Journal of Modern Applied Statistical Methods

Volume 23 | Issue 1 Article 6

Determining the Effect of Measurement Errors and Estimation of Mean by Factor-Type Estimator in Simple Random Sampling

Amita Yadav
Department of Mathematics, Statistics and Computing, Banasthali Vidyapith, India,
amitayadav73@yahoo.in

Narendra Singh Thakur
Department of Statistics, Govt. Adarsh Girls College, Jiwaji University, India,
nst_stats@yahoo.co.in

Sarla Pareek
Department of Mathematics, Statistics and Computing, Banasthali Vidyapith, India,
psarla13@gmail.com

Recommended Citation

Yadav, A., Thakur, N.S. and Pareek, S. (2024). Determining the Effect of Measurement Errors and Estimation of Mean by Factor-Type Estimator in Simple Random Sampling. Journal of Modern Applied Statistical Methods, 23(1), https://doi.org/10.56801/Jmasm.V23.i1.6

Journal of Modern Applied Statistical Methods
2024, Vol. 23, No. 1,
Doi: 10.56801/Jmasm.V23.i1.6

ISSN 1538 – 9472

Determining the Effect of Measurement Errors and Estimation of Mean by Factor-Type Estimator in Simple Random Sampling

Amita Yadav
Department of Mathematics, Statistics and Computing, Banasthali Vidyapith, India

Narendra Singh Thakur
Department of Statistics, Govt. Adarsh Girls College, Jiwaji University, India

Sarla Pareek
Department of Mathematics, Statistics and Computing, Banasthali Vidyapith, India

In sample surveys, non-sampling errors are inherent due to non-response and measurement errors. Non-response occurred when the response is not available from respondent due to not being at home, no willingness to answer etc. Another error is the difference between the observed and actual values is known as measurement error. For dealing with such kind of errors1, we advocate the use of factor-type estimator for estimation of population mean in this manuscript and the expressions of the bias, mean square error (MSE) and optimum mean square error of suggested estimator has been derived. An empirical study is also performed over a real data set and bias, mean squared error and efficiency calculated for comparison purpose by simulation.

Keywords: Estimation, non-sampling error, measurement error, simple random sampling, factor-type estimator.

1. Introduction

In the present scenario - states, industries, scientific institutions, public organizations, international agencies, etc. are the primary users of statistical data and it is necessary to collect fair data from the respondents according to requirement of the purpose. For instance, for better medical facilities, it is necessary to collect data about health status from the respondent of the country. One way of finding it is to collect the data from all the individuals of the country. This system of collection of data is called *complete enumeration* or *census.* Obviously, more labour, cost and time will be needed to obtain data for the census. However, if information about every unit is required then there is need of census. Population census, agricultural census, income tax assessment, preparation of voter list for different election purposes, etc. are examples for need of census.

But when some amount of error is permissible it may possible to select some units in the scientific manner from the population and infer the population on the basis of selected units. This set of some units is known as a *sample* and the procedure of selection of the sample is called *sampling*. Sampling is a routine experience for all of us, for example - a person makes a judgment about the quality of items by examining only a few items from a lot offered for sale.

On the other hand, in the case of destructive units, complete enumeration is impractical. If it is to obtain the average life of refrigerators, televisions, bulbs, etc., then we will have to limit the observations, up to a part of the population and inference will have to be drawn by the data in the sample. However, the outputs are different from the population values due to inference about the population made from the sample in the survey.

On the other hand, since a part of the population is examined, therefore errors are inherent and by regulating in a proper manner and operation by trained persons can reduce the same. By specified statistical principles, a sample survey is carried out to estimate the characteristics of the population.

Comparatively, sampling is a technique of saving money, time, and manpower to assess information about unknown population parameters, in a faster and more precise manner instead of complete enumeration. Sample surveys are extremely popular and the data collected, serves as one of the foremost sources of information, required for reliable predictions of unknown population parameters for investigators, researchers, planners, administrators, etc. In addition, these involve well-trained manpower for data handling which produces high-level accuracy in predictions due to the least incorporation of non-sampling errors. The sampling approach provides a dual optimization advantage, i.e., it is smaller in scale and provides an equally efficient forecast about unknown parameters.

A regular difference between answer given by the respondent and actual answer is termed as measurement bias. For example, a respondent may report lower income than actual income. [Cochran (2005), Sukhatme et al. (1984)].

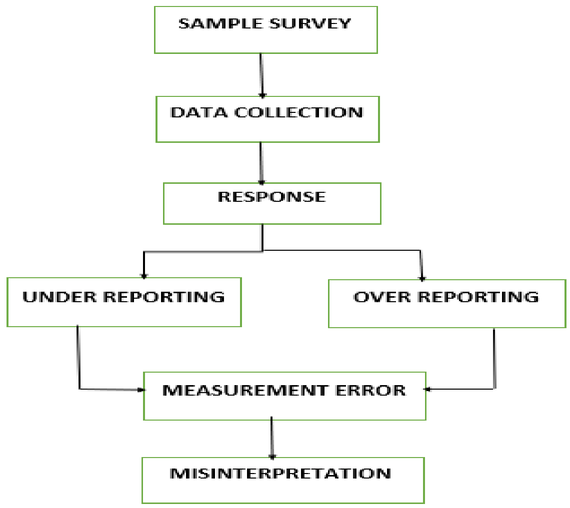

Figure. 1: Measurement errors

The methods of estimation are examined under the hypothesis that composed data are pure and consistent. Although, data collected in real life, through samples may contain errors due to under or over reporting, not willing to answer, privacy issues and other reasons of respondents (Figure. 1). The difference between inspected (observed) and actual values is known as an error and is technically named as measurement error.

