Indian Journal of Economics and Business
Vol. 20 No. 2 July-December, 2021)

http://www.ashwinanokha.com/IJEB.php

An Improved Diagonal-Cum-Linear Systematic Sampling Scheme

Muhammad Azeem[1] **& Zahid Khan**[2]

[1*]*Assistant Professor,Department of Statistics, University of Malakand, Chakdara,Dir (Lower), Pakistan.*

[2] *Lecturer, Department of Statistics, University of Malakand, Chakdara, Dir (Lower), Pakistan.*

Corresponding author Email: azeemstats@uom.edu.pk

Received: 15th March 2021
Revised: 11th August 2021
Accepted: 15th September 2021

Abstract: In survey sampling, one of the most popular methods of sample selection is the systematic sampling design widely used by survey researchers to obtain a sample from a population of interest. Azeem et al. (2021) developed a modified form of the diagonal systematic sampling design which was shown to be more efficient than some of commonly used systematic sampling methods. Following Azeem et al. (2021), this paper introduces a new modified version of systematic sampling design to get even more improvement in terms of efficiency. It is found that the suggested sampling design is more efficient than Azeem et al. (2021) and some other commonly used sampling schemes in situations where the units of a finite population exhibit a linear trend. Further, the mathematical conditions for the new sampling scheme to be more efficient than some of the available sampling designs are strong and always hold, making the proposed sampling scheme superior over the existing sampling schemes.

Keywords: survey sampling; systematic random sampling; diagonal systematic sampling; linear trend; efficiency comparison.

1. Introduction

In recent decades, systematic sampling has gotten the attention of survey statisticians due to its simplicity of use. Systematic sampling is even simpler than simple random sampling as only the first unit (or the first few units) are selected randomly from the population. The remaining units are obtained according to a pre-defined rule. First introduced by Madow and Madow (1944), many versions of systematic sampling have been developed by the researchers for use with different real-life situations. Madow and Madow (1944) introduced the novel idea of selecting the units from the population according to a pre-defined pattern, called systematic sampling. The Madow and Madow (1944) method was only applicable to those circumstances where the size of the finite population is a constant multiple of the required sample size, thus limiting its usability. To overcome this drawback, Lahiri (1951) introduced a new method called circular systematic sampling design.Later, Chang and Huang (2000)introduced anew modification of systematic random sampling which they called remainder systematic sampling which is also applicable in situations in which the size of a finite population is not a multiple of the sample size. Subramani(2000)introduced what is known as diagonal-systematic sampling. As its name suggests, the units are selected diagonally in diagonal systematic sampling. Sampath and Varalakshmi(2008)and Subramani(2009)introduced modified forms of diagonal systematic sampling design. In those situations where the sample size is an odd number, Subramani(2012) developed a new version of systematic sampling which was shown to be more efficient than other forms of systematic sampling. Another improved version of the classical systematic sampling was introducedby Subramani and Gupta(2014). The advantage of the Subramani and Gupta (2014) approach was that it didn't need the size of the finite population to be a multiple of the required sample size. More recently, Azeem (2021) studied estimation of proportion based on the diagonal systematic sampling design. Many other researchers have also analyzed various aspects of different versions of systematic sampling under different practical situations, such as studies (Madow, 1953; Yates, 1948; Bellhouse and Raom, 1975; Bellhouse, 1988; Fountain and Pathak, 1989; Sampath and Uthayakumaran, 1998; Subramani, 2013; Horvitz and Thompson, 1952)

Recently, Azeem et al. (2021) suggested a modified version of diagonal systematic random sampling which was more efficient than both linear and diagonal systematic sampling and some other existing sampling designs. Motivated by Azeem et al. (2021), this paper introduces a modified version of Azeem et al.(2021) method. It is observed that the new suggested systematic sampling design is more efficient than Azeem et al. (2021) sampling scheme and some other available sampling designs in situations where a linear trend is present in the population units. The variance of the sample mean on the basis of the new suggested sampling design is derived. The improvement in efficiency is shown for a real data set as well as for situations with a perfect linear trend.

