ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

ESTIMATION OF POPULATION MEAN USING AUXILIARY INFORMATION IN PRESENCE OF MEASUREMENT ERRORS

Sheela Misra[1], Dharmendra Kumar Yadav[*2] & Dipika[3]
[1,2&3]Department of Statistics, University of Lucknow, Lucknow-226007

DOI: 10.5281/zenodo.817860

ABSTRACT

The present article concerns with the estimation of finite population mean using auxiliary information in the presence of measurement errors. The expressions for the bias and mean squared error (MSE) of the proposed estimator are obtained up to first order of approximation. A theoretical efficiency comparison between the proposed estimator and usual linear regression estimator under measurement errors has been made. Theoretical results are supported by the simulation study using R software.

KEYWORDS: Estimation, auxiliary variable, measurement errors, simulation.

I. INTRODUCTION

Generally in statistical analysis it is assumed that observations are recorded without any error. However, in practice,this assumption may not be true and the data may be highly contaminated with measurement errors due to various reasons like interviewers, respondents,

Questionnaires or combinations of all these factors,[Cochran (1968), Sukhatme and Seth (1952) and Biemer et.al. (1991)]. When the observations are influenced by measurement errors,the estimates of population parameters (Mean, Variance, Total etc.) are not quite reliable and efficient thus provide misleading conclusions. Therefore the study of consequences of measurement errors is essential for the improved estimation of population parameters. Measurement errors are generally taken as the discrepancy between true and observed values on any desirable characteristic under study. The problem of estimating population mean (μ_Y) and variance (σ_Y^2) in presence of measurement errors have been dealt by various authors such as Shalabh (1997), Maneesha and Singh (2002),) Singh and Karpe (2009), Misra and Yadav (2015), Misra et al.(2016 a, 2016 b) and Misra et al. (2017).In the presentarticle we are dealing with the estimation of finite population mean in the presence of measurement errors.

II. NOTATIONS AND METHODOLOGY

Let us consider a population of Ndistinct and identifiable units. A sample of size n is drawn using simple random sampling without replacement technique. Let (x_i, y_i) be observed values instead of the true values (X_i,Y_i) on the auxiliary and main characteristics (X, Y) respectively for the i^{th} (i =1, 2, …, n) unit in the sample of size n.Let the measurement errors be,

$$u_i = y_i - Y_i$$
$$v_i = x_i - X_i$$

which are random in nature with mean zero and variances σ_u^2 and σ_v^2 respectively, and are independent. Further, let the population means of (x,y) be (μ_x,μ_y), population variances of (x, y) be (σ_X^2, σ_Y^2), σ_{XY} and ρ be the population covariance and the population correlation coefficient between x and y respectively.

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be the unbiased estimators of population means μ_X and μ_Y respectively. We note that in presence of measurement errors, $s_x^2 = \frac{1}{n-1}\sum_{i=n}^{n}(x_i - \bar{x})^2$ and $s_y^2 = \frac{1}{n-1}\sum_{i=n}^{n}(y_i - \bar{y})^2$ are not unbiased estimators of the population variances σ_X^2 and σ_Y^2.

In presence of measurement errors the expected value of s_y^2 and s_x^2 is given by
$E(s_y^2) = \sigma_Y^2 + \sigma_u^2$ and $E(s_x^2) = \sigma_x^2 + \sigma_v^2$.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Let error variances σ_u^2 and σ_v^2 are known aprior than unbiased estimators of population variance in presence of measurement errors are

$\hat{\sigma}_y^2 = s_y^2 - \sigma_u^2 > 0$, $\hat{\sigma}_x^2 = s_x^2 - \sigma_v^2 > 0$

