

## On the Choice of an Efficient Sampling Scheme within Strata

**Akeem O. Kareem**

Institute for Security Studies,  
Abuja,  
Nigeria

**Amos A. Adewara & Gafar M. Oyeyemi**

Department of Statistics,  
University of Ilorin,  
Ilorin, Nigeria.

### ABSTRACT

The choice of an efficient sampling scheme within strata is one of the technical operations involved with the use of Stratified Sampling design in Survey Sampling. Literature had concentrated on the use of Simple Random Sampling (SRS) as a choice scheme within the strata. This study examine systematic sampling scheme within the strata and proved to be more precise than SRS ( $V(\bar{y}_{st})_{SYS} < V(\bar{y}_{st})_{SRS}$ ) for both optimum and proportional allocation using any method of strata boundaries determination.

**KEYWORDS:** Precision, Stratified Sampling, systematic sampling, simple random sampling, optimum allocation and proportional allocation.

### INTRODUCTION

Stratified sampling is a methodology in which the elements of a heterogeneous population are classified into mutually exclusive and exhaustive homogenous subgroups called strata based on one or more important characteristics. Depending on the sampling scheme employed in selecting the samples from the stratum independently from other strata, Stratified sampling becomes Stratified random sampling when simple random sampling (SRS) is employed in selecting the samples and when systematic sampling is used it becomes Stratified systematic sampling (SYS). In practice, the former is common; therefore the latter is examined in this study in order to determine an efficient sampling scheme within the strata.

Literature adequately reports stratified sampling in terms of procedures and methods of estimation: Murthy (1967); Sukhatme and Sukhatme (1984); Mendenhall et al. (1971); Cochran (1977); Raj and Chandhok (1998) and Okafor (2002). In order to reduce the amount of variations contained within the samples and obtain more precise estimates, we employ stratified sampling technique which was described by Mendenhall et al. (1971) as a sampling

procedure which separates the population units into non-overlapping groups called strata and thereafter select sample independently from each stratum. Okafor (2002) on his part, referred to it as the procedure of drawing independent samples after grouping the whole units in the population into homogenous distinct strata as stratified sampling. When SRS is used to select sample in each stratum, the procedure is called stratified random sampling. Thus, classes of stratified sampling derived their names from the sampling schemes employed in drawing the samples from the strata. When SYS is employed, we have stratified systematic sampling.

However, we shall be satisfied with the definition of Cochran (1977) which states that “In stratified sampling, the population of  $N$  units is first divided into  $L$  subpopulations of  $N_1, N_2, \dots, N_L$  units, respectively. These subpopulations are non-overlapping and together they comprise the whole of the population, such that  $N_1 + N_2 + \dots + N_L = N$ . i. e.  $\sum_{h=1}^L N_h = N$ . The subpopulations are called strata. If SRS of sizes  $n_1, n_2, \dots, n_L$  respectively are taken from each stratum, such that  $\sum_{h=1}^L n_h = n$  the whole procedure is described as stratified random sampling.

It is pertinent to state that SRS dominated literature as the common sampling scheme used within strata to study efficiencies of methods of strata construction. In this study therefore, we examine the use of SYS sampling as an alternative sampling scheme within the strata using optimum and proportional allocations and the results compared with that of SRS.

## 1. Sampling Scheme within Strata

Choice of sampling scheme within strata is one of the design problems when stratified sampling technique is employed. Empirical investigation in the literature is dominated with the use of SRS within the strata. Okafor (2002, pp.98) dwelt only on stratified random sampling. Murthy (1967, pp. 255) dealt extensively on stratified random sampling and gave expressions for the mean and variance in Stratified Probability Proportional to Size (PPS) Sampling. Cochran (1977) exhaustively examined stratified random sampling and gave an expression for the mean and variance of stratified systematic sampling (See section 8.12, pp.226).

### 1.1 Stratified systematic sampling

Cochran (1977) describes systematic sampling thus: suppose that  $N$  units in the population are numbered 1 to  $N$  in some order. To select a sample of  $n$  units, we take a unit at random from the first  $k$  units and every  $k^{\text{th}}$  unit thereafter. The selection of the first  $k^{\text{th}}$  units determined the whole sample. This is called an every  $k^{\text{th}}$  systematic sample. Murthy (1967) states that systematic sampling is operationally more convenient and at the same time saves time while ensuring equal probability of inclusion of each unit in the sample.

Murthy (1967), Cochran (1977), Raj and Chandhok (1998) and Okafor (2002) have all mentioned that systematic sampling can be looked into in another way in relation to cluster sampling. They explained that in a population with  $N = nk$ , the population can be divided into  $k$  large systematic sampling units each containing  $n$  of the original  $n$  units. The operation of

## On the Choice of An Efficient Sampling Scheme Within Strata

choosing a randomly located systematic sample is just the operation of choosing one of these large sampling units at random. Thus, systematic sampling amounts to selecting of a simple random sample of one cluster unit from a population of  $k$  cluster units with probability  $1/k$ .