2. Proposed sampling method

Consider a finite population consisting of N units and suppose a sample of size n is needed to be drawn such that$N = nk = k \cdot k + (n - k - 1)k + k$. The proposed method selects the sample in the following steps:

1) Divide the population of size N into three non-overlapping sets of units: Set-1, Set-2 and Set-3, in such a manner that Set-1 gets the first $k \times k = k^2$ units y_i (i=1, 2, ..., k^2) thus forming a $k \times k$ square matrix, Set-2 gets the next$(n - k - 1)k$units y_i $(i = k \cdot k + 1,\ k \cdot k + 2,\ k \cdot k + 3, \ldots, (n-1)k)$, whereas Set-3 gets the last k units y_i $(i = (n-1) \cdot k + 1,\ (n-1) \cdot k + 2,\ (n-1) \cdot k + 3, \ldots, nk)$.

2) In Set-1, write the units in a $k \times k$ square matrix. In Set-2, arrange the $(n-k)k$ units in a matrix of order $(n-k-1) \times k$, and in Set-3, organize the last k units in a matrix order $1 \times k$, as shown in Table 1.

3) Select three random numbers r_1, r_2 and r_3 where $1 \le r_1 \le k$, $1 \le r_2 \le k$ and $1 \le r_3 \le k$. In Set-1, the units for the sample are chosen in such a way that the selected k units belong to the diagonal / broken diagonal of the matrix. In Set-2, the units are chosen in such a manner that the selected $n-k$ units belong to the r_2th column of the matrix. Moreover, in Set-3, one unit is randomly selected from the total k units in the set. Finally, the units selected from all three sets are merged into a single group, yielding the required sample of size n.

The difference between the Azeem et al. (2021) method and the proposed method is that the new method puts the last row of the Set-2 in Azeem et al. (2021) method into a separate set, thus forming total three sets of population units, compared to the two sets of population units in Azeem et al. (2021) method. In Section 4, we show that this approach improves the efficiency of Azeem et al. (2021) method.

Table 1. Population units organized in Set-1, Set-2 and Set-3.

Set-1					**Set-2**				
S.No.	1	2	…	k	S.No.	1	2	…	k
1	y_1	y_2	…	y_k	$k+1$	y_{kk+1}	y_{kk+2}	…	$y_{kk+k=(k+1)k}$
2	y_{k+1}	y_{k+2}	…	y_{2k}	$k+2$	$y_{(k+1)k+1}$	$y_{(k+1)k+2}$	…	$y_{(k+2)k}$
3	y_{2k+1}	y_{2k+2}	…	y_{3k}	$k+3$	$y_{(k+2)k+1}$	$y_{(k+2)k+2}$	…	$y_{(k+3)k}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	$y_{(k-1)k+1}$	$y_{(k-1)k+2}$	…	y_{kk}	n-1	$y_{(n-2)k+1}$	$y_{(n-2)k+2}$	…	$y_{(n-1)k}$
Set-3									
S.No.	1	2	…	k					
N	$y_{(n-1)k+1}$	$y_{(n-1)k+2}$	…	y_{nk}					

The total number of possible samples that can be selected under the new sampling scheme is $k \times k \times k = k^3$, each having size n. The first order and second-order probabilities of inclusion based onthe suggested sampling design are given by:

$$\pi_i = \frac{1}{k} \tag{1}$$

And

$$\pi_{ij} = \begin{cases} \frac{1}{k}, & \text{if } i\text{th and } j\text{th units belong to the same diagonal / broken diagonal of Set-1,} \\ \frac{1}{k}, & \text{if } i\text{th and } j\text{th units belong to the same column of Set-2,} \\ \frac{1}{k^2}, & \text{if } i\text{th and } j\text{th units are from two different sets,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Generally, the units selected under the proposed sampling scheme are:

$$S_{r_1 r_2 r_3} = \begin{cases} y_{r_1}, y_{(k+1)+r_1}, \ldots, y_{(k-1)(k+1)+r_1}, y_{kk+r_2}, y_{(k+1)k+r_2}, \ldots, y_{(n-2)k+r_2}, y_{r_3}. & if \quad r_1 = 1 \\ y_{r_1}, y_{(k+1)+r_1}, \ldots y_{t(k+1)+r_1}, y_{(t+1)k+1}, y_{(t+2)k+2}, \ldots, y_{(k-1)k+k-t-1}, y_{kk+r_2}, y_{(k+1)k+r_2}, \ldots, y_{(n-2)k+r_2}, y_{r_3}. & if \quad r_1 > 1 \end{cases}$$

where $r_2 = 1, 2, \ldots, k$, $r_3 = 1, 2, \ldots, k$.