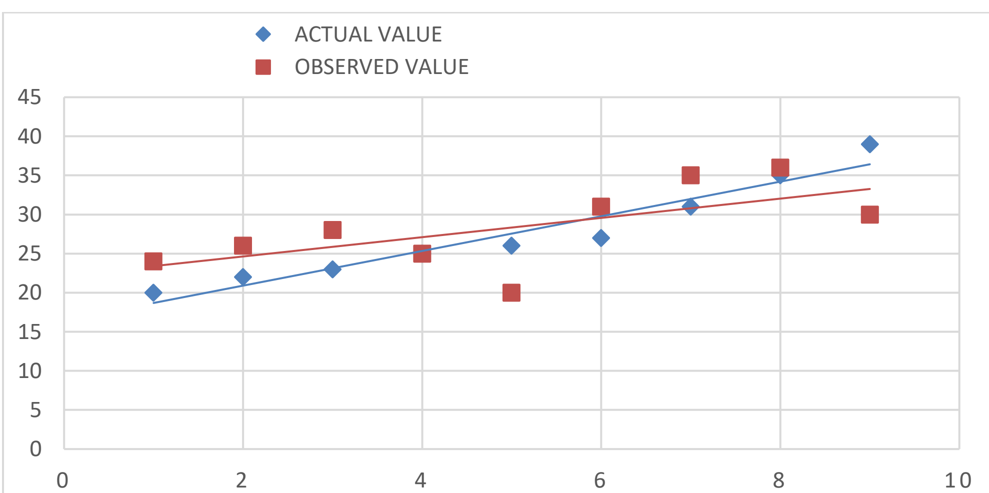

Figure. 2: Effect of measurement errors

Measurement error causes serious results for estimation. Figure.2 represents the observed (with measurement errors) and actual values from the respondent and their linear representation. Measurement error be the gap between observed values and actual values of the variable. The regression lines for observed and actual values are different, therefore, it is clear that the predictions yield by observed values will ambiguous and misleading.

Due to the importance of measurement error problems, there are a large number of research papers and books based on measurement errors are available in the literature. Hence it is quite difficult for us to discuss all the material of existing literature. Instead of reviewing all the same, we focus on the relatively important researches on measurement error-based problems in sample survey. Cochran (1968) provided a valuable text on the review of measurement errors in Statistics.

A family of factor-type ratio estimators was given by Singh and Shukla (1987) with one parameter. The estimator proposed by them is

$$\bar{y}^* = \frac{\bar{y}}{\bar{x}^*}, \tag{1}$$

$$where\ \bar{x}^* = \left[\frac{(A+C)\bar{X} + fB\bar{x}}{(A+fB)\bar{X} + C\bar{x}}\right]^{-1}$$

Where $A = (m-1)(m-2), B = (m-1)(m-4), C = (m-2)(m-3)(m-4)$, $(0 < m < \infty)$ is a constant.

The beauty of this estimator is that it provides different estimators like, ratio, product, dual to ratio, etc. on different values of parameter *m*. The factor-type estimator is bias controlled with multiple choices of parameter *m* for optimum mean squared error (MSE). Shukla et al. (2012b) introduced an estimation strategy for the purpose of optimization in presence of measurement error using factor-type estimator.

2. Notations and set up

Let *U* be a finite population of size *N* and *S* be a sample of size *n* drawn from population by SRSWOR. Let *y* and *x* represents the study and auxiliary variables, respectively under consideration. $\bar{Y}$, $\bar{X}$ be the population means and $\bar{y}$, $\bar{x}$ be the sample means of the *y* and *x*, respectively. S_Y^2 and S_X^2 represents the population variance for *y* and *x* and ρ is the coefficient of correlation between *y* and *x*.

Let the measurement errors are present in the sample while selecting, recording or processing the data. The i^{th} observed unit of *y* and *x* in the presence of measurement error in the sample is (y_i, x_i) (say) respectively. Let (u_i, v_i) be the measurement errors corresponding to the (y_i, x_i) such that $(y_i = Y_i + u_i\ , x_i = X_i + v_i)$ where Y_i and X_i be the actual values of $y_i\ and\ x_i$ respectively.

Since, sample observations are independent to each other, therefore the measurement error present on *y* and *x* are independent of each other, so *u* and *v* are also independent. Also, assume that the average of u_i and v_i are zero as the measurement error caused by under and over-reporting problems. S_u^2 and S_v^2 be the population variance for *u* and *v*, respectively. This sampling strategy is denoted by Ω and define the following symbols.

$A_\Omega = \lambda(S_Y^2 + S_u^2),\ B_\Omega = \lambda(S_X^2 + S_v^2),\ C_\Omega = \lambda\rho S_Y S_X.$

Let $\omega_y = \sum_{i=1}^{n}(Y_i - \bar{Y}),\ \ \omega_x = \sum_{i=1}^{n}(X_i - \bar{X}),\ \ \ \omega_u = \sum_{i=1}^{n} u_i\ ,\ \ \omega_v = \sum_{i=1}^{n} v_i.$

By adding ω_y and ω_u and dividing by *n*.

$$\frac{1}{n}(\omega_y + \omega_u) = \sum_{i=1}^{n} \frac{1}{n}(y_i - \bar{Y}). \tag{2}$$

So $\bar{y} = \bar{Y} + \frac{1}{n}(\omega_y + \omega_u)$.

Similarly, $\bar{x} = \bar{X} + \frac{1}{n}(\omega_x + \omega_v)$.

