Thailand Statistician
April 2025; 23(2): 420-437
http://statassoc.or.th
Contributed paper

Enhancement of Systematic Sampling with Incorporation from Consecutive Sampling; Systematic Sampling with Consecutive Approach

Mohamad Adam Bujang* [a], Tg Mohd Ikhwan Tg Abu Bakar Sidik [b] and Puzziawati Abdul Ghani [c]

[a] Clinical Research Centre, Sarawak General Hospital, Ministry of Health, Sarawak, Malaysia.
[b] Faculty of Medicine, Universiti Kebangsaan Malaysia (UKM), Kuala Lumpur, Malaysia.
[c] Faculty Computer and Mathematical Sciences, University of Technology MARA (UiTM), Selangor, Malaysia.
*Corresponding author; e-mail: adam@crc.gov.my

Received: 10 February 2022
Revised: 15 November 2022
Accepted: 31 December 2022

Abstract

This study develops a newly modified systematic sampling that allows of recruiting more than one-unit subject in each interval selection. The newly modified systematic sampling is an incorporation from both consecutive sampling and systematic sampling. Firstly, the terms and selection of the newly modified systematic sampling is identified when the sample selection derived from the new sampling technique is able to generate an unbiased estimator for the population mean. Then, the evaluation from simulation works and a real-life dataset was applied to support the evidence of sampling efficiency and consistency. This newly modified systematic sampling is named as "Systematic Sampling with Consecutive Approach" or in short SSC is indeed an innovation in sampling by the virtue of combining two sampling techniques, consecutive sampling and systematic sampling.

Keywords: Sampling technique, consecutive sampling, systematic sampling, unbiased, efficiency.

1. Introduction

Systematic sampling is considered as a type of probability sampling method where sample units are selected according to a random starting point and based on a fixed periodic interval (Madow and Madow 1944, Cochran 1946, Yates 1948). Systematic sampling shall serve as the ideal probability sampling procedure for research survey which does not require a sampling frame like simple random sampling. Hence, this sampling technique is very popular especially for research that require sample selection from an ordered population.

However, the adoption of the conventional systematic sampling procedure may lead to two major constraints. Firstly, a researcher might find it difficult obtaining the samples when the researcher will have to wait for every k^{th} sample to be selected especially for only one-unit subject in each interval

selection. And secondly, this sampling procedure will require a longer period of time for obtaining the subjects with the required sample size. Therefore, if the modified systematic sampling allows more than one-unit subject to be collected in each interval selection, then the burden from these two constraints can be reduced. So far, none of the researchers has successfully prove that an unbiased estimator can be derived based on modified systematic sampling procedures that obtain more than one-unit subject in each interval selection.

The purpose of this study is to enhance a newly systematic sampling procedure that allows to collect more than one-unit subject in each interval selection. Hence, it is necessary to first list down all the terms and assumptions of the selection process for this newly modified systematic sampling procedure. Then followed by a derivation and demonstration of the proof that the estimator obtained from the modified systematic sampling procedure is unbiased and efficient.

Table 1 Schematic diagram that showing k systematic samples in the population

	Sample number $(i = 1, 2, ..., k)$			
	S_1	S_2	S_3	S_k
Sample values $(j = 1, 2, ..., n)$	y_{11}	y_{21}	y_{i1}	y_{k1}
	y_{12}	y_{22}	y_{i2}	y_{k2}
	y_{13}	y_{23}	y_{i3}	y_{k3}
	•	•	•	•
	•	•	•	•
	•	•	•	•
	y_{ij}	y_{2j}	y_{ij}	y_{kj}
	$y_{1(j+1)}$	$y_{2(j+1)}$	•	•
	•	•	•	•
	•	•	•	•
	•	•	•	•
	y_{1n}	y_{2n}	y_{in}	y_{kn}

1.1. Sample selection for systematic sampling

The sample selection for systematic sampling is firstly introduced. Assumed that the population is arranged in sequential list from number 1 to N. Then, let N be defined in the form $N = nk$ and let the selected number be $i, (1 < i < k)$. In this sampling, the sample contains the units $i, i+k, i+2k, ..., i+(n-1)k$. This procedure will generate k systematic sample each of size n with an equal probability that can be seen in Table 1. The population is assumed to be arranged into two different ways such as; There are k clusters in the population each of size n and only one cluster is selected randomly to get a systematic sample, and there are n strata each of sized k and only one-unit from each stratum is selected to constitute the required sample. The unit of i^{th} cluster (sample) can be represented as $i+(j-1)k,\ j = 1, 2, ..., n.$

1.2. History of systematic sampling and modified systematic sampling

Systematic sampling is known as one of the probability sampling technique. Systematic sampling was first introduced by Madow and Madow (1944) and then was further explored by Cochran (1946) and Yates (1948) in the earlier development of systematic sampling. After the first development of systematic sampling, there were more modified systematic sampling techniques developed by mathematicians and statisticians (Singh et al. 1968, Hartley and Rao 1962, Singh and Singh 1977, Shiue 1960, Zinger 1980).