Further we consider the following approximations-

$$\bar{y} = \mu_Y(1+e_0), \qquad \bar{x} = \mu_X(1+e_1)$$

$$\hat{\sigma}_y^2 = \sigma_Y^2(1+e_2), \qquad \hat{\sigma}_x^2 = \sigma_X^2(1+e_3)$$

$$\hat{\sigma}_{xy} = \sigma_{XY}(1+e_4)$$

Such that $E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$

From Singh and Karpe (2009), we have

$E(e_0^2) = \frac{C_Y^2}{n\theta_Y}, C_Y = \frac{\sigma_Y}{\mu_Y}, C_X = \frac{\sigma_X}{\mu_X}, \theta_X = \frac{\sigma_X^2}{\sigma_X^2+\sigma_V^2}, \theta_Y = \frac{\sigma_Y^2}{\sigma_Y^2+\sigma_u^2} E(e_1^2) = \frac{C_X^2}{n\theta_X}, E(e_2^2) = \frac{A_Y}{n}$ and $E(e_3^2) = \frac{A_X}{n}$

where ,$A_Y = \gamma_{2Y} + \gamma_{2u}\frac{\sigma_u^4}{\sigma_Y^4} + 2\left(1+\frac{\sigma_u^2}{\sigma_Y^2}\right)^2, A_X = \gamma_{2X} + \gamma_{2v}\frac{\sigma_v^4}{\sigma_X^4} + 2\left(1+\frac{\sigma_v^2}{\sigma_X^2}\right)^2$

$$E(e_0e_1) = \rho\frac{C_XC_Y}{n}, E(e_1e_2) = \frac{\mu_{1200}}{n\sigma_Y^2\mu_X}, E(e_1e_3) = \frac{\mu_{3000}}{n\sigma_X^2\mu_X}, E(e_0e_2) = \frac{\mu_{0300}}{n\sigma_Y^2\mu_Y}, E(e_1e_4) = \frac{\mu_{2100}}{n\sigma_{XY}\mu_X}$$

where$\mu_{pqrs} = E(X-\mu_X)^p(Y-\mu_Y)^q u^r v^s$

III. ESTIMATION OF POPULATION MEAN IN PRESENCE OF MEASUREMENT ERRORS

We suggest a regression- type estimator for estimating population mean and study its performance in presence of measurement errors.

$$\bar{y}_k = \bar{y} + b(\bar{X}-\bar{x}) + k(\bar{y}^3\frac{C_Y^2}{s_y^2} - \bar{y}) \qquad (1)$$

where$b = \frac{s_{xy}}{s_x^2}$ = regression coefficient

and k be the characterizing scalar to be chosen suitably.

In presence of measurement errors (4) can be written as

$$\bar{y}_k = \bar{y} + b(\mu_X-\bar{x}) + k(\bar{y}^3\frac{C_Y^2}{\hat{\sigma}_Y^2} - \bar{y}) \qquad (2)$$

Expressing (2) in terms of e_i's

$$\bar{y}_k = \mu_Y(1+e_0) + \frac{\sigma_{XY}(1+e_4)}{\sigma_X^2(1+e_3)}(\mu_X - \mu_X - \mu_Xe_1) + k\left\{\mu_Y^3(1+e_0)^3\frac{C_Y^2}{\sigma_Y^2(1+e_2)} - \mu_Y(1+e_0)\right\}$$

$$= \mu_Y[1 + e_0 - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X(e_1 + e_1e_4 - e_3e_1) + k(2e_0 - e_2 + 3e_0^2 + e_2^2 - 3e_0e_2 + \cdots)]$$

$$(\bar{y}_k - \mu_Y) = \mu_Y\left[e_0 - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X(+e_1e_4 - e_3e_1)e_1 + k(2e_0 - e_2 + 3e_0^2 + e_2^2 - 3e_0e_2 + \cdots)\right] \qquad (3)$$

Taking expectation both sides of (3) up to order O(1/n), we get the Bias of $\bar{y}_k$

$$\text{Bias}(\bar{y}_k) = E(\bar{y}_k - \mu_Y) = \frac{\mu_Y}{n}\left[\rho\frac{\sigma_Y}{\sigma_X}\left\{\frac{\mu_{300}}{\sigma_X^2} - \frac{\mu_{2100}}{\sigma_{XY}}\right\} + k\left\{A_Y + 3\frac{C_Y^2}{\theta_Y} - 3\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right\}\right] \qquad (4)$$

