Improvement Over General And Wider Class of Estimators Using Ranked Set Sampling

V. L. Mandowara, Nitu Mehta (Ranka)

Abstract: In this paper, Improvement over general and wider class of estimators of finite population means using ranked set sampling is investigated. Ranked set sampling (RSS) was first suggested to increase the efficiency of estimator of the population mean. The first order approximation to the bias and mean square error (MSE) of the investigated estimators are obtained. Theoretically, it is shown that these suggested estimators are more efficient than the general and wider class of estimators in simple random sampling.

Key Words: Ranked set sampling, General and wider class of estimator, Auxiliary variable, Mean square error, Efficiency.

1. Introduction

Ranked set sampling was first suggested by McIntyre (1952) to increase the efficiency of estimator of population mean. Kadilar et al. (2009) used this technique to improve ratio estimator given by Prasad (1989). Here we shall improve general and wider class of estimators given by Srivastava (1971, 1980) based on auxiliary variable. Srivastava (1971) proposed a general class of estimators to estimate the population mean $\overline{Y}$ of the study variable which in the case of single mean $\overline{X}$ of the auxiliary variable is given by

$$t_g = \overline{y} H(u) \tag{1.1}$$

where $u = \frac{\overline{x}}{\overline{X}}$ and $H(\bullet)$ is a parametric function.

In the Simple random sampling, the bias and minimum mean square error of the general class of estimator, see singh (Volume 1, 2003, page 164-165) are given by

$$B(t_g) = \left(\frac{1-f}{n}\right)\overline{Y}\left[H_2 C_x^2 + H_1 \rho_{xy} C_Y C_x\right]$$

or $$B(t_g) = \frac{1}{n}\overline{Y}\left[H_2 C_x^2 + H_1 \rho_{xy} C_Y C_x\right]$$

(On ignoring $f = \frac{n}{N}$)

or $$B(t_g) = \theta\overline{Y}\left[H_2 C_x^2 + H_1 \rho_{xy} C_Y C_x\right] \tag{1.2}$$

where $\theta = \frac{1}{n}$, n and N are the sample and population sizes respectively ; $f = \frac{n}{N}; C_y^2, C_x^2$ denote the coefficient of variation of Y and X respectively and ρ_{yx} denote the correlation coefficient between Y and X . Here $H_1 = \left(\frac{\partial H}{\partial u}\right)_{u=1}$ and $H_2 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial u^2}\right)_{u=1}$ denote the first and second order partial derivatives of H with respect to u and are the known constants.

The minimum mean square error (MSE) of the general class of estimator t_g defined at (1.1), to the first order of approximation is

$$Min.MSE(t_g) = \frac{(1-f)}{n}\overline{Y}^2 C_y^2\left[1-\rho_{xy}^2\right]$$

or $$Min.MSE(t_g) = \frac{1}{n}\overline{Y}^2 C_y^2\left[1-\rho_{xy}^2\right]$$

(on ignoring $f = \frac{n}{N}$)

or $$Min.MSE(t_g) = \theta\overline{Y}^2 C_y^2\left[1-\rho_{xy}^2\right] \tag{1.3}$$

If we attach any function of $\frac{\overline{x}}{\overline{X}}$ to the sample mean $\overline{y}$, the asymptotic minimum mean square error of the resultant estimator cannot be reduced further than that given in (1.3). Thus the usual ratio estimator, product estimator and power transformation estimator are the special cases of the class of estimators defined in (1.1). While the regression estimator and difference estimator are not special cases of the general class of estimators. Then Srivastava (1980)

- ***V.L.Mandowara*** *is working as Professor in the Department of Mathematics and Statistics, University College of Science, MLSU, Udaipur.* ***Emai-*** *mandowara_vl@yahoo.co.in*
- ***Nitu Mehta (Ranka)*** *is currently pursuing Ph.D. in Statistics in the Department of Mathematics and Statistics, University College of Science, MLSU, Udaipur.* ***Email****- nitumehta82@gmail.com*

defined another class of estimators and named a wider class of estimators as

$$t_w = H\left[\bar{y}, u\right] \quad (1.4)$$

where $H\left[\bar{y}, u\right]$ is a function of $\bar{y}$ and u.

