Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 9, Issue 1 (2016): 93-100

ON IMPROVED DOUBLE SAMPLING ESTIMATORS OF POPULATION VARIANCE USING AUXILIARY INFORMATION

Peeyush Misra
Department of Statistics, D.A.V. (P.G.) College, Dehradun, India
E Mail: dr.pmisra.dav@gmail.com

Received December 09, 2015
Modified May 12, 2016
Accepted May 28, 2016

Abstract

A double sampling estimator and its generalized case representing a class of estimators using auxiliary information in the form of mean and variance both is proposed for the estimation of population variance. Bias and Mean Square Error are found and the properties of the estimators are studied. A comparative study with the estimators available in the literature is also carried out and it is shown that the proposed estimator and its generalized estimator are more efficient.

Key Words: Auxiliary Variable, Two-Phase Sampling, Taylor's Series Expansion, Bias, Mean Square Error And Efficiency.

1. Introduction

In sampling theory, it is well known that the auxiliary information is used to improve the precision of the estimators of the population parameters and if parameters of the auxiliary variables are not known in advance then double or two phase sampling technique is used. In double sampling or two-phase sampling technique, we first take a preliminary large sample of size n' (called first phase sample) from a population of size N and then a sub-sample of size n (called second phase sample) is drawn from the first phase sample of size n' by simple random sampling without replacement scheme at both the phases. At first phase sample of size n', only the auxiliary variable X be observed but at the second phase sample of size n, the study variable Y and the auxiliary variable X both are observed.

Let us denote by

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

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

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

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

$\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i{}'$ = sample mean of the first phase n' sample values on auxiliary character X

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ = sample mean of y based on second phase sample of size n

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = sample mean of auxiliary variable X based on second phase sample of size n

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ = sample variance of auxiliary variable X based on second phase sample of size n

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ = sample variance of study variable Y based on second phase sample of size n

$s_x'^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}\left(x_i' - \bar{x}'\right)^2$ = sample variance of auxiliary variable X based on second phase sample of size n

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

For estimating finite population variance, a double sampling estimator and a generalized estimator using auxiliary information in the form of mean and variance are proposed as

$$d = \hat{\theta} - \bar{y}^2\left\{1 + \frac{k_1\left(\bar{x} - \bar{x}'\right)}{\bar{x}'}\right\}\left\{1 + \frac{k_2\left(s_x^2 - s_x'^2\right)}{s_x'^2}\right\} \tag{1.1}$$

and $$d_g = \hat{\theta} - \bar{y}^2 f\left(u, v\right) \tag{1.2}$$

where $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$, $u = \frac{\bar{x}}{\bar{x}'}$ and $v = \frac{s_x^2}{s_x'^2}$.

In order to obtain bias and mean square error of the proposed estimators, let us denote by

$$\bar{y} = \overline{Y} + e_0 \qquad \bar{x} = \overline{X} + e_1$$

$$\bar{x}' = \overline{X} + e_1' \qquad s_x^2 = S_X^2 + e_2$$

$$s_x'^2 = S_X^2 + e_2' \qquad \hat{\theta} = \theta + e_3$$

where $$\theta = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \tag{1.3}$$

with $$E(e_0) = E(e_1) = E(e_1') = E(e_2) = E(e_2') = E(e_3) = 0 \tag{1.4}$$

$$E\left(e_0^2\right) = \frac{\mu_{20}}{n} \qquad E\left(e_1^2\right) = \frac{\mu_{02}}{n}$$

$$E\left(e_1'^2\right) = \frac{\mu_{02}}{n'} \qquad E\left(e_2^2\right) = \frac{\mu_{02}^{\;2}}{n}(\beta_2 - 1)$$

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

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

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

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

$$E\left(e_1 e_2\right) = \frac{\mu_{03}}{n} \qquad E\left(e_0 e_3\right) = \frac{1}{n}(\mu_{30} + 2\overline{Y}\,\mu_{20})$$

$$E\left(e_1 e_1'\right) = \frac{\mu_{02}}{n'} \qquad E\left(e_1 e_2'\right) = \frac{\mu_{03}}{n'}$$

$$E\left(e_1 e_3\right) = \frac{1}{n}(\mu_{21} + 2\overline{Y}\,\mu_{11}) \qquad E\left(e_1' e_2\right) = \frac{\mu_{03}}{n'}$$

$$E\left(e_1' e_2'\right) = \frac{\mu_{03}}{n'} \qquad E\left(e_1' e_3\right) = \frac{1}{n'}(\mu_{21} + 2\overline{Y}\,\mu_{11})$$

