Estimation of the Population Mean in the Presence of Measurement and Response Error

Olayiwola O. M [1], Ajayi A. O. [1], Akintunde A. A. [1], Wale-Orojo O. A. [1], Ajibade F. B.[2], Ojelabi O.[1]

[1] **Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria**
[2] **Department of General Studies, Mathematics units, Petroleum Training Institute, Effurun, Delta state Nigeria**

Corresponding Author: O.M. Olayiwola, olayiwolaom@funaab.edu.ng

ABSTRACT: Call back helps to reduce non-response in survey. Existing works had investigated the problem of errors in survey, but there is little or no of information in existing work where these problems exist together in survey. This work derived an estimator which incorporates the problem of non-response and measurement error. The data considered is the height and weight of students with weight as an auxiliary variable. Descriptive statistics were used to determine average height and weight while box plot was used to check for outliers across the height and weight. The average height and weight for the students were 4.488 feet and 53.95kg respectively, while the box plot indicated that there were no outliers. The Mean Square error (MSE) of the proposed estimator when there is non-response error alone was 20.5454, MSE of the proposed estimator when there is measurement error alone was 29.2543, the MSE of the new estimator when there is non-response and measurement error was gotten to be 73.3735 while for the existing es was gotten to be 206.551. It can be noticed that the MSE of the proposed estimator was lesser than the existing, hence more efficient.
KEYWORDS: Mean Square error (MSE), estimator, measurement error

1. INTRODUCTION

In practice the researcher faces the problem of measurement error while collecting information from individuals. Measurement error is the difference between the value that is recorded and the true value of a variable in the study. The reasons non-response occurs include non-availability of the respondents at home, refusal to answer the questionnaire, lack of information. Hansen and Hurwitz (1946)[6] considered the problem of non-response while estimating the population mean by taking a sub sample from the non-respondent group with the help of extra efforts and an estimator was proposed by combining the information available from the response and non-response groups.

Researchers have ignored the presence of possible measurement errors and researchers who have studied measurement errors have neglected non-response. In practice, it is possible for a researcher to face the problem of both errors at the same time. Jackman (1999) dealt with both non-response and measurement error simultaneously, in the case of voter turnout, where a reasonably large body of vote validation studies supply auxiliary information, allowing the components of bias in survey estimates of turnout rates to be isolated. Averaging over the auxiliary information provides bounds on the quantity of interest, yielding an estimate corrected for both nonresponse and measurement error. Further, Dixon (2010) studied the estimation of non-response bias and measurement error on the data from Consumer Expenditure Quarterly Interview Survey (CEQ), Current Population Survey (CPS) and National Health Interview Survey (NHIS), an attempt to measure the differences in employment status of Washington.

In estimating population parameters, sample survey experts sometimes use auxiliary information to estimate improve precision. These gave room for ratio, regression and product estimators, however these estimators assumed that the variable of interest is Normally distributed.

In this paper, putting in mind that the variable of interest may not be normally distributed, we developed new estimators for estimating the population parameter in the presence of measurement error and non-response error in the study as well as in the auxiliary variable.

2. BRIEF LITERATURE

Singh and Karpe (2008, 2010), Kumar et al. (2011), Sharma and Singh (2015) worked on population mean under measurement error. Non-response is common in mail surveys than in personal interviews.

The usual approach is call back to overcome non-response. Sharma and Singh (2015)[13] considered the ratio, product and Regression methods of estimation in the presence of measurement and non-response errors
In the [1][2][8][9][21][16][17][18][19][11][14][13], is discussed the problem of estimating population parameter using auxiliary information in presence of measurement errors.

3. METHODOLOGY

Consider population $U=(U_1, U_2, ..., U_N)$. Let Y and X be variable of interest and auxiliary variates, respectively. It is assumed that x_i and y_i for the i_{th} sampling units are observed with measurement error instead of their true values (X_i, Y_i) For a simple random sampling scheme, let (x_i, y_i) be observed values instead of the true values (X_i, Y_i) for i_{th} (i=1,2….n) unit, as

$$U_i=y_i-Y_i$$
$$V_i=x_i-X_i$$

u_i and v_i are associated measurement errors with mean zero and δ_u^2 and δ_v^2 respectively. u_i's and v_i's are uncorrelated although X_i's and Y_i's are correlated. Let the population means X and Y characteristics be μ_x and μ_y, and let ρ be the population correlation between x and y respectively [9]. Similarly, to [18], for U and V independence was assumed and uncorrelated with X and Y, in [18].
In the case of nonresponse of at initial stage, Hansen and Hurwitz in [6] considered the mail surveys at the first attempt and the personal interviews at the second attempt. In the Hansen and Hurwitz method the population is supposed to be consisting of response Stratum of size N_1 and the non-response stratum of size N_2=(N-N_1).

