Ranked set sampling versus simple random sampling in the estimation of the mean and the ratio

S. Ganeslingam *
Siva Ganesh †

Institute of Information Sciences and Technology
College of Sciences
Massey University
Private Bag 11 222
Palmerston North
New Zealand

Abstract

It is common in practice that the experimental units can be ranked easily using a cheaply measurable covariate than quantification of the main variable of interest which requires expensive measurements. In such situation ranked set sampling is more beneficial and cost effective. Environmental monitoring and assessment, for example, requires observational data where the ranked set sampling is proved to achieve observational economy when compared to the traditional simple random sampling. Ranked set sampling employs judgement ordering to obtain the actual sample and hence yield a sample of observations that is more representative of the underlying population. Therefore, either greater confidence is gained for a fixed number of observations, or for a desired level of confidence, a smaller number of observations is needed. In either way it is a big gain to the researcher. In this paper, we introduce the basic concepts of ranked set sampling and its application in the estimation of the population mean and the ratio using a real data set on body measurements.

Keywords: Ranked set sample (RSS), simple random sample (SRS), population ratio, population mean, relative precision, relative saving, empirical efficiency.

*E-mail: s.ganeslingam@massey.ac.nz
†E-mail: s.ganesh@massey.ac.nz

Journal of Statistics & Management Systems
Vol. 9 (2006), No. 2, pp. 459–472
© Taru Publications
1. Introduction

Consider a real life example such as environmental monitoring and assessment. In such situation it is not uncommon for the quantification of a sampling unit to be costly as compared to the physical acquisition of the unit. For example, consider the estimation of mean pool area for a stream in an ecological study. Here, the ranked set sample (RSS) is proved to be more economical and to better represent the population when compared to the traditional simple random sample (SRS). Pool area can be estimated visually or a team of experts can measure it more accurately. Ranked set sampling would use visual judgement to choose a few pools for expensive measurement. This is particularly useful when there is a financial constraint on the sample measurements.

Another example is selection of domestic animals for their genetic evaluation. Laboratory measurements of many genetic variables are very costly and requires expensive instruments and labour. Consequently it demand a small representative sample carefully selected to make such expensive measurements. Therefore, a cheaply measurable co-variate (eye judgement, for example) could be used to rank the sampling units before making the expensive measurements.

In such a situations considerable cost savings can be made if the number actually measured is a small fraction of the number of units available, but all units contribute to the study. Several researchers have studied ranked set sampling, but a complete review of applications and theoretical frame work on RSS is available in Patil et al (1994), Kaur et al (1995), and Johnson et al (1996). A recent study extended the application of RSS to the estimation of the distribution function using extreme and median ranked set sampling (Samawi and Al-Sagheer (2001)). Stokes (1977) studied the ranked set sampling with concomitant variables, while Halls and Dell (1966) applied RSS for forage yields to evaluate its performance. Very recently You-Gan Wank et al (2004) considered the optimal set size for ranked set sampling with fixed operational costs.

2. What is a “Ranked Set Sample” (RSS)

In ranked set sampling, separate sampling units maintain their individual identity. We must be able to rank the units in some way; as such RSS implementation requires some covariate information to perform the ranking. That is, if another characteristic is available which is highly correlated with the main characteristic of interest, but less costly or no cost
to obtain, then we may rank the units by the values of such covariate. For example, a hazardous waste site inspector may be able to reliably rank areas of soil with respect to concentration of a toxic contaminant based on features (covariates) such as surface staining, discoloration or the appearance of stressed vegetation. It is interesting to note that there is an in-built stratification taking place with RSS. McIntrye (1952) noted that in many situations, there is enough information available to enable sampling units to be ranked according to the variable of interest, but actually not measuring (quantifying) the units with respect to the variable of interest. The ranking may be done on the basis of visual inspection, using covariates that are easy to measure, or any prior information. RSS consists of the following steps: (Kaur et al (1995)).

(i) Identifying a pool of sampling units from the target population.

(ii) Randomly partitioning the pool into disjoint subsets each having a pre-assigned size – known as the set size, often taken to be less than or equal to four, for convenience of ranking and to minimize ranking error.

(iii) Ranking each sub-set.

(iv) Actually measuring one suitably selected member from each ranked sub-set.

