Applied Mathematics & Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/amisl/090101

A New form of Ratio Estimator under Rank set Sampling

Rizwan Yousuf, M. Iqbal Jeelani Bhat, Zakir Hussain Wani and Khalid ul Islam*

Division of Statistics and Computer Science SKUAST-JAMMU, 180009. India

Received: 5 Sep. 2020, Revised: 2 Nov. 2020, Accepted: 15 Dec. 2020.
Published online: 1 Jan. 2021.

Abstract: A new method of sampling suggested, namely ranked set sampling (RSS) is efficient methodology as compared to simple random sampling (SRS) technique. Initially this methodology was proposed by McIntyre (1952) in the context of estimating population mean and pastures yields. In the last few years several authors suggested different modifications of the (RSS). When observations are costly or time-consuming but the ranking of the observations without actual measurement can be done relatively easily, ranked set sampling (RSS) can be employed instead of simple random sampling (SRS) to gain more information.Its efficiency for estimating the higher population moments is better than that of simple random sampling (SRS).It is mainly used under the condition that the measurement of units are difficult or time consuming but the ranking of the units by some methods other than actual measurements (e.g. visual inspection) is relatively easy. In this paper a new form of ratio estimator has been proposed under rank set sampling by adapting the ratio estimator proposed by Jeelani *et al.* (2013) under simple random sampling. It has been seen that the proposed estimator is more efficient than the estimator suggested by Jeelani *et al.* (2013) in all situations. The results are supported by empirical study also.

Keywords: Rank set sampling, Simple random sampling, Ratio estimators, Co-efficient of kurtosis, Quartile deviation.

1 Introduction

Ratio method of estimation utilizes information on auxiliary variable in order to improve precision of the estimate of population mean. In many surveys, information on auxiliary variate which is highly correlated with variable of interest is readily available and can be used for improving the precision of sampling design.

The classical ratio estimator for population mean $\bar{Y}$ of the study variable is defined as $\bar{y}_r = {}^{\bar{y}}/_{\bar{x}}\bar{X} = \hat{R}\bar{X}$ assuming that population mean of the auxiliary variable is known, $\bar{y}$ and $\bar{x}$ are sample means of study variable and auxiliary variable, and the mean square of the classical ratio estimator is $(MSE(\bar{y}_r) = \Delta\bar{Y}^2[C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x])$. Where, $\Delta = {}^1/_n$(ignoring sampling fraction), $\hat{R}$ is population ratio, ρ_{yx} is correlation coefficient between study variable and auxiliary variable, $C_y$and C_x are coefficient of variations of study variable and auxiliary variable.

Recalling the ratio estimator proposed by Jeelani *et al.* (2013) under simple random sampling, where linear combination of β_2 (Coefficient of Kurtosis) and Qd (Quartile deviation) was used, which is given as follows;

$$\bar{y}_{\omega srs} = \omega\, {}^{\bar{y}}/_{\bar{x}}\bar{X} = \omega\,\hat{R}\bar{X} \qquad (a)$$

where,$\omega = {}^{\bar{X}\beta_2 + Qd}/_{\bar{x}\beta_2 + Qd}$

$$\beta_2 = {}^{N(N+1)\sum_{i=1}^{N}(X_i - \bar{X})^4}/_{(N-1)(N-2)S^3} - {}^{3(N-1)^2}/_{(N-2)(N-3)},$$

$\text{Qd} = \text{Q}_3 - \text{Q}_1, Q_3 = 3\,{}^{(N-1)}/_4, Q_1 = {}^{(N-1)}/_4.$

The mean squared error of the above estimator is given below;

$$MSE(\bar{y}_{\omega srs}) = \Delta\bar{Y}^2[C_y^2 + \omega^2C_x^2 - 2\omega\rho_{yx}C_yC_x] \quad (b)$$

If $\omega = 1$ then $MSE(\bar{y}_{\omega srs}) = (MSE(\bar{y}_r)$ Jeelani *et al.* (2013) shows that this estimator when compared with classical ratio estimator under simple random sampling proves to be more efficient.

