

Impact of Measurement Errors on Estimators of Parameters of a Finite Population with Linear Trend Under Systematic Sampling

Oloo Odhiambo Erick^{*}, James Kahiri, Wafula Mike Erick

Department of Statistics and Actuarial Science, Kenyatta University (KU), Nairobi, Kenya

Email address:

Ericoloo58@gmail.com (O. O. Erick)

^{*}Corresponding author

To cite this article:

Oloo Odhiambo Erick, James Kahiri, Wafula Mike Erick. Impact of Measurement Errors on Estimators of Parameters of a Finite Population with Linear Trend Under Systematic Sampling. *American Journal of Theoretical and Applied Statistics*. Vol. 6, No. 6, 2017, pp. 270-277.

doi: 10.11648/j.ajtas.20170606.12

Received: September 17, 2017; **Accepted:** October 4, 2017; **Published:** November 10, 2017

Abstract: The study involves investigating the impact of measurement errors on estimators of parameters of a finite population with linear trend among population values, under systematic sampling. The study provides deep understanding on the amount and nature of deviation introduced by errors and how these errors affect estimators of parameters of a population with linear trend. Consideration is given to measurement errors that assume a normal distribution. Systematic sampling technique is used where a sample of size n is selected randomly from a finite population with a fixed interval a . Systematic sampling is considered instead of simple random sampling in this case because of its effectiveness in dealing with linear trend. The explicit values of population totals, means and variances together with their estimates are derived. The results indicate that there can be overestimate of the population mean if the expected systematic errors tend towards positive values and underestimate if the expected systematic error tend towards negative values. When random errors are considered, there is no effect on estimated population parameters.

Keywords: Finite Population with Linear Trend, Systematic Sampling, Measurement Errors

1. Introduction

In an ideal situation, it is assumed that through some kind of probability sampling, in this case systematic sampling, the observation y_i on the i^{th} unit is the correct value for that unit, and that sampling errors may arise solely from the random sampling variation that is present when n units are measured instead of complete population of N units. Contrary to the assumption, non-sampling errors that are due to measurement or observation do occur at data collection stage.

The true score theory is a good simple model for measurement. It consists of true value and two error components; random error and systematic error.

$$y = t + e_r + \varepsilon_s \quad (1)$$

Where,

y is the measured value

t is the true value

e_r is the random error

ε_s is the systematic error

Random error is caused by unpredictable fluctuation in the reading of a measurement apparatus or experimenter's interpretation of the instrumental reading. Systematic error is caused by any physical factor that affect an experiment or measurement of the variable across the sample in a predictable direction.

According to Fuller and Carroll et al [1, 2], it is well known that measurement errors in observed data can lead researchers to draw incorrect inferences. Much of the early work in this area focused on the typical textbook model of classical measurement error. Bound et al [3] concluded the survey of measurement errors by calling for researchers to pay more attention to the possibility of non-classical measurement error, both in accessing the likely biases in the analyses that take no

account of measurement error and in devising procedure that corrects such error.

In recent years, a number of papers have examined the consequences of non-classical measurement errors in labor economics.

[4-6], all noted that non-classical measurement errors of the type typically found in income data, attenuates the role of white noise measurement error in models of earnings dynamics.

Measurement errors can best be studied if the true value is obtained. This approach is limited to items for which a feasible method for finding the true value exists. For instance majority of studies using Body Mass Index(BMI) rely on self-reported measures from survey data sets. However Conor et al shows that there is a large body of evidence which suggests that self-reported BMI tends to underestimate true BMI; this occurs both because people underreport their weight and overestimate their height [7]. Looking at measurement errors in self-reported BMI specifically, Plankey et al examines the consequences of these errors when classifying people according to obesity status [8]. Stommel et al [9] compared self-reported and recorded BMI using US data and found a substantial amount of misclassification of obesity status when using self-reported BMI, particularly in the extreme (overweight or underweight) categories. Consequences of this measurement errors were examined when analysing the impact of BMI on a range of health risks.

