text{variance of } x_{i(j)|m}) \\ \text{i.e.} \quad \sigma_x^2 &= \frac{1}{m}\sum_{j=1}^{m}\sigma_{x(j)|m}^2 + \frac{1}{m}\sum_{j=1}^{m}(\mu_{x(j)|m}-\mu_x)^2 \end{aligned} \tag{1.6}$$

Similarly,

$$\sigma_y^2 = \frac{1}{m}\sum_{j=1}^{m}\sigma_{y(j)|m}^2 + \frac{1}{m}\sum_{j=1}^{m}(\mu_{y(j)|m}-\mu_y)^2 \tag{1.7}$$

$$(\text{where } \sigma_{y(j)|m}^2 = \text{variance of } y_{i(j)|m})$$

Let us assume that the variables x and y from the same unit are correlated while from the different units are uncorrelated so that

$$Cov(x_{ij|k}, y_{ij|k}) = \mu_{xy} \quad \text{and} \quad Cov(x_{ij|k}, y_{i'j'|k'}) = 0 \quad \text{for} \quad (i,j,k) \neq (i',j',k') \tag{1.8}$$

1.3. Estimation of the mean. Let $\bar{y}_{[m]|k} = \frac{1}{m}\sum_{i=1}^{m} y_{i(i)|k}$ = arithmetic mean of the m quantified values of the variable y for the cycle k and

$$\hat{\mu}_{y(rss)} = \frac{1}{r}\sum_{k=1}^{r} \bar{y}_{[m]|k} = \frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i(i)|k} \tag{1.9}$$

is the mean of $n = mr$ quantified variables based on all the r cycles. The following theorem due to Dell and Clutter (1972)and Kaur et al. (1997) show that the estimator $\hat{\mu}_{y(rss)}$ is unbiased for μ_y and possesses a lower variance than $\hat{\mu}_{y(srs)}$, the sample mean based on an SRSWR sample of the same size n. An unbiased estimator of the variance is also presented here.

1.1. Theorem.

(i) $E(\hat{\mu}_{y(rss)}) = \mu_y$

(ii)
$$\begin{aligned} V(\hat{\mu}_{y(rss)}) &= \frac{\sigma^2_{y[m]}}{n} \\ &= \frac{1}{n}\left[\sigma_y^2 - \frac{1}{m}\sum_{i=1}^{m}(\mu_{y(j)|m} - \mu_y)^2\right] \\ &\leq \sigma_y^2/n = V(\hat{\mu}_{y(srs)}) \end{aligned}$$

(iii) *An unbiased estimator of the variance of* $V(\hat{\mu}_{y(rss)})$ *is*

$$\hat{V}(\hat{\mu}_{y(rss)}) = \frac{1}{r(r-1)}\sum_{k=1}^{r}(\bar{y}_{[m]|k} - \hat{\mu}_{y(rss)})^2.$$

where $\sigma^2_{y[m]} = \frac{1}{m}\sum_{j=1}^{m}\sigma^2_{y(j)|m}$

1.4. Precision of the rank-set sampling. The relative precision of $\hat{\mu}_{y(rss)}$ compared to $\hat{\mu}_{y(srs)}$, sample mean of an SRSWR sample of size $n = mr$ is

$$RP_{rss/srs} = \frac{V(\hat{\mu}_{y(srs)})}{V(\hat{\mu}_{y(rss)})} = \frac{\sigma_y^2}{\sigma^2_{y[m]}} \tag{1.10}$$

2. Proposed estimator of the population mean

From the ith set of the kth cycle, we construct an estimator for μ_y as follows:

$$\begin{aligned} t_{i|k} &= y_{i(i)|k} - \lambda x_{i(i)|k} + \lambda \bar{x}_{i|k} \quad \text{for} \quad i = 1,..,m \\ &= y_{i(i)|k} - \lambda(x_{i(i)|k} - \bar{x}_{i|k}) \end{aligned} \tag{2.1}$$

where $\bar{x}_{i|k} = \frac{1}{m}\sum_{j=1}^{m} x_{ij|k}$ and λ is a suitably chosen constant to be determined optimally.

