

A BETTER ESTIMATOR OF POPULATION MEAN WITH POWER TRANSFORMATION BASED ON RANKED SET SAMPLING

Nitu Mehta (Ranka)¹, V.L. Mandowara²

ABSTRACT

Ranked set sampling (RSS) was first suggested by McIntyre (1952) to increase the efficiency of estimate 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. This paper presents a modified ratio estimator using prior value of coefficient of kurtosis of an auxiliary variable x , with the intention to improve the efficiency of ratio estimator in ranked set sampling. The first order approximation to the bias and mean square error (MSE) of the proposed estimator are obtained. A generalized version of the suggested estimator by applying the Power transformation is also presented.

Key words: ranked set sampling, ratio estimator, power transformation estimator, auxiliary variable.

1. Introduction

The traditional ratio estimator for the population mean \bar{Y} of the study variable y is defined by

$$\hat{Y}_R = \bar{y} \left(\frac{\bar{X}}{\bar{x}} \right) \quad (1.1)$$

in which it is assumed that the population mean \bar{X} of the auxiliary variable x is known. Here, \bar{y} is the sample mean of the study variable and \bar{x} is the sample mean of the auxiliary variable.

¹ Research Scholar, Dept. of Mathematics & Statistics, University College of Science, M. L. Sukhadia University, Udaipur-313001, Rajasthan, India. E-mail: nitumehta82@gmail.com.

² Dept. of Mathematics & Statistics, University College of Science, M. L. Sukhadia University, Udaipur-313001, Rajasthan, India. E-mail: mandowara_vl@yahoo.co.in.

The bias and mean square error (MSE) of $\overset{\Delta}{Y}_R$ are given by

$$B(\overset{\Delta}{Y}_R) = \theta \bar{Y} C_x^2 (1 - k) \quad (1.2)$$

and $MSE(\overset{\Delta}{Y}_R) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 - 2k)]$ (1.3)

where $\theta = \frac{1}{n}$, $k = \rho \frac{C_y}{C_x}$, C_y and C_x are coefficients of variations of y and x respectively and ρ is correlation coefficient.

Singh, H.P (2004) proposed a modified ratio estimator as

$$\overset{\Delta}{Y}_M = y \left[\frac{\bar{X} + \beta_2(x)}{\bar{x} + \beta_2(x)} \right] \quad (1.4)$$

where $\beta_2(x)$ is known value of the coefficient of kurtosis of an auxiliary variable.

The Bias and MSE of this estimator were given by

$$B(\overset{\Delta}{Y}_M) = \theta \bar{Y} \lambda C_x^2 (\lambda - k) \quad (1.5)$$

and $MSE(\overset{\Delta}{Y}_M) = \theta \bar{Y}^2 [C_y^2 + \lambda C_x^2 (\lambda - 2k)]$ (1.6)

where $\lambda = \frac{\bar{X}}{\bar{X} + \beta_2(x)}$

By applying the power transformation on $\overset{\Delta}{Y}_{M\alpha}$ in (1.4), the generalized estimator is

$$\overset{\Delta}{Y}_{M\alpha} = y \left[\frac{\bar{X} + \beta_2(x)}{\bar{x} + \beta_2(x)} \right]^\alpha \quad (1.7)$$

where α is a suitably chosen scalar.

The bias and MSE of the estimator $\overset{\Delta}{Y}_{M\alpha}$ to the first degree of approximation are respectively given by

$$B(\overset{\Delta}{Y}_{M\alpha}) = \theta\alpha \left(\frac{\bar{Y}}{2} \right) \lambda C_x^2 \{ \lambda(\alpha + 1) - 2k \} \tag{1.8}$$

and $MSE(\overset{\Delta}{Y}_{M\alpha}) = \theta \bar{Y}^2 [C_y^2 + \alpha \lambda C_x^2 (\alpha \lambda - 2k)]$ (1.9)

in which the value of $\alpha = \frac{k}{\lambda}$ makes the MSE in (1.9) minimum. Comparing (1.9) when $\alpha = \frac{k}{\lambda}$ with (1.6), Singh (2004) showed that $\overset{\Delta}{Y}_{M\alpha}$ is more efficient than $\overset{\Delta}{Y}_M$.