The sample mean on the basis of the new suggested sampling design is given by:

$$\bar{y}_{msy} = w_1\bar{y}_1 + w_2\bar{y}_2 + w_3\bar{y}_3, \tag{3}$$

where,

$$\bar{y}_1 = \begin{cases} \frac{1}{k}\sum_{l=0}^{k-1} y_{l(k+1)+r_1}, & \text{if } r_1 = 1, \\ \frac{1}{k}\left(\sum_{i=0}^{t} y_{i(k+1)+r_1} + \sum_{i=1}^{k-t-1} y_{(t+i)k+i} + y_{r_3}\right), & \text{if } r_1 > 1. \end{cases} \tag{4}$$

where $t=k-r_1$,

$$\bar{y}_2 = \frac{1}{n-k}\sum_{l=k}^{n-1} y_{lk+r_2}, \quad w_1 = \frac{k}{n}, \quad w_2 = \frac{n-k-1}{n}, \quad w_3 = \frac{1}{n}, \quad w_1 + w_2 + w_3 = 1, \tag{5}$$

and,

$$\bar{y}_3 = y_{r_3}. \tag{6}$$

Theorem: The sample mean under the proposed method may be expressed in the form given by Horvitz and Thompson(1952) and is unbiased estimator of population mean with variance given by:

$$Var(\bar{y}_{msy}) = \frac{1}{N^2}\left[k^4\left\{\frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{1i}-\bar{Y}_1)^2\right\}+(n-k-1)^2k^2\left\{\frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{2i}-\bar{Y}_2)^2\right\}+k^2(\bar{y}_{3i}-\bar{Y}_3)^2\right],$$

where $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}_3$ denotethe means of the samples obtained from Set-1, Set-2 and Set-3 respectively. Moreover, $\bar{Y}_1$, $\bar{Y}_2$ and $\bar{Y}_3$ denote the means of all units of Set-1, Set-2and Set-3, respectively, and k denotes the number of all possible samples that can be drawn under the new sampling scheme.

Proof: By definition

$$\bar{y}_{msy} = \frac{k^2}{N}\bar{y}_1 + \frac{(n-k-1)k}{N}\bar{y}_2 + \frac{k}{N}\bar{y}_3 = \frac{1}{N}\left(k\sum_{i\in s_1} y_{1i} + k\sum_{i\in s_2} y_{2i} + k\sum_{i\in s_3} y_{3i}\right), \tag{7}$$

where s_1, s_2 and s_3 are the samples obtained from Set-1, Set-2 and Set-3, respectively.

$$\bar{y}_{msy} = \frac{1}{N}\left(\sum_{i\in s_1}\frac{y_{1i}}{1/k} + \sum_{i\in s_2}\frac{y_{2i}}{1/k} + \sum_{i\in s_3}\frac{y_{3i}}{1/k}\right) = \frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i} = \bar{y}_{HT}, \tag{8}$$

where 's'denote the sample drawn from the total units of population (Set-I + Set-2 + Set-3). Applying expectation on both sides of (7) gives:

$$E(\bar{y}_{msy}) = \frac{k^2}{N}E(\bar{y}_1) + \frac{(n-k-1)k}{N}E(\bar{y}_2) + \frac{k}{N}E(\bar{y}_3). \tag{9}$$

