Ratio Type Estimators of the Population Mean Based on Ranked Set Sampling

Said Ali Al-Hadhrami

Abstract—Ranked set sampling (RSS) was first suggested to increase the efficiency of the population mean. It has been shown that this method is highly beneficial to the estimation based on simple random sampling (SRS). There has been considerable development and many modifications were done on this method. When a concomitant variable is available, ratio estimation based on ranked set sampling was proposed. This ratio estimator is more efficient than that based on SRS. In this paper some ratio type estimators of the population mean based on RSS are suggested. These estimators are found to be more efficient than the estimators of similar form using simple random sample.

Keywords—Bias, Efficiency, Ranked Set Sampling, Ratio Type Estimator

I. INTRODUCTION

RANKED Set Sampling (RSS) was introduced to increase the efficiency of the estimation of population mean [1]. The method is useful when the variable of interest is very expensive or difficult to measure but it can be easily ranked at a negligible cost. The first theoretical results about this method was given in [2]. The method under imperfect ranking was investigated in [3] and [4]. Many modifications and improvements have been given for RSS and becomes well applicable method. For applications see for examples [1], [4], [5], [6], [7], [8], [9], [10], [11], [12].

There are cases in practical situation where the variable of interest Y is difficult to measure and to rank but a concomitant variable X, which is highly correlated with Y, can be easily ranked and be used for the ranking of the sampling units. This idea was first considered by [13]. Some extension is done by [14] utilized both the rank and the measure of the concomitant variable and considered ratio estimation using RSS. The ratio estimation based on RSS is more efficient compared with the SRS ratio estimator.

Let the variable of interest Y and the concomitant variable X is correlated with the coefficient of correlation \( \rho \). The population ratio of these two variable is then \( \hat{R} = \mu_y / \mu_x \) and its estimator is \( \hat{R} = \bar{Y} / \bar{X} \). Where \( \mu_y \) and \( \mu_x \) are the population means of the variables Y and X, respectively, \( \bar{X} \) and \( \bar{Y} \) are the sample mean for \( \mu_y \) and \( \mu_x \) respectively. The ratio estimator is biased but the bias is negligible when the estimator is approximated using Taylor series expansion to the first degree. The approximated Variance of \( \hat{R} \) is

\[
V\text{ar}(\hat{R}) \cong (R^2 / n)(V_y^2 + V_x^2 - 2 \rho \sigma_y \sigma_x)
\]

where \( V_x = \sigma_x / \mu_x \), \( V_y = \sigma_y / \mu_y \), and

\[
\rho = \sum_{i=1}^{N} (X_i - \mu_x)(Y_i - \mu_y) / N \sigma_x \sigma_y
\]

\( \sigma_x \) and \( \sigma_y \) are the standard deviations of the populations of the variables X and Y, respectively.

There are many ratio type estimators based on RSS has been proposed. Some of these estimators are proposed by [15]. These estimators are in the form

\[
\hat{\mu}_{RSS} = \frac{\bar{Y} + \beta (\mu_y - \bar{X})}{(\alpha \bar{X} + \gamma)}(\alpha \mu_x + \gamma)
\]

where \( \beta = \sigma_{xy} / \sigma_x^2 \).

They suggested utilizing some known parameters of the concomitant variable X. If \( \alpha = 1 \) and \( \gamma = 0 \), the estimator be

\[
\hat{\mu}_{RSS1} = \frac{\bar{Y} + \beta (\mu_y - \bar{X})}{\bar{X}} \mu_x
\]

If \( \alpha = 1 \) and \( \gamma = V_x \), then the estimator be

\[
\hat{\mu}_{RSS2} = \frac{\bar{Y} + \beta (\mu_y - \bar{X})}{(\bar{X} + V_x)} (\mu_x + V_x)
\]

where \( V_x \) is the coefficient of variation defined as \( V_x = \sigma_x / \mu_x \).