The asymptotic bias and minimum mean square error, see singh (Volume 1, 2003, page 166-167) are given by

$$B(t_w) = \left(\frac{1-f}{n}\right)\left[\bar{Y}\rho_{xy}C_Y C_x H_3 + C_x^2 H_2 + \bar{Y}^2 C_y^2 H_4\right]$$

(on ignoring $f = \frac{n}{N}$)

or $B(t_w) = \theta\left[\bar{Y}\rho_{xy}C_Y C_x H_3 + C_x^2 H_2 + \bar{Y}^2 C_y^2 H_4\right]$ (1.5)

Where $H_1 = \left(\frac{\partial H}{\partial u}\right)_{\bar{y}=\bar{Y}, u=1}$,

$H_2 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial u^2}\right)_{\bar{y}=\bar{Y}, u=1}$,

$H_3 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial \bar{y}\partial u}\right)_{\bar{y}=\bar{Y}, u=1}$ and $H_4 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial^2 \bar{y}^2}\right)_{\bar{y}=\bar{Y}, u=1}$.

The minimum mean squared error of the wider class of estimator, t_w, is given by

$$Min.MSE(t_w) = \frac{(1-f)}{n}\bar{Y}^2 C_y^2\left[1-\rho_{xy}^2\right]$$

or $Min.MSE(t_w) = \frac{1}{n}\bar{Y}^2 C_y^2\left[1-\rho_{xy}^2\right]$

(on ignoring $f = \frac{n}{N}$)

or $Min.MSE(t_w) = \theta\bar{Y}^2 C_y^2\left[1-\rho_{xy}^2\right]$ (1.6)

If we take any function of $\bar{y}$, $\bar{x}$ and $\bar{X}$ to estimate the population mean, $\bar{Y}$, the asymptotic minimum mean square error of the resultant estimator again cannot be reduced further than that given in (1.6). Thus the usual linear regression estimator and difference estimator are special cases of the wider class of estimator defined at (1.4).

2. The Suggested Estimators

In Ranked set sampling (RSS), m independent random sets, each of size m are selected with equal probability and without 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 $mr(=n)$ 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. Adapting the estimator in (1.1) to the general class of estimator for the population mean proposed by Srivastava (1971), we suggest the following estimator for general class of estimator using ranked set sampling is

$$t_{g,RSS} = \bar{y}_{[n]} H(u) \quad (2.1)$$

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

and $H(\bullet)$ is a parametric function, such that, it satisfies the following conditions:

a) $H(1) = 1$

b) The first and second order partial derivatives of H with respect to u exists and are known constants at a given point $u = 1$.

Expanding $H(u)$ about the value 1 in a second order Taylor's series, we have

$$H(u) = H[1 + (u-1)]$$

$$= H(1) + (u-1)\left(\frac{\partial H}{\partial u}\right)_{u=1} + (u-1)^2\left(\frac{1}{2}\frac{\partial^2 H}{\partial u^2}\right)_{u=1} + ...$$

Note that $|u-1| < 1$ thus the higher order terms can be neglected. Using the above two conditions, we obtain

$$t_{g,RSS} = \bar{y}_{[n]}\left[1 + (u-1)H_1 + (u-1)^2 H_2 + ...\right] \quad (2.2)$$

where $H_1 = \left(\frac{\partial H}{\partial u}\right)_{u=1}$ and $H_2 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial u^2}\right)_{u=1}$

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

$$V(\varepsilon_0)=E(\varepsilon_0^2)=\frac{V(\bar{y}_{[n]})}{\bar{Y}^2}$$

$$=\frac{1}{mr}\frac{1}{\bar{Y}^2}\left[S_y^2-\frac{1}{m}\sum_{i=1}^{m}\tau_{y[i]}^2\right]=\left[\theta C_y^2-W_{y[i]}^2\right]$$

