A Note on the Precision of Stratified Systematic Sampling

Akeem O. Kareem[1], Isaac O. Oshungade[2], Gafar M. Oyeyemi[2]

[1]Institute for Security Studies, Abuja, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria
Email: keemkareem@yahoo.com, layiosungade@yahoo.com, matanmi@unilorin.edu.ng

Abstract

Conflicting views had greeted the use of systematic sampling for sample selection and estimation in stratified sampling in terms of the precision of the population mean base on the inherent characteristics of the population. These conflicting views were analyzed using Cochran data (1977, p. 211) [1]. When the population units are ordered, variance of systematic sampling for all possible systematic samples provides equal, non-negative and most precise estimates for all the variance functions considered *i.e.* $V_1(\bar{y}_{sy}) = V_2(\bar{y}_{sy}) = V_3(\bar{y}_{sy})$, unlike when a single systematic sample is used and when variance of simple random sampling is used to estimate selected systematic samples.

Keywords

Precision, Systematic Sampling, Stratified Systematic Sampling, Systematic Random Estimator

1. Introduction

Cochran (1977) [1] 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^{th} unit thereafter. The selection of the first k^{th} units determined the whole sample. This is called an every k^{th} systematic sample.

Murthy (1967) [2] 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. He describes technique of systematic sampling as consisting of selecting every k^{th} unit starting with the unit corresponding to a number r chosen at random from 1 to k, where k is taken as the integer nearest to N/n. The random number r chosen from 1 to k is known as random start and the constant k is termed the sampling interval.

A sample selected by this procedure is termed a systematic sample with a random start *r*. Therefore, the value of *r* determines the whole sample. In other words, this procedure amounts to selecting with equal probability one of the *k* possible groups of units (samples) into which the population can be divided in a systematic manner.

Same view was expressed by Raj and Chandhok (1998) [3]. They described systematic sampling as a more convenient method of sample selection when the units were serially numbered from 1 to *N* with the assumption that $N = nk$, where *n* is the sample size desired, and *k* is an integer. A number is taken at random from the numbers 1 to *k* (using a table of random number/random number generator). Suppose the random number is *i*, then the sample contains *n* units with serial numbers $i, i+k, i+2k, \cdots, i+(n-1)k$. Thus, the sample consists of the first unit selected at random and every k^{th} unit thereafter. It is therefore called a 1-in-ksystematic sample.

Early studies on the development of theory of systematic sampling was as reported by Murthy (1967, p.134) [2] while Cochran (1977) [1] reported that Madow (1953) [4] had carried systematic sampling to its logical conclusion with his recommendation that a systematic sample be chosen at or near the center of the interval, *i.e.* instead of starting the sequence by a random number chosen between 1 and *k*, we take the starting number as $(k+1)/2$ if *k* is odd and either $k/2$ or $(k+1)/2$ if *k* is even.

Guatschi (1957) [5] investigated the efficiency of single and multiple random start systematic sampling in population exhibiting different characteristics and reported that when the population was in random order $V(\bar{y}_{sy})_S = V(\bar{y}_{sy})_M$ and $V(\bar{y}_{sy})_S < V(\bar{y}_{sy})_M$ for a population with linear trend, while in a periodic population $V(\bar{y}_{sy})_S \leq V(\bar{y}_{sy})_M$ equality results when $M = S$. He, however, concluded that, with an exponential correlelogram, single random start (S) was more precise than multiple random starts (M).

Murthy (1967) [2], Cochran (1977) [1], Raj and Chandhok (1998) [3] and Okafor (2002) [6] 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 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 $1/k$.

Thus for a population of *Y* units $Y_1, Y_2, \cdots, Y_N$ divided into *k* possible clusters, the *k* possible samples with their means are as shown in **Table 1** below.

