Original Research article

Modified Variance Estimators for non response problems in survey sampling

M. A. Bhat, T. A. Raja S. Maqbool, S.A Mir and Nageena Nazeer.

Author's contributions: This work was carried out in collaboration between all authors. 'Author M. A. Bhat' designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. 'Author T. A. Raja' and 'Author S. Maqbool' managed the analyses of the study. 'Authors Nageena Nazeer and S. H. Malik' managed the literature. All authors read and approved the final manuscript.

Abstract: In this present study, we have discussed the issue of presence and absence of non-response that we often face in survey estimation. We have suggested the estimators to estimate the finite population Variance in the absence and presence of non-response, using the linear combination of coefficient of skewness and quartiles as auxiliary information. The expression for mean square errors of suggested estimators has been derived up to the first order of approximation. The comparison of existing estimators with suggested estimators has been made through an numerical illustration to prove the efficiency of suggested estimators over existing estimators.

Key words: Simple random sampling, bias, mean square error, skewness and quartiles efficiency.

1. Introduction: The strategy of modifying the estimators through proper utilization of auxiliary information has been widely discussed by different authors in different forms when there exists a close association between auxiliary variable (X) and Study variable (Y). Several authors have proposed different estimators to improve the efficiency of estimators in absence of non-response. Some of them from the literature are as Isaki [1], who proposed ratio and regression estimators. Upadhyaya and Singh [2] used coefficient of kurtosis β_{2x} as auxiliary variable to improve the efficiency of estimator. Kadilar & Cingi, [3] utilized the coefficient of skewness C_x as auxiliary population parameter to enhance the efficiency of estimator. These efforts are now being regularly carried out by the authors to improve and develop the efficiency of estimators in absence of non-response. In presence of non-response, various authors have utilized their efforts to improve the efficiency of estimators by using different population parameters as auxiliary variables. The non-response issue has been addressed by different authors in different forms viz; M. H. Hansen and W. N Hurwtiz [4], Sarandal et al. [5] where these authors have utilized this auxiliary information to enhance the efficiency of estimators in presence of non response. Recently, Riaz et al [6], Singh et al [7] Shahzad et al [8], have also addressed the issue of non-response utilizing the auxiliary information in different ways to enhance the efficiency of estimators. Here, we consider the finite population $U = \{U_1, U_2, ..., U_N\}$, consists of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on Ui, i=1,2,3,…,N giving a vector $Y = \{y_1, y_2, ..., y_N\}$. The goal is to estimate the population mean $\bar{Y} = \frac{1}{N}\sum_{I=1}^{N} y_I$ or its variance $S_Y^2 = \frac{1}{N-1}\sum_{I=1}^{N}(y_i - \bar{y})^2$ on the basis of random sample selected from a population "U'' in the absence and presence of non-response. . In this paper our aim is to estimate the finite population variance in absence and presence of non-response, when there is presence of outliers in survey data by utilizing the linear combination of skewness and quartiles as auxiliary population parameters to improve the efficiency of estimators as quartiles are not sensitive to outliers.

2. MATERIALS AND METHODS

2.1 Notations: $N=$ Population size, $n=$ Sample size, $\gamma=\frac{1}{n}$, Y= study variable, X= Auxiliary variable. $\bar{X}$, $\bar{Y}$ = Population means. $\bar{x}$, $\bar{y}=$ Sample means. S_Y^2 , S_x^2 = Population variances. s_y^2 , s_x^2 = sample variances. C_x , C_y = . Coefficient of variation $\rho=$ coefficient of correlation, $\beta_{1(x)}=$ Skewness of auxiliary variable $\beta_{2(x)}=$ Kurtosis of the auxiliary variable, $\beta_{2(y)}=$ Kurtosis of the study variable. $Q_1=$ First quartile, $Q_2=$ second quartile $Q_3=$ third quartile $\lambda_{rs}=\frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}}$ Where $\mu_{rs}=\frac{1}{N}\sum\left(Y_i-\bar{Y}\right)^r\left(X_i-\bar{X}\right)^s$.

