Volume 65, Issue 5, 2021
Journal of Scientific Research
of
The Banaras Hindu University

Estimating Population Variance Using Median as an Auxiliary Information

Rajesh Singh[1] and Sakshi Rai*[2]

[1]Department of Statistics, Banaras Hindu University, Varanasi-22100, rsinghstat@gmail.com
*[2]Department of Statistics, Banaras Hindu University, Varanasi-221005, sakshi.rai@bhu.ac.in

Abstract: **The paper comprises an estimator for estimating population variance using auxiliary information as median, since auxiliary information correlated with the study variable boosts the efficiencies of the estimators. Simple random sampling without replacement scheme has been applied as a technique. The proposed estimator's mean squared error is derived up to the first order of approximation and the obtained efficiency is collated with some of the existing estimators mentioned in the literature. On the basis of result obtained from the simulation study that has been performed for the result's confirmation. It has seen that proposed estimator has better efficiency and hence works better than the existing estimators.**

Index Terms: **SRSWOR, Auxiliary Variable, Median, Mean Squared Error, Efficiency.**

I. Introduction

Involving auxiliary information along with the study variable is a way of amplifying the efficiencies of the estimators developed to estimate the population parameters. It is a fact that as much information get included in the estimation along with the study variable having some sort of correlation with it then that will result in minimizing the mean squared error of the estimators. Thus in this paper along with the variance, median of the auxiliary variable has been used for the purpose of increasing the efficiency of the proposed estimator .Variance being a measure of dispersion shows the average deviation of observations from mean. Authors like Das (1978), Isaki (1983), Prasad & Singh (1990), Singh et al (2009), Singh & Solanki (2013), Singh & Malik (2014) had worked on the concept of improving the estimator with the involvement of auxiliary information.

Median which is a positional average divides the data into two equal parts and it is an appropriate measure when the data is skewed like age-income. Subramani & Kumarapandiyan (2012b) has worked in estimation of variance utilizing information regarding median.

A. Notations & Terminologies

Let (x_i, y_i) where (i =1, 2,….n) be the pair of sample observations of size n drawn by applying the simple random sampling scheme (SRSWOR) from a finite population of size N. (X_i, Y_i) ,(i=1,2,……N) be the pair of population observations of the auxiliary variable and study variable respectively. S_y^2 and S_x^2 are the population variances of study variable and auxiliary variable respectively while s_y^2 and s_x^2 are the sample variances of study variable and auxiliary variable that are the unbiased estimators of S_y^2 and S_x^2 respectively. M and m are the population and sample median of the auxiliary variable.

The error terms are defined as-

$$s_y^2 = S_y^2(1+e_0) \tag{1.1}$$

$$s_x^2 = S_x^2(1+e_1) \tag{1.2}$$

$$m = M(1+e_2) \tag{1.3}$$

The expectation of error terms

$$E(e_0) = E(e_1) = 0 \tag{1.4}$$

$$E(e_2) = \left(\frac{\bar{M}-M}{M}\right) \tag{1.5}$$

$$E(e_0^2) = \gamma(\partial_{400} - 1) \tag{1.6}$$

$$E(e_1^2) = \gamma(\partial_{040} - 1) \tag{1.7}$$

$$E(e_2^2) = \gamma C_m^2 \tag{1.8}$$

$E(e_0e_1) = \gamma(\partial_{220} - 1)$ (1.9)
$E(e_1e_2) = \gamma\partial_{021m}C_m$ (1.10)
$E(e_0e_2) = \gamma\partial_{201m}C_m$ (1.11)

Where,

$$\partial_{pqr} = \frac{\mu_{pqr}}{\mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2}}$$

$$\mu_{pqr} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^p\,(X_i - \bar{X})^q(X_{1i} - \bar{X}_{1i})^r$$

p, q, r *be any non* $-$ *negative integers*

$$\partial_{pqrm} = \frac{m\mu_{pqr}}{m\mu_{200}^{p/2}m\mu_{020}^{q/2}m\mu_{002}^{r/2}}$$

$$m\mu_{pqr} = \frac{1}{\binom{N}{n}}\sum_{i=1}^{\binom{N}{n}}(\bar{y}_i - \bar{Y})^p(\bar{x}_i - \bar{X})^q(m_i - M)^r$$

p, q, r *be any non* $-$ *negative integers*

$$\gamma = \left(\frac{1}{n} - \frac{1}{N}\right)$$

$$\bar{M} = \frac{1}{\binom{N}{n}}\sum_{i=1}^{\binom{N}{n}} m_i$$

$m_i, \left(i = 1,2, \ldots\ldots\ldots \binom{N}{n}\right)$ is i^{th} sample median