And the expected values are $E\left(\frac{1}{n}\ (\omega_y + \omega_u)\right)^2 = A_\Omega \bar{Y}^{-2}, E\left(\frac{1}{n}\ (\omega_x + \omega_v)\right)^2 = B_\Omega \bar{X}^{-2}$ and E $\left(\frac{1}{n}(\omega_y + \omega_u)\frac{1}{n}(\omega_x + \omega_v)\right) = C_\Omega \bar{X}^{-1}\ \bar{Y}^{-1}$, $\lambda = \frac{1}{n} - \frac{1}{N}$, $R = \frac{\bar{Y}}{\bar{X}}$.

3. Some existing estimators

Some existing estimators of population mean $\bar{Y}$ in the setup **Ω** are discussed in this section.

3.1 Sample mean estimator

Sample mean estimator is very well – known and in setup **Ω**, is defined as

$$t_1 = \bar{y} = \frac{1}{n}\sum y \tag{3}$$

t_1 is an unbiased estimator and its variance is by

$$V(t_1) = A_\Omega \tag{4}$$

3.2 Shalabh (1997) estimator

Shalabh (1997) suggested a ratio-type estimator in setup **Ω** and it is defined as

$$t_2 = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{5}$$

The estimator t_2 is biased and the expression of bias is

$$B(t_2) = \bar{X}^{-1}[RB_\Omega - C_\Omega] \tag{6}$$

and its MSE is given by

$$M(t_2) = [A_\Omega + R^2 B_\Omega - 2RC_\Omega] \tag{7}$$

3.3 Manisha and Singh (2001) estimator

An estimator as a combination of different estimators suggested by Manisha and Singh (2001) as

$$t_3 = \phi t_2 + (1 - \phi) t_1 \tag{8}$$

where ϕ is a constant so that MSE of t_3 is minimum. The estimator t_3 is biased and the expression of bias is

$$B(t_3) = \phi \, \bar{X}^{-1} [R B_\Omega - C_\Omega] \tag{9}$$

The expression of MSE of t_3 is

$$M(t_3) = A_\Omega + \phi^2 R^2 B_\Omega - 2\phi \, R C_\Omega \tag{10}$$

and the optimum MSE of t_3 at $\phi_{opt} = (R \, B_\Omega)^{-1} \, C_\Omega$ is

$$M(t_3)_{min} = {B_\Omega}^{-1} [A_\Omega B_\Omega - C_\Omega^2] \tag{11}$$

3.4 Grover and Kaur (2011) estimator

Grover and Kaur (2011) proposed an exponential ratio type under sampling strategy Ω as

$$t_4 = \{p_1 \bar{y} + p_2 (\bar{X} - \bar{x})\} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \tag{12}$$

where p_1, p_2 are constants so that MSE of t_4 is minimum. t_4 is biased and the expression of bias is

$$B(t_4) = (p_1 - 1) \, \bar{Y} + 0.375 \, p_1 R \frac{B_\Omega}{\bar{X}} - 0.5 \, p_1 \frac{C_\Omega}{\bar{X}} + 0.5 \, p_2 \frac{B_\Omega}{\bar{X}} \tag{13}$$

The expression of MSE of t_4 is

$$M(t_4) = \bar{Y}^2 (p_1 - 1)^2 + p_2^2 A_\Omega + B_\Omega (p_2 + 0.5 R p_1)^2 + 2(p_1 - 1)(0.5 \, p_2 \bar{X} + 0.375 \, p_1 \bar{Y}) R B_\Omega \bar{X}^{-1} - C_\Omega [p_1 R (2 \, p_2 + p_1 R) - p_1 (p_1 - 1) R] \tag{14}$$

And the optimum MSE of t_4 at

$$p_1 = \frac{\bar{Y}^2 B_\Omega}{\bar{Y}^2 B_\Omega + A_\Omega B_\Omega - C_\Omega^2}, \quad p_2 = \frac{\bar{Y}^2 (2 C_\Omega - R B_\Omega)}{2(\bar{Y}^2 B_\Omega + A_\Omega B_\Omega - C_\Omega^2)}$$

is

$$M(t_4)_{min} = \bar{Y}^2 - \frac{4 \bar{Y}^2 B_\Omega - R^2 B_\Omega C_\Omega^2 + R^2 A_\Omega B_\Omega^2 + 0.0625 R^4 B_\Omega^3}{4 \bar{Y}^2 B_\Omega + 4 A_\Omega B_\Omega - 4 C_\Omega^2} \tag{15}$$

3.5 Shukla et al. (2012a) estimator

An estimator in the presence of measurement errors for mean estimation is proposed by Shukla et al. (2012) is given by

$$t_5 = \varphi \bar{y}^* + (1 - \varphi) \bar{y} \tag{16}$$

where $\bar{y}^* = \bar{y} (\bar{X} - f \bar{x}) (1 - f)^{-1} \bar{X}^{-1}$

where φ is a constant so that MSE of t_5 is minimum. t_5 is biased and the expression is

$$B(t_5) = \varphi \, f \{(1 - f) \bar{X}\}^{-1} \, C_\Omega \tag{17}$$

the MSE of t_5 is given by

$$M(t_5) = B_\Omega \varphi^2 f^2 R^2 (1-f)^{-2} + A_\Omega - 2f\varphi R C_\Omega (1-f)^{-1} \tag{18}$$

The minimum MSE of t_5 at $\varphi = (1-f) C_\Omega {B_\Omega}^{-1} f^{-1} R^{-1}$ is

$$M(t_5)_{min} = {B_\Omega}^{-1}[A_\Omega B_\Omega - C_\Omega^2] \tag{19}$$

3.6 Ekpenyong and Enang (2014) estimator

Under sampling strategy Ω, Ekpenyong and Enang (2014) suggested an estimator as:

$$t_6 = d_1 \bar{y} + d_2(\bar{X} - \bar{x}) \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \tag{20}$$

where d_1, d_2 are constants so that MSE of t_6 is minimum. t_6 is biased and the expression is

$$B(t_6) = (d_1 - 1)\bar{Y} - 0.5 d_2 \frac{B_\Omega}{\bar{X}} \tag{21}$$

The expression of MSE of t_6 is

$$M(t_6) = A_\Omega d_1^2 + \bar{Y}^2(d_1^2 + 1 - 2d_1) + d_2^2 B_\Omega - 2[RB_\Omega + d_1 C_\Omega - 0.5 d_1 R B_\Omega] \tag{22}$$