$\bar{Y}=\sum_{i=1}^{N}\frac{y_i}{N}$ and

$S_y^2=\sum_{i=1}^{N}\frac{(y_i-\bar{Y})^2}{N-1}$ denote the mean and the population variance of the study variable y.

$\bar{Y}_1=\sum_{i=1}^{N_1}\frac{y_i}{N_1}$ and $S_{y(1)}^2=\sum_{i=1}^{N_1}\frac{(y_i-\bar{Y})^2}{N_1-1}$ denote the mean and variance of response group. Similarly, let

$\bar{Y}_2=\sum_{i=1}^{N_2}\frac{y_i}{N_2}$ and $S_{y(2)}^2=\sum_{i=1}^{N_2}\frac{(y_i-\bar{Y})^2}{N_2-1}$ denote the mean and variance of the non-response group. The population mean can be written as $\bar{Y}=W_1\bar{Y}_1+W_2\bar{Y}_2$ where $W_1=\frac{N_1}{N}$ and $W_2=\frac{N_2}{N}$. The sample mean $\bar{y}_1=\sum_{i=1}^{N_2}\frac{y_i}{n_1}$ is an unbiased for $\bar{Y}_1$, but has a bias equal to $W_2(\bar{Y}_1-\bar{Y}_2)$ in estimating the population mean $\bar{Y}$. The sample mean $\bar{y}_{2r}=\sum_{i=1}^{r}\frac{y_i}{r}$ is unbiased for the mean y_2 for the n_2 units.

Let $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$, be the unbiased estimator of population means $\bar{X}$ and $\bar{Y}$, respectively but $S_x^2=\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}$ and $S_y^2=\sum_{i=1}^{n}\frac{(y_i-\bar{Y})^2}{n-1}$ are not unbiased estimator of (σ_x^2,σ_y^2), respectively. The expected values of S_x^2 and S_y^2 in the presence of measurement error are, given by,

$E(S_x^2)=\sigma_x^2+\sigma_v^2$,
$E(S_y^2)=\sigma_y^2+\sigma_v^2$ and for non-response group
$E(S_{x_2}^2)=\sigma_{x_2}^2+\sigma_{v_2}^2$
$E(S_{y_2}^2)=\sigma_{y_2}^2+\sigma_{u_2}^2$

When the error variance $\sigma_{v_2}^2$ is known, the unbiased estimator of σ_x^2 is $\sigma_x^2=s_x^2-\sigma_v^2>0$, and when σ_u^2 is known, then the unbiased estimator of σ_y^2 is $\sigma_y^2=s_y^2-\sigma_u^2>0$.
Similarly, for the non-response group the unbiased estimator of $\sigma_{x_2}^2$ is $\sigma_{x_2}^2=s_{x_2}^2-\sigma_{v_2}^2>0$, and when $\sigma_{u_2}^2$ is known, then the unbiased estimator of $\sigma_{y_2}^2$ is $\sigma_{y_2}^2=s_{y_2}^2-\sigma_{u_2}^2>0$.

Define, Singh and Karpe in [18],
$\bar{y}=\mu_y(1+e_0)$, $\bar{x}=\mu_x(1+e_1)$

Such that
$E(e_0)=E(e_1)=0$,

$$E(e_0^2)=\frac{C_y^2}{n}\left(1+\frac{S_u^2}{S_y^2}\right)+\frac{W_2(k-1)}{n}C_{y_2}^2\left(1+\frac{S_{u_2}^2}{S_{y_2}^2}\right)$$

$$E(e_1^2)=\frac{C_y^2}{n}\left(1+\frac{S_v^2}{S_x^2}\right)+\frac{W_2(k-1)}{n}C_{x_2}^2\left(1+\frac{S_{v_2}^2}{S_{x_2}^2}\right)$$

$$E(e_0e_1)=\frac{\rho_{yx}C_yC_x}{n}+\frac{W_2(k-1)}{n}\rho_{yx_2}C_{y_2}C_{x_2}$$