2. The suggested estimator

In Ranked set sampling (RSS), m independent random sets, each of size m , are selected with equal probability and with replacement from the population. The members of each random set are ranked with respect to the characteristic of the study variable or auxiliary variable. Then, the smallest unit is selected from the ordered set and the second smallest unit is selected from the second ordered set. By this way, this procedure is continued until the unit with the largest rank is chosen from the m^{th} set. This cycle may be repeated r times, so mr units have been measured during this process.

When we rank on the auxiliary variable, let $(y_{[i]}, x_{(i)})$ denote a i^{th} judgment ordering in the i^{th} set for the study variable and i^{th} set for the auxiliary variable.

Swami and Muttalak (1996) defined the estimator of the population ratio using RSS as

$$\overset{\Delta}{R}_{RSS} = \frac{\bar{y}_{[n]}}{\bar{x}_{(n)}} \tag{2.1}$$

where $\bar{y}_{[n]} = \frac{1}{n} \sum_{i=1}^n y_{[i]}$ and $\bar{x}_{(n)} = \frac{1}{n} \sum_{i=1}^n x_{(i)}$

As Swami and Muttlak (1996) remind that this estimator can also be used for the population total and mean. We can write the following estimator for the population mean as

$$\frac{\Delta}{\bar{Y}_{R,RSS}} = \bar{y}_{[n]} \left(\frac{\bar{X}}{\bar{x}_{(n)}} \right) \quad (2.2)$$

To obtain bias and MSE of $\frac{\Delta}{\bar{Y}_{R,RSS}}$, we put $\bar{y}_{[n]} = \bar{Y}(1 + \varepsilon_0)$ and $\bar{x}_{(n)} = \bar{X}(1 + \varepsilon_1)$ so that $E(\varepsilon_0) = E(\varepsilon_1) = 0$

$$\begin{aligned} V(\varepsilon_0) &= E(\varepsilon_0^2) = \frac{V(\bar{y}_{[n]})}{\bar{Y}^2} \\ &= \frac{1}{mr} \frac{1}{\bar{Y}^2} \left[S_y^2 - \frac{1}{m} \sum_{i=1}^m \tau_{y[i]}^2 \right] = [\theta C_y^2 - W_{y[i]}^2] \end{aligned}$$

similarly, $V(\varepsilon_1) = E(\varepsilon_1^2) = [\theta C_x^2 - W_{x(i)}^2]$

$$\begin{aligned} \text{and } Cov(\varepsilon_0, \varepsilon_1) &= E(\varepsilon_0, \varepsilon_1) = \frac{Cov(\bar{y}_{[n]}, \bar{x}_{(n)})}{\bar{X}\bar{Y}} \\ &= \frac{1}{\bar{X}\bar{Y}} \frac{1}{mr} \left[S_{yx} - \sum_{i=1}^m \tau_{yx(i)} \right] = [\theta \rho_{yx} C_y C_x - W_{yx(i)}] \end{aligned}$$

where $\theta = \frac{1}{mr}$, $C_y^2 = \frac{S_y^2}{\bar{Y}^2}$, $C_x^2 = \frac{S_x^2}{\bar{X}^2}$, $C_{yx} = \frac{S_{yx}}{\bar{X}\bar{Y}} = \rho_{yx} C_y C_x$,

$$W_{x(i)}^2 = \frac{1}{m^2 r} \frac{1}{\bar{X}^2} \sum_{i=1}^m \tau_{x(i)}^2, \quad W_{y[i]}^2 = \frac{1}{m^2 r} \frac{1}{\bar{Y}^2} \sum_{i=1}^m \tau_{y[i]}^2 \quad \text{and}$$

$$W_{yx(i)} = \frac{1}{m^2 r} \frac{1}{\bar{X}\bar{Y}} \sum_{i=1}^m \tau_{yx(i)}.$$

Here, we would also like to remind that $\tau_{x(i)} = \mu_{x(i)} - \bar{X}$, $\tau_{y[i]} = \mu_{y[i]} - \bar{Y}$ and $\tau_{yx(i)} = (\mu_{x(i)} - \bar{X})(\mu_{y[i]} - \bar{Y})$.

Further, to validate first degree of approximation, we assume that the sample size is large enough to get $|\varepsilon_0|$ and $|\varepsilon_1|$ as small so that the terms involving ε_0 and or ε_1 in a degree greater than two will be negligible.

