NEW RATIO METHOD OF ESTIMATION UNDER RANKED SET SAMPLING

M. Iqbal Jeelani*[1] and Carlos N Bouza**
*Division of Agricultural Statistics, SKUAST-K, India
**Facultad de Matemática y Computación, Universidad de La Habana, Cuba

ABSTRACT
In this paper a new form of ratio estimator using the linear combination of Median and Quartile Deviation of auxiliary variable under rank set sampling has been suggested. We have proved theoretically that the proposed estimator is more efficient than the classical estimator of ranked set sampling. The results are supported by empirical study also.

KEYWORDS: Rank set sampling, Median, Quartile deviation.

MSC: 62D05

RESUMEN
En este trabajo sugerimos una nueva forma del estimador de razón usando una combinación lineal de la Mediana y la Desviación Cuartílica de la variable auxiliar bajo muestreo por rangos ordenados. Hacemos una propuesta teórica sobre el estimador su mayor eficiencia con respecto al estimador clásico de muestreo por rangos ordenados. Los resultados son soportados por los resultados de un estudio empírico.

1. INTRODUCTION:

The ratio method of estimation is an important type of estimation in sample surveys which utilizes information on the auxiliary variable with a view of increasing the precision of the estimate of the population mean. The classic 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 the population mean of the auxiliary variable is known, $\bar{y}$ and $\bar{x}$ are sample means of the study variable and auxiliary variable, and the mean square of the classical ratio estimator is ($MSE(\bar{y}_r) = \delta\bar{Y}^2\left[C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\right]$). We denote $\delta = {}^1/_n$(ignoring sampling fraction), $R = \frac{\bar{Y}}{\bar{X}}$ is the population ratio, ρ_{yx} is the correlation coefficient between X and Y, C_y and C_x are the coefficients of variations of the study variable and the auxiliary variable respectively.

The next section presents the basics on the use of a ratio estimator when ranked set sampling is used. The proposed estimator is developed in section 3 and the last section is concerned with an empirical study.

2. RATIO ESTIMATION UNDER RANKED SET SAMPLING (RSS)

An alternative method to Simple random sampling (SRS) known as Ranked Set Sampling (RSS) was introduced to increase the efficiency of the estimation of population mean by McIntyre (1952). 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 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, without measuring Y. Then the unit with rank 1 is identified and taken for the measurement. 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 using 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

[1] jeelani.miqbal@gmail.com

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).
Takahasi-Wakimoto (1968) gave mathematical support to his claims. Dell-Clutter (1972) established that even if the ranking is not perfect the proposed estimator is still unbiased. The use of RSS is the theme a growing number of papers. Patil et.al. (2002) and Bouza (2005) Alomari-Bouza (2014) gave reviews of the theme as well as a large list of papers.

Different developments of ratio type estimators has been developed for RSS, see for example Al-Omari et al (2009, Bouza (2011).The classical ratio estimator under RSS proposed by Samawi and Muttlak (1996) is given below;

$$\hat{\bar{y}}_{rss} = \frac{\bar{y}_{rss}}{\bar{x}_{rss}} \tag{2.1}$$

where $\bar{y}_{rss} = \frac{1}{mk}\sum_{i=1}^{k} Y_i$ and $\bar{x}_{rss} = \frac{1}{mk}\sum_{i=1}^{k} X_i$, assuming that the population mean of the auxiliary variable is known the equation (2.1) changes to ;

$$\hat{\bar{y}}_{rss} = \frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} = \hat{R}_{rss}\bar{X} \tag{2.2}$$

The mean square error of this estimator is given by;

$$MSE(\hat{\bar{y}}_{rss}) = \left[\delta\{C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\} - \{\alpha_{y[i]} - \alpha_{x(i)}\}^2\right] \tag{2.3}$$

where,

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

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

$$\alpha_{x(i)}^2 = \frac{1}{k^2m}\frac{1}{\bar{X}^2}\sum_{i=1}^{k}\lambda_{x(i)}^2,\quad \alpha_{y[i]}^2 = \frac{1}{k^2m}\frac{1}{\bar{X}^2}\sum_{i=1}^{k}\lambda_{y(i)}^2,$$

$$\alpha_{yx(i)} = \frac{1}{k^2m}\frac{1}{\bar{Y}\bar{X}}\sum_{i=1}^{k}\lambda_{yx(i)}\ ,\ \lambda_{x(i)} = \left(\mu_{x(i)} - \bar{X}\right), \lambda_{y[i]} = \left(\mu_{y[i]} - \bar{Y}\right)$$

and

$$\lambda_{yx(i)} = \left(\mu_{x(i)} - \bar{X}\right)\left(\mu_{y[i]} - \bar{Y}\right)$$

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.