Now,

$$E(\bar{y}_1) = E\left(\frac{1}{k}\sum_{i=1}^{k} y_{1i}\right) = \frac{1}{k}\sum_{i=1}^{k} E(y_{1i}).$$

$$= E(\frac{1}{k}\sum_{i=1}^{k} y_{1i}) = E(y_{1i}) = \bar{Y}_1 \tag{10}$$

Similarly,

$$E(\bar{y}_2) = E\left(\frac{1}{n-k-1}\sum_{i=1}^{n-k-1} y_{2i}\right) = E(y_{2i}) = \bar{Y}_2, \tag{11}$$

and,

$$E(\bar{y}_3) = E(y_{3i}) = \bar{Y}_3 \tag{12}$$

Now using (10), (11) and (12) in (9) and on simplification, we get $E(\bar{y}_{msy}) = \bar{Y}$.

Now applying variance on both sides of (3) gives:

$$Var\left(\bar{y}_{msy}\right)=\frac{k^4}{N^2}Var\left(\bar{y}_1\right)+\frac{\left(n-k-1\right)^2k^2}{N^2}Var\left(\bar{y}_2\right)+\frac{k^2}{N^2}Var\left(\bar{y}_3\right), \tag{13}$$

where,

$$Var\left(\bar{y}_1\right)=\frac{1}{k}\sum_{i=1}^{k}\left(\bar{y}_{1i}-\bar{Y}_1\right)^2, \tag{14}$$

since each possible sample in Set-1 has probability equal to *1/k*. Similarly,

$$Var\left(\bar{y}_2\right)=\frac{1}{k}\sum_{i=1}^{k}\left(\bar{y}_{2i}-\bar{Y}_2\right)^2, \tag{15}$$

$$Var\left(\bar{y}_3\right)=\frac{1}{k}\sum_{i=1}^{k}\left(\bar{y}_{3i}-\bar{Y}_3\right)^2. \tag{16}$$

Using (14), (15) and (16) in (13), the variance of the mean $\bar{y}_{msy}$ under the suggested sampling scheme is obtained as:

$$Var\left(\bar{y}_{msy}\right)=\frac{1}{N^2}\left[k^4\left\{\frac{1}{k}\sum_{i=1}^{k}\left(\bar{y}_{1i}-\bar{Y}_1\right)^2\right\}+\left(n-k-1\right)^2k^2\left\{\frac{1}{k}\sum_{i=1}^{k}\left(\bar{y}_{2i}-\bar{Y}_2\right)^2\right\}+k^2\left(\bar{y}_{3i}-\bar{Y}_3\right)^2\right]. \tag{17}$$

Remark 1: Using the approach given by Sen-Yates-Grundy(Sen, 1953; Yates and Grundy, 1953), the variance of $\bar{y}_{msy}$ can be expressed as:

$$Var\left(\bar{y}_{mdsy}\right)=\frac{1}{N^2}\left\{\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\right\}=Var_{SYG}\left(\bar{y}_{HT}\right) \tag{18}$$

Remark 2: The Sen-Yates-Grundy type estimator of the sampling variance given in (18), is given as:

$$var\left(\bar{y}_{mdsy}\right)=\frac{1}{N^2}\left\{\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\right\}=var_{SYG}\left(\bar{y}_{HT}\right) \tag{19}$$

One can use the values of π_i and π_{ij} given in (1) and (2) in equation (18) and (19) to get the sampling variance of the sample mean as well as its estimator based on the proposed method.

3. Linear Trend

Linear trend refers to the arithmetic progression which the population units may follow, which may either in increasing or decreasing order. In some practical circumstances, a moderate to a high degree of linear trend can occur. As an illustration, educational institutes almost everywhere in the world offer admissions on merit basis in various academic departments. Many universities tend to allocate roll numbers to their students on the basis on their quantified academic scores during admission process. In such cases, intelligent students tend to occupy the top enrollment numbers. Thus, upon admission, if the examination marks obtained by students are observed in order of students' enrollment roll numbers, one can expect a high degree of increasing linear trend in this data set since the top roll-numbered students, being the merit toppers, tend to perform better in subsequent examinations.

Likewise, to observe linear trend in practice, one can also consider the daily record of the milk yield data, starting from calving. One can naturally expect that the daily milk yield quantity will tend to decrease over time, thus creating a linear trend.

