International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 3 Issue 8 || December. 2015 || PP-23-28

On Estimation of Population Variance Using Auxiliary Information

Peeyush Misra
Department of Statistics, D.A.V.(P.G.) College, Dehradun- 248001, Uttarakhand, India

Abstract: *A generalized class of estimator representing a class of estimators using auxiliary information in the form of mean and variance is proposed. The expression for bias and mean square error are found and it is shown that the proposed generalized class of estimator is more efficient than few of the estimators available in the literature. An empirical study is also included as an illustration.*

Keywords : *Auxiliary information, Bias, Mean square error and Taylor's Series Expansion.*

I. INTRODUCTION

It is well known that the use of auxiliary information in sample surveys results in substantial improvement in the precision of the population parameters. By using the auxiliary information in different forms, estimators for population parameters mainly population mean and variance are studied and are available in the literature. Consider a finite population U with N units (U_1, U_2, . . . , U_N) for each of which the information is available on auxiliary variable X, Y being the study variable.
Let us denote by

$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ *= population mean of study variable Y*

$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ *= population mean of auxiliary variable X*

$S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2$ *= population variance of study variable Y*

$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2$ *= population variance of auxiliary variable X*

and $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^r\left(X_i - \overline{X}\right)^s$.

Also, let a sample of size n be drawn with simple random sample without replacement to estimate the population variance of the study variable Y.

Let $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ *= sample mean of study variable Y*

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ *= sample mean of auxiliary variable X*

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$ *= sample variance of study variable Y*

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ *= sample variance of auxiliary variable X.*

For simplicity, we assume that N is large enough as compared to n so that the finite population correction terms are ignored.

In order to have an estimate of population variance of the study variable Y, assuming the knowledge of mean and variance of auxiliary character X, proposed generalized class of estimator is given by

$$d_g = \hat{\theta} - g\left(\bar{y}^2, \bar{x}, s_x^2\right) \tag{1.1}$$

where $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$ *satisfying the validity conditions of Taylor's series expansion is a bounded function of* $\left(\bar{Y}^2, \bar{X}, S_X^2\right)$ *such that*

(i) $$g\left(\bar{Y}^2, \bar{X}, S_X^2\right) = \bar{Y}^2 \tag{1.2}$$

(ii) first order partial differential coefficient of $g\left(\bar{y}^2, \bar{x}, s_x^2\right)$ *with respect to* $\bar{y}^2$ *at T =* $\left(\bar{Y}^2, \bar{X}, S_X^2\right)$ *is unity, that is*

$$g_0 = \left(\frac{\partial}{\partial\left(\bar{y}^2\right)} g\left(\bar{y}^2, \bar{x}, s_x^2\right)\right)_T = 1 \tag{1.3}$$

(iii) second order partial differential coefficient of $g\left(\bar{y}^2, \bar{x}, s_x^2\right)$ *with respect to* $\bar{y}^2$ *at T =* $\left(\bar{Y}^2, \bar{X}, S_X^2\right)$ *is zero, that is*

$$g_{00} = \left(\frac{\partial^2}{\partial\left(\bar{y}^2\right)^2} g\left(\bar{y}^2, \bar{x}, s_x^2\right)\right)_T = 0 \tag{1.4}$$

II. BIAS AND MEAN SQUARE ERROR OF THE PROPOSED ESTIMATOR

In order to obtain bias and mean square error, let us denote by

$$\bar{y} = \bar{Y} + e_0$$

$$\bar{x} = \bar{X} + e_1$$

$$s_x^2 = S_X^2 + e_2$$

$$\hat{\theta} = \theta + e_3 \qquad \text{where} \quad \theta = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \tag{2.1}$$

with $E(e_0) = E(e_1) = E(e_2) = E(e_3) = 0$ *(2.2)*

$$E\left(e_0^2\right) = \frac{\mu_{20}}{n}$$

$$E\left(e_1^2\right) = \frac{\mu_{02}}{n}$$

$$E\left(e_2^2\right) = \frac{\mu_{02}^2}{n}(\beta_2 - 1)$$

$$E\left(e_3^2\right) = \frac{1}{n}(\mu_{40} + 4\bar{Y}\mu_{30} + 4\bar{Y}^2\mu_{20} - \mu_{20}^2)$$

$$E\left(e_0 e_1\right) = \frac{\mu_{11}}{n}$$

$$E\left(e_0 e_2\right) = \frac{\mu_{12}}{n}$$

$$E\left(e_1 e_2\right) = \frac{\mu_{03}}{n}$$

$$E\left(e_0 e_3\right) = \frac{1}{n}(\mu_{30} + 2\overline{Y}\,\mu_{20})$$

$$E\left(e_1 e_3\right) = \frac{1}{n}(\mu_{21} + 2\overline{Y}\,\mu_{11})$$

$$E\left(e_2 e_3\right) = \frac{1}{n}(\mu_{22} + 2\overline{Y}\,\mu_{12} - \mu_{02}\,\mu_{20}) \qquad (2.3)$$

