Original Research Article

Efficient Handling of Non-Response in Sample Surveys: Development of Ratio-Type Variance Estimators

Abstract: The present study has been under taken to address the issue of presence and absence of non-response problems in Sample surveys. We have developed some new efficient ratio-type variance estimators for the computation of the finite population variance in presence and absence of non-response, by incorporating Positional parameters as auxiliary information. The expression for mean square errors of proposed estimators has been derived up to the first order of approximation. To test the efficiency of new developed estimators, practical demonstration has been carried to ascertain the performance of suggested Estimators.

Keywords: Bias, $MSE_{mim.}$ Sample Random Sampling, Quartiles, Deciles, Trimmed Meanand efficiency.

Introduction: "The problem of non-response from respondents in survey data is one of the biggest problems while ascertaining the information from the source. The most popular Sub-sampling scheme introduced by Hansen and Hurwitz in 1946 is widely used to handle the non-response problems in survey data. In our present study, we will adapt this sub-sampling technique to handle the non-response problems in survey data. The technique is based on two-way stratification, one for respondent group and the other for non-respondent group in phase first of survey. Then in phase-II, by making intensive efforts, a sub sample from non-respondent strata is drawn to recover the non-response. The strategy of developing efficient estimators through proper utilization of supplementary information has been widely addressed by different authors, when the strong association exists between the auxiliary variable and study variable" [11]. To enhance the efficiency of estimators in presence and absence of non-response by utilizing auxiliary information, different authors have developed different Estimators like as Isaki [1] who has introduced ratio and regression estimators. Similarly, Upadhaya and Singh [2] have used "coefficient of Kurtosis as supplementary information to improve the efficiency of Estimators". Kadilar and Cingi[3] used "coefficient of skewness. On the same lines, several authors have contributed their research efforts to improve the efficiency of estimators". From the present literature, the issue of non response problems in survey sampling is being regularly addressed by the authors like as M.H.Hansen and W.N Hurwitz [3], Sarandal et el[4], Riaz et el[5], Singh et el[6] and Shahzad et el[7].

1. Review of estimators in literature:

Let the finite population under survey be $Z=\{Z_1, Z_2, ... Z_N\}$, which consists of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on $Z_i, i=1,2,3......N$, giving a vector $Y=\{y_1, y_2, ... y_N\}$. The aim of our effort is to estimate the populations mean $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N} y_i$ and 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 afinite population Z.

2. Notations:

$N = population\ size,\ n = sample\ size,$

$\gamma=\frac{N-n}{n},\ Y = Study\ variable,\ X = Auxiliary\ variable,\ \bar{X}\ and\ \bar{Y} = Population\ mean$

$\bar{x}\ and\ \bar{y} = Sample\ mean,\ S_Y^2\ and\ S_x^2 = Population\ variance$

$s_y^2\ and\ s_x^2 = Sample\ variance \quad C_x\ and\ C_y = Cofficient\ of\ variation$

$\rho = Cofficient\ of\ Correlation,\ \beta_{2(x)} = Cofficient\ of\ kurtosis$

$\beta_{2(y)} = Cofficient\ of\ kurtosis$

$HL = \text{Hodges Lehmann}, MR = \text{Midrange}, TM = \text{Trimean}, Q_3 = \text{Third Quartile}$

$Q_r = \text{Quartile range}, Q_d = \text{Quartile deviation}, \lambda_{rs} = \dfrac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}}$

$\mu_{rs} = \dfrac{1}{N-1}\sum_{i=1}^{N}\left(Y-\bar{Y}\right)^r\left(X-\bar{X}\right)^s = \text{Covariance between study variable and auxiliary variable}$

3. Existing Estimators in absence of non-response from the literature

3.1 Ratio type Variance estimator proposed by Isaki [1]:

$$\hat{S}_{R_1}^2 = s_y^2\frac{S_x^2}{s_x^2}$$

Bias ($\left(\hat{S}_{R_1}^2\right) = \gamma S_y^2\left[\left(\beta_{2(x)}-1\right)-\left(\lambda_{22}-1\right)\right]$)

MSE ($\left(\hat{S}_{R_1}^2\right) = \gamma S_y^4\left[\left(\beta_{2(y)}-1\right)+\left(\beta_{2(x)}-1\right)-2\left(\lambda_{22}-1\right)\right]$)

3.2 Ratio type Variance estimator proposed by Upadhaya and Singh[2]:

$$\hat{S}_{R_2}^2 = s_y^2\left[\frac{S_x^2+\beta_{2x}}{s_x^2+\beta_{2x}}\right]$$

Bias ($\left(\hat{S}_{R_2}^2\right) = \gamma S_y^2\theta_2\left[\theta_2\left(\beta_{2(x)}-1\right)-\left(\lambda_{22}-1\right)\right]$)