The proposed estimator of the population mean μ_y based on the kth cycle is

$$\begin{aligned} t_k &= \frac{1}{m}\sum_{i=1}^{m} t_{i|k} \\ &= \left(\frac{1}{m}\sum_{i=1}^{m} y_{i(i)|k}\right) - \lambda\left(\frac{1}{m}\sum_{i=1}^{m} x_{i(i)|k} - \frac{1}{m}\sum_{i=1}^{m}\bar{x}_{i|k}\right) \end{aligned} \tag{2.2}$$

and the overall estimator for μ_y is

$$\bar{t} = \frac{1}{r}\sum_{k=1}^{r} t_k \tag{2.3}$$

2.1. Mean and variance of $\bar{t}$.

$$\begin{aligned} E(t_{i|k}) &= E(y_{i(i)|k}) - \lambda E(x_{i(i)|k} - \bar{x}_{i|k}) \\ &= \mu_{y(i)|k} - \lambda(\mu_{x(i)|k} - \mu_x) \\ &= \mu_{y(i)} - \lambda(\mu_{x(i)} - \mu_x) \end{aligned} \tag{2.4}$$

(noting $\mu_{y(i)|k} = \mu_{y(i)}$ for every k)
Now using (2.2), we get

$$\begin{aligned} E(t_k) &= \frac{1}{m}\sum_{i=1}^{m}[(\mu_{y(i)} - \lambda\mu_{x(i)}) + \lambda\mu_x] \\ &= \mu_d + \lambda\mu_x \quad (\text{where } \mu_d = \mu_y - \lambda\mu_x) \\ &= \mu_y \end{aligned} \tag{2.5}$$

The variance of t_k is

$$\begin{aligned} V(t_k) &= V\left(\frac{1}{m}\sum_{i=1}^{m} t_{i|k}\right) \\ &= \frac{1}{m^2}\sum_{i=1}^{m} V(t_{i|k}) \end{aligned} \tag{2.6}$$

Now

$$V(t_{i|k}) = V(y_{i(i)|k}) + \lambda^2 V(x_{i(i)|k} - \bar{x}_{i|k}) - 2\lambda Cov(y_{i(i)|k}, x_{i(i)|k} - \bar{x}_{i|k}) \tag{2.7}$$

Further,

$$\begin{aligned} V(x_{i(i)|k} - \bar{x}_{i|k}) &= V(x_{i(i)|k}) + V\left(\bar{x}_{i|k}\right) - 2Cov(x_{i(i)|k}, \bar{x}_{i|k}) \\ &= \sigma^2_{x(i)} + \frac{\sigma^2_x}{m} - \frac{2}{m}\left[V(x_{i(i)|k}) + \sum_{j(\neq i)} Cov(x_{i(i)k}, x_{i(j)|k})\right] \end{aligned} \tag{2.8}$$

$$\begin{aligned} Cov(y_{i(i)|k}, x_{i(i)|k} - \bar{x}_{i|k}) &= Cov(y_{i(i)|k}, x_{i(i)|k}) - Cov\left(y_{i(i)|k}, \frac{1}{m}\sum_{j=1}^{m} x_{i(j)||k}\right) \\ &= \sigma_{xy(i)|k} - \frac{1}{m}\left[Cov(x_{i(i)|k}, y_{i(i)|k}) + \frac{1}{m}\sum_{j(\neq 1)}^{m} Cov(y_{i(i)|k}, x_{i(j)|k})\right] \end{aligned} \tag{2.9}$$

where $\sigma_{xy(i)}$ is the covariance between $x_{i(i)k}$ and $y_{i(i)k}$.
Now substituting(2.8) and (2.9) in (2.7), we get

$$\begin{aligned} V(t_{i|k}) = \ & \sigma^2_{y(i)} + \lambda^2\left[\sigma^2_{x(i)} + \frac{\sigma^2}{m} - \frac{2}{m}\left\{V(x_{i(i)}) + \sum_{j(\neq i)} Cov(x_{i(i)}, x_{i(j)})\right\}\right] \\ & -2\lambda\left[\sigma_{xy(i)} - \frac{1}{m}\left\{Cov(x_{i(i)}, y_{i(i)}) + \frac{1}{m}\sum_{j(\neq i)}^{m} Cov(y_{i(i)}, x_{i(j)})\right\}\right] \end{aligned}$$

