J. Statistics
1998(1), 23-32

IMPROVEMENT OVER RATIO AND RATIO-TYPE ESTIMATOR

By

JAVID SHABBIR
(Department of Statistics, Government College Bahawalnager)

and

MUNIR AKHTAR
(Department of Statistics, Islamia university Bahawalpur)

ABSTRACT

We propose an estimator that is more efficient than the ratio estimator and, under certain conditions, than Ray and Sing's (1981) estimator up to the first degree approximation. The mean square error (MSE) is also studied by using the two phase sampling technique.

Key Words: Ratio-type estimator, Bias, MSE, Two Phase Sampling

INTRODUCTION

Let a sample of size n be drawn by simple random sample (SRS) from a population of size N. Two variables, y and x, are measured on each member of the sample, giving means $\bar{y}$ and $\bar{x}$; the corresponding population means are $\bar{Y}$ and $\bar{X}$. Ray and Singh (1981) proposed a difference-cum-ratio-type estimator for $\bar{Y}$, namely $\bar{y}_s = (\bar{y} + \bar{x}^w - \bar{X}^w)(\bar{X}/\bar{x})$ in which w is the weight. Along the same lines we propose an estimator

$\bar{y}_j = (\bar{y} + \bar{x}/\bar{X} - \bar{x}^{w+1}/\bar{X}^{w+1})(\bar{X}/\bar{x})$ which seems to be an improvement over ratio estimator $\bar{y}_r = (\bar{y}/\bar{x})\bar{X}$ as well as over $\bar{y}_s$ under suitable certain for w. Our estimator $\bar{y}_j$ includes the special case of $\bar{y}_r$ when w = 0, and for w = $\pm\frac{1}{2}$ two classes of ratio-type estimators are obtained. Before making a comparison, we define the following.

Let

$$\xi_o = (\bar{y} - \bar{Y})/\bar{Y}, \xi_1 = (\bar{x} - \bar{X})/\bar{X}, E(\xi_o) = E(\xi_1) = 0$$
$$E(\xi_o^{\ 2}) = (f/n)C_y^{\ 2}, E(\xi_1^{\ 2}) = (f/n)C_x^{\ 2}, E(\xi_o\xi_1) = (f/n)\rho C_y C_x$$

Where

$$C_y^{\ 2} = S_y^{\ 2}/\bar{Y}^2, C_x^{\ 2} = S_x^{\ 2}/\bar{X}^2, S_x^{\ 2} = \sum_{i=1}^{N}(X_i - \bar{X})^2/N-1$$

$$S_y^{\ 2} = \sum_{i=1}^{N}(Y_i - \bar{Y})^2/N-1, f = (N-n)/N, \rho = S_{xy}/S_x S_y,$$

$$S_{xy} = \sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})/[\sum_{i=1}^{N}(X_i - \bar{X})^2\sum_{i=1}^{N}(Y_i - \bar{Y})^2]^{1/2}$$

We assume that sample size is large enough so as to make $|\xi_o|$ and $|\xi_1|$ small. Let us now solve the estimator $\bar{y}_j$ for Bias and MSE by using the first degree approximation (i.e. up to the terms of order n^{-1}).

$$\bar{y}_j = (\bar{y} + \bar{x}/\bar{X} - \bar{X}^{w+1})(\bar{X}/\bar{x}) \quad (1)$$

or

$$\bar{y}_j = \bar{y}(\bar{X}/\bar{x}) - (\bar{x}/\bar{X})^w + 1$$
$$= \bar{Y}(1+\xi_o)(1+\xi_1)^{-1} - (1+\xi_1)^w + 1$$
$$= \bar{Y} + \bar{Y}(\xi_o - \xi_1 - \xi_o\xi_1 + \xi_1^{\ 2}) - \left[w\xi_1 + w(w-1)/2\xi_1^2\right]$$