B (.)=Bias of the estimator and MSE (.)= Mean square error.

Section-1

Existing estimators in absence of non-response:

2.2 Ratio type variance estimator proposed by Isaki [1]

$$\hat{S}_{Is}^2=s_y^2\left[\frac{S_x^2}{s_x^2}\right] \tag{1}$$

The bias and mean square error of the estimator up to first order of approximation is given by the following expression

$$\text{Bias }(\hat{S}_{IS}^2)=\gamma S_y^2\left[\left(\beta_{2(x)}-1\right)-(\lambda_{22}-1)\right] \tag{2}$$

$$\text{MSE }((\hat{S}_{IS}^2)=\gamma S_y^4\left[\left(\beta_{2(y)}-1\right)+(\beta_{2(x)}-1)-2(\lambda_{22}-1)\right] \tag{3}$$

2.3 Ratio type variance estimator proposed by Upadhyaya and Singh [2]

$$\hat{S}_{US}^2=s_y^2\left[\frac{S_x^2+\beta_2(x)}{s_x^2+\beta_2(x)}\right] \tag{4}$$

The bias and mean square error of the estimator up to first order of approximation is give the following expression

$$\text{Bias}((\hat{S}_{US}^2)=\gamma S_y^2 A_{US}\left[A_{US}\left(\beta_{2(x)}-1\right)-(\lambda_{22}-1)\right] \tag{5}$$

$$\text{MSE}((\hat{S}_{US}^2)=\gamma S_y^4\left[\left(\beta_{2(y)}-1\right)+A_{US}^2(\beta_{2(x)}-1)-2A_{US}(\lambda_{22}-1)\right] \tag{6}$$

2.4 Ratio type variance estimator proposed by Kadilar and Cingi [3]

$$\hat{S}_{kc1}^2=s_y^2\left[\frac{S_x^2+C_x}{s_x^2+C_x}\right] \tag{7}$$

The bias and mean square error of the estimator up to first order of approximation is given by the following expression

$$\text{Bias}(\hat{S}^2_{kc1}) = \gamma S^2_y A_{kc} \left[A_1\left(\beta_{2(x)} - 1\right) - (\lambda_{22} - 1) \right] \quad (8)$$

$$\text{MSE}(\hat{S}^2_{kc1}) = \gamma S^4_y \left[\left(\beta_{2(y)} - 1\right) + A^2_{kc} \left(\beta_{2(x)} - 1\right) - 2A_{kc} (\lambda_{22} - 1) \right] \quad (9)$$

2.5. Ratio type Variance estimator proposed by Singh et al [7]

$$\hat{S}_s = S'^2_y \left[\frac{S^2_x + Q^2_1}{s^2_x + Q^2_1} \right] \quad (10)$$

The bias and mean square error of the estimator up to first order of approximation is given by the following expression

$$\text{Bias}(\hat{S}^2_s) = \gamma S^2_y A_s \left[A_s \left(\beta_{2(x)} - 1\right)' - (\lambda_{22} - 1) \right] \quad (11)$$

$$\text{MSE}(\hat{S}^2_s) = \gamma S^4_y \left[\left(\beta_{2(y)} - 1\right) + A^2_s \left(\beta_{2(x)} - 1\right) - 2A_s (\lambda_{22} - 1) \right] \quad (12)$$

3. Proposed estimators in the absence of non-response:

We have suggested, new modified ratio type variance estimators to estimate the finite population variance in absence of non-response which are given below:-

$$\hat{S}^2_{p_1} = Ks^2_y \left[\frac{S^2_x + (\beta_{1x})(Q^2_1)}{s^2_x + (\beta_{1x})(Q^2_1)} \right] \quad \hat{S}^2_{p_2} = Ks^2_y \left[\frac{S^2_x + (\beta_{1x})(Q^2_2)}{s^2_x + (\beta_{1x})(Q^2_2)} \right] \quad \hat{S}^2_{p_3} = Ks^2_y \left[\frac{S^2_x + (\beta_{1x})(Q^2_3)}{s^2_x + (\beta_{1x})(Q^2_3)} \right]$$

Where, ‘K’ is a Characterizing scalar to be determined such that the MSE of the proposed estimators is minimized. The bias and mean square error of the proposed estimators has been carried out by the following mathematical expression.