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

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

$$E\left(e_2'e_3\right)=\frac{1}{n'}(\mu_{22}+2\overline{Y}\,\mu_{12}-\mu_{02}\mu_{20}) \tag{1.5}$$

2. Bias and Mean Square Error of the proposed estimator d

The proposed estimator d given by (1.1) is

$$d=\hat{\theta}-\bar{y}^2\left\{1+\frac{k_1\left(\bar{x}-\bar{x}'\right)}{\bar{x}'}\right\}\left\{1+\frac{k_2\left(s_x^2-s_x'^2\right)}{s_x'^2}\right\}$$

In terms of e_i's, $i=0,1,2$ and by taking first degree of approximation, the proposed estimator reduces to

$$\begin{aligned}d-\sigma_Y^2 &= e_3-2\overline{Y}e_0-\frac{\overline{Y}^2k_2}{\mu_{02}}e_2+\frac{\overline{Y}^2k_2}{\mu_{02}}e_2'-\frac{\overline{Y}^2k_1}{\overline{X}}e_1+\frac{\overline{Y}^2k_1}{\overline{X}}e_1' \\ &-e_0^2-\frac{\overline{Y}^2k_2}{\mu_{02}^2}e_2'^2+\frac{\overline{Y}^2k_2}{\mu_{02}^2}e_2e_2'-\frac{2\overline{Y}k_2}{\mu_{02}}e_0e_2+\frac{2\overline{Y}k_2}{\mu_{02}}e_0e_2'-\frac{2\overline{Y}k_1}{\overline{X}}e_0e_1 \\ &+\frac{2\overline{Y}k_1}{\overline{X}}e_0e_1'-\frac{2\overline{Y}k_1k_2}{\overline{X}\mu_{02}}e_1e_2+\frac{2\overline{Y}k_1k_2}{\overline{X}\mu_{02}}e_1e_2'-\frac{2\overline{Y}^4k_1k_2}{\overline{X}\mu_{02}}e_1'e_2 \\ &+\frac{2\overline{Y}^4k_1k_2}{\overline{X}\mu_{02}}e_1'e_2'+\frac{\overline{Y}^2k_1k_2}{\overline{X}\mu_{02}}e_1'e_2-\frac{\overline{Y}^2k_1k_2}{\overline{X}\mu_{02}}e_1'e_2'\end{aligned} \tag{2.1}$$

Now taking expectation on both the sides of (2.1) and then using the values of the expectations given from (1.4) to (1.5), the bias in d to the first degree of approximation is given by

$$\text{Bias}\left(d\right)=E\left(d-\sigma_Y^2\right)=-\frac{\mu_{20}}{n}-\left(\frac{1}{n}-\frac{1}{n'}\right)\left(\frac{2\overline{Y}\mu_{11}}{\overline{X}}k_1+\frac{2\overline{Y}\mu_{12}}{\mu_{02}}k_2+\frac{\overline{Y}^2\mu_{03}}{\overline{X}\mu_{02}}k_1k_2\right) \tag{2.2}$$

Now squaring (2.1) on both the sides and then taking expectation, the mean square error of d to the first degree of approximation is given by

