Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 7, Issue 2 (2014): 65-70

VARIANCE ESTIMATION IN PRESENCE OF RANDOM NON-RESPONSE

Sunil Kumar
Alliance School of Business, Alliance University, Bangalore
E Mail: sunilbhougal06@gmail.com

Received April 29, 2014
Modified November 13, 2014
Accepted November 25, 2014

Abstract

The goal of the paper is to advocate the problem of estimating the population variance of the study variance of the study variable using information on certain known parameters of the auxiliary variable in the presence of non-response. The proposed estimator is considered under two situations: i) when random non-response on study as well as auxiliary variable and population variance of auxiliary variable is known, and ii) when random non response on study variable and information on auxiliary variable is known with known population variance. Asymptotic expressions for bias and mean squared error of the proposed estimator have been obtained. Comparison of the proposed estimators with usual unbiased estimator has been carried out.

Key Words: Variance, Random Non Response, Study Variable, Auxiliary Variable.

1. Introduction

Consider a finite population $U = \{U_1, U_2, \dots, U_N\}$ of N identifiable units taking values $(y_1, y_2, \dots, y_N)$ on a study variable. The use of auxiliary variable x in survey sampling has its own eminent role. The ratio, product and regression estimator are well known examples. Following Olkin (1958), Isaki (1983) has considered the use of auxiliary variable in building up ratio and regression estimators for estimating the population variance.
Let $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(X_i - \bar{X})^2$ be the known population mean and variance of the auxiliary variable X. Assuming that a simple random sample of size n is drawn from U. Defining $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $s_y^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$. From a finite population of N identifiable units from which a simple random sample of size n is drawn without replacement. If $r(= 0, 1, 2, \dots, n-2)$ denote the number of sampling units on which information could not be obtained due to random non-response, then the remaining $(n - r)$ units in the sample can be treated as simple random sampling without replacement (SRSWOR) sample from the population U. Since, considering the problem of unbiased estimation of finite population variance, therefore assuming that r should be less than $(n-1)$, i.e. $0 \le r \le (n-r)$. Assuming that p denotes the probability of non-response among the $(n-2)$ possible values of non-response, Singh and Joarder (1998) have given the following discrete distribution given by $P(r) = \frac{n-r}{nq+2p}\binom{n-2}{r}p^r q^{n-2-r}$, where $p + q = 1; r = 0,1,2, \dots, n-2;$ and $\binom{n-2}{r}$ denotes the total number of ways of r non responses out of total possible $(n-2)$responses. It is interesting that under this distribution of random non-response the

exact bias and mean square error expressions, up to first order of approximation exists for the proposed strategies (see Singh and Joarder, 1998). Tracy and Osahan (1994), and Singh et al. (2000), Singh and Tracy (2001), Singh (2003), Singh et al. (2003), Ahmed et al. (2005), Misra et al. (2008), Singh and Solanki (2011), Singh et al. (2012), Subramani and Kumarapandiyan (2012), Yadav and Kadilar (2013) and Tiwari and Chilwal (2014) have studied the effect of random non-response on the study and auxiliary variables on several estimators of variance. In the present study, the effect of random non-response on estimating the population variance of the study variable studied in two situations:

Situation I

When random non response exists on the study as well as auxiliary variable and population variance S_x^2 of the auxiliary variable is known.

Situation II

When information on study variable could not be obtained for r units while information on auxiliary variable is available and population variance S_x^2 of the auxiliary variable is known.

Notations Define

Singh and Joarder (1998) have given the distribution of r units given by

$$P(r) = \left(\frac{n-r}{nq+2p}\right)\binom{n-2}{r} p^r q^{n-2-r}. \tag{1.1}$$

Let us define

$$s_y^{*2} = S_y^2(1+\epsilon);\ s_x^{*2} = S_x^2(1+\eta);\ s_x^2 = S_x^2(1+\omega);\ s_{yx} = S_{yx}(1+\lambda);\ s_{yx}^* = S_{yx}(1+\delta)$$

where

$S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2\ ; S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})\ ;\ \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i; \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i;$

$s_y^{*2} = \frac{1}{n-r-1}\sum_{i=1}^{n-r}(y_i - \bar{y}^*)^2\ ;\ s_x^{*2} = \frac{1}{n-r-1}\sum_{i=1}^{n-r}(x_i - \bar{x}^*)^2\ ;$

$s_{yx}^* = \frac{1}{n-r-1}\sum_{i=1}^{n-r}(y_i - \bar{y}^*)(x_i - \bar{x}^*)$are conditionally unbiased estimators of S_y^2; S_x^2 and S_{yx} respectively, and where $\bar{y}^* = \frac{1}{n-r}\sum_{i=1}^{n-r} y_i$ and $\bar{x}^* = \frac{1}{n-r}\sum_{i=1}^{n-r} x_i$.

