

Effects of Measurement Errors on Population Estimates from Samples Generated from a Stratified Population through Systematic Sampling Technique

Abel OUKO^{1*}, Cheruiyot W. KIPKOECH², Emily KIRIMI³

¹The East African University, Kenya

²Maasai Mara University, Kenya

³Technical University of Kenya

In various surveys, presence of measurement errors has led to misleading results in estimation of various population parameters. This study indicates the effects of measurement errors on estimates of population total and population variance when samples are drawn using systematic sampling technique from a stratified population. A finite population was generated through simulation. The population was then stratified into four strata followed by generation of ten samples in each of them using systematic sampling technique. In each stratum a sample was picked at random. The findings of this work indicated that systematic errors affected the accuracy of the estimates by overestimating both the population total and the population variance. Random errors only added variability to the data but their effect on the estimates of the population total and population variance was not that profound.

Keywords: Measurement errors, Systematic Errors, Stratified population, Population Total and Population Variance

JEL Classification: P42, R23

1. Introduction

The ultimate goal of each survey is to obtain information about the population under study. The theory of sample survey as developed during the past several decades provides us with various kinds of scientific tools for drawing samples and making valid inference about the population parameters of interest. According to Koninj (1973), in measurements as the best available. Measurement errors cannot be completely eliminated but minimized to an extent which their effects on survey results are not that profound. Basic contributions to the methodology of measurement error models were given by Mahalanobis (1946), Hansen (1946) and Sukhatme and Seth (1952) have examined the question of non sampling errors in census and survey work and they have furnished mathematical models for such errors. The objective of this study was to investigate the effects of measurement errors on the estimates of population total and population variance when samples are drawn from a stratified population using systematic sampling technique. The contribution of this study is to establish more weight as to why systematic errors should be minimized if at all valid results are to be obtained.

* Corresponding Author:

Abel Ouko, Department of Mathematics, The East African University, Kenya

Article History:

Received 24 October 2014 | Accepted 4 November 2014 | Available Online 25 November 2014

Cite Reference:

Ouko, A., Kipkoech, C.W., and Kirimi, E., 2014. Effects of Measurement Errors on Population Estimates from Samples Generated from a Stratified Population through Systematic Sampling Technique. *Expert Journal of Economics*, 2(3), pp. 120-132

2. Systematic sampling

This is a sample selection technique in which sample members are selected from a given population according to a random starting point and a fixed periodic interval. Systematic sampling is still thought of as being random, as long as the periodic interval is determined beforehand and the starting point is random. A common way of selecting members into the sample using systematic sampling is simply by dividing the total number of units in the population by the desired number of units for the sample. The result of the division serves as the marker for selecting sample units from within the given population. Systematic sampling is to be applied only when the given population is logically homogeneous because systematic sample units are uniformly distributed over the population. In some cases systematic sampling is preferred since it spreads the sample more evenly over the population and easier to conduct.

Table 1. Composition of the k systematic samples.

Sample no.	1	2	...	i	...	K
	y_1	y_2	...	y_i	...	y_k
	y_{k+1}	y_{k+2}	...	y_{k+i}	...	y_{2k}

	$y_{(n-1)k+1}$	$y_{(n-1)k+2}$...	$y_{(n-1)k+3}$...	y_{nk}
Means	\bar{y}_1	\bar{y}_2	...	\bar{y}_i	...	\bar{y}_k

2.1. Stratified Systematic sampling.

This is whereby the finite population under study is divided into relatively homogeneous groups referred to as strata and then systematic sampling is carried out in each stratum to generate samples.

Notations:

N_h - Total number of units in stratum h where $\{h=1, 2, \dots, H\}$

n_h - Number of units in a sample drawn from stratum h

μ_{hk} - The true value of the k^{th} unit in stratum h .

y_{hk} - is the observed value of the k^{th} in stratum h .

Note that $N = N_1 + N_2 + \dots + N_H$

3. Sampling design

In the theory of finite population sampling, a sampling design specifies for every possible sample its probability of being drawn. It is convenient to have special notation for this probability which in this case will be $P(s)$.

