# A MODIFIED REGRESSION ESTIMATOR FOR DOUBLE SAMPLING

[1] Olayiwola O.M., [1] Popoola R. A., [2] Bisira H.O.

[1] Department of Statistics, College of Physical Sciences,
Federal University of Agriculture, Abeokuta, Ogun State, Nigeria
[2] Department of Mathematics and Statistics, Lagos State Polytechnic, Ikorodu, Nigeria

Corresponding author: Olayiwola O. M., laniyimathew@yahoo.com

***ABSTRACT:*** Regression is a criterion for estimating the relationships among variables. Though many authors have derived regression estimator on double sampling by making use of auxiliary variables without stratification and obtained their variances. Hence, there is need to determine a modified regression estimator on double sampling by making use of auxiliary variables with stratification for estimating the population estimates. This study proposed an estimator for estimating the population estimates of double sampling involving stratification. The data were collected on cost of House Rent, Transportation, Feeding, Fueling of Car, PHCN Bill, Toiletries, Fueling of Generator, Recharge Card and total expenditure of some randomly selected staff members of Federal University of Agriculture, Abeokuta, Ogun State, Nigeria. A random sample of size 138 was randomly selected from 256 respondents and sub sample of 81 was taken from the first sample. Auxiliary variables in the first sample and sub-sample (cost of House Rent, Transportation, Feeding, Fueling of Car; PHCN Bill, Toiletries, Fueling of Generator and Recharge Card) were stratified into two vector groups using level of expenditure and income as stratifying factors respectively. The mean and the minimum variance of the double sampling for the existing estimator are 92,634 and 2,672,766,891 respectively, while the mean and variance of the proposed estimator are 196,330.1 and 3,114,477 respectively. The minimum variance of the proposed estimator is smaller than the minimum variance of the existing estimator; hence, the proposed estimator is an efficient estimator.
***KEYWORDS:*** Double sampling, Stratification, Auxiliary variables.

## 1. INTRODUCTION

Most powerful techniques used for the stratification purpose which was firstly introduced by Neyman ([Ney38]) are referred to as double sampling. Double sampling is based on the idea of a sampling design in which nature (specifically the size) of sampling units does not differ at any phase of sampling. Double sampling is generally employed when number of units, required to give the desired precision on different items, is widely different. This technique is employed to utilize the information collected at the first phase in order to improve the precision of the information to be collected at the second phase.

In double sampling, also known as two-phase sampling, observations are made on a vector of auxiliary variables in a large sample, called the first phase sample. Then observations are made on the variables of interest using a smaller sample, called the second phase sample. The smaller sample is typically a subsample of the original sample, and often the vector of auxiliary variables is a vector of indicator variables defining subgroups (strata) of the original sample used in selecting the second phase sample.

Double sampling is cost effective sampling design, and precision of ratio and regression estimates of study variable. Double sampling increases if there is a high degree of correlation between the auxiliary variable and study variable Hidiroglou and Sarndal ([HS98]). The fact that precision of estimators of the mean of study variable "*y*" is increased by proper attachment of highly correlated auxiliary variables. In some situations where auxiliary information is available at population level and cost per unit of collecting study variable "*y*" is affordable then single-phase sampling is more appropriate. But in a situation where prior information of auxiliary variable is lacking then it is neither practical nor economical to conduct a census for this purpose. The appropriate technique used to get estimates of those auxiliary variables on the basis of samples is two-phase sampling. In such cases we take large preliminary sample and from that auxiliary variables are computed. The main sample is independently sub-sampled from that large sample. In this study we modified regression estimator by Muhammad Hanif, Shahbaz and Ahmad ([HSA2010]) by grouping the auxiliary variable into homogeneous group before selecting first and the second stage.

## 2. LITERATURE REVIEW

Several estimators have been developed in single and double phase sampling which utilizes information on auxiliary variables as well as auxiliary attributes. Freese ([Fre62]) presents detailed description of its application in forestry. He (1962) selects auxiliary variable to achieve primary objective of two-phase sampling i.e. to reduce total inventory time without sacrificing the precision about the point estimate. Basal area is commonly utilized as auxiliary variable with two-phase sampling for volume estimates. This is because of high correlation between basal area and volume and also the fact that basal area can be determined very quickly. Rao ([Rao73]) applies auxiliary variables to stratification, non-response problems and analytic comparisons.