$$\begin{aligned}\text{MSE}\left(d\right)&=E\left(d-\sigma_Y^2\right)^2 \\ &=E\left(e_3\right)^2+4\overline{Y}^2E\left(e_0\right)^2+\frac{\overline{Y}^4k_2^2}{\mu_{02}^2}E\left(e_2^2\right)-\frac{\overline{Y}^4k_2^2}{\mu_{02}^2}E\left(e_2'^2\right)+\frac{\overline{Y}^4k_1^2}{\overline{X}^2}E\left(e_1^2\right) \\ &-\frac{\overline{Y}^4k_1^2}{\overline{X}^2}E\left(e_1'^2\right)-4\overline{Y}E\left(e_0e_3\right)-\frac{2\overline{Y}^2k_2}{\mu_{02}}E\left(e_2e_3\right)+\frac{2\overline{Y}^2k_2}{\mu_{02}}E\left(e_2'e_3\right) \\ &-\frac{2\overline{Y}^2k_1}{\overline{X}}E\left(e_1e_3\right)+\frac{2\overline{Y}^2k_1}{\overline{X}}E\left(e_1'e_3\right)+\frac{4\overline{Y}^3k_2}{\mu_{02}}E\left(e_0e_2\right)-\frac{4\overline{Y}^3k_2}{\mu_{02}}E\left(e_0e_2'\right)\end{aligned}$$

$$+\frac{4\bar{Y}^3k_1}{\bar{X}}E(e_0e_1)+\frac{4\bar{Y}^3k_1}{\bar{X}}E(e_0e_1')+\frac{2\bar{Y}^4k_1k_2}{\bar{X}\mu_{02}}E(e_1e_2)-\frac{2\bar{Y}^4k_1k_2}{\bar{X}\mu_{02}}E(e_1e_2')$$

$$-\frac{2\bar{Y}^4k_1k_2}{\bar{X}\mu_{02}}E(e_1'e_2)+\frac{2\bar{Y}^4k_1k_2}{\bar{X}\mu_{02}}E(e_1'e_2')$$

Now using the values of the expectations given from (1.4) to (1.5), the mean square error in d to the first degree of approximation is given by

$$\text{MSE }(d)=\frac{1}{n}\left(\mu_{40}-\mu_{20}^2\right)+\left(\frac{1}{n}-\frac{1}{n'}\right)\left(\frac{\bar{Y}^4\mu_{02}}{\bar{X}^2}k_1^2+\bar{Y}^4(\beta_2-1)k_2^2\right.$$

$$\left.-\frac{2\bar{Y}^2\mu_{21}}{\bar{X}}k_1+2\bar{Y}^2\mu_{20}k_2-\frac{2\bar{Y}^2\mu_{22}}{\mu_{02}}k_2+\frac{2\bar{Y}^4\mu_{03}}{\bar{X}\mu_{02}}k_1k_2\right) \qquad (2.3)$$

which attains the minimum for the optimum values of k_1 and k_2 given by

$$k_1=\frac{\bar{X}\left\{\mu_{02}^2\mu_{21}(\beta_2-1)-\mu_{03}(\mu_{22}-\mu_{20}\mu_{02})\right\}}{\bar{Y}^2(\beta_2-\beta_1-1)\mu_{02}^3} \text{ and} \qquad (2.4)$$

$$k_2=\frac{\left(\mu_{02}\mu_{22}-\mu_{03}\mu_{21}-\mu_{02}^2\mu_{20}\right)}{\bar{Y}^2(\beta_2-\beta_1-1)\mu_{02}^2} \qquad (2.5)$$

substituting the values of k_1 and k_2 given by (2.4) and (2.5) in (2.3), the minimum mean square error of d is given by

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

(2.6)

where $\gamma_1=\sqrt{\beta_1}$, $\beta_1=\frac{\mu_{03}^2}{\mu_{02}^3}$ and $\beta_2=\frac{\mu_{04}}{\mu_{02}^2}$.

3. Bias and Mean Square Error of the proposed estimator d_g

The proposed generalized estimator d_g in (1.2) is

$$d_g=\hat{\theta}-\bar{y}^2f(u,v) \qquad (3.1)$$

where $u = \frac{\bar{x}}{\bar{x}'}$ and $v = \frac{s_x^2}{s_x'^2}$, f_1 and f_2 being the first order partial derivatives of $f(u,v)$ with respect to u and v respectively at the point $(1,1)$ i.e. $\left(\frac{\partial}{\partial u} f(u,v)\right)_{(1,1)} = f_1$ and $\left(\frac{\partial}{\partial v} f(u,v)\right)_{(1,1)} = f_2$.