The bias and mean square error of proposed estimators up to first order of approximation has been carried out by the following expression

Let $e_0 = \dfrac{s^2_y - S^2_y}{S^2_y}$ and $e_1 = \dfrac{s^2_x - S^2_x}{S^2_x}$. Further we can write $s^2_y = S^2_y(1 + e_0)$ and $s^2_x = S^2_x(1 + e_0)$ and from the definition of e_0 and e_1 we obtain:

$$E[e_0] = E[e_1] = 0, \quad E[e^2_0] = \frac{1-f}{n}(\beta_{2(y)} - 1), \quad E[e^2_1] = \frac{1-f}{n}(\beta_{2(x)} - 1), \quad E[e_0 e_1] = \frac{1-f}{n}(\lambda_{22} - 1)$$

The proposed estimator can be written as $\hat{R}_{MS} = Ks_y^2(1+e_0)(1+R_i e_1)^{-1}$ (1)

Expanding the right hand side of above equation up to the first order approximation we get

$$\hat{R}_{MS} = Ks_y^2(1+e_0 - R_i e_1 - R_i e_0 e_1 + R_i^2 e_1^2) \quad (2)$$

After subtracting the population variance S_y^2 of study variable on both sides of above equation we get

$$\hat{R}_{MS} - S_y^2 = Ks_y^2(1+e_0 - R_i e_1 - R_i e_0 e_1 + R_i^2 e_1^2) - S_y^2 \quad (3)$$

By taking expectations on both sides of above equation, we get the bias of the proposed estimators

$$\text{Bias}= \gamma K S_y^2 \left[R_i^2(\beta_{2y}-1) - R_i(\lambda_{22-1})\right] + S_y^2(K-1) \quad (4)$$

The mean square error is obtained by squaring both sides of equation (1) and taking expectations on both sides up to first order of approximation as:-

$$\text{MSE}= S_y^4\left[K^2\gamma(\beta_{2y}-1) + (3K^2-2K)R_i^2(\beta_{2x}-1) - 2(2K^2-K)R_i\gamma(\lambda_{22}-1) + (K-1)^2\right] \quad (5)$$

Where $\gamma = \dfrac{1-f}{n}$, MSE is minimum for $K= \dfrac{1+R_i^2\gamma(\beta_{2x}-1) - R_i\gamma(\lambda_{22}-1)}{1+\gamma(\beta_{2y}-1) + 3R_i^2\gamma(\beta_{2x}-1) - 4R_i\gamma(\lambda_{22}-1)}$

The Minimum MSE for the estimator, optimum value of K is:

$$\text{MSE } MSE_{\min.}\hat{R}_{MS} = S_y^4\left[1 - \frac{\{(1+R_i^2\gamma(\beta_{2x}-1) - R_i\gamma(\lambda_{22}-1))^2\}}{1+\gamma(\beta_{2y}-1) + 3R_i^2\gamma(\beta_{2x}-1) - 4R_i\gamma(\lambda_{22}-1)}\right]$$

4. Numerical Illustration: We use the data set presented in Sarandal et al (1992) concerning (P85) 1985 population considered as Y and RMT85 revenue from 1985 municipal taxation in millions of kronor considered as X. Descriptive statistics is given below.