Considering all the *k* possible samples, the sample mean $\bar{y}_{sy}$ is obtained thus:

$$E(\bar{y}_{sy}) = \frac{1}{k}\sum_{i=1}^{K} \bar{y}_i = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{1}{n}\sum_{j=1}^{n} y_{ij}\right)$$
$$= \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} \tag{1}$$
$$= \frac{1}{N}(Y_1 + Y_2 + \cdots + Y_N) = \bar{Y},$$

Showing that when $N = nk$, $\bar{y}_{sy}$ is unbiased for $\bar{Y}$. It should also be noted that systematic sampling has no repetition of sampling unit and therefore related to simple random sampling without replacement (SRSWOR).

Table 1. Compositions of systematic samples of *k* clusters (such that $N = nk$).

	Random Start (Sample Number)				
	1	2	3	*i*	*k*
	y_1	y_2	y_3	y_i	y_k
	y_{k+1}	y_{k+2}	y_{k+3}	y_{k+i}	y_{2k}
	$y_{(n-1)k+1}$	$y_{(n-1)k+2}$	$y_{(n-1)k+3}$	$y_{(n-1)k+i}$	y_{nk}
Means	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_i$	$\bar{y}_k$

Above is the applicable systematic sampling in a situation in which $N = nk$. In practice, it is common to encounter situations in which $N \neq nk$, and various suggestions have been made on how to handle such a situation.

2. Approaches When $N \neq nk$

2.1. Circular Systematic Sampling (CSS)

Lahiri (1952) [7] suggests the Circular Systematic Sampling (CSS) which consists of taking a random number from 1 to N and selecting the unit corresponding to this random start and every k^{th} unit thereafter in a cyclical manner until a sample of n units is obtained, k being the nearest integer to N/n, *i.e.* If r is a random number selected from 1 to N, the sample consists of the units corresponding to the number.

$$\text{Sample} \begin{cases} r + jk, & \text{if } r + jk \leq N; \\ r + jk - N, & \text{if } r + jk > N. \end{cases} \tag{2}$$

for $j = 0, 1, 2, \cdots, (n-1)$

It implies from CSS, therefore, that the usual procedure of selecting a random start r from 1 to k and including in the sample the units corresponding to $r + jk \leq N$ for $j = 0, 1, 2, \cdots, (n-1)$ reflected above may be termed Linear Systematic Sampling (LSS).

2.2. Murthy's Approach

Murthy (1967) [2] suggested that when $N \neq nk$, *i.e.*, the population units N cannot be divided into k clusters of equal size, therefore we choose the interval k to be the nearest integer to N/n resulting in n' which may not necessarily be equal to n, the required sample size. He stated further that if $N \neq nk$, if q and r' were the quotient and remainder obtained respectively on dividing N/n, then, N can be written as $N = nq + r'$ and the sampling interval k can be taking as:

$$k = q \ \text{ if } \ r' \leq n/2 \ \text{ or } \ k = q+1 \ \text{ if } \ r' > n/2$$

Then, the units' n' that can be expected in the sample would be given by:

$$n' = N/q' \ \text{ or } \ \left[N/q'\right]+1 \ \text{ where } \ k = q'(= q \text{ or } q+1)$$

This approach is suitable in situations in which the sample size n is not fixed or predetermined and the sampler is free to adjust the sample to suit the above application. Therefore, Murthy's approach to handle $N \neq nk$ is not suitable for fixed sample size or when stratum sample sizes are determined using the standard procedures for allocating samples into the strata.

2.3. Fractional Interval Approach

Another approach when $N \neq nk$ is the use of fractional interval reported by Murthy (1967) [2]. This approach called for taking $k = N/n$ as k without rounding it off to the nearest integer, *i.e.*, the ith unit is selected in the sample if $i-1 < r + jk \leq i$ for any $j = 0, 1, 2, \cdots, (n-1)$. It is equivalent to associating different numbers with each unit such that the first gets the number 1 to n, the second gets from $n+1$ to $2n$ and so on and thus selecting units corresponding to a LSS sample of n numbers selected from 1 to Nn with N as the sampling interval. This approach involves a long process of iteration to satisfy the equation; hence it wastes time.