II. EXISTING ESTIMATORS

1. $t_0 = s_y^2$ (2.1)

Expression of its variance up to the first order of approximation is given by-

$$mse(t_0) = \gamma S_y^4(\partial_{400} - 1) \quad (2.2)$$

2. $t_r = s_y^2\left(\frac{S_x^2}{s_x^2}\right)$ (2.3)

Expression of its mean squared error (MSE) up to the first order of approximation is given by-

$$mse(t_r) = \gamma S_y^4[(\partial_{400} - 1) + (\partial_{040} - 1) - 2(\partial_{220} - 1)] \quad (2.4)$$

3. $t_{lr} = s_y^2 + \beta(S_x^2 - s_x^2)$ (2.5)

Expression of its mean squared error up to the first order of approximation is given by-

$$mse(t_{lr}) = \gamma S_y^4[(\partial_{400} - 1)(1 - \rho^2)] \quad (2.6)$$

And $\rho = \frac{(\partial_{220}-1)}{\sqrt{(\partial_{400}-1)(\partial_{040}-1)}}$

4. Upadhyaya and Singh (1999) estimator t_2 as

$$t_2 = s_y^2\left[\frac{S_x^2 + \beta_{2(x)}}{s_x^2 + \beta_{2(x)}}\right] \quad (2.7)$$

5. Kadilar and Cingi (2006) proposed estimators t_2 *to* t_4 incorporating the information regarding coefficient of variation and measure of kurtosis

$$t_2 = s_y^2\left[\frac{S_x^2 + C_x}{s_x^2 + C_x}\right] \quad (2.8)$$

$$t_3 = s_y^2\left[\frac{S_x^2\beta_{2(x)} + C_x}{s_x^2\beta_{2(x)} + C_x}\right] \quad (2.9)$$

$$t_4 = s_y^2\left[\frac{S_x^2C_x + \beta_{2(x)}}{s_x^2C_x + \beta_{2(x)}}\right] \quad (2.10)$$

6. Subramani and Kumarpandiyan (2012a) proposed estimators of the form t_5 *to* t_9 including the auxiliary information of first and third kurtosis, inter quartile range, semi-quartile range and quartile average.

$$t_5 = s_y^2\left[\frac{S_x^2 + Q_1}{s_x^2 + Q_1}\right] \quad (2.11)$$

$$t_6 = s_y^2\left[\frac{S_x^2 + Q_3}{s_x^2 + Q_3}\right] \quad (2.12)$$

$$t_7 = s_y^2\left[\frac{S_x^2 + Q_r}{s_x^2 + Q_r}\right] \quad (2.13)$$

$$t_8 = s_y^2\left[\frac{S_x^2 + Q_d}{s_x^2 + Q_d}\right] \quad (2.14)$$

$$t_9 = s_y^2\left[\frac{S_x^2+Q_a}{s_x^2+Q_a}\right] \quad (2.15)$$

The expression of mean squared error (MSE) for the estimators $t_1, t_2, \ldots\ldots\ldots\ldots\ldots .. t_9$ up to the first order of approximation is given by-

$$mse(t_i) = \gamma S_y^4[(\partial_{400} - 1) + \omega_i^2(\partial_{040} - 1) - 2\omega_i(\partial_{220} - 1)]\,, i = 1,2,3, \ldots\ldots\ldots .9 \quad (2.16)$$

$$\omega_1 = \frac{S_x^2}{S_x^2 + \beta_{2(x)}}$$

where $\beta_{2(x)}$ *is measure of skewness based on X observations*

$$\omega_2 = \frac{S_x^2}{S_x^2 + C_x}$$

where C_x is the coefficient of variation based on X observations

$$\omega_3 = \frac{S_x^2\beta_{2(x)}}{S_x^2\beta_{2(x)} + C_x}$$

$$\omega_4 = \frac{S_x^2C_x}{S_x^2C_x + \beta_{2(x)}}$$

$$\omega_5 = \frac{S_x^2}{S_x^2 + Q_1}$$

where Q_1 is first quartile

$$\omega_6 = \frac{S_x^2}{S_x^2 + Q_3} \text{ where } Q_3 \text{ is third Quartile}$$