In other words we assume there is a function $P(\cdot)$ such that $P(s)$ gives the probability of selecting specified samples under the scheme in use. The function $P(\cdot)$ will be referred to as sampling design.

4. The Simple Measurement Model

In this case we would like to formulate a statistical model for measurements made on elements of a sample from a finite population. Consider a finite population, $U = \{1, \dots, k, \dots, N\}$. It is assumed that for each element $k \in U$, there exists a true value μ_k and that the objective is to estimate the population total of these true values,

$$t_\mu = \sum_U \mu_k$$

A sample s of size n_s is selected from U by a given probability sampling design $p(\cdot)$. The idea is to obtain the true value μ_k for each element $k \in s$, but what we actually obtain through the measurement procedure are the observed values y_k for $k \in s$. The observed y_k is composed of the true value μ_k and a random error or both random error and systematic error $y_k - \mu_k$. For lack of better values, y_k is used in the estimation. For a given sample s , the random variables $y_k (k \in s)$ are assumed to have a certain joint probability distribution (conditional on s), called a measurement model denoted by m . In this case we consider our survey as a two stage process whereby the first stage involves the sample selection, which results in a selected sample s and the second stage involves the measurement procedure, which generates an observed value y_k for each $k \in s$. When evaluating expectations and variances with respect to the two stages jointly, the conditional argument is useful. As for the expected values,

$$E_{pm}(\cdot) = E_p \left[E_m(\cdot / s) \right]$$

Where $E_m(\cdot / s)$ denotes conditional expectation with respect to the measurement model m , for a given sample s , $E_p(\cdot)$ denotes expectation with respect to the sample design $p(\cdot)$ and $E_{pm}(\cdot)$ denotes expectation with respect to sampling design and measurement model jointly. Similarly, for the joint variance, called the *pm-variance* or the *total variance*, we have

$$V_{pm}(\cdot) = E_p \left[V_m(\cdot / s) \right] + V_p \left[E_m(\cdot / s) \right]$$

Where $V_m(\cdot / s)$ denotes conditional variance with respect to the model m , given s , $V_p(\cdot)$ denotes variance with respect to $P(\cdot)$ and $V_{pm}(\cdot)$ denotes variance with respect to $P(\cdot)$ and m jointly.

We specify further the model m . For element k and l belonging to the same sample s , the first and second moments are

$$\mu_k = E_m(y_k / s)$$

$$\sigma_k^2 = V_m(y_k / s)$$

$$\text{And } \sigma_{kl} = C_m(y_k, y_l / s)$$

5. Measurement errors.

Measurement is the basis of any scientific study. All measurements are, however, approximate values (not true values) within the limitation of measuring device, measuring environment, process of measurement and human error. Several measurements of the same quality on the same subject will not in general be the same.

Measurement errors refer to errors in survey responses arising from the method of data collection, the respondent or the questionnaire. They include the errors in a survey response as a result of respondents confusion, ignorance, carelessness, or dishonesty; the errors attributable to the interviewer, perhaps as a consequence of the poor or inadequate training, prior expectations regarding respondents' response etc. These measurement errors end up causing a considerable effect on survey estimates. These errors are broadly classified in two categories which are systematic errors and random errors

Systematic errors are biases in measurement which lead to the situation where the mean of many separate measurements differs significantly from the actual value of the measured attribute. All measurements are prone to systematic errors, often of several different types. The errors of this category are characterized by deviation in one direction from the true value. Systematic errors may result from; Usage of faulty instrument, Usage of faulty measuring process and Personal bias. Clearly this type of error cannot be minimized by repeated measurements. Systematic errors can therefore lead to either overestimation or underestimation of the desired population parameters.

Random errors are errors in measurement that lead to measurable values being inconsistent when repeated measures of a constant attribute or quantity are taken. Random errors unlike systematic errors are not unidirectional i.e. some measurements are higher while some are lower than the true value. Another distinguishing aspect of random errors is that it is not biased. It is normally present because of the limitation of the instrument in hand and the limitation of the part of the human ability. No human being can repeat an

action in exactly the same manner. Hence it is likely that the same person reports different values with the same instrument, which measures quality correctly. This error is caused by any factor that randomly affects the sample. Random errors add variability to the data but do not affect the average performance for the group. This is why at times it is regarded as ‘noise’.