The minimum MSE of t_6 at $d_1 = \frac{4\bar{Y}^2 + 4C_\Omega - R^2 B_\Omega}{4\bar{Y}^2 + 4A_\Omega - R^2 B_\Omega}$ and $d_2 = \frac{2R(A_\Omega - C_\Omega)}{4\bar{Y}^2 + 4C_\Omega - R^2 B_\Omega}$ is

$$M(t_6)_{min} = \bar{Y}^2 - \frac{4\bar{Y}^4 B_\Omega - R^2 \bar{Y}^2 B_\Omega^2 + R^2 A_\Omega B_\Omega^2 + 4R\bar{Y}^2 B_\Omega C_\Omega}{4\bar{Y}^2 B_\Omega + 4A_\Omega B_\Omega - (RB_\Omega - 2C_\Omega)^2} \tag{23}$$

3.7 Singh and Pal (2015) estimator

Singh and Pal (2015) proposed an estimator as:

$$t_7 = \frac{\bar{y}}{\bar{x}} \bar{X} \, exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \tag{24}$$

t_7 is a biased estimator and its bias is given by

$$B(t_7) = \bar{Y}(0.625 B_\Omega - 1.5 C_\Omega) \tag{25}$$

and its MSE is given by

$$M(t_7) = [A_\Omega + 2.25 R^2 B_\Omega - 3RC_\Omega] \tag{26}$$

3.8 Gupta and Yadav (2018) estimator

Gupta and Yadav (2018) have used the information of sample size to improve the estimation of the mean of y and recommended a generalized ratio type estimator for estimation purpose as:

$$t_8 = \bar{y}\,[q + (1-q)\left(\frac{\bar{X}+n}{\bar{x}+n}\right)] \tag{27}$$

where constant q is chosen so that MSE of t_8 is minimum. t_8 is biased and the expression of bias is

$$B(t_8) = (q-1)(\bar{X}+n)^{-1}(C_\Omega - RB_\Omega) \tag{28}$$

The expression of MSE of t_8 is

$$M(t_8) = A_\Omega + \bar{Y}^2(\bar{X}+n)^{-2}(1-q)^2\, B_\Omega + 2(q-1)\bar{Y}(\bar{X}+n)^{-1}C_\Omega \tag{29}$$

The minimum MSE of t_8 at $q = 1 - \frac{C_\Omega\,(\bar{X}+n)}{B_\Omega\,\bar{Y}}$ is

$$M(t_8)_{min} = \left[A_\Omega - \frac{C_\Omega^2}{B_\Omega}\right] \tag{30}$$

3.9 Tiwari et al. (2022) estimator

Tiwari et al. (2022) proposed a family of estimator under the set up Ω as

$$t_9 = [a\,\bar{y} + b(\bar{X}-\bar{x})]\left[\frac{\delta\bar{X}+\eta}{\delta\bar{x}+\eta}\right] \exp\left[\frac{3(\bar{X}-\bar{x})}{\bar{X}+\bar{x}}\right] \tag{31}$$

where $a, b,$ are constants so that MSE of t_9 is minimum. The estimator t_9 is bias and the expression of bias is

$$B(t_9) = (a-1)\bar{Y} - a\bar{Y}^{-1}(R_1+1.5R)C_\Omega + [b(R_1+1.5R)$$
$$+ a(R_1^2 + 1.875R^2 + 1.5RR_1)]\bar{Y}^{-1}B_\Omega \tag{32}$$

The expression of MSE of t_9 is

$$M(t_9) = \bar{Y}^2 + aM_1 + a^2M_2 + bM_3 + b^2M_4 + abM_5 \tag{33}$$

and the minimum MSE of t_9 at $a = \frac{M_3M_5 - 2M_1M_4}{4\,M_2M_4 - M_5^2}$, b = $\frac{M_1M_5 - 2M_2M_3}{4\,M_2M_4 - M_5^2}$ is

$$M(t_9)_{min} = \bar{Y}^2 - \frac{M_1^2M_4 + M_3^2M_2 - M_1M_3M_5}{4M_2M_4 - M_5^2} \tag{34}$$

Let $M_1 = (R_1 + 1.5\,R)C_\Omega - (2R_1^2 + 3.75R^2 + 3RR_1)B_\Omega - 2\bar{Y}^2$,

$M_2 = (3R_1^2 + 6R^2 + 6RR_1)B_\Omega + A_\Omega + \bar{Y}^2 - (3R_1 + 4.5R)C_\Omega$,

$M_3 = -(2R_1 + 3R)\,B_\Omega$, $M_4 = B_\Omega$, $M_5 = (4R_1 + 6R)\,B_\Omega - C_\Omega$, $R_1 = \frac{\delta\,\bar{Y}}{\delta\bar{X}+\eta}$

4. Proposed estimators and its properties

Earlier, several authors described the problem of measurement errors in different sampling strategies and suggested procedures for removing the same. It appears that by using these procedures, measurement errors influence the results in sample surveys. As a result, we expanded the same using the factor-type estimator advocated by Singh and Shukla (1987).

