Journal of Scientific Research
Banaras Hindu University, Varanasi

Vol. 60, 2016 : 97-109
ISSN : 0447-9483

VARIANCE ESTIMATION USING ARITHMETIC MEAN, GEOMETRIC MEAN AND HARMONIC MEAN UNDER SIMPLE RANDOM SAMPLING

Prabhakar Mishra and *Rajesh Singh
Department of Statistics, Banaras Hindu University, Varanasi, 221005
*Corresponding author

Abstract

In this paper we have suggested some estimators of population variance using auxiliary information based on arithmetic mean, geometric mean and harmonic mean.We have also suggested an almost unbiased estimator for estimating population variance. The expressions of mean squared error (MSE) have been derived up to the first order of approximation. It has been shown that almost unbiased estimator gives better result than estimators included in the paper. Numerical illustrations are given in support of the theoretical study.

Keywords: Auxiliary information, arithmetic mean, geometric mean, harmonic mean, Mean square error, unbiased estimator and simple random sampling.

Introduction

Use of auxiliary variables is very common in estimating various population parameters with greater efficiency. Let x be the auxiliary variable, highly correlated to the study variable y. If information on an auxiliary variable is readily available then it is a well-known fact that the ratio-type and regression-type estimators can be used for estimation of parameters of interest, due to increase in efficiency of these estimators.

Variations are present everywhere in our daily life. It is the law of nature that no two things or individuals are exactly alike. For instance, a physician needs a full understanding of variations in the degree of human blood pressure, body temperature and pulse rate for adequate prescription (see Singh and Solanki (2012). Consider the problem of estimating the population variance S_y^2 of the study variable y assuming S_x^2 is known. Isaki (1983) presented the ratio estimator for the population variance using the auxiliary information. Several authors including Singh and Singh (2001, 2003), Jhajj et al. (2005), Kalidar and Cingi (2007), Shabbir and Gupta (2007), Grover (2010), Singh and Solanki (2012), Singh et al. (2014) and Singh and Singh (2015) suggested improved estimators of S_y^2.

Let $U = (U_1, U_2, ..., U_N)$ be the finite population of size N out of which a sample of size n is drawn according to simple random sampling without replacement (SRSWOR) technique. Let Y and X denote the study variable and auxiliary variable taking values y_i and x_i respectively on the i^{th} unit U_i of the population U. Let $(\bar{y}, \bar{x})$ be the sample mean estimator of $(\bar{Y}, \bar{X})$, the population means of yand x respectively.In order to have a survey estimate of the population variance S_y^2 of the study character Y assuming the knowledge of the population variance S_x^2 of the auxiliary character X,the usual estimator of population variance of the study variable Y is defined by

$$P_o = s_y^2 \tag{1}$$

Isaki (1983) suggested a ratio type estimator of population variance, given as

$$P_{IR} = s_y^2 \left[\frac{S_x^2}{s_x^2} \right] \tag{2}$$

Bahl and Tuteja (1991) suggested a ratio type exponential estimator of the population variance as

$$P_{exp} = s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right] \tag{3}$$

The dual to ratio estimator and dual to ratio type exponential estimator of population variance is given by

$$P_1 = s_y^2 \left[\frac{s_x^{*2}}{S_x^2} \right] \tag{4}$$

$$P_2 = s_y^2 \exp\left[\frac{s_x^{*2} - S_x^2}{S_x^2 + s_x^{*2}} \right] \tag{5}$$

where, $s_x^{*2} = \dfrac{NS_x^2 - ns_x^2}{N - n}$

Isaki (1983) also suggested the regression type estimator of finite population variance , given by

$$P_{reg} = s_y^2 + k_o \left[S_x^2 - s_x^2 \right] \tag{6}$$

where $s_y^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$, $s_x^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$,

$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2$ and $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$.