Now expanding $t = g\left(\overline{y}^2, \overline{x}, s_x^2\right)$ *in the third order Taylor's series about the point* $T = \left(\overline{Y}^2, \overline{X}, S_X^2\right)$, *we have*

$$t = g\left(\overline{Y}^2, \overline{X}, S_X^2\right) + \left(\overline{y}^2 - \overline{Y}^2\right)g_0 + \left(\overline{x} - \overline{X}\right)g_1 + \left(s_x^2 - S_X^2\right)g_2$$

$$+ \frac{1}{2!}\left\{ \left(\overline{y}^2 - \overline{Y}^2\right)^2 g_{00} + \left(\overline{x} - \overline{X}\right)^2 g_{11} + \left(s_x^2 - S_X^2\right)g_{22} + 2\left(\overline{y}^2 - \overline{Y}^2\right)\left(\overline{x} - \overline{X}\right)g_{01}\right.$$

$$\left. + 2\left(\overline{y}^2 - \overline{Y}^2\right)\left(s_x^2 - S_X^2\right)g_{02} + 2\left(\overline{x} - \bar{\bar{X}}\right)\left(s_x^2 - S_X^2\right)g_{12}\right\}$$

$$+ \frac{1}{3!}\left\{\left(\overline{y}^2 - \overline{Y}^2\right)\frac{\partial}{\partial\left(\overline{y}^2\right)} + \left(\overline{x} - \overline{X}\right)\frac{\partial}{\partial\overline{x}} + \left(s_x^2 - S_X^2\right)\frac{\partial}{\partial s_x^2}\right\}^3 g\left(\overline{y}^{2^*}, \overline{x}^*, s_x^{*2}\right)$$

where g_0 , g_{00} *are already defined earlier and*

$$g_1 = \left(\frac{\partial}{\partial\overline{x}} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

$$g_2 = \left(\frac{\partial}{\partial s_x^2} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

$$g_{11} = \left(\frac{\partial^2}{\partial\overline{x}^2} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

$$g_{22} = \left(\frac{\partial^2}{\partial\left(s_x^2\right)^2} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

$$g_{01} = \left(\frac{\partial^2}{\partial\left(\overline{y}^2\right)\partial\overline{x}} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

$$g_{02} = \left(\frac{\partial^2}{\partial\left(\overline{y}^2\right)\partial s_x^2} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

$$g_{12} = \left(\frac{\partial^2}{\partial\overline{x}\,\partial s_x^2} g\left(\overline{y}^2, \overline{x}, s_x^2\right)\right)_T$$

and $\overline{y}^{2^*} = \overline{Y}^2 + h\left(\overline{y}^2 - \overline{Y}^2\right)$

$$\overline{x}^* = \overline{X} + h\left(\overline{x} - \overline{X}\right)$$

$s_x^{2*} = S_X^2 + h\left(s_x^2 - S_X^2\right)$ *for* $0 < h < 1$.

Now using the conditions given in (1.2), (1.3) and (1.4), we have to the first degree of approximation

$$t = \overline{Y}^2 + e_0^2 + 2\overline{Y}e_0 + e_1 g_1 + e_2 g_2 + \frac{1}{2!}\left\{ e_1^2 g_{11} + e_2^2 g_{22} + 4\overline{Y}e_0 e_1 g_{01} + 4\overline{Y}e_0 e_2 g_{02} + 2e_1 e_2 g_{12}\right\} \qquad (2.4)$$

Now using (2.4) in (1.1), we have

$$d_g = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 + e_3 - \Big[\bar{Y}^2 + e_0^2 + 2\bar{Y}e_0 + e_1 g_1 + e_2 g_2 + \frac{1}{2!}\left\{ e_1^2 g_{11} + e_2^2 g_{22} + 4\bar{Y}e_0 e_1 g_{01} + 4\bar{Y}e_0 e_2 g_{02} + 2e_1 e_2 g_{12} \right\}\Big]$$

$$d_g - \sigma_Y^2 = e_3 - 2\bar{Y}e_0 - e_1 g_1 - e_2 g_2 - e_0^2 - \frac{1}{2}\left(e_1^2 g_{11} + e_2^2 g_{22} + 4\bar{Y}e_0 e_1 g_{01} + 4\bar{Y}e_0 e_2 g_{02} + 2e_1 e_2 g_{12}\right) \quad (2.5)$$