The proposed class of estimators for population mean under Ω as given below:

$$\bar{y}_{FT} = f\bar{y}^* + (1-f)\bar{y} \tag{35}$$

Where $\bar{y}^* = \frac{\bar{y}}{\bar{x}^*}$, $where\ \bar{x}^* = \left[\frac{(A+C)\bar{X}+fB\bar{x}}{(A+fB)\bar{X}+C\bar{x}}\right]^{-1}$

and $A = (m-1)(m-2),\ B = (m-1)(m-4),\ C = (m-2)(m-3)(m-4)$,

$f = \frac{n}{N}$ $and\ m \in (0, \infty)$ is a constant.

Remark 4.1 If sample size is too small ($n \to 0$)

then $\bar{y}_{FT} = \bar{y} = t_1$ (36)

and when the sample size is large enough ($n \to N$)

then $\bar{y}_{FT} = \bar{y}^* = \frac{\bar{y}}{\bar{x}^*}$ (37)

the usual factor-type estimator of Singh and Shukla (1987).

4.1 Some special cases

The $\bar{y}_{FT}$ for first four constant values of *m* is as below:

(i) If *m* = 1, then

$$\bar{y}_{FT} = f\bar{y}\ \frac{\bar{X}}{\bar{x}} + (1-f)\bar{y} = ft_2 + (1-f)t_1 \quad (38)$$

(ii) If *m* = 2, then

$$\bar{y}_{FT} = f\ \bar{y}\ \ \frac{\bar{x}}{\bar{X}} + (1-f)\bar{y} \quad (39)$$

(iii) If *m* = 3, then

$$\bar{y}_{FT} = f\bar{y}\ \left[\frac{\bar{X}-f\ \bar{x}}{(1-f)\bar{X}}\right] + (1-f)\bar{y} \quad (40)$$

(iv) If *m* = 4, then

$$\bar{y}_{FT} = \bar{y} = t_1 \quad (41)$$

5. Properties of proposed estimator

Using the concept of large sample approximation i.e. if $n \to N$

Let $\bar{y} = (1+e_0)\bar{Y} \Rightarrow e_0 = \bar{Y}^{-1}(\bar{y}-\bar{Y})$,

$\bar{x} = (1+e_1)\bar{X} \Rightarrow e_1 = \bar{X}^{-1}(\bar{x}-\bar{X})$,

$E(e_0) = 0$, $E(e_1) = 0$, $E(e_0^2) = A_\Omega\ \bar{Y}^{-2}$, $E\ (e_1^2) = B_\Omega\ \bar{X}^{-2}$, and $E(e_0e_1) = C_\Omega\ \bar{X}^{-1}\ \bar{Y}^{-1}$,

Let $\alpha = \frac{fB}{A+fB+C}$ and $\beta = \frac{C}{A+fB+C}$ and $\theta = \alpha - \beta$

Theorem 5.1: The estimator $\bar{y}_{FT}$ in terms of $e_0\ and\ e_1$ could be expressed as:

$$\bar{y}_{FT} = \bar{Y}[1 + e_0 + f(e_1\theta + e_0e_1\theta - \beta\theta e_1^2)] \quad (42)$$

Proof: We have

$\bar{y}_{FT} = f\bar{y}^* + (1-f)\bar{y}$

where $\bar{y}^* = \frac{\bar{y}}{\bar{x}^*}$, $where\ \bar{x}^* = \left[\frac{(A+C)\bar{X}+fB\bar{x}}{(A+fB)\bar{X}+C\bar{x}}\right]^{-1}$

$\bar{y}_{FT} = f\frac{\bar{y}}{\bar{x}^*} + (1-f)\bar{y}$

$\bar{y}_{FT} = f\bar{Y}(1+e_0)\left[\frac{(A+C)\bar{X}+fB\bar{X}(1+e_1)}{(A+fB)\bar{X}+C\,\bar{X}(1+e_1)}\right] + (1-f)\bar{Y}(1+e_0)$

$$\bar{y}_{FT} = f\bar{Y}(1+e_0)\,(1+\alpha e_1)(1+\beta e_1)^{-1} + (1-f)\bar{Y}(1+e_0) \quad (43)$$

Expanding by Taylor's theorem and neglecting the terms having e_i's degree greater than two. We have

$\bar{y}_{FT} = f\bar{Y}[(\alpha-\beta)e_1 - \beta(\alpha-\beta)e_1^2 + (\alpha-\beta)\,e_0\,e_1] + \bar{Y}(1+e_0)$

$$\bar{y}_{FT} = \bar{Y}[(1+e_0) + f(e_1\theta + e_0e_1\theta - \beta\theta e_1^2)] \quad (44)$$

Theorem 5.2: The bias of $\bar{y}_{FT}$ is $B(\bar{y}_{FT}) = f\theta R\bar{X}^{-1}\,\bar{Y}^{-1}(C_\Omega\bar{X} - \beta B_\Omega\bar{Y})$ (45)

Proof: The bias for the proposed estimator using the result obtained in (5.1) can be written as

$\bar{y}_{FT} - \bar{Y} = \bar{Y}\,[e_0 + f(e_1\theta + e_0e_1\theta - \beta\theta e_1^2)]$

$$B(\bar{y}_{FT}) = E(\bar{y}_{FT} - \bar{Y}) = E[\bar{Y}\,\{e_0 + f(e_1\theta + e_0e_1\theta - \beta\theta e_1^2)\}] \quad (46)$$

Taking expectations and ignoring higher terms of o(n^{-1})

$$B(\bar{y}_{FT}) = f\theta R\bar{X}^{-1}\,\bar{Y}^{-1}(C_\Omega\bar{X} - \beta B_\Omega\bar{Y}) \quad (47)$$