To obtain the Bias and MSE, we write

$s_y^2 = S_y^2[1+\omega_o]$ and $s_x^2 = S_x^2[1+\omega_1]$

such that $E(\omega_o) = E(\omega_1) = 0$, $E(\omega_o^2) = f(\lambda_{40}-1)$, $E(\omega_1^2) = f(\lambda_{04}-1)$

and $E(\omega_o\omega_1) = f(\lambda_{22}-1)$.

where $f = \frac{1}{n} - \frac{1}{N}$, $\lambda_{pq} = \frac{\mu_{pq}}{\mu_{20}^{\frac{p}{2}}\mu_{02}^{\frac{q}{2}}}$ and $\mu_{pq} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^p\left(x_i - \overline{X}\right)^q$

(where p, q being non-negative integers)

Then, $s_x^{*2} = (1-g\omega_1)S_x^2$; $g = \frac{n}{N-n}$.

The variance of the usual estimator s_y^2 under SRSWOR is given as

$$Var(P_o) = Var(s_y^2) = fS_y^4(\lambda_{40}-1) \tag{7}$$

To the first order of approximation, the Bias and MSE expressions of the estimators P_1 and P_2 are respectively given by

$$Bias(P_1) = -S_y^2 gf(\lambda_{22}-1) \tag{8}$$

$$Bias(P_2) = -S_y^2 f\left[\frac{g^2}{8}(\lambda_{04}-1) + \frac{g}{2}(\lambda_{22}-1)\right] \tag{9}$$

$$MSE(P_1) = S_y^4 f\left[(\lambda_{40}-1) + g^2(\lambda_{04}-1) - 2g(\lambda_{22}-1)\right] \tag{10}$$

$$MSE(P_2) = S_y^4 f\left[(\lambda_{40}-1) + \frac{g^2}{4}(\lambda_{04}-1) - g(\lambda_{22}-1)\right] \tag{11}$$

To the first order of approximation the MSE's of the estimators P_{IR} , P_{exp} and P_{reg} are respectively given by

$$MSE(P_{IR}) = S_y^4 f\left[(\lambda_{40}-1) + (\lambda_{04}-1) - 2(\lambda_{22}-1)\right] \tag{12}$$

$$MSE(P_{\exp}) = S_y^4 f\left[(\lambda_{40}-1) + \frac{1}{4}(\lambda_{04}-1) - (\lambda_{22}-1)\right] \tag{13}$$

$$MSE_{\min.}\left(P_{reg}\right)=S_y^4 f\left[\left(\lambda_{40}-1\right)-\frac{\left(\lambda_{22}-1\right)^2}{\left(\lambda_{04}-1\right)}\right] \tag{14}$$

Suggested Estimators

Motivated by Singh et al. (2014), we have proposed some estimators of population variance of the study variable Y based on arithmetic mean, geometric mean and harmonic mean (AM,GM,HM) of the estimators $\left(P_o, P_1\right), \left(P_o, P_2\right), \left(P_1, P_2\right)$ and $\left(P_o, P_1, P_2\right)$. The properties of suggested estimators have been studied up to first order of approximation.

The estimators based on P_o and P_1

Taking the AM, GM and HM of the estimators P_o and P_1, we get the following estimators of the population variance respectively as,

$$P_3^{AM}=\frac{P_o+P_1}{2}=\left(\frac{s_y^2}{2}\right)\left(1+\frac{s_x^{*2}}{S_x^2}\right) \tag{15}$$

$$P_3^{GM}=\left(P_oP_1\right)^{\frac{1}{2}}=s_y^2\left(\frac{s_x^*}{S_x}\right) \tag{16}$$

$$P_3^{HM}=\frac{2}{\left(\frac{1}{P_o}+\frac{1}{P_1}\right)}=\frac{2s_y^2}{\left(1+\frac{S_x^2}{s_x^{*2}}\right)} \tag{17}$$

To the first degree of approximation, the mean squared errors of P_3^j ; (j= AM,GM,HM) are respectively given by

$$MSE\left(P_3^j\right)=S_y^4 f\left[\left(\lambda_{40}-1\right)+\frac{g^2}{4}\left(\lambda_{04}-1\right)-g\left(\lambda_{22}-1\right)\right] \tag{18}$$