In practice, it is common to encounter situations in which  $N \neq nk$ . Various suggestions had been made on how to handle such a situation. Lahiri (1952) suggested the Circular Systematic Sampling (CSS); Murthy (1967) suggested that when  $N \neq nk$ , i.e., the population units  $N$  cannot be divided into  $K$  clusters of equal size, and therefore choose the interval  $k$  to be the nearest integer to  $\frac{N}{n}$  resulting in  $n'$  which may not necessarily be equal to  $n$ , the required sample size. Another approach when  $N \neq nk$  is the use of fractional interval also reported by Murthy (1967). Leu and Tsui (1996) developed the New Partially Systematic Sampling (NPSS) in order to derive an unbiased estimator of the variance of  $V(\bar{y}_{st})$ . Thus, NPSS combines SRS with systematic sample to obtain its estimates thereby deviating from the objective of this study as we intend to observe the performance of stratified systematic sampling over stratified random sampling as a choice scheme within strata. Remainder Linear Systematic Sampling (RLSS) due to Chang and Huang (2000) is also an approach when  $N \neq nk$ . Competing methods therefore are; CSS, NPSS, and RLSS. Due to its greater efficiency over the CSS and NPSS as reported by Chang and Huang (2000), RLSS is used in this study when  $N_h \neq n_h k_h$  in each stratum. Therefore when any of these procedures is applied in selecting samples from each stratum, whether  $N = nk$ , or not we have stratified systematic sample.

Murthy (1967), Cochran (1977), Raj and Chandkok (1998) and Okafor (2002) have all itemized reasons why stratification technique is commonly used to include:

- i. the need for estimates with known precision for subpopulations;
- ii. administrative conveniences may dictate the use of stratified sampling; most especially when there are sub-stations in a statistical organization and such offices could be directed to collect data on a subject of interest in their domain;
- iii. differences between strata may necessitate different sampling methods; and
- iv. when increased precision for the whole population is required.

## 2. Design Problems Associated with Stratified Sampling

Dalenius and Hodges (1959), Hess et al. (1966), Wang and Aggrawal (1984), Horgan (2006) and Okafor (2002) itemized the following as specific design problems involved in stratification processes:

- (a) the choice of a stratification variable;
- (b) the choice of number of strata  $L$  to be formed;
- (c) mode of stratification i.e. the way/manner in which strata boundaries are determined;

- (d) the choice of sample size  $n_h$  to be taken from the  $h^{\text{th}}$  stratum. i.e. the problem of allocation of sample size to strata; and
- (e) choice of sampling design within strata.

Solutions have also been suggested to most of these problems by different authors as examined below. On the choice of stratification variable, Cochran (1977) enjoined the use of the frequency distribution of the study variate itself as stratification variable if available or that of variable  $X$  (an auxiliary variable) which is highly correlated with the study variate and perhaps the value of variable  $Y$  at a recent census. Same view was expressed by Sukhatme & Sukhatme (1984). Raj and Chandhok (1998) opined that the basic consideration in stratification is to ensure that strata are internally homogenous; therefore all prior knowledge, personal intuition and judgment have to be brought to play in ensuring similarity within strata. Hess et al. (1966) examined the significance of a highly correlated auxiliary variable as choice of stratification variable while Kish and Anderson (1977) took it to a logical conclusion in a multivariate stratification study. Almost all references cited emphasized the need for the use of a highly correlated auxiliary variable for the purpose of stratification. However in practice, the estimation variable is commonly used for the purpose of stratification, e.g., Dalenius (1959), Ghosh (1963), Hess et al. (1966) and Hedlin (2000).

Concerning the number of strata  $L$  to be constructed, in most of the surveys, the number of strata is predetermined; while in others, optimum number of strata is believed to have been attained when there is no further gain in precision by increasing the number of strata. This study allows for optimum number of strata and the stratification process continued until when deep stratification occurs, i.e.,  $N_h = 1, \forall h = 1, 2, \dots, L$  (at least one population units in one or more stratum). However, Cochran (1977) developed a model representing the approximate reduction in the precision of stratified sample mean compared to that obtained with SRS and concluded that beyond six strata,  $L \geq 6$ , there is little or no further gain, in terms of precision. This was premised on the following two basic questions Cochran (1977) said should be considered to efficiently determine the number of strata:

- (a) at what rate does the variance of  $V(\bar{y}_{st})$  decrease as  $L$  is increases; and
- (b) how is the cost of the survey affected by an increase in  $L$ ?

Thus, when there is no appreciable gain in precision with additional strata, then optimum number of strata has been reached and when the cost of sampling additional strata is already overshooting the survey budget, it is obvious that the number of strata to be surveyed should be limited to the one covered by the survey budget.

Hess et al. (1966) confirmed that efficient number ( $L$ ) of strata can be arrived at by observing the  $Var(L)/Var(L - 1)$  reduction in variance attained when additional stratum is considered, with the remark that in many multipurpose investigations, only marginal stratification gains could be expected from the use of more than six strata. Ghosh (1963), Hess et al. (1966), Hidirolou (1986), Hedlin (2000), Gunning and Horgan (2004), Kozark (2004),

### On the Choice of An Efficient Sampling Scheme Within Strata

Keskintur and Er (2007) etc, have all studied methods of strata construction using predetermined number of strata.

On the issue of placing the population unit into its exact stratum, that is, strata boundary determination, various methods have been proposed by different authors in the literature. Dalenius (1950) took the lead, Equalization of Strata Totals (EST) was due to Mahalanobis (1952), Dalenius and Hodge (1959) suggested  $\text{cum}\sqrt{f(y)}$  method, here referred to as Dalenius and Hodge's Rule (DHR). Others are Ekman's Rule (1959) (EKR), Durbin's Rule (1959) (DUR), Sethi's Rule (1963) (STR), Thompson Rule (1976) (TNR), Lavalley and Hidirolou Method (1988) (LHM) and Extended Ekman's Rule (EEKR) by Hedlin (2000). Random Search Method (RSM) was due to Kozark (2004), Geometric Stratification (GMS) to Gunning and Horgan (2004) and Genetic Algorithm (GA) to Keskintur and Er (2007).