Use of double sampling is necessary if the value of auxiliary variable is obtained by performing a non-destructive experiment whereas to obtain a value of study variable of a unit, a destructive experiment has to be performed Unnikrishan and Kunte ([UK95]). Armstrong and St-Jean ([AS93]) describe the use of two-phase sampling design for sampling tax records.

The classical estimators which use information on auxiliary variables are the ratio and regression estimators as given in Hansen et al. ([HHM93]) as:

$$\bar{y}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x}) \quad (1)$$

where $\beta$ is the regression coefficient of Y on X. The value of $\beta$ for which the variance of (1) is minimum is $\beta = S_{xy}/S_x^2$. The minimum variance of (2.1) is given as:

$$Var(\bar{y}_{lr}) = \theta S_y^2(1 - \rho_{yx}^2); \quad (2)$$

where $\theta = n^{-1} - N^{-1} and\ \rho_{yx}$ is the correlation coefficient between X and Y,

where $\rho_{yx}^2$ is the squared multiple correlation coefficient between Y and the combined effect of all the auxiliary variables. The regression estimator has been effectively used in two phase and multiphase sampling. The estimator of mean in two phase sampling has been discussed by Hansen *et al* ([HHM93]) and is given as:

$$\bar{y}_{lr(2)} = \bar{y}_2 + \beta(\bar{x}_1 - \bar{x}_2); \quad (3)$$

where $\bar{x}_1$ and $\bar{x}_2$ are first phase and second phase mean of auxiliary variable X based upon the samples of sizes $n_1$ and $n_2$ respectively; $\bar{y}_2$ is mean of Y for the second phase sample of size $n_2$. The variance of regression estimator for two phase sampling is given as:

$$Var(\bar{y}_{lr(2)}) = S_y^2\{\theta_2(1 - \rho_{yx}^2) + \theta_1\rho_{yx}^2\} \quad (4)$$

where $\theta_h = n_h^{-1} - N^{-1} and\ n_h$ is sample size at $h^{th}$ phase.

Mukerjee et al. ([MRV87]) has also proposed a regression type estimator using two auxiliary variables. The estimator is:

$$\bar{y}_M = \bar{y}_2 + \beta_{yx}(\bar{x}_1 - \bar{x}_2) + \beta_{yw}(\bar{w}_1 - \bar{w}_2); \quad (5)$$

The variance of $\bar{y}_M$ is given as:

$$[Var(\bar{y}]_M) = S_y^2\{\theta_2 - (\theta_2 - \theta_1)(\rho_{yw}^2 + \rho_{yx}^2 - 2\rho_{yx}\rho_{yw}\rho_{xw})\} \quad (6)$$

Sahoo ([SSM93]) suggested a regression type estimator by using an additional auxiliary variable for two-phase sampling when population mean of main auxiliary variable was unknown and his class of estimators covered a large number of estimators. A slightly modified regression type estimator has been proposed by Sahoo et al. ([SSM93]) by using information of two auxiliary variables. The proposed estimator is:

$$\bar{y}_{ssm} = \bar{y}_2 + \beta_1(\bar{x}_1 - \bar{x}_2) + \beta_2(\bar{W}_1 - \bar{w}_2) \quad (7)$$

The variance of estimator given by Sahoo et al. ([SSM93]) is:

$$Var(\bar{y}_{ssm}) = S_y^2\{\theta_2(1 - \rho_{yx}^2) + [\theta_1(\rho]_{yx}^2 - \rho_{yw}^2)\} \quad (8)$$

where $\rho_{yw}^2$ is squared correlation coefficient between Y and W.

Roy ([Roy03]) constructed a regression-type estimator of population mean of the main variable in the presence of available knowledge on second auxiliary variable, when the population mean of the first auxiliary variable was not known. Roy ([Roy03]) has given the following regression type estimator:

$$\bar{y}_R = \bar{y}_2 + \alpha[\bar{x}_1 + \beta(\overline{W} - \bar{w}_1) - \{\bar{x}_2 + \gamma(\overline{W} - \bar{w}_2)\}]; \quad (9)$$

The variance of estimator given by Roy ([Roy03]) is:

$$Var(\bar{y}_R) = S_y^2[\theta_2(1 - \rho_{y.wx}^2) + \theta_1\rho_{yx.w}^2(1 - \rho_{yw}^2)] \quad (10)$$

where $\rho_{y.wx}^2$ is the squared multiple correlation between Y and combined effects of X and W, $\rho_{yw.x}^2$ is the squared partial correlation between Y and X keeping W at a constant level and $\rho_{yw}^2$ is the squared correlation between Y and W.