The estimators based on P_o and P_2

The estimators of S_y^2 based on AM, GM and HM of the estimators P_o and P_2 are respectively defined as

$$P_4^{AM} = \frac{P_o + P_2}{2} = \left(\frac{s_y^2}{2}\right)\left(1 + \exp\left(\frac{s_x^{*2} - S_x^2}{s_x^{*2} + S_x^2}\right)\right) \tag{19}$$

$$P_4^{GM} = \left(P_o P_2\right)^{\frac{1}{2}} = s_y^2 \exp\left(\frac{1}{2}\left(\frac{s_x^{*2} - S_x^2}{s_x^{*2} + S_x^2}\right)\right) \tag{20}$$

$$P_4^{HM} = \frac{2}{\left(\frac{1}{P_o} + \frac{1}{P_2}\right)} = \frac{2s_y^2}{\left(1 + \exp\left(\frac{S_x^2 - s_x^{*2}}{s_x^{*2} + S_x^2}\right)\right)} \tag{21}$$

To the first degree of approximation, the mean squared errors of P_4^j ; (j= AM,GM,HM) are respectively given by

$$MSE\left(P_4^j\right) = S_y^4 f\left[\left(\lambda_{40} - 1\right) + \frac{g^2}{16}\left(\lambda_{04} - 1\right) - \frac{g}{2}\left(\lambda_{22} - 1\right)\right] \tag{22}$$

The estimators based on P_1 and P_2

We propose the following estimators of S_y^2 based on AM, GM and HM of the estimators P_1 and P_2 respectively as

$$P_5^{AM} = \frac{P_1 + P_2}{2} = \left(\frac{s_y^2}{2}\right)\left(\frac{s_x^{*2}}{S_x^2} + \exp\left(\frac{s_x^{*2} - S_x^2}{s_x^{*2} + S_x^2}\right)\right) \tag{23}$$

$$P_5^{GM} = \left(P_1 P_2\right)^{\frac{1}{2}} = s_y^2\left(\frac{s_x^*}{S_x}\right)\exp\left(\frac{1}{2}\left(\frac{s_x^{*2} - S_x^2}{s_x^{*2} + S_x^2}\right)\right) \tag{24}$$

$$P_5^{HM} = \frac{2}{\left(\frac{1}{P_1} + \frac{1}{P_2}\right)} = \frac{2s_y^2}{\left(\frac{S_x^2}{s_x^{*2}} + \exp\left(\frac{S_x^2 - s_x^{*2}}{s_x^{*2} + S_x^2}\right)\right)} \tag{25}$$

To the first degree of approximation, the mean squared errors of P_5^j ; (j= AM, GM, HM) are respectively given by

$$MSE\left(P_5^j\right)=S_y^4 f\left[\left(\lambda_{40}-1\right)+\frac{9g^2}{16}\left(\lambda_{04}-1\right)-\frac{3g}{2}\left(\lambda_{22}-1\right)\right] \tag{26}$$

The estimators based on P_o, P_1 and P_2

We suggest the following estimators of S_y^2 based on AM, GM and HM of the estimators P_o , P_1 and P_2 respectively as

$$P_6^{AM}=\frac{P_o+P_1+P_2}{3}=\left(\frac{s_y^2}{3}\right)\left(1+\frac{s_x^{*2}}{S_x^2}+\exp\left(\frac{s_x^{*2}-S_x^2}{s_x^{*2}+S_x^2}\right)\right) \tag{27}$$

$$P_6^{GM}=\left(P_oP_1P_2\right)^{\frac{1}{2}}=s_y^2\left(\frac{s_x^*}{S_x}\right)^{\frac{2}{3}}\exp\left(\frac{1}{3}\left(\frac{s_x^{*2}-S_x^2}{s_x^{*2}+S_x^2}\right)\right) \tag{28}$$

$$P_6^{HM}=\frac{3}{\left(\frac{1}{P_o}+\frac{1}{P_1}+\frac{1}{P_2}\right)}=\frac{3s_y^2}{\left(1+\frac{S_x^2}{s_x^{*2}}+\exp\left(\frac{S_x^2-s_x^{*2}}{s_x^{*2}+S_x^2}\right)\right)} \tag{29}$$