Suppose the $N=nk=k\cdot k+(n-k-1)k+k$ population units exhibit a linear trend. Thus,

$$y_i = a + ib, \quad \text{for } i = 1, 2, 3, \ldots, N. \tag{20}$$

The variance of the sample mean of simple random sampling under perfect linear trend is given as:

$$Var(\bar{y}_r) = (k-1)(N+1)\frac{b^2}{12}. \tag{21}$$

The variance of the sample mean under linear systematic sampling design is as follows:

$$Var(\bar{y}_{sy}) = (k-1)(k+1)\frac{b^2}{12}. \tag{22}$$

Further, the variance in the case of diagonal systematic sampling method is given as:

$$Var(\bar{y}_{dsy}) = (k-n)\left[n(k-n)+2\right]\frac{b^2}{12n}. \tag{23}$$

Where $N=nk+r$. The variance under the modified systematic sampling suggested by Subramani (2012) is given as:

$$Var(\bar{y}_{ssy}) = \left(\frac{(n-1)^2+1}{n^2}\right)(k-1)(k+1)\frac{b^2}{12}. \tag{24}$$

The variance of the Azeem et al. (2021) sampling scheme is given as:

$$Var(\bar{y}_{mdsy}) = \left(\frac{n-k}{n}\right)^2 (k-1)(k+1)\frac{b^2}{12}. \tag{25}$$

Finally, under perfect linear trend in the population units, the variance of the suggested sampling scheme can be derived as:

$$Var\left(\bar{y}_{msy}\right) = w_1^2 Var\left(\bar{y}_1\right) + w_2^2 Var\left(\bar{y}_2\right) + w_3^2 Var\left(\bar{y}_3\right). \tag{26}$$

In order to obtain the variance of the proposed sampling scheme under perfect linear trend, we first need to obtain the variance expressions for $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}_3$. Since the total number of units in Set-1 are $k \times k = k^2$, so using $k=n$ in (23) leads to:

$$Var\left(\bar{y}_1\right) = 0. \tag{27}$$

Moreover, Set-2 contains total$(n-k-1)k$ units where the linear systematic sampling method is utilized. Since the right-hand side of (22) does not depend on n, thus

$$Var\left(\bar{y}_2\right) = \left(k-1\right)\left(k+1\right)\frac{b^2}{12}. \tag{28}$$

Also,

$$Var\left(\bar{y}_3\right) = \left(k-1\right)\left(k+1\right)\frac{b^2}{12}. \tag{29}$$

Substituting (27), (28) and (29) in (26), and after simplification, the variance of $\bar{y}_{msy}$ is obtained as:

$$Var\left(\bar{y}_{msy}\right) = \left[\frac{\left(n-k-1\right)^2 + 1}{n^2}\right]\left(k-1\right)\left(k+1\right)\frac{b^2}{12}. \tag{30}$$

4. Theoretical Comparison

In this section, the variances of proposed method and some of the available popular sampling schemes are compared and the efficiency conditions are derived.

4.1 Comparison with Simple Random Sampling

In situations where the units follow a perfect linear trend, the suggested sampling scheme will be more efficient than the simple random sampling if:

$$Var\left(\bar{y}_{msy}\right) < Var\left(\bar{y}_r\right) \tag{31}$$

Substituting (21) and (30) in (31) and simplification yields:

$$\left[\frac{(n-k-1)^2+1}{n^2}\right](k+1) < N+1. \tag{32}$$

Condition (32) is strong and always holds since $N = nk > k$. This means that the suggested sampling procedure is more efficient than the simple random sampling procedure.

4.2 Comparison with Linear Systematic Sampling

The suggested sampling scheme will be more efficient than the linear systematic sampling scheme if:

$$Var\left(\bar{y}_{msy}\right) < Var\left(\bar{y}_{sy}\right) \tag{33}$$

Substituting (22) and (30) in (33) yields:

$$\left[\frac{(n-k-1)^2+1}{n^2}\right](k-1)(k+1)\frac{b^2}{12} < (k-1)(k+1)\frac{b^2}{12}.$$

The above inequality on further simplification reduces to:

$$\left[\frac{(n-k-1)^2+1}{n^2}\right] < 1. \tag{34}$$

which is always true, thus the suggested method is more efficient than linear systematic sampling.