Then, under model (1.1), one can obtain the following expectations:

$$E(\epsilon) = E(\eta) = E(\omega) = E(\lambda) = E(\delta) = 0$$

and

$$E(\epsilon^2) = \theta(\lambda_{40}-1); E(\eta^2) = \theta(\lambda_{04}-1);\ E(\omega^2) = \gamma(\lambda_{04}-1);\ E(\lambda^2) = \gamma(\lambda_{22}-1);$$

$$E(\delta^2) = \theta(\lambda_{22}-1);\ E(\epsilon\eta) = \theta(\lambda_{22}-1);\ E(\epsilon\omega) = \gamma(\lambda_{22}-1); E(\epsilon\lambda) = \gamma(\lambda_{22}-1);$$

$$E(\epsilon\delta) = \theta(\lambda_{22}-1);\ E(\eta\omega) = \gamma(\lambda_{04}-1); E(\eta\lambda) = \gamma(\lambda_{22}-1); E(\eta\delta) = \theta(\lambda_{22}-1);$$

$$E(\omega\lambda) = \gamma(\lambda_{22}-1);\ E(\omega\delta) = \gamma(\lambda_{22}-1); E(\lambda\delta) = \gamma(\lambda_{22}-1);$$

where $\hat{\beta} = S_{yx}/S_x^2\ ;\ \hat{\beta}^* = s_{yx}^*/s_x^{*2}\ ;\ K_{yx} = \frac{S_{yx}}{S_x^2 S_y^2}\ ;\ \hat{\lambda}_{ls}^* = \frac{\hat{\mu}_{ls}^*}{(\hat{\mu}_{20}^*)^{l/2}(\hat{\mu}_{20}^*)^{s/2}}\ ;$

$\hat{\mu}^*_{ls} = \frac{1}{n-r-1}\sum_{i=1}^{n-r}(y_i - \bar{y}^*)^l(x_i - \bar{x}^*)^s$; $k_{yx} = \frac{s^*_{yx}}{s^{*2}_x s^{*2}_y}$; $\theta = \left(\frac{1}{nq+2p} - \frac{1}{N}\right)$; $\gamma = \left(\frac{1}{n} - \frac{1}{N}\right)$; $\lambda_{ls} = \frac{\mu_{ls}}{\mu_{20}^{l/2}\mu_{02}^{s/2}}$ and $\mu_{ls} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^l(X_i - \bar{X})^s$ have their usual meanings.

It is to be noted if $p = 0$ i.e. if there is no non-response, the above expected values coincide with the usual results.

2. Suggested Estimators

Strategy I

When random non-response exists on both study variable y and auxiliary variable x and population variance S^2_x of the auxiliary variable x is known. Under this situation, proposed estimator of finite population variance as

$$\hat{v}_{sk} = \left\{s^{*2}_y + \hat{\beta}^*\left(S^2_y - s^{*2}_x\right)\right\}\left(\frac{S^2_x}{s^{*2}_x}\right) \qquad (2.1)$$

Thus the following theorems are stated as

Theorem 2.1: The bias in the proposed estimator $\hat{v}_{sk}$ to the first degree of approximation, is given by

$$B(\hat{v}_{sk}) = \theta S^2_y\left(1 + K_{yx}\right)(\lambda_{04} - \lambda_{22}) \qquad (2.2)$$

Proof: Expressing the proposed estimator $\hat{v}_{sk}$ in terms of ϵ, η and δ, we have

$$\left(\hat{v}_{sk} - S^2_y\right) = S^2_y\{\epsilon - \eta - \epsilon\eta + \eta^2 - K_{yx}(\eta - 2\eta^2 + \eta\delta)S^2_x\}. \qquad (2.3)$$

Taking expectations of both sides of (2.3) and using the results on the expectations from section 2, we get (2.2). Hence the theorem holds.

