Global Journal of Modeling and Intelligent Computing (GJMIC)

Vol. 1, No. 2, July, 2021 ISSN: 2767-1917

A GENERALIZED CLASS OF DIFFERENCE AND REGRESSION TYPE ESTIMATORS OF FINITE POPULATION VARIANCE USING AUXILIARY INFORMATION

Abhijeet Vishwakarma , University of Lucknow, India (abhijeetvishwakarma12jan@gmail.com)
S. A. H. Rizvi, Shia P.G. College, India (rizvi_athar@hotmail.com)

ABSTRACT

A generalized estimator representing a class of difference and regression type alternative estimators of finite population variance using information on known mean of an auxiliary variable is proposed, its bias and mean square error are found and a comparative study is made to establish the existence of some superior estimators in the proposed class in the sense of having lesser Mean Square Error. The optimum values minimizing the mean square error of the generalized class of difference type estimator depend upon some unknown parameters which give rise to the practical utility; hence , the alternative is to use the estimated optimum values depending on sample observations and putting the estimated optimum values in place of optimum values of difference type estimator, we get the generalized class of regression type estimator attaining the same minimum value of its mean square error to be equal to the minimum mean square error of the difference type estimator for its unknown optimum values. As an illustration, an empirical study is also considered to support the theoretical findings.

Keywords: Difference and Regression type estimators, Auxiliary information, Bias and Mean Square Error (MSE) , Optimum and Estimated Optimum Values, Efficiency.

1. INTRODUCTION

Auxiliary information is widely used in sampling theory at both the stages of selection as well as estimation. Various sampling schemes are designed at the selection stage by using auxiliary information while at the estimation stage it is used by formulating a variety of estimators of different population parameters in order to get increased efficiency. Various estimation procedures for the estimation of parameters such as ratio, mean, square of mean, total, variance etc, are studied and their properties are analyzed by several authors. To improve the estimation procedures over the existing ones in the sense of having lesser risk, prior information is utilized in some form or the other. In sampling theory, prior information in the form of some known parameters like mean, variance, coefficient of variation etc. of one or several auxiliary variables is extensively used in estimation procedures to develop ratio type, product type, difference type or regression type estimators.

In real life examples, difference and regression type estimators are widely used in areas like engineering sciences, biological sciences, medical sciences, geosciences, agriculture sciences, weather forecasting, sensex predictions etc. as in any system in which variable quantities change, it is of interest to examine the effects that some variable exert (or appear to exert) on others. There may in fact be a simple functional relationship between variables. In case, we wish to approximate to this functional relationship by some simple mathematical function, such as a polynomial, which contains the appropriate variable and which graduates to the true function over some limited ranges of the variables involved. In this way we may be able to learn more about the underlying true relationship and to appreciate the separate and joint effects produced by changes in certain important variables.

In many surveys, information on an auxiliary variable which is highly correlated with variable under study is readily available and can be used for improving sampling design. Stratified sampling and PPS scheme are two such examples in which information on auxiliary variable is used. In situation when data on auxiliary variable for individual sampling units are not available but only the aggregate value for all the units of auxiliary variable is available, the above two schemes cannot be used. Two such methods of estimation when the aggregated data on auxiliary variable can still be used at the time of estimation of the parameters under study provided the information on auxiliary variable for the sampled units can easily be obtained are known as Ratio method of estimation and Regression method of estimation.

Though a lot of work has been done on improving upon different type of estimators of population parameters with increased efficiency using auxiliary information, some of them are ratio or product type estimators which are widely used in practice for their simplicity and easy computability, in contrast difference or regression type estimators, being laborious to compute have not been used so extensively. Hence, still there remains enough work to be done in this direction of investigating new estimators with increased efficiency in the sense of having lesser mean square error and extending them to classes and analyzing their properties.

In this paper, the properties of the proposed generalized classes of estimators are studied. It has been shown theoretically as well as empirically that the proposed classes of estimators are more efficient than some existing estimators.

Let (Y_i, X_i) be the values on the variables (y, x) for the i^{th} $(i = 1,2, \ldots, N)$ unit of a finite population of size N where y and x are the study and auxiliary variables respectively. Let $(\bar{Y}, \bar{X})$ be the population means of (y, x) respectively and

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

Further, let a simple random sample of size n without replacement be drawn from the population with sample observations $y_1, y_2, \ldots, y_n$ on y and $x_1, x_2, \ldots, x_n$ on x. Also let $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ be the sample means of y values and x values respectively. For simplicity, we assume that N is large enough as compared to n so that the finite population correction terms may be ignored.

Background of the study

We know that finite population variance of the study variable y is

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \overline{Y})^2$$

$$= \theta - \overline{Y}^2 \quad (2)$$

Where , $\theta = \frac{1}{N} \sum_{i=1}^{N} {Y_i}^2$.