Belloc [10], compared data on hospitalization as reported in household interviews with the hospital records for the individuals. Hospital record produces the true values which are then compared to the observed values from household interviews. Gray [11] compared employee's statements of sick leave with the personal office records. The comparison of data was to determine the presence of measurement errors if any. [12-13] compares respondents' illness with either doctors' records on the respondents or with the results of a complete medical examination.

According to Särndal measurement errors arise during data collection stage, and may have a considerable impact on the estimates [14]. In recent studies; Nyabwanga [15] studied Effect of measurement errors on population in random order. Rosella et al [16] studied the influence of measurement error on calibration, discrimination and overall estimation of risk prediction model. O'Neil et al [17] examined the consequences of measurement errors in self-reported BMI when estimating the relationship between obesity and income. Subramani et al [18] studied Estimation of Populations Means in the Presence of Linear Trend among population values under circular systematic sampling. Ouko et al [19] studied the effects of Measurement Errors on Population Estimates from Samples Generated from a Stratified Population through Systematic Sampling Technique. Grellety et al [20] examined the Effect of random error on diagnostic accuracy illustrated with anthropometric diagnosis of malnutrition.

In this project further study is based on *The Impact of Measurement Errors on Estimators of Parameters for a finite Population with Linear Trend under Systematic Sampling*.

Finite population with linear trend consists of N units identified by the label $1, 2, \dots, N$ ordered in increasing size. Through systematic sampling, the population is then divided into a samples of n units each.

The table below shows sets of all possible samples.

Table 1. Sets of all possible samples.

	Samples, s					
	s_1	s_2	...	s_h	...	s_a
	y_1	y_2	...	y_h	...	y_a
y values	y_{a+1}	y_{a+2}	...	y_{a+h}	...	y_{2a}

	$y_{(n-1)a+1}$	$y_{(n-1)a+2}$...	$y_{(n-1)a+h}$...	y_{na}
Sample totals	t_{s1}	t_{s2}	...	t_{sh}	...	t_{sa}

Let systematic sampling be denoted as SY.

The sets of all possible samples denoted S_{sy} consists of a different sets that are non-overlapping that can be obtained as

$$S_{sy} = (s_1, \dots, s_i, \dots, s_a) \quad (2)$$

Sampling design of SY is thus given as

$$p(s) = \begin{cases} \frac{1}{a} & \text{if } s \text{ belongs to } S_{sy} \\ 0 & \text{otherwise} \end{cases}$$

Each sample or a cluster is selected with probability $1/a$ and observed completely as per the design.

2. Parameter Estimation

2.1. Estimation of Parameters for a Finite Population with Linear Trend Under Systematic Sampling

Let N be the size of a finite population. Suppose the finite population is such that the observed values assume a hypothetical trend as

$$y_i = \mu + \beta i \quad (3)$$

where μ and β are constants and $i = 1, 2, \dots, N$ are ordered in increasing size of the label

The population is then said to possess linear trend among its values.

Let population size N be a multiple of n , $N = an$. The estimate of population mean under linear systematic sampling in which case a single random start is taken is obtained. as

$$\begin{aligned} \bar{Y} &= \frac{1}{N} \sum_{i=1}^N y_i \\ &= \frac{1}{N} \sum_{i=1}^N (\mu + \beta i) \\ &= \mu + \frac{N+1}{2} \beta \end{aligned} \quad (4)$$

According to Mukhopadhyay [20], population with linear trend has a systematic sample given by

$$\mu + \beta[r + (l-1)a] \text{ , where } r \text{ is the random start,}$$

$l = 1, \dots, n$

The sample total

$$t_s = \sum_{i=1}^n \left\{ \mu + \beta \left[r + (l-1)a \right] \right\} \quad (5)$$

Then the sample mean is written as

$$\bar{y}_s = \mu + \beta \left(r + a \frac{n-1}{2} \right) \quad (6)$$

\bar{y} is the mean of a systematic sample and through probability sampling, it is the unbiased estimator of \bar{Y}

Similarly, population total is denote as $t_\theta = \sum_U y_i$

Since the interest is in estimating population total, from a design-based approach, Horvitz-Thompson estimator, HTE, Horvitz and Thompson [22] is used.