2 Proposed Estimator

The basic premise for Rank set sampling (RSS) is an infinite population under study and the assumption that a set of sampling units drawn from the population can be ranked by certain means rather cheaply without the actual measurement of the variable of interest which is costly and/or time-consuming. This assumption may look rather

*Corresponding author e-mail:rizwanwar50@gmail.com

restrictive at first sight, but it turns out that there are plenty of situations in practice where this is satisfied. The original form of RSS conceived by McIntyre (1952) can be described as follows. First, a simple random sample of size *k* is drawn from the population and the *k* sampling units are ranked with respect to the variable of interest, say X, by judgment without actual measurement. Then the unit with rank 1 is identified and taken for the measurement of *X*. 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 of *X* 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 of *X*. This whole process is referred to as a cycle. The cycle then repeats *m* times and yields a ranked set sample of size *n* = *mk.* In recent past a lot of research has been done in RSS by Jeelani *et al.* (2013, 2014a, 2014b, 2014c). The essence of RSS is conceptually similar to the classical stratified sampling. RSS can be considered as post-stratifying the sampling units according to their ranks in a sample. Although the mechanism is different from the stratified sampling, the effect is the same in that the population is divided into sub-populations such that the units within each sub-population are as homogeneous as possible. In fact, we can consider any mechanism, not necessarily ranking the units according to their *X* values, which can post-stratify the sampling units in such a way that it does not result in a random permutation of the units. The mechanism will then have similar effect to the ranking mechanism considered above.

The classical ratio estimator under rank set sampling proposed by Samawi and Muttalak (1996) is given below;

$$\hat{\bar{y}}_{rss} = \frac{\bar{y}_{rss}}{\bar{x}_{rss}} \qquad (c)$$

where $\bar{y}_{rss} = \frac{1}{mk}\sum_{i=1}^{k} Y_i$ and $\bar{x}_{rss} = \frac{1}{mk}\sum_{i=1}^{k} X_i$, also assuming the population mean of the auxiliary variable is known the equation (c) changes to ;

$$\hat{\bar{y}}_{rss} = \frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} \qquad (d)$$

The mean square error of this estimator is given by;

$$MSE(\hat{\bar{y}}_{rss}) = \bar{Y}\left[\Delta\{C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\} - \{Z_{y[i]} - Z_{x(i)}\}^2\right] \qquad (e)$$

where, $\Delta = 1/_{mk}$, $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_X}{\bar{X}}$, $\rho_{yx} = \frac{S_{yx}}{S_yS_x}$, $S_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N-1}}$, $S_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{X})^2}{N-1}}$,

$S_{yx} = \sqrt{\frac{\sum(y_i-\bar{Y})(x_i-\bar{X})}{N-1}}$, $Z_{x(i)}^2 = \frac{1}{k^2m}\frac{1}{\bar{X}^2}\sum_{i=1}^{k}\pi_{x(i)}^2$,

$Z_{y[i]}^2 = \frac{1}{k^2m}\frac{1}{\bar{X}^2}\sum_{i=1}^{k}\pi_{y(i)}^2$, $Z_{yx(i)} = \frac{1}{k^2m}\frac{1}{\bar{Y}\bar{X}}\sum_{i=1}^{k}\pi_{yx(i)}$, $\pi_{x(i)} = (\mu_{x(i)} - \bar{X})$, $\pi_{y[i]} = (\mu_{y[i]} - \bar{Y})$ and $\pi_{yx(i)} = (\mu_{x(i)} - \bar{X})(\mu_{y[i]} - \bar{Y})$

ρ_{yx} is correlation coefficient between study variable and auxiliary variable, C_y and C_x are coefficient of variations of study variable and auxiliary variable.

Now if we incorporate the ratio estimator of equation (a) proposed by Jeelani *et al.* (2013) in equation (d) given by Samawi and Muttalak (1996), then we will have a following new type of ratio estimator under rank set sampling, which is given below;

$$\hat{\bar{y}}_{\omega rss} = \omega\frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} = \hat{R}_{\omega rss}\bar{X} \qquad (f)$$

Then the mean square of $\hat{R}_{\omega rss}$ will follow as ;

$$MSE(\hat{R}_{\omega rss}) = E(\hat{R}_{\omega rss} - R)^2 \qquad (g)$$

$$\because MSE(X) = E(\bar{x} - \mu)^2$$

$$MSE(\hat{R}_{\omega rss}) = E(\hat{R}_{\omega rss} - R)^2 = E\left(\omega\frac{\bar{y}_{rss}}{\bar{x}_{rss}} - R\right)^2 = E\left(\frac{\omega\bar{y}_{rss} - R\bar{x}_{rss}}{\bar{x}_{rss}}\right)^2$$

Applying Taylor series approximation, we have

$$MSE(\hat{R}_{\omega rss}) \cong E\frac{1}{\bar{X}^2}(\omega\bar{y}_{rss} - R\bar{x}_{rss})^2$$