Diana and Tommasi ([DT04]) proposed a two phase sampling scheme. In the first phase, an initial simple random sample (without replacement) $s' \subset U$ on fixed size $n'$ is selected to observed auxiliary variable x. in the second phase, a simple random sample (without replacement) s of fixed size n is drawn from $s'$ to observe the variable of interest y. The regression type estimator of finite population variance $S_y^2$ in two phase sampling takes the form

$$S_{yreg}^{*2} = s_y^{2} + \beta_{22}(y,x)\left(s_x'^{2} - s_x^{2}\right) \quad (11)$$

where $s_x'^2$ and $s_x^2$ are estimates of finite population variance of x using first phase and second phase samples respectively, $s_y^2$ is an unbiased estimate of the finite population of y based on the second phase sample and further $\beta_{22}(y,x) = \frac{Cov(S_y^2, S_x^2)}{Var(S_x^2)}$ under bivariate normality of $(y,x)$, $\beta_{22}(y,x) = \beta_{yx}^2$, where $\beta_{yx}$ represents regression coefficient of y on x; and hence to first order approximation

$$V(S_{yreg}^{*2}) = 2\left(1-\rho_{yx}^4\right)\frac{s_y^4}{n} + 2\rho_{yx}^4\frac{s_y^4}{n'} \quad (12)$$

where $\rho_{yx}$ is the correlation coefficient.
Another regression type estimator has been proposed by Samiuddin and Hanif ([SH07]). The estimator is given as:

$$\bar{y}_{sh(2)} = \bar{y}_2 + \alpha(\bar{x}_1 - \bar{x}_2) + \beta(\bar{W} - \bar{w}_1) + \gamma(\bar{W} - \bar{w}_2); \quad (13)$$

Samiuddin and Hanif ([SH07]) has shown that the variance of (12) as:

$$Var(\bar{y}_{sh(2)}) = S_y^2\left\{\theta_2\left(1-\rho_{y.wx}^2\right) + \theta_1\rho_{yx.w}^2\left(1-\rho_{yw}^2\right)\right\}$$

Muhammad Hanif et al. ([HHS10]) also proposed the following estimator of population mean in two phase sampling:

$$\bar{y}_{sh1} = \bar{y}_2 + \alpha_1(\bar{x}_1 - \bar{x}_2) + \alpha_2(\bar{w}_1 - \bar{w}_2) + \alpha_3(\bar{W} - \bar{w}_1) \quad (14)$$

where $\alpha_1$; $\alpha_2$ and $\alpha_3$ are the unknowns to be determined by minimizing the variance of eqn (14) where

$$\alpha_1 = \frac{S_y\left(\rho_{xy} - \rho_{xw}\rho_{yw}\right)}{S_w\left(1-\rho_{wx}^2\right)} = \beta_{yx.w}$$

$$\alpha_2 = \frac{S_y\left(\rho_{wy} - \rho_{xw}\rho_{yx}\right)}{S_w\left(1-\rho_{wx}^2\right)} = \beta_{yw.x}$$

$$\alpha_3 = \frac{S_y\rho_{yw}}{S_w} = \beta_{yw}$$

Using value of $\alpha_1$; $\alpha_2$ and $\alpha_3$ in (14); and simplified variance of (14) is given as:

$$Var(\bar{y}_{sh1}) = S_y^2\left[\theta_2\left(1-\rho_{y.wx}^2\right) + \theta_1\left(\rho_{y.wx}^2 - \rho_{yw}^2\right)\right] \quad (15)$$

The variance given in (15) is same as the variance of the regression type estimator given by Roy ([Roy03]), but the formation of Roy's estimator is relatively complex as compared with the estimator proposed in (14). Further, by using the relation $1-\rho_{y.wx}^2 = \left(1-\rho_{yw}^2\right)\left(1-\rho_{yx.w}^2\right)$ the variance of $\bar{y}_{sh1}$ can also be written as:

$$Var(\bar{y}_{sh1}) = S_y^2\left[\theta_2\left(1-\rho_{y.wx}^2\right) + \theta_1\rho_{yx.w}^2\left(1-\rho_{yw}^2\right)\right] \quad (16)$$