The estimator is defined as,

$$\hat{t}_\pi = \sum_U \frac{I_i y_i}{\pi_i} = \sum_s \frac{y_i}{\pi_i} \quad (7)$$

Where $\pi_i > 0$ is the first order inclusion probability.

Under Systematic Sampling, SY design with sampling interval a , and the response variable y_i , the population total estimator for $\pi_i = p(I_i = 1) = \frac{1}{a}$ is given as

$$\hat{t}_\pi = \sum_s \frac{y_i}{\pi_i} = at_s \quad (8)$$

Where $\pi_i = p(I_i = 1) = \frac{1}{a}$ is the probability that the i^{th} unit of the population is included in the sample.

Finite population variance is given as

$$\begin{aligned} S^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2 \\ &= \frac{N(N+1)}{12} \beta^2 \end{aligned}$$

Variance of mean from a simple random sample is given as

$$\begin{aligned} V_{ran}(\bar{y}) &= \frac{N-n}{N} \frac{s^2}{n} \\ &= \frac{(a-1)(na+1)}{12} \beta^2 \end{aligned}$$

Under systematic sampling, the variance of mean is given as

$$Var(\bar{y}_{sy}) = \frac{(a^2 - 1)}{12} \beta^2 \quad (9)$$

According to Daroga et al [23], for removing effect of linear

trend, systematic sampling is much more efficient than simple random sampling.

2.2. Estimation of Variance from a Single Systematic Sample

According to Särndal [14], a major drawback of SY is that there is no unbiased estimator for the variance of the estimator of population mean except for some cases of circular systematic sampling. This is because SY is equivalent to cluster sampling with only one cluster selected.

However, under some assumptions about the nature of the population, it is possible to propose estimators that are approximately unbiased for the design variance.

In this case, the appropriate variance estimator for the population with linear trend in which case values of the units are steadily increasing by a constant amount is considered.

Many biased estimators have been proposed for this kind of population;

Wolter [22], made analytical studies on population with linear trend and proposed the following estimator for the variance of the estimator of population total.

Assume $n = 2m$. Since a systematic sample can be looked upon as grouping the population in m groups and choosing 2 units from each group of size $2a$, an estimator of the mean of the g^{th} group is

$$\bar{y}_g = \frac{y_{i2-1} - y_{i2g}}{2}, g = 1, \dots, m$$

with the variance estimator

$$v = \frac{a-1}{a} \left(\frac{y_{i2g-1} - y_{i2g}}{2} \right)^2$$

Hence estimator of $V_{sy}(\hat{t}_\pi)$ is

$$\begin{aligned} v_1 &= N^2 \hat{v} \left(\frac{1}{m} \sum_g \bar{y}_g \right) \\ v_1 &= N^2 \frac{1-f}{n^2} \sum_{g=1}^{\frac{n}{2}} (y_{i2g-1} - y_{i2g})^2, f = \frac{n}{N} \end{aligned} \quad (10)$$

Cochran [25], suggested the estimator below to be appropriate

$$v = \hat{v}(\hat{t}_\pi) = N^2 \frac{a-1}{a} \frac{n'}{n^2} \frac{\sum_{i=1}^{n-2} (y_i - 2y_{i+1} + y_{i+2})^2}{6(n-2)}$$

for

$$1 < i \leq n-2$$

Where $S_{sy}^2 = \frac{\sum_{i=1}^{n-2} (y_i - 2y_{i+1} + y_{i+2})^2}{6(n-2)}$ is the variance of a systematic sample.

a is the sampling interval.

The estimate is based on successive quadratic terms in the sequence y_i . The factor 6 is the sum of squares of the coefficients in the $n-2$ differences. The term $\frac{n'}{n^2}$ is the sum of squares of the weights in \bar{y}_{swy} .

n' is due to the weight $\frac{2r-a-1}{2(n-1)a}$ applied to the first and the last sample values.

Unless n is small, $\frac{n'}{n^2}$ can be replaced by the factor $\frac{1}{n}$.

