AN EMPIRICAL INVESTIGATION INTO THE EFFICIENCY OF NEYMAN'S ALLOCATION PROCEDURE OVER OTHER ALLOCATION PROCEDURES IN STRATIFIED RANDOM SAMPLING

Oguagbaka Samuel Kennedy[1] and Udobi Joy I.[2]
[1, 2] Department of Statistics,
Federal Polytechnic Oko,
Anambra State
samkengolden2003@gmail.com[1], wonderoly@gmail.com[2]

Abstracts
Empirical Investigation into The Efficiency of Neyman's Allocation Procedure over other Allocation Procedures in Stratified Random Sampling investigated the efficiency of Neyman's allocation procedure over equal and proportional allocations in stratified sampling techniques. Secondary data collected from the Department of Examinations and Records of Federal polytechnic Oko, Atani campus was employed in the study. The data comprised of grade point average (GPA) of the second-semester result 2015/2016 academic session for National Diploma (ND), Higher National Diploma (HND), of Mathematics/Statistics, and Computer Science departments. A Stratified Random Sampling scheme was used in selecting 120 samples from 1550 students' result. Each department and programme of study stands as a stratum. The Independent sample was selected randomly based on equal, proportional and Neyman's Allocation Procedures. The mean, variance and the mean squared error for Neyman's allocation procedure were 2.90, 0.001075 and 0.0010 respectively. For the proportional allocation procedure, the mean was 2.90, the variance was 0.001982 and the mean squared error was 0.0320, while the equal allocation procedure gave a mean of 2.96, variance of 0.001138 and mean squared error of 0.2865. Neyman's Allocation Procedure gave the least mean squared error. This was followed by proportional allocation and equal allocation. Hence for estimating the average, variance and the mean squared error of student's academic performance of all the three sample allocation procedures considered in this study, Neyman's Allocation Procedure is the best and hence the most efficient using the mean squared criterion. Recommendations were made to researchers employing a stratified sampling scheme to use Neyman's Allocation Procedures.

Keywords: Efficiency, Empirical Investigation, stratified sampling, proportional allocation, Equal allocation

1.0 Introduction

Sampling is concerned with the proper selection of good representative of the whole within a population to estimate characteristics of the whole population (Okafor, 2002). Probability samples are usually designed to be measurable, that is, so designed that statistical inference to population values can be based on measures of variability, usually standard errors, computed from the sample data.

In general, there is a need to devise a sampling scheme that is economical and easy to operate, that yields unbiased estimates, and minimizes the effects of sampling variation. Stratification is a means of sample design by which the population of interest is divided into groups, called strata, according to some known characteristic(s). Stratified sample designs are employed for several reasons. The precision of information obtained from a survey depends on many factors viz: the size of the sample, sampling technique, variation within the strata, among others.
Nevertheless, in the context of stratified sampling technique, Horgan (2006) and Okafor (2002) itemized the following as specific design problems involved in stratification processes: the choice of a stratification variable; the choice of several strata L to be formed; mode of stratification; the problem of allocation of a sample size to strata; and choice of sampling design within strata. On the problem of allocation of sample to strata, Okafor (2002) stated that the following are very important: variability within the stratum, stratum population size and the cost of obtaining information per element in each stratum.

However, how do we draw stratified samples? First, before a sample is drawn it is necessary to determine the strata boundaries. After the definition of the strata boundaries, then the researcher is challenged with how many samples should be drawn from/allocated to the various strata given the total sample size n. Although there are various methods of allocating samples to the strata when the sample size n have been given viz: equal allocation procedure, proportional allocation procedure, optimal/Neyman's allocation procedure, but out of the outlined allocation procedures in stratified technique, which one should be seen as the most efficient allocation procedure. Olayiwola et al. (2013) verified in their study that Neyman's allocation procedure is the most efficient allocation technique based on the minimum variance criterion i.e. (maximized precision $\mathbf{V(\bar{y}_{st})}$) when compared with other methods. Yet, minimum variance criterion remains the basis of appraisal of the performance of procedures of allocating samples to strata which fails to account for the bias associated with each procedure. Thus, the most precise method may not actually be the most accurate.
Meanwhile, Okafor (2002) defined a sampling strategy by the pair $\mathbf{T[D,\hat{\phi}]}$ and that for two different estimators $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ of the parameter ϕ obtained using the same sampling design, $\hat{\boldsymbol{\theta}}_1$ is said to be uniformly better than or more efficient than $\hat{\boldsymbol{\theta}}_2$ if and only if $\mathbf{MSE(\hat{\theta}_1) <}$ $\mathbf{MSE(\hat{\theta}_2}$