The equation (2.6) yields

$$\begin{aligned} & V(t_k) \\ = & \frac{1}{m^2}\sum_{i=1}^{m}\sigma^2_{y(i)} + \frac{\lambda^2}{m^2}\left[\sum_{i=1}^{m}\sigma^2_{x(i)} + \sigma^2_x - \frac{2}{m}\right. \\ & \left.\left\{\sum_{i=1}^{m} V(x_{i(i)|k}) + \sum_{i=1}^{m}\sum_{j(\neq i)} Cov(x_{i(i)k}, x_{i(j)|k})\right\}\right] \\ & -2\frac{\lambda}{m^2}\left[\sum_{i=1}^{m}\sigma_{xy(i)} - \frac{1}{m}\right. \\ & \left.\left\{\sum_{i=1}^{m} Cov(x_{i(i)|k}, y_{i(i)k}) + \frac{1}{m}\sum_{i=1}^{m}\sum_{j(\neq i)}^{m} Cov(y_{i(i)|k}, x_{i(j)|k})\right\}\right] \\ = & \frac{1}{m^2}\sum_{i=1}^{m}\sigma^2_{yi} + \frac{\lambda^2}{m^2}\left[\sum_{i=1}^{m}\sigma^2_{x(i)} - \sigma^2_x\right] - 2\frac{\lambda}{m^2}\left[\sum_{i=1}^{m}\sigma_{xy(i)} - \sigma_{xy}\right] \end{aligned} \tag{2.10}$$

Further, the equation (2.3) yields the variance of $\bar{t}$ as

$$\begin{aligned} & V(\bar{t}) \\ = & \frac{1}{r^2}\sum_{k=1}^{r} V(t_k) \\ = & \frac{1}{rm^2}\left[\sum_{i=1}^{m}\sigma^2_{y(i)} + \lambda^2\left(\sum_{i=1}^{m}\sigma^2_{x(i)} - \sigma^2_x\right) - 2\lambda\left(\sum_{i=1}^{m}\sigma_{xy(i)} - \sigma_{xy}\right)\right] && (2.11) \\ = & \frac{1}{n}\left[\frac{1}{m}\sum_{i=1}^{m}\sigma^2_{y(i)} + \lambda^2\frac{1}{m}\sum_{i=1}^{m}\sigma^2_{x(i)} - 2\frac{\lambda}{m}\sum_{i=1}^{m}\sigma_{xy(i)}\right] && (2.12) \\ & -\frac{1}{nm}\left[\lambda^2\sigma^2_x - 2\lambda\sigma_{xy}\right] \end{aligned}$$

(noting $n = rm$)

Now using (1.6), (1.7) and (1.8),we get

$$\begin{aligned} V(\bar{t}) = & \frac{1}{n}\left[\left(\sigma^2_y - \frac{1}{m}\sum_{i=1}^{m}\lambda^2_{y(i)}\right) + \lambda^2\left(\sigma^2_x - \frac{1}{m}\sum_{i=1}^{m}\mu^2_{x(i)}\right)\right. \\ & \left. -2\lambda\left(\sigma_{xy} - \frac{1}{m}\sum_{i=1}^{m}\mu_{xy(i)}\right)\right] - \frac{1}{rm^2}[\lambda^2\sigma^2_x - 2\lambda\sigma_{xy}] \\ = & \frac{1}{n}\left[\left(\sigma^2_d - \frac{1}{m}\sum_{i=1}^{m}\mu^2_{d(i)}\right) - \frac{1}{rm}\left(\lambda^2\sigma^2_x - 2\lambda\sigma_{xy}\right)\right] \end{aligned} \tag{2.13}$$

The above results are summarized as follows:

2.1. Theorem.

(i) *The estimator $\bar{t}$ is unbiased for μ_y*

(ii) *The variance of $\bar{t}$ is*

$$\begin{aligned} V(\bar{t}) &= \frac{1}{nm}\left[\sum_{i=1}^{m}\sigma_{y(i)}^2 + \lambda^2\left(\sum_{i=1}^{m}\sigma_{x(i)}^2 - \sigma_x^2\right) - 2\lambda\left(\sum_{i=1}^{m}\sigma_{xy(i)} - \sigma_{xy}\right)\right] \\ &= \frac{1}{rn}\left[\sigma_d^2 - \frac{1}{m}\sum_{i=1}^{m}(\mu_{d(i)} - \mu_d)^2 + \frac{2\lambda\rho\sigma_x\sigma_y - \lambda^2\sigma_x^2}{m}\right] \end{aligned}$$

where $\mu_{d(i)} = \mu_{y(i)} - \lambda\mu_{x(i)}$.