Expanding $f(u,v)$ in (3.1) in the third order Taylor's series about (1, 1), we have

$$d_g = \hat{\theta} - \bar{y}^2 \left[f(1,1) + (u-1)f_1 + (v-1)f_2 + \frac{1}{2!}\left\{ \begin{array}{l} (u-1)^2 f_{11} + (v-1)^2 f_{22} \\ +2(u-1)(v-1)f_{12} \end{array} \right\} + \frac{1}{3!}\left\{ (u-1)\frac{\partial}{\partial u} + (v-1)\frac{\partial}{\partial v} \right\}^3 f(u^*, v^*) \right]$$

where f_1 and f_2 are already defined and f_{11}, f_{22}, f_{12} are the second order partial derivatives given by $f_{11} = \left(\frac{\partial^2}{\partial u^2} f(u,v)\right)_{(1,1)}$, $f_{22} = \left(\frac{\partial^2}{\partial v^2} f(u,v)\right)_{(1,1)}$ and $f_{12} = \left(\frac{\partial^2}{\partial u \partial v} f(u,v)\right)_{(1,1)}$ and $u^* = 1 + h(u-1)$, $u^* = 1 + h(v-1)$ for $0 \prec h \prec 1$.

In terms of e_i's, $i = 0,1,2$ and by taking first degree of approximation, the proposed estimator reduces to

$$\begin{aligned} d_g - \sigma_Y^2 = {} & e_3 - 2\bar{Y}e_0 - \frac{\bar{Y}^2 f_2}{\mu_{02}} e_2 + \frac{\bar{Y}^2 f_2}{\mu_{02}} e_2' - \frac{\bar{Y}^2 f_1}{\bar{X}} e_1 + \frac{\bar{Y}^2 f_1}{\bar{X}} e_1' - e_0^2 \\ & - \frac{\bar{Y}^2 f_2}{\mu_{02}^2} e_2'^2 + \frac{\bar{Y}^2 f_2}{\mu_{02}^2} e_2 e_2' - 2\frac{\bar{Y} f_2}{\mu_{02}} e_0 e_2 + 2\frac{\bar{Y} f_2}{\mu_{02}} e_0 e_2' - 2\frac{\bar{Y} f_1}{\bar{X}} e_0 e_1 \\ & + 2\frac{\bar{Y} f_1}{\bar{X}} e_0 e_1' - \frac{\bar{Y}^2 f_{12}}{\bar{X}\mu_{02}} e_1 e_2 + \frac{\bar{Y}^2 f_{12}}{\bar{X}\mu_{02}} e_1 e_2' + \frac{\bar{Y}^2 f_{12}}{\bar{X}\mu_{02}} e_1' e_2 - \frac{\bar{Y}^2 f_{12}}{\bar{X}\mu_{02}} e_1' e_2' \\ & - \frac{\bar{Y}^2}{2!}\left\{ \frac{f_{11}}{\bar{X}^2}\left(e_1^2 + e_1'^2 - 2e_1 e_1'\right) + \frac{f_{22}}{\mu_{02}^2}\left(e_2^2 + e_2'^2 - 2e_2 e_2'\right) \right\} \end{aligned} \tag{3.2}$$

Now taking expectation on both the sides of (3.2) and using values of the expectation given in (1.4) and (1.5), the bias in d_g to the first degree of approximation is given by

$$\text{Bias}\left(d_g\right) = E\left(d_g - \sigma_Y^2\right) = -\frac{\mu_{20}}{n} - \left(\frac{1}{n} - \frac{1}{n'}\right)\begin{pmatrix} \dfrac{2\bar{Y}\mu_{12}}{\mu_{02}} f_2 + \dfrac{2\bar{Y}\mu_{11}}{\bar{X}} f_1 + \dfrac{\bar{Y}^2\mu_{03}}{\bar{X}\mu_{02}} f_{12} \\ + \dfrac{\bar{Y}^2\mu_{02}}{2\bar{X}^2} f_{11} + \dfrac{\bar{Y}^2(\beta_2 - 1)}{2} f_{22} \end{pmatrix} \tag{3.3}$$