Hence, this empirical investigation seeks to verify the claim that Neyman's allocation procedure is more efficient than other procedures using mean square error (MSE) criterion.

The major objective of this study is to verify the efficiency of Neyman's allocation procedure over other procedures. However the following specific objectives were identified also, viz:

- To estimate the mean performance of the students using; equal, proportional and neyman's allocation procedures respectively.
- To estimate the variance of the mean performance of the students using; equal, proportional and neyman's allocation procedures respectively.
- To estimate the mean square error of the mean performance of the students using; equal, proportional and neyman's allocation procedures respectively.
- To compute 95% confidence interval for the mean performance of the students using neyman's allocation procedure.

The research hypotheses for the present study was developed and tested at 5% level of significance.

$\mathbf{H_0}$**:** The estimated mean performance of the students using Neyman's allocation procedure is equal to the stratified population mean.

$\mathbf{H_1}$**:** The estimated mean performance of the students using neyman's allocation procedure is not equal to the stratified population mean.

2.0 Allocation Procedures in Stratified Sampling

Stratified sampling definition does not specify a particular size of sample in each stratum; therefore, the samples can be allocated equally to each stratum or allocated in some other way. Once we select at least a sample from each stratum the procedure of stratified sampling is satisfied. After the sample size **n** is chosen, there are many ways of allocating **n** into individual stratum sizes $\mathbf{n_1}, \mathbf{n_2}, \ldots, \mathbf{n_L}$ with the aim of using an allocation method that gives a specified amount of information at minimum cost. But allocation scheme is affected by the total number of units in each stratum, the variability of observations within each stratum and the cost of obtaining an observation from each stratum. Available and various methods of allocating sample size to strata are discussed viz:

2.2 Equal Allocation

For administrative convenience or otherwise, it is found desirable to select equal sample of size $\mathbf{n_h}$ from each stratum and this is determined using the relation.

$$\mathbf{n_h} = \frac{\mathbf{n}}{\mathbf{L}}, \text{ for, } \mathbf{h = 1,2,\ldots,L}$$

The variance, using equal allocation in stratified random sampling, is given as:

$$\mathbf{V(\bar{y}_{eq})} = \frac{\mathbf{L}}{\mathbf{n}} \sum_{h=1}^{\mathbf{L}} \mathbf{W_h^2 S_h^2} - \frac{\mathbf{1}}{\mathbf{N}} \sum_{h=1}^{\mathbf{L}} \mathbf{W_h\, S_h^2} \qquad \text{(Cochran,1977)}$$

2.3 Optimum/Neyman's Allocation

With optimum allocation, samples may be selected to minimize the overall cost of the survey for a specified value of $\mathbf{V(\bar{y}_{st})}$ or to minimize the $\mathbf{V(\bar{y}_{st})}$ for a given overall cost of the survey. The cost function can be expressed as a linear cost function:

$$\mathbf{Cost = C = C_o} + \sum_{h=1}^{\mathbf{L}} \mathbf{C_h\, n_h}$$

Where:

C is the total cost of the survey

$\mathbf{C_0}$ is the overhead cost and

$\mathbf{C_h}$ is the cost per unit in stratum h

The cost is always proportional to the size of sample, but the cost per unit, $\mathbf{C_h}$, may vary from stratum to stratum. The above cost function is appropriate when the major item of cost is that of taking measurements on each unit.