3. THE PROPOSED ESTIMATOR

Now using the linear combination of Median and Quartile Deviation of the auxiliary variable we propose the following estimator under rank set sampling;

$$\hat{\bar{y}}_{\tau rss} = \tau\frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} = \hat{R}_{\tau rss}\bar{X} \tag{3.1}$$

where, $\tau = \frac{\bar{X}Md+Qd}{\bar{x}Md+Qd}$, $Md =$ Median and Qd $= Q_3 - Q_1$, $Q_3 = 3\frac{(N-1)}{4}$, $Q_1 = \frac{(N-1)}{4}$.
Then ;

$$\hat{\bar{y}}_{\tau rss} = \tau\frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} = \hat{R}_{\tau rss}\bar{X} \tag{3.2}$$

The mean square of $\hat{R}_{\tau rss}$ will be estimated by the procedure given below ;

$$MSE\left(\hat{R}_{\tau rss}\right) = E(\hat{R}_{\tau rss} - R_{rss})^2 \tag{3.3}$$

Note that

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

Applying Taylor Series approximation, we have

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

$$\because \left({}^{1}\!/_{\bar{x}_{rss}} = {}^{1}\!/_{\bar{X} + \bar{x}_{rss} - \bar{X}} = {}^{1}\!/_{\bar{X}}\left(1 + {}^{\bar{x}_{rss} - \bar{X}}\!/_{\bar{X}}\right)^{-1} \cong {}^{1}\!/_{\bar{X}}\right)$$

as per Wolter (1985).

$$MSE\left(\hat{R}_{\tau rss}\right) \cong \frac{1}{\bar{X}^2}E\{(\tau\bar{y}_{rss} - \bar{Y}) - R_{rss}(\bar{x}_{rss} - \bar{X})\}^2 \cong \frac{1}{\bar{X}^2}\{E(\tau\bar{y}_{rss} - \bar{Y})^2 - 2R_{rss}E(\tau\bar{y}_{rss} - \bar{Y})(\bar{x}_{rss} - \bar{X}) + R_{rss}{}^2E(\bar{x}_{rss} - \bar{X})^2\}$$

Then, if C is a constant

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

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

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

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

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

Now the mean square error of $\hat{\bar{y}}_{\tau rss}$ in equation (2.1) will be ;

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

where ,

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

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

Suppose that

$$\psi = \frac{1}{k^2 m}\left(\tau^2\sum_{i=1}^k \lambda_{y[i]}^2 - 2R_{rss}\tau\sum_{i=1}^k \lambda_{yx(i)}^2 + {R_{rss}}^2\sum_{i=1}^k \lambda_{x(i)}^2\right),$$

where ψ is non-negative value, then the increase in accuracy of the proposed estimator is readily obtained.

$$MSE(\hat{\bar{y}}_{\tau rsss}) \cong MSE(\hat{\bar{y}}_{rss}) - \psi.$$

From this condition it can easily be seen that the MSE of proposed estimator is less than the classical ratio estimator of rank set sampling proposed by Samawi and Muttlak (1996), provided $\psi > 1$. The bias of the proposed estimator is given below;

$$\bar{y}_{rss} = \bar{Y}(1+\iota_0) \quad \bar{x}_{rss} = \bar{X}(1+\iota_1)\ ,$$

such that the expectation of ι_0 and ι_1 is equal to zero

$$E(\iota_0) = E(\iota_1) = 0$$

Hence

$$V(\iota_0) = E({\iota_0}^2) = \frac{V(\bar{y}_{rss})}{\bar{Y}^2} = \frac{1}{km}\frac{1}{\bar{Y}^2}\left[\frac{1}{k}\sum_{i=1}^k \pi_{y[i]}^2\right] = \left\lfloor \delta C_y^2 - \alpha_{y[i]}^2 \right\rfloor$$

Similarly

$$V(\iota_1) = E({\iota_1}^2) = \left\lfloor \delta C_x^2 - \alpha_{x(i)}^2 \right\rfloor$$

and

$$Cov\ (\iota_0, \iota_1) = E\ (\iota_0, \iota_1) = \frac{Cov(\bar{y}^*\bar{x}^*)}{\bar{Y}\bar{X}} = \frac{1}{\bar{Y}\bar{X}}\frac{1}{km}\left[\sigma_{yx} - \frac{1}{k}\sum_{i=1}^k \lambda_{yx(i)}\right] = [\delta\rho C_y C_x - \alpha_{yx(i)}]$$

where,

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

also,

$$\hat{\bar{y}}_{\tau rss} = \tau\frac{\bar{y}_{rss}}{\bar{x}_{rss}}\bar{X} = \tau\frac{\bar{Y}(1+\iota_0)}{\bar{X}(1+\iota_1)}\bar{X} = \tau\bar{Y}(1+\iota_0)(1+\iota_1)^{-1}$$

Taking expectations on both sides the bias of $\hat{\bar{y}}_{\tau rss}$,is

$$Bias\ (\hat{\bar{y}}_{\tau rss}) \cong (\tau-1)\bar{Y} - \tau\bar{Y}(\delta C_x^2 - \alpha_{x(i)}^2 - \delta C_{yx} + \alpha_{yx(i)})$$