Sarandal population [5]

$$N = 234, n = 35, \bar{Y} = 29.3626, \bar{X} = 245.088, S_y = 51.556, S_x = 596.332, \rho = 0.96, \beta_{2y} = 89.231, \beta_{2x} = 89.189$$
$$\lambda_{22} = 4.041, \beta_{1x} = 8.83, \beta_{1y} = 8.27, TM = 167.4, Q_1 = 67.75, Q_2 = 113.5, Q_3 = 230.25, C_x = 2.43, D_1 = 49.0,$$
$$D_2 = 63.0, D_3 = 75.0, D_4 = 90.0, D_5 = 113.5, D_6 = 145.9, D_7 = 197.9, D_8 = 271.1, D_9 = 467.5, D_{10} = 6720.0$$

We consider 20% weight for non-response (missing values) and have considered last 47 values as non-respondents results are as follows:-

$l = 2, S_{y2}^2 = 2.9167, \beta_2(y_2) = 11.775, N_2 = 47$. We apply the proposed and existing estimators to this data set and the data statistics is given below:

Table-1 Mean square errors of existing estimators in absence of non response

Existing Estimators	Mean square error
Isaki [1]	29216846.227
Upadhyaya and Singh [2]	29187686.03
Kadilar and Cingi [3]	29216846.224
Singh et al [7]	28839478.44

Table-2 Mean square errors of proposed estimators in absence of non response

Proposed Estimators	Mean square error
$\hat{S}_{p_1}^2 = Ks_y^2\left[\frac{S_x^2 + (\beta_{1x})(Q_1^2)}{s_x^2 + (\beta_{1x})(Q_1^2)}\right]$	2690922.57
$\hat{S}_{p_2}^2 = Ks_y^2\left[\frac{S_x^2 + (\beta_{1x})(Q_2^2)}{s_x^2 + (\beta_{1x})(Q_2^2)}\right]$	2817688.72
$\hat{S}_{p_3}^2 = Ks_y^2\left[\frac{S_x^2 + (\beta_{1x})(Q_3^2)}{s_x^2 + (\beta_{1x})(Q_3^2)}\right]$	3930301.6

Table-3 Percent relative efficiency of existing estimators with proposed estimators in absence of non-response

Existing Estimators $\rightarrow$ Pr*oposed Estimators* $\downarrow$	Isaki [1]	Upadhyaya and Singh [2]	Kadilar and Cingi [3]	Singh et al [7]
$\hat{S}_{p_1}^2 = Ks_y^2\left[\frac{S_x^2 + (\beta_{1x})(Q_1^2)}{s_x^2 + (\beta_{1x})(Q_1^2)}\right]$	1085.75	1084.67	1085.75	1071.73
$\hat{S}_{p_2}^2 = Ks_y^2\left[\frac{S_x^2 + (\beta_{1x})(Q_2^2)}{s_x^2 + (\beta_{1x})(Q_2^2)}\right]$	1036.90	1035.87	1036.90	1023.51
$\hat{S}_{p_3}^2 = Ks_y^2\left[\frac{S_x^2 + (\beta_{1x})(Q_3^2)}{s_x^2 + (\beta_{1x})(Q_3^2)}\right]$	743.37	742.63	743.37	733.77

Section-2

5. Existing estimators in presence of non response

Hansen and Hurwitz [4] sub sampling scheme is the most popular scheme, used for the Non-response problems let us consider a finite population consisting of *N* units. Let *y* be the character under study and a simple random sample of size *n* is drawn without replacement, of which n_1 units respond and n_2 units do not respond. From the n_2 non-respondents we select a sample of size $r = \frac{n_2}{k}, (k \geq 1)$ where *k* is the inverse sampling rate at the second phase sample of size *n* (fixed in advance) and from whom we collect the required information. It is assumed here that all the *r* units respond fully this time.

Let N_1 and $N_2 = N - N_1$ be the sizes of the responding and non-responding units from the finite population $N; W_1 = \frac{N_1}{N}, W_2 = \frac{N_2}{N}$ are the corresponding weights

Hansen and Hurwitz (1946) unbiased estimator under non-response is given by

$Var(\hat{T}') = S_y^4(\beta_{2y} - 1) + WS_{y2}^4\beta_2(y_2)^*$ where W= $\frac{N_2(l-1)}{nN}$, Where, l= sampling inverse rate