Now taking expectation on both the sides of (2.5), the bias in d_g to the first degree of approximation is given by

Bias in $d_g = E(d_g) - \sigma_Y^2$

$$= E(e_3) - 2\bar{Y}E(e_0) - E(e_1)g_1 - E(e_2)g_2 - E(e_0^2) - \frac{1}{2}\left\{ E(e_1^2)g_{11} + E(e_2^2)g_{22} + 4\bar{Y}E(e_0 e_1)g_{01} + 4\bar{Y}E(e_0 e_2)g_{02} + 2E(e_1 e_2)g_{12} \right\}$$

using the values of the expectation given in (2.2) and (2.3), we have

Bias in $d_g = -\frac{\mu_{20}}{n} - \frac{1}{2n}\left(\mu_{02}g_{11} + \mu_{02}^2(\beta_2 - 1) + 4\bar{Y}g_{01}\mu_{11} + 4\bar{Y}g_{02}\mu_{12} + 2g_{12}\mu_{03}\right)$ *(2.6)*

Now squaring (2.5) on both the sides and then taking expectation, the mean square error to the first degree of approximation is given by

$$MSE(d_g) = E(d_g - \sigma_Y^2)^2 = E\left(e_3 - 2\bar{Y}e_0 - e_1 g_1 - e_2 g_2\right)^2$$

$$= E(e_3^2) + 4\bar{Y}^2 E(e_0^2) + g_1^2 E(e_1^2) + g_2^2 E(e_2^2) - 4\bar{Y}E(e_0 e_3) - 2g_1 E(e_1 e_3) - 2g_2 E(e_2 e_3) + 4\bar{Y}g_1 E(e_0 e_1) + 4\bar{Y}g_2 E(e_0 e_2) + 2g_1 g_2 E(e_1 e_2)$$

using values of the expectation given in (2.2) and (2.3), we have

$$MSE(d_g) = \frac{1}{n}\left(\mu_{40} + 4\bar{Y}\mu_{30} + 4\bar{Y}^2\mu_{20} - \mu_{20}^2\right) + 4\bar{Y}^2\frac{\mu_{20}}{n} + g_1^2\frac{\mu_{02}}{n} + g_2^2\frac{\mu_{02}^2}{n}(\beta_2 - 1) - 4\bar{Y}\frac{1}{n}\left(\mu_{30} + 2\bar{Y}\mu_{20}\right) - 2g_1\frac{1}{n}\left(\mu_{21} + 2\bar{Y}\mu_{11}\right) - 2g_2\frac{1}{n}\left(\mu_{22} + 2\bar{Y}\mu_{12} - \mu_{02}\mu_{20}\right) + 4\bar{Y}g_1\frac{\mu_{11}}{n} + 4\bar{Y}g_2\frac{\mu_{12}}{n} + 2g_1 g_2\frac{\mu_{03}}{n}$$

$$MSE(d_g) = MSE(s_y^2) + \frac{1}{n}\left[\mu_{20}g_1^2 - 2\mu_{21}g_1 + \mu_{02}^2(\beta_2 - 1)g_2^2 + 2(\mu_{02}\mu_{20} - \mu_{22})g_2 + 2\mu_{03}g_1 g_2\right] \quad (2.7)$$

For minimizing (2.7) in two unknowns g_1 and g_2, the normal equations after differentiating (2.7) partially with respect to g_1 and g_2 are

$$\mu_{02}g_1 + \mu_{03}g_2 - \mu_{21} = 0 \quad \text{and} \quad (2.8)$$

$$\mu_{03}g_1 + \mu_{02}^2(\beta_2 - 1)g_2 + (\mu_{02}\mu_{20} - \mu_{22}) = 0 \quad . \quad (2.9)$$

Solving (2.8) and (2.9) for g_1 and g_2, we get the minimizing optimum values to be

$$g_1^* = \frac{\mu_{21}\mu_{02}^2(\beta_2 - 1) + \mu_{03}(\mu_{02}\mu_{20} - \mu_{22})}{(\beta_2 - \beta_1 - 1)\mu_{02}^3} \quad \text{and} \quad (2.10)$$

$$g_2^* = -\frac{\mu_{21}\mu_{03} + \mu_{02}\left(\mu_{02}\mu_{20} - \mu_{22}\right)}{\left(\beta_2 - \beta_1 - 1\right)\mu_{02}^3} \tag{2.11}$$