By using the above results. Bias and MSE is given below

$$\text{Bias } (\bar{y}_j) = E\left[\bar{y}_j - \bar{Y}\right]$$
$$= \bar{Y}E\{\xi_o - \xi_1 - \xi_o\xi_1 + \xi_1^2\} - wE\{\xi_1 + (w-1)/2\xi_1^2\}$$
$$= (f/n)\{\bar{Y}(C_x^2 - \rho C_y C_x) - w(w-1)/2C_x^2\} \quad (2)$$

$$\text{MSE } (\bar{y}_j) = E[\bar{y}_j - \bar{Y}]^2$$

$$= \bar{Y}^2 E\{\xi_o - \xi_1 - \xi_o\xi_1 + \xi_1^2\}^2 + w^2 E\{\xi_1 + (w-1)/2\xi_1^2\}^2$$
$$- 2\bar{Y}wE\{\xi_0 - \xi_1 - \xi_0\xi_1 + \xi_1^2\}\{\xi_1 + (w-1)/2\xi_1^2\}$$
$$= \bar{Y}^2 E\{\xi_0^2 + \xi_1^2 - 2\xi_0\xi_1\} + w^2 E\{\xi_1^2\} - 2\bar{Y}wE\{\xi_0\xi_1 - \xi_1^2\}$$
$$= \bar{Y}^2 (f/n)\{C_y^2 + C_x^2 - 2\rho C_y C_x\} + w^2 (f/n) C_x^2$$
$$- 2\bar{Y}w(f/n)\{\rho C_y C_x - C_x^2\} \quad (3)$$

For finding the optimum value of w, we differentiate (3) w.r.tw and get

$$w_{opt} = \bar{Y}(C-1), where C = \rho C_y / C_x$$

putting the value of w_{opt} in (3), we get

$$\text{MSE}(\bar{Y}_j)_{opt} = \bar{Y}^2 (f/n)[\{C_y^2 + C_x^2 - 2\rho C_y C_x\} + \{\rho C_y / C_x - 1\}^2 C_x^2$$
$$- 2\{\rho C_y / C_x - 1\}\{\rho C_y C_x - C_x^2\}]$$
$$= S_y^2 (f/n)\{1-\rho^2\}$$

Which is equal to the corresponding expression for the linear regression estimator and is less than that for the ratio estimator $\bar{y}_r$.

Putting w=0 and w= $\pm$ 1/2 in (2) and (3), we get the following Bias and MSEs of $\bar{y}_r$ and, $\bar{y}_{j1}$ and $\bar{y}_{j2}$, where $\bar{y}_{j1}$ and $\bar{y}_{j2}$ are for w = 1/2 and w = -1/2 respectively.

SAMPLING BIAS

$$\text{Bias } (\bar{y}_r) = \bar{Y}(f/n)\{C_x^2 - \rho C_y C_x\} \quad (4)$$

$$\text{Bias } (\bar{y}_{j1}) = \bar{Y}(f/n)\{C_x^2 - \rho C_y C_x\} + 1/8(f/n)C_x^2 \quad (5)$$

$$\text{Bias } (\bar{y}_{j2}) = \bar{Y}(f/n)\{C_x^2 - \rho C_y C_x\} - 3/8(f/n)C_x^2 \quad (6)$$

and also from Ray and Singh (1981)

$$\text{Bias } (\bar{y}_.) = \bar{Y}(f/n)\{C_x^2 - \rho C_y C_x\} - 5/8(f/n)\bar{X}^{-0.5} C_x^2 \quad (7)$$

$$\text{Bias } (\bar{y}^*) = \bar{Y}(f/n)\{C_x^2 - \rho C_y C_x\} + 7/8(f/n)\bar{X}^{0.5} C_x^2 \quad (8)$$

SAMPLING MEAN SQUARE ERROR (MSE)

$$\text{MSE }(\bar{y}_r) = \bar{Y}^2 (f/n)\{C_y^2 + C_x^2 - 2\rho C_y C_x\} \tag{9}$$

$$\text{MSE }(\bar{y}_{j1}) = \bar{Y}^2 (f/n)\{C_y^2 + C_x^2 - 2\rho C_y C_x\} + 1/4(f/n)C_x^2 - \bar{Y}(f/n)\{\rho C_y C_x - C_x^2\} \tag{10}$$

$$\text{MSE }(\bar{y}_{j2}) = \bar{Y}^2 (f/n)\{C_y^2 + C_x^2 - 2\rho C_y C_x\} + 1/4(f/n)C_x^2 + \bar{Y}(f/n)\{\rho C_y C_x - C_x^2\} \tag{11}$$