As a simple illustration of the concept of RSS consider the following example. Let’s say we wish to estimate the mean height of Olympic game participants in a particular year, using a random sample of only three participants. Furthermore, in order to acknowledge the inherent uncertainty, we need to present this estimate as a confidence interval within which we expect the true population mean height to lie with desired confidence.

The simplest way to obtain our sample is to randomly select three participants from the entire set of participants, and measure their heights. While the mean, $\bar{X}_{SRS}$, of the three heights we measured is an unbiased point estimate of the population mean, the associated confidence interval can be very large, reflecting the high degree of uncertainty with estimating the mean with only three measurements from a large population. This is because we have no control over which individuals of the population enter the sample. This high degree of uncertainty can be reduced only by taking a very good representative sample of three participants or by increasing the sample size of the SRS, which we have restricted to only three, for
economic reasons. RSS comes to the rescues and assists us in increasing the representativeness of the sample.

To do this, we may randomly invite three participants to a free breakfast and visually rank them with respect to their heights. We then select the participant we judge as the shortest, and then actually measure this person’s height. When we repeat this process for lunch with another three randomly selected participants, we actually measure the height of the middle ranked person. Similarly, we measure the height of the tallest ranked person at dinner with another fresh set of three randomly selected participants. Note that we deal with nine participants, but only three are actually measured. The resulting three measurements (the shortest, medium, and the tallest) form the RSS. The mean of these three measurements, \( \bar{X}_{RSS} \) is also an unbiased point estimate of the population mean height. However, the standard error of this sample mean will be smaller compared to that of the \( \bar{X}_{SRS} \). This is because these RSS measurements are likely to be more regularly spaced than those of SRS; and therefore better represent the population. In effect, the RSS procedure induces stratification of the whole population at the sample level.

Now let us examine the mathematical formulation of RSS. For the RSS scheme we select \( m \) random sets each of size \( m \) from the target population. In practice, \( m \) usually takes values such as 2, 3, or 4. Each set is then ranked by a convenient (cheap) method. In matrix notation, we have

\[
\begin{pmatrix}
X_{11} & X_{12} & \cdots & X_{1m} \\
X_{21} & X_{22} & \cdots & X_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
X_{m1} & X_{m(2)} & \cdots & X_{mm}
\end{pmatrix} \rightarrow \begin{pmatrix}
X_{1(1)} & X_{1(2)} & \cdots & X_{1(m)} \\
X_{2(1)} & X_{2(2)} & \cdots & X_{2(m)} \\
\vdots & \vdots & \ddots & \vdots \\
X_{m(1)} & X_{m(2)} & \cdots & X_{m(m)}
\end{pmatrix}.
\]

Here, \( x_{ij} \) denotes the \( j \)th observation in the \( i \)th set and \( x_{i(j)} \) is the \( j \)th ordered statistic in the \( i \)th set.

After ranking, select only the diagonal units and actually measure them. This constitutes the ranked set sample, RSS. That is, we have only measures \( x_{1(1)}, x_{2(2)}, \ldots, x_{m(m)} \), by obtaining the unit with the smallest rank from the first row, the second smallest rank from the second row and so on until the largest unit from the \( m \)th row. This represents one cycle of RSS. We can repeat the whole procedure \( r \) times to get a RSS of size \( n = rm \). It is to be noted here that RSS requires \( m^2 \) units to be taken, but only \( m \) of them are actually measure.
3. **Aim of study and the description of the data set**

In this paper we consider a real data set on body measurements to evaluate the performance of the RSS estimator compared to SRS estimator of the population mean height and the ratio of the mean height versus the mean weight. Thus the height \(X\) is the main variable of interest and the Weight \(Y\) is considered as a covariate. The data set consists of 25 body measurements of 507 New Zealand individuals in their twenties and thirties, with a scattering of older men and women, all exercising several hours a week. For our information, it is known that there are 247 men and 260 women in the data set. This information is useful to obtain a stratified random sample using gender as a stratifying covariate. Suppose our interest is to estimate the average height of the whole population of 507 individuals. A stratified random sample using gender as the covariate will give us a better estimate of the population mean compared to SRS and RSS methods. We would expect the RSS estimate to fall in between the SRS estimate and the stratified random sample estimate. Our interest here is to only compare the RSS estimate with the SRS estimate. An efficiency measure for this purpose is defined, computed and compared.

