ISSN 2250-3153

Some Modified Unbiased Estimators of Population Mean

***Shashi Bahl, **Bhawna Mehta**

* Professor, Deptt. of Statistics, MDU, Rohtak
**Research Scholar, MDU, Rohtak

I. Abstract

The use of supplementary information on auxiliary variables in sample surveys was extensively discussed by Cochran and Jessen. In their work they showed that regression estimator is superior to the other estimators (viz. ratio and mean per unit estimators etc.).This paper proposes a new kind of estimator based on appropriate weighing of the sample means of the main and the auxiliary variables. It is shown that the proposed estimator is more efficient when compared with the regression, ratio and the mean per unit estimator under certain restrictions on the correlation coefficient between main and the auxiliary variables.

II. Introduction

Let a simple random sample of size n be drawn without replacement from a finite population of N units and the variables observed are Y and X. The variable Y is the variable of interest so-called main variable and the variable X is called the auxiliary variable. The population means of the variables Y and X are $\overline{Y}$ and $\overline{X}$ respectively, where $\overline{X}$ is assumed to be known. Let $\overline{y}$ and $\overline{x}$ be the unbiased sample estimators of $\overline{Y}$ and $\overline{X}$. Let ρ be the correlation coefficient between the variables Y and X, given by the relation

$$\rho = \frac{S_{XY}}{S_X S_Y} \qquad (2.1)$$

where

$S_X^2 =$ Variance of X,

$S_Y^2 =$ Variance of Y,

$S_{XY} =$ Covariance of X and Y.

The usual estimators of the simple, ratio, and the regression for estimating the mean of the main variable Y with the help of auxiliary variable are given by Cochran and Jessen as follows:

Simple: $\overline{y} = [(1).\,\overline{y} + 0.(\overline{X} - \overline{x}\,)]$ (2.2)

Ratio: $\overline{y}_{rat} = [(0).\,\overline{y} + \frac{\overline{y}}{\overline{x}}.\,(\,\overline{X} - 0)$ (2.3)

Regression: $\overline{y}_{reg} = [(1).\,\overline{y} + (b).(\,\overline{X} - \overline{x})$ (2.4)

where 0 and 1 are the weights, and

$$\mathrm{b} = \frac{\mathrm{S_{XY}}}{\mathrm{S_X^2}} \qquad (2.1.5)$$

III. Generalized Estimators

An estimator denoted by $\overline{y}_{gen}$ is proposed:

$\overline{y}_{gen} = W_1(\overline{y}) + W_2(\text{ ratio})$ (3.1)

$\overline{y}_{gen} = W_1(\overline{y}) + W_2(\frac{\overline{y}}{\overline{x}}\overline{X})$

where W_1 *and* W_2 are constants such that

ISSN 2250-3153

$W_1 + W_2 = 1$ (3.2)

$E\left(\overline{y}_{gen}\right) = E\ (W_1(\overline{y}) + W_2(\frac{\overline{y}}{\overline{X}}\overline{X}))$

$\mathrm{E}(\overline{y}_{gen}) = (W_1 + W_2)\ \overline{Y}$ (3.3)

$\mathrm{Var}(\overline{y}_{gen}) = E[(\ \overline{y}_{gen} - E\left(\overline{y}_{gen}\right)]^2$

$= \mathrm{E}[W_1(\overline{y}) + W_2(\frac{\overline{y}}{\overline{X}}\overline{X}) - (W_1 + W_2)\ \overline{y}\ ]^2$

$= \mathrm{E}[W_1(\ \overline{y} - \overline{Y}\ )^2 + W_2(\frac{\overline{y}}{\overline{X}}\overline{X} - \ \overline{Y}\ )]^2$

$= \mathrm{E}[{W_1}^2(\ \overline{y} - \overline{Y}\ )^2 + {W_2}^2(\frac{\overline{y}}{\overline{X}}\overline{X} - \ \overline{Y}\ )]^2 + 2W_1W_2\ \left(\ \overline{y} - \overline{Y}\ \right)(\frac{\overline{y}}{\overline{X}}\overline{X} - \ \overline{Y}\ )]$