MSE ($\left(\hat{S}_{R_2}^2\right) = \gamma S_y^4\left[\left(\beta_{2(y)}-1\right)+\theta_2^2\left(\beta_{2(x)}-1\right)-2\theta_2\left(\lambda_{22}-1\right)\right]$)

3.3 Ratio type Variance estimator proposed by Kadilar and Cingi[3]:

$$\hat{S}_{R_3}^2 = s_y^2\left[\frac{S_x^2+C_x}{s_x^2+C_x}\right]$$

Bias ($\left(\hat{S}_{R_3}^2\right) = \gamma S_y^2\tau_3\left[\tau_3\left(\beta_{2(x)}-1\right)-\left(\lambda_{22}-1\right)\right]$)

MSE ($\left(\hat{S}_{R_3}^2\right) = \gamma S_y^4\left[\left(\beta_{2(y)}-1\right)+\tau_3^2\left(\beta_{2(x)}-1\right)-2\tau_3\left(\lambda_{22}-1\right)\right]$)

3.4 Ratio type Variance estimator proposed by Singh et al[8]:

$$\hat{S}_{R_4}^2 = s_y^2\left[\frac{S_x^2+Q_d^2}{s_x^2+Q_d^2}\right]$$

Bias ($\left(\hat{S}_{R_4}^2\right) = \gamma S_y^2\omega_4\left[\omega_4\left(\beta_{2(x)}-1\right)-\left(\lambda_{22}-1\right)\right]$)

MSE ($\left(\hat{S}_{R_4}^2\right) = \gamma S_y^4\left[\left(\beta_{2(y)}-1\right)+\omega_4^2\left(\beta_{2(x)}-1\right)-2\omega_4\left(\lambda_{22}-1\right)\right]$)

Remark: "θ_2, τ_3 and ω_4" are Constants

'R_i' Stands for ratio estimator and $i = 1,2,3,4$

4. Proposed estimators

$$1.\ \hat{S}^2_{R_1} = K s_y^2 \left[\frac{S_x^2 + (\beta_{1x})\times(D_3)^2}{s_x^2 + (\beta_{1x})\times(D_3)^2}\right]$$

$$2.\ \hat{S}^2_{R_1} = K s_y^2 \left[\frac{S_x^2 + (\beta_{1x})\times(D_4)^2}{s_x^2 + (\beta_{1x})\times(D_4)^2}\right]$$

$$3.\ \hat{S}^2_{R_1} = K s_y^2 \left[\frac{S_x^2 + (\beta_{1x})\times(D_5)^2}{s_x^2 + (\beta_{1x})\times(D_5)^2}\right]$$

Where 'K' is Characterizing scalar to be determined such that the MSE of the proposed estimators is minimized.

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

Let $e_0 = \frac{s_y^2 - S_y^2}{S_y^2}$ and $e_1 = \frac{s_x^2 - S_x^2}{S_x^2}$. Further we can write $s_y^2 = S_y^2(1+e_0)$ and $s_x^2 = S_x^2(1+e_1)$ and from the definition of e_0 and e_1 we obtain:

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

The proposed estimator can be written as:

$$\hat{R}_{MS} = K s_y^2 (1+e_0)(1+R_i e_1)^{-1} \qquad (4.2)$$

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

$$\hat{R}_{MS} = K s_y^2 (1 + e_0 - R_i e_1 - R_i e_0 e_1 + R_i^2 e_1^2) \qquad (4.3)$$

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

$$\hat{R}_{MS} - S_y^2 = K s_y^2 (1 + e_0 - R_i e_1 - R_i e_0 e_1 + R_i^2 e_1^2) - S_y^2 \qquad (4.4)$$

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

$$Bias = \gamma K S_y^2 R_i \left[R_i(\beta_{2x} - 1) - R_i(\lambda_{22} - 1)\right] + S_y^2 (K-1) \qquad (4.5)$$

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:-

$$MSE = S_y^4 \left[K^2 \gamma(\beta_{2y} - 1) + (3K^2 - 2k) R_i^2 (\beta_{2x} - 1) - 2(2K^2 - K) R_i \gamma(\lambda_{22} - 1) + (K-1)^2\right] \qquad (4.6)$$

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

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

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

$$Where, (\beta_{2y}-1)'=(\beta_{2y}-1)' \tag{4.7}$$

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:

$$N=234, n=35, \bar{Y}=29.3626, \bar{X}=245.088, S_y=51.556, S_x=596.332, \rho=0.96, \beta_{2y}=89.231, \beta_{2x}=89.18$$
$$\lambda_{22}=4.041, \beta_{1x}=8.83, \beta_{1y}=8.27, TM=167.4, Q_1=67.75, Q_2=113.5, Q_3=230.25, C_x=2.43, D_1=49.0,$$
$$D_2=63.0, D_3=75.0, D_4=90.0, D_5=113.5, D_6=145.9, D_7=197.9, D_8=271.1, D_9=467.5, D_{10}=672.0$$