If \( \alpha = 1 \) and \( \gamma = K_x \), then the estimator be

\[
\hat{\mu}_{RSS3} = \frac{\bar{Y} + \beta (\mu_y - \bar{X})}{(\bar{X} + K_x)} (\mu_x + K_x)
\]

where \( K_x \) is the coefficient of Kurtosis defined as \( K_x = \mu_{4x} / \mu_x^2 \), where \( \mu_{4x} = E(X^4 - \mu_x^4) \).

If \( \alpha = K_x \) and \( \gamma = V_x \), then the estimator be

\[
\hat{\mu}_{RSS4} = \frac{\bar{Y} + \beta (\mu_y - \bar{X})}{(K_x \bar{X} + V_x)} (K_x \mu_x + V_x)
\]

If \( \alpha = V_x \) and \( \gamma = K_x \), then the estimator be

\[
\hat{\mu}_{RSS5} = \frac{\bar{Y} + \beta (\mu_y - \bar{X})}{(V_x \bar{X} + K_x)} (V_x \mu_x + K_x)
\]

The mean square error (MSE) of the above estimators are approximately

Said Ali Al-Hadhrami, assistant professor, college of Applied Sciences, Nizwa. Oman (e-mail: abur1972@yahoo.co.uk).
MSE(\( \hat{\mu}_{RSS} \)) \approx \frac{1 - f}{n} \left[ R_i^2 \sigma_x^2 + \sigma_y^2 (1 - \rho^2)\right]

\text{for}

\text{where } R_1 = \frac{\mu_i}{\mu_x}, R_2 = \frac{\mu_y}{\mu_x + V_x}, R_3 = \frac{\mu_y}{\mu_x + K_x},

R_4 = \frac{\mu_x K_x}{\mu_x K_x + V_x}, \text{ and } R_5 = \frac{V_x}{\mu_x} + K_x.

In this paper we suggest to use similar form of estimators as above based on RSS. We assume that the population mean of the auxiliary variable is known beforehand. For some estimators, we need to know some other parameters such as coefficient of variation and coefficient of kurtosis. We also assume that the relation between X and Y is positive and approximately linear.

II. SAMPLING METHOD

Let Y be the variable of interest and X be a suitable concomitant variable which is correlated to Y and easy to rank. The summary of The RSS procedure is then as following:

1- Select randomly \( m^2 \) bivariate units \( (X, Y) \) from the population.
2- Allocate the chosen units into \( m \) sets each of size \( m \).
3- From the first set, the smallest \( X \) and the associated \( Y \) are measured. From the second set, the second smallest of \( X \) and the associated \( Y \) are measured. We continue in this way until the last set where the largest \( X \) and the associated \( Y \) are measured.
4- Repeat the steps above \( r \) times until getting the required number of elements.

We assume that ranking on the auxiliary variable, X, is perfect. The associated variable, Y, is then with error unless the relation between X and Y is perfect. Let us denote \( (X_{j(1)}, Y_{j(1)}), \ldots, (X_{j(w)}, Y_{j(w)}) \) as the pair of the \( i^{th} \) order statistics of X and the associated element Y in the \( j^{th} \) cycle. Then the ranked set sample is

\( (X_{1(1)}, Y_{1(1)}), \ldots, (X_{1(m)}, Y_{1(m)}), \)

\( (X_{2(1)}, Y_{2(1)}), \ldots, (X_{2(m)}, Y_{2(m)}), \)

\( \vdots \)

\( (X_{r(1)}, Y_{r(1)}), \ldots, (X_{r(m)}, Y_{r(m)}) \)

Then we define the sample means based on RSS

\( \bar{x}^* = \frac{1}{m} \sum_{i=1}^{m} x_{i} \) and \( \bar{y}^* = \frac{1}{m} \sum_{i=1}^{m} y_{i} \) with

variances are \( Var(\bar{x}^*) = \frac{\sigma_x^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_x - \mu_x)^2 \),

\( Var(\bar{y}^*) = \frac{\sigma_y^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_y - \mu_y)^2 \) and

\( Cov(\bar{x}^*, \bar{y}^*) = (1/m) \sigma_{xy} - (1/m^2) \sum_{i=1}^{m} T_{xy[i]} \)

with \( T_{xy[i]} = (\mu_x - \mu_x)(\mu_y - \mu_y) \).