5.1 Ratio type variance estimator proposed by Isaki [1]

$$\hat{S}_{Is}^2 = s_y'^2\left[\frac{S_x^2}{s_x^2}\right] \quad (1)$$

The bias and mean square error of the estimator up to first order of approximation is given by

the following expression

$$\text{Bias } (\hat{S}_{IS}^2) = \gamma S_y^2\left[\left(\beta_{2(x)} - 1\right)' - (\lambda_{22} - 1)\right] \quad (2)$$

$$\text{MSE}((\hat{S}_{IS}^2) = \gamma S_y^4\left[\left(\beta_{2(y)} - 1\right)' + (\beta_{2(x)} - 1) - 2(\lambda_{22} - 1)\right] \quad (3)$$

Where, $(\beta_{2y} - 1)' = \left(\beta_{2y} - 1\right) + \frac{WS_{y2}^4(\beta_{2y})^*}{S_y^4} = \frac{Var(\hat{T}')}{S_y^4}$

5.2 Ratio type variance estimator proposed by Upadhyaya and Singh [2]

$$\hat{S}_{US}^2 = s_y'^2\left[\frac{S_x^2 + \beta_2(x)}{s_x^2 + \beta_2(x)}\right] \quad (4)$$

The bias and mean square error of the estimator up to first order of approximation is give the following expression

$$\text{Bias}((\hat{S}^2_{US}) = \gamma S^2_y A_{US} \left[A_{US}\left(\beta_{2(x)} - 1\right)' - (\lambda_{22} - 1) \right] \quad (5)$$

$$\text{MSE}((\hat{S}^2_{US}) = \gamma S^4_y \left[\left(\beta_{2(y)} - 1\right)' + A^2_{US} \ (\beta_{2(x)} - 1) - \ 2A_{US} \ (\lambda_{22} - 1) \right] \quad (6)$$

Where $(\beta_{2y} - 1)' = \left(\beta_{2y} - 1\right) + \dfrac{WS^4_{y2}(\beta_{2y})^*}{S^4_y} = \dfrac{Var(\hat{T}')}{S^4_y}$

5.3 Ratio type variance estimator proposed by Singh [7]

$$\hat{S}^2_s = s'^2_y \left[\frac{S^2_x + Q^2_1}{s^2_x + Q^2_1} \right] \quad (7)$$

The bias and mean square error of the estimator up to first order of approximation is given by the following expression

$$\text{Bias}((\hat{S}^2_s) = \gamma S^2_y A_1 \left[A_s\left(\beta_{2(x)} - 1\right)' - (\lambda_{22} - 1) \right] \quad (8)$$

$$\text{MSE}((\hat{S}^2_s) = \gamma S^4_y \left[\left(\beta_{2(y)} - 1\right)' + A^2_s \ (\beta_{2(x)} - 1) - \ 2A_s \ (\lambda_{22} - 1) \right] \quad (9)$$

Where $(\beta_{2y} - 1)' = \left(\beta_{2y} - 1\right) + \dfrac{WS^4_{y2}(\beta_{2y})^*}{S^4_y} = \dfrac{Var(\hat{T}')}{S^4_y}$

6**. Proposed Estimators in presence of non-response**

$$\hat{S}^2_{p_1} = Ks'^2_y \left[\frac{S^2_x + (\beta_{1x})(Q^2_1)}{s^2_x + (\beta_{1x})(Q^2_1)} \right] \quad \hat{S}^2_{p_2} = Ks'^2_y \left[\frac{S^2_x + (\beta_{1x})(Q^2_2)}{s^2_x + (\beta_{1x})(Q^2_2)} \right] \quad \hat{S}^2_{p_3} = Ks'^2_y \left[\frac{S^2_x + (\beta_{1x})(Q^2_3)}{s^2_x + (\beta_{1x})(Q^2_3)} \right],$$

where ‘K’ is a Characterizing scalar to be determined such that the MSE of the proposed estimators is minimized. The bias and mean square error of the proposed estimators has been carried out by the following mathematical expression.