2.4. New Partially Systematic Sampling (NPSS)

Leu and Tsui (1996) [8] developed the New Partially Systematic Sampling (NPSS) in order to derive an unbiased estimator of the variance of systematic sampling $V(\bar{y}_{sy})$. The population size N need not be a multiple of sample size n; therefore, it is a suitable procedure when $N \neq nk$. The procedure entails selection of SRS of size a and the remaining sample of size $(n-a)$ systematically, these samples are combined to derive an unbiased estimate of $V(\bar{y}_{sy})$. Thus, NPSS combines SRS with systematic sample to obtain its estimates thereby deviating from the objective of this study as we intend to observe performances when systematic sampling is employed as a choice scheme within strata and not when SRS is combined with systematic samples.

2.5. Remainder Linear Systematic Sampling (RLSS)

Also reviewed in this section is Remainder Linear Systematic Sampling (RLSS) due to Chang and Huang (2000) [9]. This procedure is a modification of the LSS. It is developed for situation when $N \neq nk$, and depends only on the remainder. It involves dividing the population into two strata, the sampling interval k is taken as the nearest integer to N/n such that $N = nk + r$, where r is the remaining population units, where $0 \leq r < n$; N, n, k, and rare integers. When the remainder r is zero, $r = 0$ the procedure reduces to LSS. Procedures for the RLSS are:

a) Divide the population units into two strata with the first stratum containing the front $(n-r)k$ units and second stratum housing the remaining $r(k+1)$ units. From stratum I, a random start k_1 is selected from the interval 1 to k and every k^{th} units thereafter, from the $(n-r)$ group of units forming stratum I. Thus samples from stratum I contained in a sample space S' are:

$$S' = \{kl + k_1, \ l = 0, 1, 2, \cdots, (n-r)-1\};$$

b) From stratum II, random start k_2 is taken from interval 1 to $k+1$, starting with $\left[(n-r)k+k_2\right]$ units and every $(k+1)$ units thereafter from the r group forming the second stratum. Samples from stratum II are contained in the sample space S'' are:

$$S'' = \{(k+1)l' + (n-r)k + k_2, \ l' = 0, 1, 2, \cdots, (r-1)\}$$

Sample of size n is the combination of S' and S'' units.

Therefore, in stratified systematic sampling when $N_h \neq n_h k_h$, competing methods are: CSS, NPSS, and RLSS. Due to its greater efficiency over the CSS and NPSS as reported by Chang and Huang (2000) [9], RLSS was used by Kareem *et al* (2015) [10] in stratum where $N_h \neq n_h k_h$. The mean and variance of RLSS is as given below (see relation 2.2 and 2.3, p. 251 of Chang and Huang (2000) [9]).

3. Estimation Procedures in Systematic Sampling

Estimation of the population mean of a systematic sample over all possible samples is as given by relation (1). For the variance of the population mean, Murthy (1967) [2], while assuming $N = nk$ for a sample of size n and k sampling interval, states that there are k possible samples and $\bar{y}_i$ be the sample mean of ith possible sample $(i = 1, 2, \cdots, k)$. The sampling variance of the systematic sample is given as:

$$V_1\left(\bar{y}_{sy}\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\bar{y}_i - \bar{Y}\right)^2 \tag{3}$$

where $\bar{y}_i = \frac{1}{n}\sum_{i=1}^{k} y_{ij}$

Equivalent to

$$V_1\left(\bar{y}_{sy}\right) = \frac{1}{n^2 k}\left[n\bar{y}_{i.} - n\bar{Y}\right]^2$$

It is simplified as

$$V_1\left(\bar{y}_{sy}\right) = \frac{1}{n^2 k}\left[\sum\left(y_{i.}\right)^2 - \frac{\sum\left(y_{ij}\right)^2}{k}\right], \quad i = 1, 2, \cdots, k$$

where $\sum y_{i.}$ is the sum of systematic sample in the ith group, y_{ij} is the jth variate of the ith systematic sample.