Since we know the actual values of the height and the weight, a perfect ranking is possible and hence enables us to compare the estimate of the mean height obtained with the perfect ranking with that of the judgement ranking using the weight as the covariate. The ratio of the mean height to the mean weight can also be obtained in both ranking which are compared with the actual ratio of the mean height to the mean weight.

4. **Mathematical underpinning**

Suppose we have a simple random simple, \(x_1, x_2, \ldots, x_n\) of size \(n\), from a population with mean \(\mu\) and finite variance \(\sigma^2\). Then the traditional non-parametric estimator of \(\mu\) is given by

\[
\hat{\mu}_{SRS} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

with the variance

\[
V(\hat{\mu}_{SRS}) = \frac{\sigma^2}{n}
\]
Now with RSS, \( n = rm \) as explained in Section 2.

\[
\bar{X}_{RSS} = \frac{1}{rm} \sum_{k=1}^{r} \sum_{i=1}^{m} x_{(i)k}
\]

and

\[
V(\bar{X}_{RSS}) = \frac{\sigma^2}{n} - \frac{1}{rm^2} \sum_{i=1}^{m} (\mu_{(i)} - \mu)^2,
\]

where \( \mu_{(i)} \) is the mean of the \( i \)th ranked set, and is given by \( \mu_{(i)} = \frac{1}{r} \sum_{k=1}^{r} x_{(i)k} \).

We can see a variance reduction factor of \( \frac{1}{rm^2} \sum_{i=1}^{m} (\mu_{(i)} - \mu)^2 \) in the expression for \( V(\bar{X}_{RSS}) \) above, associated with \( \bar{X}_{RSS} \).

Takhasi and Wakimoto (1968) were the first to study the mathematical theory of RSS in detail. They defined two quantities namely relative precision (RP) and relative savings (RS) by

\[
RP = \frac{\text{Var}(\bar{X}_{SRS})}{\text{Var}(\bar{X}_{RSS})}
\]

and

\[
RS = \frac{\text{Var}(\bar{X}_{SRS}) - \text{Var}(\bar{X}_{RSS})}{\text{Var}(\bar{X}_{SRS})}.
\]

For a set size of 3, we have \( RS = \frac{1}{3} \sum_{i=1}^{3} \left( \frac{\mu_{(i)} - \mu}{\sigma} \right)^2 \).

The authors showed that \( RS \) and \( RP \) satisfy the following relations.

\[
0 \leq RP \leq \frac{m + 1}{2},
\]

\[
0 \leq RS \leq \frac{m - 1}{m + 1}.
\]

and

\[
RP = \frac{1}{1 - RS}.
\]

The above two quantities are used to compare the performance of the two estimators \( \bar{X}_{SRS} \) and \( \bar{X}_{RSS} \) as an estimator of the population mean \( \mu \), in our case study considered in sections. Under the equal allocation of
each order statistic, RSS will always result in as precise an estimate as SRS if not better. The degree to which RSS exceeds SRS will depend upon the amount of information gained about the distribution from ranking.

In a recent study on estimation of milk yield using RSS, Al-Saleh and Al-Shrafat (2001) used 402 sheep with \( m = 3 \) and showed that \( RP = 1.80 \) for a perfect ranking. That is, the accuracy of the estimator of \( \mu \) using a RSS of size 100 is equivalent to the accuracy of the estimator of \( \mu \) using a SRS of size 180. They also showed that \( RP = 1.50 \) for the judgement ranking. This difference is due to the ranking error. Takahasi and Wakimoto (1968) warned that in practice the number of units, which are easily ranked, should not be more than four.

5. Case study

In our data set we have 507 individuals whose actual weight and height are known. Hence we can calculate the population mean height and population mean weight and the corresponding population variances. These will be compared with our corresponding estimates obtained using SRS and RSS.