We can further write

$$\sigma_y^2 = \overline{Y}^2 \left(\frac{\theta}{\overline{Y}^2} - 1 \right) \quad (3)$$

where, $\left(\frac{\theta}{\overline{Y}^2} - 1 \right) > 0$.

Replacing θ and $\overline{Y}^2$ in (1.3) by their some consistent or unbiased estimators, we may get an alternative estimator of the population variance σ_y^2 . In particular, replacing θ by its unbiased estimator $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} y_i^2$, $\overline{Y}^2$ by its consistent estimator $\overline{y}^2$ and using the ratio, product or difference type methodology, Rizvi and Singh (2016) proposed the following two estimators of σ_y^2 :

$$s^2 = \left\{ \overline{y}^2 + k(\overline{x} - \overline{X}) \right\} \left(\frac{\hat{\theta}}{\overline{y}^2} - 1 \right) \quad (4)$$

and $$\hat{s}^2 = \left\{ \overline{y}^2 + \hat{k}(\overline{x} - \overline{X}) \right\} \left(\frac{\hat{\theta}}{\overline{y}^2} - 1 \right) \quad (5)$$

where k is the optimum value minimizing the mean square error of s^2 and $\hat{k}$ is the estimated optimum value of k based on sample observations.

The optimum value of k for which MSE (s^2) is minimized is given by

$$k_{opt} = -\frac{\overline{Y}^2 \mu_{21}}{\mu_{20} \mu_{02}} \quad (6)$$

and the minimum mean square error of s^2 to the first degree of approximation is

$$\mathrm{MSE}(s^2)_{min} = \mathrm{MSE}(s_y^2) - \frac{1}{n}\frac{\mu_{21}^2}{\mu_{02}} \tag{7}$$

Where, $\mathrm{MSE}(s_y^2) = \frac{\mu_{20}^2}{n}\{\beta_{2(y)} - 1\}$ is the mean square error of the convectional usual estimator $s_y^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2$ of the population variance σ_y^2 and $\beta_{2(y)} = \frac{\mu_{40}}{\mu_{20}^2}$.

The optimum value $k_{opt.} = -\frac{\bar{Y}^2\mu_{21}}{\mu_{20}\mu_{02}}$ given by (6) involving unknown parameters may not be known in practice; hence the alternative is to replace the parameters involved in the optimum value $k_{opt.}$ by their estimated values to get the estimated optimum as

$$\hat{k} = -\frac{\bar{y}^2\hat{\mu}_{21}}{\hat{\mu}_{20}\hat{\mu}_{02}} \tag{8}$$

Where $\hat{\mu}_{21} = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2\ (x_i - \bar{X})$,

$\hat{\mu}_{20} = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2$,

$\hat{\mu}_{02} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$,

and $\hat{\bar{Y}} = \bar{y}$ (9)

are consistent or unbiased estimators of respective parameters.

Using the estimated optimum value $\hat{k}$ given by (1.8) in (4), we obtain the regression type estimator given by (5) as

$$\hat{s}^2 = \{\bar{y}^2 + \hat{k}(\bar{x} - \bar{X})\}\left(\frac{\hat{\theta}}{\bar{y}^2} - 1\right)$$

and the mean square error of $\hat{s}^2$ up to first degree of approximation comes out to be

$$MSE(\hat{s}^2) = MSE(s_y^2) - \frac{1}{n}\frac{\mu_{21}^2}{\mu_{02}} \tag{10}$$

which is the same expression of $\mathrm{MSE}(s^2)$.

The Proposed Class of Estimators

A class of estimators to which s^2 and $\hat{s}^2$ belong to, is defined as

$$\hat{\sigma}_g^2 = g(\bar{y}, \bar{x}, \hat{\theta}) \tag{11}$$

where $g(\bar{y}, \bar{x}, \hat{\theta})$ satisfying the validity conditions of Taylor's series expansion and having first, second and third order partial derivatives bounded, is a bounded function of $(\bar{y}, \bar{x}, \hat{\theta})$ such that at the point $P = (\bar{Y}, \bar{X}, \theta)$,

(□) $g(\bar{Y}, \bar{X}, \theta) = \sigma_y^2$; (12)

(□) first order partial derivative of $g(\bar{y}, \bar{x}, \hat{\theta})$ with respect to $\bar{y}$ at point P is

$$g_0 = \left.\frac{\partial\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{y}}\right]_P = -2\bar{Y}. \tag{13}$$

(iii) for first order partial derivatives $g_1 = \left.\frac{\partial\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{x}}\right]_P = 1$ and $g_2 = \left.\frac{\partial\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \hat{\theta}}\right]_P = 1$ (14)

(iv) for second order partial derivatives $g_{00} = \left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{y}^2}\right]_P = -2$ (15)

$$g_{01} = \left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{y}\,\partial \bar{x}}\right]_P = 0 \quad ,$$

$$g_{02} = \left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{y}\,\partial \hat{\theta}}\right]_P = 0$$

$$g_{11} = \left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{x}^2}\right]_P = 0$$

$$g_{12} = \left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \bar{x}\,\partial \hat{\theta}}\right]_P = 0$$

and $$g_{22} = \left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta})}{\partial\, \hat{\theta}^2}\right]_P = 0 \tag{16}$$