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

and $Cov(\varepsilon_0,\varepsilon_1)=E(\varepsilon_0,\varepsilon_1)=\dfrac{Cov\left(\bar{y}_{[n]},\bar{x}_{(n)}\right)}{\bar{X}\bar{Y}}$

$$=\frac{1}{\bar{X}\bar{Y}}\frac{1}{mr}\left[S_{yx}-\frac{1}{m}\sum_{i=1}^{m}\tau_{yx(i)}\right]=\left[\theta\rho_{yx}C_yC_x-W_{yx(i)}\right]$$

where $\theta=\dfrac{1}{mr}$, $C_y^2=\dfrac{S_y^2}{\bar{Y}^2}$, $C_x^2=\dfrac{S_x^2}{\bar{X}^2}$

$$C_{yx}=\frac{Syx}{\bar{X}\bar{Y}}=\rho_{yx}C_yC_x,\ W_{x(i)}^2=\frac{1}{m^2r}\frac{1}{\bar{X}^2}\sum_{i=1}^{m}\tau_{x(i)}^2$$

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

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. The proposed estimator $t_{g,RSS}$ given in (2.1) can easily be written in terms of ε_0 and ε_1 as

$$t_{g,RSS}=\bar{Y}(1+\varepsilon_0)\left[1+\varepsilon_1H_1+\varepsilon_1^2H_2+..........\right]$$

$$=\bar{Y}\left[1+\varepsilon_0+\varepsilon_1H_1+\varepsilon_1^2H_2+\varepsilon_0\varepsilon_1H_1+o(\varepsilon)\right]$$

Now Bias and MSE of the estimator $t_{g,RSS}$ to the first degree of approximation are respectively given by

$$B(\ t_{g,RSS})=E(\ t_{g,RSS})-\bar{Y}$$

Here $E(t_{g,RSS})=E\left[\varepsilon_0+\varepsilon_1H_1+\varepsilon_1^2H_2+\varepsilon_0\varepsilon_1H_1\right]$

$$\Rightarrow B(t_{g,RSS})=\bar{Y}\left[\theta\left(H_2C_x^2+H_1\rho_{xy}C_yC_x\right)-\left(H_2W_{x(i)}^2+W_{yx(i)}H_1\right)\right] \quad (2.3)$$

Now $MSE(t_{g,RSS})=E\left[t_{g,RSS}-\bar{Y}\right]^2$

again neglecting higher order terms, we have

$$MSE(t_{g,RSS})=\bar{Y}^2E\left[\varepsilon_0^2+H_1^2\varepsilon_1^2+2H_1\varepsilon_0\varepsilon_1\right]^2$$

$$\Rightarrow MSE(t_{g,RSS})=\bar{Y}^2\begin{bmatrix}\theta\left(C_y^2+H_1^2C_x^2+2H_1\rho_{xy}C_yC_x\right)\\-\left(W_{y[i]}^2+H_1^2W_{x(i)}^2+2H_1W_{yx(i)}\right)\end{bmatrix} \quad (2.4)$$

The optimum value of H_1 to minimize the MSE of $t_{g,RSS}$ can be easily found as follows

$$\frac{\partial MSE(t_{g,RSS})}{\partial H_1}=0$$

$$\Rightarrow H_1^*=-\frac{\left(\theta\rho_{xy}C_xC_y-W_{yx(i)}\right)}{\left(\theta C_x^2-W_{x(i)}^2\right)}$$

$$=-\frac{Cov[\bar{x}_{(n)},\bar{y}_{[n]}]\big/\bar{X}\bar{Y}}{V[\bar{x}_{(n)}]\big/\bar{X}}$$