Theorem 2.2: The mean square error (MSE) of the proposed estimator $\hat{v}_{sk}$ to the first degree of approximation, is given by

$$MSE(\hat{v}_{sk}) = \theta S^4_y\left[\left(1 + K_{yx}S^2_x\right)\left\{(\lambda_{04} - 1)\left(1 + K_{yx}S^2_x\right) - 2(\lambda_{22} - 1)\right\} + (\lambda_{40} - 1)\right] \qquad (2.4)$$

Proof: Squaring both sides of equation (2.3) and taking expectations to the first degree of approximation, one obtains

$$MSE(\hat{v}_{sk}) = E\left(\hat{v}_{sk} - S^2_y\right)^2 = S^4_y E\left(\epsilon - \eta - \eta K_{yx}S^2_x\right)^2$$

$$MSE(\hat{v}_{sk}) = S^4_y E\left(\epsilon^2 + \eta^2 + \eta^2 K^2_{yx}S^4_x - 2\eta\epsilon - 2\eta\epsilon K_{yx}S^2_x + 2\eta^2 K_{yx}S^2_x\right)$$

$$MSE(\hat{v}_{sk}) = \theta S^4_y\left[\left(1 + K_{yx}S^2_x\right)\left\{(\lambda_{04} - 1)\left(1 + K_{yx}S^2_x\right) - 2(\lambda_{22} - 1)\right\} + (\lambda_{40} - 1)\right]$$

Hence the theorem holds.

Theorem 2.3: An estimator of the $MSE(\hat{v}_{sk})$ is given by

$$\widehat{MSE}(\hat{v}_{sk}) = \left(\frac{1}{n\hat{q}+2\hat{p}} - \frac{1}{N}\right)s^{*4}_y\left[\left(1 + k_{yx}s^{*2}_x\right)\left\{\left(\hat{\lambda}^*_{04} - 1\right)\left(1 + k_{yx}s^{*2}_x\right) - 2\left(\hat{\lambda}^*_{22} - 1\right)\right\} + \left(\hat{\lambda}^*_{40} - 1\right)\right] \qquad (2.5)$$

Strategy II

When information on study variable y could not be obtained from r units while information on auxiliary variable x is available and population variance S_x^2 of the auxiliary variable is known. The proposed estimator under such situation is as follows:

$$\hat{v}_{sk1} = \left\{s_y^{*2} + \hat{\beta}(S_x^2 - s_x^2)\right\}\left(\frac{S_x^2}{s_x^2}\right). \tag{2.6}$$

Thus, the following theorems with proofs are obvious:

Theorem 2.4: To the first degree of approximation, the bias in the estimator $\hat{v}_{sk1}$ is given by

$$B(\hat{v}_{sk1}) = -\gamma S_y^2\{(\lambda_{22} - 1) + K_{yx}(\lambda_{22} - 2\lambda_{04} + 1)S_x^2\} \tag{2.7}$$

Theorem 2.5: The mean square error, up to terms of order $O(n^{-1})$, is given by

$$MSE(\hat{v}_{sk1}) = \gamma\{(\lambda_{04} - 1)K_{yx}^2 S_x^2 - 2(\lambda_{22} - 1)\}S_y^4 + \theta(\lambda_{40} - 1)S_y^4 \tag{2.8}$$

Theorem 2.6: An estimator of mean square error of $\hat{v}_{sk1}$ is given by

$$\widehat{MSE}(\hat{v}_{sk1}) = \gamma s_y^{*4}\{(\hat{\lambda}_{04} - 1)k_{yx}^2 s_x^{*2} - 2(\lambda_{22}^* - 1)\} + \theta(\hat{\lambda}_{40}^* - 1)s_y^{*4}. \tag{2.9}$$

When information on study variable y could not be obtained for r units but information on auxiliary variable x is available and the population variance S_x^2 of the auxiliary variable is known. One can collect information on auxiliary variable while collecting information on study variable y. Using this information, another proposed estimator is given as

$$\hat{v}_{sk2} = \left\{s_y^* + \hat{\beta}^*(S_x^2 - s_x^{*2})\right\}\left(\frac{S_x^2}{s_x^{*2}}\right)\left(\frac{s_x^{*2}}{S_x^2}\right)^{\alpha}, \tag{2.10}$$

whereα is any suitable constant.