Some more special cases of the generalized estimate σ_g^2 are:

(i) $s_1^2 = \overline{y}\left\{\overline{y} + k\left(\overline{x} - \overline{X}\right)\right\}\left(\frac{\hat{\theta}}{\overline{y}^2} - 1\right)$

(ii) $s_2^2 = \overline{y}\left\{\overline{y} + \left(\overline{x}^k - \overline{X}^k\right)\right\}\left(\frac{\hat{\theta}}{\overline{y}^2} - 1\right)$

Where k is the characterizing scalar to be chosen suitably. It may be checked that the regularity conditions mentioned for the generalized estimator $\hat{\sigma}_g^2$ are satisfied for above two estimators.

2. LITERATURE REVIEW

For the estimation of some parameters of interest, one may see the estimation procedures and their properties in Murthy (1967), Liu (1974), Cochran (1977), Sukhatme et.al. (1984), Das et al. (1978, 1980), Ray et al. (1981), Srivastava et al.(1981), Singh et al. (1983), Bisht et al. (1990). Rizvi et.al.(1996), Rizvi et.al.(2009), Singh et.al. (2012), Rizvi et.al. (2016) considered several estimators of population variance of the variable under study, when the information in the form of known mean or variance of the auxiliary variable is utilized. Some of the estimators along with their bias and mean square error are given below:

S. No.	Estimator	Bias	MSE
1.	$d_g = \hat{\theta} f\left(\frac{\overline{x}_s}{\overline{X}}\right) - \bar{y}_s^2 = \theta f(u) - \bar{y}_s^2$ Singh et al. (1995)	$\frac{\mu_{20}}{n} + \frac{1}{n\overline{X}}(\mu_{21} - 2\overline{Y}\mu_{11})f'(1) + \frac{\mu_{02}}{2n\overline{X}^2}(\mu_{02} + \overline{Y}^2)f''(1)$	$\frac{1}{n}\left[\mu_{20}^2\{\beta_{2(y)} - 1\} - \frac{\mu_{21}^2}{\mu_{02}}\right]$
2.	$d_1 = \hat{\theta}\left\{1 + k_1\left(\frac{\bar{x}_{1s}}{\overline{X}_1} - 1\right) + k_2\left(\frac{\bar{x}_{2s}}{\overline{X}_2} - 1\right)\right\} - \bar{y}_s^2$ Singh et al. (1996)	$\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{N}{N-1}\right)[-\mu_{200} + \frac{(\mu_{200} + \overline{Y}^2)}{2}\left\{\frac{2k_1k_2}{\overline{X}_1\overline{X}_2}\mu_{011} + \frac{k_1(k_1-1)}{\overline{X}_1^2}\mu_{020} + \frac{k_2(k_2-1)}{\overline{X}_2^2}\mu_{002}\right\} + \frac{k_1(\mu_{210} + 2\overline{Y}\mu_{110})}{\overline{X}_1} + \frac{k_2(\mu_{201} + 2\overline{Y}\mu_{101})}{\overline{X}_2}$	$\left(\frac{1}{n} - \frac{1}{N}\right)[\mu_{200}^2(\beta_2 - 1) - \frac{(\mu_{201}^2\mu_{020} + \mu_{210}^2\mu_{002} - 2\mu_{011}\mu_{201}\mu_{210})}{(\mu_{002}\mu_{020} - \mu_{011}^2)}]$. It may be mentioned here that $\left(\frac{1}{n} - \frac{1}{N}\right)(\mu_{201}^2\mu_{020} + \mu_{210}^2\mu_{002} - 2\mu_{011}\mu_{201}\mu_{210})$ $= E(\mu_{201}^2e_1^2 + \mu_{210}^2e_2^2 - 2e_1e_2\mu_{201}\mu_{210})$ $= E(\mu_{201}e_1 - \mu_{210}e_2)^2 > 0$
3.	$s_k^2 = s_y^2 + k\left[\bar{y}\frac{C_y}{s_y} - 1\right]$ Chandel et al. (1999)	$k\left[\frac{(\beta_2 - 1)}{8n} - \frac{\gamma_1\sigma_y}{2n\overline{Y}}\right]$	$MSE(s_y^2) - \frac{\sigma_y^4\left[\gamma_1C_y - \frac{1}{2}(\beta_2 - 1)\right]^2}{n\left[C_y^2 + \frac{1}{4}(\beta_2 - 1) - \gamma_1C_y\right]}$
4.	$\hat{\sigma}_a^2 = \sum_{h=1}^{L}\left(\frac{W_h^2}{n_h}\right)s_{yh}^2a_h(u_h)$ Singh H.P. et al. (2008)	$\sum_{h=1}^{L}W_h^2\left(\frac{\sigma_{yh}^2}{2n_h^2}\right)[2\lambda_{21h}C_{xh}a_{h1}(1) + C_{xh}^2a_{h22}(1)]$	$\sum_{h=1}^{L}\frac{(W_h\sigma_{yh})^4}{n_h^3}[(\lambda_{40h} - 1) + C_{xh}^2a_{h1}^2(1) + 2\lambda_{21h}C_{xh}a_{h1}(1)]$