$= \mathrm{E}[{W_1}^2(\ \overline{y} - \overline{Y}\ )^2 + {W_2}^2(\frac{\overline{y}}{\overline{X}}\overline{X} - \ \overline{Y}\ )]^2 + 2W_1W_2\ \left(\ \overline{y} - \overline{Y}\ \right)(\frac{\overline{y}}{\overline{X}}\overline{X} - \ \overline{Y}\ )]$

$= {W_1}^2\mathrm{E}\left(\ \overline{y} - \overline{Y}\ \right)^2 + {W_2}^2\mathrm{E}\left(\hat{R}\overline{X} - E\left(\hat{R}\overline{X}\ \right)\right)^2 + 2W_1W_2\ E\left(\hat{R}\overline{X} - \ \overline{Y}\ \right)$

$\left(\ \overline{y} - \overline{Y}\ \right)]$

$= {W_1}^2 var(\overline{Y}\ ) + {W_2}^2\mathrm{var}(\hat{R}\overline{X}) + 2W_1W_2\ E\ (\hat{R}\overline{X} - \ \overline{Y}\ )\left(\ \overline{y} - \overline{Y}\ \right)]$

$= {W_1}^2 var(\overline{Y}\ ) + {W_2}^2\mathrm{var}(\hat{R}\overline{X}) + 2W_1W_2\ E\ [\hat{R}\overline{X}\left(\ \overline{y} - \overline{Y}\ \right) - \overline{Y}\left(\ \overline{y} - \overline{Y}\ \right)]$

$= {W_1}^2 var\left(\overline{Y}\ \right) + {W_2}^2\mathrm{var}\left(\hat{R}\overline{X}\right) + 2W_1W_2\ \overline{X}\ cov\left(\hat{R}, \overline{Y}\ \right)$

$= \frac{(1-f)}{n}(\ W_1\ + W_2\ )^2 S_y^2 - \ 2\ R\rho\ \mathrm{S_x}\ \mathrm{S_y} W_2\ (W_1 + W_2) + \ {W_2}^2 S_y^2\ R^2)$

$\mathrm{Var}\ (\ \overline{y}_{gen}) \ = \frac{(1-f)}{n}(\ S_y^2 - \ 2\ R\rho\ \mathrm{S_x}\ \mathrm{S_y} W_2 \ + \ {W_2}^2 S_y^2\ R^2)$ (3.4)

where $f = \frac{n}{N}$

The constants W_1 and W_2 are chosen such that the variance of the proposed estimator is minimum. The following theorem obtains the value of $W_1\ and\ W_2$

This variance of the proposed estimator given by equation (2.2.4) is minimum if

$-\ 2\ R\rho\ \mathrm{S_x}\ \mathrm{S_y} + \ 2W_2 S_y^2\ R^2 = 0$

or if $W_2 = \frac{\rho}{R}\left(\frac{\mathrm{S_x}}{\mathrm{S_y}}\right)$ (3.5)

and $W_1 = (1 - W_2)$ (3.6)

Thus, the **minimum value** of the proposed estimator is given by

$\mathrm{Var}\ (\ \overline{y}_{gen}) = \frac{(1-f)}{n}(S_y^2 - \ \rho^2 S_x^2\ )$

IV. COMPARISON

To Compare the efficiencies of various estimators, the following variances of the mean per unit, ratio and regression estimators are required

Simple: var($\bar{y}$) = $\frac{(1-f)}{n}(S_y^2)$ (4.1)

Ratio: var($\bar{y}_{rat}$) = $\frac{(1-f)}{n}(S_y^2 + R^2S_x^2 - 2\rho R\, S_y\, S_x)$ (4.2)

Regression: var($\bar{y}_{reg}$)= $\frac{(1-f)}{n}(1 - \rho^2)S_y^2$ (4.3)

Theorem 4.1: The proposed estimator is more efficient than the mean per unit estimator if

Var($\bar{y}_{gen}$) < var($\bar{y}$) (4.4)

This implies
$\frac{(1-f)}{n}(S_y^2 - \rho^2S_x^2) < \frac{(1-f)}{n}(S_y^2)$

$(S_y^2 - \rho^2S_x^2) < S_y^2$

$or\ S_{xy} > 0$ (4.5)

Hence the proposed estimator is more efficient than the mean per unit estimator if $S_{xy} > 0$