$$\omega_7 = \frac{S_x^2}{S_x^2 + Q_r}\,, \text{where } Q_r = Q_3 - Q_1$$

$$\omega_8 = \frac{S_x^2}{S_x^2 + Q_d} \text{ where, } Q_d = \frac{Q_3 - Q_1}{2}$$

$$\omega_9 = \frac{S_x^2}{S_x^2+Q_a} \text{ where, } Q_a = \frac{Q_3+Q_1}{2}$$

III. PROPOSED ESTIMATOR

Motivated by Mishra et al. (2017), we have proposed a log type estimator involving median as an auxiliary information adopting SRSWOR sampling scheme. The MSE is derived up to the first order of approximation.

$$t_{prop} = \left[(\tau_1 + 1)s_y^2 + \tau_2 log\left(\frac{s_x^2}{S_y^2}\right)\right] exp\left(\frac{m - M}{m + M}\right) \quad (3.1)$$

The expression of minimum mean squared error up to the first order of approximation is given by (3.2)

$$\min mse(t_{prop}) = A\tau_{1opt}^2 + B\tau_{1opt} + C\tau_{2opt}^2 + D\tau_{2opt} + E\tau_{1opt}\tau_{2opt} + F \quad (3.2)$$

Where,

$$A = S_y^4\left[1 + \gamma(\partial_{400} - 1) + \left(\frac{\bar{M} - M}{M}\right) + 2\gamma\partial_{201m}C_m\right]$$

$$B = S_y^4\left[\left(\frac{\bar{M} - M}{M}\right) + \frac{1}{4}\gamma C_m^2 + 2\gamma(\partial_{400} - 1) + 3\gamma\partial_{201m}C_m\right]$$

$$C = \gamma(\partial_{040} - 1)$$

$$D = \gamma S_y^2[2(\partial_{220} - 1) + \partial_{021m}C_m]$$

$$E = \gamma S_y^2[(\partial_{040} - 1) + 2\partial_{021m}C_m + 2(\partial_{220} - 1)]$$

$$F = \gamma S_y^4\left[(\partial_{400} - 1) + \frac{1}{4}C_m^2 + \partial_{201m}C_m\right]$$

$$\tau_{1opt} = \frac{2BC - DE}{E^2 - 4AC} \quad (3.3)$$

$$\tau_{2opt} = \frac{2AD - BE}{E^2 - 4AC} \quad (3.4)$$

IV. SIMULATION STUDY

In this section a simulation study has been performed for comparing the efficiency of proposed estimator with the existing estimators in literature and for this we have followed Reddy et al. (2010). The algorithm for the simulation study is as follows-

1. Generate two independent random variables X and X_1 from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively using Box-Muller method (Jhonson (1987)) of size 2000 where $\mu_1 = 5, \sigma_1 = 3$ and $\mu_2 = 5, \sigma_2 = 3$.

2. Obtain the value of Y as
$Y = \rho X + \left(\sqrt{1 - \rho^2}\right)X_1$ where $0 < \rho < 1$
For the study purpose the value of $\rho = 0.7, 0.8, 0.9$.

3. Get 2000 pairs of (Y, X)

4. From 2000 pairs of (Y, X) draw n sized simple random samples without replacement scheme $(y_i, x_i), (i = 1,2, \dots \dots \dots . . n)$.
For the study purpose the value of n=30, 50, 80.

5. Replicate the samples $(y_i, x_i), i = 1, 2, \dots \dots \dots \dots . . n$ to the no. times so that PRE (percentage relative efficiency) get stabilized.