Table-4 Mean square errors of existing estimators in presence of non response

Existing Estimators	Mean square error
Isaki [1]	29215130.96
Upadhyaya and Singh [2]	29185970.72
Singh et al [7]	28837763.13

Table-5 Mean square errors of proposed estimators in presence of non response

Proposed Estimators	Mean square error
$\hat{S}^2_{p_1} = Ks'^2_y\left[\frac{S^2_x+(\beta_{1x})(Q^2_1)}{s^2_x+(\beta_{1x})(Q^2_1)}\right]$	22343807.45
$\hat{S}^2_{p_2} = Ks'^2_y\left[\frac{S^2_x+(\beta_{1x})(Q^2_2)}{s^2_x+(\beta_{1x})(Q^2_2)}\right]$	2012750.82
$\hat{S}^2_{p_3} = Ks'^2_y\left[\frac{S^2_x+(\beta_{1x})(Q^2_3)}{s^2_x+(\beta_{1x})(Q^2_3)}\right]$	3855736.26

Table-6 Percent relative efficiency of existing estimators with proposed estimators in presence of non-response

Existing Estimators $\rightarrow$ Pr*oposed Estimators* $\downarrow$	Isaki [1]	Upadhyaya and Singh [2]	Singh et al [7]
$\hat{S}^2_{p_1} = Ks'^2_y\left[\frac{S^2_x+(\beta_{1x})(Q^2_1)}{s^2_x+(\beta_{1x})(Q^2_1)}\right]$	1246.48	1245.23	1230.38
$\hat{S}^2_{p_2} = Ks'^2_y\left[\frac{S^2_x+(\beta_{1x})(Q^2_2)}{s^2_x+(\beta_{1x})(Q^2_2)}\right]$	1451.50	1450.05	1432.75
$\hat{S}^2_{p_3} = Ks^2_y\left[\frac{S^2_x+(\beta_{1x})(Q^2_3)}{s^2_x+(\beta_{1x})(Q^2_3)}\right]$	757.70	756.94	731.68

7. Conclusion: In this manuscript, suggested estimators for the estimation of finite population variance in absence and presence of non-response clearly shows from the tables viz, table-1, table-2, table-3, table-4, table-5 and table-6 that the suggested estimators are more efficient than the mentioned existing estimators. Hence suggested estimators may be preferred over existing estimators.

Literature cited:

[1] Isaki, C.T. (1983) Variance estimation using auxiliary information. *Journal of the American Statistical Association, 78,117-123*.

[2]*Upadhyaya, L. N. and Singh, H. P, (1999) : use of auxiliary variable in the estimation of population variance, Mathematical forum, 4, 33-36 (1936).*

[3] Kadilar, C. & Cingi, H.(2006a).Improvement in Variance estimation using auxiliary information, *Hacettepe Journal of mathematics and Statistics, 35(1).117-115.*

[4] M.H. Hansen and W.N. Hurtiz. The problem of non response in sample surveys. Journal of the American Statistical Association, 41:236 (1946) 517-529.

[5] C. E. Sarandal, B. Swenson and. Wretman, Model Assisted Survey Sampling, New York: Springer (1992).

[6] S. Riaz. Diana and j. Shabbir. A general class of estimators for the population means using multiphase sampling with the non-respondents. *Hacettepe journal of Mathematics Statistics, 43:39(2014) 511-527*

[7] B. K..Singh, W. W. Chanu and Y. Tato. *A class of ratio cum dual to ratio estimator in presence of non-response using two phase sampling scheme. GUJ Sci, 29: 1 (2016)213-221.*

[8].Shahzad & Hanif, M & Koyuncu, Nursel & Luengo, A.V.G.. (2017). Estimation of population variance using quartiles in presence and absence of non-response. Gazi university *Journal of science 30.205-218(2017).*

[9] M .a . Bhat[*1], S . Maqbool[2], S .A .Saraf[2], Ab. Rouf[2]and S .H Malik[2]. *Journal of Advances in Research 13(2): 1-6 , 2018 Article no.37321. ISSN: 2348-0394 , NLMID: 101666096*

[10].Shahzad & Hanif, M & Koyuncu, Nursel & Luengo, A.V.G.. (2017*). Estimation of population mean under systematic random sampling in absence and presence of non-response. Istatistikciler Dergisi: Istatistik &Aktuerya,2017,1,23-39*