Squaring and Taking expectation both sides of (3) up to order O(1/n), we get the Mean Squared Error of $\bar{y}_k$

$$MSE(\bar{y}_k) = (1-\rho^2)\frac{\sigma_Y^2}{n} + \frac{1}{n}\left[\sigma_u^2 + \rho^2\left(\frac{\sigma_Y}{\sigma_X}\right)^2\sigma_v^2\right] + \frac{k^2\mu_Y^2}{n}\left[4\frac{C_Y^2}{\theta_Y} + A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right] + 2\frac{k\mu_Y^2}{n}\left[\rho\left(\frac{\sigma_Y}{\sigma_X}\right)\frac{\mu_{1200}}{\sigma_Y^2} + 2\frac{C_Y^2}{\theta_Y} - \frac{\mu_{0300}}{\sigma_Y^2\mu_Y} - 2\mu_X\rho^2C_XC_Y\right] \qquad (5)$$

The optimum value of k minimizing the mean square error of $\bar{y}_k$ in (5) is given as

$$k = \frac{-\left[\rho\left(\frac{\sigma_Y}{\sigma_X}\right)\frac{\mu_{1200}}{\sigma_Y^2} + 2\frac{C_Y^2}{\theta_Y} - \frac{\mu_{0300}}{\sigma_Y^2\mu_Y} - 2\mu_X\rho^2 C_X C_Y\right]}{\left[4\frac{C_Y^2}{\theta_Y} + A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]} \qquad (6)$$

The Minimum Mean Squared Error of $\bar{y}_k$ for the optimum value of k is

$$\mathrm{MSE}(\bar{y}_k)_{\min} = (1-\rho^2)\frac{\sigma_Y^2}{n} + \frac{1}{n}\left[\sigma_u^2 + \rho^2\left(\frac{\sigma_Y}{\sigma_X}\right)^2\sigma_v^2\right] - \frac{\mu_Y^2\left[\rho\frac{\mu_{1200}}{\sigma_X\sigma_Y} + 2\frac{C_Y^2}{\theta_Y} - \frac{\mu_{0300}}{\sigma_Y^2\mu_Y} - 2\mu_X\rho^2 C_X C_Y\right]^2}{n\left[\frac{4C_Y^2}{\theta_Y} - A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]} \qquad (7)$$

IV. THEORETICAL EFFICIENCY COMPARISON

Under this section the efficiency of the proposed estimator has been compared with the usual linear regression estimator in presence of measurement errors

From Maneesha and Singh (2002), the MSE of usual linear regression estimator $\bar{y}_{lr}$ in presence of measurement errors is given by

$$\mathrm{MSE}(\bar{y}_{lr}) = (1-\rho^2)\frac{\sigma_Y^2}{n} + \frac{1}{n}\left[\sigma_u^2 + \rho^2\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\sigma_v^2\right] \qquad (8)$$

From equation (8) and equation (7), we have,

$\mathrm{MSE}(\bar{y}_{lr}) - \mathrm{MSE}(\bar{y}_k)_{\min} > 0$, if

$$\frac{\mu_Y^2\left[\rho\frac{\mu_{1200}}{\sigma_X\sigma_Y} + 2\frac{C_Y^2}{\theta_Y} - \frac{\mu_{0300}}{\sigma_Y^2\mu_Y} - 2\mu_X\rho^2 C_X C_Y\right]^2}{n\left[\frac{4C_Y^2}{\theta_Y} - A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]} > 0 \qquad (9)$$

If any data set satisfy the optimal condition (9), then our proposed estimator will be more efficient than the usual linear regression estimator in presence of measurement errors for that data set.