Bias and MSE of the estimator $\bar{Y}_{R,RSS}$ to the first degree of approximation are respectively given by

$$B(\bar{Y}_{R,RSS}) = E(\bar{Y}_{R,RSS}) - \bar{Y}$$

Here $\bar{Y}_{R,RSS} = \bar{Y}(1 + \varepsilon_0)(1 + \varepsilon_1)^{-1}$

$$= \bar{Y}[1 + \varepsilon_0 - \varepsilon_1 - \varepsilon_0\varepsilon_1 + \varepsilon_1^2 + o(\varepsilon_1)]$$

Now $E(\bar{Y}_{R,RSS}) = \bar{Y}[1 + E(\varepsilon_1^2) - E(\varepsilon_0\varepsilon_1)]$

$$\Rightarrow B(\bar{Y}_{R,RSS}) = \bar{Y}[\{\theta C_x^2 - W_{x(i)}^2\} - \{\theta \rho_{yx} C_y C_x - W_{yx(i)}\}]$$

$$\Rightarrow B(\bar{Y}_{R,RSS}) = \bar{Y}[\theta C_x^2(1 - k) - (W_{x(i)}^2 - W_{yx(i)})]$$

where $k = \rho \frac{C_y}{C_x}$ (2.3)

Now $MSE(\bar{Y}_{R,RSS}) = E(\bar{Y}_{R,RSS} - \bar{Y})^2$

$$= \bar{Y}^2 E[\varepsilon_0 - \varepsilon_1 - \varepsilon_0\varepsilon_1 + \varepsilon_1^2]^2 = \bar{Y}^2 [\varepsilon_0^2 - \varepsilon_1^2 - 2\varepsilon_0\varepsilon_1]$$

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

$$= \bar{Y}^2 [\theta\{C_y^2 + C_x^2(1 - 2k)\} - \{W_{y[i]}^2 + W_{x(i)}^2 - 2W_{yx(i)}\}]$$

$$MSE(\bar{Y}_{R,RSS}) = \bar{Y}^2 [\theta\{C_y^2 + C_x^2(1 - 2k)\} - \{W_{y[i]} - W_{x(i)}\}^2] \tag{2.4}$$

Adapting the estimator in (2.1) to the modified ratio estimator for the population mean suggested by Singh (2004), given in (1.4), we develop the following estimator

$$\bar{Y}_{M,RSS} = y_{[n]} \left(\frac{\bar{X} + \beta_2(x)}{x_n + \beta_2(x)} \right) \tag{2.5}$$

The Bias and MSE of $\bar{Y}_{M,RSS}$ can be found as follows:

$$B(\bar{Y}_{M,RSS}) = E(\bar{Y}_{M,RSS}) - \bar{Y}$$

Here $\bar{Y}_{M,RSS} = \bar{Y}(1 + \varepsilon_0)(1 + \lambda\varepsilon_1)^{-1}$ where $\lambda = \frac{\bar{X}}{\bar{X} + \beta_2(x)}$

Suppose $|\lambda\varepsilon_1| < 1$ so that $(1 + \lambda\varepsilon_1)^{-1}$ is expandable.

$$\begin{aligned} \text{So } B(\bar{Y}_{M,RSS}) &= \bar{Y} \left[\lambda^2 E(\varepsilon_1^2) - \lambda E(\varepsilon_0 \varepsilon_1) \right] \\ &= \bar{Y} \left[\lambda^2 \{ \theta C_x^2 - W_{x(i)}^2 \} - \lambda \{ \theta \rho_{yx} C_y C_x - W_{yx(i)} \} \right] \end{aligned}$$

$$\text{and } B(\bar{Y}_{M,RSS}) = \bar{Y} \left[\theta \lambda C_x^2 (1-k) - \lambda (W_{x(i)}^2 - W_{yx(i)}) \right] \quad (2.6)$$