Thus

$$v_2 = \hat{V}(\hat{t}_\pi) = N^2 \frac{a-1}{an} \frac{\sum_{i=1}^{n-2} (y_i - 2y_{i+1} + y_{i+2})}{6(n-2)}, \quad (11)$$

for $1 < i \leq n-2$

Yates [26], suggested the following estimator among others based successively on second and higher order differences.

$$v_3 = N^2 \frac{1-f}{3.5n(n-4)} \sum_{i=1}^{n-4} \left(\frac{y_i}{2} - y_{i+1} + y_{i+2} - y_{i+3} + \frac{y_{i+4}}{2} \right)^2 \quad (12)$$

The sum of squares contains $(n-4)$ terms.

2.3. Population Total Estimator and Its Variance in Presence of Random Errors

Through measurement procedure, the i^{th} individual observed is accompanied by the random error term e_i .

The observed value is thus given as

$$y = \mu + \beta i + e_i \quad (13)$$

The model can thus be expressed as

$$y_i = \mu + e_i \quad (14)$$

Where

y_i - is the observed value for the i^{th} individual,
 $\mu + \beta i = \mu_i$ - is the true value for the i^{th} individual,
 e_i - is the random error for the i^{th} individual.
 For the true function value μ_i , population total is

$$t_\theta = \sum \mu_i$$

Define $\hat{t}_{\pi/s} = \sum_{i=1}^N \frac{I_i y_i}{\pi_i} = \frac{\sum_s y_i}{\pi_i}$ as the population total

estimator (Horvitz-Thompson estimator), with $\pi_i > 0$

In SY design, $\pi_i = p(I_i = 1) = \frac{1}{a}$

The joint expectation of the population total estimator is

obtained as;

$$E_{pm}(\hat{t}_\pi) = E_p \left[E_m(\hat{t}_\pi/s) \right] = \sum_U \mu_i$$

The total variance of \hat{t}_π with respect to sampling design $p(\cdot)$ and the measurement model m according to Särndal [14] is given as

$$V_{pm}(\hat{t}_\pi) = E_p \left[V_m(\hat{t}_\pi/s) \right] + V_p \left[E_m(\hat{t}_\pi/s) \right] \quad (15)$$

The variance of population estimator is the sum of the expected value of the conditional variance and the variance of conditional expected value.

Therefore total variance consists of measurement variance and sampling variance respectively.

Measurement variance when decomposed is expressed as follows;

$$E_p \left[V_m(\hat{t}_\pi/s) \right] = a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij}$$

Sampling variance when decomposed is also expressed as

$$V_p \left[E_m(\hat{t}_\pi/s) \right] = \sum_{ij} \sum (a-1) \mu_i \mu_j$$

Combining the results we have

$$V_{pm}(\hat{t}_\pi) = a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij} + \sum_{ij} \sum (a-1) \mu_i \mu_j \quad (16)$$

This is the derived equation for total variance in presence of random error with respect to sampling design $p(\cdot)$ and measurement model m jointly.

The measurement variance $a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij}$ has simple response variance and the correlated response variance as its components respectively.

$\sum_{ij} \sum (a-1) \mu_i \mu_j$ is the sampling variance. Simple response variance reflects the random variation in a respondent's answer to a survey over repeated measurements.

Correlated response variance also known as interviewer variance occurs because response errors are correlated for sample units interviewed by the same interviewer.

2.4. Mathematical Model for Errors of Measurement

Suppose measurement could be independently repeated many times on unit $i \in s$, we could generate different y_i -values.

Let y_i be the realized value in the repeated observation, then

$$y_i = \mu + \beta i + e_i + \varepsilon_i$$

Let $\mu + \beta i = \mu_i$

Thus

$$y_i = \mu_i + e_i + \varepsilon_i \quad (17)$$

μ_i is the true value of unit i

Both e_i and ε_i are random variables where $e_i (i \in s)$ are independent random variables with $E(e_i) = 0$ and variance δ_e^2 .