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

$$l=2, S_{y2}^2=2.9167, \beta_2(y_2)=11.775, N_2=47.$$

Table-1: Mean square error of existing estimators and proposed estimators in absence of non-response

Existing Estimators	Mean square error
Isaki [1]	29216846.22
Upadhyaya and Singh [2]	29187686.03
Kadilar and Cingi [3]	29216846.22
Singh *et al.* [4]	28839478.44
Proposed [1]	993482.49
Proposed[2]	1368635.80
Proposed [3]	1960434.83

Table-2: Percent relative efficiency of existing and proposed estimators in absence of non-response

Existing estimators → Proposed estimators↓	Isaki [1]	Upadhyaya and Singh [2]	Singh *et al.*[3]
Proposed [1]	2940 %	2937 %	2902 %
Proposed[2]	2134 %	2132 %	2107 %
Proposed [3]	1490 %	1488 %	4171 %

5.0 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]

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

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

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

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

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

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

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

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

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

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

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

5.3 Ratio type variance estimator proposed by Singh [7]

$$\hat{S}_s^2 = s_y'^2 \left[\frac{S_x^2 + Q_1^2}{s_x^2 + Q_1^2} \right] \quad (7)$$

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

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

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

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

6. PROPOSED ESTIMATORS IN PRESENCE OF NON-RESPONSE

, $1.\ \hat{S}_{R_1}^2 = K s_y^{*2} \left[\frac{S_x^2 + (\beta_{1x}) \times (D_3)^2}{s_x^2 + (\beta_{1x}) \times (D_3)^2} \right]$, $2.\ \hat{S}_{R_1}^2 = K s_y^{*2} \left[\frac{S_x^2 + (\beta_{1x}) \times (D_4)^2}{s_x^2 + (\beta_{1x}) \times (D_4)^2} \right]$,

$3.\ \hat{S}_{R_1}^2 = K s_y^{*2} \left[\frac{S_x^2 + (\beta_{1x}) \times (D_5)^2}{s_x^2 + (\beta_{1x}) \times (D_5)^2} \right]$

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

Table-3: 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-4: Mean square errors of proposed estimators in presence of non response

Proposed estimators	Mean square error
$\hat{S}_{p_1}^2 = K s_y'^2 \left[\frac{S_x^2 + (\beta_{1x})(D_3)^2}{s_x^2 + (\beta_{1x})(D_3)^2} \right]$	661069.84
$\hat{S}_{p_1}^2 = K s_y'^2 \left[\frac{S_x^2 + (\beta_{1x})(D_4)^2}{s_x^2 + (\beta_{x})(D_4)^2} \right]$	1083247.49
$\hat{S}_{p_1}^2 = K s_y'^2 \left[\frac{S_x^2 + (\beta_{1x})(D_5)^2}{s_x^2 + (\beta_{1x})(D_5)^2} \right]$	2497795.24

Table-5: Percent relative efficiency of existing estimators with proposed estimators in presence of non-response

Existing Estimator → *Proposed Estimator* ↓	**Isaki [1]**	**Upadhyaya and Singh [2]**	**Singh et al [7]**
$\hat{S}^2_{p1} = K s_y'^2 \left[\frac{S_x^2 + (\beta_{1x})(D_3)^2}{s_x^2 + (\beta_{1x})(D_3)^2}\right]$	4419.37 %	4414.96 %	4362.28 %
$\hat{S}^2_{p1} = K s_y'^2 \left[\frac{S_x^2 + (\beta_{1x})(D_4)^2}{s_x^2 + (\beta_{1x})(D_4)^2}\right]$	2696.99 %	2694.30 %	2662.15 %
$\hat{S}^2_{p1} = K s_y'^2 \left[\frac{S_x^2 + (\beta_{1x})(D_5)^2}{s_x^2 + (\beta_{1x})(D_5)^2}\right]$	1169.63 %	1168.46 %	1154,52 %

7. CONCLUSION

In this manuscript, suggested estimators for the estimation of finite population variance in absence and presence of non-response have clearly determined that proposed Estimators have shown excellent performance over the Existing Estimators, which can be easily accessed from the tables viz, Table-1, Table-2, Table-3, Table-4, Table-5. Hence suggested estimators may be preferred over existing estimators.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

REFERENCES

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. IstatistikcilerDergisi: Istatistik&Aktuerya,2017,1,23-39
11. Bhat M, Raja T, Maqbool S, Mir S, Nazeer N. Modified variance estimators for non response problems in survey sampling. Advances in Research. 2018 Mar 12;14(2):1-7.