First our attention was focused on the estimation of the mean height. For the RSS protocol, these 507 individuals were divided into 169 sets each of size three. The individuals in each set were ranked perfectly (using their actual known heights) as lowest (rank 1), medium (rank 2) and the tallest (rank 3). A similar judgement ranking was done using the weight as a covariate. Thus we have two sets of RSS data, one obtained by perfect ranking and the other obtained by the judgement ranking using the weight as the covariate. In real practice these two variables are reasonably correlated and hence the judgement sample will not differ very much from the sample with perfect ranking.

5.1 Summary of the data and the results based on perfect ranking

Let \( x_i; i = 1, 2, \ldots, 507 \), be the actual height of the \( i \)th individual in the population. The population mean height and the population variance are given by

\[
\mu = \frac{1}{507} \sum_{i=1}^{507} x_i = 171.14 \quad \text{and} \quad \sigma^2 = \frac{1}{507} \sum_{i=1}^{507} (x_i - \mu)^2 = 88.50 .
\]
Now $\mu_{(i)}$, $(i = 1, 2, 3)$ the mean of all the individuals with the $i$th rank is given by,

$$\mu_{(1)} = 163.33, \quad \mu_{(2)} = 170.86, \quad \mu_{(3)} = 179.25. $$

If $n = 3r$, then $RS$ of using $\bar{X}_{RSS}$ with respect to $\bar{X}_{SRS}$ is given by

$$RS = \frac{1}{3} \sum_{i=1}^{3} \frac{(\mu_{(i)} - \mu)^2}{\sigma^2} = 0.4778.$$ 

This implies that, with the perfect ranking, the saving in the sample size in using $RSS$ with set size three is 48% as compared to $SRS$.

Also, $RP = \frac{1}{1 - RS} = 1.914$ indicates that the accuracy of the estimator of $\mu$ using a $RSS$ of size 100 is equivalent to the accuracy of the estimator of $\mu$ using a $SRS$ of size 192.

5.2 Result based on judgement ranking using the covariate-weight

Here we select a pair $(x_i, y_i)$; where $x_i$ represents the height and the $y_i$ represents the weight, which is to be used as the covariate in the judgement ranking.

Let $\mu_{(i)}^*$ be the mean height of the individuals with $i$th judgement rank, $i = 1, 2, 3$. Then we have

$$\mu_{(1)}^* = 165.18, \quad \mu_{(2)}^* = 171.43, \quad \mu_{(3)}^* = 176.81.$$ 

If $n = 3r$ as in Section 5.1 we find $RS = \frac{1}{3} \sum_{i=1}^{3} \left( \frac{\mu_{(i)}^* - \mu}{\sigma} \right)^2 = 0.2552$.

That is, with the judgement ranking, the saving in the sample size by using $RSS$ with set size three is 26% as compared to $SRS$.

Also, $RP = \frac{1}{1 - RS} = 1.34$ indicates that the accuracy of the estimator of $\mu$ using a $RSS$ of size 100 is equivalent to the accuracy of the estimator of $\mu$ using a $SRS$ of size 134.

In our mixed population of males and females we have a moderate correlation, $\rho_{(xy)} = 0.72$. Therefore, ranking the heights using weights in this case will have errors and hence explain a notable difference between the $RP$ in the case of perfect ranking and judgement ranking using the weight as the covariate. However, a relative efficiency of 34% is attained. A highly correlated covariate will definitely increase this efficiency. Hence this leads to the idea of using a cheaply measurable but highly correlated covariate to rank the units to quantify the expensive variable of interest.
These results given in Sections 5.1 and 5.2 are based on the infinite population. In the following Section 5.3 we use our population of 507 individuals and empirically derived relative precision in perfect ranking and judgement ranking using the weight as the covariate.

5.3 Empirical efficiency

Here, we wish to see how the theoretical relative precisions obtained in the Section 5.1 and 5.2 compares with the empirical efficiencies. So we will find the efficiencies empirically based on 507 pairs of (heights and weights) values we have. This is done via the following steps:

**Step 1.** Three sets of size three are taken randomly from 507 pairs of observations \((x_i, y_i), i = 1, \ldots, 507\), which is our target population. These sets are ranked using the \(x_i\), the main variable of interest, which we call a perfect ranking. From the first set the lowest value of \(x\) is obtained, from the second set the medium value is obtained and from the third set the largest value is obtained.

**Step 2.** Step 1 is repeated \(r\) times so that we have a RSS sample of size \(n = mr\) (\(m = 3\)). The mean of these observations are calculated.