## 3. METHODOLOGY

Muhammad Hanif, Muhammad Qaiser Shahbaz and Zahoor Ahmad ([HSA10]), proposed the following estimator of population mean by using information of several auxiliary variables:

$$\bar{y}_{hsa} = \bar{y}_2 + \alpha'(\bar{x}_1 - \bar{x}_2) + \beta'(\bar{Z} - \bar{z}_1) \quad (17)$$

where $\bar{x}_1$ is the vector of first phase mean of X's auxiliary variables, $\bar{x}_2$ is the vector of second phase mean of X's auxiliary variables, $\bar{z}_1$ is vector of first phase mean of Z's auxiliary variables and $\bar{Z}$ is the vector of population means for variable Z's. The vectors $\alpha$ and $\beta$ are the vectors of unknown parameters whose values are to be determined so that the variance of equation (17) is minimum. Now, using: $\bar{y}_2 = \bar{Y} + \bar{e}_{y2}$, $\bar{x}_1 = \bar{X} + \bar{e}_{x1}, \bar{x}_2 = \bar{X} + \bar{e}_{x2}, \bar{z}_1 = \bar{Z} + \bar{e}_{z1}$ in eqn (17) we have

$$\bar{y}_{lr} - \bar{Y} = \bar{e}_{y2} + \alpha\left(\left(\bar{X} + \bar{e}_{x1}\right) - \left(\bar{X} - \bar{e}_{x1}\right)\right) + \beta\left(\bar{Z} - \left(\bar{Z} + \bar{e}_{z1}\right)\right)$$

$$\bar{y}_{hsa} - \bar{Y} = \bar{e}_{y2} + \alpha'(\bar{e}_{x1} - \bar{e}_{x2}) + \beta'\bar{e}_{z1}$$

$$\left[[\bar{y}]_{hsa} - \bar{Y}\right]^2 = \left[\bar{e}_{y2} + \alpha'(\bar{e}_{x1} - \bar{e}_{x2}) + \beta'\bar{e}_{z1}\right]^2$$

$$E\left[[\bar{y}]_{hsa} - \bar{Y}\right]^2 = E\left[\bar{e}_{y2} + \alpha'(\bar{e}_{x1} - \bar{e}_{x2}) + \beta'\bar{e}_{z1}\right]^2$$

Squaring and applying expectation, the variance of $\bar{y}$ is given as:

$$Var(\bar{y}_{hsa}) = \theta_2 S_y^2 + \left[(\theta_2 - \theta_1)\right]\alpha' S_x\alpha + \theta_1\beta' S_z\beta - 2\theta_1\left[(\theta_2 - \theta_1)\right]\alpha' S_{xy} - 2\theta_1\beta' S_{zy}. \quad (18)$$

where $S_x$ is the covariance matrix of x, $S_z$ is the covariance matrix of w, $s_{xy}$ is vector of covariance between Y and x and $s_{zy}$ is vector of covariance between Y and z. Partially differentiating (18) with respect to $\alpha$ and $\beta$ and setting the derivatives to zero we obtain $\alpha = S_x^{-1} s_{xy}$ and $\beta = S_z^{-1} s_{zy}$. Further, by using the value of $\alpha$ and $\beta$ in (18), the variance of (17) is:

$$Var(\bar{y}_{hsa}) = S_y^2[\theta_2(1-\rho_{yx}^2) + \theta_1(\rho_{yx}^2 + \rho_{yz}^2)] \quad (19)$$

## 4. PROPOSED ESTIMATOR

$$\bar{y}_{slr} = \sum_{i=1}^{h} W\left(\bar{y}_2 + \alpha'(\bar{x}_1 - \bar{x}_2) + \beta'(\bar{z}_1 - \bar{z}_2)\right); \quad (20)$$

- **Variance Estimation**

Let

$$\theta_1 = \frac{1}{n_1} - \frac{1}{N}, \theta_2 = \frac{1}{n_2} - \frac{1}{N}$$