$$\text{where } k = \rho \frac{C_y}{C_x}$$

$$\begin{aligned} \text{MSE}(\bar{Y}_{M,RSS}) &= E(\bar{Y}_{M,RSS} - \bar{Y})^2 \\ &= \bar{Y}^2 \left[\varepsilon_0^2 + \lambda^2 \varepsilon_1^2 - 2\lambda \varepsilon_0 \varepsilon_1 \right] \\ &= \bar{Y}^2 \left[\theta C_y^2 - W_{y[i]}^2 + \lambda^2 (\theta C_x^2 - W_{x(i)}^2) - 2\lambda (\theta \rho_{yx} C_y C_x - W_{yx(i)}) \right] \\ &= \bar{Y}^2 \left[\theta \{ C_y^2 + \lambda C_x^2 (\lambda - 2k) \} - \{ W_{y[i]}^2 + \lambda^2 W_{x(i)}^2 - 2\lambda W_{yx(i)} \} \right] \end{aligned}$$

$$\text{MSE}(\bar{Y}_{M,RSS}) = \bar{Y}^2 \left[\theta \{ C_y^2 + \lambda C_x^2 (\lambda - 2k) \} - \{ W_{y[i]} - \lambda W_{x(i)} \}^2 \right] \quad (2.7)$$

By applying the power transformation on $\bar{Y}_{M,RSS}$, the generalized estimator is given by

$$\bar{Y}_{M\alpha,RSS} = \bar{y}_{[n]} \left(\frac{\bar{X} + \beta_2(x)}{x_n + \beta_2(x)} \right)^\alpha \quad (2.8)$$

The bias and MSE of the estimator $\bar{Y}_{M\alpha,RSS}$ to the first degree of approximation, are respectively given by

$$B(\bar{Y}_{M\alpha,RSS}) = E(\bar{Y}_{M\alpha,RSS}) - \bar{Y}$$

$$\text{Here } \bar{Y}_{M\alpha,RSS} = \bar{Y} (1 + \varepsilon_0) (1 + \lambda \varepsilon_1)^{-\alpha}$$

$$= \bar{Y} \left[(1 + \varepsilon_0) \left\{ 1 - \lambda \alpha \varepsilon_1 + \frac{\alpha(\alpha+1)}{2} \lambda^2 \varepsilon_1^2 + o(\varepsilon_1) \right\} \right]$$

$$\begin{aligned} B(\bar{Y}_{M\alpha,RSS}) &= \bar{Y} \left[\lambda^2 \frac{\alpha(\alpha+1)}{2} \{ \theta C_x^2 - W_{x(i)}^2 \} - \lambda \alpha \{ \theta \rho_{yx} C_y C_x - W_{yx(i)} \} \right] \\ &= \theta \alpha \left(\frac{\bar{Y}}{2} \right) \lambda C_x^2 \{ \lambda(\alpha+1) - 2k \} - \left(\frac{\bar{Y}}{2} \right) \lambda \alpha \{ \lambda(\alpha+1) W_{x(i)}^2 - 2W_{yx(i)} \} \end{aligned}$$

$$\Rightarrow B(\overset{\Delta}{Y}_{M\alpha,RSS}) = \left(\frac{\bar{Y}}{2}\right) \lambda\alpha \left[\theta C_x^2 \{ \lambda(\alpha + 1) - 2k \} - \{ \lambda(\alpha + 1)W_{x(i)}^2 - 2W_{yx(i)} \} \right] \quad (2.9)$$

where $k = \rho \frac{C_y}{C_x}$

$$\begin{aligned} \text{and } MSE(\overset{\Delta}{Y}_{M\alpha,RSS}) &= E(\overset{\Delta}{Y}_{M\alpha,RSS} - \bar{Y})^2 \\ &= \bar{Y}^2 E \left[\varepsilon_0 - \lambda\alpha\varepsilon_1 + \lambda^2 \frac{\alpha(\alpha + 1)}{2} \varepsilon_1^2 - \lambda\alpha\varepsilon_0\varepsilon_1 \right]^2 \\ &= \bar{Y}^2 E \left[\varepsilon_0^2 + \lambda^2 \alpha^2 \varepsilon_1^2 - 2\lambda\alpha\varepsilon_0\varepsilon_1 \right] \\ &= \bar{Y}^2 \left[\theta C_y^2 - W_{y[i]}^2 + \lambda^2 \alpha^2 (\theta C_x^2 - W_{x(i)}^2) - 2\lambda\alpha(\theta\rho_{yx} C_y C_x - W_{yx(i)}) \right] \end{aligned}$$

$$MSE(\overset{\Delta}{Y}_{M\alpha,RSS}) = \bar{Y}^2 \left[\theta \{ C_y^2 + \alpha\lambda C_x^2 (\alpha\lambda - 2k) \} - \{ W_{y[i]} - \lambda\alpha W_{x(i)} \}^2 \right] \quad (2.10)$$

let $A = (W_{y[i]} - \lambda\alpha W_{x(i)})^2$

By this way, we can write (2.10) as

$$MSE(\overset{\Delta}{Y}_{M\alpha,RSS}) \cong MSE(\overset{\Delta}{Y}_{M\alpha}) - A \quad (2.11)$$