III. RATIO TYPE ESTIMATORS USING RSS

Samawi & Muttlak(1996) proposed a ratio estimator which is based on RSS as \( \hat{\mu} = \bar{y}^*/\bar{x}^* \)

Where \( \bar{x}^* = \frac{1}{mr} \sum_{k=1}^{r} \sum_{i=1}^{m} x_{k(i)} \) and

\( \bar{y}^* = \frac{1}{mr} \sum_{k=1}^{r} \sum_{i=1}^{m} y_{k(i)} \) and the ratio estimator of the population mean of \( Y \) is \( \hat{\mu} = \hat{\mu} \).

Using one degree of Taylor series expansion, they showed that this estimator is more efficient than that from SRS with similar form.

Based on RSS, we suggest ratio-type estimators for the mean in the form

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(\alpha \bar{x}^* + \gamma)} \)

where \( \alpha \) and \( \gamma \) are positive constants, and \( \beta = \sigma_{xy} / \sigma_x^2 \).

Let us take some special cases of this kind of ratio type estimators. If the coefficient of variation, and kurtosis of the concomitant variable are available, we may choose these parameters to be values for \( \alpha \) and \( \gamma \) in the estimator above. For examples:

If \( \alpha = 1 \) and \( \gamma = V_x \), then we have the estimator

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(\bar{x}^* + V_x)} \)

If \( \alpha = 1 \) and \( \gamma = K_x \), then we have the estimator

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(\bar{x}^* + K_x)} \)

If \( \alpha = K_x \) and \( \gamma = V_x \), then we have the estimator

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(K_x \bar{x}^* + V_x)} \)

If \( \alpha = V_x \) and \( \gamma = K_x \), then we have the estimator

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(V_x \bar{x}^* + K_x)} \)

if \( \alpha = 1 \) and \( \gamma = 0 \), then we have the estimator

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{\bar{x}^*} \)

Lemma(1): Let

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(\alpha \bar{x}^* + \gamma)} \)

and

\( \hat{\mu}_{RSS} = \frac{\bar{y}^* + \beta(\mu_x - \bar{x}^*)}{(\alpha \bar{x}^* + \gamma)} \).

Using one degree of Taylor series expansion, then
\[ \text{MSE}(\hat{\mu}_{\text{RSS}}) \leq \text{MSE}(\hat{\mu}_{\text{RSS}}) \]

\textbf{proof}

Using the first order of Taylor series expansion of \( \hat{\mu}_{\text{RSS}} \) about \( \mu_x, \mu_y \), then

\[ \hat{\mu}_{\text{RSS}} \cong \mu_y + (\bar{X} - \mu_x) - D(\bar{X} - \mu_x), \]

where

\[ D = \beta + \alpha \mu_x + (\mu_x + \gamma) \]

and the variance is equal to

\[ \text{Var}(\hat{\mu}_{\text{RSS}}) \cong \text{Var}(\bar{Y}) + D^2 \text{Var}(\bar{X}) \]

Since the bias in this expansion is zero then

\[ \text{MSE}(\hat{\mu}_{\text{RSS}}) \cong \mu_y + (\bar{Y} - \mu_y), \]

Using \( \text{Var}(\bar{X}) = \frac{\sigma^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_{x(i)} - \mu_x)^2 \),

\[ \text{Var}(\bar{Y}) = \frac{\sigma^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_{y(i)} - \mu_y)^2 \]

and

\[ \text{Cov}(\bar{X}, \bar{Y}) = (1/m \sigma_{xy} - (1/m^2) \sum_{i=1}^{m} T_{xy(i)} \](with

\[ T_{xy(i)} = (\mu_{x(i)} - \mu_x)(\mu_{y(i)} - \mu_y) \].