It is pertinent to mention that of all the above methods, two are popularly in use: the DHR and the GMS because of simplicity in their application. In addition, EST and STR are also considered in this study due to their simplicity in application. Detailed procedures of these four methods and their applications are discussed in section 3.

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, if the variability and cost per element are not known in advance, the general rule is to take a large sample from the stratum that has a large population unit (size) and a small sample from the stratum that has a small stratum population size. Literature had extensively dwelt on the subject matter with the following result: Optimum allocations due to Neyman (1934), Sethi (1963) determines his optimum points of stratification (OPS) for various distributions using equal, proportional and Neyman allocations. Hess et al. (1966) used equal, Neyman and proportional allocations in their studies; same allocation procedure was employed by Kish and Anderson (1977). Compromise allocation was by Chatterjee (1968); it was used and improved upon by Khan and Ahsan (2003). While Judez and Chaya (1999, 2000), Hedlin (2000), Kozark (2004), and Gunning and Horgan (2004) used the Neyman allocation with a predetermined sample size  $n$ . Power allocation was used by Lavalley and Hidirolou (1988), Keskinturk and Er (2007) examined Genetic allocation in comparison with equal, Neyman and proportional allocations. This study makes use of the optimum and proportional allocations for its empirical investigation. Details of these allocation methods are discussed in section 4. Sample size  $n$  is predetermined for all the sets of data used in this study and sample estimation was restricted to strata formation in which  $n_h \geq 2$ .

With all the aforementioned, it is hereby assumed that the strata had been formed and next is the sample selection from each stratum. SRS is the commonly used sampling scheme within the strata due to its simplicity in terms of sample selection and estimation. This study, without deviating from the common practice, examines another sampling scheme within the strata by making use of systematic sampling. Sample selection and estimation procedures for this new approach are reflected in detail in section 5.

### 3. Method of Strata Boundary Determination

#### 3.1 Dalenius and Hodge Rule (DHR)

DHR requires us to choose equal class interval, obtain the cumulative square root of the frequency ( $\text{cum}\sqrt{f(y)}$ ) of the study variate and determine the strata boundaries by dividing the total cumulative square root of the frequency by the required number of strata  $L$  and the boundary is placed at this division point.

Dalenius (1950) was credited with the first statistical study into the problem of strata boundary determination. He found the optimum stratification points for Neyman allocation to be those which satisfy the equation:

$$\frac{\sigma_h^2 + (X_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (X_h - \mu_{h+1})^2}{\sigma_{h+1}}, h = 1, 2, \dots, L-1, L \quad (1)$$

and for proportional allocation to the number of units, the general expression for the simultaneous equation is:

$$X_h = \frac{(\mu_h + \mu_{h+1})}{2}; h = 1, 2, \dots, L-1 \quad (2)$$

while Cochran (1961) and Sethi (1963) reported that the general equation to be satisfied in order to obtain optimum stratification points for equal allocation is given

as: 
$$W_h [\sigma_h^2 + (X_h - \mu_h)^2] = W_{h+1} [\sigma_{h+1}^2 + (X_h - \mu_{h+1})^2], h = 1, 2, \dots, L-1, L \quad (3)$$

These equations are solved through various steps of iteration to obtain Optimum Points of Stratification (OPS) and these equations are derived on the assumptions that the estimation variable is the stratification variable and that the frequency distribution is continuous.

For details on derivation of these sets of general equations, see Dalenius (1957, 1959), Murthy (1967, section 10.7a, pp.262), Sukhatme and Sukhatme (1970, section 3.11, pp. 108), Cochran (1977, section 5A.7, pp.127), Raj and Chandhok (1998, section 4.8, pp 107) and Okafor (2002, section 4.6, pp. 120)

#### 3.2 Equalization of Strata Totals (EST)

Mahalanobis (1952) suggested Equalisation of Strata Totals (EST) which requires us to form stratum boundaries for equal allocation by making the product of the stratum weight and the mean equal to a constant:

$$W_h \mu_h = C, h = 1, 2, \dots, L-1, L \quad (4)$$

### 3.3 Sethi's Rule (STR)

Sethi (1963) solved the sets of general equations (1) – (3) in section 3.1 for some standard distributions (Normal, Beta, Gamma and various Chi-squares) with a view to evolve a good stratification system and obtained corresponding OPS for the said distributions. He also tabulated the optimum boundaries for Neyman, equal and proportional allocations for  $L \leq 6$  (for Gamma distribution) and  $L \leq 10$  for Normal distribution. (See comprehensive Sethi's Table for these distributions in Sethi (1963), Hess et al. (1966) and Murthy (1967, pp.267-269). Sethi's procedure is referred to as Sethi's Rule (STR).

### 3.4 Geometric Stratification (GMS)

Gunning and Horgan (2004) introduced the new and now the most commonly used method of strata boundary determination called Geometric Stratification (GMS). It was applied to positively skewed populations and results compared with DHR. Stratum boundaries are automatically formed with GMS once the geometric ratio  $r$  is determined.

$$r = [\max Y_i / \min Y_i]^{1/L}$$

$$r = [Y_L / Y_0]^{1/L} \quad (5)$$

where  $Y_L$  is the highest value and  $Y_0$  is the smallest value of the study variate  $Y$ . The boundaries are at the points:

Minimum  $K_0 = a, ar, ar^2, \dots, ar^L = \text{Maximum } K_L$ .

The general term is:

$$K_h = ar^h, h = 0, 1, 2, \dots, L - 1$$

Details of the GMS algorithm are in section 2 of Gunning and Horgan (2004). The simplicity of GMS had been extended to Pareto distribution by Gunning and Keogh (2006) and was found to be more efficient than DHR.