5.	$\hat{R}_{gd} = g(\hat{R}, \bar{x}_2, \bar{x}_2')$	$\left(\frac{f_1}{n}\right) R(C_1^2 - \rho_{01} C_0 C_1) + \frac{1}{2}\left[\left(\frac{f_1}{n}\right) \bar{X}_2^2 C_2^2 g_{11} + \left(\frac{f_1'}{n'}\right) \bar{X}_2^2 C_2^2 (g_{22} + 2g_{12}) \; 2\left(\frac{1}{n} - \frac{1}{n'}\right) R\bar{X}_2(\rho_{02} C_0 C_2 - \rho_{12} C_1 C_2) g_{01}\right]$	$MSE(\hat{R}) - \left(\frac{1}{n} - \frac{1}{n'}\right) R^2 C_2^2 C^2$
	$\hat{P}_{gd} = h(\hat{P}, \bar{x}_2, \bar{x}_2')$ Singh et al. (2012)	$\left(\frac{f_1}{n}\right) P(C_1^2 - \rho_{01} C_0 C_1) + \frac{1}{2}\left[\left(\frac{f_1}{n}\right) \bar{X}_2^2 C_2^2 h_{11} + + \left(\frac{f_1'}{n'}\right) \bar{X}_2^2 C_2^2 (h_{22} + 2h_{12}) + 2\left(\frac{1}{n} - \frac{1}{n'}\right) P\bar{X}_2(\rho_{02} C_0 C_2 - \rho_{12} C_1 C_2) h_{01}\right]$	$MSE(\hat{P}) - \left(\frac{1}{n} - \frac{1}{n'}\right) P^2 C_2^2 C^{*2}$
6.	$s^2 = \{\bar{y}^2 + k(\bar{x} - \bar{X})\}\left(\frac{\hat{\theta}}{\bar{y}^2} - 1\right)$ Rizvi et al. (2016)	$-\frac{\mu_{20}}{n} + \frac{k}{n\bar{Y}^2}\left(\mu_{21} - \frac{2\mu_{20}}{\bar{Y}}\right),$	$\frac{1}{n}\left[\mu_{20}^2\{\beta_{2(y)} - 1\} - \frac{\mu_{21}^2}{\mu_{02}}\right]$

3. BIAS AND MEAN SQUARE ERROR (MSE) OF $\hat{\sigma}_g^2$

For simplicity, it is assumed that the population size N is large enough as compared to sample size n so that finite population correction terms may be ignored.

Let $\bar{y} = \bar{Y} + e_0$, $\bar{x} = \bar{X} + e_1$, and $\hat{\theta} = \theta + e_2$ so that

$E(e_0) = E(e_1) = E(e_2) = 0$ (17)

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

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

$E(e_2^2) = \frac{1}{n}(\mu_{40} + 4\bar{Y}\mu_{30} + 4\bar{Y}^2\mu_{20} - \mu_{20}^2)$

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

$E(e_0 e_2) = \frac{1}{n}(\mu_{30} + 2\bar{Y}\mu_{20})$

$E(e_1 e_2) = \frac{1}{n}(\mu_{21} + 2\bar{Y}\mu_{11})$ (18)

Expanding $g(\bar{y}, \bar{x}, \hat{\theta})$ about the point $P = (\bar{Y}, \bar{X}, \theta)$ in third order Taylor's, we have

$$\hat{\sigma}_g^2 = g(\bar{Y}, \bar{X}, \theta) + (\bar{y} - \bar{Y})g_0 + (\bar{x} - \bar{X})g_1 + (\hat{\theta} - \theta)g_2 + \frac{1}{2!}\Big\{(\bar{y} - \bar{Y})^2 g_{00} + (\bar{x} - \bar{X})^2 g_{11} + (\hat{\theta} - \theta)^2 g_{22}$$
$$+2(\bar{y} - \bar{Y})(\bar{x} - \bar{X})g_{01} + 2(\bar{y} - \bar{Y})(\hat{\theta} - \theta)g_{02} + 2(\bar{x} - \bar{X})(\hat{\theta} - \theta)g_{12}\Big\}$$
$$+\frac{1}{3!}\left\{(\bar{y} - \bar{Y})\frac{\partial}{\partial \bar{y}} + (\bar{x} - \bar{X})\frac{\partial}{\partial \bar{x}} + (\hat{\theta} - \theta)\frac{\partial}{\partial \hat{\theta}}\right\}^3 g(\bar{y}^*, \bar{x}^*, \hat{\theta}^*) \quad (19)$$

where first and second order partial derivatives are already defined in (13) to (16) and

$\bar{y}^* = \bar{Y} + h(\bar{y} - \bar{Y})$, $\bar{x}^* = \bar{X} + h(\bar{x} - \bar{X})$, $\hat{\theta}^* = \theta + h(\hat{\theta} - \theta)$ for $0 < h < 1$.