Now, using

$$\bar{y}_2 = \bar{Y} + \bar{e}_{y2},$$
$$\bar{x}_1 = \bar{X} + \bar{e}_{x1}, \bar{x}_2 = \bar{X} + \bar{e}_{x2}, \bar{z}_1 = \bar{Z} + \bar{e}_{z1},$$
$$\bar{z}_2 = \bar{Z} + \bar{e}_{z2}$$

in eqn (20) we have:

$$\bar{y}_{Slr} = \sum W_h \left(\bar{y}_2 + \alpha(\bar{x}_1 - \bar{x}_2) + \beta(\bar{z}_1 - \bar{z}_2)\right) \quad (20)$$

$$V(\bar{y}_{Slr}) = V\left(\sum W_h \left(\bar{y}_2 + \alpha(\bar{x}_1 - \bar{x}_2) + \beta(\bar{z}_1 - \bar{z}_2)\right)\right)$$

$$V(\bar{y}_{Slr}) = \sum W_h^2 \, V\left(\bar{y}_2 + \alpha(\bar{x}_1 - \bar{x}_2) + \beta(\bar{z}_1 - \bar{z}_2)\right)$$

$$Var(\bar{y}_{Slr}) = \sum W_h^2 \left(\theta_2 S_y^2 + (\theta_2 - \theta_1)\alpha' S_x \alpha + (\theta_2 - \theta_1)\beta' S_z \beta - 2(\theta_2 - \theta_1)\alpha' s_{xy} - 2(\theta_2 - \theta_1)\beta' S_{zy} + 2(\theta_2 - \theta_1)\alpha'\beta' S_{xz}\right) \quad (21)$$

where $S_x$ is the covariance matrix of X, $S_z$ is the covariance matrix of Z, $S_{xy}$ is vector of covariance between Y and X, $S_{zy}$ is vector of covariance between Y and Z, and $S_{xz}$ is vector of covariance between X and Z. Partially differentiating (21) with respect to $\alpha$ and $\beta$ and setting the derivatives to zero we obtain

$$\alpha = \frac{S_y(\rho_{xy} - \rho_{xz}\rho_{zy})}{S_x(1-\rho_{xz}^2)} = \beta_{yx.z} \text{ . and } \beta = \frac{S_y(\rho_{zy} - \rho_{xz}\rho_{xy})}{S_z(1-\rho_{xz}^2)} = \beta_{yz.x}.$$

Further, by using the value of $\alpha$ and $\beta$ in (21), the variance of (20) is:

$$\therefore Var(\bar{y}_{Slr}) = \sum W_h^2 \left(\theta_2 S_y^2 + (\theta_2 - \theta_1)\alpha' S_x \alpha + (\theta_2 - \theta_1)\beta' S_z \beta - 2(\theta_2 - \theta_1)\alpha' s_{xy} - 2(\theta_2 - \theta_1)\beta' S_{zy} + 2(\theta_2 - \theta_1)\alpha'\beta' S_{xz}\right)$$
$1-\rho_x$ (22)

Substitute for $s_{xy} = \rho_{yx} s_x s_y$, $s_{zy} = \rho_{yz} s_z s_y$, and $s_{xz} = \rho_{xz} s_x s_z$ in equation (22)

V(ȳ↓Slr) = ∑ [W↓h]↑2 (θ↓2 S↓y↑2 + [(θ] ↓2 − θ↓1) (S↓y↑2 (ρ↓xy − ρ↓xz ρ↓zy)↑2)/(1 − ρ↓xz↑2)↑2 + [(θ] ↓2 − θ↓1)(S↓y↑2 (ρ↓zy − ρ↓xz ρ↓xy)↑2)/(1 − ρ↓xz↑2)↑2 − 2 [(θ] ↓2 − θ↓1)(ρ↓xy − ρ↓xz ρ↓zy)ρ↓yx s↓x s↓y − [2(θ] ↓2 − θ↓1)

V(ȳ↓Slr) = ∑W↓h↑2 (θ↓2 S↓y↑2 + [(θ] ↓2 − θ↓1) S↓y↑2 (1/(1 − ρ↓xz↑2)↑2 ( [ρ↓xy] ↑2 + [ρ↓xz] ↑2 [ρ↓zy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz + [ρ↓zy] ↑2 + [ρ↓xz] ↑2 [ρ↓xy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz + 2ρ↓xz (ρ↓xy ρ↓zy − [ρ↓zy] ↑2 ρ↓xz −

V(ȳ↓Slr) = ∑W↓h↑2 (θ↓2 S↓y↑2 + [(θ] ↓2 − θ↓1) S↓y↑2 (1/(1 − ρ↓xz↑2)↑2 ( [ρ↓xy] ↑2 + [ρ↓xz] ↑2 [ρ↓zy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz + [ρ↓zy] ↑2 + [ρ↓xz] ↑2 [ρ↓xy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz + 2ρ↓xy ρ↓zy ρ↓xz − [ [2ρ] ↓xz] ↑2 [ρ↓z