In stratified random sampling with a linear cost function of the form above, the appropriate allocation which minimizes cost for a fixed value of $\mathbf{V(\bar{y}_{st})}$ or minimizes $\mathbf{V(\bar{y}_{st})}$ for a given cost is given as:

$$\mathbf{n_h} = \frac{\mathbf{nW_h\, S_h}/\sqrt{\mathbf{C_h}}}{\sum_{h=1}^{\mathbf{L}} \mathbf{W_h\, S_h}/\sqrt{\mathbf{C_h}}}$$

The use of this allocation method leads to following rules of conduct. In a given stratum, take a larger sample if:

- The stratum is larger.
- The stratum is more variable internally.
- The sampling is cheaper in the stratum.

On the other hand, if the costs are unknown or constant (the same in each stratum i.e. $\mathbf{C_h = C}$ for **h = 1, 2, . . . ,L,** then the expression above reduces to:

$$\mathbf{n_h} = \frac{\mathbf{nN_h\,S_h}}{\sum_{h=1}^{L}\mathbf{N_h\,S_h}}$$

This method of allocation of total sample size **n** to strata is called **optimum** or minimum variance allocation and it was due to **Neyman** (1934). Hence, it is often referred to as **Neyman allocation**, while the expression of its variance is given as:

$$\mathbf{V(\bar{y}_{opt}) = V_{min}\,(\bar{y}_{st})} = \frac{(\mathbf{W_h\,S_h})^2}{\mathbf{n}} - \frac{\sum_{h=1}^{L}\mathbf{W_h\,S_h^2}}{\mathbf{N}}$$

Cochran (1977) stated that optimum allocation is better than proportional allocation in a population that consists of large and small institutions, stratified by some measure of size. Variance $\mathbf{S_h^2}$ is usually much greater for the large institutions than for the smaller ones, making proportional allocation inefficient. In some survey in which some strata are more expensive to sample than the others, the influence of the factor $\sqrt{\mathbf{C_h}}$ may make proportional allocation poor. (Cochran,1977)

2.4 Proportional Allocation

This allocation calls for making the stratum sample size $\mathbf{n_h}$ proportional to stratum size $\mathbf{N_h}$. That is it requires us to allocate large sampling units to large stratum and small sampling units to a small stratum. Thus, a representative sample of the population units is obtained with this allocation since the sample sizes $\mathbf{n_1, n_2, \dots, n_L}$ are **proportional** to the stratum sizes $\mathbf{N_1, N_2, . . . ,N_L}$ i.e.

$\frac{\mathbf{n_h}}{n} = \frac{\mathbf{N_h}}{\mathbf{N}}$ or $\frac{\mathbf{n_h}}{\mathbf{N_h}} = \frac{n}{\mathrm{N}}$ or $\mathbf{f_h} = \mathrm{f}$. Proportional allocation is given by the expression:

$$\mathbf{n_h} = \frac{n\mathbf{N_h}}{\mathbf{N}} = n\mathbf{W_h}\mathbf{,\ h = 1, 2, \dots, L}$$

This allocation method could also be deduced from **Neyman's** allocation when the cost $\mathbf{C_h}$ and variances $\mathbf{S_h^2}$ for **h = 1, 2, …, L,** are constant (the same) through all the = 1, 2, . . . , L, are constant (the same) through all the strata.

The variance $\mathbf{V(\bar{y}_{st})}$ of the mean in stratified random sampling when proportional allocation is used is given as:

$$\mathbf{V(\bar{y}_{st})_{prop}} = \frac{(1-\mathbf{f_h})}{n}\sum_{h=1}^{L}\mathbf{W_h\,S_h^2} = \frac{\sum_{h=1}^{L}\mathbf{W_h\,S_h^2}}{\mathbf{n}} - \frac{\sum_{h=1}^{L}\mathbf{W_h\,S_h^2}}{\mathbf{N}}$$

If stratified random sampling with proportional allocation is adopted, and the variance in all the strata have the same value $\mathbf{S_w^2}$**,** then:

$$\mathbf{V(\bar{y}_{st})} = \left(\frac{\mathbf{N-n}}{N}\right)\left(\frac{\mathbf{S_w^2}}{\mathbf{n}}\right) \qquad \text{(Cochran,1977)}$$

2.5 Criteria for Evaluation

In line with the stated objectives of this study these statistical tools are used to evaluate various methods of allocating sample to strata considered in this research for a fixed sample size **n**. The **MSE** of the estimates is of paramount importance among the criteria considered for evaluation. The one with minimum **MSE** is adjudged the best method to use when stratification technique is desired.