(iii) *An unbiased estimator of $V(\bar{t})$ is*

$$\hat{V}(\bar{t}) = \frac{1}{r(r-1)}\sum_{k=1}^{r}(t_k - \bar{t})^2$$

The part (iii) of the Theorem 2.1 follows from the fact that the estimators $t_k(k = 1, \ldots, r)$ are independently identically distributed random variables.

2.2. Optimum value of λ. The optimum value of λ that minimizes $V(\bar{t})$ is obtained from the equation

$$\frac{\partial V(\bar{t})}{\partial\lambda} = 0 \tag{2.14}$$

and it is given by

$$opt\lambda = \lambda_0 = \frac{\sum_{i=1}^{m}\sigma_{xy(i)} - \sigma_{xy}}{\sum_{i=1}^{m}\sigma_{x(i)}^2 - \sigma_x^2} \tag{2.15}$$

$$= \delta\frac{\sqrt{\sum_{i=1}^{m}\sigma_{y(i)}^2}}{\sqrt{\sum_{i=1}^{m}\sigma_{x(i)}^2 - \sigma_x^2}} \tag{2.16}$$

where δ is the correlation coefficient between $\frac{1}{rm}\sum_{k=1}^{r}\sum_{i=1}^{k} y_{i(i)|k}$ and $\frac{1}{rm}\sum_{k=1}^{r}\sum_{i=1}^{k}\left(x_{i(i)|k} - \bar{x}_{i|k}\right)$.

Finally, the variance $\bar{t}_0$, the optimum value of $\bar{t}$ with $\lambda = \lambda_0$ is given by

$$\begin{aligned} V_0 &= (1-\delta^2)\frac{1}{m^2 r}\sum_{i=1}^{m}\sigma_{y(i)}^2 \\ &= (1-\delta^2)\frac{1}{n}\left[\sigma_y^2 - \frac{1}{m}\sum_{i=1}^{m}(\mu_{y(i)} - \mu_y)^2\right] \end{aligned} \tag{2.17}$$

2.3. Precision of the proposed optimum estimator $\bar{\mathbf{t}}_0$. The relative precision of $\bar{t}_0$ with respect to the conventional estimator $\hat{\mu}_{y(rss)}$ based on an SRSWR sample mean of size $n = mr$ is given by

$$RP_{t_0|rss} = \frac{V(\hat{\mu}_{y(rss)})}{V(\bar{t}_0)} = \frac{1}{1-\delta^2} \tag{2.18}$$

From the expression (2.18), we note that the modified estimator is more efficient than the conventional RSS estimator $\hat{\mu}_{y(rss)}$ since $\delta^2 \leq 1$.

2.4. Estimator of λ_0. The optimum estimator t_0 cannot be used in practice since the value λ_0 is generally unknown. The following estimators for λ_0 may be used

$$\hat{\lambda}_0 = \frac{\sum_{k=1}^{r}(g_k - \bar{g})(h_k - \bar{h})}{\sum_{k=1}^{r}(h_k - \bar{h})^2} \tag{2.19}$$

and

$$\hat{\lambda}_1 = \frac{\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i(i)|k}x_{i(i)|k} - \left(\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i(i)|k}\right)\left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k}\right)/(rm)}{\sum_{k=1}^{r}\sum_{i=1}^{m} x^2_{i(i)|k} - \left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k}\right)^2/rm} \tag{2.20}$$

where $g_k = \frac{1}{m}\sum_{i=1}^{m} y_{i(i)|k}$, $h_k = \frac{1}{m}\sum_{i=1}^{m}(x_{i(i)|k} - \bar{x}_{i|k})$, $\bar{g} = \frac{1}{r}\sum_{k=1}^{r} g_k$ and $\bar{h} = \frac{1}{r}\sum_{k=1}^{r} h_k$.