V(ȳ↓Slr) = ∑W↓h↑2 (θ↓2 S↓y↑2 + [(θ] ↓2 − θ↓1) S↓y↑2 (1/(1 − ρ↓xz↑2)↑2 ( [ρ↓xy] ↑2 − [ρ↓xz] ↑2 [ρ↓zy] ↑2 + [ρ↓zy] ↑2 − [ρ↓xz] ↑2 [ρ↓xy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz + 2 [ρ↓xy ρ↓zy ρ↓xz] ↑3) − 2/((1 − ρ↓xz↑2))( [ρ↓xy] ↑2 + [ρ↓zy] ↑2 − [2

V(ȳ↓Slr) = ∑W↓h↑2 (θ↓2 S↓y↑2 + [(θ] ↓2 − θ↓1) S↓y↑2 (1/(1 − ρ↓xz↑2)↑2 ( [ρ↓xy] ↑2 − [ρ↓xz] ↑2 [ρ↓zy] ↑2 + [ρ↓zy] ↑2 − [ρ↓xz] ↑2 [ρ↓xy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz + 2 [ρ↓xy ρ↓zy ρ↓xz] ↑3) − 2/((1 − ρ↓xz↑2))( [ρ↓xy] ↑2 + [ρ↓zy] ↑2 −

$$V(\bar{y}_{slr}) = \sum W_h^2 \left(\theta_2 S_y^2 + [(\theta]_2 - \theta_1) S_y^2 \left(\frac{1}{(1-\rho_{xz}^2)^2(\rho_{xy}{}^2(1-\rho_{xz}^2) + \rho_{zy}{}^2(1-\rho_{xz}^2) - 2\rho_{xy}\rho_{zy}\rho_{xz}(1-\rho_{xz}^2))} - \frac{2}{((1-\rho_{xz}^2))(\rho_{xy}{}^2 + \rho_{zy}{}^2 - 2\rho_{xy}\rho_{zy}\rho_{xz})}\right)\right)$$

V(ȳ↓Slr) = ∑W↓h↑2 (θ↓2 S↓y↑2 + [(θ] ↓2 − θ↓1) S↓y↑2 (1/((1 − ρ↓xz↑2)) ( [ρ↓xy] ↑2 + [ρ↓zy] ↑2 − [2ρ] ↓xy ρ↓zy ρ↓xz − 2 [ρ↓xy] ↑2 − [2ρ↓zy] ↑2 + [4ρ] ↓xy ρ↓zy ρ↓xz ))

$$V(\bar{y}_{Slr}) = \sum W_h^2 \left(\theta_2 S_y^2 + (\theta_2 - \theta_1) S_y^2 \left(1/(1-\rho_{xz}^2)\right)\left(\rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xy}\rho_{zy}\rho_{xz}\right)\right)$$

$$V(\bar{y}_{Slr}) = \sum W_h^2 \left(\theta_2 S_y^2 + (\theta_2 - \theta_1) S_y^2 (\rho_{y.xz}^2)\right)$$

$$V(\bar{y}_{Slr}) = \sum W_h^2 \left(\theta_2 S_y^2 - \theta_2 S_y^2 \rho_{y.xz}^2 + \theta_1 S_y^2 \rho_{y.xz}^2\right)$$

$$\therefore V(\bar{y}_{Slr}) = \sum W_h^2 \left(S_y^2\left(\theta_2\left(1-\rho_{y.xz}^2\right) + \theta_1\rho_{y.xz}^2\right)\right) \quad (23)$$

## 5. DISCUSSION

The data were collected from FUNAAB staff members to estimate the total monthly expenditure of the staffs from their monthly expenses (Cost of House Rent ($X_1$), Transportation ($X_2$), Feeding ($X_3$), Fueling of Car ($X_4$), PHCN Bill ($Z_1$), Toiletries ($Z_2$); Fueling of Generator ($Z_3$); Recharge Card ($Z_4$))