EFFICIENCY: - Okafor (2002) defined a sampling strategy by the pair $\mathbf{T[D,\hat{\phi}]}$ and that for two different estimators $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ of the parameter ϕ obtained using the same sampling design, $\hat{\boldsymbol{\theta}}_1$ is said to be uniformly better than or more efficient than $\hat{\boldsymbol{\theta}}_2$ if and only if $\mathbf{MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)}$. The efficiency of any strategy $\mathbf{T}_1$ compared to $\mathbf{T}_2$ for a parameter ϕ or the relative efficiency of $\mathbf{T}_1$ over $\mathbf{T}_2$ is expressed as:

RE = MSE ($\mathbf{T_2}$) / MSE ($\mathbf{T_1}$)

ACCURACY:- This is a measure of closeness of an estimate to its parameter. This is judged by the **MSE** as it includes the effect of bias. The error in the estimation of θ is $\hat{\theta} - \theta$. The Loss function $\mathbf{L}(\hat{\theta} - \theta)$ is the loss incurred through the error $(\hat{\theta} - \theta)$.The commonly used loss function is the squared error loss function denoted as:

$$\mathbf{L}(\hat{\theta} - \theta) = (\hat{\theta} - \theta)^2$$

$$\mathbf{E}[\mathbf{L}(\hat{\theta} - \theta)] = \mathbf{E}(\hat{\theta} - \theta)^2$$

$$= \sum_{r \epsilon S} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 \ \mathbf{P}(\mathbf{S})$$

(Okafor,2002) and is called the Mean Squared Error **(MSE)** and $\mathbf{P}(\boldsymbol{S})$ is the probability over the sample space **S**. It has been proved by Cochran (1977) that $\mathbf{MSE}(\hat{\boldsymbol{\theta}}) = \mathbf{V}(\hat{\boldsymbol{\theta}}) + \{\mathbf{B}(\hat{\boldsymbol{\theta}})\}^2$

where $\mathbf{B}(\hat{\boldsymbol{\theta}})$ is called the bias of the estimator $\hat{\boldsymbol{\theta}}$ and it is the difference between the expected value of the estimator and the parameter **θ**. When the estimator $\hat{\boldsymbol{\theta}}$ is unbiased for **θ**, then $\mathbf{MSE}(\hat{\boldsymbol{\theta}}) = \mathbf{V}(\hat{\boldsymbol{\theta}})$.

Thus, the MSE of stratified sampling is given as:

$$\text{MSE}\ (\bar{y}_{\mathbf{st}}) = \sum_{h=1}^{\mathbf{L}} \mathbf{W_h^2} \ \frac{\mathbf{S_h^2}}{\mathbf{n_h}} (\mathbf{1} - \mathbf{f_h}) + [\sum_{h=1}^{\mathbf{L}} (\mathbf{w_h} - \mathbf{W_h}) \overline{\mathbf{Y}}_{\mathbf{h}}]^2$$

The second part in the RHS gives the bias associated with a particular method.

when Neyman's allocation is used:

$$\mathbf{MSE}\ (\bar{y}_{\mathbf{st}})_{\mathbf{ne}} = \mathbf{V}(\bar{y}_{\mathbf{st}})_{\mathbf{ne}} + [\sum_{h=1}^{\mathbf{L}} (\mathbf{w_h} - \mathbf{W_h}) \overline{\mathbf{Y}}_{\mathbf{h}}]^2$$

also, when equal allocation is used:

$$\mathbf{MSE}\ (\bar{y}_{\mathbf{st}})_{\mathbf{eq}} = \mathbf{v}\ (\bar{y}_{\mathbf{st}})_{\mathbf{eq}} + [\sum_{h=1}^{\mathbf{L}} (\mathbf{w_h} - \mathbf{W_h}) \overline{\mathbf{Y}}_{\mathbf{h}}]^2$$

while for Proportional allocation:

$$\mathbf{MSE}\ (\bar{y}_{\mathbf{st}})_{\mathbf{prop}} = \mathbf{v}\ (\bar{y}_{\mathbf{st}})_{\mathbf{prop}} + [\sum_{h=1}^{\mathbf{L}} (\mathbf{w_h} - \mathbf{W_h}) \overline{\mathbf{Y}}_{\mathbf{h}}]^2$$