## 4. Allocation of Samples to Strata

Once we select at least a sample from each stratum the procedure of stratified sampling is satisfied. However, this study requires at least two units in each stratum for estimation purposes. After the sample size  $n$  is chosen, there are many ways of allocating  $n$  into individual stratum sizes  $n_1, n_2, \dots, n_L$  with the aim of using an allocation method that gives a specified amount of information at minimum cost. Available allocation methods have being highlighted in section 2. The two methods used in this study are discussed below.

#### 4.1 Optimum Allocation

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

$$Cost = C = C_0 + \sum_{h=1}^L C_h n_h \quad (6)$$

where  $C$  is the total cost of the survey  
 $C_0$  is the overhead cost and  
 $C_h$  is the cost per unit in stratum  $h$

The cost is always proportional to the size of sample, but the cost per unit in each stratum,  $C_h$ , may vary from stratum to stratum. The above cost function in (6) 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 (6), the appropriate allocation which minimizes cost for a fixed value of  $V(\bar{y}_{st})$  or minimizes  $V(\bar{y}_{st})$  for a given cost is given as:

$$n_h = \frac{nW_h S_h / \sqrt{C_h}}{\sum W_h S_h / \sqrt{C_h}} \quad (7)$$

Cochran (1977 pp.98) stated that the use of this allocation method leads to following rules of conduct. In a given stratum, take a larger sample if:

1. The stratum is larger.
2. The stratum is more variable internally
3. 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.  $C_h = C$  for  $h = 1, 2, \dots, L$ ), then expression (7) above reduces to:

$$n_h = \frac{nN_h S_h}{\sum N_h S_h} \quad (8)$$

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:

$$V_{opt} = V_{min}(\bar{y}_{st}) = \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^2}{N} \quad (9)$$

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  $S_h^2$  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  $\sqrt{C_h}$  may make proportional allocation poor.

#### 4.2 Proportional Allocation

This allocation calls for making the stratum sample size  $n_h$  proportional to stratum size  $N_h$ . 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  $n_1, n_2, \dots, n_L$  are proportional to the stratum sizes  $N_1, N_2, \dots, N_L$  i.e.  $\frac{n_h}{n} = \frac{N_h}{N}$  or  $\frac{n_h}{N_h} = \frac{n}{N}$  or  $f_h = f$ . Proportional allocation is given by the expression:

$$n_h = \frac{nN_h}{N} = nW_h, \quad h = 1, 2, \dots, L \quad (10)$$

This allocation method could also be deduced from Neyman allocation when the cost  $C_h$  and variances  $S_h^2$  for  $h = 1, 2, \dots, L$ , are constant (the same) through all the strata. Expression (7) reduces to (10). i.e.

$$n_h = \frac{nN_h S_h}{\sum N_h S_h} \text{ becomes } n_h = \frac{nN_h}{\sum N_h} = \frac{nN_h}{N} = nW_h, h = 1, 2, \dots, L$$

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

$$V(\bar{y}_{st})_{prop} = \frac{(1-f)}{n} \sum W_h S_h^2 = \frac{\sum W_h S_h^2}{n} - \frac{\sum W_h S_h^2}{N} \quad (11)$$

This allocation is often employed due to its simplicity. It is also good when  $S_h^2$  is the same in all strata and when there is no knowledge of relative size of within strata variances.

### 5. Estimation Procedures in Stratified Sampling

#### 5.1 Stratified Random Sampling

This section discusses estimation procedure in stratified random sampling while the subsequent section discusses that of stratified systematic sampling. Symbols and notations of Cochran (1977, pp.90) were adopted in this study.

##### Notations

The subscript  $h$  denotes the stratum and  $i$  the unit within the stratum, for  $h = 1, 2, \dots, L$ .

- $L$  = Number of strata.
- $N_h$  = Total number of population units in stratum  $h$ .
- $n_h$  = Number of sample units taken in stratum  $h$ .
- $N$  = Total number of population units in all the  $L$  strata
- $n$  = Sample size of the study
- $Y_{hi}$  = The observed value of the  $i^{\text{th}}$  unit in the  $h^{\text{th}}$  stratum

$W_h = N_h/N = \text{stratum weight (population units)}$

$w_h = n_h/n = \text{stratum weight (sample units)}$

$$\text{Sample mean} = \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \quad (12)$$

$$\text{True mean} = \bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi} \quad (13)$$

$$\text{Sample Variance} = S_h^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \quad (14)$$

$$\text{True Variance} = S_h^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2 \quad (15)$$

$$\text{Population Mean} = \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h \quad (16)$$

$$\text{Variance of the population mean} = V(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 V(\bar{y}_h) = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{S_h^2}{n_h} \quad (17)$$

## 5.2 Estimation in Stratified Systematic Sampling

Literature had dealt on the significance of the arrangement of the population units on the precision of SYS  $V(\bar{y}_{st})$ , Madow (1953), Guatschi (1957), Murthy (1967, section 5.8, pp.153), Mendenhall et al. (1971, pp.151) Cochran (1977, pp.227), Raj and Chandhok (1998). Cochran (1977, pp.208) has given a corollary that the mean of a systematic sample will be more precise than that of SRS if and only if  $S_{wsy}^2 > S^2$ , where  $S_{wsy}^2$  is the weighted variance of all possible systematic samples.

### Notations

Cochran (1977, pp.91) stated that expressions for the mean and variance of stratified sampling applied generally to all classes of stratified sampling and not restricted to stratified random sampling. Therefore, all notations in section 4 above are also valid for stratified systematic sampling.