Note that $\sigma^2 = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij} - \bar{Y}\right)^2$ is the population variance of SRS and can be written as the sum of σ_b^2 and σ_w^2, which are the between and the within sample variances, respectively.

Therefore $V(\bar{y}_{sy})$ which is σ_b^2 can be expressed as

$$\sigma_b^2 = V(\bar{y}_{sy}) = \sigma^2 - \sigma_w^2 \tag{4}$$

where $\sigma_w^2 = \dfrac{1}{k}\sum_{i=1}^{k}\sigma_{wi}^2$, $\sigma_{wi}^2 = \dfrac{1}{n}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2$.

Other expressions for the estimation of variance of the mean of systematic samples by various authors are reported by Murthy (1967, Section 5.8, pp. 153-155) [2] and Cochran (1977, pp. 213-226) [1]. Cochran, however, remarked "that no dearth of formulae for the estimated variance, but all appeared to have limited applicability".

On the efficiency of systematic sampling in relation to other sampling scheme, literature agreed that efficiency of systematic sampling was strongly anchored on the arrangement of the population units. Cochran (1977) [1] stated that it greatly depended on the properties of the population. For some population, systematic sampling is extremely precise and for others, SRS is more precise than systematic sampling, not even with increase in sample size n. Thus, it is difficult to give general advice about the situation in which systematic sampling is to be recommended. However, the knowledge of the population structure is necessary for its most effective use.

Same view was expressed by Murthy (1967) [2], that a good arrangement of the population units may yield a better estimate while a bad arrangement may lead to inefficient estimate and therefore, warned that one had to be careful with the use of systematic sampling and to ensure first, that the existing arrangement did not lead to inefficient estimates before using it. One way suggested is to ensure that the units are arranged either in increasing or decreasing order and this directly suits our investigation in this study, since application of methods of strata construction requires that the population units be arranged in order of magnitude to avoid overlapping of units.

Cochran (1977) [1] stated that several formulae had been developed for $V(\bar{y}_{sy})$. Three of such formulae given by Cochran under the assumption that $N = nk$ and could be applied to any kind of cluster sampling in which the clusters contain n elements, and the sample consists of one cluster, are stated below.

1) The variances of the mean of systematic sample given by Cochran are:

$$V_2(\bar{y}_{sy}) = \frac{(N-1)S^2}{N} - \frac{k(n-1)S_{wsy}^2}{N} \tag{5}$$

where

$$S_{wsy}^2 = \frac{1}{k(n-1)}\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 \tag{6}$$

This can further be expressed as

$$\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{(y_{ij} - \bar{y}_i)^2}{n-1} = \frac{1}{k}\sum_{i=1}^{k}S_i^2$$

which is the weighted variance over all possible systematic samples generated by random start $(i = 1, 2, \cdots, k)$. It implies therefore from relation (4)

$$\sigma_b^2 = V(\bar{y}_{sy}) = \sigma^2 - \sigma^2 \tag{4}$$

$(N-1)S^2 = \sigma^2$ while $\dfrac{(N-1)S^2}{N} = \dfrac{\sigma^2}{N} = V(\bar{y})$

$\therefore V(\bar{y}_{sy}) = V(\bar{y}) - \sigma_w^2$ hence, relation (5) above.