5.1. Mathematical model for measurement errors.

According to Cochran (1977), we assume a large number of independent repetitions of the measurement on the i th unit are possible. Let $y_{i\alpha}$ be the value obtained in the α^{th} repetition.

Then $y_{i\alpha} = \mu_i + e_{i\alpha}$
 Where $\mu_i =$ true value
 $e_{i\alpha} =$ measurement error

Where the expectation of e_{ij} is zero and variance is σ_i^2 .

$$E[e_{i\alpha}] = 0$$

$$Var[e_{i\alpha}] = \sigma_i^2$$

Therefore,

$$E[y_{i\alpha} / i] = \mu_i$$

$$V[y_{i\alpha} / i] = \sigma_i^2$$

6. Inclusion probability in a stratified population

Suppose we have a stratified population containing H number of strata. If we take one stratum denoted by h , our indicator variable becomes,

$$I_{hk}(s) = \begin{cases} 1 & \text{if the } k^{th} \text{ unit is included in the sample from stratum } h \\ 0 & \text{otherwise} \end{cases}$$

Our first order inclusion probability is denoted by π_{hk} which denotes the probability that element k from stratum h is included into the sample. In the case of systematic sampling, because each element k belongs to one of the ' a_h ' equally probable systematic samples in stratum ' h ', (where a_h is the sampling interval in stratum h .)

$$\pi_{hk} = \frac{1}{a_h}$$

The second order inclusion probability is denoted by π_{hkl} which refers to the probability that both elements ' k ' and ' l ' from stratum h are included into the sample. Under systematic sampling in stratum h

$$\pi_{hkl} = \begin{cases} \frac{1}{a_h} & \text{if } k \text{ and } l \text{ from stratum } h \text{ are in same systematic sample} \\ 0 & \text{otherwise} \end{cases}$$

The expectation and variance of $I_{hk}(s)$ are obtained as follows,

$$E[I_{hk}(s)] = \pi_{hk}$$

$$V[I_{hk}(s)] = \pi_{hk}(1 - \pi_{hk})$$

Covariance

$$\text{COV}[I_{hk} I_{hl}] = \pi_{hkl} - \pi_{hk} \pi_{hl}$$

Since in stratum h our sampling interval is a_h we will have a_h systematic samples in the stratum indicated by s_r where $r = 1, 2, \dots, a_h$

7. Horvitz Thompson estimator in a Stratified Population

According to Horvitz and Thompson (1952), the estimator of the population total is given by

$$\hat{t}_{HT} = \sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}}$$

Its variance is given by,

$$V\left(\hat{t}_{HT}\right) = \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{y_{hk}^2}{\pi_{hk}} (1 - \pi_{hk}) + \sum_{h=1}^H \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{y_{hk} y_{hl}}{\pi_{hk} \pi_{hl}} (\pi_{hkl} - \pi_{hk} \pi_{hl})$$

According to Sarndal (1992) there is no unbiased estimator of the variance of estimated population total when a sample is generated using systematic sampling technique. A precise estimator below was then chosen

$$\hat{V}\left(\hat{t}_{HT}\right) = \sum_{h=1}^H \hat{V}_h\left(\hat{t}_{hHT}\right)$$

$$\hat{V}\left(\hat{t}_{HT}\right) = \sum_{h=1}^H N_h^2 \frac{1 - f_h}{n_h} S_{hs_r}^2$$

where $S_{hs_r}^2 = \frac{1}{n_h - 1} \sum_{hs_r} \left(y_{hk} - \bar{y}_{hs_r} \right)^2$

Where $f_h = \frac{n_h}{N_h}$ represents the sampling fraction

8. The Mean Square Error (MSE)

The mean square error (MSE) of an estimator is one of the many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. The difference occurs because of randomness or because the estimator does not account for information that could produce a more accurate estimate. In order to compare a biased estimate with unbiased estimate or two estimates that have different amounts of bias, a useful criterion is the ‘mean square error’ of the estimate measured from the population value that is being estimated. Formally,

$$\begin{aligned} \text{MSE}\left(\hat{\mu}\right) &= E\left[\hat{\mu} - \mu\right]^2 = E\left[\left(\hat{\mu} - m\right) + \left(m - \mu\right)\right]^2 \\ &= E\left(\hat{\mu} - m\right)^2 + 2\left(m - \mu\right) E\left(\hat{\mu} - m\right) + \left(m - \mu\right)^2 \\ &= \left(\text{Variance of } \hat{\mu}\right) + (\text{bias})^2 \end{aligned}$$