The subscript  $h$  denotes the stratum and  $i$  the unit within the stratum.

The subscript “sy” in this section denotes systematic sample.

$$\bar{y}_{syh} = \frac{1}{k_h} \sum_{i=1}^{k_h} \left( \sum_{j=1}^{n_h} \frac{y_{ij}}{n_h} \right) \quad (18)$$

is the mean of systematic sample in stratum  $h$ ,

$$\bar{y}_{stsy} = \sum W_h \bar{y}_{syh} \quad (19)$$

is the population mean of the stratified systematic sample.

For the Variance,

- i. When  $N_h = n_h k_h$

### On the Choice of An Efficient Sampling Scheme Within Strata

$$V(\bar{y}_{syh}) = \frac{(N_h - 1)S_h^2}{N_h} - \frac{k_h(n_h - 1)S_{wsyh}^2}{N_h} \quad (20)$$

is the variance of systematic sample in stratum h, where

$$S_{wsyh}^2 = \frac{1}{k_h(n_h - 1)} \sum_{i=1}^{K_h} \sum_{j=1}^{n_h} (y_{ijh} - \bar{y}_{ht})^2$$

and

$$V(\bar{y}_{stsy}) = \sum W_h^2 V(\bar{y}_{syh}) \quad (21)$$

is the variance of the population mean of stratified systematic samples.

ii. when  $N_h \neq n_h k_h$  the mean and the variance of RLSS is used and given below:  
(See relation 2.2 and 2.3 pp.251 of Chang and Huang (2000)).

$$\bar{y}_{sy(RLSS)} = \frac{(n-r)k*\bar{y}_{1i} + r(k+1)*\bar{y}_{2i}}{N} \quad (22)$$

$$V(\bar{y}_{sy(RLSS)}) = \frac{1}{N^2} \left\{ (n-r)^2 k^2 \left[ \frac{1}{k} \sum_{i=1}^k (\bar{y}_{1i} - \bar{Y}_1)^2 \right] + r^2 (k+1)^2 \left[ \frac{1}{(k+1)} \sum_{i=1}^{k+1} (\bar{y}_{2i} - \bar{Y})^2 \right] \right\} \quad (23)$$

To suite our applications, expression (22) and (23) are modified as follows:

$$\bar{y}_{syh(RLSS)} = \frac{(n_h - r_h)k_h*\bar{y}_{1i} + r_h(k_h+1)*\bar{y}_{2i}}{N_h} \quad (24)$$

$$V(\bar{y}_{syh(RLSS)}) = \frac{1}{N_h^2} \left\{ (n_h - r_h)^2 k_h^2 \left[ \frac{1}{k_h} \sum_{i=1}^{k_h} (\bar{y}_{1i} - \bar{Y}_1)^2 \right] + r_h^2 (k_h + 1)^2 \left[ \frac{1}{(k_h + 1)} \sum_{i=1}^{k_h+1} (\bar{y}_{2i} - \bar{Y})^2 \right] \right\} \quad (25)$$

## 6. Analysis and Results

The four (4) sets of data below, whose features are reflected in Table 1 are used for this study.

Data used in this study are;

- i. Overall cumulative average scores of 145 students that graduated from the Faculty of Engineering University of Ilorin 1989/90 set.
- ii. Data of Kano State Ministry of Commerce and Industry Survey (2008) on manpower strength of companies and industries in the six (6) industrial Estates of Kano, Nigeria.
- iii. Grants allocation to 774 Local Government's Council in the country for the month of December, 2008 shared in January 2009.  
(See [www.fmf.gov.ng](http://www.fmf.gov.ng))
- iv. Population Census figures for the 774 Local Government Areas of Nigeria

during the year 2006 census.(see [www.nigeriastat.gov.ng](http://www.nigeriastat.gov.ng))

| S/N | N   | n   | Range         | Skewness | Mean   | Variance | Standard Deviation |
|-----|-----|-----|---------------|----------|--------|----------|--------------------|
| 1   | 145 | 48  | 44.7 - 68.8   | 0.712    | 55.48  | 20.05    | 4.48               |
| 2   | 171 | 57  | 3 – 3756      | 6.581    | 166    | 163923   | 405                |
| 3   | 774 | 258 | 72.2 - 365.0  | 3.239    | 108.96 | 700.61   | 26.47              |
| 4   | 774 | 258 | 11.7 - 1277.7 | 3.218    | 180    | 10281    | 101                |

**Table 1: Summary Statistics of the data used in this study.**

The four (4) methods of strata boundary determination highlighted in section 4 above where used to determine strata boundaries, population units placed in their respective stratum and simple random and systematic samples of fixed sample sizes 48, 57, 258 and 258 are selected for data 1 to 4 respectively in order to obtain relevant statistics for the purpose of estimating the population parameters using **R**packages (generating seed of 123). For the systematic samples, units within strata are in order of magnitude. Results obtained for data 1 to 4 using optimum and proportional allocation are as shown in Tables 2 to 9.

| No. of Strata | Stratification Methods |          |          |          |          |          |          |          |
|---------------|------------------------|----------|----------|----------|----------|----------|----------|----------|
|               | DHR                    |          | GMS      |          | EST      |          | STR      |          |
|               | SRS                    | SYS      | SRS      | SYS      | SRS      | SYS      | SRS      | SYS      |
| 2             | 0.098690               | 0.005383 | 0.089851 | 0.006235 | 0.094785 | 0.007162 | 0.106414 | 0.011471 |
| 3             | 0.048783               | 0.004871 | 0.067589 | 0.003410 | 0.056735 | 0.006482 | 0.061245 | 0.006283 |
| 4             | 0.024145               | 0.003284 | 0.034797 | 0.002092 | 0.035492 | 0.002894 | 0.030784 | 0.003929 |
| 5             | 0.015230               | 0.002483 | 0.023983 | 0.001599 | 0.017760 | 0.003554 |          |          |
| 6             |                        |          | 0.016269 | 0.001585 | 0.014280 | 0.001881 |          |          |
| 7             |                        |          |          |          | 0.012348 | 0.001932 |          |          |