We can write the MSE of \( \hat{\mu}_{\text{RSS}} \) as

\[ \text{MSE}(\hat{\mu}_{\text{RSS}}) \cong \left[ \frac{\sigma^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_{x(i)} - \mu_x)^2 \right] + D^2 \]

\[ + D^2 \left[ \frac{\sigma^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_{y(i)} - \mu_y)^2 \right] \]

\[ - 2D \left[ (1/m \sigma_{xy} - (1/m^2) \sum_{i=1}^{m} T_{xy(i)} \right] \]

Therefore, \( \text{MSE}(\hat{\mu}_{\text{RSS}}) \leq \text{MSE}(\hat{\mu}_{\text{RSS}}) \cdot \)**

Let \( W_{x(i)} = \mu_{x(i)} - \mu_x, W_{y(i)} = \mu_{y(i)} - \mu_y \).

\[ \text{MSE}(\hat{\mu}_{\text{RSS}}) \cong \text{MSE}(\hat{\mu}_{\text{RSS}}) \]

Then

\[ - (1/m^2) \left[ \sum_{i=1}^{m} W_{x(i)}^2 - D W_{x(i)} \right]^2 \]

Therefore, \( \text{MSE}(\hat{\mu}_{\text{RSS}}) \leq \text{MSE}(\hat{\mu}_{\text{RSS}}) \).

The second order bivariate Taylor expansion of \( h(X, Y) \) about \( \mu_x, \mu_y \) is in the form

\[ h(X, Y) \cong h(\mu_x, \mu_y) + (X - \mu_x)h_x + (Y - \mu_y)h_y \]

\[ + \frac{1}{2!}(X - \mu_x)^2 h_{xx} \]

\[ + \frac{1}{2!}(Y - \mu_y)^2 h_{yy} \]

where \( h_x = \frac{\partial h(X, Y)}{\partial X} \|_{\mu_x, \mu_y}, h_y = \frac{\partial h(X, Y)}{\partial Y} \|_{\mu_x, \mu_y} \).

\[ h_{xx} = \frac{\partial^2 h(X, Y)}{\partial X^2} \|_{\mu_x, \mu_y}, h_{yy} = \frac{\partial^2 h(X, Y)}{\partial Y^2} \|_{\mu_x, \mu_y} \]

Using this expansion, we get

\[ \hat{\mu}_{\text{RSS}} \cong \mu_y + (\bar{Y} - \mu_y) \]

Now, the approximated bias of \( \hat{\mu}_{\text{RSS}} \) is

\[ \text{Bias}(\hat{\mu}_{\text{RSS}}) \cong - \frac{1}{2 \alpha \mu_x + \gamma} E \left[ (\bar{X} - \mu_x)(\bar{Y} - \mu_y) \right] \]

\[ + \frac{\alpha(\beta + \alpha \mu_x + \alpha \beta \mu_y)}{(\alpha \mu_x + \gamma)^2} E \left( \bar{X}^2 - \mu_x^2 \right) \]

which can be written as

\[ \text{Bias}(\hat{\mu}_{\text{RSS}}) \cong \frac{\alpha(\beta + \alpha \mu_x + \alpha \beta \mu_y)}{(\alpha \mu_x + \gamma)^2} V a r \left( \bar{X} \right) \]

\[ - \frac{1}{2 \alpha \mu_x + \gamma} \text{Cov}(\bar{X}, \bar{Y}) \]

We have from the theory of RSS that

\[ \text{Var}(\bar{X}) = \frac{\sigma^2}{m} - \frac{1}{m^2} \sum_{i=1}^{m} (\mu_{x(i)} - \mu_x)^2 \]

\[ \text{Cov}(\bar{X}, \bar{Y}) = \frac{\sigma_{xy}}{m} - \frac{1}{m^2} \sum_{i=1}^{m} T_{xy(i)} \]

with \( T_{xy(i)} = (\mu_{x(i)} - \mu_x)(\mu_{y(i)} - \mu_y) \).