Theorem4.2 :The proposed estimator is more efficient than the regression estimator if

Var($\bar{y}_{gen}$) < var($\bar{y}_{reg}$) (4.6)

This implies

$\frac{(1-f)}{n}(S_y^2 - \rho^2S_x^2) < \frac{(1-f)}{n}(1 - \rho^2)S_y^2$ (4.7)

$S_y^2 - \rho^2S_x^2 < (S_y^2 - \rho^2S_y^2)$

$or \qquad S_x > S_y$ (4.8)

Hence the proposed estimator is more efficient than the regression estimator if $S_x > S_y$

The equation (4.7) clearly suggests that if $S_x = S_y$ then the proposed estimator becomes the regression estimator.

V. CONSIDERING ANOTHER ESTIMATOR

Considering the estimator as:

$\bar{y}_{reg}$= W_1($\bar{y}$)+W_2(regression) (5.1)

$\bar{y}_{reg}$= $W_1(\bar{y})+W_2(\bar{y} + b(\bar{X} - \bar{x}))$

$where\ W_1 and\ W_2$ are constants such that
$W_1 + W_2 = 1$ (5.2)

Also we consider the estimator based on paired observations as follows:

$\overline{y}_{mg} = \overline{y} + W_1 (\overline{y} - \lambda_1 \overline{x}))$ (5.3)

$\overline{y}_{mm} = \overline{y} + W_2 (\overline{y} - \alpha_0 - \beta_0 \overline{x}))$ (5.4)

$\overline{y}_{sg} = \overline{y} + W_1 (\overline{x} - \lambda)$ (5.5)

$where\ W_1, W_2, \alpha_0, \beta_0, \lambda_1$ are constants with no restrictions.

$E\left(\overline{y}_{reg}\right) = E\ [W_1(\overline{y}) + W_2 (\overline{y} + b (\overline{X} - \overline{x}))]$

$E(\overline{y}_{reg}) = (W_1 + W_2)\ \overline{Y} = \overline{Y}$ (5.6)

$E\left(\overline{y}_{mg}\right) = E\ [\overline{y} + W_1 (\overline{y} - \lambda_1 \overline{x})]$

$= \overline{Y} + W_1 (\overline{Y} - \lambda_1 \overline{X})$ (5.7)

$E(\overline{y}_{mm}) = E\ [\overline{y} + W_2 (\overline{y} - \alpha_0 - \beta_0 \overline{x})]$

$= \overline{Y} + W_2 (\overline{Y} - \alpha_0 - \beta_0 \overline{X})$ (5.8)

$E\left(\overline{y}_{sg}\right) = E\ [\overline{y} + W_1 (\overline{x} - \lambda)]$

$= \overline{Y} + W_1 (\overline{X} - \lambda)$ (5.9)

$\mathrm{Var}(\overline{y}_{reg}) = E[(W_1 + W_2)\overline{y} + W_2\, b \left(\overline{X} - \overline{x}\right) - (W_1 + W_2)\overline{Y}\ ]^2$

$= \mathrm{E}[\left(\overline{y} - \overline{Y}\right) + W_2\, b \left(\overline{X} - \overline{x}\right)]^2$

$= var(\overline{y}) + {W_2}^2 \mathrm{b}^2 \mathrm{var}(\overline{x}) - 2W_2 b\ E\ (\overline{x} - \overline{X})\left(\overline{y} - \overline{Y}\right)]$

$= var\left(\overline{Y}\right) + {W_2}^2 \mathrm{b}^2 \mathrm{var}(\overline{x}) - 2W_2 b\ cov(\overline{x}, \overline{y})$

$\mathrm{Var}(\overline{y}_{reg}) = \frac{(1-f)}{n}(S_y^2 - 2\, W_2\, b\rho\, \mathrm{S_x}\, \mathrm{S_y} + {W_2}^2 S_x^2 b^2)$ (5.10)

where $f = \frac{n}{N}$

$\mathrm{Var}(\overline{y}_{mg}) = \frac{(1-f)}{n}(S_y^2 + {W_1}^2 S_y^2 + {\lambda_1}^2 S_x^2 + 2W_1 S_y^2 - 2\,\lambda_1 W_1\, \rho\, S_x\, S_y - 2\lambda_1 \rho\, S_x\, S_y)$ (5.11)