The cross product term vanishes since $E\left(\hat{\mu} - m\right) = 0$. The use of the MSE as a criterion of the accuracy of an estimate amounts to regarding two estimates that have the same MSE as equivalent. This is not strictly correct because the frequency distribution of errors $\left(\hat{\mu} - \mu\right)$ of different sizes will not be the same for the two estimates if they have different amounts of bias.

9. Total variance of the estimator of the population total in absence of systematic errors

The total variance in the absence of systematic errors is obtained as follows. Since our population is stratified our measurement model is $y_{hk} = \mu_{hk} + e_{hk}$

$$\begin{aligned} E_m(y_{hk} / s) &= \mu_{hk} & k \in s \\ V_m(y_{hk} / s) &= \sigma_{hk}^2 & k \in s \\ \text{Cov}_m(y_{hk} y_{hl} / s) &= \sigma_{hkl} & k, l \in s \end{aligned}$$

The total variance is given by

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

Where

$$\begin{aligned} E_p\left[V_m\left(\hat{t}_{HT}\right)\right] &= E_p\left[V_m\left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}}\right)\right] \\ &= E_p\left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{\sigma_{hk}^2}{\pi_{hk}^2} + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{n_h} \sum_{k \neq l}^{n_h} \frac{\sigma_{hkl}}{\pi_{hk} \pi_{hl}}\right) \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\sigma_{hk}^2}{\pi_{hk}} + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\sigma_{hkl} \pi_{hkl}}{\pi_{hk} \pi_{hl}} \\ V_p\left[E_m\left(\hat{t}_{HT}\right)\right] &= V_p\left[E_m\left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}}\right)\right] \\ &= V_p\left[\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{\mu_{hk}}{\pi_{hk}}\right] \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\mu_{hk}^2}{\pi_{hk}} (1 - \pi_{hk}) + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk} \pi_{hl})}{\pi_{hk} \pi_{hl}} \mu_{hk} \mu_{hl} \end{aligned}$$

Therefore the total variance in absence of systematic errors is given by

$$V_{pm}\left(\hat{t}_{HT}\right) = \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\sigma_{hk}^2}{\pi_{hk}} + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\sigma_{hkl} \pi_{hkl}}{\pi_{hk} \pi_{hl}} + \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\mu_{hk}^2}{\pi_{hk}} (1 - \pi_{hk}) + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk} \pi_{hl})}{\pi_{hk} \pi_{hl}} \mu_{hk} \mu_{hl}$$

By replacing both first and second order inclusion probabilities with $\frac{1}{a_h}$ the variance becomes

$$V_{pm}(\hat{t}_{HT}) = \sum_{h=1}^H \sum_{k=1}^{N_h} a_h \sigma_{hk}^2 + \sum_{h=1}^H \sum_{k, l \in S_{h-1}} \sum_{l=1}^{N_h} \sum_{k \neq l} a_h \sigma_{hkl}^2 + \sum_{h=1}^H \sum_{k=1}^{N_h} (a_h - 1) \mu_{hk}^2 + \sum_{h=1}^H \sum_{k, l \in S_{h-1}} \sum_{k=1}^{N_h} \sum_{k \neq l} (a_h - 1) \mu_{hk} \mu_{hl}$$