Then the bias is
Bias(μ_{RSS}) \approx \frac{\alpha(\beta + \alpha \mu_x + \alpha \beta \mu_y)}{(\mu_x + \gamma)^2} \\
\times \left\{ \frac{\sigma^2}{m} - \frac{1}{m} \sum_{i=1}^{n} (\mu_{(i)} - \mu_x)^2 \right\} \\
+ \frac{1}{2} \alpha \mu_x + \gamma \left\{ \frac{1}{m} \sum_{i=1}^{n} T_{\nu[i]} \right\}

which can be further written as

\text{Bias} (\hat{\mu}_{RSS}) \equiv \left[ \alpha(\beta + \alpha \mu_x + \alpha \beta \mu_y) \right] \left\{ \frac{1}{m} \sum_{i=1}^{n} (\mu_{(i)} - \mu_x)^2 \right\}

Therefore,

\text{Bias} (\hat{\mu}_{RSS}) \equiv \text{Bias} (\hat{\mu}_{RSS}) \\
- \frac{\alpha}{m^{2}} (\mu_x + \gamma) \left\{ \frac{1}{2} \sum_{i=1}^{n} T_{\nu[i]} \right\}

where \hat{\mu}_{RSS} is the ratio-type estimator in the similar form as our estimator based on SRS.

From the last result we may get that \text{Bias} (\hat{\mu}_{RSS}) < \text{Bias} (\hat{\mu}_{RSS}) if

D \sum_{i=1}^{n} (\mu_{(i)} - \mu_x)^2 > \frac{1}{2} \sum_{i=1}^{n} T_{\nu[i]} with \ D = \frac{\alpha(\beta + \alpha \mu_x + \alpha \beta \mu_y)}{\mu_x + \gamma}

To illustrate the amount of bias and efficiency, a simulation study under an example from a bivariate normal distribution is provided in the next section.

IV. NUMERICAL COMPARISON

The behavior of the above estimators is studied and compared with the corresponding estimators from SRS. Let us assume that the variable of interest Y and a concomitant variable X are correlated with a correlation coefficient \( \rho \). Assume also that X and Y have a bivariate normal distribution with parameters \( \mu_x, \mu_y, \sigma_x^2, \sigma_y^2 \), and \( \rho \). These parameter are represented in the following example.

Let us generate samples using SRS and RSS from a bivariate normal distribution with parameters \( \mu_x = 20, \mu_y = 10, \sigma_x^2 = \sigma_y^2 = 1 \) and coefficient of kurtosis of the variable X is \( K_x = 3 \). From this distribution we generated 5000 samples based on RSS and another 5000 samples using SRS. For each sample, the estimate of \( \mu_y \) is calculated. Then the average of \( \hat{\mu}_y \)'s and the mean square error are computed respectively using \( \hat{\mu}(\hat{\mu}_y) = (1/5000) \sum_{i=1}^{5000} \hat{\mu}_y \) and \( \text{MSE} \hat{\mu}(\hat{\mu}_y) = (1/5000) \sum_{i=1}^{5000} (\hat{\mu}_y - \mu_y)^2 \). Different values of \( \rho \) and \( m \) are used and the results are shown in Tables I and II.

<table>
<thead>
<tr>
<th>Table I</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
</tr>
<tr>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
</tr>
<tr>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

From Tables I and II the following concluding can be derived:

1. The bias of the estimators are small and is independent of \( \rho \). The fluctuation is due to simulation error.
2. The RSS estimators considered are more efficient then the estimators based on SRS of similar forms.
3. The efficiency of RSS estimators decreases as the correlation coefficient decreases.
4. The efficiency is increasing as the set size \( m \) is increasing.
REFERENCES