**Step 3.** Steps 1 and 2 are repeated 5000 times so that we obtain 5000 values of RSS estimates of the parameter \(\mu\). The mean and variance of these values are computed.

**Step 4.** A SRS of size \(n\) is obtained and the mean is calculated.

**Step 5.** Step 4 is repeated 5000 times and the mean and variance of these 5000 mean values are computed.

**Step 6.** The value of \(RP\) is obtained as the variance in Step 5 divided by the variance in Step 3.

**Step 7.** Steps 1 to 6 are repeated for the judgement ranking using the weight as the covariate. Tables 1 and 2 give the results for \(n = 9, 12, 18\) and 30.

It can be seen from the Tables 1 and 2, that the empirical relative precision \((RP)\) is reasonably close to the theoretical one. Also, it is interesting to note that the sample size has no effect on \(RP\). Furthermore, we can note that both (RSS and SRS) estimates are very close to the actual mean height value (171.1cm), but the RSS estimator (in both perfect and judgement sampling) has smaller variability than the SRS estimator, hence is more efficient. We should also note that as expected, judgement sampling is less efficient (has more variability) than the perfect sampling (see \(V(\bar{x}_{RSS})\) in Tables 1 and 2).
Table 1
Empirical mean, variance and the relative precision of RSS with respect to SRS in the case of perfect ranking

<table>
<thead>
<tr>
<th>n = 3r</th>
<th>Mean of $\bar{X}_{RSS}$</th>
<th>Mean of $\bar{X}_{SRS}$</th>
<th>$\text{Var}(\bar{X}_{RSS})$</th>
<th>$\text{Var}(\bar{X}_{SRS})$</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>171.17</td>
<td>173.13</td>
<td>5.02</td>
<td>9.57</td>
<td>1.91</td>
</tr>
<tr>
<td>12</td>
<td>171.18</td>
<td>172.12</td>
<td>3.72</td>
<td>7.10</td>
<td>1.91</td>
</tr>
<tr>
<td>18</td>
<td>171.17</td>
<td>171.22</td>
<td>2.47</td>
<td>4.64</td>
<td>1.88</td>
</tr>
<tr>
<td>30</td>
<td>171.20</td>
<td>171.24</td>
<td>1.48</td>
<td>2.81</td>
<td>1.90</td>
</tr>
</tbody>
</table>

Table 2
Empirical mean, variance and the relative precision of RSS with respect to SRS in the case of judgement ranking using the covariate weight

<table>
<thead>
<tr>
<th>n = 3r</th>
<th>Mean of $\bar{X}_{RSS}$</th>
<th>Mean of $\bar{X}_{SRS}$</th>
<th>$\text{Var}(\bar{X}_{RSS})$</th>
<th>$\text{Var}(\bar{X}_{SRS})$</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>171.14</td>
<td>173.13</td>
<td>7.31</td>
<td>9.57</td>
<td>1.31</td>
</tr>
<tr>
<td>12</td>
<td>171.15</td>
<td>172.12</td>
<td>5.41</td>
<td>7.10</td>
<td>1.31</td>
</tr>
<tr>
<td>18</td>
<td>171.13</td>
<td>171.22</td>
<td>3.67</td>
<td>4.64</td>
<td>1.26</td>
</tr>
<tr>
<td>30</td>
<td>171.14</td>
<td>171.24</td>
<td>2.18</td>
<td>2.81</td>
<td>1.29</td>
</tr>
</tbody>
</table>

In real life, for example, in a forest survey situation where we wish to estimate the mean height of pine trees to estimate the total timber volume, measuring heights is expensive and cumbersome compared to measuring the girth of the tree at shoulder height. Thus, girth can be used as a covariate to rank the trees for the purpose of estimating the average height. In such a situation, the RSS method will be ideal for obtaining a reliable estimate of the mean height, which otherwise is expensive and time consuming.

6. Estimation of the ratio of the mean height versus the mean weight, $\frac{\mu(X)}{\mu(Y)}$

In our case study considered in the Section 5, our attention was focused on the estimation of the mean height. Suppose in the same population we want to estimate the population ratio $\frac{\mu(X)}{\mu(Y)}$. We defined

$$R = \frac{\mu(X)}{\mu(Y)}.$$
Similar to the univariate RSS sample, we take the bivariate RSS sample as follows.