2.5. Ratio and difference estimators. Instead of the optimum value of λ_0, one may use the following ratio and difference estimators:

$$\bar{t}_R = \left(\frac{\hat{\mu}_{y(rss)}}{\hat{\mu}_{x(rss)}}\right)\left(\frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\bar{x}_{i|k}\right) \tag{2.21}$$

and

$$\bar{t}_d = \hat{\mu}_{y(rss)} - (\hat{\mu}_{x(rss)} - \bar{x}) \tag{2.22}$$

where $\hat{\mu}_{x(rss)} = \frac{1}{mr}\left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k}\right)$, $\hat{\mu}_{y(rss)} = \frac{1}{mr}\left(\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i(i)|k}\right)$ and $\bar{x} = \left(\frac{1}{m^2r}\sum_{k=1}^{r}\sum_{j=1}^{m}\sum_{i=1}^{m} x_{ij|k}\right)$

For large $n = mr$, the ratio estimator is appropriately unbiased and an approximate estimator of the mean square of $\hat{\mu}_{x(rss)}$ is obtained by using Cochran (1977) as

$$\begin{aligned} M(\bar{t}_R) &\cong \mu_x^2 V\left(\frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i(i)|k} - \theta\frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k}\right) \\ &\cong \frac{\mu_x^2}{n}\Big[\left(\sigma_y^2 - 2\theta\rho_{xy}\sigma_x\sigma_y + \theta^2\sigma_x^2\right) \\ &\quad -\frac{1}{m}\sum_{i=1}^{m}\left\{\left(\mu_{y(i)} - \theta\mu_{x(i)}\right) - (\mu_y - \theta\mu_x)\right\}^2\Big] \end{aligned} \tag{2.23}$$

where $\theta = \frac{\mu_y}{\mu_x}$.

From the expression (2.23), we note that the ratio estimator based on ranked set sample is more precise than the conventional ratio estimator based on the same sample size.

A reasonably good estimator of $M(\bar{t}_R)$ is

$$\hat{M}(\bar{t}_R) \cong \hat{\mu}^2_{x(rss)}\frac{1}{n-1}\sum_{k=1}^{r}\sum_{i=1}^{m}\left(z_{i(i)|k} - \bar{z}\right)^2 \tag{2.24}$$

where $z_{i(i)|k} = y_{i(i)|k} - \hat{\theta}x_{i(i)|k}$, $\bar{z} = \sum_{k=1}^{r}\sum_{i=1}^{m} z_{i(i)|k}/n$ and $\hat{\theta} = \frac{\hat{\mu}_{y<rss>}}{\hat{\mu}_{x(rss)}}$.

It is easy to note that the difference estimator $\bar{t}_d$ is always unbiased and it is more efficient than the conventional difference estimator of the same sample size.

2.6. Judgment ranking. Sometimes ranking may be imperfect. Let $y_{i<j>|k}$ be the smallest jth "judgment order statistic" corresponding to order statistic $x_{i(j)|k}$ in the ith set of the cycle k. In case the judgment ranking is perfect $y_{i<j>|k}$ becomes equal to $y_{i(j)k}$, otherwise if the judgment process is imperfect, we find $y_{i<j>|k} \neq y_{i(j)|k}$. Here we assume that the expectation of $y_{i<j>|k}$ over the judgment process is the true ranking so that $E(y_{i<j>|k}) = y_{i(j)|k}$. In this case we modify the estimators $\hat{\mu}_{y(rss)}, \bar{t}_0,\ \bar{t}_1,\ \bar{t}_R$ and $\bar{t}_d$ by replacing $y_{i(j)|k}$ with $y_{i<j>|k}$. The modified estimators become respectively as follows:

$$\begin{aligned}
\hat{\mu}_{y<rss>} &= \frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i<i>|k} \\
\bar{t}_{<0>} &= \frac{1}{mr}\left[\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i<i>|k} - \lambda_{<0>}\left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k} - \sum_{k=1}^{r}\sum_{i=1}^{m} \bar{x}_{i|k}\right)\right], \\
\bar{t}_{<1>} &= \frac{1}{mr}\left[\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i<i>|k} - \lambda_{<1>}\left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k} - \sum_{k=1}^{r}\sum_{i=1}^{m} \bar{x}_{i|k}\right)\right], \\
\bar{t}_{<R>} &= \left(\frac{\hat{\mu}_{y\langle rss\rangle}}{\hat{\mu}_{x(rss)}}\right)\left(\frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m} \bar{x}_{i|k}\right) \\
&\text{and} \\
\bar{t}_d &= \hat{\mu}_{y<rss>} - (\hat{\mu}_{x(rss)} - \bar{x}) \qquad (2.25)
\end{aligned}$$