10. The mathematical model for measurement of errors in stratified population

Since our population is stratified, the model becomes

$$y_{hk} = \mu_{hk} + e_{hk}$$

μ_{hk} - represents the true value of unit k in stratum h

e_{hk} - represents the measurement error term of unit k in stratum h

$$\therefore E_m(y_{hk} / s) = b_{hk} + \mu_{hk} \quad k \in s$$

Where b_{hk} refers to the bias term

$$V_m(y_{hk} / s) = \sigma_{hk}^2 \quad k \in s$$

The covariance between k^{th} and l^{th} unit is

$$Cov_m(y_{hk} y_{hl} / k, l \in s) = E[y_{hk} - E(Y_{hk})][y_{hl} - E(y_{hl})]$$

$$= E(e_{hk} - b_{hk})(e_{hl} - b_{hl})$$

$$= Cov(e_{hk} e_{hl})$$

$$\therefore Cov_m(y_{hk} y_{hl} / k, l \in s) = \sigma_{hkl}$$

11. Decomposition of the mean square error.

In this case we will consider the Horvitz Thompson estimator and the effect of measurement errors on its accuracy. We will decompose the mean square error into components, assuming that the measurements obey the simple measurement model 'm' as stated before. The mean square error of \hat{t}_{HT} can be written as the sum of the total variance and squared bias.

$$MSE_{pm}(\hat{t}_{HT}) = V_{pm}(\hat{t}_{HT}) + \left[B_{pm}(\hat{t}_{HT}) \right]^2$$

The total variance is given by $V_{pm}(\hat{t}_{HT}) = E_{pm} \left[\hat{t}_{HT} - E_{pm}(\hat{t}_{HT}) \right]^2$

The bias is given by $B_{pm}(\hat{t}_{HT}) = E_{pm}(\hat{t}_{HT}) - t_{\mu}$

This is called the measurement bias, which arises when expected measurement values on elements do not agree with true values. Variance term can be decomposed as follows

$$V_{pm}(\hat{t}_{HT}) = V_p \left[E_m(\hat{t}_{HT} / s) \right] + E_p \left[V_m(\hat{t}_{HT} / s) \right] = V_1 + V_2$$

The first component V_1 is referred to as the sampling variance which is zero in the case of complete enumeration whereas the second component V_2 is referred to as the measurement variance. The Horvitz Thompson estimator for the population total in this case will be

$$\hat{t}_{HT} = \sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}}$$

But,

$$\begin{aligned} E_{pm} \left(\hat{t}_{HT} \right) &= E_p \left[E_m \left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}} \right) \right] \\ &= E_p \left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{g_{hk}}{\pi_{hk}} \right) \quad \text{where } g_{hk} = \mu_{hk} + b_{hk} \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{g_{hk}}{\pi_{hk}} E_p [I_{hk}] \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} g_{hk} \quad \text{since } E_p [I_{hk}] = \pi_{hk} \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} \mu_{hk} + \sum_{h=1}^H \sum_{k=1}^{N_h} b_{hk} \end{aligned}$$

From the decomposition of the variance term, we have V_1 & V_2 where

$$\begin{aligned} V_1 &= V_p \left[E_m \left(\hat{t}_{HT} \right) \right] & V_2 &= E_p \left[V_m \left(\hat{t}_{HT} \right) \right] \\ \therefore V_1 &= V_p \left[E_m \left(\hat{t}_{HT} \right) \right] = V_p \left[E_m \left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}} \right) \right] \\ &= V_p \left[\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{g_{hk}}{\pi_{hk}} \right] \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{g_{hk}^2}{\pi_{hk}^2} V [I_{hk}] + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{ah}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{g_{hk}}{\pi_{hk}} \frac{g_{hl}}{\pi_{hl}} \text{cov} (I_{hk} I_{hl}) \end{aligned}$$

But $V [I_{hk}] = \pi_{hk} (1 - \pi_{hk})$ and $\text{cov}(I_{hk} I_{hl}) = (\pi_{hkl} - \pi_{hk} \pi_{hl})$

$$\begin{aligned} \therefore V_1 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{g_{hk}^2}{\pi_{hk}^2} \pi_{hk} (1 - \pi_{hk}) + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{ah}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{g_{hk}}{\pi_{hk}} \frac{g_{hl}}{\pi_{hl}} (\pi_{hkl} - \pi_{hk} \pi_{hl}) \\ &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{g_{hk}^2}{\pi_{hk}} (1 - \pi_{hk}) + \sum_{h=1}^H \sum_{k, l \in s_{r=1}}^{s_{ah}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk} \pi_{hl})}{\pi_{hk} \pi_{hl}} g_{hk} g_{hl} \end{aligned}$$