**Original bivariate observations**

\[
\begin{bmatrix}
  x_{11} & y_{11} & \cdots & x_{1m} & y_{1m} \\
  x_{21} & y_{21} & \cdots & x_{2m} & y_{2m} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  x_{m1} & y_{m1} & \cdots & x_{mm} & y_{mm}
\end{bmatrix}
\]

**After ranking using the variable x (perfect ranking)**

\[
\begin{bmatrix}
  x_{1(1)} & y_{1[1]} & \cdots & x_{1(m)} & y_{1[m]} \\
  x_{2(1)} & y_{2[1]} & \cdots & x_{2(m)} & y_{2[m]} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  x_{m(1)} & y_{m[1]} & \cdots & x_{m(m)} & y_{m[m]}
\end{bmatrix}
\]

Now to obtain the RSS bivariate sample take only the diagonal elements \([x_{1(1)}, y_{1[1]}], [x_{2(2)}, y_{2[2]}], \ldots, [x_{m(m)}, y_{m[m]}]\) of the matrix of the ranked set; where \(x_{i(i)}\) is the \(i\)th perfect order statistic in the \(i\)th sample for the variable \(X\) and \(y_{i[i]}\) is the \(i\)th judgement order statistic in the \(i\)th sample for the variable \(Y\).

The whole process can be repeated \(r\) times to obtain a sample of size \(n = rm\).

### 6.1 Definitions of \(r_{SRS}\), \(r_{RSS1}\) and \(r_{RSS2}\)

The ratio estimator using SRS is given by:

\[
r_{SRS} = \left( \frac{\bar{X}}{\bar{Y}} \right)_{SRS}.
\]

The ratio estimator using RSS (with perfect ranking on \(X\)) is given by

\[
r_{RSS1} = \left( \frac{\bar{X}}{\bar{Y}} \right)_{RSS1}.
\]

And the ratio estimator using RSS (with perfect ranking on \(Y\)) is given by:

\[
r_{RSS2} = \left( \frac{\bar{X}}{\bar{Y}} \right)_{RSS2}.
\]

Hansen (1953) has given the approximate expression for the \(\text{var}(r_{SRS})\) in terms of the coefficient of variations of \(X\), \(Y\) and the correlation coefficient \(\rho(x, y)\). Samawi and Muttlak (1996) has given the expressions for the \(\text{var}(r_{RSS1})\) and \(\text{var}(r_{RSS2})\) and shown that,

(i) \(\text{var}(r_{RSS1}) \leq \text{var}(r_{SRS})\),

(ii) \(\text{var}(r_{RSS2}) \leq \text{var}(r_{SRS})\),

(iii) \(\text{var}(r_{RSS2}) \leq \text{var}(r_{RSS1})\).
Note that $r_{RSS_2}$ is defined on the basis of the perfect ranking on $Y$, the variable appears in the denominator of the ratio estimator.

Since $R$ is known perfectly, we wish to compare $r_{SRS}$, $r_{RSS_1}$, and $r_{RSS_2}$ with $R$ and wanted to see which is the best estimator in terms of their associated mean squared errors. We performed the simulation similar to the one given in Section 5, but now we consider the ratio of the means and the variance of these ratios over 5000 simulations undertaken. We define the relative efficiency as

$$RE = \frac{\text{var}(r_{SRS})}{\text{var}(r_{RSS(i)})}, \quad i = 1, 2.$$  

The results are given in Tables 3 and 4.

