Ratio-type Estimator of Population Mean Based on Ranked Set Sampling

Zahid Khan
Department of Statistics, COMSATS University Islamabad, Lahore Campus
Pakistan
zahid.uom1@gmail.com

Muhammad Ismail
Department of Statistics, COMSATS University Islamabad, Lahore Campus
 Pakistan
drismail39@gmail.com

Abstract

In this paper, a generalized ratio-type estimator based on ranked set sampling (RSS) is proposed for estimating population mean using the known population parameters of auxiliary variable. The mean square error (MSE) of the proposed estimator is derived and compared with some existing ratio estimators under RSS. It is proved that MSC of proposed ratio estimator is less than MSC of some existing ratio estimators based on RSS under some conditions. The MSE of proposed estimators along with some existing estimators are also calculated numerically which showed that the proposed ratio estimators is more efficient than some existing ratio estimators under RSS.

Keywords; ratio estimator, rank set sampling, efficiency, mean square error.

1. Introduction

McIntyre (1952) was the first who introduced rank set sampling (RSS) to estimate the population mean. Takahasi and Wakimoto (1968) proved that sample mean under RSS is unbiased estimator of population mean. Dell and Clutter (1972) showed that RSS-based sample mean is not only unbiased estimator of population mean but also it is at least as efficient as the sample mean with SRS, regardless of whether the ranking is perfect or not. Stokes (1977) demonstrated the case where judgment ranking of variable of interest Y is difficult but it can be easily ranked by concomitant variable X. Samawi and Muttlak (1996) defined the ratio estimator of population mean under RSS design. The ratio estimator is modified by many researchers, see Mehta and Mandowara (2012); Kadilar and Cingi (2007); Jeelani and Bouza (2015); Ismail et al., (2015); Mehta an Mandowara (2016); Khan and Shabbir (2016). Procedure of RSS can be briefly defined as, select m^2 simple random samples from the target population, then, distribute the selected units into m sets each of size m units. Rank the units within each set by eye or by other economical method. Finally, select i^{th} ranked unit from i^{th} set for actual measurement, for $i = 1,2,\dots m$. In a result, m ranked set sample are obtained. In order to select more samples the procedure of RSS is repeated r times, which yield rm identified units.

The sample mean of X and Y under RSS is define as;

$$\bar{X}_{(RSS)} = \frac{1}{rm}\left(\sum_{i=1}^{m}\sum_{j=1}^{r} X_{i(i:m)j}\right) \tag{1}$$

$$\bar{Y}_{(RSS)} = \frac{1}{rm}\left(\sum_{i=1}^{m}\sum_{j=1}^{r} Y_{i(i:m)j}\right) \tag{2}$$

With variance

$$Var\left(\bar{X}_{(RSS)}\right) = \frac{\sigma^2}{rm} - \frac{1}{rm^2}\sum_{i=1}^{m}\tau_{x(i)}^2 \tag{3}$$

$$Var\left(\bar{Y}_{(RSS)}\right) = \frac{\sigma^2}{rm} - \frac{1}{rm^2}\sum_{i=1}^{n}\tau_{y(i)}^2 \tag{4}$$

and covariance

$$Cov\left(\bar{X}_{(RSS)}, \bar{Y}_{(RSS)}\right) = \frac{\sigma^2}{rm} - \frac{1}{rm^2}\sum_{i=1}^{m}\mu_{xy(i)} \tag{5}$$

Where $\tau_{y(i)} = \left(\mu_{y(i)} - \bar{Y}\right)$, $\tau_{x(i)} = \left(\mu_{x(i)} - \bar{X}\right)$, $\mu_{xy(i)} = \left(\mu_{x(i)} - \bar{X}\right)\left(\mu_{y(i)} - \bar{Y}\right)$

Samawi and Muttalak (1996) define the ratio estimator under RSS is as follows;

$$\bar{y}_R{}^{rss1} = \bar{y}_{(rss)}\frac{\bar{X}}{\bar{x}_{(rss)}} \tag{6}$$

With MSC

$$MSC\left(\bar{y}_R{}^{rss1}\right) \cong \bar{Y}^2\left[\theta\{C_y^2 + C_x^2 - 2\rho_{xy}C_x C_y\} - \{W_{y(i)}^2 + W_{x(i)}^2 - 2W_{yx(i)}\}\right] \tag{7}$$