Now squaring (3.2) on both the sides and then taking expectation and using values of the expectation given in (1.4) and (1.5), the mean square error in d_g to the first degree of approximation is given by

$$\text{MSE}\left(d_g\right) = E\left(d_g - \sigma_Y^2\right)^2$$

$$= \frac{1}{n}\left(\mu_{40} - \mu_{20}^2\right) + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(\frac{\bar{Y}^4\mu_{02}}{\bar{X}^2} f_1^2 + \bar{Y}^4(\beta_2 - 1) f_2^2 - \frac{2\bar{Y}^2\mu_{21}}{\bar{X}} f_1 \right.$$

$$\left. + 2\bar{Y}^2\mu_{20} f_2 - \frac{2\bar{Y}^2\mu_{22}}{\mu_{02}} f_2 + \frac{2\bar{Y}^4\mu_{03}}{\bar{X}\mu_{02}} f_1 f_2\right) \tag{3.4}$$

which attains the minimum for the optimum values of f_1 and f_2 given by

$$f_1 = \frac{\bar{X}\left\{\mu_{02}^2\mu_{21}(\beta_2 - 1) - \mu_{03}(\mu_{22} - \mu_{20}\mu_{02})\right\}}{\bar{Y}^2(\beta_2 - \beta_1 - 1)\mu_{02}^3} \quad \text{and} \tag{3.5}$$

$$f_2 = \frac{\left(\mu_{02}\mu_{22} - \mu_{03}\mu_{21} - \mu_{02}^2\mu_{20}\right)}{\bar{Y}^2(\beta_2 - \beta_1 - 1)\mu_{02}^2} \tag{3.6}$$

substituting the values of f_1 and f_2 given by (3.5) and (3.6) in (3.4), the minimum mean square error in d_g is given by

$$\text{MSE}\left(d_g\right)_{\min} = \text{MSE}\left(s_y^2\right) - \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\frac{\mu_{21}^2}{\mu_{02}} + \frac{\mu_{21}^2}{(\beta_2 - \beta_1 - 1)\mu_{02}^2}\left(\frac{\mu_{02}\mu_{20}}{\mu_{21}} - \frac{\mu_{22}}{\mu_{21}} + (\mu_{02})^{1/2}\gamma_1\right)^2\right\} \tag{3.7}$$

4. Efficiency comparison with the available estimators

For comparing the efficiency of the proposed estimators d and d_g, let us consider the following

(i) Usual conventional unbiased estimator of population variance in case of SRSWOR

$$\hat{d}_1 = s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \text{with } MSE\left(\hat{d}_1\right) = \frac{1}{n}\left(\mu_{40} - \mu_{20}^2\right) \tag{4.1}$$

from (4.1) and (2.6) and (3.7) it is clear that the proposed estimators d and d_g are having mean square error lesser than the usual conventional unbiased estimator.

(ii) Estimator of population variance given by Peeyush Misra and R. Karan Singh

$$\hat{d}_2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}.f\left(\bar{y}, \bar{x}, \bar{x}'\right) \quad \text{with} \quad MSE\left(\hat{d}_2\right) = \frac{1}{n}\left(\mu_{40} - \mu_{20}{}^2\right) - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{\mu_{21}{}^2}{\mu_{02}} \tag{4.2}$$

from (4.2) and (2.6) and (3.7), it is clear that the proposed estimators d and d_g are having mean square error lesser than the mean square error of the estimator of population variance given by Peeyush Misra and R. Karan Singh (2014).

5. Conclusion

The comparative study of the proposed estimators of population variance establishes their 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 (2014).

Acknowledgements

The author is very indebted to the editor in chief and also to the referees for their valuable suggestions leading to the improvement of the contents and presentation of the original manuscript.

References

1. Cochran, W.G. (1977). Sampling Techniques, 3rd edition, John Wiley and Sons, New York.
2. Murthy, M. N. (1967). Sampling Theory and Methods, 1st edition, Statistical Publishing Society, Calcutta (India).
3. Peeyush Misra and R. Karan Singh (2014). Estimation of population variance using a generalized double sampling estimator, Sri Lankan Journal of Applied Statistics, Vol. 15, No. 3, p. 211-220.
4. 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.