$\varepsilon_1, \dots, \varepsilon_a$ are independent and identically distributed random variables with the expected value ε and variance δ_ε^2

The random variables $\varepsilon_i (i = 1, \dots, a)$ are independent of the random variables $e_i (i \in s)$

From the selected sample consisting of y_i -values, the first and second-order moments are;

$$E_m(y_i/s) = E(\mu_i) + E(e_i) + E(\varepsilon_i) = \mu_i + \varepsilon = \alpha_i$$

$$V_m(y_i/s) = \delta_e^2 + \delta_\varepsilon^2 = \delta_i^2$$

$$\text{Cov}_m(y_i/s) = C_m(\varepsilon_i, \varepsilon_j) = \delta_{ij}$$

ε_i is the systematic error on the individual measured value and $E(\varepsilon_i) = \varepsilon$ is the bias in measurement.

Unlike random error, systematic error tend to be consistently either positive or negative - because of this, systematic error is sometimes considered to be bias in measurement.

2.5. Measurement Bias and Expectation of π -estimator

Measurement bias arises when expected measurement value on elements do not agree with true element values.

$$B_{pm}(\hat{t}_\pi) = E_{pm}(\hat{t}_\pi) - t_\theta$$

The derived expected measurement value is expressed as

$$E_{pm}(\hat{t}_\pi/s) = t_\theta + \sum_U \varepsilon$$

The derive total measurement bias with respect to sampling design $p(\cdot)$ and measurement model m respectively is thus expressed as follows

$$B_{pm}(\hat{t}_\pi) = t_\theta + \sum_U \varepsilon - t_\theta = \sum_U \varepsilon \quad (18)$$

$$V_{pm}(\hat{t}_\pi) = a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij} + (a-1) \sum_{i,j} (\mu_i \mu_j) + (a-1) \sum_{i,j} (\mu_i \varepsilon_l + \mu_j \varepsilon_k + \varepsilon_k \varepsilon_l) \quad (22)$$

The term $(a-1) \sum_{i,j} (\mu_i \varepsilon_l + \mu_j \varepsilon_k + \varepsilon_k \varepsilon_l)$ consists of

2.6. Decomposing Variance of Population Total Estimator When Systematic Errors Are Present

According to Särndal [14] the total variance of \hat{t}_π with respect to sampling design $p(\cdot)$ and the measurement model m is given as

$$V_{pm}(\hat{t}_\pi) = E_p[V_m(\hat{t}_\pi/s)] + V_p[E_m(\hat{t}_\pi/s)] \quad (19)$$

Measurement variance when decomposed is expressed as follows;

$$E_p[V_m(\hat{t}_\pi/s)] = a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij} \quad (20)$$

Similarly, sampling variance is decomposed as follows

$$V_p[E_m(\hat{t}_\pi/s)] = V_p[E_m(a \sum_s y_i)]$$

$$E_m(a \sum_s y_i) = a \sum_s E_m(y_i) = a \sum_s \alpha_i$$

Now

$$\begin{aligned} V_p[E_m(a \sum_s y_i)] &= V_p[a \sum_s \alpha_i] \\ &= \sum_U (a \alpha_i)^2 V_p(I_i(s)) + \sum_{i \neq j} a^2 \alpha_i \alpha_j \text{cov}(I_i(s), I_j(s)) \\ &= \sum_U a^2 \pi_i (1 - \pi_i) \alpha_i^2 + a^2 (\pi_{ij} - \pi_i \pi_j) \sum_{i \neq j} \alpha_i \alpha_j \\ &= \sum_U \frac{a^2 (a-1)}{a^2} \alpha_i^2 + \sum_{i \neq j} \frac{a^2 (a-1)}{a^2} \alpha_i \alpha_j \\ &= \sum_U (a-1) \alpha_i^2 + \sum_{i \neq j} (a-1) \alpha_i \alpha_j \\ &= \sum_{i,j} (a-1) \alpha_i \alpha_j \end{aligned}$$

$$V_p[E_m(\hat{t}_\pi/s)] = \sum_{i,j} (a-1) \alpha_i \alpha_j$$

Combining the results the total variance becomes

$$V_{pm}(\hat{t}_\pi) = a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij} + \sum_{i,j} (a-1) \alpha_i \alpha_j \quad (21)$$

We let $\alpha_i = \mu_i + \varepsilon_k$ and $\alpha_j = \mu_j + \varepsilon_l$

Substituting 17 & 18 into 16 above

$$\begin{aligned} V_{pm}(\hat{t}_\pi) &= a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij} + \sum_{i,j} (a-1) (\mu_i + \varepsilon_k) (\mu_j + \varepsilon_l) \\ &= a \sum_U \delta_i^2 + a \sum_{i \neq j} \delta_{ij} + (a-1) \sum_{i,j} (\mu_i \mu_j + \mu_i \varepsilon_l + \mu_j \varepsilon_k + \varepsilon_k \varepsilon_l) \end{aligned}$$

systematic error and it causes variation in total variance.