$\Rightarrow H_1^*=-\rho_{xy}\dfrac{C_y}{C_x}$ [because $Cov(\bar{x}_{(n)},\bar{y}_{[n]})=\beta V\left(\bar{x}_{(n)}\right)$]

when we replace H_1 by H_1^* in (2.4), we obtain minimum MSE of the proposed estimator as follows

$$\Rightarrow Min.MSE(t_{g,RSS})=\bar{Y}^2\theta\left[C_y^2\left(1-\rho_{xy}^2\right)\right]-A \quad (2.5)$$

where $A=\bar{Y}^2\left(W_{y[i]}-H_1^*W_{x(i)}\right)^2$

By this way , we can write (2.5) as

$$MSE(t_{g,RSS})\cong MSE\left(t_g\right)-A$$

It is easily shown that minimum mean square error of the proposed estimator given in (2.5) using ranked set sampling is always smaller than the MSE of estimator, suggested by Srivastava in (1.3), because A is a non negative value. Thus the ratio estimator, product estimator and power transformation estimator using ranked set sampling are the special cases of the proposed estimators defined in (2.1). While the regression estimator and difference estimator are not special cases of the estimator $t_{g,RSS}$ .Then adapting the estimator in (1.4) to the wider class of estimator for the population mean proposed by Srivastava (1980), we develop the following estimator for wider class of estimator using ranked set sampling is

$$t_{w,RSS} = H\left[\bar{y}_{[n]}, u\right] \quad (2.6)$$

where $H\left[\bar{y}_{[n]}, u\right]$ is a function of $\bar{y}_{[n]}$ and u ,satisfies the following regularity conditions:

- The point $\left[\bar{y}_{[n]}, u\right]$ assumes the value in a closed convex subset R_2 of two dimensional real space containing the point $(\bar{Y},1)$.
- The function $H\left[\bar{y}_{[n]}, u\right]$ is continuous and bounded in R_2.
- $(\bar{Y},1)=\bar{Y}$ and $H_0(\bar{Y},1)=1$, where $H_0(\bar{Y},1)$ denotes the first order partial derivative of H with respect to $\bar{y}_{[n]}$.
- The first and second order partial derivatives of $H\left[\bar{y}_{[n]}, u\right]$ exist and are continuous and bounded in R_2.

Expanding $H\left[\bar{y}_{[n]}, u\right]$ about the point in a second order Taylor series, we have

$$t_{w,RSS} = H\left[\bar{y}_{[n]}, u\right] = H\left[\bar{Y} + \left(\bar{y}_{[n]} - \bar{Y}\right), 1 + (u-1)\right]$$

Using the above regularity conditions and $\left(\frac{\partial H}{\partial \bar{y}_{[n]}}\right)_{\bar{y}_{[n]}=\bar{Y}, u=1}$ =1, we have

$$t_{w,RSS} = \begin{bmatrix} \bar{y}_{[n]} + (u-1)H_1 + (u-1)^2 H_2 \\ + \left(\bar{y}_{[n]} - \bar{Y}\right)(u-1)H_3 + \left(\bar{y}_{[n]} - \bar{Y}\right)^2 H_4 + \ldots \end{bmatrix} \quad (2.7)$$

where $H_1 = \left(\frac{\partial H}{\partial \bar{y}_{[n]}}\right)_{\bar{y}_{[n]}=\bar{Y}, u=1}$, $H_2 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial u^2}\right)_{\bar{y}_{[n]}=\bar{Y}, u=1}$

$$H_3 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial \bar{y}_{[n]} \partial u}\right)_{\bar{y}_{[n]}=\bar{Y}, u=1} \;\&\; H_4 = \left(\frac{1}{2}\frac{\partial^2 H}{\partial^2 \bar{y}_{[n]}^2}\right)_{\bar{y}_{[n]}=\bar{Y}, u=1}$$

Now Bias and MSE of the estimator $t_{w,RSS}$ to the first degree of approximation are respectively given by

$$B(t_{w,RSS}) = E(t_{w,RSS}) - \bar{Y}$$

The estimator $t_{w,RSS}$ can easily be written in terms of ε_0 and ε_1 as $t_{w,RSS}$

$$= \bar{Y}[1+\varepsilon_0] + \varepsilon_1 H_1 + \varepsilon_1^2 H_2 + \bar{Y}\varepsilon_0\varepsilon_1 H_3 + \bar{Y}^2\varepsilon_0^2 H_4 + \ldots$$