Where

$$\theta = \frac{1}{rm}, C_y^2 = \frac{s_y^2}{\bar{Y}^2}, C_x^2 = \frac{s_x^2}{\bar{X}^2}, W_{y(i)}^2 = \frac{1}{rm^2}\frac{1}{\bar{Y}^2}\sum_{i=1}^{m}\tau_{y(i)}^2, W_{x[i]}^2 = \frac{1}{m^2}\frac{1}{\bar{X}^2}\sum_{i=1}^{m}\tau_{x(i)}^2$$

Mehta and Mandowara (2016) proposed a modified ratio estimator under ranked set sampling is as follows,

$$\bar{y}_R{}^{rss2} = \bar{y}_{(rss)}\left[\frac{\bar{X}+C_x}{\bar{x}_{(rss)}+C_x}\right] \tag{8}$$

With MSC

$$MSC\left(\bar{y}_R{}^{rss2}\right) \cong \bar{Y}^2\left[\theta\{C_y^2 + \lambda_m{}^2 C_x^2 - 2\lambda_m\rho_{xy}C_x C_y\} - \{W_{y(i)}^2 + \lambda_m{}^2 W_{x(i)}^2 - 2\lambda_m W_{yx(i)}\}\right] \tag{9}$$

Where

$$\lambda_m = \frac{\bar{X}}{\bar{X} + C_x}$$

In the present study, we suggest a generalized ratio-type estimator of the population mean using known value of coefficient of variation, coefficient of skewness and coefficient of kurtosis of auxiliary variable under rank set sampling.

2. The proposed estimator

Motivated by Yan and Tain (2010), the proposed generalized ratio-type estimator of $\bar{Y}$ under RSS is as follows,

$$\bar{y}_R{}^{m,rss} = \bar{y}_{(rss)}\left[\frac{a\bar{X}+b}{a\bar{x}_{(rss)}+b}\right] \tag{10}$$

Where a and b are the known value of population parameters; $\bar{y}_{(rss)}$ and $\bar{x}_{(rss)}$ are the sample mean of study variable and auxiliary variable respectively, and $\bar{X}$ denotes the population mean of auxiliary variable.

One can easily obtain different estimators of population mean by choosing different values of a and b. For example,

   1. For $a = 1$, $b = \beta_1$, $\bar{y}_R{}^{m,rss}$ reduced to $\bar{y}_R{}^{m,rss1} = \bar{y}_{(rss)}\left[\frac{\bar{X}+\beta_1}{\bar{x}_{(rss)}+\beta_1}\right]$

2. For $a = \beta_2$, $b = \beta_1$, $\bar{y}_R{}^{m,rss}$ reduced to $\bar{y}_R{}^{m,rss2} = \bar{y}_{(rss)}\left[\dfrac{\beta_2\bar{X}+\beta_1}{\beta_2\bar{x}_{(rss)}+\beta_1}\right]$.

3. For $a = C_x$, $b = \beta_1$, $\bar{y}_R{}^{m,rss}$ reduced to $\bar{y}_R{}^{m,rss3} = \bar{y}_{(rss)}\left[\dfrac{C_x\bar{X}+\beta_1}{C_x\bar{x}_{(rss)}+\beta_1}\right]$

To get bias and MSC of $\bar{y}_R{}^{m,rss}$ we put $\bar{y}_{(rss)} = \bar{Y}(1 + e_0)$ and $\bar{x}_{(rss)} = \bar{X}(1 + e_1)$, so that $E(e_0) = E(e_1) = 0$

$V(e_0) = E(e_0^2) = \dfrac{V(\bar{y}_{(rss)})}{\bar{Y}^2} = \dfrac{1}{m}\dfrac{1}{\bar{Y}^2}\left[s_y^2 - \dfrac{1}{rm}\sum_{i=1}^{m}\tau_{y(i)}^2\right] = \left[\theta C_y^2 - W_{y(i)}^2\right]$