$\mathrm{Var}(\overline{y}_{mm}) = \frac{(1-f)}{n}(S_y^2 + {W_2}^2 S_y^2 + {W_2}^2 \beta_0^2 S_x^2 + 2W_2 S_y^2 - 2\, {W_2}^2 \beta_0 \rho\, \mathrm{S_x}\, \mathrm{S_y} - 2\beta_0 W_2 \rho\, \mathrm{S_x}\, \mathrm{S_y})$ (5.12)

$\mathrm{Var}(\overline{y}_{sg}) = \frac{(1-f)}{n}(S_y^2 + {W_1}^2 S_x^2 + 2W_1 \rho\, \mathrm{S_x}\, \mathrm{S_y})$ (5.13)

The constants W_1 and W_2 are chosen such that the variance of the proposed estimator is minimum. The following theorem obtains the value of $W_1\ and\ W_2$

This variance of the proposed estimator $\overline{y}_{reg}$ given by equation (5.10) is **minimum** if

$2W_2 S_x^2 b^2 - 2\,\rho\, \mathrm{b}\mathrm{S_x}\, \mathrm{S_y} = 0$

or if $W_2 = \frac{\rho}{b}\left(\frac{S_y}{S_x}\right)$ (5.14)

and $W_1=(1-W_2)$ (5.15)

Thus, the minimum value of the estimator $\bar{y}_{reg}$ is given by

Var($\bar{y}_{reg}$)= $\frac{(1-f)}{n}(1-\rho^2)S_y^2$ which is same as the variance of the regression estimator.

To find the **minimum variance of the proposed estimator** $\boldsymbol{\bar{y}_{mg}}$ given by equation (5.11) we differentiate the equation w.r.t. W_1 and equate it to zero. Thus we get

$$W_1 S_y + S_y = \lambda_1 \rho S_x$$

Or $W_1 = \left(\frac{\lambda_1 \rho S_x - S_y}{S_y}\right)$ (5.16)

Thus, the minimum value of the estimator $\bar{y}_{mg}$ is obtained by substituting $W_1 = \left(\frac{\lambda_1 \rho S_x - S_y}{S_y}\right)$ in equation (2.4.11) and we get the minimum value as

Var($\bar{y}_{mg}$)= $\frac{(1-f)}{n}(S_y^2 - {\lambda_1}^2 \rho^2 {S_x}^2 + {\lambda_1}^2 {S_x}^2)$ (5.17)

Also the minimum value of the estimator $\bar{y}_{mg}$ is obtained if we differentiate the equation w.r.t. λ_1 and solving for λ_1

$$\lambda_1 = (1 + W_1)\left(\frac{\rho S_y}{S_x}\right)$$

and the minimum variance as

Var($\bar{y}_{mg}$)= $\frac{(1-f)}{n}(1-\rho^2)(1 + W_1)^2$ (5.18)

Similarly the minimum variance of the estimator $\bar{y}_{sg}$ is given by substituting

$W_1 = -\frac{\rho S_y}{S_x}$ in equation (5.13)

And we get the minimum variance of $\bar{y}_{sg}$ as

Var($\bar{y}_{sg}$) = $\frac{(1-f)}{n}(1-\rho^2)S_y^2$ which is same as the variance of the regression estimator.

VI. COMPARISON

6.1: The proposed estimator $\bar{y}_{reg}$ is more efficient than the mean per unit estimator if

Var($\bar{y}_{reg}$) < var($\bar{y}$) (6.1)

This implies

$$\frac{(1-f)}{n}(1-\rho^2)S_y^2 < \frac{(1-f)}{n}(S_y^2)$$

$$(1-\rho^2) < 0$$

$$i.e \qquad \rho^2 > 1$$

$or \qquad \rho > \pm 1$ (6.2)

Hence the proposed estimator is more efficient than the mean per unit estimator if

$$\rho > \pm 1$$

6.2: The proposed estimator $\overline{y}_{mg}$ is more efficient than the mean per unit estimator if

$\text{Var}(\overline{y}_{mg}) < \text{var}(\overline{y})$ (6.3)