4.3 Comparison with Subramani's(2012)modified sampling scheme

Under perfect linear trend, the suggested sampling scheme will be more efficient than Subramani's(2012) sampling procedure if:

$$Var\left(\bar{y}_{msy}\right) < Var\left(\bar{y}_{ssy}\right) \tag{35}$$

Using (24) and (30) in (35),

$$\left[\frac{(n-k-1)^2+1}{n^2}\right](k-1)(k+1)\frac{b^2}{12} < \left(\frac{(n-1)^2+1}{n^2}\right)(k-1)(k+1)\frac{b^2}{12},$$

or

$$\left[\frac{(n-k-1)^2+1}{n^2}\right] < \left(\frac{(n-1)^2+1}{n^2}\right),$$

or,

$$\left(n-k-1\right)^2<\left(n-1\right)^2. \tag{36}$$

which always holds. This implies that the suggested method is always more efficient than Subramani's(2012) modified sampling.

4.4 Comparison with Diagonal Systematic Sampling

The suggested sampling design will be more efficient than diagonal systematic random sampling if:

$$Var\left(\bar{y}_{msy}\right)<Var\left(\bar{y}_{dsy}\right) \tag{37}$$

Using (23) and (30) in (37),

$$\left[\frac{\left(n-k-1\right)^2+1}{n^2}\right]\left(k-1\right)\left(k+1\right)\frac{b^2}{12}<\left(k-n\right)\left[n\left(k-n\right)+2\right]\frac{b^2}{12n},$$

or,

$$\left[\frac{\left(n-k-1\right)^2+1}{n}\right]\left(k-1\right)\left(k+1\right)<\left(k-n\right)\left[n\left(k-n\right)+2\right],$$

or,

$$\left[\left(n-k-1\right)^2+1\right]\left(k-1\right)\left(k+1\right)<n\left(k-n\right)\left[n\left(k-n\right)+2\right]. \tag{38}$$

Since the proposed method uses $n>k$, which implies condition (38) is strong and always holds.

4.5 Comparison with Azeem et al. (2021) sampling scheme

The suggested sampling design will be more efficient than the Azeem et al. (2021) sampling scheme if:

$$Var\left(\bar{y}_{msy}\right)<Var\left(\bar{y}_{mdsy}\right) \tag{39}$$

Using (25) and (30) in (39),

$$\left[\frac{\left(n-k-1\right)^2+1}{n^2}\right]\left(k-1\right)\left(k+1\right)\frac{b^2}{12}<\left(\frac{n-k}{n}\right)^2\left(k-1\right)\left(k+1\right)\frac{b^2}{12},$$

or

$$(n-k-1)^2+1<(n-k)^2,$$

or

$$(m-1)^2+1<m^2, \text{ where } m=n-k. \qquad (40)$$

Since the proposed method uses $n>k$, so condition (40) is strong and always hold for $m=n-k>1 \Rightarrow n>k+1$.

5 Empirical Results

The improvement in efficiency of the proposed method over the existing method is as sense dusing the milk yield data taken from Pandey and Kumar (2014). The results in Table 2clearly indicate that the new suggested sampling design is more precise than some of the commonly used sampling methods. The milk yield (in liters) related to the Sahiwal cows for a 252 consecutive days period, starting from the day of calving, has been obtained from Pandey and Kumar(2014). One can clearly observe a decreasing linear trend in the data set where the milk yield follows a decreasing pattern over time. The variances of various systematic sampling methods for milk-yield data set are presented in Table 2. The findings clearly indicate that the new suggested sampling procedure is more precise than the available sampling designs discussed in Section-3.As the population size in the milk yield data is N = 252 and since systematic sampling sheme needs $N=nk$, so for various choices of the values of n and k, a few units from the population in the milk yield data were randomly removed in order to compromise between the choice of values of N, n, and k. As an example, if we choose$n$ = 10 and k = 25, two population units were deleted randomly in order to reduce the population size to$N$ = 250 units in place of using N = 252 units. This will make efficiency comparison feasible for milk yield data.