Theorem 5.3: The MSE of $\bar{y}_{FT}$ (up to first order of approximation) is

$$M(\bar{y}_{FT}) = f^2R^2\theta^2B_\Omega + A_\Omega + 2fR\theta C_\Omega \quad (48)$$

Proof: $E(\bar{y}_{FT} - \bar{Y})^2 = E[\bar{Y}\,\{e_0 + f(e_1\theta + e_0e_1\theta - \beta\theta e_1^2)\}]^2$

Taking expectations and ignoring higher terms of o(n^{-1})

$$M(\bar{y}_{FT}) = E(\bar{y}_{FT} - \bar{Y})^2 = f^2R^2\theta^2B_\Omega + A_\Omega + 2fR\theta C_\Omega \quad (49)$$

Theorem 5.4: The minimum MSE $\bar{y}_{FT}$ occurs when $\theta = \frac{-C_\Omega}{fRB_\Omega}$ and expression is:

$$\mathrm{M}(\bar{y}_{FT}) = A_\Omega - \frac{C_\Omega^2}{B_\Omega} \quad (50)$$

Proof: Differentiating MSE $(\bar{y}_{FT})$ in (5.3) with respect to θ and equating to zero (assuming $\theta \neq 0$), we have

$$\theta_{opt} = \frac{-C_\Omega}{fRB_\Omega} = -\mathrm{V}\ (\mathrm{let}) \quad (51)$$

Substituting optimum value of θ in (5.3), the expression of minimum mean squared error is $M(\bar{y}_{FT})_{min} = A_\Omega - \frac{C_\Omega^2}{B_\Omega}$

Note 5.1: The optimality condition $\theta = -V$ provides the equation

AV + (V+1) f B + (V – 1) C = 0 (52)

which is cubic in terms of *m*. One can get maximum three values of *m* for which MSE is optimal. For the best choice of *m*, the following strategy can be adopted:

STEP I: Compute $|B(\bar{y}_{FT})_i|$, i = m_1, m_2, m_3

STEP II: Select i as:

$$|B(\bar{y}_{FT})_i| = min[|B(\bar{y}_{FT})_i|] = min\left[\left|B(\bar{y}_{FT})_{m_1}\right|, \left|B(\bar{y}_{FT})_{m_2}\right|, \left|B(\bar{y}_{FT})_{m_3}\right|\right] \quad (53)$$

Obviously, this strategy provides bias controlled optimum MSE.

Note 5.2: For the pair of values (V, f), one can generate a trivariate table for m_1, m_2, m_3 so as to achieve solution quickly.

Note 5.3: Sometimes one can obtain only one optimum(real) value of *m* and other two complex values. In this case, there will be no choices for minimum bias.

6. Comparisons

Comparison between $\bar{y}_{FT}$ with other estimators under measurement error will be discussed in this section.

6.1 Comparison between $\bar{y}_{FT}$ and t_1

$$M(t_1) - M(\bar{y}_{FT})_{min} = A_\Omega - \left(A_\Omega - \frac{C_\Omega^2}{B_\Omega}\right) = \frac{C_\Omega^2}{B_\Omega} \quad (54)$$

$\bar{y}_{FT}$ is better than $t_1, if\ M(t_1) - M(\bar{y}_{FT})_{min} = \frac{C_\Omega^2}{B_\Omega} > 0$

Therefore, the estimator $\bar{y}_{FT}$ is better than t_1.

6.2 Comparison between $\bar{y}_{FT}$ and t_2

$$M(t_2) - M(\bar{y}_{FT})_{min} = [A_\Omega + R^2 B_\Omega - 2RC_\Omega] - \left(A_\Omega - \frac{C_\Omega^2}{B_\Omega}\right) > 0 \quad (55)$$

$(C_\Omega - RB_\Omega)^2 > 0$

Therefore, estimator $\bar{y}_{FT}$ is better than estimator t_2.

6. 3 Comparison between $\bar{y}_{FT}$ and t_3

$$M(t_3)_{min} - M(\bar{y}_{FT})_{min} = A_\Omega - \frac{C_\Omega^2}{B_\Omega} - \left[A_\Omega - \frac{C_\Omega^2}{B_\Omega}\right] = 0 \quad (56)$$

The estimator $\bar{y}_{FT}$ and estimator t_3 are equal efficient.

6.4 Comparison between $\bar{y}_{FT}$ and t_4

$$M(t_4)_{min} - M(\bar{y}_{FT})_{min} = \left[\bar{Y}^2 - \frac{4\bar{Y}^2 B_\Omega - R^2 B_\Omega C_\Omega^2 + R^2 A_\Omega B_\Omega^2 + \frac{R^4 B_\Omega^3}{16}}{4\bar{Y}^2 B_\Omega + 4A_\Omega B_\Omega - 4C_\Omega^2}\right] - A_\Omega - \frac{C_\Omega^2}{B_\Omega} > 0 \quad (57)$$

$\bar{y}_{FT}$ is better than t_4, if $\bar{Y}^2\,(\bar{Y}^2 - 1) > \left(\frac{R^2 B_\Omega}{8} + A_\Omega - \frac{C_\Omega^2}{B_\Omega}\right)^2$

so, estimator $\bar{y}_{FT}$ is more efficient than estimator t_4.

6.5 Comparison between $\bar{y}_{FT}$ and t_5

$$M(t_5)_{min} - M(\bar{y}_{FT})_{min} = A_\Omega - C_\Omega^2 - (A_\Omega - C_\Omega^2) = 0 \quad (58)$$

The estimator $\bar{y}_{FT}$ and estimator t_5 are equal efficient.