Similarly,

$V(e_1) = E(e_1^2) = \left[\theta C_x^2 - W_{x(i)}^2\right]$

$E(e_0e_1) = cov(e_0e_1) = \left[\theta\rho_{xy}C_x C_y - W_{yx(i)}\right]$

The bias and MSC of $\bar{y}_R{}^{rss}$ can be found as follows;

$$B(\bar{y}_R{}^{m,rss}) = E(\bar{y}_R{}^{rss}) - \bar{Y}$$

Here $\bar{y}_R{}^{m,rss} = \bar{Y}(1 + e_0)(1 + \lambda e_1)^{-1}$

Where,

$$\lambda = \dfrac{-a\bar{X}}{a\bar{X} + b}$$

Now expended $(1 + \lambda e_1)^{-1}$ as

$$\bar{y}_R{}^{m,rss} = \bar{Y}(1 + e_0)(1 - \lambda e_1 + \lambda^2 e_1{}^2 + \cdots)$$

Using Taylor series expansion for obtaining bias of $\bar{y}_R{}^{m,rss}$ as,

$B(\bar{y}_R{}^{m,rss}) = \bar{Y}\left[\lambda^2 E(e_1{}^2) - \lambda E(e_0e_1)\right]$

$\qquad = \bar{Y}\left[\lambda^2\{\theta C_x^2 - W_{x(i)}^2\} - \lambda\{\theta\rho_{xy}C_x C_y - W_{yx(i)}\right]$

$\qquad = \bar{Y}\left[\theta\{\lambda^2 C_x^2 - \lambda\rho_{xy}C_x C_y\} - \{\lambda^2 W_{y(i)}^2 - \lambda W_{yx(i)}\}\right]$ $\hfill (11)$

Now

$MSC(\bar{y}_R{}^{m,rss}) = E(\bar{y}_R{}^{m,rss} - \bar{Y})^2$

$\qquad \cong E(e_0 - \lambda e_1)^2$

$\qquad \cong \bar{Y}^2 E\left[e_0{}^2 + \lambda^2 e_1{}^2 - 2\lambda e_0 e_1\right]$

$\qquad \cong \bar{Y}^2\left[E(e_0{}^2) + \lambda^2 E(e_1{}^2) - 2\lambda E(e_0e_1)\right]$

$\qquad \cong \bar{Y}^2\left[\theta C_y^2 - W_{y(i)}^2 + \lambda^2(\theta C_x^2 - W_{x(i)}^2) - 2\lambda(\theta\rho_{xy}C_x C_y - W_{yx(i)})\right]$

$MSC(\bar{y}_R{}^{m,rss}) \cong \bar{Y}^2\left[\theta\{C_y^2 + \lambda^2 C_x^2 - 2\lambda\rho_{xy}C_x C_y\} - \{W_{y(i)}^2 + \lambda^2 W_{x(i)}^2 - 2\lambda W_{yx(i)}\}\right]$ $\hfill (12)$

3. Efficiency comparison

On comparing (7) and (9) with (12), we obtain;

$$MSC(\bar{y}_R{}^{m,rss}) < MSC(\bar{y}_R{}^{rss1})$$

$\bar{Y}^2\left[\theta\{C_y^2 + \lambda^2 C_x^2 - 2\lambda\rho_{xy}C_x C_y\} - \{W_{y[i]}^2 + \lambda^2 W_{x(i)}^2 - 2\lambda W_{yx(i)}\}\right]$

$\qquad < \bar{Y}^2\left[\theta\{C_y^2 + C_x^2 - 2\rho_{xy}C_x C_y\} - \{W_{y(i)}^2 + W_{x(i)}^2 - 2W_{yx(i)}\}\right]$

$$(1 + \lambda)(C_x^2 - W_{x[i]}^2) > 2(\rho_{xy}C_x C_y + W_{yx(i)}) \hfill (13)$$

and

$$MSC(\bar{y}_R{}^{m,rss}) < MSC(\bar{y}_R{}^{rss2})$$

$$\overline{Y}^2\big[\theta\{C_y^2 + \lambda^2 C_x^2 - 2\lambda\rho_{xy}C_x\, C_y\} - \{W_{y(i)}^2 + \lambda^2 W_{x(i)}^2 - 2\lambda W_{yx(i)}\}\big]$$
$$< \overline{Y}^2\big[\theta\{C_y^2 + \lambda_m^{\;2}C_x^2 - 2\rho_{xy}C_x\, C_y\} - \{W_{y(i)}^2 + \lambda_i^{\;2}W_{x(i)}^2 - 2\lambda_m W_{yx(i)}\}\big]$$
$$2\big(\rho_{xy}C_x\, C_y + W_{yx(i)}\big) > (\lambda - \lambda_m)\big(\theta C_x^2 - W_{x(i)}^2\big) \tag{14}$$

Thus when (13) and (14) satisfied then the suggested estimators are more efficient than $\overline{y}_R^{\;m,rss1}$ and $\overline{y}_R^{\;m,rss2}$.