### Table 3

<table>
<thead>
<tr>
<th>$n = 3r$</th>
<th>Mean of $\frac{\bar{X}}{\bar{Y}}_{RSS}$</th>
<th>Variance of $\frac{\bar{X}}{\bar{Y}}_{RSS}$</th>
<th>Mean of $\frac{\bar{X}}{\bar{Y}}_{SRS}$</th>
<th>Variance of $\frac{\bar{X}}{\bar{Y}}_{SRS}$</th>
<th>Relative efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>2.4837</td>
<td>0.0146</td>
<td>2.4873</td>
<td>0.0177</td>
<td>1.21</td>
</tr>
<tr>
<td>12</td>
<td>2.4815</td>
<td>0.0108</td>
<td>2.4818</td>
<td>0.0126</td>
<td>1.17</td>
</tr>
<tr>
<td>18</td>
<td>2.4786</td>
<td>0.0072</td>
<td>2.4810</td>
<td>0.0082</td>
<td>1.14</td>
</tr>
<tr>
<td>30</td>
<td>2.4767</td>
<td>0.0043</td>
<td>2.4774</td>
<td>0.0048</td>
<td>1.12</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>$n = 3r$</th>
<th>Mean of $\frac{\bar{X}}{\bar{Y}}_{RSS}$</th>
<th>Variance of $\frac{\bar{X}}{\bar{Y}}_{RSS}$</th>
<th>Mean of $\frac{\bar{X}}{\bar{Y}}_{SRS}$</th>
<th>Variance of $\frac{\bar{X}}{\bar{Y}}_{SRS}$</th>
<th>Relative efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>2.4817</td>
<td>0.0094</td>
<td>2.4873</td>
<td>0.0177</td>
<td>1.88</td>
</tr>
<tr>
<td>12</td>
<td>2.4802</td>
<td>0.0070</td>
<td>2.4818</td>
<td>0.0126</td>
<td>1.86</td>
</tr>
<tr>
<td>18</td>
<td>2.4776</td>
<td>0.0047</td>
<td>2.4810</td>
<td>0.0082</td>
<td>1.75</td>
</tr>
<tr>
<td>30</td>
<td>2.4761</td>
<td>0.0028</td>
<td>2.4774</td>
<td>0.0048</td>
<td>1.71</td>
</tr>
</tbody>
</table>

For the data set (the population of 507 height, and weight measurements) considered in this study the actual population ratio of the mean height to the mean weight

$$\left(\frac{\mu_x}{\mu_y}\right) = 2.475.$$
It can be seen from Tables 3 and 4, that RSS gives a better estimate of the above ratio when compared to the SRS estimate. The relative efficiency of RSS based ratio estimate is large when \( n \) is small and this decreases as \( n \) increases although this decrease in efficiency is very marginal. One way to explain this, is that SRS may performs better as the sample size increases and hence behave close to RSS. However, it requires further simulation studies to see the actual effect of the set size and hence the sample size on the efficiency. It is to be pointed out here that as the set size increases the efficiency of the RSS will increase, because the representativeness of the population by RSS increases with the increase of set size, Samawi and Muttlak (1996). But, when the set size is large more sampling units will be discarded and more cost will be incurred. Therefore a balance between these two aspects should be stricken.

Nevertheless, it should be noted that the RSS is used mainly for the purpose of the observational economy. Therefore our focus must be on the behaviour of RSS when the sample size is small. Undoubtedly it is very clear from our case study that the RSS is much more efficient when the sample size is small such as \( n = 9 \) (see Tables 3 and 4).

Comparing the relative efficiencies in the Tables 3 and 4, the relative efficiency values in Table 4, are larger than that in Table 3. This leads to the suggestion of performing ranking on the basis of \( Y \), the covariate appearing in the denominator of the ratio. The estimation of the ratio based on the RSS sample obtained by ranking based on the covariate \( Y \) appears to be better than that of based on \( X \), the actual ranking based on the main variable of interest. Samawi and Muttlak (1996).

This is understandable, for example in a regression context, the covariate \( (Y \text{ in our case}) \) is more controllable (usually fixed by the experimenter and hence can be ranked more exactly) than the response variable \( (X \text{ in our case which is a random outcome}) \). Further, for the population considered in our case study, \( \rho(x, y) = 0.72 \) (a moderate correlation) there is a notable difference in the efficiencies. If \( \rho(x, y) \) is close to one, the both efficiencies will be high and close to each other.

This increased sampling efficiency is achieved by exploiting auxiliary information involving an acquired field sample – a characteristic of double sampling procedure. With RSS however, the auxiliary information does not have to be a quantitative concomitant variable. In fact, it can be purely judgement, and thus, in the spirit of total quality management, (TQM), it stimulates and utilize a productive cross disciplinary dialogue among those responsible for sampling and assessment.
References


Received September, 2005