6.6 Comparison between $\bar{y}_{FT}$ and t_6

$$M(t_6)_{min} - M(\bar{y}_{FT})_{min} = \left[\bar{Y}^2 - \frac{4\bar{Y}^4 B_\Omega - R^2\bar{Y}^2 B_\Omega^2 + R^2 A_\Omega B_\Omega^2 + 4R\bar{Y}^2 B_\Omega C_\Omega}{4\bar{Y}^2 B_\Omega + 4A_\Omega B_\Omega - (RB_\Omega - 2C_\Omega)^2}\right] - A_\Omega - \frac{C_\Omega^2}{B_\Omega}) > 0 \quad (59)$$

$\bar{y}_{FT}$ is better than t_6, if $4A_\Omega B_\Omega^2\,(A_\Omega - C_\Omega R) < (2\,C_\Omega^2 - RB_\Omega C_\Omega)^2$

so, estimator $\bar{y}_{FT}$ is more efficient than estimator t_6.

6.7 Comparison between $\bar{y}_{FT}$ and t_7

$$M(t_7) - M(\bar{y}_{FT})_{min} = \left[A_\Omega + \frac{9}{4}R^2 B_\Omega - 3RC_\Omega\right] - A_\Omega - \frac{C_\Omega^2}{B_\Omega} > 0 \quad (60)$$

$\bar{y}_{FT}$ is better than t_7, if $C_\Omega(RB_\Omega)^{-1} > 1.5$

so, estimator $\bar{y}_{FT}$ is more efficient than estimator t_7.

6.8 Comparison between $\bar{y}_{FT}$ and t_8

$$M(t_8)_{min} - M(\bar{y}_{FT})_{min} = A_\Omega - \frac{C_\Omega^2}{B_\Omega} - \left[A_\Omega - \frac{C_\Omega^2}{B_\Omega}\right] = 0 \quad (61)$$

The estimator $\bar{y}_{FT}$ and estimator t_8 are equal efficient.

6.9 Comparison between $\bar{y}_{FT}$ and t_9

$$M(t_9)_{min} - M(\bar{y}_{FT})_{min} = \left[\bar{Y}^2 - \frac{M_1^2 M_4 + M_3^2 M_2 - M_1 M_3 M_5}{4M_2 M_4 - M_5^2}\right] - (A_\Omega - \frac{C_\Omega^2}{B_\Omega}) > 0 \quad (62)$$

$\bar{y}_{FT}$ is better than t_9,if

$(\bar{Y}^2 B_\Omega - A_\Omega B_\Omega - C_\Omega^2)(4M_2 M_4 - M_5^2) > B_\Omega(M_1^2 M_4 + M_3^2 M_2 - M_1 M_3 M_5)$

so, estimator $\bar{y}_{FT}$ is more efficient than estimator t_9.

7. Empirical study

For testing the performance of suggested estimator, a population of size N = 25000 holding study variable Y and auxiliary variable X which is the synthetic records of human heights (in inches) and weights (in pounds) of 18 years old children. This population is taken from statistics online computational resource(SOCR) website.

The characteristics of this population are given as:

$\bar{Y}$ = 127.04839, $\bar{X} = 68.085086$, $S_Y^2 = 135.9765320$, $S_X^2 = 3.616382$,

$\rho = 0.5028585$, $\lambda = 0.0019600$, V = 39.2539870 and

$\phi = 0.7850797$, $\varphi = 38.46890$, ($p_1 = 0.999985$, $p_2 = 0.5319576$),

$q = -5.55051$, ($d_1 = 0.9999843$, $d_2 = 0.0000146$),

($m_1 = 2.05618$, $m_2 = 3.69878$, $m_3 = 4.86821$)

The following steps are followed by simulation procedure:

1) Select a random sample by SRSWOR from the population.
2) Put measurement error u and v in Y and X respectively.
3) After repeating above 60,000 times, it delivers sample estimates $\bar{y}_1, \bar{y}_2, \bar{y}_3 \ldots\ldots \bar{y}_{50000}$.
4) Bias of $\bar{y}$ is obtained by $B(\bar{y}) = \frac{1}{50000}\sum_{i=1}^{50000} (\bar{y}_i - \bar{Y})$
5) M. S. E. of $\bar{y}$ is obtained by $\text{M}(\bar{y}) = \frac{1}{50000}\sum_{i=1}^{50000}(\bar{y}_i - \bar{Y})^2$
6) Efficiency is measured as $e(t) = \frac{M(t_1)}{M(t)} \times 100$, with M(t) the MSE of any estimator t.

The condition bias and MSE of estimators are calculated over 60000 repeated samples drawn by SRSWOR from population N = 25000.

Table 1: Efficiency comparisons of the estimators

Estimator	Bias	MSE	Efficiency
t_1	0.094573339	0.302098176	100
t_2	0.102603524	0.27782892	108.7353239
t_3	0.091206145	0.271933683	111.0925918
t_4	0.090275474	0.279119853	108.2324218
t_5	0.091086293	0.271929184	111.0944299
t_6	0.064859797	0.299997673	100.7001731
t_7	0.129175912	0.281299365	107.3938351

t_8	0.091100643	0.279297223	108.1636877
t_9 (1,1)	0.069585149	0.286823955	105.3252951
t_9 (1,2)	0.069583407	0.286822338	105.3258888
t_9 (1,3)	0.069581714	0.286820767	105.3264657
t_9 (1,4)	0.069580068	0.28681924	105.3270265
t_9 (2,1)	0.069586039	0.286824781	105.3249916
t_9 (2,3)	0.069584272	0.28682314	105.3255941
t_9 (3,1)	0.069586339	0.286825059	105.3248895
t_9 (3,2)	0.069585741	0.286824504	105.3250933
t_9 (3,4)	0.069584563	0.286823410	105.3254949
t_9 (4,1)	0.069586489	0.286825199	105.3248382
t_9 (4,3)	0.069585592	0.286824366	105.3251439
$\bar{y}_{FT}(m_1)$	0.050832545	0.270211850	111.8004915
$\bar{y}_{FT}(m_2)$	0.049880218	0.275985694	109.4615347
$\bar{y}_{FT}(m_3)$	0.049487639	0.268658651	112.4468447