From Ray and Singh (1981), we have

$$\text{MSE }(\bar{y}_s) = \bar{Y}^2 (f/n)\{C_y^2 + C_x^2 - 2\rho C_y C_x\} + 1/4\bar{X}(f/n)C_x^2 \quad \bar{Y}(f/n)\bar{X}^{0.5}\{\rho C_y C_x - C_x^2\} \tag{12}$$

$$\text{MSE}(y^*_s) = \bar{Y}^2 (f/n)\{C_y^2 + C_x^2 - 2\rho C_y C_x\} + 1/4(f/n)\bar{X}^{-1} C_x^2 - \bar{Y}(f/n)\bar{X}^{0.5}\{\rho C_y C_x - C_x^2\} \tag{13}$$

COMPARISON

By (4) and (5)

$$\text{Bias }(\bar{y}_{j1}) = Bias(\bar{y}_r) + 1/8(f/n)C_x^2$$

$$\left[Bias(\bar{y}_{j1})\right]^2 < \left[Bias(\bar{y}_r)\right]^2 \quad if\, Bias(\bar{y}_r) < 0 \tag{14}$$

By (4) and (6)

$$\text{Bias }(\bar{y}_{j2}) = Bias(\bar{y}_r) - 3/8(f/n)C_x^2$$

$$\left[Bias(\bar{y}_{j2})\right]^2 < \left[Bias(\bar{y}_r)\right]^2 \, if\, Bias(\bar{y}_r) < 0 \tag{15}$$

Similarly by (9) and (10)

$$\text{MSE }(\bar{y}_{j1}) = MSE(\bar{y}_r) + 1/4(f/n)C_x^2 - \bar{Y}(f/n)\{\rho C_y C_x - C_x^2\}$$

$$\text{MSE } (\bar{y}_{j1}) < MSE(\bar{y}_r) \text{ if } C > 1 + 1/(4\bar{Y}) \tag{16}$$

By (9) and (11)

$$\text{MSE } (\bar{y}_{j2}) = MSE(\bar{y}_r) + 1/4(f/n)C_x^2 + \bar{Y}(f/n)\{\rho C_y C_x - C_x^2\}$$

$$\text{MSE } (\bar{y}_{j2}) < MSE(\bar{y}_r) \text{ if } C < 1 - 1/(4\bar{Y}) \tag{17}$$

As $1/(4\bar{Y})$ is a small quantity so MSE $(\bar{y}_{j1})$ and MSE $(\bar{y}_{j2})$ are less than MSE $(\bar{y}_r)$ under the conditions stated above.

Also

$$MSE(\bar{y}_{j1}) < MSE(\bar{y}_s) \text{ if } C > 1 + (1 - \bar{X}) / \left[4\bar{Y}\left\{1 + \bar{X}^{0.5}\right\}\right] \tag{18}$$

and

$$MSE(\bar{y}_{j2}) \angle MSE(\bar{y}^*_s) \text{ if } C < 1 + (1 - \bar{X}^{-1}) / \left[4\bar{Y}\left\{1 + \bar{X}^{0.5}\right\}\right] \tag{19}$$

Both quantities

$(1 - \bar{X}) / \left[4Y\left\{1 + \bar{X}^{0.5}\right\} and\,(1 - \bar{X}^{-1}) / 4\bar{Y}\left\{1 + \bar{X}^{0.5}\right\}\right]$ are small,

therefore we have an improvement over estimators y_s and y^*_s under the above conditions.

Examples

All calculations are made by ignoring the sampling fraction.

Population 1, from Coachran (1977, pp. 173, 196)

$S_y^2 = 6409, S_x^2 = 3898, \rho = 0.887, \bar{X} = 44.45, \bar{Y} = 56.47$

Population 2, from Biradar and Singh (1992)

$S_y^2 = 18.71, S_x^2 = 12.50, S_{yx} = 13.77, \rho = 0.90, \bar{X} = 5.66, \bar{Y} = 7.47$

Population 3, from Murthy (1967, (33-48), pp. 178)

$C_y = 0.47, Cx = 0.09, \rho = 0.29, \bar{X} = 277.19, \bar{Y} = 106.69.$

Results of these populations are given in Table 1.