3.0 Research Methodology

Sample survey was the adopted research design. The population of this study includes all the students of Federal Polytechnic Oko, Atani Campus, which is one thousand, five hundred and fifty students (1550) from two departments viz: Mathematics/Statistics and Computer Science, and two programmes of study-National Diploma(ND) and Higher National Diploma (HND). The researcher used stratified random sampling technique to select 120 samples. The population size was assumed to be in four strata according to students department and program of study:ND-Computer Science, ND-Mathematics/Statistics, HND-Computer Science, HND-Mathematics/Statistics. Equal allocation procedure, Proportional allocation procedure and Neyman's allocation procedure were used to allocate sample size to the various strata respectively, which are summarized in the table below:

Table 1: Neyman's Sample size allocation

Neyman's Stratum allocation procedure	**Equal allocation**	**Proportional allocation procedure**	**procedure**
ND-Mathematics/Statistics	25	2	3
ND-Computer Science	25	47	50
HND-Mathematics/Statistics	25	7	8
HND- Computer Science	25	64	59
TOTAL	120	120	120

The researcher collected secondary data from the department of Examinations and Records of Federal Polytechnic Oko, Atani Campus. The secondary data collected were comprised of Grade Point Average (GPA) of Second Semester Result for 2015/2016 Academic Session for: The estimated mean, variance and mean square error (MSE) of students' performance using equal allocation, Proportional allocation and Neyman's allocation respectively were pictorially presented. An estimate of the mean, variance and mean square error (MSE) of students' performance using Equal allocation, Proportional allocation and Neyman's allocation respectively were obtained. Also Z-Test Statistics was used to test the hypothesis at 5% level of significance.

4.0 DATA ANALYSIS

The estimated mean, estimated variance and the estimated mean square error (MSE) were presented on the table, according to the research question and was analyzed.

Table 2 Estimated Mean Performance of Students' for Stratified Sampling Using Different Allocation Procedures:

Equal Allocation Procedure	2.96
Proportion Allocation Procedure	2.90
Neyman's Allocation Procedure	2.90

Figure 1 Pie chart showing the Estimated Mean Performance of Students' for Stratified Sampling Using Different Allocation Procedures

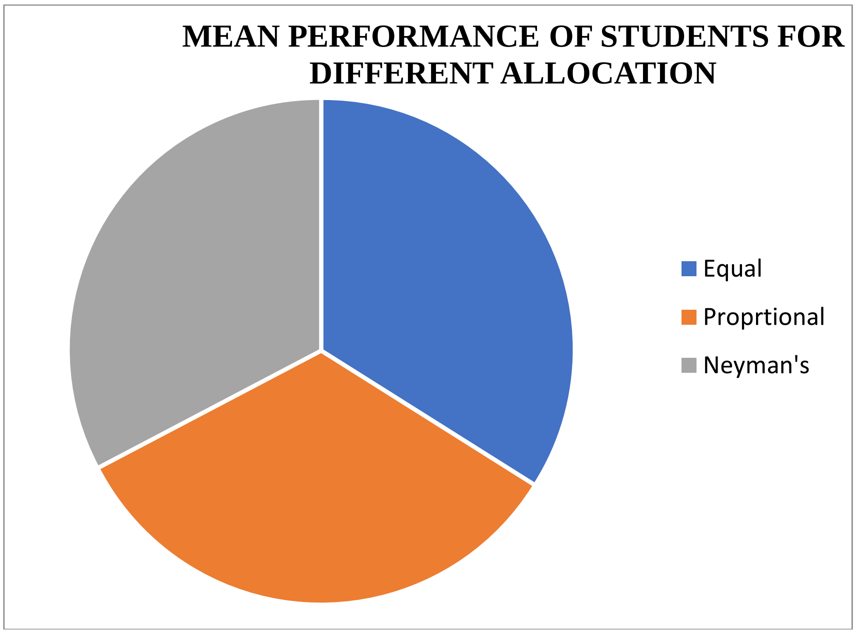


Table 3 Estimated Variance of the Mean Performance of Students Per Stratum Based On Different Allocation Procedures