**Table 2: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 1 Using Optimum Allocation.**

**On the Choice of An Efficient Sampling Scheme Within Strata**

| No. of Strata | Stratification Methods |          |          |          |          |          |          |          |
|---------------|------------------------|----------|----------|----------|----------|----------|----------|----------|
|               | DHR                    |          | GMS      |          | EST      |          | STR      |          |
|               | SRS                    | SYS      | SRS      | SYS      | SRS      | SYS      | SRS      | SYS      |
| 2             | 0.111925               | 0.009091 | 0.099633 | 0.008992 | 0.110154 | 0.008748 | 0.110305 | 0.009063 |
| 3             | 0.060491               | 0.006506 | 0.070774 | 0.006031 | 0.079164 | 0.006931 | 0.071611 | 0.006526 |
| 4             | 0.040585               | 0.002700 | 0.034715 | 0.002831 | 0.045240 | 0.006137 | 0.042008 | 0.002778 |
| 5             | 0.015952               | 0.003011 | 0.026129 | 0.001928 | 0.048436 | 0.004606 |          |          |
| 6             |                        |          | 0.016136 | 0.002009 | 0.043932 | 0.004826 |          |          |
| 7             |                        |          |          |          | 0.040995 | 0.004443 |          |          |

**Table 3: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 1 Using Proportional Allocation.**

| Number of Strata | Stratification Methods |         |        |          |        |         |       |          |
|------------------|------------------------|---------|--------|----------|--------|---------|-------|----------|
|                  | DHR                    |         | GMS    |          | EST    |         | STR   |          |
|                  | SRS                    | SYS     | SRS    | SYS      | SRS    | SYS     | SRS   | SYS      |
| 2                | 34.78                  | 2.16791 | 128.76 | 41.68460 | 104.99 | 4.35991 | 42.67 | 1.28302  |
| 3                | 20.79                  | 1.12741 | 36.51  | 4.06372  | 43.43  | 2.26057 | 41.04 | 14.22326 |
| 4                | 12.46                  | 3.54900 | 23.03  | 1.51544  | 20.45  | 1.25324 |       |          |
| 5                | 13.11                  | 0.83763 | 3.43   | 0.71966  | 20.90  | 1.80413 |       |          |
| 6                |                        |         | 5.33   | 0.92732  |        |         |       |          |
| 7                |                        |         | 3.12   | 0.47215  |        |         |       |          |
| 8                |                        |         | 4.34   | 0.46705  |        |         |       |          |

**Table 4: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 2 Using Optimum Allocation.**

|                  | Stratification Methods |        |         |        |        |        |         |        |
|------------------|------------------------|--------|---------|--------|--------|--------|---------|--------|
|                  | DHR                    |        | GMS     |        | EST    |        | STR     |        |
| Number of Strata | SRS                    | SYS    | SRS     | SYS    | SRS    | SYS    | SRS     | SYS    |
| 2                | 1495.84                | 638.56 | 1715.50 | 531.92 | 1624.6 | 442.63 | 1447.34 | 506.39 |
| 3                | 1260.81                | 481.43 | 1192.55 | 482.78 | 483.79 | 48.17  | 143.29  | 62.79  |
| 4                | 1264.85                | 360.21 | 1279.28 | 360.16 | 187.35 | 35.15  |         |        |
| 5                | 89.33                  | 365.56 | 100.20  | 370.57 |        |        |         |        |

Table 5: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 2 Using Proportional Allocation.

|                  | Stratification Methods |          |         |          |         |          |         |          |
|------------------|------------------------|----------|---------|----------|---------|----------|---------|----------|
|                  | DHR                    |          | GMS     |          | EST     |          | STR     |          |
| Number of Strata | SRS                    | SYS      | SRS     | SYS      | SRS     | SYS      | SRS     | SYS      |
| 2                | 0.57960                | 0.057222 | 0.62554 | 0.037251 | 0.46866 | 0.060540 | 0.71987 | 0.062450 |
| 3                | 0.22953                | 0.001531 | 0.36151 | 0.005036 | 0.21272 | 0.049105 | 0.20474 | 0.037198 |
| 4                | 0.11481                | 0.002408 | 0.24619 | 0.004881 | 0.09441 | 0.002400 | 0.10832 | 0.001636 |
| 5                | 0.09489                | 0.002076 | 0.20642 | 0.003907 | 0.05126 | 0.001199 | 0.07967 | 0.002184 |
| 6                | 0.08035                | 0.001423 |         |          | 0.03598 | 0.000871 | 0.06319 | 0.001270 |
| 7                |                        |          |         |          | 0.02830 | 0.000734 | 0.05336 | 0.001760 |
| 8                |                        |          |         |          | 0.02155 | 0.000771 |         |          |
| 9                |                        |          |         |          | 0.01529 | 0.000699 |         |          |
| 10               |                        |          |         |          | 0.01340 | 0.000567 |         |          |

Table 6: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 3 Using Optimum Allocation.