Table 2. Calculations of variances of various sampling designs using milk yield data set.

n	k	$Var(\bar{y}_r)$	$Var(\bar{y}_{sy})$	$Var(\bar{y}_{dsy})$	$Var(\bar{y}_{ssy})$	$Var(\bar{y}_{mdsy})$	$Var(\bar{y}_{msy})$
83	3	1.7329	2.0410	1.6124	0.9154	0.8653	0.6317
63	4	1.5051	1.1436	1.0070	0.7900	0.5180	0.3964
49	5	1.0628	0.7286	0.6023	0.5826	0.3496	0.2386
42	6	0.9291	0.6542	0.5317	0.4604	0.2604	0.1804
35	7	0.7690	0.5881	0.4781	0.3713	0.2267	0.1438
31	8	0.7157	0.5017	0.4496	0.3418	0.1807	0.1365
28	9	0.6143	0.4210	0.3305	0.2501	0.1268	0.1006
25	10	0.5243	0.3781	0.2851	0.1947	0.1097	0.0835
22	11	0.5034	0.3535	0.2641	0.1693	0.0988	0.0718
21	12	0.4632	0.3130	0.2345	0.1471	0.0923	0.0701
19	13	0.4153	0.3042	0.2160	0.1260	0.0891	0.0691
18	14	0.3613	0.2896	0.1990	0.1063	0.0856	0.0643
16	15	0.3650	0.2775	0.1630	0.0993	0.0833	0.0594

Now consider the cases where the units exhibit a perfect linear trend. We have performed efficiency comparison for different choices of the values of N, n, and k. The variances of the sample mean of the suggested and other popular sampling methods have been given in Table 3.The different values of N, n, and $k$for the purpose of efficiency comparison have been selected in such a manner that $N = nk$ and $n > k$. Note that the constant b^2 is a multiplication factor in the variance expressions of all of the sampling procedures discussed in Section-3, so in order to make the analysis easier, we have b = 1 in the calculation of variances. The findings from Table-3clearly show that the new suggested systematic sampling scheme is superior in terms of efficiency over the available sampling schemes, including the one suggested by Azeem et al. (2021).

6 Conclusion

As opposed to the Azeem et al. (2021) sampling scheme where the population is divided into two disjoint sets, the suggest sampling scheme partitions the population into three mutually exclusive subpopulations. In order to select a sample, the proposed method utilizes the diagonal systematic sampling method in the first set, linear systematic sampling method in the second set, and the selection of a single unit from the third set. Then a weighting approach has been used to estimate the population mean based on the new sampling scheme. Efficiency comparison has been carried out to analyze the performance of the new sampling scheme with other existing sampling schemes, using a real data set as well as perfect linear trend, and the improvement in efficiency has been shown. Based on empirical and theoretical findings in the current study, the proposed method is recommended for use in practical situations where a high degree of an increasing or decreasing linear trend exists.

Table 3. Linear trend – based variances of different sampling designs

n	k	$Var(\bar{y}_r)$	$Var(\bar{y}_{sy})$	$Var(\bar{y}_{dsy})$	$Var(\bar{y}_{ssy})$	$Var(\bar{y}_{mdsy})$	$Var(\bar{y}_{msy})$
10	4	10.25	1.25	2.90	1.03	0.45	0.33
	6	25.42	2.92	1.27	2.39	0.47	0.29
	8	47.25	5.25	0.30	4.31	0.21	0.11
30	5	50.33	2.00	51.94	1.87	1.39	1.28
	10	225.75	8.25	33.22	7.72	3.67	3.32
	15	526.17	18.67	18.67	17.46	4.67	4.09
	20	951.58	33.25	8.28	31.11	3.69	3.03
	25	1502.00	52.00	2.06	48.65	1.44	0.98
50	10	375.75	8.25	133.20	7.93	5.28	5.02
	20	1584.92	33.25	74.90	31.95	11.97	11.20
	30	3627.42	74.92	33.27	71.98	11.99	10.85
	40	6503.25	133.25	8.30	128.03	5.33	4.37
100	20	3168.25	33.25	533.20	32.59	21.28	20.75
	40	13003.25	133.25	299.90	130.61	47.97	46.40
	60	29504.92	299.92	133.27	293.98	47.99	45.65
	80	52673.25	533.25	33.30	522.69	21.33	19.30
500	100	412508.25	833.25	13333.20	829.92	533.28	530.62
	200	1658349.92	3333.25	7499.90	3319.94	1199.97	1192.00