Employing regularity conditions from (12) to (16) in (19), we have

$$\hat{\sigma}_g^2 = \sigma_y^2 - 2\,\overline{Y}(\overline{y}-\overline{Y}) + (\overline{x}-\overline{X})g_1 + (\hat{\theta}-\theta)g_2 + \frac{1}{2!}\{-2(\overline{y}-\overline{Y})^2 + (\overline{x}-\overline{X})^2 g_{11} + 2(\overline{y}-\overline{Y})(\overline{x}-\overline{X})g_{01} + 2(\overline{x}-\overline{X})(\hat{\theta}-\theta)g_{12}\} + \frac{1}{3!}\left\{(\overline{y}-\overline{Y})\frac{\partial}{\partial\,\overline{y}} + (\overline{x}-\overline{X})\frac{\partial}{\partial\,\overline{x}} + (\hat{\theta}-\theta)\frac{\partial}{\partial\,\hat{\theta}}\right\}^3 g(\overline{y}^*,\overline{x}^*,\hat{\theta}^*)$$

or,

$$\hat{\sigma}_g^2 - \sigma_y^2 = -2\overline{Y}e_0 + e_1 g_1 + e_2 + \frac{1}{2!}\{-2e_0^2 + e_1^2 g_{11} + 2e_0 e_1 g_{01} + 2e_1 e_2 g_{12}\} + \frac{1}{3!}\left\{(\overline{y}-\overline{Y})\frac{\partial}{\partial\,\overline{y}} + (\overline{x}-\overline{X})\frac{\partial}{\partial\,\overline{x}} + (\hat{\theta}-\theta)\frac{\partial}{\partial\,\hat{\theta}}\right\}^3 g(\overline{y}^*,\overline{x}^*,\hat{\theta}^*) \quad . \qquad (20)$$

Taking expectation on both sides of (20), to the first degree of approximation and retaining terms up to order $O\left(\frac{1}{n}\right)$, we have

$$E(\hat{\sigma}_g^2 - \sigma_y^2) = E\left(-2\overline{Y}e_0 + e_1 g_1 + e_2 - e_0^2 + \frac{1}{2}e_1^2 g_{11} + e_0 e_1 g_{01} + e_1 e_2 g_{12}\right)$$

or $$Bias(\hat{\sigma}_g^2) = \frac{1}{n}\left\{-\mu_{20} + \frac{\mu_{02}}{2}g_{11} + \mu_{11}g_{01} + (\mu_{21} + 2\overline{Y}\mu_{11})g_{12}\right\} \qquad (21)$$

Squaring both sides of (20), taking expectation to the first degree of approximation and retaining terms up to order $O\left(\frac{1}{n}\right)$, we have

$$E(\hat{\sigma}_g^2 - \sigma_y^2)^2 = \mathrm{E}(4\overline{Y}^2 e_0^2 + e_1^2 g_1^2 + e_2^2 - 4\overline{Y}e_0 e_1 g_1 - 4\overline{Y}e_0 e_2 + e_1 e_2 g_1)$$

or $$MSE(\hat{\sigma}_g^2) = \frac{\mu_{20}^2}{n}\{\beta_{2(y)} - 1\} + \frac{1}{n}(\mu_{02}g_1^2 + 2\mu_{21}g_1)$$

where $\beta_{2(y)} = \frac{\mu_{40}}{\mu_{20}^2}$

or $$MSE(\hat{\sigma}_g^2) = MSE(s_y^2) + \frac{1}{n}(\mu_{02}g_1^2 + 2\mu_{21}g_1) \qquad (22)$$

where, $MSE(s_y^2) = \frac{\mu_{20}^2}{n}\{\beta_{2(y)} - 1\}$ is the mean square error of the conventional usual estimator $s_y^2 = \frac{1}{(n-1)}(y_i - \overline{y})^2$ of the population variance σ_y^2.

4. OPTIMUM AND ESTIMATED OPTIMUM VALUE

From (22), we can see that the value of g_1 for which $MSE\ (\hat{\sigma}_g^2)$ is minimized is given by,

$$g_{1*} = -\frac{\mu_{21}}{\mu_{02}} \qquad (23)$$

and the minimum mean square error is

$$MSE(\hat{\sigma}_g^2)_{min.} = MSE(s_y^2) - \frac{1}{n}\left(\frac{\mu_{21}^2}{\mu_{02}}\right) \qquad (24)$$

Practically, the optimum value g_{1*} in (23) may not be available always, hence the alternative is to replace the parameters involved therein be their unbiased or consistent estimators and thus get the estimated optimum value.