**On the Choice of An Efficient Sampling Scheme Within Strata**

|                  | Stratification Methods |          |         |          |         |          |         |          |
|------------------|------------------------|----------|---------|----------|---------|----------|---------|----------|
|                  | DHR                    |          | GMS     |          | EST     |          | STR     |          |
| Number of Strata | SRS                    | SYS      | SRS     | SYS      | SRS     | SYS      | SRS     | SYS      |
| 2                | 1.05755                | 0.108500 | 0.69610 | 0.086420 | 0.75910 | 0.111123 | 1.76997 | 0.113350 |
| 3                | 0.57920                | 0.095809 | 0.37780 | 0.074565 | 1.06532 | 0.082455 | 0.58863 | 0.100827 |
| 4                | 0.28540                | 0.085411 | 0.29809 | 0.051942 | 0.75310 | 0.099657 | 0.38693 | 0.075293 |
| 5                | 0.20067                | 0.076763 | 0.22977 | 0.010085 | 0.42796 | 0.095615 | 0.19153 | 0.069257 |
| 6                | 0.17339                | 0.076118 |         |          | 0.39227 | 0.093355 | 0.20044 | 0.047051 |
| 7                |                        |          |         |          | 0.34057 | 0.075736 | 0.19483 | 0.009301 |
| 8                |                        |          |         |          | 0.31786 | 0.089197 |         |          |
| 9                |                        |          |         |          | 0.27504 | 0.085933 |         |          |
| 10               |                        |          |         |          | 0.22209 | 0.084034 |         |          |

**Table 7: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 3 Using Proportional Allocation.**

|                  | Stratification Methods |         |         |         |        |         |         |         |
|------------------|------------------------|---------|---------|---------|--------|---------|---------|---------|
|                  | DHR                    |         | GMS     |         | EST    |         | STR     |         |
| Number of Strata | SRS                    | SYS     | SRS     | SYS     | SRS    | SYS     | SRS     | SYS     |
| 2                | 8.4273                 | 0.90257 | 18.2072 | 2.39406 | 9.5739 | 0.96503 | 11.0460 | 1.01598 |
| 3                | 3.0593                 | 0.06419 | 6.9140  | 0.79715 | 3.3461 | 0.52980 | 3.6232  | 0.77195 |
| 4                | 2.1646                 | 0.02676 | 6.0417  | 0.08440 | 1.7310 | 0.03112 | 1.8361  | 0.03642 |
| 5                | 1.1201                 | 0.02953 | 3.4276  | 0.11740 | 1.0244 | 0.02481 | 1.3645  | 0.03711 |
| 6                | 0.8765                 | 0.02374 | 3.3028  | 0.08610 | 0.7462 | 0.01831 | 1.2312  | 0.02660 |
| 7                |                        |         | 2.1526  | 0.03584 | 0.5288 | 0.02161 | 0.8623  | 0.01593 |
| 8                |                        |         | 1.7707  | 0.04724 | 0.4854 | 0.01527 |         |         |
| 9                |                        |         |         |         | 0.3486 | 0.01355 |         |         |
| 10               |                        |         |         |         | 0.2952 | 0.01413 |         |         |

**Table 8: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 4 Using Optimum Allocation.**

|                  | Stratification Methods |         |         |         |         |         |         |         |
|------------------|------------------------|---------|---------|---------|---------|---------|---------|---------|
|                  | DHR                    |         | GMS     |         | EST     |         | STR     |         |
| Number of Strata | SRS                    | SYS     | SRS     | SYS     | SRS     | SYS     | SRS     | SYS     |
| 2                | 11.3638                | 2.26073 | 26.3685 | 2.51967 | 19.6785 | 2.29634 | 25.3089 | 2.61715 |
| 3                | 7.6474                 | 2.13573 | 8.3237  | 1.93014 | 14.3550 | 1.95777 | 7.5338  | 2.40526 |
| 4                | 4.8101                 | 1.95892 | 7.8909  | 1.72459 | 5.7938  | 2.08423 | 5.3639  | 2.39719 |
| 5                | 2.5160                 | 1.81447 | 8.7369  | 1.46527 | 4.0064  | 1.87954 | 2.8495  | 2.28774 |
| 6                | 2.2992                 | 1.62204 | 7.9777  | 1.27294 | 3.8869  | 1.85589 | 2.7487  | 2.12626 |
| 7                |                        |         |         |         | 2.0633  | 1.86822 |         |         |
| 8                |                        |         |         |         | 7.5922  | 1.78097 |         |         |
| 9                |                        |         |         |         | 1.5761  | 1.77796 |         |         |
| 10               |                        |         |         |         | 1.1873  | 1.79319 |         |         |

Table 9: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 4 Using Proportional Allocation.

Tables 2 to 9 present the result when simple random samples and systematic samples are selected from each stratum for the four methods of stratification considered in this study. For all the methods of strata construction studied  $V(\bar{y}_{st})_{OPT} < V(\bar{y}_{st})_{PROP}$ . This is in agreement with common experience. On the choice of sampling scheme within the stratum, SYS is more precise than SRS for optimum and proportional allocation for all the datasets used. Thus,  $(V(\bar{y}_{st})_{SYS} < V(\bar{y}_{st})_{SRS}$ .

### Conclusion

This study has empirically confirmed that on the choice of sampling scheme within the Strata when stratified sampling is in use, systematic samples provides more precise estimates than the commonly used simple random sample within the stratum when the sampling units in each stratum are in order of magnitude. It further confirms that whether  $N = nk$  or otherwise, estimation is possible using the appropriate formulae recommended in respective cases in this study. Thus the use of systematic sampling is recommended for use when most precise estimate is required.