$\because (1/_{\bar{x}_{rss}} = 1/_{\bar{X} + \bar{x}_{rss} - \bar{X}} = 1/_{\bar{X}}\left(1 + {\bar{x}_{rss} - \bar{X}}/_{\bar{X}}\right)^{-1} \cong 1/_{\bar{X}})$ as per Wolter (1985).

$$MSE(\hat{R}_{\omega rss}) \cong \frac{1}{\bar{X}^2}E\{(\omega\bar{y}_{rss} - \bar{Y}) - R(\bar{x}_{rss} - \bar{X})\}^2$$

$$\cong \frac{1}{\bar{X}^2}\{E(\omega\bar{y}_{rss} - \bar{Y})^2 - 2RE(\omega\bar{y}_{rss} - \bar{Y})(\bar{x}_{rss} - \bar{X}) + R^2E(\bar{x}_{rss} - \bar{X})^2\}$$

$$\because (a-b)^2 = a^2 - 2ab + b^2$$

$$= \frac{1}{\bar{X}^2}\{\omega^2E(\bar{y}_{rss})^2 - 2\bar{Y}\omega E(\bar{y}_{rss}) + \bar{Y}^2 + R^2Var(\bar{x}_{rss}) - 2R[\omega E(\bar{y}_{rss}\bar{x}_{rss}) - \bar{Y}E(\bar{x}_{rss}) + \bar{Y}\bar{X}]\}$$

$\because E(CX) = cE(X), Var(CX) = C^2Var(X)$, if C is constant

$$= \frac{1}{\bar{X}^2}\{\omega^2[var(\bar{y}_{rss}) + \bar{Y}^2] + \bar{Y}^2(1-2\omega) + R^2Var(\bar{x}_{rss}) - 2R\omega cov(\bar{y}_{rss}\bar{x}_{rss})\}$$

$$= \frac{1}{\bar{X}^2}\{\omega^2var(\bar{y}_{rss}) + \bar{Y}^2(1-2\omega)^2 + R^2Var(\bar{x}_{rss}) - 2R\omega cov(\bar{y}_{rss}\bar{x}_{rss})\}$$

Now the mean square error of $\hat{\bar{y}}_{\omega rss}$ will be ;

$$MSE(\hat{\bar{y}}_{\omega rss}) \cong \{\omega^2var(\bar{y}_{rss}) + \bar{Y}^2(1-2\omega)^2 + R^2Var(\bar{x}_{rss}) - 2R\omega cov(\bar{y}_{rss}\bar{x}_{rss})\}$$

Where $var(\bar{y}_{rss}) = 1/_{km}(\sigma_y^2 - 1/_k\sum_{i=1}^{k}\pi_{y[i]}^2)$, $var(\bar{x}_{rss}) = 1/_{km}(\sigma_x^2 - 1/_k\sum_{i=1}^{k}\pi_{x(i)}^2)$, $cov(\bar{y}_{rss}\bar{x}_{rss}) = 1/_{km}(\sigma_{yx}^2 - 1/_k\sum_{i=1}^{k}\pi_{yx(i)}^2)$, $\pi_{x(i)} = (\mu_{x(i)} - \bar{X})$, $\pi_{y[i]} = (\mu_{y[i]} - \bar{Y})$ and $\pi_{yx(i)} = (\mu_{x(i)} - \bar{X})(\mu_{y[i]} - \bar{Y})$

$$\therefore MSE(\hat{\bar{y}}_{\omega rss}) \cong \frac{1}{km}(\omega^2\sigma_y^2 - R\omega\sigma_{yx} + R^2\sigma_x^2) + \bar{Y}^2(\omega-1)^2 - \frac{1}{k^2m}\left(\omega^2\sum_{i=1}^{k}\pi_{y[i]}^2 - 2R\omega\sum_{i=1}^{k}\pi_{yx(i)}^2 + R^2\sum_{i=1}^{k}\pi_{x(i)}^2\right)$$

Let us suppose $\varphi = \frac{1}{k^2m}(\omega^2\sum_{i=1}^{k}\pi_{y[i]}^2 - 2R\omega\sum_{i=1}^{k}\pi_{yx(i)}^2 + R^2\sum_{i=1}^{k}\pi_{x(i)}^2)$, where φ is non-negative value, then $MSE(\hat{\bar{y}}_{\omega rss}) \cong MSE(\bar{y}_{jsrs}) - \varphi$. From this condition it can easily be seen that the MSE of proposed estimator is less than the estimator proposed by Jeelani *et al.* (2013)., provided $k > 1$. The bias of the proposed estimator is given below;