$C_y = {S_y}/{\bar{Y}}$, $C_x = {S_x}/{\bar{X}}$, $C_{y_2} = {S_{y_2}}/{\bar{Y}}$, $C_{x_2} = {S_{x_2}}/{\bar{X}}$,

$\rho_{xy} = {S_{xy}}/{S_x S_y}$

When there is non-response and response error both are present, a ratio type estimator for estimating population mean was Singh and Sharma [20] given by

$$t_r = \frac{\bar{y}^*}{\bar{x}^*}\bar{X} \qquad (1)$$

With mean square error to be

$$MSE(t_r) = \frac{1}{n}\left[S_y^2\left(1+\frac{\sigma_u^2}{S_y^2}\right)+S_x^2\left(1+\frac{\sigma_v^2}{S_x^2}\right)-2\rho_{xy}S_xS_y\right]+ A\left[S_{y2}^2\left(1+\frac{\sigma_{u2}^2}{S_{y2}^2}\right)+S_x^2\left(1+\frac{\sigma_{v2}^2}{S_{x2}^2}\right)-2\rho_{xy2}S_{x2}S_{y2}\right]$$

where $A = \dfrac{(k-1)W_2}{n}$ (2)

3.1. Proposed Estimator

$$t_{er} = \bar{y}^* \exp\left(\frac{\bar{X}-\bar{x}^*}{\bar{X}+\bar{x}^*}\right) \text{(ratio)} \qquad (3)$$

The Bias and MSE of the estimator t_{r2}, given by

$$\left(t_{er}-\bar{Y}\right) = \bar{Y}\left(e_0 - \tfrac{1}{2}e_0e_1 + \tfrac{3}{8}e_1^2\right)$$
$$\left(t_{er}-\bar{Y}\right)^2 = \bar{Y}^2\left(e_0^2 - \tfrac{1}{4}e_1^2 + e_0e_1\right) \qquad (4)$$

$$B(t_{er}) = E\left(t_{r1}-\bar{Y}\right) = \bar{Y}\left(E(e_0) - \tfrac{1}{2}E(e_0e_1) + \tfrac{3}{8}E(e_1^2)\right) \qquad (5)$$

$$B(t_{er}) = \bar{Y}\left\{\frac{3}{8}\left[\frac{S_x^2}{n}\left(1+\frac{S_v^2}{S_x^2}\right)+\frac{W_2(k-1)}{n}S_{x_2}^2\left(1+\frac{S_{v_2}^2}{S_{x_2}^2}\right)\right] - \frac{1}{2}\left[\frac{\rho_{yx}S_yS_x}{n}+\frac{W_2(k-1)}{n}\rho_{yx_2}S_{y_2}S_{x_2}\right]\right\} \qquad (6)$$

$$MSE(t_{r2}) = E\left(t_{r2}-\bar{Y}\right)^2 = \bar{Y}^2\left(E(e_0^2) + \tfrac{1}{4}E(e_1^2) - E(e_0e_1)\right) \qquad (7)$$

$$MSE(t_{r2}) = \left\{\begin{array}{l}\left[\frac{S_y^2}{n}\left(1+\frac{S_u^2}{S_y^2}\right)+\frac{W_2(k-1)}{n}S_{y_2}^2\left(1+\frac{S_{u_2}^2}{S_{y_2}^2}\right)\right] \\ +\frac{1}{4}\left[\frac{S_x^2}{n}\left(1+\frac{S_v^2}{S_x^2}\right)+\frac{W_2(k-1)}{n}S_{x_2}^2\left(1+\frac{S_{v_2}^2}{S_{x_2}^2}\right)\right] \\ -\left[\frac{\rho_{yx}S_yS_x}{n}+\frac{W_2(k-1)}{n}\rho_{yx_2}S_{y_2}S_{x_2}\right]\end{array}\right\} \qquad (8)$$

3.2. Class of Estimator

$$t_{cp} = m_1\bar{y}^\bullet + m_1 t_{er} \qquad (9)$$

It should be noted that

(i) For $(m_1, m_2) = (1,0)$ $t_{c1} = \bar{y}^\bullet$ (usual unbiased estimator)