3. Numerical Results and Discussion

A finite population of size N is generated for a population without errors, a population with random errors and a population with systematic errors.

The population total variances, the population means and the population totals are then computed.

In the selection of a systematic sample of size n , a random start r is selected between 1 and a inclusive in which case a is the sampling interval.

To estimate parameters, simulation of data is done 10 times in each case the estimate is obtained. The results are then averaged to get the estimates of all parameters required in the study.

Estimation of variance is done using the three estimators

Case 1

Let $N=800$, $n=32$, $a=25$

below simplified to reflect systematic sampling.

$$v_1 = a(a-1) \sum_{g=1}^{n/2} (y_{i2g-1} - y_{i2g})^2 \quad (23)$$

$$v_2 = N(a-1) \frac{\sum_{i=1}^{n-2} (y_i - 2y_{i+1} + y_{i+2})^2}{6(n-2)} \text{ for } (1 \leq i \leq n-2) \quad (24)$$

$$v_3 = N(a-1) \frac{\sum_{i=1}^{n-4} \left(\frac{y_i}{2} - y_{i+1} + y_{i+2} - y_{i+3} + \frac{y_{i+4}}{2} \right)^2}{3.5(n-4)} \text{ for } (1 \leq i \leq n-2) \quad (25)$$

The tables from case 1 to case 3 consists of parameters and their estimates for populations without errors and populations with errors.

Table 2. Population Parameters and Their Estimates for $e_r \sim N(0,1)$ and $\varepsilon_i \sim N(-0.6,1.5)$.

Pop	Parameters			Estimates				
	Mean	Total	Pop.var	Estimated mean	Estimated total	V_1	V_2	V_3
No error	423.525	338, 820	36, 691, 200	422.475	337, 980	25, 412, 184	0	0
With e_r	423.530	338, 824	36, 716, 492	421.102	336, 880	25, 556, 994	18, 875	18, 956
With ε_i	422.940	338, 352	37, 594, 498	418.767	335, 013	26, 027, 490	50, 734	52, 305

Table 3. Population Parameters and Their Estimates for $e_r \sim N(0,1)$ and $\varepsilon_i \sim N(0.5,1.5)$.

Pop	Parameters			Estimates				
	Mean	Total	Pop.var	Estimated mean	Estimated total	V_1	V_2	V_3
No error	423.525	338, 820	36, 691, 200	422.475	337, 980	25, 412, 184	0	0
With e_r	423.563	338, 824	36, 883, 067	420.752	336, 601	25, 662, 927	17, 652	18, 720
With ε_i	424.020	339, 216	36, 740, 158	427.770	342, 216	25, 450, 518	44, 746	37, 078

From tables 2 & 3, the results shows that:

1. The population mean is underestimated if a population with systematic errors has a negative systematic bias.
2. The population mean is overestimated if a population with systematic errors has a positive systematic bias.
3. Population mean in presence of random errors almost

Case 2

Let $N=800$, $n=40$, $a=20$

conforms to the mean of population without errors.

4. Estimators of variance of population total estimator underestimate the population total variance.
5. v_2 and v_3 for population with true values are zeros but very small values for population with errors.

Table 4. Population Parameters and Their Estimates for $e_r \sim N(0,1)$ and $\varepsilon_i \sim N(-0.6,1.5)$.