Equal Allocation Procedure	0.001138
Proportion Allocation Procedure	0.001982
Neyman's Allocation Procedure	0.001075

Figure 2 Bar chart showing the Estimated Variance of Mean Performance of Students' for Stratified Sampling Using Different Allocation Procedures

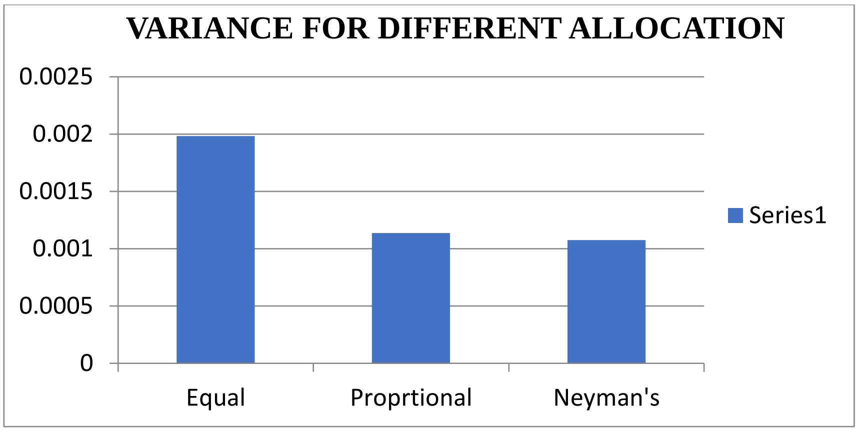
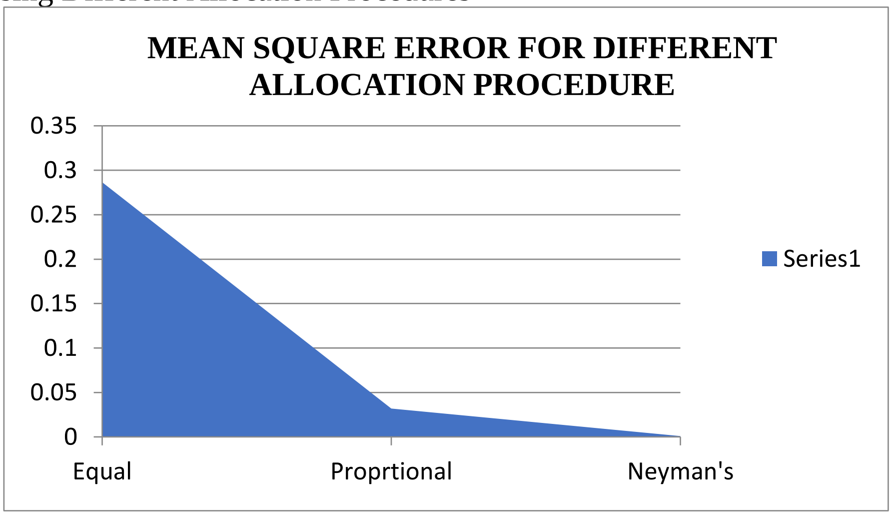


Table 5 Estimated MSE of the Mean Performance of Students Per Stratum Based On Different Allocation Procedures

Equal Allocation Procedure	0.2865
Proportion Allocation Procedure	0.0320
Neyman's Allocation Procedure	0.0010

Figure 3 Area chart showing the MSE of Mean Performance of Students' for Stratified Sampling Using Different Allocation Procedures



Also, the 95% Confidence Interval for the estimated mean performance using Neyman's allocation procedure yielded a lower and upper limit of 2.83 and 2.97, respectively. Using Z-test statistics, the null hypothesis was not rejected and a conclusion was made that the estimated mean performance of the students using Neyman's allocation procedure is equal to the stratified population mean.