This implies

$$\frac{(1-f)}{n}(S_y^2 + {W_1}^2 S_y^2 + {\lambda_1}^2 S_x^2 + 2W_1 S_y^2 - 2\lambda_1 W_1 \rho S_x S_y - 2\lambda_1 \rho S_x S_y) < \frac{(1-f)}{n}(S_y^2)$$

Hence the proposed estimator is more efficient than the mean per unit estimator if $\rho > \frac{({W_1}^2 S_y^2 + 2W_1 S_y^2 + {\lambda_1}^2 S_x^2)}{2\lambda_1 \rho S_x S_y (1+W_1)}$

(6.4)

6.3. The proposed estimator $\overline{y}_{mm}$ is more efficient than the mean per unit estimator if

$\text{Var}(\overline{y}_{mm}) < \text{var}(\overline{y})$ (6.5)

This implies

$$\frac{(1-f)}{n}(S_y^2 + {W_2}^2 S_y^2 + {W_2}^2 \beta_0^2 S_x^2 + 2W_2 S_y^2 - 2{W_2}^2 \beta_0 \rho S_x S_y - 2\beta_0 W_2 \rho S_x S_y) < \frac{(1-f)}{n}(S_y^2)$$

$$W_2 < \frac{2(\beta_0 \rho S_x S_y - S_y^2)}{(S_y^2 + \beta_0^2 S_x^2 - 2\beta_0 \rho S_x S_y)} \qquad (6.6)$$

Hence the proposed estimator is more efficient than the mean per unit estimator if $W_2 < \frac{2(\beta_0 \rho S_x S_y - S_y^2)}{(S_y^2 + \beta_0^2 S_x^2 - 2\beta_0 \rho S_x S_y)}$

(6.7)

6.4. The proposed estimator $\overline{y}_{sg}$ is more efficient than the mean per unit estimator if

$\text{Var}(\overline{y}_{sg}) < \text{Var}(\overline{y})$ (6.8)

This implies

$$\frac{(1-f)}{n}(S_y^2 + {W_1}^2 S_x^2 + 2W_1 \rho S_x S_y) < \frac{(1-f)}{n}(S_y^2)$$

$$W_1 < -2\rho \ S_y \qquad (6.9)$$

Hence the proposed estimator is more efficient than the mean per unit estimator if

$W_1 < -2\rho \ S_y$

VII. EMPIRICAL STUDY

For Empirical Study we consider the example dealing with a complete enumeration of 256 commercial pea orchards in North Carolina in June 1946 as given in Cochran (1983) , pp. 174-175,200. In example we have

$S_y^2 = 6009$ $S_{xy} = 4439$ $S_x^2 = 3898$ r= 0.887

$b = 1.14$ n=100 N=256 $\overline{X} = 44.45$ $\overline{Y} = 56.47$

β	λ	b	W1	W2	$\overline{Y}$ (gen)	$\overline{Y}$ (reg)	$\overline{Y}$ (mg)	$\overline{Y}$ (mm)
0.5	1	1.138789	-0.9	1.9	119.3773	30.7645	18.70958	180.343
0.5	1	1.138789	0	1	26.98626	5.81292	6.270469	98.30742
0.5	1	1.138789	1	0	36.61734	36.61734	62.02219	36.61734
1	2	1.138789	0.1	0.9	22.63047	6.120965	20.30005	58.91736
0.2	1	1.138789	0.5	0.5	17.02709	13.51403	24.99199	74.51151
0.5	0.5	1.138789	0.6	0.4	18.58119	16.90251	56.39851	57.57204

VIII. CONCLUSION

From the above study we have the conclusion that the proposed estimators are better than simple mean per unit estimators in many practical situations. Also $\overline{Y}$ (mg) is better than any other estimators if W_1 is negative or close to zero & $\overline{Y}$ (reg) is better if $W_2 > W_1$. Thus theoretically and from above study we see that the proposed estimators are better than existing ones in many practical situations under certain conditions.

REFERENCES

[1] 1. Cochran, William G. (1977), Sampling Techniques, 3rd edition. New York : John Wiley and Sons.
[2] 2. Jessen, R.J. (1978). Statistical surveys techniques. John Wiley and Sons.