Pop	Parameters			Estimates				
	Mean	Total	Pop.var	Estimated mean	Estimated total	V_1	V_2	V_3
No error	423.525	338, 820	23, 461, 200	425.625	340, 500	12, 744, 459	0	0
With e_r	423.480	338, 824	23, 645, 877	424.97	339, 976	12, 755, 957	16, 807	17, 127
With ε_i	422.940	338, 352	23, 446, 698	421.586	337, 268	12, 817, 490	67, 835	62, 6298

Table 5. Population Parameters and Their Estimates for $e_r \sim N(0,1)$ and $\varepsilon_i \sim N(0.5,1.5)$.

Pop	Parameters			Estimates				
	Mean	Total	Pop.var	Estimated mean	Estimated total	V_1	V_2	V_3
No error	423.525	338, 820	23, 461, 200	426.938	341, 550	12, 744, 459	0	0
With e_r	423.545	338, 824	25, 514, 150	420.900	336, 720	12, 706, 411	17, 792	18, 229
With ε_i	424.020	339, 216	23, 419, 259	424.404	339, 523	12, 812, 648	37, 570	33, 006

From tables 4 & 5, the results shows that:

1. The population mean is underestimated if the expectation of systematic errors is negative.

2. The population mean is overestimated if the expectation of systematic errors is positive.

3. Estimates from v_1 for population with systematic errors

Case 3

Let $N=800$, $n=100$, $a=8$

exceeds the corresponding estimates from the population without errors.

4. Estimates from v_1 exceeds estimates from v_2 and v_3 .

5. The mean of population with random errors is closer to the actual population mean.

Table 6. Population Parameters and Their Estimates for $e_r \sim N(0,1)$ and $\varepsilon_i \sim N(-0.6,1.5)$.

Pop	Parameters			Estimates				
	Mean	Total	Pop.var	Estimated mean	Estimated total	V_1	V_2	V_3
No error	423.525	338, 820	3, 764, 400	422.77	338, 216	694, 575	0	0
With e_r	423.230	338, 824	3, 663, 845	423.589	338, 871	698, 297	6, 509	6, 787
With ε_i	422.940	338, 352	3, 680, 080	422.702	338, 161	723, 733	19, 702	19, 937

Table 7. Population Parameters and Their Estimates for $e_r \sim N(0,1)$ and $\varepsilon_i \sim N(0.5,1.5)$.

Pop	Parameters			estimates				
	Mean	Total	Pop.var	Estimated mean	Estimated total	V_1	V_2	V_3
No error	423.525	338, 820	3, 764, 400	423.42	348, 736	694, 575	0	0
With e_r	423.530	338, 824	3, 616, 050	422.816	338, 252	692, 4021	6, 887	7, 490
With ε_i	424.020	338, 216	3, 677, 402	424.586	339, 668	703, 750	18, 466	16, 872

From tables 6 & 7, the results shows that:

1. When the sample size is increased, both the population variances and estimated variances are reduced.

2. Estimates from v_1 are much higher than the respective estimates from v_2 and v_3 .

3. Positive expected systematic errors overestimate population means and totals while negative expected systematic errors underestimate population means and total.

4. Summary

From the study, it is observed that:

1. The population means and hence the population totals are overestimated for the case where expectation of systematic errors is positive.

2. The population means and hence the population totals are underestimated for the case where expectation of systematic errors is negative.

3. Impact of random errors on population mean and population total is minimal and inconsistent.

4. The variances of population total estimator are all underestimated using the three estimators, v_1 , v_2 and v_3 .

5. Increase in sample size leads to decrease in estimated variance of population total estimator.

6. For population with systematic errors, the estimated variances are over represented. Estimator v_1 gives higher variance than estimators v_2 and v_3 .

5. Conclusions

The study has shown that:

Impact of random errors on population mean, population

total and estimated variance of population total estimator is very minimal.

Systematic errors produces systematic bias that overestimate the population mean when the bias is positive and underestimate the population mean when the bias is negative.

All the three estimators underestimate population variances and therefore they are biased. Among the three, v_1 is better because it gives values closer to the population variance.

Generally systematic errors lead to over representation of the estimated variance while random errors have no impact on estimates of population variance.