5.0 Conclusion

This study attempted to estimate the mean, variance and Mean Square Error (MSE) of the academic performance of the students, using three different allocation procedures in stratified sampling technique viz: equal allocation, proportional allocation and Neyman's allocation procedures. To verify the claim, that Neyman's allocation procedure is more efficient than other procedures using mean square error (MSE) criterion In the selection of sample in each stratum, for equal allocation, 25 samples were taken from each stratum making a total of 120 samples across 4 strata (students department and programme of study). For proportional allocation, stratum sample size varies depending on the total number of students in each stratum (students department and programme of study). Variability of samples from each stratum (students department and programme of study) with the equal cost of obtaining the samples were put into consideration in Neyman's allocation. Few samples were selected from the stratum where the variability is low, while many samples were selected from the stratum where the variability is high.

The mean, variance and mean square error under Neyman's allocation procedure were 2.90, 0.001075 and 0.0010 respectively. For the proportional allocation procedure, the mean was 2.90, the variance was 0.001982 and the mean square error was 0.0320, while the equal allocation procedure gave a mean of 2.90, variance of 0.001138 and mean square error of 0.2865. Neyman's allocation procedure gave the least mean square error. This was followed by Proportional allocation and Equal allocation.
Neyman's allocation procedure was verified to be the best selection procedure in this paper when the mean square error criterion was used. The result obtained from this investigation does not differ from the assertion of (Olayiwola et al., 2013)

Further investigation of this research was made through 95% confidence interval computation and hypothesis testing. The 95% confidence interval computed indicated mean that the estimated mean performance of the students using Neyman's allocation procedure lies between 2.83 and 2.97. The hypothesis testing showed also, that the estimated mean performance of the students using Neyman's allocation procedure is equal to the stratified population mean, which verified the proof that sample mean, is an unbiased estimator of a population mean. The researcher concluded that for estimating the mean, variance and mean square error (MSE) of the academic performance of students, out of all the three sample allocation procedures considered in this empirical study, Neyman's allocation procedure is the best and hence the most efficient when Mean Square Error (MSE) criterion is used.

REFERNCES

Adebola, F.B and Ajayi, O.S (2014) Comparison of allocation procedures in a stratified random sampling of skewed populations under different distributions and sample sizes *IJISET - International Journal of Innovative Science, Engineering & Technology,*

Kareem, A.O(2012) *An Empirical Investigation On The Choice Of An Efficient Strata Boundary Determination Method.*

Cochran, W.G. (1977). *Sampling Techniques, Third edition: John Wiley and Sons, New York.*

Dalenius, T. and Hodges, J.L., Jr. (1959) *"*Minimum Variance Stratification" *JASA, 54, pp. 88 – 101.*

Diaz-Garcia, J.A and Cortez, L.U. 2006. Multi-Objective Optimisation For Optimum Allocation
In Multivariate Stratified Sampling. *Survey Methodology,*
Horgan, J. M. (2006). "Stratification Of Skewed Populations: *A Review". International Statistical Review, 74, 1, Pp.67-76.*

Ogunyinka, P. I., Sodipo, A. A. (2006), Efficiency of Ratio and Regression Estimators Using double; Sampling ;*Journal of Natural Sciences Research* www.iiste.org *ISSN 2224-3186*
(Paper) ISSN 225-0921 (Online).

Okafor, F. C. (2002) Sample Survey Theory with Applications *Afro-Orbis Publications Ltd. Nsukka, Nigeria.*

Olayiwola Olaniyi, Apantaku Fadeke, Bisira Hammed, Adewara Adedayo (2013). Efficiency of
Neyman allocation procedure over other allocation procedures in stratified random sampling, *American Journal of Theoretical and Applied Statistics2013;*
.
Satterthwaite, F. (1946), An Approximate Distribution of Estimates of Variance Components,Biometrics, 2.

Ugwu, P. (2003), Digestive Principle and Techniques of Statistics,*23 Neni Street Ogui-New layout Enugu.*