7.1 Almost unbiased estimation:

In terms of expression (5.2), the bias of $\bar{y}_{FT}$ (to the first order of approximation) could be made zero.

$$B(\bar{y}_{FT}) = f\theta R\bar{X}^{-1}\,\bar{Y}^{-1}(C_\Omega\bar{X} - \beta B_\Omega\bar{Y}) = 0 \tag{63}$$

the solution appears either $\theta = 0$

$$\text{or } (C_\Omega\bar{X} - \beta B_\Omega\bar{Y}) = 0 \tag{64}$$

After simplifying (1) and (2) one can obtain cubic equations in the form of m and there will be six possible values of m. For the population under consideration, we get different values of m for unbiased $\bar{y}_{FT}$ as shown in table 2.

Table 2: Almost unbiased estimator $\bar{y}_{FT}$

Values of m	Bias	MSE
$m_1{}'$= 4 = $m_6{}'$	0.0495733	0.3000982
$m_2{}'$ = 3.00081	0.0495945	0.3000815
$m_3{}'$ = 1.99960	0.0485671	0.3008991
$m_4{}'$ = 3.039245	0.0495733	0.3000982
$m_5{}'$= 1.980755	0.0495733	0.3000982

8. Discussion and conclusion

This manuscript discusses the importance of improving estimation under measurement errors set up in survey data. Measurement errors can occur due to various factors such as equipment errors, errors in data recording, false information, self-interest and they can significantly affect survey results, leading to an ill data set. Therefore, it is necessary to develop estimation methods that can adjust for measurement errors. After studying several research papers, the decision was made to extend the factor-type estimator in the setup Ω. Expressions for bias, MSE, and optimum MSE of existing estimators and the suggested estimators are derived under large sample approximation up to first order. An empirical study is conducted using a data set of size 25000, and the bias and optimum MSE of the proposed estimator are compared to the mean per unit estimator, evaluating their efficiency. The factor-type estimator performed better than some other existing estimators, as indicated by the results presented in Table 1. Specifically, the estimator $\bar{y}_{FT}$ shows greater efficiency and a significant reduction in mean squared error compared to other estimators,t_i; i = 1, 2,….9. Thus, the proposed estimator is considered more useful and advantageous. A key feature of the factor-type estimator is that there are multiple values of the m for which the MSE is optimal. By selecting the values with minimum bias, one can obtain an almost unbiased estimator. The proposed strategies focus on bias control at the optimum level of MSE, allowing for the selection of m values that provide the lowest MSE. Consequently, the proposed factor-type estimator is useful and has an advantage over other existing methods.

Overall, this manuscript highlights the significance of accounting for measurement errors in survey data and demonstrates the superiority of the factor-type estimator in addressing this issue.

References

Cochran, W. G. (1968): Errors of measurement in statistics, Technometrics 10, 637–666. doi:10.2307/ 1267450.

Cochran, W.G. (2005): Sampling Techniques, Second edition. Wiley Eastern Private Limited, New Delhi.

Ekpenyong, E. J. and Enang, E. I. (2014): “Efficient exponential ratio estimator for estimating the population mean in simple random sampling,” Hacettepe Journal of Mathematics and Statistics, 44(19), 1

Grover, L. K. and Kaur, P. (2011): “An improved estimator of the finite population mean in simple random sampling,” Model Assisted Statistics and Applications, 6(1), 47–55.

Gupta, R.K. and Yadav, S. K. (2018): Improved estimation of population mean using information on size of the sample, American Journal of Mathematics and Statistics 8(2),27-35

Manisha and Singh, R.K. (2001): An estimation of population mean in the presence of measurement errors, Journal of the Indian Society of Agricultural Statistics 54(1), 13–18.

Singh, H. P. and Pal, S. K. (2015): “A new chain ratio-ratio-type exponential estimator using auxiliary information in sample surveys,” International Journal of Machine Intelligence and Applications, 3(4),37–46, http://www.ijmaa.in/v3n4-b/37-46.pdf.

Shalabh (1997): Ratio method of estimation in the presence of measurement errors, Journal of the Indian Society of Agricultural Statistics, 50(2), 150–55.

Shukla, D., Pathak, S. and Thakur, N.S. (2012a): An estimator for mean estimation in presence of measurement error, Research & Reviews: A Journal of Statistics, 2(1) 1-8.

Shukla, D., Pathak, S. and Thakur, N.S. (2012b): Class(es) of Factor-type estimator in presence of measurement error, Journal of Modern Applied Statistical Methods, 11(2), 336-347.

Singh, H. P. and Karpe, N. (2008): Ratio-product estimator for population mean in presence of measurement errors, Journal of Applied Statistical Sciences 16(4), 49–64.

Singh, V.K. and Shukla, D. (1987): One parameter family of factor-type ratio estimator, Metron, 45, 1-2, 273-283.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Ashok, C. (1984): Sampling Theory of Surveys with Applications, Iowa State University Press, I.S.A.S. Publication, New Delhi.

Tiwari, K. K., Bhougal, S., Kumar, S. and Onyango R. (2022): Assessing the Effect of Nonresponse and Measurement Error Using a Novel Class of Efficient Estimators, Journal of Mathematics 2022 (9)

Statistics online computational resource (SOCR) http://socr.ucla.edu/docs/resources/SOCR_Data/SOCR_Data_Dinov_020108_HeightsWeights.html