References

- [1] Fuller, W. (1987). *Measurement Error Models*. Wiley and Sons.
- [2] Carroll, R., J. R. D. and Stefanski, L. (1994). *Measurement Error in nonlinear models*, Chapman and Hall, London.
- [3] Bound, J., Brown, C., and Mathiowetz, N (2001). Measurement Error in Survey data. *American Journal of Theoretical and Applied Statistics*, 5.
- [4] Pischke, J. (1995). Measurement error and earnings dynamics: some estimates from the psid validation study. *Journal Business of Economics Statistics*, 13(3):305-314.
- [5] O' Neil, D., Sweetman, O., and Van der gaer, D. (2007). The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility. *Journal of Economic Inequality*, 5(2):159-178.
- [6] Gottschalk, P. and Huynh, M. (2010). Are earnings inequality and morbidity overstated? The Impact of Nonclassical Measurement Error. *Review of Economics and Statistics*, 92(2): 302-315.

- [7] Conor, G., Temblay, M., Moher, D., and Gorver, B. (2007). A comparison of Direct vs Self-reported Measures for Assessing Height, Weight and Body Mass Index: a systematic review. *Obesity Rev.* 8:307-326.
- [8] Plankey, M., Stevens, J., Flegal, K., and Rust, P. (1997). Prediction equations do not eliminate systematic error in self-reported body mass index. *Obesity Research*, 5(4):308-314.
- [9] Stommel, M. Schoenborn, C. (2009). Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the nhanes and nhis 20012006, *BMC Public Health*, 9(421).
- [10] Belloc, N. (1954). Validation of morbidity survey data by comparison with hospital records, *Journal of American Statistics Associations*, 49:832-846.
- [11] Gray, P. (1955). The Memory Factor in Social Survey. *Journal of American Statistics Associations*, 50:344-363.
- [12] Sagen, O. K, D. R and Simmons, W (1959). Health Statistics from record sources and household interview compared. *Proceedings of the social Statistics election of American Statistics Associations*, pages 6-15.
- [13] Trusell, R and Elinson, J. (1959). *Chronic Illness in a Large City*. Harvard University Press Cambridge Mass.
- [14] Särndal, C. (1992). *Model assisted Survey Sampling*. Springer-verlag New York, Inc, USA.
- [15] Nyabwanga, R. (2010). *Effect of measurement errors on population in random order when sampling systematically*. Unpublished project department of Mathematics Kenyatta University.
- [16] Rosella, L. C. Corey, P. Stukel, T. Mustard, C. Hux, J. and Manuel, D. G. (2012). The influence of measurement error on calibration, discrimination and overall estimation of a risk prediction model. *Population Health Metr*; 10:20. doi:10.1186/1478-7954-10-20. [PMC 3545925].
- [17] O' Neil, D. and Olive, S. (2013). The consequences of measurement error when estimating the impact of obesity on income. *IZA Journal of Labor Economics*, 2(3).
- [18] Subramani, J. and Singh, S. (2014). Estimation of population mean in the presence of linear trend. *Communications in the Statistics-Theory and Methods*, 43.
- [19] Ouko, A., Cheruiyot, W., and Emily, K. (2014). Effects of measurement errors on population estimates from samples generated from a stratified population through systematic sampling technique. *Expert Journal of Economics*, 2:120-132.
- [20] Grellety, E. Golden, M. H (2016). The effect of Random error on diagnosis accuracy illustrated with the anthropometric diagnosis of malnutrition. *PLoS ONE* 11(12): e0168585 doi 10.1371.
- [21] Mukhopadhyay, P. (1998). *Theory and Methods of Survey Sampling*. Prentice-Hall of India Private Ltd.
- [22] Horvitz, D. and Thompson, D. (1952). A generalization of Sampling Without Replacement From a Finite Universe. *Journal of American Statistics Associations*, 47:663-685.
- [23] Daroga, S. and Chaudhary, F. S (1986). *Theory and Analysis of Sample Survey Designs*. New Age International (P) Ltd.
- [24] Wolter, K. (1985). *Introduction to Variance Estimation*, Springer-Verlag, New York.
- [25] Cochran, W. G. (1977). *Sampling Techniques*, John Wiley and Sons, third edition.
- [26] Yates, F. (1948). Systematic sampling, *Philosophical Transaction of the Royal society of London*, A241:345-377.