4. EMPIRICAL STUDY :

In this section we use the data of Jeelani et al. (2014). The data was collected on Apple production from the district Ganderbal of Kashmir valley from 420 orchards in 30 villages. The variables chosen for the study, where Yield

(MT), Area (ha). We take equal samples for each ratio estimator for the sake of comparison. The sample sizes considered were 16, 40, 72 . The set size for the ranked set ratio estimator is $k = 2, 5, 9$ and the number of cycles $m = 8$, that is ($n = km = 6 \times 10 = 60$. A simulation study has been carried out by taking three combinations of set sizes and three combinations of correlation coefficients, for efficiency comparison. Using these inputs poputlion were generated.
Table 1 gives the results obtained. It is clear that the proposed estimator is more efficient as compared to the classical in all the cases.

Table 1 Efficiency comparison of $\bar{y}_{rss}$ and $\hat{\bar{y}}_{\tau rss}$

	$(\bar{y}_{rss})$	$(\hat{\bar{y}}_{\tau rss})$	k(set size)	m (no.of cycles)	$n = km$ (Sample size)
0.99	1.33	1.56	2	8	16
	1.61	1.78	5		40
	1.78	2.11	9		72
0.70	1.25	1.45	2		16
	1.53	1.64	5		40
	1.60	1.80	9		72
0.50	1.12	1.38	2		16
	1.41	1.55	5		40
	1.48	1.71	9		72

5. CONCLUSIONS

From the above derived results we conclude that the proposed ratio estimator under RSS is more efficient than classical ratio estimator proposed by Samawi and Muttalak (1996) under rank set sampling. The efficiency of the proposed RSS estimator decreases as the correlation coefficient decreases. These benefits of RSS need not be restricted to simple random sampling; replacing SRS with RSS in the final stage of any survey design or method of estimation in surveys will greatly improve the efficiency of the estimators.

Acknowledgments: The authors thanks the criteria of two anonymous referees for their helpful comments. The research was benefited by the support of the PNB-CITMA Modelos Matemáticos para el Estudio de Medio Ambiente, Salud y Desarrollo Humano. and by Modelos e Metodologias para Apoio a Decisão na Gestão da Saúde e do Meio Ambiente, PROJETO CAPES-MES-CUBA 209/13.

RECEIVED NOVEMBER, 2014
REVISED JANUARY, 2015

REFERENCES

[1] AL-OMARI, A.I., A. A. JEMAIN and I. KAMARULZAMAN (2009): New ratio estimators of the mean using simple random sampling and ranked set sampling methods. **Revista Investigación Operacional** . 30, 97-108.
[2] AL-OMARI, A.I and BOUZA, C. N. (2014): Review Of Ranked Set Sampling: Modifications and Applications. **Revista Investigación Operacional,** 3, 215-240**.**
[3] BOUZA, C. N. (2005): Sampling Using Ranked Sets:Concepts, Results and Perspectives**. Revista Investigación Operacional.** 26, No. 3,275-293
[4] BOUZA, C. N (2001): Model assisted ranked survey sampling. **Biometrical J**., 43, 249-259.
[5] DELL, T.R. and CLUTTER, J.L.(1972): Ranked set sampling theory with order statistics. Background. **Biometrics** 28, 545-555.

[6] JEELANI, M.I., ,MAQBOOL, S., MIR,S.A., KHAN,I., NAZIR,N., and JEELANI,F. .(2013).A class of Modified Ratio Estimator Using Linear Combination of Co-efficient of Kurtosis and Quartile Deviation**. International Journal of Modern Mathematical Sciences,** 8, 149-153.
[7] JEELANI, M.I., MIR, S.A., KHAN,I.,NAZIR,N and JEELANI,F.(2014 a): Non-Response Problems In Ranked Set Sampling. **Pakistan Journal of Statistics,** 30, 555-562.
[8] JEELANI, M.I., MIR, S.A., NAZIR,N and JEELANI,F. (2014b):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, 723-728.
[9] JEELANI, M.I., MIR, S.A., NAZIR,N and JEELANI,F. (2014c).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 , 21-25.
[10] MCINTYRE, G. A. (1952): A Method of Unbiased Selective Sampling Using Ranked Set. **Australian, J. Agricultural Research**, 3, 385-390.
[11] PATIL, G.P (2002): Ranked set sampling. In **Encyclopedia of Enviromentrics**. (A.H. El-Shaarawi and W.W. Pieegoshed, (Editors). Vol. 3,1684-1690. Wiley, Chichester.
[12] SAMAWI, H. M. and MUTTLAK, H. A. (1996) Estimation of Ratio Using Rank Set Sampling. **Biom. Journal,**38, 753-764.
[13] TAKAHASI K. and WAKIMOTO, K.(1968): On unbiased estimates of the population mean based on sample stratified by means of ordering. **Annals of the Inst. of Statistical Mathematics**. 20, 1-31