### REFERENCE

1. Chang, H. J. and Huang, K. C. "Reminder Linear Systematic Sampling" Sankya: The Indian Journal of Statistics 62, series B, (2000), pp. 249 - 256.

# **On the Choice of An Efficient Sampling Scheme Within Strata**

2. Cochran, W.G. "Comparison of Method for Determining Stratum Boundaries". Bulletin of the International Statistical Institute. 38, No. 2, (1961), pp.345-358.
3. Cochran, W.G. Sampling Techniques, Third edition: John Wiley and Sons, New York. (1977).
4. Dalenius, T. "The Problem of Optimum Stratification", SkandinaviskAkturietidskrift, 33, (1950). pp.203-213.
5. Dalenius, T. and Hodges, J.L., Jr. "Minimum Variance Stratification" JASA, 54, (1959), pp. 88 – 101.
6. Durbin, J., "Review of sampling in Sweden". Journal of Royal Statistical Societies. A. 122, (1959), pp. 246-248.
7. Ekman, G. (1959). "An Approximation Useful in Univariate Stratification". Annals of Mathematical Statistics, 30, pp. 210-229.
8. Ghosh, S. P. "Optimum Stratification with Two Characters". Annals of Mathematical Statistics, 34, (1963) , pp. 866-872.
9. Gunning P, and Horgan J. M "A New algorithm for the construction of stratum boundaries in skewed population" Survey Methodology, 30, No. 2, (2004), pp.159-166.
10. Gunning P, Horgan J. M. and Keogh G. "Efficient Pareto Stratification". Mathematical Proceedings of Royal Irish Academy 106A (2), (2006), PP. 131-138.
11. Gunning P, Horgan, J. M. and Yancey W. "Geometric Stratification of Accounting data". J.de contaduria Y. Administration, 214, Septiembre – Diciembre. (2004).
12. Hedlin D. "A procedure for Stratification by an Extended Ekman rule". Journal of Official Statistics 6, No. 1, (2000) , pp.15-29.
13. Hess I, Sethi V.K. and Balakrishnan, T.R. "Stratification: A practical investigation" JASA 61, (1966). pp. 74-90.
14. Hidirolou, M. A. "The Construction of self-representing stratum of large units of Survey Design" 40, No.1, (1986). pp. 27-31.
15. Horgan J. M. "Stratification of Skewed Populations: A review". International Statistical Review, 74, 1, (2006). pp.67-76.
16. Judez, L. and Chaya, C. "Effects of Geographical Stratification in a Farm Accountancy Data Network on the Accuracy of the Estimates. Journal of Agricultural Economics, 50, Number 3, (1999), pp. 388-399.
17. Judez, L. and Chaya, C. "Comparative Analysis of Alternative Sampling plans to create a farm Accountancy Data Network for the Agricultural Section of Navarra". Questio, 21, (2000) , pp. 137-150.

18. Judez L, Chaya C, Miguel T. M and Bru R Stratification & sample size of data sources for the agricultural mathematical Programming models. Report of the commissioned study by the EU and Spanish Government. (2005).
19. Keskinturk T. and Er S. "A Genetic algorithm approach to determine boundaries and Sample size of each stratum in stratified Sampling". South Pacific Journal of Natural Sciences, B., 21, (2007) , pp 91-95.
20. Kharn, M. G. M and Ahsan, M. J, A note on optimum allocation in Multivariate Sampling. South Pacific Journal of Natural Science B., 21 (2003) , pp 91-95
21. Kish, L. and Anderson, A. W. "Multivariate and Multipurpose Stratification".JASA 73, No. 361. (1978). pp. 24-34.
22. Kozak, M. "Optimal Stratification Using Random Search Method in Agricultural Surveys".Statistics in Transition, 6, No. 5, (2004) , pp.797-806.
23. Lahiri, D. B. (1952). "NSE instruction to field workers" See Murthy (1967, pp.140).
24. Lavalle, P. and Hidirolou, M.A. "On the Stratification of Skewed Populations".Survey Methodology 14, 1, (1988) , pp.33-43.
25. Leu, C. H and Tsui, K. W. "New Partially Systematic Sampling".Statisticasinica 6, (1996), pp. 617-630.
26. Mahalanobis, P.C. "Some Aspects of the Design of Sample Surveys".Sankhya, 12, (1952). pp 1-7.
27. Mendehall, W, Ott, L. and Scheafffer, R. L.; Elementary Survey Sampling; Duxbary Press, Belmont, California. (1971).
28. Murthy, M. N. Sampling Theory and methods 2<sup>nd</sup> Edition Statistical Publishing Society, Calcutta - 35, India. (1967)
29. Neyman, J. "On the Two different aspects of the representative method: The method of Stratified sampling and the method of purposive selection". Journal of Royal Statistical Society, 97, (1934). pp.558-606.
30. Okafor, F. C. Sample Survey Theory with Applications Afro-Orbis Publications Ltd. Nsukka, Nigeria. (2002).
31. Raj, D. and Chandhok, P. Sample Survey Theory; Narosa publishing House, 6, Community Centre, Panchsheel Park, New Delhi 100 017. (1998).
32. Sethi, V. K. "A Note on Optimum Stratification for Estimating the Population Means". The Australian Journal of Statistics, 5, (1963). pp. 20-33.
33. Sukhatme P.V., Sukhatme B.V. and Asok, C. Sampling Theory with Applications.3rd Edition, Iowa University Press, USA. (1984).
34. Wang, W. C. and Aggarwal, V. "Stratification under a particular pareto distribution".Commun.Statist. - Theory. Meth.13, (6), (1984). pp.711-35.