but $g_{hk} = \mu_{hk} + b_{hk}$ and $g_{hl} = \mu_{hl} + b_{hl}$

$$\begin{aligned}
 V_1 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} (\mu_{hk} + b_{hk})^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk}\pi_{hl})}{\pi_{hk}\pi_{hl}} (\mu_{hk} + b_{hk})(\mu_{hl} + b_{hl}) \\
 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} (\mu_{hk}^2 + 2\mu_{hk}b_{hk} + b_{hk}^2) + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk}\pi_{hl})}{\pi_{hk}\pi_{hl}} (\mu_{hk}\mu_{hl} + \mu_{hk}b_{hl} + \mu_{hl}b_{hk} + b_{hk}b_{hl})
 \end{aligned}$$

$$\begin{aligned}
 V_1 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} \mu_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk}\pi_{hl})}{\pi_{hk}\pi_{hl}} \mu_{hk}\mu_{hl} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} b_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk}\pi_{hl})}{\pi_{hk}\pi_{hl}} b_{hk}b_{hl} \\
 &+ 2 \left[\sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} \mu_{hk}b_{hl} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{(\pi_{hkl} - \pi_{hk}\pi_{hl})}{\pi_{hk}\pi_{hl}} \mu_{hk}b_{hl} \right]
 \end{aligned}$$

$$\begin{aligned}
 V_2 &= E_p \left[V_m \left(\hat{t}_{HT} \right) \right] \\
 &= E_p \left[V_m \left(\sum_{h=1}^H \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_{hk}} \right) \right] \\
 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\sigma_{hk}^2}{\pi_{hk}} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\sigma_{hkl}\pi_{hkl}}{\pi_{hk}\pi_{hl}}
 \end{aligned}$$

Thus the total variance in presence of measurement errors is

$$V_{pm}(\hat{t}_{HT}) = V_p \left[E_m(\hat{t}_{HT}/s) \right] + E_p \left[V_m(\hat{t}_{HT}/s) \right] = V_1 + V_2$$

$$\begin{aligned}
 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\sigma_{hk}^2}{\pi_{hk}} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\sigma_{hkl}\pi_{hkl}}{\pi_{hk}\pi_{hl}} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} \mu_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\pi_{hkl} - \pi_{hk}\pi_{hl}}{\pi_{hk}\pi_{hl}} \mu_{hk}\mu_{hl}
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} b_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\pi_{hkl} - \pi_{hk}\pi_{hl}}{\pi_{hk}\pi_{hl}} b_{hk} b_{hl} \\
 & + 2 \left(\sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} \mu_{hk} b_{hk} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\pi_{hkl} - \pi_{hk}\pi_{hl}}{\pi_{hk}\pi_{hl}} \mu_{hk} b_{hl} \right)
 \end{aligned}$$

Under systematic sampling in populations with stratification as previously obtained the inclusion probabilities are:

$$\pi_{hk} = \frac{1}{a_h}$$

$$\pi_{hkl} = \begin{cases} \frac{1}{a_h} & \text{if } k \text{ and } l \text{ from stratum } h \text{ are in same systematic sample} \\ 0 & \text{if } k \text{ and } l \text{ from stratum } h \text{ are in different systematic samples} \end{cases}$$

We then substitute these probabilities in

$$\begin{aligned}
 V_{pm}(\hat{t}_{HT}) &= V_p \left[E_m(\hat{t}_{HT}/s) \right] + E_p \left[V_m(\hat{t}_{HT}/s) \right] \\
 &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\sigma_{hk}^2}{\pi_{hk}} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\sigma_{hkl} \pi_{hkl}}{\pi_{hk} \pi_{hl}} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} \mu_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\pi_{hkl} - \pi_{hk}\pi_{hl}}{\pi_{hk}\pi_{hl}} \mu_{hk} \mu_{hl} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} b_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\pi_{hkl} - \pi_{hk}\pi_{hl}}{\pi_{hk}\pi_{hl}} b_{hk} b_{hl} \\
 &+ 2 \left(\sum_{h=1}^H \sum_{k=1}^{N_h} \frac{(1-\pi_{hk})}{\pi_{hk}} \mu_{hk} b_{hk} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\pi_{hkl} - \pi_{hk}\pi_{hl}}{\pi_{hk}\pi_{hl}} \mu_{hk} b_{hl} \right)
 \end{aligned}$$