300	3737524.92	7499.92	3333.27	7469.98	1199.99	1188.05
400	6650033.25	13333.25	833.30	13280.02	533.33	522.77

Acknowledgments

The author received no funds for this study.

Conflict of interest

The author has no conflict of interest to declare.

References

1. Azeem, M., Asif, M., Ilyas, M., Rafiq, M. and Ahmad, S.(2021). An efficient modification to diagonal systematic sampling for finite populations. *AIMS Mathematics*,6(5): 5193-5204.

2. Madow,W.G., Madow, L.H.(1944). On the theory of systematic sampling, I. *Ann. Math. Statist.*,15: 1–24.

3. Lahiri, D.B. (1951). A method for selection providing unbiased estimates. *International Statistical Association Bulletin*, 33: 133-140.

4. Chang, H.J., Huang, K.C.(2000). Remainder linear systematic sampling, *Sankhya: Indian J. Stat.Ser. B*,62:249–256.

5. Subramani, J. (2000). Diagonal systematic sampling scheme for finite populations. *Journal of the Indian Society of Agricultural Statistics*, 53(2): 187-195.

6. Sampath, S. (2008). Varalakshmi V. Diagonal circular systematic sampling, *Model Assisted Stat. Appl.*, 3: 345-352.

7. Subramani, J.(2009). Further results on diagonal systematic sampling for finite populations ,*J. Indian Soc. Agric. Stati.*, 63: 277-282.

8. Subramani, J.(2012). A modification on linear systematic sampling for odd sample size,*Bonfring Int. J. Data Min.*, 2: 32-36.

9. Subramani, J., Gupta, S.N.(2014). Generalized modified linear systematic sampling scheme for finite populations,*Hacet. J. Math. Stat.*, 43: 529-542.

10. Azeem, M.(2021). On estimation of population proportion in diagonal systematic sampling. *Life Cycle Reliability and Safety Engineering*, 10: 249-254.

11. Madow, W.G.(1953). On the theory of systematic sampling III-comparison of centered and random start systematic sampling, *Ann. Math. Statist.*, 24: 101-106.

12. Yates, F.(1948). Systematic sampling. *Philosophical Transactions of the Royal Society,* A 241, 345-377.

13. Bellhouse, D.R., Raom, J.N.K. (1975). Systematic sampling in the presence of linear trends, *Biometrika*, 62: 694-697.

14. Bellhouse, D.R. (1988). Systematic sampling. *Handbook of Statistics*, 6: 125-145.

15. Fountain, R.L., Pathak, P.L.(1989). Systematic and non-random sampling in the presence of linear trends,*Commun. Stat –Theory* M., 18: 2511-2526.

16. Sampath, S., Uthayakumaran, N. (1998). Markov systematic sampling, *Biometrical J.*, 40: 883-895.

17. Subramani, J. (2013). A modification on linear systematic sampling, *Model Assisted Stat. Appl.*, 8: 215-227.

18. Horvitz, D.G., Thompson, D.J.(1952). A generalization of sampling without replacement from finite universe,*J. Am. Stat. Assoc.*, 47: 663-685.

19. Sen, A.R. (1953). On the estimate of variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5: 119-127.

20. Yates, F., Grundy, P.M.(1953). Selection without replacement from within strata with probability proportional to size,*J. R. Stat. Soc.* Ser. B, 15:253-261.

21. Pandey, T.K., Kumar, V.(2014). Systematic sampling for non-linear trend in milk yield data, *J. Reliab. Stat. Stud.*,7: 157-168.