$\bar{y}_{rss} = \bar{Y}(1+\eta_0)$ and $\bar{x}_{rss} = \bar{X}(1+\eta_1)$, such that the expectation of η_0 and η_1 is equal to zero

$E(\eta_0) = E(\eta_1) = 0$

$$V(\eta_0) = E(\eta_0{}^2) = \frac{V(\bar{y}_{rss})}{\bar{Y}^2} = \frac{1}{km}\frac{1}{\bar{Y}^2}\left[\frac{1}{k}\sum_{i=1}^{k}\pi_{y[i]}^2\right] = \left[\Delta C_y^2 - Z_{y[i]}^2\right]$$

Similarly

$V(\eta_1) = E(\eta_1{}^2) = \left[\Delta C_x^2 - Z_{x(i)}^2\right]$

And Cov (η_0,η_1)=E (ε_0,ε_1)= Cov(y ̅^* x ̅^*)/(Y ̅X ̅) = 1/(Y ̅X ̅) 1/km [σ_yx-1/k ∑_(i=1)^k▒π_(yx(i))]=[ΔρC_yC_x-Z_(yx(i))]

where, $\Delta = \frac{1}{km}, C_y^2 = \frac{\sigma_y^2}{\bar{Y}^2}, C_x^2 = \frac{\sigma_x^2}{\bar{X}^2}, C_{yx} = \frac{\sigma_{yx}}{\bar{X}\bar{Y}} = \rho C_y C_x, Z_{x(i)}^2 = \frac{1}{k^2m}\frac{1}{\bar{X}^2}\sum_{i=1}^{k}\pi_{x(i)}^2, Z_{y[i]}^2 = \frac{1}{k^2m}\frac{1}{\bar{X}^2}\sum_{i=1}^{k}\pi_{y(i)}^2, Z_{yx(i)} = \frac{1}{k^2m}\frac{1}{\bar{Y}\bar{X}}\sum_{i=1}^{k}\pi_{yx(i)}$, $\pi_{x(i)} = (\mu_{x(i)} - \bar{X}), \pi_{y[i]} = (\mu_{y[i]} - \bar{Y})$ and $\pi_{yx(i)} = (\mu_{x(i)} - \bar{X})(\mu_{y[i]} - \bar{Y})$

also, $\hat{\bar{y}}_{\omega rss} = \omega\frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} = \omega\frac{\bar{Y}(1+\eta_0)}{\bar{X}(1+\eta_1)}\bar{X} = \omega\bar{Y}(1+\eta_0)(1+\eta_1)^{-1}$

Taking expectations on both sides the bias of $\hat{\bar{y}}_{\omega rss}$ will be (See, Wolter (1985) ;

$$Bais\ (\hat{\bar{y}}_{\omega rss}) \cong (\omega-1)\bar{Y} - \omega\bar{Y}(\Delta C_x^2 - Z_{x(i)}^2 - \Delta C_{yx} + Z_{yx(i)})$$

3 Empirical Study

In this section we use the data of Jeelani *et al.* (2014). The data was collected on Apple production from district Ganderbal of Kashmir valley from 420 orchards in 30 villages. The variables chosen for the study, where Yield (MT), Area (ha). The summary statistics of the data is given in Table 1 below. We take equal samples for each ratio estimator for the sake of comparison. The sample sizes considered were 60. The set size for rank set ratio estimator is $k = 6$ and number of cycles $m = 10$ that is ($n = km = 6 \times 10 = 60$. From Table 2it is clear that the proposed estimator has less MSE as compared to others. Also simulation study has been carried out by taking three combinations of set sizes and three combinations of correlation coefficients, for efficiency comparison in Table

Table 1: Parameters of the study.

$\bar{Y}$	23.10	σ_y	33.42
$\bar{X}$	43.81	σ_x	2251.00
N	420	C_y	1.49
n	60	C_x	1.66
k	6	β_2	13.10
m	10	Qd	1056
ρ	0.90		

Table 2: Mean square errors of estimators.

Ratio Estimators	MSE values
$MSE(\bar{y}_r)$ classical ratio estimator under SRS	272.10
$MSE(\bar{y}_{\omega srs})$ Jeelani et al. (2013)	230.11
$MSE(\hat{\bar{y}}_{rss})$ classical ratio estimator under RSS	168.10
$MSE(\hat{\bar{y}}_{\omega rss})$ Suggested ratio estimator.	105.08

Table 3: Efficiency comparison of estimators.