Table 1: Results of above three populations

Biases	pop.1	pop.2	pop.3	MSEs	pop.1	pop.2	pop.3
Bias $(\bar{y}_r)$	11.67	0.4850	-0.4445	MSE $(\bar{y}_r)$	1466.24	4.15	2327.39
Bias $(\bar{y}_{j1})$	11.92	0.5338	-0.4435	MSE $(\bar{y}_{j1})$	1475.92	4.73	2326.95
Bias $(\bar{y}_{j2})$	10 93	0.3385	-0.4476	MSE $(\bar{y}_{j2})$	1455.06	3.76	2327.84
Bias $(\bar{y}_s)$	3.46	-0.0958	-0.5288	MSE $(\bar{y}_s)$	1410.31	3.09	2334.79
Bias $(\dot{\bar{y}}_s)$	23.16	1.2982	-0.3366	MSE $(\dot{\bar{y}}_s)$	1544.07	5.32	2320.00

In populations 1 and 2, the condition for inequalities (15), (17) and (19) are satisfied whilst in population 3, the remaining conditions (14), (!6) and (18) are fulfilled.

For these three populations we also calculate the condition values in Table 2 for the conditions mentioned before.

Table 2: Condition values for the populations

Pop	C=pCy/Cx	$1+1/(4\bar{Y})$	$1-1/(4\bar{Y})$	$1+(1-\bar{X})/[4\bar{Y}(1+\bar{X}^{0.5})]$	$1+(1-\bar{X}^{-1})/[4\bar{Y}(1+\bar{X}^{0.5})]$
1	0.90	1.00	0.99	0.97	1.00
2	0.84	1.03	0.97	0.95	1.00
3	1.51	1.00	1.00	0.96	1.00

For a large sample size, the effect of bias is negligible. We observed that our proposed estimator effectively gives precision over ratio and, Ray and Sing's estimator under certain conditions. So in that circumstances we prefer our proposed estimator $\bar{y}_j$. We also use this estimator in two phase sampling.

TWO PHASE SAMPLING

In this section, we shall evaluate the Bias and MSE of the proposed estimator $\bar{y}_j$ (equation (1)) when information on the auxiliary character $\bar{X}$ is unknown. Therefore it is cheaper to take a sample in two cases. In first case we shall take a sample size n' to estimate $\overline{X}$ and then in second case a sub sample of it of size $n(i.e.\ n < n')$ or another independent of it.

Let we define the following relations.

$$\xi_1' = (\overline{x'} - \overline{X})/\overline{X}, E(\xi_1') = 0, E(\xi_1'^2) = (f'/n')C_x^2, f'/n' = (N-n')/(Nn')$$

$$E(\xi_0\xi'_1) = (f'/n')\rho C_y C_x, E(\xi_1\xi'_1) = (f'/n')C_x{}^2] \quad (20)$$

$\xi_0\xi_1 E(\xi_0), E(\xi_1), E(\xi_0^2), E(\xi_1^2)$ have already defined be before.

Let us define the estimator $\overline{y}_j'$ like $\overline{y}_j$ (equation (1)), just replacing $\overline{X}$ by $\overline{x}'$

$$\overline{y}'_j = \overline{y}(\overline{x}'/\overline{x}) - (\overline{x}/\overline{x}')^w + 1 \quad (21)$$

Where $\overline{x'}$ is the first phase large sample mean based on n' units, $\overline{x}$ and $\overline{y}$ are the second phase sample means based on n units.

Using the relation (20) in (21), we get

$$\overline{y}'_j = \overline{y}(1+\xi_0)(1+\xi_1')(1+\xi_1)^{-1} - \left[(1+\xi_1)^w(1+\xi'_1)^{-w}\right] + 1$$

$$= \overline{Y} + \overline{Y}\left[\xi_0 - \xi_1 + \xi^2{}_1 + \xi'_1 - \xi_1\xi_1' - \xi_0\xi_1 + \xi_0\xi'_1\right]$$

$$- w\left[\xi_1 - \xi'_1 + (w+1)/2\xi'^2{}_1 - w\xi_1\xi_1' + (w-1)/2\xi^2{}_1\right] \quad (22)$$