$$\therefore \qquad \mu_{02}\, g_1^2 = \frac{1}{\left(\beta_2 - \beta_1 - 1\right)^2 \mu_{02}^5}\left\{\mu_{21}\mu_{02}{}^2\left(\beta_2 - 1\right) + \mu_{03}\left(\mu_{02}\mu_{20} - \mu_{22}\right)\right\}^2 \tag{2.12}$$

$$\mu_{02}{}^2\left(\beta_2 - 1\right) g_2^2 = \frac{\left(\beta_2 - 1\right)}{\left(\beta_2 - \beta_1 - 1\right)^2 \mu_{02}^4}\left\{\mu_{21}\mu_{03} + \mu_{02}\left(\mu_{02}\mu_{20} - \mu_{22}\right)\right\}^2 \tag{2.13}$$

and

$$2\mu_{03}\, g_1 g_2 = -\frac{2\mu_{03}}{\left(\beta_2 - \beta_1 - 1\right)^2 \mu_{02}^6}\left\{\mu_{21}\mu_{02}{}^2\left(\beta_2 - 1\right) + \mu_{03}\left(\mu_{02}\mu_{20} - \mu_{22}\right)\right\}\cdot \left\{\mu_{21}\mu_{03} + \mu_{02}\left(\mu_{02}\mu_{20} - \mu_{22}\right)\right\} \tag{2.14}$$

Now adding (2.12), (2.13) and (2.14), we get

$$\mu_{02}\, g_1^2 + \mu_{02}{}^2\left(\beta_2 - 1\right) g_2^2 + 2\mu_{03}\, g_1 g_2$$

$$= \frac{\mu_{21}{}^2}{\mu_{02}} + \frac{1}{\left(\beta_2 - \beta_1 - 1\right)\mu_{02}{}^2}\left[\left(\mu_{02}\mu_{20} - \mu_{22}\right) + \mu_{21}\left(\mu_{02}\right)^{1/2}\gamma_1\right]^2 \tag{2.15}$$

also

$$-2\mu_{21}\, g_1 = -\frac{2\mu_{21}}{\left(\beta_2 - \beta_1 - 1\right)\mu_{02}{}^3}\left\{\mu_{21}\mu_{02}{}^2\left(\beta_2 - 1\right) + \mu_{03}\left(\mu_{02}\mu_{20} - \mu_{22}\right)\right\} \tag{2.16}$$

$$2\left(\mu_{02}\mu_{20} - \mu_{22}\right) g_2 = -\frac{2\left(\mu_{02}\mu_{20} - \mu_{22}\right)}{\left(\beta_2 - \beta_1 - 1\right)\mu_{02}{}^3}\left\{\mu_{21}\mu_{03} + \mu_{02}\left(\mu_{02}\mu_{20} - \mu_{22}\right)\right\} \tag{2.17}$$

on adding (2.16) and (2.17), we get

$$-2\mu_{21}\, g_1 - 2\left(\mu_{02}\mu_{20} - \mu_{22}\right) g_2 = -2\,\frac{\mu_{21}{}^2}{\mu_{02}} - \frac{2}{\left(\beta_2 - \beta_1 - 1\right)\mu_{02}{}^2}\left[\left(\mu_{02}\mu_{20} - \mu_{22}\right) + \mu_{21}\left(\mu_{02}\right)^{1/2}\gamma_1\right]^2 \tag{2.18}$$

putting (2.15) and (2.18) in (2.7), we get

$$MSE\left(d_g\right)_{\min} = MSE\left(s_y^2\right) - \frac{\mu_{21}{}^2}{n\,\mu_{02}} - \frac{\mu_{21}{}^2}{n\left(\beta_2 - \beta_1 - 1\right)\mu_{02}{}^2}\left(\frac{\mu_{02}\mu_{20}}{\mu_{21}} - \frac{\mu_{22}}{\mu_{21}} + \left(\mu_{02}\right)^{1/2}\gamma_1\right)^2 \tag{2.19}$$

III. EFFICIENCY COMPARISON WITH THE AVAILABLE ESTIMATORS

For comparing the efficiency of the proposed generalized estimator, let us consider the following

(i) Usual Conventional unbiased Estimator of Population Variance in case of SRSWOR

$$\hat{d}_1 = s_y^2 = \frac{1}{n-1}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{y}_{\mathrm{i}} - \bar{y}\right)^2 \quad \text{with} \quad MSE\left(\hat{d}_1\right) = \frac{1}{n}\left(\mu_{40} - \mu_{20}{}^2\right) \tag{3.1}$$

from (3.1) and (2.19), it is clear that the proposed generalized class of estimator has mean square error lesser than the usual conventional unbiased estimator.