where $\hat{\lambda}_{<0>} = \dfrac{\sum_{k=1}^{r}(g_{<k>} - \bar{g}_{<>})(h_k - \bar{h})}{\sum_{k=1}^{r}(h_k - \bar{h})^2}$,

$$\hat{\lambda}_{<1>} = \frac{\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i<i>|k} x_{i(i)|k} - \left(\sum_{k=1}^{r}\sum_{i=1}^{m} y_{i<i>|k}\right)\left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k}\right)/(rm)}{\sum_{k=1}^{r}\sum_{i=1}^{m} x^2_{i(i)|k} - \left(\sum_{k=1}^{r}\sum_{i=1}^{m} x_{i(i)|k}\right)^2/(rm)},$$

$g_{<k>} = \dfrac{1}{m}\sum_{i=1}^{m} y_{i<i>|k}$, $\quad \bar{g}_{<>} = \dfrac{1}{r}\sum_{k=1}^{r} g_{<k>}$ with h_k and $\bar{h}$ as defined in section 2.4.

The modified estimator $\hat{\mu}_{y<rss>}$ remains exactly unbiased for μ_y while the remaining modified estimators based on the judgment order statistics remains approximately unbiased for μ_y.

3. Simulation studies

In the proposed simulation study we consider the tree data set originally collected by Platt et al. (1988) and cited by Chen et al. (2003). The data comprises of diameters in centimetre (cm) at breastheights (x) and entire height (y) in feet of 396 trees. The mean diameter and height of the 396 trees are $\mu_x = 20.9641$ and $\mu_y = 52.6768$ respectively. Treating the 396 trees as a population, initially a sample of m^2 trees is selected by SRSWR sampling procedures. The selection of the sample (cycle) is repeated r times. Since, for this data y does not always increase with x, we have compared performances with the proposed five estimators $\hat{\mu}_{y<rss>}, \bar{t}_{<0>},\ \bar{t}_{<1>},\ \bar{t}_{<R>}$ and $\bar{t}_{<d>}$ based on judgement order statistic. However, as per suggestions from one of the referees, we have considered the following ratio estimator (Kadilar et al., 2009) and regression estimators when the population mean μ_x is known

$$\bar{t}^*_{<R>} = \frac{\hat{\mu}_{y<rss>}}{\hat{\mu}_{x(rss)}}\mu_x \quad \text{and} \quad \bar{t}^*_{<1>} = \hat{\mu}_{y<rss>} - \hat{\lambda}_{<1>}(\hat{\mu}_{x(rss)} - \mu_x) \tag{3.1}$$

We call the process of selection of m^2 trees and replication r times as an iteration. The iteration is repeated R = 100,000 times. Let the values of the $\hat{\mu}_{y<rss>}$, $\bar{t}_{<0>}$, $\bar{t}_{<1>}$, $\bar{t}_{<R>}$, $\bar{t}_{<d>}$, $\bar{t}^*_{<R>}$ and $\bar{t}^*_{<1>}$ based on the q^{th} iteration be denoted by $\hat{\mu}_{y<rss>}(q)$, $\bar{t}_{<0>}(q)$, $\bar{t}_{<1>}(q)$, $\bar{t}_{<R>}(q)$, $\bar{t}_{<d>}(q)$, $\bar{t}^*_{<R>}(q)$ and $\bar{t}^*_{<1>}(q)$ respectively.

The percentage relative biases (RB) and mean square errors (MSE) of the seven estimators are computed by the following formula:

$$RB(\hat{\theta}) = \frac{1}{\mu_y}\left(\frac{1}{R}\sum_{q=1}^{R}\hat{\theta}\,(q) - \mu_y\right) \quad \text{and} \quad MSE(\hat{\theta}) = \frac{1}{R}\sum_{q=1}^{R}(\hat{\theta}(q) - \mu_y)^2 \tag{3.2}$$

where $\mu_y = 52.6768$ and $\hat{\theta} = \hat{\mu}_{y<rss>},\ \bar{t}_{<0>},\ \bar{t}_{<1>},\ \bar{t}_{<R>},\ \bar{t}_{<d>},\ \bar{t}^*_{<R>},\ \bar{t}^*_{<1>}$.