2) The second one is given as

$$V_3(\bar{y}_{sy}) = \frac{S^2}{n}\left(\frac{(N-1)}{N}\right)\left[1 + (n-1)\rho_w\right] \tag{7}$$

where ρ_w is the correlation coefficient between pairs of units that are in the same systematic sample, other references referred to it as intra-class correlation coefficient and denoted by ρ_c

$$\rho_w \text{ or } \rho_c = \frac{E\left(y_{ij} - \overline{Y}\right)\left(y_{iu} - \overline{Y}\right)}{E\left(y_{ij} - \overline{Y}\right)^2}$$

where the numerator is averaged overall $kn(n-1)/2$ distinct pairs, and the denominator over all N values of y_{ij}. Since the denominator is $(N-1)S^2/N$, this gives

$$\rho_w = \frac{2}{(n-1)(N-1)S^2} \sum_{i=1}^{K} \sum_{j<u} \left(y_{ij} - \overline{Y}\right)\left(y_{iu} - \overline{Y}\right) \tag{8}$$

The two expressions of $V\left(\overline{y}_{sy}\right)$ above are expressed in terms of S^2, hence it relates to the variance of SRS.

3) The third is expressed in terms of variance of stratified random sample in which the strata are composed of the first k units, the second k units and so on.

The subscript j in y_{ij} denotes the stratum and the stratum mean is written as $\overline{y}_{.j}$.

$$V_4\left(\overline{y}_{sy}\right) = \frac{S_{wst}^2}{n}\left(\frac{N-n}{N}\right)\left[1+(n-1)\rho_{wst}\right] \tag{9}$$

where $S_{wst}^2 = \frac{1}{n(k-1)} \sum_{j=1}^{n} \sum_{i=1}^{k} \left(y_{ij} - \overline{y}_{.j}\right)^2$

This is the variance among units that lie in the same stratum. The divisor $n(k-1)$ is used because each of the n strata contributes $(k-1)$ degrees of freedom and

$$\rho_{wst} = \frac{E\left(y_{ij} - \overline{y}_{.j}\right)\left(y_{iu} - \overline{y}_{.u}\right)}{E\left(y_{ij} - \overline{y}_{.j}\right)^2}$$

This quantity is the correlation between the deviations from the stratum means of pairs of units that are in the same systematic sample.

$$\rho_{wst} = \frac{2}{n(n-1)(k-1)} \sum_{i=1}^{K} \sum_{j<u} \frac{\left(y_{ij} - \overline{y}_{.j}\right)\left(y_{iu} - \overline{y}_{.u}\right)}{S_{wst}^2} \tag{10}$$

It implies therefore from relation (9) above that a systematic sample has the same precision as that of a stratified random sampling sample with one unit per stratum if $\rho_{wst} = 0$, thus relation (9) reduces to

$$V\left(\overline{y}_{st}\right) = \frac{S_{wst}^2}{n}\left(\frac{N-n}{N}\right) \tag{11}$$

Thus, we have examined systematic sampling in terms of procedure and estimation process. But our concern is taking a systematic sample of fixed sample size n from each stratum for estimation purpose.

3.1. Estimation in Stratified Systematic Sampling

Much have been said in Section 2 on the significance of the arrangement of the population units on the precision $V\left(\overline{y}_{sy}\right)$, while Cochran (1977, p. 208) [1] 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 as defined by relation (6) above and S^2 is the variance of the population mean.

Notations

Cochran (1977, p. 91) [1] has stated that expressions for the mean and variance of stratified sampling applied generally to all classes of stratified sampling and are not restricted to stratified random sampling. Therefore, all notations in Cochran (1977, p. 90) [1] 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.

$$\overline{y}_{syh} = \frac{1}{k_h}\sum_{i=1}^{k_h}\left(\sum_{i=1}^{n_h}\frac{y_{ij}}{n_h}\right) \tag{12}$$

is the mean of systematic sample in stratum h, equivalent to relation (1).

$$\overline{y}_{stsy} = \sum W_h \overline{y}_{syh} \tag{13}$$

is the population mean of the stratified systematic sample.

$$V\left(\overline{y}_{syh}\right) = \frac{(N_h-1)S_h^2}{N_h} - \frac{k_h(n_h-1)S_{wsyh}^2}{N_h} \tag{14}$$

is the variance of stratified systematic samples in stratum h when $N_h = n_h k_h$.