	$(\bar{y}_r)$	$(\bar{y}_{\omega srs})$	$(\bar{y}_{rss})$	$(\bar{y}_{\omega rss})$	k(set size)	m (no.of cycles)	$n = km$ (Sample size)
0.95	1.52	1.7	1.88	2.17	3	10	30
	1.58	2.01	2.04	2.39	6		60
	1.27	1.37	1.61	1.83	8		80
0.75	1.29	1.46	1.55	1.57	3		30
	1.38	1.62	1.71	1.75	6		60
	1.17	1.34	1.44	1.56	8		80
0.65	1.25	1.41	1.48	1.53	3		30
	1.25	1.53	1.62	1.67	6		60
	1.12	1.19	1.35	1.44	8		80

4 Conclusions

Hence it is concluded that RSS procedures are more attractive than its counter parts as they increase the efficiency of population mean, RSS has its practical implications. These benefits of RSS need not be restricted to simple random sampling, replacing SRS with RSS in the final stage of any survey design ort method of estimation in surveys will greatly improve the efficiency of the estimators.

References

[1] Jeelani, M.I., Maqbool, S., Mir,S.A., Khan,I., Nazir,N., and Jeelani,F.A class of Modified Ratio Estimator Using Linear Combination of Co-efficient of Kurtosis and Quartile Deviation. *International Journal of Modern Mathematical Sciences, USA.Florida*., **8(3)**, pp.149-153,2013

[2] Jeelani, M.I., Mir, S.A., Khan, I.,Nazir,N and Jeelani,F. NON-RESPONSE PROBLEMS IN RANKED SET SAMPLING. Pakistan Journal of Statistics.,**30(4)**, 555-562,2014a.

[3] Jeelani, M.I., Mir, S.A., Nazir,N and Jeelani,F.Modified ratio estimators using linear combination of Co-efficient of Skewness and Median of auxiliary variable under Rank set sampling and Simple random sampling. Indian Journal of Science and Technology.,**7(5)**, 723-728,2014b.

[4] Jeelani, M.I., Mir, S.A., Nazir,N and Jeelani,F. Role of Rank set sampling in Improving the estimates of Population under Stratification. American Journal of Mathematics and Statistics. Indian Journal of Science and Technology., **4(1)**,.21-25, 2014c.

[5] McIntyre, G. A. A Method of Unbiased Selective Sampling Using Ranked Set. Australian, J. Agricultural Research., **3(2)**. 385-390, 1952.

[6] Samawi, H. M. and Muttlak, H. A. Estimation of Ratio Using Rank Set Sampling, Biom. Journal., **38(1)**, 753-764,1996.

[7] Singh, H.P. and Tailor,R Use of known correlation co efficient in estimating the finite population means, Statistics in Transition.,**6(4)**, 555-560,2003.

[8] Frey, J. Recursive computation of inclusion probabilities in ranked set sampling. Journal of Statistical Planning and Inference ,**141**, 3632–3639, 2011.

[9] Frey, J. Nonparametric mean estimation using partially ordered sets. Environmental and Ecological Statistics.,**19**, 309–326, 2012.

[10] Gao, J., Ozturk, O. Two sample distribution-free inference based on partially rank-ordered set sample. Statistics & Probability Letters.,**82**, 876–884, 2012.

[11] Gokpinar, F., Ozdemir, Y.A. Generalization of inclusion probabilities in ranked set sampling. Hacettepe Journal of Mathematics and Statistics 39, 89–95,2010.

[12] Sohail, M.U., Shabbir, J. and Ahmed, S. A class of ratio type estimators for imputing the missing values under rank set sampling. Journal of Statistical Theory and Practice., **12(4)**, 704-717, 2018.

[13] Jeelani, M.I., Bouza, C.N. and Sharma, M. Modified ratio estimator under rank set sampling. Investigación Operacional., **38(1)**, 103-106, 2018.

[14] Abid, M., Nazir, H.Z., Tahir, M. and Riaz, M. On designing a new cumulative sum Wilcoxon signed rank chart for monitoring process location. PloS one., **13(4)**, p.e0195762. 2018.

[15] Subzar, M. and Bouza-Herrera, C.N. RATIO ESTIMATOR UNDER RANK SET SAMPLING SCHEME USING HUBER M IN CASE OF OUTLIERS. Investigación Operacional., **42(4)**, 469-477, 2021.