(ii) For $(m_1, m_2) = (0,1)$ $t_{c2} = \bar{y}^* \exp\left(\dfrac{\bar{X}-\bar{x}^*}{\bar{X}+\bar{x}^*}\right)$ (usual exponential-ratio estimator)

Thus, the proposed class of estimators is generalized version of usual unbiased estimator and exponential-ratio estimator. Expressing the estimator t_{cp} in terms of e's we have

$$t_{cp} = m_1\bar{Y}(1+e_0) + m_2\bar{Y}(1+e_0)\exp\left[\frac{-e_1}{2+e_1}\right] \qquad (10)$$

$$\left(t_{cp}-\bar{Y}\right) = \bar{Y}\left[e_0 + m_2\left(\frac{-e_1}{2}+\frac{3}{8}e_1^2-\frac{1}{2}e_0e_1\right)\right] \qquad (11)$$

Squaring both sides

$$\left(t_{cp}-\bar{Y}\right)^2 = \bar{Y}^2\left[e_0^2 + m_2\frac{e_1^2}{4} - m_2e_0e_1\right] \qquad (12)$$

$$MSE\left(t_{cp}\right) = M + m_2^2R^2\frac{N}{4} - m_2RO \qquad (13)$$

Where $M = \left(\dfrac{1-f}{n}\right)S_y^2 + \dfrac{W_2(k-1)}{n}S_{y(2)}^2$,

$N = \dfrac{S_x^2}{n}\left(1+\dfrac{S_v^2}{S_x^2}\right) + \dfrac{W_2(k-1)}{n}C_{x_2}^2\left(1+\dfrac{S_{v_2}^2}{S_{x_2}^2}\right)$ and

$O = \dfrac{\rho_{yx}S_yS_x}{n} + \dfrac{W_2(k-1)}{n}\rho_{yx_2}S_{y_2}S_{x_2}$

The optimum value of m_2 is obtained by minimizing (13), given by

$$m_2 = \frac{4O}{RN} \qquad (14)$$

Substituting the optimum value of m_2 will give the minimum MSE of the estimator.

4. DATA ANALYSIS

μ_{ynon} is the estimator when the data contains just nonresponse error

$\mu_{ymeasure}$ is the estimator when the data contains just measurement error

t_{r1} existing estimator when the data contains nonresponse and measurement error

t_{r2} proposed estimator when the data contains nonresponse and measurement error

The data on the height and weight of students of the department of statistics, Igbajo Polytechnic were extracted from the student files in the record unit. This data was measured when they got admission. The students were recalled their height and weights were taken again. The new heights and weights were now taken as measured variables. Text messages were sent to the 75 students. 50 students responded and questionnaires were administered to them. After a day remainder was sent to 15 students out of the remaining 25 students. Samples were then taken from the first 50 who responded which were added to the 15 students who later responded.
These consist of the Age, Sex and Marital status of the students.

Table 1: Socio Demographic Data

	Frequency	Percent
Sex		
Female	48	64.0
Male	27	36.0
Total	75	100.0
Age		
Under 16	6	8.0
17-20	28	37.3
21-30	37	49.3
Above 31	4	5.3
Total	75	100.0
Marital Status		
Single	48	64.0
Married	26	34.7
Divorced	1	1.3
Total	75	100.0

Table 2: Parameter Estimation

	Mini	Max	Mean	Std. Deviation
Height	3.90	5.40	4.49	0.41
Weight	40.00	75.00	53.95	8.62
Measured Height	3.44	5.73	4.49	0.17
U	-0.66	1.95	0.00	6.88
Measure Weight	40.00	75.08	53.91	54.32
V	-8.17	5.31	0.00	23.11

$N = 75$; $\mu_Y = 4.488$ $\mu_X = 53.947$, $\sigma_U^2 = 722.635$, $\sigma_V^2 = 534.271$, $\sigma_y^2 = 0.16834$, $\sigma_x^2 = 74.263$, $\rho_{yx} = 0.719$

Table 2.1

N_1	N_2	σ_{y2}^2	σ_{x2}^2	ρ_{yx_2}	σ_{U2}^2	σ_{V2}^2
50	25	2372.479	1981.238	0.892	556.69	411.35

Table 2.2

N_1	N_2	Estimators	Mean Square Errors
50	25	μ_{ynon}	20.5454
		$\mu_{ymeasure}$	29.2543
		t_{r1}	73.3735
		t_{r2}	206.551

5. CONCLUSIONS

When N_1=50 and N_2=25, the MSE of the new estimator when there is non-response error alone (μ_{ynon}) was 20.5454, the MSE of the new estimator when there is measurement error alone ($\mu_{ymeasure}$) was 29.2543, the new estimator when there is non-response and measurement error (t_{r2}) was gotten to be 73.3735 while for the existing was gotten to be 206.551 as can be seen in table 2.2. It was noticed that the MSE of the new estimator was lesser than the existing, hence it can be said to be more efficient.