We then replace both the first and second order inclusion probabilities with $\frac{1}{a_h}$ to obtain

$$\begin{aligned}
 V_{pm} \left[\hat{t}_{HT} \right] &= \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\sigma_{hk}^2}{\left[\frac{1}{a_h} \right]} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\sigma_{hkl}}{\left[\frac{1}{a_h} \right] \left[\frac{1}{a_h} \right]} \left[\frac{1}{a_h} \right] \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\left[1 - \frac{1}{a_h} \right]}{\left[\frac{1}{a_h} \right]} \mu_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\left[\left[\frac{1}{a_h} \right] - \left[\frac{1}{a_h} \right] \left[\frac{1}{a_h} \right] \right]}{\left[\frac{1}{a_h} \right] \left[\frac{1}{a_h} \right]} \mu_{hk} \mu_{hl} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\left[1 - \frac{1}{a_h} \right]}{\left[\frac{1}{a_h} \right]} b_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\left[\left[\frac{1}{a_h} \right] - \left[\frac{1}{a_h} \right] \left[\frac{1}{a_h} \right] \right]}{\left[\frac{1}{a_h} \right] \left[\frac{1}{a_h} \right]} b_{hk} b_{hl} \\
 &+ 2 \left[\sum_{h=1}^H \sum_{k=1}^{N_h} \frac{\left(1 - \frac{1}{a_h} \right)}{\frac{1}{a_h}} \mu_{hk} b_{hk} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} \frac{\left(\frac{1}{a_h} - \frac{1}{a_h} \frac{1}{a_h} \right)}{\frac{1}{a_h} \frac{1}{a_h}} \mu_{hk} b_{hk} \right]
 \end{aligned}$$

This then simplifies to

$$\begin{aligned}
 V_{pm} \left(\hat{t}_{HT} \right) &= \sum_{h=1}^H \sum_{k=1}^{N_h} a_h \sigma_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} a_h \sigma_{hkl} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} (a_h - 1) \mu_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} (a_h - 1) \mu_{hk} \mu_{hl} \\
 &+ \sum_{h=1}^H \sum_{k=1}^{N_h} (a_h - 1) b_{hk}^2 + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} (a_h - 1) b_{hk} b_{hl} \\
 &+ 2 \left[\sum_{h=1}^H \sum_{k=1}^{N_h} (a_h - 1) \mu_{hk} b_{hk} + \sum_{h=1}^H \sum_{k,l \in s_{r=1}}^{s_{a_h}} \sum_{k=1}^{N_h} \sum_{k \neq l}^{N_h} (a_h - 1) \mu_{hk} b_{hl} \right]
 \end{aligned}$$

12. Simulation of a finite population

Visual basic programming language along with Microsoft access were used to generate a finite population of size N=1000. A population was simulated containing true values ranging from 20 to 90 inclusive. The true values had a normal distribution with a mean of 55 and a variance of 100. The population was then subdivided into four strata. This was done by first arranging all the population units in ascending order. The first 300 units were selected to constitute the first stratum and the following 250 units were selected to constitute the second stratum. The third and fourth strata were also selected to contain 250 and 200 units respectively. A sampling interval of 10 was used in each stratum resulting to 10 systematic samples in each stratum. In each stratum a sample was selected at random. In this case an assumption was made that systematic error b_{hk} is proportional to the true value $b_{hk} = d\mu_{hk}$

13. Results

Table 2. Estimated population total subjected to systematic errors with proportions of size 'd'

		TRUE VALUES	ESTIMATES WITH SYSTEMATIC ERROR					
			d=0.05	d=0.10	d=0.15	d=0.20	d=0.25	d=0.30
STRATUM	1	13320	14215	14881	15547	16213	16879	17545
	2	13350	14030.5	14698	15365.5	16033	16700.5	17368
	3	15050	16034.5	16787	17539.5	18292	19044.5	19797
	4	13840	14369	15061	15723	16445	17137	17829
Estimated pop Total		55560	58649	61427	64205	66983	69761	72539
population Total		55000	57865.3	60615.3	63365.3	66115.3	68865.3	71615.3
Variance of estimator of population Total		1343800	2462003	2620951	2786619	2959005	3138111	3323935
Estimate of the Variance of pop Total		99688.7	188425	199168	210410	222151	234390	247127