To the first degree of approximation, the mean squared errors of P_6^j ; (j= AM, GM, HM) are respectively given by

$$MSE\left(P_6^j\right)=S_y^4 f\left[\left(\lambda_{40}-1\right)+\frac{g^2}{4}\left(\lambda_{04}-1\right)-g\left(\lambda_{22}-1\right)\right] \tag{30}$$

Almost Unbiased Estimator

Motivated by Singh et al. (2007), we suggest a class of estimator of population variance.

Suppose $P_o=s_y^2$, $P_1=s_y^2\left[\frac{s_x^{*2}}{S_x^2}\right]$ and $P_2=s_y^2\exp\left[\frac{s_x^{*2}-S_x^2}{S_x^2+s_x^{*2}}\right]$,

such that $\left(P_o,P_1,P_2\right)\in T$, where T denotes the set of all possible estimators for estimating the population variance S_y^2. By definition, the set T is a linear variety if

$$P_t = \sum_{i=0}^{2} t_i P_i \quad (\in T) \tag{31}$$

for

$$\sum_{i=0}^{2} t_i = 1 \ (t_i \in \Re) \tag{32}$$

where $t_i ; (i = 0,1,2)$ denotes the scalar constants used for reducing the bias in the set of real numbers.

Expressing the estimator P_t in terms of ω's, we have

$$P_t = S_y^2 \left(1 + \omega_o - \omega_1 g \left(t_1 + \frac{t_2}{2} \right) - \frac{1}{8} \omega_1^2 g^2 t_2 - \omega_o \omega_1 g \left(t_1 + \frac{t_2}{2} \right) \right) \tag{33}$$

Subtracting S_y^2 from equation (33), we get

$$P_t - S_y^2 = S_y^2 \left(\omega_o - \omega_1 gt - \frac{1}{8} \omega_1^2 g^2 t_2 - \omega_o \omega_1 gt \right) \tag{34}$$

where

$$t = t_1 + \frac{t_2}{2} \Rightarrow 2t = 2t_1 + t_2 \tag{35}$$

After squaring both sides of equation (34) and taking expectations, we get MSE of the estimator P_t up to the first order of approximation as

$$MSE(P_t) = S_y^4 f \left[(\lambda_{40} - 1) + g^2 t^2 (\lambda_{04} - 1) - 2gt(\lambda_{22} - 1) \right] \tag{36}$$

Differentiating equation (36) with respect to t, we get minimum $MSE(P_t)$ at

$$t = \frac{(\lambda_{22} - 1)}{g(\lambda_{04} - 1)} = L(say) \tag{37}$$

Thus the minimum MSE of P_t is given by

$$MSE_{\min.}(P_t) = S_y^4 f\left[(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{38}$$

From equation (35) and (37), we have

$$2t_1 + t_2 = 2L \tag{39}$$

From equation (32) and (39), we have only two equations in three unknowns, so we can not find unique values of t_i's, $(i = 0,1,2)$. In order to get unique values of t_i's , we shall impose the linear restriction :

$$\sum_{i=0}^{2} t_i B(P_i) = 0 \tag{40}$$

where $B(P_i)$ denotes the bias in the i^{th} estimator.

Equations (32), (39) and (40) can be written in the matrix from as

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & B(P_1) & B(P_2) \end{bmatrix} \begin{bmatrix} t_o \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2L \\ 0 \end{bmatrix} \tag{41}$$

Using equation (41), we get the unique values of t_i's, $(i = 0,1,2)$ as

$$\left.\begin{aligned} t_o &= 1 - L + 4L^2 \\ t_1 &= L + 4L^2 \\ t_2 &= -8L^2 \end{aligned}\right\} \tag{42}$$

Use of these t_i's, $(i = 0,1,2)$ remove the bias of the estimator P_t up to terms of first order at (31).