6. Compute mean squared error (MSE) and percent relative efficiency (PRE) of the estimator-

$$MSE(t_{prop}) = E(t_{prop} - S_y^2)^2$$

$$AMSE(t_{prop}) = \frac{1}{\varphi}\sum_{j=1}^{\varphi} MSE_j, \quad (j = 1,2, \dots \dots \dots \dots . . \varphi)$$

Where φ is the number of replications and AMSE is the average mean squared error.

$$PRE = \frac{V(t_0)}{AMSE(t_{prop})} \times 100$$

Table I. Table comparing the percent relative efficiencies of proposed estimator and existing estimators

For N=2000, n=30

Estimators	PRE at $\rho = 0.7$	PRE at $\rho = 0.8$	PRE at $\rho = 0.9$
t_0	100	100	100
t_r	112.472	158.005	289.821
t_{lr}	139.956	180.75	306.329
t_{prop}	153.047	200.072	373.103
t_1	134.014	180.155	296.372
t_2	118.474	165.672	300.703
t_3	114.557	160.744	294.116
t_4	138.998	179.384	268.921
t_5	134.422	180.321	295.141
t_6	139.955	174.864	245.242
t_7	136.828	180.732	285.285
t_8	128.965	176.784	305.145
t_9	138.922	179.496	269.766

For N=2000, n=50

Estimators	PRE at $\rho = 0.7$	PRE at $\rho = 0.8$	PRE at $\rho = 0.9$
t_0	100	100	100
t_r	112.472	158.005	289.821
t_{lr}	139.956	180.75	306.329
t_{prop}	153.921	202.765	372.852
t_1	134.014	180.155	296.372

t_2	118.474	165.672	300.703
t_3	114.557	160.744	294.116
t_4	138.998	179.384	268.921
t_5	134.422	180.321	295.141
t_6	139.955	174.864	245.242
t_7	136.828	180.732	285.285
t_8	128.965	176.784	305.145
t_9	138.922	179.496	269.766

For N=2000, n=80

Estimators	**PRE at $\rho = 0.7$**	**PRE at $\rho = 0.8$**	**PRE at $\rho = 0.9$**
t_0	100	100	100
t_r	112.472	158.005	289.821
t_{lr}	139.956	180.75	306.329
t_{prop}	152.034	200.877	370.827
t_1	134.014	180.155	296.372
t_2	118.474	165.672	300.703
t_3	114.557	160.744	294.116
t_4	138.998	179.384	268.921
t_5	134.422	180.321	295.141
t_6	139.955	174.864	245.242
t_7	136.828	180.732	285.285
t_8	128.965	176.784	305.145
t_9	138.922	179.496	269.766

Conclusion

In the above table it can be seen that the proposed estimator t_{prop} works with greater efficiency than the existing estimators. So, it can be recommended to prefer the proposed estimator t_{prop} with median as an auxiliary variable and simple random sampling without replacement as a sampling scheme.

References

1) Das, A. K. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya, c*, *40*, 139-148.
2) Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, *78*(381), 117-123.
3) Johnson, M. E. (2013). *Multivariate statistical simulation: A guide to selecting and generating continuous multivariate distributions*. John Wiley & Sons.
4) Kadilar, C., & Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling. *Applied Mathematics and Computation*, *173*(2), 1047-1059.
5) Mishra, P., & Singh, R. (2017). A new log. product type estimator using auxiliary information. *Jour. Sci. Res*, *61*(179.183).
6) Prasad, B., & Singh, H. P. (1990). Some improved ratio-type estimators of finite population variance in sample surveys. *Communications in Statistics-Theory and Methods*, *19*(3), 1127-1139.
7) Reddy, M. K., Rao, K. R., & Boiroju, N. K. (2010). Comparison of ratio estimators using Monte Carlo simulation. *International Journal of Agriculture and Statistical Sciences*, *6*(2), 517-527.
8) Singh, H. P., & Solanki, R. S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statistical Papers*, *54*(2), 479-497.
9) Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2009). *Improved exponential estimator for population variance using two auxiliary variables*. Infinite Study.
10) Singh, R., & Malik, S. (2014). Improved estimation of population variance using information on auxiliary attribute in simple random sampling. *Applied mathematics and computation*, *235*, 43-49.
11) Subramani, J., & Kumarapandiyan, G. (2012a). Variance estimation using quartiles and their functions of an auxiliary variable. *Ratio*, *1*, 1.
12) Subramani, J., & Kumarapandiyan, G. (2012b). Variance estimation using median of the auxiliary variable. *Ratio*, *1*, 1.
13) Upadhyaya, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *41*(5), 627-636.

***