4. Numerical Example

We consider the data of Cochran (1977) for practical application of the proposed estimators. The data was collected on the total number of inhabitants (in 1000's) in 49 cities during 1920 and 1930. The example considers the data of inhabitants in 1930 as the study variable Y, and inhabitant in 1920 is auxiliary variable X. The parameters of this population are presented in table 1.

From inhabitant of 49 cities, for both of study variable and auxiliary variable, we take 16 simple random samples each, and then it is distributed into 4 sets each of size 4. After ranking the data within each set, finally i^{th} ranked unit from i^{th} set is drawn, which yield $m = 4$ ranked set samples. This process is repeated again and $m = 4$ units are drawn once more. In this way, 8 ranked set samples are selected.

Table 2 shows the estimated MSE's discussed in our study. The table reveals that the MSC of proposed estimators are smaller than MSE of Samawi and Muttalak (1996), and Mehta and Mandowara (2016).

Table 1: Data Statistics

$N = 49$	$\overline{X} = 103.1428571$	$\mu_{2(x)} = 42$	$\mu_{3(y)} = 93.5$
$n = 8$	$\overline{Y} = 127.7959$	$\mu_{3(x)} = 76$	$\mu_{4(y)} = 103.5$
$m = 4$	$s_y = 32.8720$	$\mu_{4(x)} = 144$	$\beta_{1(x)} = 2.20$
$r = 2$	$s_x = 56.2342$	$\mu_{1(y)} = 70$	$\beta_{2(x)} = 7.22$
$\rho = 0.18$	$\mu_{1(x)} = 70$	$\mu_{2(y)} = 87$	

Table 2: MSC value of ratio estimators

Estimators	MSC
$\overline{y}_R^{\;rss1}$	376.7045
$\overline{y}_R^{\;rss2}$	367.5767
$\overline{y}_R^{\;m,rss1}$	359.1788
$\overline{y}_R^{\;m,rss2}$	364.2033
$\overline{y}_R^{\;m,rss3}$	359.6667

5. Conclusion

The study concludes that using RSS in ratio estimator of population mean improve the efficiency of the estimator. Moreover, the proposed ratio estimators can be extended to stratified random sampling.

448

Pak.j.stat.oper.res. Vol.XV No.2 2019 pp445-449

References

1. Cochran, W. G. (1977). Sampling techniques. Wiley, New York.
2. Dell, T. R. and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. Biometrics, 28: 545–553.
3. Ismail, M., Hanif, Shahbaz. M. Q. (2015). Generalized Estimators for Population Mean in the Presence of Non-Response for Two-Phase Sampling. Pak. J. Statist, 31(3): 295-306.
4. Jeelani, M. I. and Bouza, C. N. (2015). New ratio method of estimation under ranked set sampling. Revista Investigacion Operacional, 36(2):151-156.
5. Kadilar, C., Unyazici, Y. and Cingi, H. (2009). Ratio estimator for the population mean using ranked set sampling. Statistical Papers, 50(2): 301-309.
6. Khan, L., and Shabbir, J. (2016). Improved ratio-type estimators of population mean in ranked set sampling using two concomitant variables. Pakistan Journal of Statistics and Operation Research, 12(3): 507-518.
7. Mclntyre, G. A. (1952). A method for unbiased selective sampling using ranked sets. Australian Journal of Agricultural Research, 3:385–390.
8. Mehta, N. and Mandowara, V, L. (2012). A better estimator of population mean with power transformation based on ranked set sampling. Sampling methods and estimation, 13(3):551-558.
9. Mehta, N. and Mandowara, V. L. (2016). A modified ratio-cum-product estimator of finite population mean using ranked set sampling. Communications in Statistics-Theory and Methods, 45(2):267-276.
10. Samawi, H. M. and Muttlak, H. A. (1996). Estimation of ratio using rank set sampling. Biometrical Journal, 38(6):753-764.
11. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. Annals of the institute of statistical mathematics, 20(1): 1-31.
12. Yan, Z. and Tian, B. (2010, October). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. In International Conference on Information Computing and Applications (pp. 103-110). Springer, Berlin, Heidelberg.