Efficiency Comparisons

From equations (7), (10), (11), (12), (13), (14), (18), (22), (26), (30) and (38), we have

1. $MSE(P_o) \geq MSE(P_1)$ if $g \leq \dfrac{2(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$

2. $MSE(P_o) \geq MSE(P_2)$ if $g \leq \dfrac{4(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$

3. $MSE(P_o) \geq MSE(P_{IR})$ if $\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} \geq \frac{1}{2}$

4. $MSE(P_o) \geq MSE(P_{exp})$ if $\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} \geq \frac{1}{4}$

5. $MSE(P_{IR}) \geq MSE(P_1)$ if

Case 1 $g \geq 1$ then $g \leq \frac{2(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1$

Case 2 $g \leq 1$ then $g \geq \frac{2(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1$

6. $MSE(P_2) \leq MSE(P_{exp})$ if

Case 1 $g \geq 1$ then $g \leq \frac{4(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1$

Case 2 $g \leq 1$ then $g \geq \frac{4(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1$

7. $MSE(P_o) \geq MSE_{min.}(P_{reg})$ if $\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} \geq 0$

8. $MSE(P_o) \geq MSE(P_3^j)$ if $g \leq \frac{4(\lambda_{22}-1)}{(\lambda_{04}-1)}$; j = (AM, GM, HM)

9. $MSE(P_o) \geq MSE(P_4^j)$ if $g \leq \frac{8(\lambda_{22}-1)}{(\lambda_{04}-1)}$; j = (AM, GM, HM)

10. $MSE(P_o) \geq MSE(P_5^j)$ if $g \leq \frac{8(\lambda_{22}-1)}{3(\lambda_{04}-1)}$; j = (AM, GM, HM)

11. $MSE(P_o) \geq MSE(P_6^j)$ if $g \leq \frac{4(\lambda_{22}-1)}{(\lambda_{04}-1)}$; j = (AM, GM, HM)

12. $MSE(P_o) \geq MSE_{min.}(P_t)$ if $\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} \geq 0$

Empirical Study

For empirical study, we consider three natural population data sets.

Population 1: Murthy (1967, p.226)

Y: Output

X: Number of workers.

N = 80, n = 10, $\lambda_{40} = 2.2667$, $\lambda_{04} = 3.65$, $\lambda_{22} = 2.3377$.

Population 2: Murthy (1967, p.127)

Y: Cultivated area (in acres).

X: Area in square miles.

N = 80, n = 10, $\lambda_{40} = 2.373$, $\lambda_{04} = 2.0193$, $\lambda_{22} = 1.6757$.

Population 3: Sukhatme and Sukhatme (1970, p.185)

Y: Area under wheat in 1937 (in acres).

X: Cultivated area in 1931.

N = 80, n= 10, $\lambda_{40} = 3.5469$, $\lambda_{04} = 3.2816$, $\lambda_{22} = 2.6601$.

Table 1: Value of t_i's, $(i = 0,1,2)$

Scalars	Population		
	I	II	III
t_o	47.410275	82.490732	99.670500
t_1	53.477369	90.771415	108.856949
t_2	-99.887644	-172.262147	-207.52745

Using these values of t_i's, $(i = 0,1,2)$ given in Table 1, one can reduce the bias to the first order approximation in the estimators P_t at (31).