Therefore, $V_2\left(\overline{y}_{sy}\right)$ variance of systematic samples given by Cochran in relation (5) above when $N_h = n_h k_h$, is adopted for our sample estimation and modified for the stratified systematic samples as shown in relation (14) above. However, it should be noted that each of $V_1\left(\overline{y}_{sy}\right)$, $V_2\left(\overline{y}_{sy}\right)$, and $V_3\left(\overline{y}_{sy}\right)$ would yield the same estimates when $N_h = n_h k_h$.

$$V\left(\overline{y}_{stsy}\right) = \sum W_h^2 V\left(\overline{y}_{syh}\right) \tag{15}$$

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

$$MSE\left(\overline{y}_{stsy}\right) = V\left(\overline{y}_{stsy}\right) + \left[\sum(w_h - W_h)\overline{Y}_h\right]^2 \tag{16}$$

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

The mean and the variance of RLSS are given below (see relation 2.2 and 2.3, p. 251 of Chang and Huang (2000) [9]).

$$\overline{y}_{sy(\text{RLSS})} = \frac{(n-r)k*\overline{y}_{1i} + r(k+1)*\overline{y}_{2i}}{N} \tag{17}$$

$$V\left(\overline{y}_{sy(\text{RLSS})}\right) = \frac{1}{N^2}\left\{(n-r)^2 k^2\left[\frac{1}{k}\sum_{i=1}^{k}\left(\overline{y}_{1i}-\overline{Y}_1\right)^2\right] + r^2(k+1)^2\left[\frac{1}{(k+1)}\sum_{i=1}^{k+1}\left(\overline{y}_{2i}-\overline{Y}\right)^2\right]\right\} \tag{18}$$

To suite our applications, expression (17) and (18) are modified as follows:

$$\overline{y}_{syh(\text{RLSS})} = \frac{(n_h-r_h)k_h*\overline{y}_{1i} + r_h(k_h+1)*\overline{y}_{2i}}{N_h} \tag{19}$$

$$V\left(\overline{y}_{syh(\text{RLSS})}\right) = \frac{1}{N_h^2}\left\{(n_h-r_h)^2 k_h^2\left[\frac{1}{k_h}\sum_{i=1}^{k_h}\left(\overline{y}_{1i}-\overline{Y}_1\right)^2\right] + r_h^2(k_h+1)^2\left[\frac{1}{(k_h+1)}\sum_{i=1}^{k_h+1}\left(\overline{y}_{2i}-\overline{Y}\right)^2\right]\right\} \tag{20}$$

It should be noted that that expressions $\frac{1}{k_h}\sum_{i=1}^{k_h}\left(\overline{y}_{1i}-\overline{Y}_1\right)^2$ and $\frac{1}{(k_h+1)}\sum_{i=1}^{k_h+1}\left(\overline{y}_{2i}-\overline{Y}\right)^2$ in relation (20) are

equivalent to relation (3) above, *i.e.* $V_1\left(\overline{y}_{sy}\right)$.

3.2. Empirical Investigation

Systematic samples are easy to draw and to execute but may not be simple in term of estimation as there are competing estimators. This drew our attention for an empirical investigation to ensure the right choice of estimator in the face of conflicting reports. Murthy (1967, section 5.8, p. 153) [2] stated that "it is not possible to estimate unbiasedly the variance of the population mean and total on the basis of a single sample, but it is possible to build up some biased but useful variance estimators on the basis of systematic samples". Same view was expressed by Mendenhall *et al.* (1971, p. 151) [11] that "an unbiased estimate of $V\left(\overline{y}_{sy}\right)$ cannot be obtained us-

ing data from only one systematic sample and that for random population, systematic sampling is equivalent to SRS", *i.e.* $V(\bar{y}_{sy})$ is approximately equal to $V(\bar{y})$. This is referred to as conservative estimator of $V(\bar{y}_{sy})$ by them, but referred to as Systematic Random Estimator (SRE) $V(\bar{y}_{SRE})$ in this study, *i.e.* when $V(\bar{y})$ is used to estimate $V(\bar{y}_{sy})$.