V. SIMULATION STUDY

We demonstrate the performance of all estimators by generating a sample from Normal distribution by using R software. The auxiliary information on variable X has been generated by N (5,10) population. This type of population is very relevant in most socio -economic situations with one interest and one auxiliary variable.
The description of this data is as follows

X = N(5,10), Y = X + N(0,1), y = Y+N(1,3), x =X+N(1,3),n=5000,$\mu_X = 4.95$, $\mu_Y = 4.93$, $\sigma_X^2 = 99.38$, $\sigma_Y^2 = 100.12$, $\sigma_u^2 = 25.57$, $\sigma_v^2 = 24.28$, $\rho_{XY} = 0.99$, $C_X = 2.012$, $C_Y = 2.029$, $\lambda = -0.038$, $A_X = 3.05$, $A_Y = 3.11$,
By using these values, the mean square errors (MSE) of the estimators of our interest are given as follows,
$\mathrm{MSE}(\bar{y}_{lr}) = 3.98$and $\mathrm{MSE}(\bar{y}_k) = 3.81$

From above, the percent relative efficiency(PRE) of the proposed estimator$\bar{y}_k$over the usual linear regression estimator$\bar{y}_{lr}$in presence of measurement errors is 104%, showing that the proposed estimator has enhanced efficiency than the usual linear regression estimator.

VI. CONCLUSION

From the above results**,** it is observed that the proposed estimator $\bar{y}_k$ is more efficient than the usual linear regression estimator $\bar{y}_{lr}$ in the presence of measurement errors in the sense of having lesser mean square error. Therefore the proposed estimator is recommendedto survey practitioners over the usual linear regression estimator in the situation when observations are affected by measurement errors.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

VII. REFERENCES

[1] Cochran, W.G.(1968). Errors of measurement in statistics, *Technometrics*, 10,637-666

[2] Sukhatme, P. V. and G.R. Seth (1952). Non sampling errors in surveys, *Journalof the Indian Society of Agricultural Statistics*, 4, 5-41.

[3] Paul P. Biemer, Robert M.Groves, Lars E.Lyberg, Nancy A. Mathiowetz andSeymour Sudman (1991). *Measurement Errors in Surveys*, New York: Wiley.

[4] Shalabh,(1997). Ratio method of estimation in the presence of measurementerrors, *Jour. Ind. Soc. Ag. Statistics*,Vol.I, No-2, 150-155.

[5] Maneesha and Singh R.K.(2002). Role of regression estimator involving measurement errors, *Brazilian Journal of Probability and Statistics*, 16, 39-46.

[6] Singh H.P. and Karpe , N. (2009). A general procedure for estimating thegeneral parameter using auxiliary information in presence of measurementerrors, *Communication of the Korean Statistical Society*,16(5), 821-840.

[7] Misra S. and Yadav. D.K.(2015). Estimating Population Mean Using Known Coefficient of Variation under Measurement Errors in the edited book Statistics and Informatics in Agricultural Research, Excel India Publisher, New Delhi.

[8] Misra S., Dipika and Yadav, D.K.,(2016), Some Improved Estimators for Estimating Population Variance in the Presence of Measurement Errors, Journal of Statistics Applications and Probability, 5(2) 311-320.

[9] Misra S., Yadav D.K, and Dipika, (2016). An Efficient Estimator for Estimating Population Variance in Presence of Measurement Errors, International Journal of Mathematics And its Applications, Vol.4, Issue-2(D), 23-28, ISSN-2347-1557.

[10] Misra S., Dipika and Yadav D.K,(2017) An Improved Estimator of Population Variance using Known Coefficient of Variation Journal of Statistics Application and Probability Letters, *Natural Science Publishing*, New York, USA,ISSN- 2090-8458,No-1, Vol-4, 11-16.

CITE AN ARTICLE

Misra, S., Yadav, D. K., & D. (2017). ESTIMATION OF POPULATION MEAN USING AUXILIARY INFORMATION IN PRESENCE OF MEASUREMENT ERRORS. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 6*(6), 499-502. doi:10.5281/zenodo.817860