The relative efficiency of the estimator $\hat{\theta}$ compared with the conventional estimator $\mu_{y<rss>}(q)$ is given by

$$RE(\hat{\theta}) = 100 \times MSE(\hat{\mu}_{y<rss>})/MSE(\hat{\theta})\% \tag{3.3}$$

The values of $RB(\hat{\theta})$ and $RE(\hat{\theta})$ are computed for different combinations of $m(= 3, 4, 6, 10)$ and $r = (3, 6, 8, 9, 12, 15, 18, 20, 36)$. These are presented in the following Table-1 and Table-2. The simulation study shows for unknown, μ_x, the population mean of x, all the proposed estimators are approximately unbiased. The maximum absolute relative bias was 1.25. The minimum standard error (which is approximately $\sqrt{MSE}$) is 3.69 (not shown in the table). The biases of all the estimators are ignorable since the maximum of the ratio of bias of an estimator to its standard error is $0.0034 << 0.1$(see Cochran (1977)). For a given sample size $n(= mr)$ the biases of all the estimators increase with m. As per efficiency, all the proposed estimators are more efficient than the conventional estimator $\hat{\mu}_{y<rss>}$ in all situations considered here. The estimator $\bar{t}_{<1>}$ performed the best, closely followed by $t_{<R>}$ and $t_{<0>}$. The estimator $t_{<d>}$ performed least among the proposed five estimators. The maximum relative efficiency 147.70 was attained by $\bar{t}_{<1>}$ with $m = 10$, $r = 9$ and it attained the minimum 133.66 when $m = 3$, $r = 12$. This shows that the estimator $\bar{t}_{<1>}$ brings gains in efficiency over the conventional estimator $\hat{\mu}_{y<rss>}$ between 33% and 48% for estimating the population mean μ_y. However, in case μ_x, the population mean is known one should use the estimators $t^*_{<R>}$ and $\bar{t}^*_{<1>}$ as they perform much better than all the proposed estimators with respect to bias and mean square errors.

Table-1: Relative Bias of the proposed estimators

Sample			$B\left(\hat{\theta}\right)$						
size n	m	r	$\hat{\mu}_{y<rss>}$	$\bar{t}_{<0>}$	$\bar{t}_{<1>}$	$\bar{t}_{<R>}$	$\bar{t}_{<d>}$	$\bar{t}^*_{<1>}$	$\bar{t}^*_{<R>}$
36	3	12	0.29	−0.35	0.20	0.21	0.30	−0.24	−0.19
	4	9	0.43	0.38	0.29	0.31	0.42	−0.14	−0.09
	6	6	0.67	0.45	0.52	0.54	0.67	−0.05	0.03
54	3	18	0.26	−0.15	0.20	0.20	0.27	−0.15	−0.11
	6	9	0.66	0.51	0.55	0.56	0.66	0.04	0.10
60	4	15	0.44	0.17	0.33	0.34	0.43	−0.02	0.02
	10	6	1.19	1.14	1.12	1.12	1.19	0.06	0.21
72	3	24	0.30	−0.02	0.24	0.25	0.30	−0.11	−0.07
	4	18	0.42	0.20	0.34	0.35	0.41	−0.00	0.04
	6	12	0.66	0.54	0.58	0.59	0.66	0.07	0.13
80	4	20	0.41	0.21	0.34	0.34	0.41	0.00	0.04
	10	8	1.19	1.18	1.15	1.15	1.19	0.10	0.24
90	3	30	0.30	0.06	0.28	0.28	0.31	−0.08	−0.04
	6	15	0.65	0.57	0.59	0.59	0.65	0.10	0.16
	10	9	1.21	1.20	0.16	1.17	1.21	0.11	0.26
108	3	36	0.32	0.12	0.30	0.30	0.33	−0.05	−0.01
120	6	20	0.65	0.57	0.60	0.60	0.64	0.13	0.19
	10	12	1.21	1.21	1.19	1.19	1.22	0.14	0.28