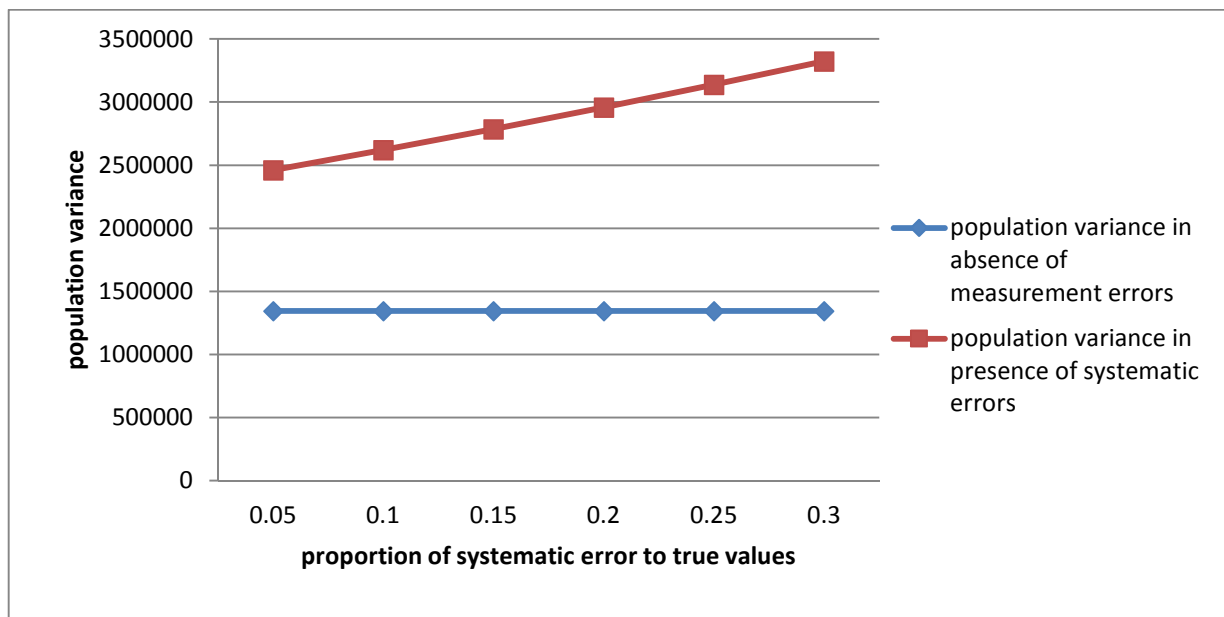


Figure 1. Effects of systematic errors on variance of estimated population total

14. Conclusion

The table 2 and figure 1 indicated that there was an increase in population total and population variance with increase in systematic error. The findings of the study indicated that systematic errors had a significant impact on the accuracy of the estimates of both population total and population variance.

15. References

Cochran, W.G., 1977. *Sampling techniques*. 3rd Edition, John Willey and Sons, New York.
 Hansen, M. H., and Hurwitz, W. N., 1946. The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41, pp.517-529
 Horvitz D. G. and Thompson D. J., 1952. A Generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), pp.663-685

- Konijn H.S., 1973. *Statistical theory of sample survey design and analysis*. North Holland Publishing Company, Amsterdam.
- Mahalanobis, P.C., 1946. Recent experiments in statistical sampling in the Indian statistical institute. *Journal of the Royal Statistical Society*, 325-370.
- Sarndal, C.E., Swensson, B., and Wretman, J., 1992. *Model assisted survey sampling*. New York, Springer Verlag.
- Suhkatme, P.V., and Seth, G.R., 1952. Non-sampling errors in surveys. *Journal of Indian Society of Agricultural Statistics*, 5-41.



Creative Commons Attribution 4.0 International License.
CC BY