It is easily shown that MSE of the proposed estimator using ranked set sampling is always smaller than the estimator, suggested by Singh (2004) given in (1.7), because A is a non-negative value. As a result, it is shown that the

proposed estimator $\overset{\Delta}{Y}_{M\alpha,RSS}$ is more efficient than the estimator $\overset{\Delta}{Y}_{M\alpha}$.

3. Optimality of α

The optimum value of α to minimize the MSE of $\overset{\Delta}{Y}_{M\alpha,RSS}$ can easily be found as follows:

$$\begin{aligned} \frac{\partial MSE(\overset{\Delta}{Y}_{M\alpha,RSS})}{\partial \alpha} &= 0 \\ \Rightarrow \theta C_x^2 (2\lambda^2 \alpha - 2k\lambda) - (2\lambda^2 \alpha W_{x(i)}^2 - 2\lambda W_{yx(i)}) &= 0 \end{aligned}$$

$$\begin{aligned}
&\Rightarrow \theta C_x^2 (2\lambda^2 \alpha - 2k\lambda) - (2\lambda^2 \alpha W_{x(i)}^2 - 2\lambda W_{yx(i)}) = 0 \\
&\Rightarrow \lambda \alpha (\theta C_x^2 - W_{x(i)}^2) + (W_{yx(i)} - \theta C_x^2 k) = 0 \\
&\Rightarrow \alpha' = \frac{\theta k C_x^2 - W_{yx(i)}}{\lambda (\theta C_x^2 - W_{x(i)}^2)} \quad (3.1)
\end{aligned}$$

When we replace α by α' in (2.10), we obtain minimum MSE of the proposed estimator as follows:

$$\begin{aligned}
\min .MSE(\bar{Y}_{M\alpha, RSS}) &\cong \bar{Y}^2 [\theta \{C_y^2 + \alpha' \lambda C_x^2 (\lambda \alpha' - 2k)\} - \{W_{y[i]}^2 + \lambda \alpha' (\lambda \alpha' W_{x(i)}^2 - 2W_{yx(i)})\}] \\
\Rightarrow MSE(\bar{Y}_{M\alpha, RSS}) &= \bar{Y}^2 [\theta \{C_y^2 + \alpha' \lambda C_x^2 (\alpha' \lambda - 2k)\} - \{W_{y[i]} - \lambda \alpha' W_{x(i)}\}^2] \quad (3.2)
\end{aligned}$$

REFERENCES

- KADILAR, C., UNYAZICI, Y. and CINGI, H. (2009). Ratio estimator for the population mean using ranked set sampling, *Stat. papers*, 50, 301-309.
- KOWALCZYK B. (2004). Ranked set sampling and its Applications in Finite population studies, *Statistics in transition*, Vol.6, No.7 pp.1031-1046.
- MCINTYRE, G. A. (1952). A method of unbiased selective sampling using ranked sets, *Australian Journal of Agricultural Research*, 3, 385-390.
- SAMAWI, H. M. and MUTTLAK, H. A. (1996). Estimation of Ratio Using Rank Set Sampling, *Biom. J.* 38, 6, 753-764.
- SINHA, A. K. (2005). On some recent developments in ranked set sampling, *Bulletin of Informatics & Cybernetics*, Research Association of Statistical Sciences, Vol. 37, 136-160.
- SINGH, H. P., TAILOR RAJESH, TAILOR RITESH and KAKRAN, M. S. (2004). An improved estimator of population mean using Power Transformation, *J. Ind. Soc. Agril. Statist.* 58(2), 223-230.
- SINGH, S. (2003). *Advanced Sampling Theory with Application*, Vol. I, Kluwer Academic Publishers, Netherlands.