Defining $m_{rs} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^r (x_i - \overline{X})^s$, replacing μ_{21} and μ_{02} by their estimates

$\hat{\mu}_{21} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2(x_i - \overline{X})$ and

$\hat{\mu}_{20} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$, we get

$$\hat{g}_{1*} = -\frac{\hat{\mu}_{21}}{\hat{\mu}_{02}} = \hat{G} \qquad (25)$$

The generalized estimator $\hat{\sigma}_g^2$ attains the minimum mean square error in (24) if the conditions from (12) to (16) and (23) are satisfied for the estimator $\hat{\sigma}_g^2$.

This means that the function $\hat{\sigma}_g^2 = g(\overline{y},\overline{x},\hat{\theta})$ as an estimator of σ_y^2 should not involve only $(\overline{y},\overline{x},\hat{\theta})$ but also g_{1*} for the condition (23) to be satisfied. Thus we get the resulting estimator as a function $g(\overline{y},\overline{x},\hat{\theta},g_{1*})$ satisfying the condition (12) to (16) along with condition (23) to attain the minimum mean square error in (24). Replacing unknown g_{1*} in $g(\overline{y},\overline{x},\hat{\theta},g_{1*})$, we get the estimator as a function $\hat{\sigma}_{ge}^2 = g^*(\overline{y},\overline{x},\hat{\theta},\hat{G})$ depending upon estimated optimum value. Let $(\hat{G}-G)=e_3$, now expanding $g^*(\bar{y},\bar{x},\hat{\theta},\hat{G})$ about the point $P^*=(\bar{Y},\bar{X},\theta,G)$ in Taylor's series, we have

$$\hat{\sigma}_{ge}^2=g^*(\bar{y},\bar{x},\hat{\theta},\hat{G})=g^*(P^*)+(\bar{y}-\bar{Y})g_0+(\bar{x}-\bar{X})g_1+(\theta-\hat{\theta})g_2+(\hat{G}-G)g_3+\frac{1}{2!}\{(\bar{y}-\bar{Y})^2g_{00}$$
$$+(\bar{x}-\bar{X})^2g_{11}+(\hat{\theta}-\theta)^2g_{22}+(\hat{G}-G)^2g_{33}+2(\bar{y}-\bar{Y})(\bar{x}-\bar{X})g_{01}+2(\bar{y}-\bar{Y})(\hat{\theta}-\theta)g_{02}$$
$$+2(\bar{y}-\bar{Y})(\hat{G}-G)g_{03}+2(\bar{x}-\bar{X})(\hat{\theta}-\theta)g_{12}+2(\bar{x}-\bar{X})(\hat{G}-G)g_{13}+2(\hat{\theta}-\theta)(\hat{G}-G)g_{23}\}$$
$$+\frac{1}{3!}\left\{(\bar{y}-\bar{Y})\frac{\partial}{\partial\bar{y}}+(\bar{x}-\bar{X})\frac{\partial}{\partial\bar{x}}+(\hat{\theta}-\theta)\frac{\partial}{\partial\hat{\theta}}+(\hat{G}-G)\frac{\partial}{\partial\hat{G}}\right\}^3g^*(\bar{y}^*,\bar{x}^*,\hat{\theta}^*,\hat{G}^*) \quad (26)$$

Where $g^*(P^*)=\sigma_y^2$, $g_3=\left.\frac{\partial\, g(\bar{y},\bar{x},\hat{\theta},\hat{G})}{\partial\,\hat{G}}\right]_{P^*}=0$, $g_{13}=\left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta},\hat{G})}{\partial\bar{x}\partial\,\hat{G}}\right]_{P^*}$,

$$g_{23}=\left.\frac{\partial^2\, g(\bar{y},\bar{x},\hat{\theta},\hat{G})}{\partial\hat{\theta}\partial\,\hat{G}}\right]_{P^*},\ g_{33}=\left.\frac{\partial^2 g(\bar{y},\bar{x},\hat{\theta},\hat{G})}{\partial\,\hat{G}^2}\right]_{P^*} \quad (27)$$

or $\hat{\sigma}_{ge}^2-\sigma_y^2=-2\bar{Y}(\bar{y}-\bar{Y})+(\bar{x}-\bar{X})g_1+(\hat{\theta}-\theta)g_2+(\hat{G}-G)g_3+\frac{1}{2!}\{-2(\bar{y}-\bar{Y})^2+(\bar{x}-\bar{X})^2g_{11}$

$$+(\hat{\theta}-\theta)^2g_{22}+(\hat{G}-G)^2g_{33}+2(\bar{y}-\bar{Y})(\bar{x}-\bar{X})g_{01}+2(\bar{y}-\bar{Y})(\hat{\theta}-\theta)g_{02}+2(\bar{y}-\bar{Y})(\hat{G}-G)g_{03}$$
$$+2(\bar{x}-\bar{X})(\hat{\theta}-\theta)g_{12}+2(\bar{x}-\bar{X})(\hat{G}-G)g_{13}+2(\hat{\theta}-\theta)(\hat{G}-G)g_{23}\}+\cdots \quad (28)$$