(ii) Estimator of Population Variance given by Peeyush Misra and R. Karan Singh

$$\hat{d}_2 = \hat{\theta} - \bar{y}\,.\,f\left(\bar{y}, \bar{x}\right) \qquad \text{and} \qquad \hat{d}_3 = \hat{\theta}\; f\left(u\right) - \bar{y}^2$$

$$\text{with } MSE\left(\hat{d}_2\right) = MSE\left(\hat{d}_3\right) = \frac{1}{n}\left(\mu_{40} - \mu_{20}{}^2\right) - \frac{\mu_{21}{}^2}{n\mu_{02}} \tag{3.2}$$

from (3.2) and (2.19), it is clear that the proposed generalized class of estimator has mean square error lesser than the mean square error of the estimator of population variance given by Peeyush Misra and R. Karan Singh.

IV. EMPIRICAL STUDY

For comparing efficiency of the proposed generalized class of estimator, let us consider the data given in, William G. Cochran (1977), Sampling Techniques, 3rd Edition, John Wiley and Sons, New York, dealing with Paralytic Polio Cases 'Placebo' (Y) group, Paralytic Polio Cases in not inoculated group (X), we have

$n = 34$

$\overline{Y} = 2.58$

$\overline{X} = 8370.6$

$\mu_{20} = 9.8894$

$\mu_{30} = 47.015235$

$\mu_{40} = 421.96088$

$\mu_{21} = 93.464705 \times 10^{3}$

$\mu_{11} = 19.34352945 \times 10^{3}$

$\mu_{02} = 7.1865882 \times 10^{7}$

$\mu_{03} = 1.4510955 \times 10^{12}$

$\mu_{04} = 4.5961952 \times 10^{16}$

$\mu_{12} = 3.443287 \times 10^{8}$

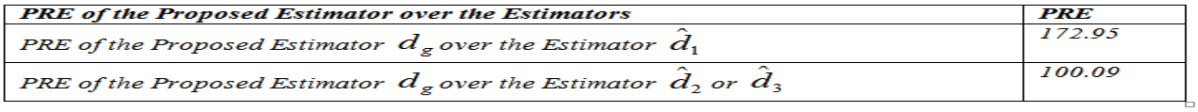
$\mu_{22} = 3.0156658 \times 10^{9}$.

We have $MSE\left(\hat{d}_1\right) = 9.534136695.$

$MSE\left(\hat{d}_2\right) = MSE\left(\hat{d}_3\right) = 5.958992179.$

$MSE\left(d_g\right) = 5.512540843.$

Table 4.1: PRE of the Proposed Estimator over the Estimators Described Above



PRE of the Proposed Estimator over the Estimators	*PRE*
PRE of the Proposed Estimator d_g *over the Estimator* $\hat{d}_1$	*172.95*
PRE of the Proposed Estimator d_g *over the Estimator* $\hat{d}_2$ *or* $\hat{d}_3$	*100.09*

V. CONCLUSION

The comparative study and empirical study of the proposed generalized sampling estimator of population variance establishes its superiority in the sense of having minimum mean square error over the usual conventional unbiased estimator of population variance in case of SRSWOR and the estimator of population variance given by Peeyush Misra and R. Karan Singh.

REFERENCES

[1] *Cochran, W.G. (1977): Sampling Techniques, 3rd edition, John Wiley and Sons, New York.*

[2] *Das, A. K. and Tripathi, T. P (1978): Use of auxiliary information in estimating the finite population variance, Sankhya, C, 40, 139-148.*

[3] *Liu, T. P (1974): A general unbiased estimator for the variance of a finite population, Sankhya, C, 36, 23-32.*

[4] *Murthy, M. N. (1967): Sampling Theory and Methods, 1st edition, Statistical Publishing Society, Calcutta (India).*

[5] *Peeyush Misra and R. Karan Singh (2015): Ratio-Product-Difference type estimators for finite population variance using auxiliary information, Accepted for publication in 'Progress in Mathematics'.*

[6] *R. P. Gupta (1983): Unbiased estimation of finite population variance using auxiliary information, Statistics and Probability Letters, 1, 121-124.*

[7] *Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling theory of surveys with applications, 3rd revised edition, IOWA State University Press, Ames, U.S.A.*

[8] *Srivastava, S.K. and Jhajj, H. S. (1980): A class of estimators using auxiliary information for estimating finite population variance, Sankhya, C, 42, 87-96.*

[9] *Singh, R. K., Zaidi, S. M. H. and Rizvi, S. A. H. (1995): Estimation of finite population variance using auxiliary information, journal of Statistical Studies, Volume 15, 17-28.*