Case (i) : When the second phase sample is a subset of first phase sample. By using (20) and (22), the Bias and MSE of (21) to the first degree approximation is given as

$$\text{Bias}\ (\overline{y}'_j) = E\left[\overline{y}'_j - \overline{y}\right]$$

$$= \overline{Y}(1/n - 1/n')\left[C_x^2 - \rho C_y C_x\right] \quad (23)$$

$$\text{MSE}\ (\overline{y'}_j) = E\left[\overline{y'}_j - \overline{Y}\right]^2$$

By using (22), we get

$$\text{MSE}\ (\overline{y}'_j) = \overline{Y}^2 E\left[\xi_0^2 + \xi^2{}_1 + \xi'^2{}_1 - 2\xi_0\xi_1 + 2\xi_0\xi'_1 - 2\xi_1\xi'_1\right]$$

$$+w^2E\left[\xi^2{}_1+\xi'^2{}_1-2\xi_1\xi'_1\right]-2\bar{Y}w\left[\xi_0\xi_1-\xi_0\xi'_1-\xi^2{}_1+2\xi_1\xi'_1-\xi'^2{}_1\right] \quad (24)$$

By using (20) in above equation, we get

$$\text{MSE }(\bar{y}'_j)=\bar{Y}^2\left[(1/N-1/n)C_y^2+(1/n-1/n')C_x^2-2(1/n-1/n')\rho C_yC_x\right]$$

$$+w^2(1/n-1/n')C_x^2-2\bar{Y}w(1/n-1/n')[\rho C_yC_x-C_x^2]$$

The optimum value is the same as we have calculated before i.e.

$$w_{opt}=\bar{Y}\text{(C-1), where C}=\rho\text{ Cy/Cx} \quad (25)$$

therefore by using (24) and (25), ,we have

$$\text{MSE }(\bar{y}'_j)_{opt}=S_y^2\left[(1/n-1/N)-(1/n-1/n')\rho^2\right] \quad (26)$$

which is equal to the variance of double sampling linear regression estimator.

Case (ii) : when the second phase sample is independent of first phase sample.

For independence, we use the following relation

$$E(\xi_0\xi_1')=E(\xi_1\xi'_1)=0$$

By using (20) and (22), we get the Bias and MSE as we have calculated for Case (i)

$$\text{Bias }(\bar{y}'_j)=\bar{Y}\left[(1/N-1/n)\right]\left[C_x^2-\rho C_yC_x\right]-w^2(1/n+1/n'-2/N)C_x^2 \quad (27)$$

$$\text{MSE }(\bar{y}'_j)=\bar{Y}^2[(1/N-1/n)C_y^2+(1/n+1/n'-2/N)C_x^2-2\left[(1/N-1/n)\rho C_yC_x\right]$$

$$+w^2(1/n+1/n'-2/N)C_x^2-2\bar{Y}w[(1/N-1/n)\rho C_yC_x$$

$$-(1/n+1/n'-2/N)C_x^2] \quad (28)$$

we get the optimum value of w by minimizing (28)

$$W_{opt} = \bar{Y}(C-1), where C = (1/N - 1/n)/(1/n + 1/n' - 2/N)\rho Cy/Cx \tag{29}$$

By (28) and (29), we have

$$\text{MSE}\,(\bar{Y}'_j)_{opt} = S_y^2(1/N - 1/n)\left[1 - (1/N - 1/n)/(1/n + 1/n' - 2/N)\rho^2\right]$$

It is noted that for two phase sampling if n' =N, then both estimators give same precision i.e. MSE $(\bar{y}_j) = MSE(\bar{y}'_j)$

REFERENCES

[1] **Cochran, W.G (1977).** *Sampling Techniques, third edt. John Willey & Sons, New York.*

[2] **Biradar R. S. and Housila, P.S (1992).** *On a class of almost unbiased ratio Estimator, Biometrical Journal 8, 937-944.*

[3] **Murthy, M.N (1977).** *Sampling Theory and Methods, Statistical Publishing Society, India.*

[4] **Ray, S.K. and Singh, R. K (1981).** *Difference-Cum ration type estimators, Journal of Indian Statistical Association, 19, 147-151.*

NOTES