Thus the following theorems are stated as

Theorem 2.7: To the first degree of approximation, the bias in the proposed estimator $\hat{v}_{sk2}$ is given by

$$B(\hat{v}_{sk2}) = S_y^2\left[\gamma\{(\lambda_{04} - 1) - (\lambda_{22} - 1)(1 + \alpha)\} + \theta\left\{\alpha(\lambda_{22} - 1) + (\lambda_{04} - 1)\left(\frac{\alpha(\alpha+1)}{2} + (1 - \alpha)S_x^2\right)\right\}\right] \tag{2.11}$$

Theorem 2.8: The minimum mean square error of the proposed estimator $\hat{v}_{sk2}$ is given by

$$min.MSE(\hat{v}_{sk2}) = S_y^4\left[\theta(\lambda_{40} - 1) + \gamma(\lambda_{04} - 2\lambda_{22} + 1) - \frac{(\theta-\gamma)^2(1-\lambda_{22})^2}{\theta(\lambda_{04}-1)}\right] \tag{2.12}$$

Proof: We have

$$\begin{aligned} MSE(\hat{v}_{sk2}) &= E\left(\hat{v}_{sk2} - S_y^2\right)^2 = S_y^4 E(\epsilon - \omega + \alpha\eta - \eta S_x^2)^2 \\ &= S_y^4[\theta\{\lambda_{40} - 1 + (\lambda_{04} - 1)(\alpha - S_x^2)^2\} + \gamma(\lambda_{04} - 2\lambda_{22} + 1) + \\ &\quad 2(\theta - \gamma)(\lambda_{22} - 1)(\alpha - S_x^2)] \end{aligned} \tag{2.13}$$

which is minimum when

$$\alpha = \frac{(\theta-\gamma)(1-\lambda_{22})}{\theta(\lambda_{04}-1)} + S_x^2 = \alpha_{10}(say) \tag{2.14}$$

Substituting (2.14) in (2.13), one can easily obtain (2.12).

Theorem 2.9: An estimator of $min.MSE(\hat{v}_{sk2})$ is given by

$$min.MSE(\hat{v}_{sk2}) = s_y^{*4}\left[\theta\left(\hat{\lambda}_{40}^{*} - 1\right) + \gamma\left(\hat{\lambda}_{04}^{*} - 2\hat{\lambda}_{22}^{*} + 1\right) - \frac{(\theta-\gamma)^2\left(1-\hat{\lambda}_{22}^{*}\right)^2}{\theta\left(\hat{\lambda}_{04}^{*}-1\right)}\right] \quad (2.15)$$

3. Efficiency comparisons

It is well known

$$Var\left(s_y^{*2}\right) = \theta S_y^4(\lambda_{40} - 1) \quad (3.1)$$

From (2.4), (2.8), (2.12) and (3.1), we have

$Var\left(s_y^{*2}\right) - MSE(\hat{v}_{sk}) \geq 0$

if $\lambda_{22} \geq \frac{1}{2}\left\{(1 + \lambda_{04}) - (1 + \lambda_{40})K_{yx}S_x^2\right\}$ (3.2)

$Var\left(s_y^{*2}\right) - MSE(\hat{v}_{sk1}) \geq 0$

if $\lambda_{22} \geq \frac{1}{2} + (\lambda_{04} - 1)K_{yx}S_x^2$ (3.3)

$Var\left(s_y^{*2}\right) - min.MSE(\hat{v}_{sk2}) \geq 0$

if $\lambda_{22} \leq 1 - \left(\frac{1}{\theta-\gamma}\right)\{\gamma\theta(\lambda_{04} - 1)(\lambda_{04} - 2\lambda_{22} + 1)\}^{1/2}$, (3.4)