Raj and Chandhok (1998) [3] stated that "when units are deliberately ordered, the formula for estimating variance of SRS will not apply to systematic sampling". However, Cochran (1977, p. 227) [1] stated that if there were many strata, one systematic sample can be used in most of them.

In view of the above, the question is: should a single systematic sample be used to estimate $V(\bar{y}_{sy})$ or all possible systematic samples? To reach a conclusion, we explore Table 8.3, p. 211 of Cochran (1977) [1]. When all possible systematic samples are considered $V_1(\bar{y}_{sy}) = V_2(\bar{y}_{sy}) = V_3(\bar{y}_{sy}) = 11.6256$, $V(\bar{y}_{SRE}) = 30.66$, $V(\bar{y}_{st}) = 3.04$ as obtained by Cochran.

Empirical investigation reveals that when we select a single systematic sample, the result is as shown in **Table 2** below.

Since the efficiency of systematic sampling depends on the arrangement of the population units, an attempt is also made to rearrange Cochran (1977) [1] data (Table 8.3, p. 211), in order of magnitude; same sample of size $n = 4$ was taking, $N = 40$ and $k = 10$. With this arrangement, for all possible systematic samples our estimates are:

$$V_1(\bar{y}_{sy}) = V_2(\bar{y}_{sy}) = V_3(\bar{y}_{sy}) = 9.513125$$

while $V(\bar{y}_{SRE}) = 30.696396$ and $V(\bar{y}_{st}) = 2.4481$.

Table 3 below gives the estimates for single systematic samples when the units are arranged in order of magnitude before sample selection.

4. Conclusion

This analysis brings to the lime light the caution by Murthy (1967, p. 145) [2] in the application of systematic sampling that "one has to be careful in using systematic sampling and should at least ensure that the existing arrangement do not lead to inefficient estimates". From the empirical investigation, it could be observed that when population units are arranged in order of magnitude, a more precise estimate is obtained for $V(\bar{y}_{Sy})$ and $V(\bar{y}_{st})$ when compared with the use of SRS estimator $V(\bar{y}_{SRE})$. It also reveals that, even when units are not in order of magnitude, it may be more precise than that of SRS, *i.e.* $V_2(\bar{y}_{Sy}) < V(\bar{y}_{SRE})$ as shown in **Table 2**. Furthermore, while $V_1(\bar{y}_{sy}) = V_2(\bar{y}_{sy}) = V_3(\bar{y}_{sy})$ for all possible systematic samples, this is not true for single systematic samples as $V_1(\bar{y}_{sy}) \neq V_2(\bar{y}_{sy}) \neq V_3(\bar{y}_{sy})$ and in some instances reporting negative variances as

Table 2. Variance of single systematic samples using Cochran's data.

Groups	$V_1(\bar{y}_{sy})$	$V_2(\bar{y}_{sy})$	$V_3(\bar{y}_{sy})$	$V(\bar{y}_{SRE})$
g_1	−314.7056	30.0944	33.0578	30.825
g_2	−309.3056	11.5944	31.1928	36.375
g_3	−307.0744	4.6569	30.6481	38.4563
g_4	−305.5244	14.6569	30.6975	35.4563
*g_5	−295.1744	19.6569	30.4062	33.9563
g_6	−298.8244	19.6569	30.3950	33.9563
g_7	−288.3056	22.0944	30.8478	33.225
g_8	−281.9306	2.3444	31.0459	39.15
g_9	−278.5744	−0.8431	31.4389	40.1063
g_{10}	−281.9306	−7.6556	30.7959	42.15

*In **Table 2**, g_5 indicates the center for systematic sample estimates when Madow's procedure is used while the subscript $i = 1, \cdots, k = 10$ is the random start in the interval 1 to 10.