Squaring both sides of (28) and taking expectation, we see that the mean square error $E(\hat{\sigma}_{ge}^2-\sigma_y^2)^2=MSE(\hat{\sigma}_{ge}^2)$ to the first degree of approximation becomes equal to $MSE(\hat{\sigma}_g^2)_{min.}$ given by (24) if $g_3=\left.\frac{\partial\, g(\bar{y},\bar{x},\hat{\theta},\hat{G})}{\partial\,\hat{G}}\right]_{P^*}=0$ and thus the estimator taken as a function $\hat{\sigma}_{ge}^2=g^*(\bar{y},\bar{x},\hat{\theta},\hat{G})$ depending upon estimated optimum values attains the same minimum mean square error given by (24).

5. NUMERICAL ILLUSTRATION

Considering the data given in Cochran (1997) dealing with paralytic polio cases 'Placebo'(y) group ,paralytic polio cases in 'not inoculated'(x) group, computations of required values of μ_{rs} have been done and comparison among different estimators of finite population variance is made for a simple random sample.

For the data considered here, we have $\bar{Y}=2.58$, $\bar{X}=8370.6$, $\mu_{20}=9.8894$, $\mu_{02}=7.1865882*10^7$, $\mu_{40}=421.96088$, $\mu_{21}=93.464705*10^3$.

Comparison with the conventional estimator

For sample size $n=15$, using the above values, we have $MSE(s_y^2)=21.61034$ and

$$MSE(\hat{\sigma}_g^2)=MSE(\hat{\sigma}_{ge}^2)=13.506654 \quad (29)$$

whereas the percent relative efficiency (PRE) of the proposed estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ over the conventional estimator s_y^2 comes out to be

$$PRE(\hat{\sigma}_g^2)=PRE(\hat{\sigma}_{ge}^2)=159.99753 \quad (30)$$

which shows that the proposed estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ are more efficient with high percent relative efficiency over the usual conventional estimator s_y^2 of the population variance σ_y^2.

Comparison with other estimator

Renu Chandel (1999) proposed the following estimator;

$$s_k^2=s_y^2+k\left[\bar{y}\frac{C_y}{s_y}-1\right]. \quad (31)$$

Considering the same data given in Cochran (1997) dealing with paralytic polio cases 'Placebo'(y) group, paralytic polio cases in 'not inoculated'(x) group, the percent relative efficiency (PRE) of the estimator s_k^2 given by (31) over the conventional estimator s_y^2 as calculated by Chandel comes out to be

$$PRE(s_k^2)=130.67269\ . \quad (32)$$

which shows that the proposed estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ are more efficient with high percent relative efficiency over the estimator s_k^2 of the population variance σ_y^2 given in (31).

6. CONCLUDING REMARKS

(a) From (24), any estimator belonging to the class $\hat{\sigma}_g^2$ of estimators cannot have its mean square error (to the first degree of approximation) less than

$$\frac{1}{n}\left[\mu_{20}^2\{\beta_{2(y)}-1\}-\frac{\mu_{21}^2}{\mu_{02}}\right] \tag{33}$$

(b) The optimum estimator $\hat{\sigma}_g^2$ in the sense of having minimum mean square error involves the function $g(\bar{y},\bar{x},\hat{\theta})$ such that

$$g(\bar{Y},\bar{X},\theta)=\sigma_y^2 \text{ and } g_{1*}=-\frac{\mu_{21}}{\mu_{02}} \tag{34}$$

Thus the estimators $s^2=\{\bar{y}^2+k(\bar{x}-\bar{X})\}\left(\frac{\hat{\theta}}{\bar{y}^2}-1\right)$, $s_1^2=\bar{y}\{\bar{y}+k(\bar{x}-\bar{X})\}\left(\frac{\hat{\theta}}{\bar{y}^2}-1\right)$ and

$$s_2^2=\bar{y}\left\{\bar{y}+(\bar{x}^k-\bar{X}^k)\right\}\left(\frac{\hat{\theta}}{\bar{y}^2}-1\right) \tag{35}$$

belonging to the class $\hat{\sigma}_g^2$ of estimators and having the value of g_{1*} to be equal to

$$k\text{ }, k \text{ and } k \tag{36}$$

respectively will attain the minimum mean square error given by (33) for the optimum values of k equal to $\frac{\mu_{20}}{\bar{Y}^2}k$, $\frac{\mu_{20}}{\bar{Y}^2}k$ and $\frac{\mu_{20}}{\bar{Y}^2}k$ respectively.