REFERENCES

[1] **Allen, J., Singh, H. & Smarandache, F.** - *A family of estimators of population mean using multi-auxiliary information in presence of measurement errors*, International Journal of Social Economics 30, 837–849, 2003

[2] **Fuller, W.** - *Estimation in the presence of measurement error*, International Statistical Review 63, 121–147, 1995

[3] **Gregoire, T. G.** - *Design-based and model-based inference in survey sampling: Appreciating the difference*, Canadian Journal of Forestry Research 28, 1429–1447, 1998

[4] **Gregoire, T. G. and Salas, C.** - *Ratio estimation with measurement error in the auxiliary variate*, Biometrics 65, 590–598, 2009

[5] **Groves, R. M.** - *Survey Errors and Survey Costs*, Wiley, New York, 1989

[6] **Hansen, M. and Hurwitz, W.** - *The problem of non-response in sample surveys*, Journal of American Statistical Association 41, 517–529, 1946

[7] **Jackman, S.** - *Correcting surveys for non-response and measurement error using auxiliary information*, Electoral Studies 18, 7–27, 1999

[8] **Manisha and Singh, R.** - *An estimation of population means in the presence of measurement errors*, Journal of the Indian Society of Agricultural Statistics 54, 13–18, 2001

[9] **Manisha and Singh, R.** - *Role of regression estimator involving measurement errors*, Brazilian Journal of Probability and Statistics 16, 39–46, 2002

[10] **Rao, P.** - *Ratio estimation with sub sampling the non-respondents*, Survey Methodology 12, 217–230, 1986

[11] **Salas, C. and Gregoire, T.** - *Statistical analysis of ratio estimators and their estimators of variances when the auxiliaryvariate is measured with error*, European Journal of Forest Research 129, 847–861, 2010

[12] **Sarndal, C. E., Swensson, B. & Wretman, J.** - *Model Assisted Survey Sampling*, Springer, New York, 1992

[13] **Sharma, P. and Singh, R**. - *A generalized class of estimators for finite population variance in presence of measurement errors*, Journal of Modern Applied Statistical Methods 12, 231–241, 2013

[14] **Shukla, D., Pathak, S. and Thakur, N.S.** - *Class(es) of factor-type estimator(s) in presence of measurement error*, Journal of Modern Applied Statistical Methods 11, 336–347, 2012

[15] **Shukla, D., Pathak, S. and Thakur, N.S.** - *An estimator for mean estimation in presence of measurement error*, Research and Reviews: A Journal of Statistics 1, 1–8, 2012

[16] **Singh, H. and Karpe, N.** - *Effect of measurement errors on a class of estimators of population mean using auxiliary information in sample surveys*, Journal of Statistical Research of Iran 4, 175–189, 2007

[17] **Singh, H. and Karpe, N**. - *Estimation of population variance using auxiliaryinformation in the presence of measurement errors*, Statistics in Transition- New Series 9, 443–470, 2008

[18] **Singh, H. and Karpe, N.** - *On the estimation of ratio and product of two population means using supplementary information in presence of measurement errors*, Statistica, Anno 69, 27–47, 2009

[19] **Singh, H. and Karpe, N.** - *Estimation of mean, ratio and product using auxiliary information in the presence of measurement errors in sample surveys*, Journal of Statistical Theory and Practice 4, 111–136, 2010

[20] **Singh, R. and Sharma, P.** - *Method of Estimation in the Presence of Non-response and Measurement Errors Simultaneously*, Journal of Modern Applied Statistical Methods, Vol. 14, No. 1, 107-121, 2015

[21] **Wang, L.** - *A simple adjustment for measurement errors in some dependent variable models*, Statistics & Probability Letters 58, 427–433, 2002