Table-2: Relative Efficiencies of the proposed estimators

Sample size n	m	r	$E\left(\hat{\theta}\right)$ $\hat{\mu}_{y<rss>}$	$\bar{t}_{<0>}$	$\bar{t}_{<1>}$	$\bar{t}_{<R>}$	$\bar{t}_{<d>}$	$\bar{t}^*_{<1>}$	$\bar{t}^*_{<R>}$
36	3	12	100	119.29	133.66	132.87	115.94	427.32	410.01
	4	9	100	118.63	139.52	138.29	117.99	394.42	380.69
	6	6	100	109.33	145.55	143.67	119.70	356.66	342.76
54	3	18	100	125.59	134.50	133.27	116.05	434.01	409.9
	6	9	100	125.52	146.00	143.67	119.70	363.26	343.26
60	4	15	100	129.13	140.14	138.46	118.05	401.35	379.73
	10	6	100	111.93	147.67	144.93	119.97	320.61	301.68
72	3	24	100	127.54	134.09	132.81	115.94	440.94	412.61
	4	18	100	130.61	139.48	137.83	117.87	402.2	378.78
	6	12	100	132.17	146.41	143.77	119.69	366.35	343.37
80	4	20	100	132.37	140.18	138.39	118.05	403.1	378.85
	10	8	100	124.22	147.28	144.39	119.78	323.22	302.52
90	3	30	100	129.53	134.71	133.33	116.13	441.11	411.54
	6	15	100	135.51	146.19	143.50	119.63	368.25	344.19
	10	9	100	127.61	147.70	144.69	119.86	322.81	302.05
108	3	36	100	130.82	134.74	133.26	116.06	443.02	411.38
120	6	20	100	138.59	145.93	143.15	119.49	366.84	341.58
	10	12	100	133.46	146.66	143.75	119.59	325.82	303.93

Acknowledgement
The authors are grateful to the referees and the editor for their fruitful suggestions which led to the sufficient improvement of the earlier version of the manuscript.

References

[1] Chen, Z., Bai, Z. Sinha, B.K. (2003). *Ranked Set Sampling theory and Applications*, Springer, New York.

[2] Cochran, W.G. (1977). *Sampling Techniques*, third ed, John Wiley & Sons, Inc.

[3] Dell, T.R. and Clutter, J.L. (1972). *Ranked-set sampling: theory with order statistic background*, Biometrics, **28**, 545-555.

[4] Kadilar, C., Unyazici, Y. and Cingi, H. (2009). *Ratio estimator for the population mean using ranked set sampling*, Statist. Papers, **50**, 301-309.

[5] Kaur, A., Patil, G.P. and Taillie, C. (1997). *Unequal allocation model for ranked set sampling with skew distribution*, Biometrics **53**, 123-130.

[6] McIntyre, G.A. (1952). *A method of unbiased selective sampling, using ranked sets.* Austral. J. Agri. Res., **3**, 385-390.

[7] Platt, W.J., Evans, G.M., and Rathbun, S.L. (1988). *The population dynamics of long-lived conifer (pinus plustris)*, American Naturalist, **131**, 491-525.

[8] Prasad, B. (1989). *Some improved ratio type estimators of population mean and ratio in finite population sample surveys*, Communi. Statist. Theory Methods **18**, 379-392.

[9] Singh, H.P., Tailor, R. and Singh, S. (2014). *General procedure for estimating the population mean using ranked set sampling*, Journal of Statistical Computation and Simulation, **84**, 931-945.

[10] Stokes , S.L. (1988). *Characterization of a ranked-set sample with application to estimating distribution functions*, J. Amer. Statist. Assoc., **83**, 374-381.

[11] Stokes, S.L. (1977). *Ranked set sampling with concomitant variables*, Communi. Statist. Theory. Methods A *6, 12,* 1207-1211.

[12] Stokes, S.L. (1980 a). *Inference on correlation coefficient in bivariate normal populations from ranked-set sampling*, J. Amer. Statist. Assoc., **75**, 989-995.

[13] Stokes, S.L. (1980). *Estimation of variance using judgment ordered ranked-set samples*, Biometrics, **36**, 35-42.