**Table 1: Data Statistics**

| STRATUM 1 | STRATUM 2 | EXISTING |
|---|---|---|
| N = 256 | N = 256 | N = 256 |
| $n_1$= 138 | $n_1$= 138 | $n_1$= 138 |
| $n_2$=81 | $n_2$=81 | $n_2$=81 |
| è$_1$= 0.0033 | è$_1$= 0.0033 | è$_1$= 0.0033 |
| è$_2$= 0.0084 | è$_2$= 0.0084 | è$_2$= 0.0084 |
| $h_1$=61 | $h_2$=20 | |
| $W_1$= 0.7407 | $W_2$= 0.2593 | |
| $W_1^2$= 0.5487 | $W_2^2$= 0.0672 | |
| $S_{y1}^2$ = 411710734 | $S_{y2}^2$ = 1654761905 | $S_y^2$ = 2764577160 |
| MEAN = 56733.9 | MEAN = 139596.2 | MEAN = 92634 |
| VARIANCE = 1932363 | VARIANCE = 1182114 | VARIANCE = 2,672,766,891 |

A comparison of estimates between the existing estimator and the proposed estimator is presented beneath:

**Table 2: Estimator Comparison**

| ESTIMATOR | PROPOSED | EXISTING |
|---|---|---|
| **MEAN** | 196,330.1 | 92,634 |
| **VARIANCE** | 3,114,477 | 2,672,766,891 |

From table 2, the mean and the minimum variance of the double sampling for the existing estimator are 92,634 and 2,672,766,891 respectively, while the mean and variance of the proposed estimator are 196,330.1 and 3,114,477 respectively.

## CONCLUSION

The variance of the proposed estimator is smaller than the variance of the existing estimator hence, the proposed estimator is an efficient estimator.

## REFERENCES

[AS93] **J. Armstrong, H. St-Jean** - *Generalized Regression Estimator for Two Phase Sample of Tax Records*. Survey Methodology, 20, 1993, pp. 91-105.

[DT04] **G. Diana, C. Tommasi** - *Estimation for Finite Population Variance in Double Sampling*, Metron 62, no. 2, 2004, pp 223 – 232.

[Fre62] **F. Freese** - *Elementary Forest Sampling*. U. S. Department of Agriculture, Forest Service Agriculture Handbook, 232. 1962.

[HS98] **M. A. Hidirogou, C.E. Sarndal** - *Use of Auxiliary Information for Two-Phase Sampling*. Survey Methodology, 24(1), 1998, pp. 11-20.

[HHM93] **M. H. Hansen, W. N. Hurwitz, W. G. Madow** - *Sample Survey Methods and Theory*. Vol.I, Reprint of the 1953 original, Wiley, New York, 1993.

[HHS10] **M. Hanif, N. Hamad, M. Q. Shahbaz** - *Some New Regression Type Estimators in Two-Phase Sample*, World Applied Sciences Journal, Vol. 8, No. 7, 2010, pp. 799 – 803.

[HSA10] **M Hanif, M. Q. Shahbaz, Z. Ahmad** - *Some Improved Estimators in Multiphase Sampling*, Pakistian Journal of Statistician, 2010, Vol. 26 pg 1, 195 – 202.

[MRV87] **R. Mukerjee, T. J. Rao, K. Vijayan** - *Regression Type Estimators Using Multiple Auxiliary Information*, Australian Journal of Statistics, Vol. 29, No. 3, 1987, pp. 244 – 254.

[Ney38] **J. Neyman** - *Contribution to the Theory of Sampling Human in Populations*, Journal of the American

Statistical Association, Vol. 33, No. 201, 1938, pp. 101-116.

[Rao73] **J. N. K. Rao** - *On double sampling for stratification and analytical surveys,* Biometrika 60, pp. 125-133, 1973.

[Roy03] **D. C. Roy** - *A Regression Type Estimator In Two Phase Sampling Using Two Auxiliary Variables,* Pakistian Journal of Statistics, 19(3), 281-290, 2003.

[SH07] **M. Samiuddin, M. Hanif** - *Estimation Of Population Mean In Single And Two Phase Sampling With Or Without Additional Information,* Pakistan Journal of Statistics. 23, 2007, pg 99-118.

[SSM93] **J. Sahoo, L. N. Sahoo, S. Mohanty** - *A Regression Approach To Estimation In Two Phase Sampling Using Two Auxiliary Variables.* Current Sciences, 65, 1993, pg 73-75.

[UK95] **N. K. Unnikrishnan, S. Kunte** - *Optimality of an Analogue of Basu's Estimator under a Double Sampling Design,* The Indian Journal of Statistics, Volume 57, Series B, Pt. 1, 1995, pp. 103-111.