$MSE(\hat{v}_{sk}) - min.MSE(\hat{v}_{sk2}) \geq 0$

if $\left(\frac{\theta-\gamma}{\theta}\right)\left\{\frac{\theta(\lambda_{04}-\lambda_{22})^2-\gamma(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right\} + \theta\left\{2(\lambda_{04} - 2\lambda_{22} + 1) + (\lambda_{04} - 1)K_{yx}S_x^2\right\}K_{yx}S_x^2 \geq 0$ (3.5)

$$MSE(\hat{v}_{sk1}) - min.MSE(\hat{v}_{sk2}) \geq 0$$

if $\gamma\left(K_{yx}^2S_x^2 - 1\right) + \frac{1}{\theta}\left\{\frac{(\theta-\gamma)(1-\lambda_{22})}{(\lambda_{04}-1)}\right\}^2 \geq 0$ (3.6)

It follows from the above expressions that the proposed estimators $\hat{v}_{sk}$, $\hat{v}_{sk1}$ and $\hat{v}_{sk2}$ are more efficient than the usual unbiased estimator s_y^{*2}. It is further observed that the proposed estimator $\hat{v}_{sk2}$ at its optimum is more efficient than $\hat{v}_{sk}$and $\hat{v}_{sk1}$, only if (3.5) and (3.6) holds, respectively.

4. Conclusion

Auxiliary information is used in estimating finite population parameters is a cliché. In this article, I studied a well-known problem of improving estimation of the population mean under the account of random non response. The proposed estimator performs efficient in terms of MSE and should perform very well in practical surveys. Future studies should focus on developing more efficient estimators for the situations in which random non response occurs also in random non response sample, a situation that is quite possible in survey practice.

References

1. Ahmeda, M. S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2005). Estimation of finite population variance in presence of random non-response using auxiliary variables, Information and Management Sciences, 16, 2, p. 73-82.
2. Isaki, C. T. (1983). Variance estimation using auxiliary information, Journal of the American Statistical Association, 78, p. 117-123.
3. Misra, S., Yadav, S. K. and Pandey, A. (2008). Ratio type estimator of square of coefficient of variation using qualitative auxiliary information, Journal of Reliability and Statistical Studies, Vol. 1, Issue 1, p. 42-47.
4. Olkin, I. (1958). Multivariate ratio estimation for finite population, Biometrika, 45, p. 154-165.
5. Singh, H. P., Chandra, P. and Singh, S. (2003). Variance estimation using multi auxiliary information for random non-response in survey sampling, Statistica, LX III, 1, p. 23-40.
6. Singh. S. (2003). Advanced sampling theory with applications: How Michael "selected" Amy. p. 1-1194. "Kluwer Academic Publisher, The Netherlands" (In press).
7. Singh, S. and Joarder, A. H. (1998). Estimation of finite population variance using random non-response in survey sampling, Metrika, 47, p. 241-249.
8. Singh, H P and Solanki, RS (2011). Estimation of finite population mean using random non response in survey sampling, PJSOR, Vol. 7, No. 1, p. 21-41.
9. Singh, H P and Tracy, DS (2001). Estimation of population mean in presence of random non response in sample surveys, Statistica, LXI, p. 231-248.
10. Singh, S., Joarder, A H. and Tracy, D. S. (2000). Regression type estimators for random non-response in survey sampling, Statistica, LX, 1, p. 39-44.
11. Singh, H. P., Tailor,R., Kim, J. and Singh, S. (2012). Families of estimators of finite population variance using a random non-response in survey sampling, The Korean Journal of Applied Statistics, 25(4), p. 681-695.
12. Subramani, J. and Kumarapandiyan (2012). Variance estimation using median of auxiliary variable, International Journal of Probability and Statistics, 1(3), p. 62-66.
13. Tiwari, N. and Chilwal, A. (2014). Empirical comparison of various approximate estimatorsof the variance of Horwitz Thompson estimator under split method of sampling, Journal of Reliability and Statistical Studies, Vol. 7, Issue 1, p. 51-60.
14. Tracy, D. S. and Osahan, S. S. (1994). Random non-response on study variable versus on study as well as auxiliary variables, Statistica, LIV, 2, p. 163-168.
15. Yadav, S. K. and Kadilar, C. (2013). A class of ratio-cum-dual to ratio estimator of population variance, Journal of Reliability and Statistical Studies, Vol. 6, Issue 1, p. 29-34.