Table 2 : Percent relative efficiency of different estimators with respect to S_y^2

Scalars	Population		
	I	II	III
P_o	100.0000	100.0000	100.0000
P_1	134.9590	114.3457	120.1850
P_2	116.3049	170.1257	109.7147
P_{IR}	102.0462	131.9050	168.8589
P_{exp}	214.1504	144.2037	174.7804
P_{reg}	214.1725	148.4205	190.2083
P_3^j	116.3049	107.1257	109.7147
P_4^j	107.8474	103.5417	104.7576
P_5^j	125.3574	110.7340	114.8620
P_6^j	116.3049	107.1226	109.7147
P_t	214.1725	148.4205	190.2083

Conclusion

From Table 2, we observe that the estimators constructed by taking AM,GM and HM of the usual estimator S_y^2, dual to Isaki (1983) estimator and dual to exponential ratio type variance estimator performs better than S_y^2. From Table 2, we also observe that dual to exponential ratio type variance estimator performs inferior than exponential ratio type estimator . Reason is that the condition (6) in efficiency comparison is rarely met in practice. It is not advisable to construct dual to exponential ratio type variance estimators. The unbiased estimator P_t performs better than all other estimators considered here.

References

1. Adichwal, N.K., Sharma, P. and Singh, R. (2015). Generalized class of estimators for population variance using information on two auxiliary variables. Int. J. Appl. Comput. Math.

2. Bahl, S., and Tuteja, R.K. (1991). Ratio and product type exponential estimators. Journal of Information and Optimization sciences, 12(1) ; 159-164.

3. Cochran, W.G. (1997). Sampling Techniques.New York, John Wiley and Sons.

4. Grover, L. K. (2010). A correction note on improvement in variance estimation using auxiliary information. Communications in Statistics Theory and Methods, 39, 753-764.

5. Isaki, C.T. (1983). Variance estimation using auxiliary information. Jour. Amer. Stat. Assoc., 78, 117-123.

6. Jhajj, H. S., Sharma, M. K. and Grover, L. K. (2005). An efficient class of chain estimators of population variance under sub-sampling scheme. J. Japan Statist. Soc. Vol 35, 273-286.

7. Kadilar, C. and Cingi, H.(2007). Improvement in Variance Estimation in Simple Random Sampling.Communications in Statistics-Theory and Methods, 36(11): 2075-2081.

8. Murthy, M.N., (1967). Sampling Theory and Methods.Statistical Pub.Society. Calcutta, India.

9. Sukhatme, P.V., Sukhatme, B.V. (1970). Sampling Theory of Surveys with Applications. Asia Publishing House, New Delhi.

10. Singh, S. (2003). Advanced Sampling Theory with Applications: How Michael "Selected" Amy. Vol. 1 and 2, pp. 1-1247, Kluwer Academic Publisher, The Netherlands.

11. Singh, H.P., Singh, A.K. and Solanki, R.S. (2014). Estimation of finite population variance using auxiliary information in Sample Surveys. Statistica.

12. Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys.Biometrika, 67, 199-204.

13. Singh, R., Chauhan, P. and Sawan, N. (2007). Auxiliary information anda priori values in construction of improved estimators.(edited Book), p. 41-53.

14. Shabbir, J. and Gupta, S. (2007). On improvement in variance estimation using auxiliary information. Communications in Statistics- Theory and Methods, 36(12): 2177-2185.

15. Singh, H. P. and Solanki, R.S. (2012): A new procedure for variance estimation in simple random sampling using auxiliary information. Stat. Paps. DOI10.1007/s00362-012-0445-2.

16. Singh, H.P. and Singh, R. (2001): Improved ratio-type estimator for variance using auxiliary information. J Indian SocAgric Stat 54(3):276–287.

17. Singh,H.P. and Singh, R. (2003): Estimation of variance through regression approach in two phase sampling. Aligarh Journal of Statistics, 23, 13-30.

18. Singh, R. and Singh, V. K. (2015). Improved class of estimators for variance under single and two phase sampling. J. Stat. Appl. Pro. 4, 161-171.

19. Vishwakarma, G.K., Gangele, R. K. and Singh, R. (2014). An efficient variant of dual to ratio and product estimator in sample survey. The Philippine Statistician, Vol. 63, p.2129.