But in practice, $\frac{\mu_{20}}{\bar{Y}^2}k$ may be rarely known, the alternative is to replace $\frac{\mu_{20}}{\bar{Y}^2}\ k$ be their estimates from sample values. The estimators depending upon estimated optimum values are

$$\hat{s}^2=\{\bar{y}^2+\hat{k}(\bar{x}-\bar{X})\}\left(\frac{\hat{\theta}}{\bar{y}^2}-1\right)$$

$$\hat{s}_1^2=\bar{y}\{\bar{y}+\hat{k}(\bar{x}-\bar{X})\}\left(\frac{\hat{\theta}}{\bar{y}^2}-1\right) \text{ and } \hat{s}_2^2=\bar{y}\left\{\bar{y}+(\bar{x}^{\hat{k}}-\bar{X}^{\hat{k}})\right\}\left(\frac{\hat{\theta}}{\bar{y}^2}-1\right) \tag{37}$$

which may belong to class $\hat{\sigma}_{ge}^2$ and satisfy the conditions in (27) , and also attain the minimum mean square error given by (33) to the first degree of approximation.

(c) The conventional estimator $s_y^2=\sum_{i=1}^n\frac{1}{(n-1)}(y_i-\bar{y})^2$ of the population variance σ_y^2 has its mean square error

$$MSE(s_y^2)=\frac{1}{n}\left[\mu_{20}^2\{\beta_{2(y)}-1\}\right] \tag{38}$$

Further , the proposed estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ have their mean square error

$$\frac{1}{n}\left[\mu_{20}^2\{\beta_{2(y)}-1\}-\frac{\mu_{21}^2}{\mu_{02}}\right] \tag{39}$$

From (38) and (39), it is clear that $MSE(\hat{\sigma}_g^2)$ and $MSE(\hat{\sigma}_{ge}^2)$ is less than $MSE(s_y^2)$, showing that the proposed estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ are more efficient than the conventional estimator s_y^2.

(d) An empirical study in support of theoretical findings as illustration shows that the $PRE(\hat{\sigma}_g^2)=PRE(\hat{\sigma}_{ge}^2)=159.99753$ given by (30) indicates that proposed generalized classes of estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ are more efficient with high percent relative efficiency over the usual estimator s_y^2 of the population variance σ_y^2.

(e) A comparative study shows that the $PRE(\hat{\sigma}_g^2)=PRE(\hat{\sigma}_{ge}^2)=159.99753$ given by (30) of the proposed generalized classes of estimators $\hat{\sigma}_g^2$ and $\hat{\sigma}_{ge}^2$ is greater than the $PRE(s_k^2)=130.67269$ given by (32) of the estimator of population variance proposed by Chandel (1999) which implies that the proposed generalized classes of estimators are better in the sense of having lesser mean square error.

REFERENCES

Arcos Cebnan, A. and Rueda Gracia, M. (1997). Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator. *Metrika*. 45,171-178.

Chandel, Renu (1999, December). Estimation of Population Variance Using Coefficient of Variation, Ch. VI of the Ph.D. thesis entitled Some Estimation Procedures in Sampling Theory and their Applications, awarded by University of Lucknow (pp. 110-128), Lucknow.

Cochran, W.G. (1977). Sampling Techniques (3rd Edition), John Wiley & Sons, Inc., New York.

Cramer, H. (1946). Mathematical Methods of Statistics, Princeton, Princeton University Press.

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

Rizvi, S.A.H. and Singh, R. Karan (2016). An alternative estimator of finite population variance using auxiliary information. Chapter 21, *Recent Advances in Applied Statistics and its Applications*. p.p. 173-176 Lambert Academic Publishing.

Singh H.P., S. K. Pal, and R. S. Solanki (2014) . A new procedure for estimation of finite population variance using auxiliary information, *Journal of Reliability and Statistical Studies*, vol. 7, no. 2, pp. 149–160.

Singh H.P. and Vishwakarma G.K. (2008). Some families of estimators of variance of stratified random sample mean using auxiliary information. *Journal of Statistical Theory and Practice*, vol. 2 , No. 1, pp. 21-43.

Singh, R. K., Zaidi, S. M. H. and Rizvi, S. A. H. (1995). Estimation of finite population variance using auxiliary information, *Jour. Stat. Stud*., 15, 17-28.

Singh, R. K., Zaidi, S. M. H. and Rizvi, S. A. H. (1996). Some estimators of finite population variance using information on two auxiliary variables, *Microelectron and Reliab.*, 36(5), 667-670.

Singh, R. Karan, and Singh, G. (1984). A class of estimators with estimated optimum values in sample surveys, Stat. and Prob. Lectt., 2, 319-321.

Singh, R. Karan, Rizvi, S. A. H. and Rizvi, S. A. H. (2012). Generalized Double Sampling Estimators of Ratio (Product) of Parameters. *International Journal of Statistics and Analysis*, Vol. 2 , No. 4 p.p. 391-402 Research India Publications, New Delhi, India.

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

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok, C. (1984). Sampling theory of Surveys with Application. 3rd Ed., Ames, Iowa (USA) and Indian Society of Agricultural Statistics.