Now

$$B(t_{w,RSS}) = H_2 E(\varepsilon_1^2) + H_3\bar{Y}E(\varepsilon_0\varepsilon_1) + H_4\bar{Y}^2 E(\varepsilon_0^2)$$

$$\Rightarrow B(t_{w,RSS}) = \begin{bmatrix} \theta\left(H_2C_x^2 + H_4\bar{Y}^2C_y^2 + \bar{Y}\rho_{xy}C_xC_y\right) \\ -\left(H_2W_{x(i)}^2 + H_4\bar{Y}^2W_{y[i]}^2 + H_3\bar{Y}W_{yx(i)}\right) \end{bmatrix} \quad (2.8)$$

and $MSE(t_{w,RSS}) = E\left[t_{w,RSS} - \bar{Y}\right]^2 = E\left[\bar{Y}\varepsilon_0 + \varepsilon_1 H_1\right]^2$

$$\Rightarrow MSE(t_{w,RSS}) = \begin{bmatrix} \theta\left(\bar{Y}^2C_y^2 + H_1^2C_x^2 + 2H_1\bar{Y}\rho_{xy}C_yC_x\right) \\ -\left(\bar{Y}^2W_{y[i]}^2 + H_1^2W_{x(i)}^2 + 2H_1\bar{Y}W_{yx(i)}\right) \end{bmatrix} \quad (2.9)$$

On differentiating (2.9) with respect to H_1 and equating to zero, we obtain the optimum value of H_1 denoted by H_1^{**} is

$$\frac{\partial MSE(t_{w,RSS})}{\partial H_1} = 0$$

$$\Rightarrow H_1^{**} = -\bar{Y}\frac{\left(\theta\rho_{xy}C_xC_y - W_{yx(i)}\right)}{\left(\theta C_x^2 - W_{x(i)}^2\right)}$$

$$\Rightarrow H_1^{**} = -\bar{Y}\rho_{xy}\frac{C_y}{C_x}$$

Replacing H_1 by H_1^{**} in (2.9), we obtain minimum MSE of the estimator $t_{w,RSS}$ as follows

$$Min.MSE(t_{w,RSS}) = \bar{Y}^2 \begin{bmatrix} \theta C_y^2\left(1-\rho_{xy}^2\right) \\ -\left(W_{y[i]} - \rho_{xy}\frac{C_y}{C_x}W_{x(i)}\right)^2 \end{bmatrix}$$

$$\Rightarrow Min.MSE(t_{w,RSS}) = \bar{Y}^2\theta\left[C_y^2\left(1-\rho_{xy}^2\right)\right] - A$$
(2.10)

Again we can write (2.10) as

$$MSE(t_{w,RSS}) \cong MSE(t_w) - A$$

So it is proved that the Min. MSE of the estimator $t_{w,RSS}$ using ranked set sampling is always smaller than the MSE of estimator, suggested by Srivastava given in (1.6), because A is a non negative value. As a result, show that the suggested estimator $t_{w,RSS}$ is more efficient than the estimator t_w.

References

1) Kadilar, C.,Unyazici, Y. and Cingi H.(2009), Ratio estimator for the population mean using ranked set sampling, Stat. papers,50,301-309.

2) McIntyre, G.A.(1952), A method of unbiased selective sampling using ranked sets, Australian Journal of Agricultural Research, 3, 385-390.

3) Prasad,B. (1989), Some improved ratio type estimators of population mean and ratio in finite population sample surveys. Commun Stat Theory Methods 18:379–392

4) Singh, Sarjinder (2003), Advanced Sampling Theory with Application, Vol. 1, Kluwer Academic Publishers, Netherlands.

5) Srivastava, S.K.(1971), A generalized estimator for the mean of a finite population using mutli auxiliary information, J. Amer. Statist. Assoc., 66, 404-407.

6) Srivastava, S.K.(1980), A class of estimator using auxiliary information in sample surveys. Canad. J. Statist., 8(2), 253-254.