Table 3. Variance of single systematic samples using Cochran's data when sampling units are arranged in order of magnitude.

Groups	$V_1\left(\overline{y}_{sy}\right)$	$V_2\left(\overline{y}_{sy}\right)$	$V_3\left(\overline{y}_{sy}\right)$	$V\left(\overline{y}_{SRE}\right)$
g_1	−311.4244	14.6569	31.7250	35.4563
g_2	−309.3056	11.5944	31.1928	36.375
g_3	−307.0744	4.6569	30.6481	38.4563
g_4	−305.5244	14.6569	30.6975	35.4563
*g_5	−298.8244	19.6569	30.3950	33.9563
g_6	−295.1744	19.6569	30.4062	33.9563
g_7	−289.3244	15.1569	30.5918	35.3063
g_8	−285.1744	0.1569	30.6031	39.8063
g_9	−281.9306	−2.6556	30.9209	40.65
g_{10}	−279.7056	−2.4056	31.2328	40.575

*In Table 3, g_5 indicates the center for systematic sample estimates when Madow's procedure is used while the subscript $i = 1, \cdots, k = 10$ is the random start in the interval 1 to 10.

shown in **Table 2** and **Table 3** above. Therefore, when systematic sampling is the choice design within strata, estimates for all possible systematic samples should be used and the sampling units arranged in order of magnitude within the stratum. Kareem *et al.* (2015) [10] used this procedure and reported higher efficiency of systematic sampling within stratum over the popularly used SRS. It is hereby recommended that $V_2\left(\overline{y}_{sy}\right)$ given by Cochran (1977) [1] should be used for estimation purpose when $N_h = n_h k_h$ and that of Chang and Huang (2000) [9] $V\left(\overline{y}_{syh(\text{RLSS})}\right)$ when $N_h \neq n_h k_h$ when systematic sampling is employed within strata.

References

[1] Cochran, W.G. (1977) Sampling Techniques. 3rd Edition, John Wiley and Sons, New York.

[2] Murthy, M.N. (1967) Sampling Theory and Methods. 2nd Edition, Statistical Publishing Society, Calcutta.

[3] Raj, D. and Chandhok, P. (1998) Sample Survey Theory. Narosa Publishing House, New Delhi.

[4] Madow, W.G. (1953) On the Theory of Systematic Sampling III. *Annals of Mathematical Statistics*, **24**, 101-106.
http://dx.doi.org/10.1214/aoms/1177729087

[5] Gautschi, W. (1957) Some Remarks on Systematic Sampling. *Annals of Mathematical Statistics*, **28**, 385-394.
http://dx.doi.org/10.1214/aoms/1177706966

[6] Okafor, F.C. (2002) Sample Survey Theory with Applications. Afro-Orbis Publications Ltd., Nsukka.

[7] Lahiri, D.B. (1952) NSE Instruction to Field Workers. See Murthy (1967, p. 140).

[8] Leu, C.-H. and Tsui, K.-W. (1996) New Partially Systematic Sampling. *Statistica Sinica*, **6**, 617-630.

[9] Chang, H.J. and Huang, K.C. (2000) Reminder Linear Systematic Sampling Sankya. *The Indian Journal of Statistics*, **62** (**Series B**), 249-256.

[10] Kareem, A.O, Oyeyemi, G.M. and Adewara, A.A (2015) On the Choice of an Efficient Sampling Scheme within Strata ICASTOR. *Indian Journal of Mathematical Science*, **9**. (In Press)

[11] Mendehall, W., Ott, L. and Scheafffer, R.L. (1971) Elementary Survey Sampling. Duxbary Press, Belmont.