The other development was by Khan et al. (2013). Their technique was the improved version of systematic sampling when Linear Systematic Sampling (LSS) is no longer applicable. Before that, few techniques were introduced such as Balanced Circular Systematic Sampling (BCSS) by Uthayakumaran (1998), Diagonal Systematic Sampling (DSS) by Subramani (2000), Diagonal Circular Systematic Sampling by Sampath and Varalakshmi (2008) and Generalization of Diagonal Systematic Sampling by Subramani (2010). The recent development was by Khan et al. (2019). Their study developed Systematic ranked set-sampling scheme (SRSS) for symmetric distributions which is the improved version of Ranked set-sampling (RSS) which was introduced by Mclntyre (1952).

In general, majority of these modified systematic sampling has the same similarities where all these sampling techniques only allow one-unit subject is to be selected in each interval selection which is similar with the selection from conventional systematic sampling. Although Bujang et al. (2012) had proposed taking more than one-unit subject in each interval selection, but this paper did not portray the structured random selection and also did not discuss from the aspects of the unbiasedness of the estimator. Therefore, the sample selection was considered more as non-probability sampling or convenience sampling instead.

1.3. Development of a newly modified systematic sampling

The selection techniques for consecutive sampling and systematic sampling are already known. Therefore, we named the newly modified sampling technique as “systematic sampling with consecutive approach” or in short “SSC”. The idea of SSC is that to incorporate the consecutive sampling within the interval of systematic sampling. Previously, there is only one-unit subject can be recruited in every interval selection with respect to systematic sampling. But now, SSC will allows more than one-unit subject to be recruited in every interval selection. Therefore, the process in developing the terms and selection for SSC is started with the incorporation between consecutive sampling and systematic sampling.

During the process, there were many trials and error have been made to develop the terms and selection of SSC. The terms and selection of SSC is finalized when the terms and selection is able to produce an unbiased estimator for population mean. This is one of the important requirements of probability sampling. To do this, a selection of sample is first obtained by adopting this new SCC sampling technique and then a mathematical equation were applied to prove the estimator such as the sample mean that derives from SSC is an unbiased estimator.

In designing the sample selection, the development of this modified systematic sampling technique involves multiple rounds of trial-and-error. This is because this newly modified sampling technique will have to produce an unbiased estimator, before it can be adopted as a scientifically-valid probability sampling technique. To understand terms and selection of SSC, first of all, there are new terminology and definitions to be introduced such as N, m, q, v and w. The definitions of these terminology are as follow:

N : total number in the entire population,

m : the targeted sample size,

q : the number of samples (e.g., sample 1, sample 2, etc.) of size m that can be obtained from the population where q is always an even number,

c : the number of cluster where $c = q/2$,

v : the number of batch of consecutive sample until all samples unit have been obtained (e.g., $v = 4$ indicates that the selection needs 4 batches until all the desired sample size is achieved),

w : the total number of consecutive samples that will be obtained and also the total number of samples to be skipped (e.g., $w = 4$ indicates that 4 consecutive samples unit will be taken, then skip another 4 samples unit before select another 4 consecutive samples. The selection of consecutive samples continues until v times).

Let the terms are expressed in the form:

$$\frac{N}{m} = q \text{ (where } \frac{q}{2} = c) \quad \text{and} \quad \frac{m}{w} = v,$$

where $q \in 2\mathbb{N}$, $v \in \mathbb{N}$ and $\mathbb{N} = 1, 2, 3, ...$ ($\mathbb{N}$ is a natural number).

For the SSC, the following modified formula is proposed, in which w will stand for the total number of consecutive samples to be selected and also as the skip of interval selection. This method will obtain w samples in each interval selection, whereas the conventional systematic sampling technique shall only obtain one-unit subject in each interval selection. The sample will be obtained at every w^{th} selection after the first sample has been obtained.

So, there will be v times for each w sample to be obtained which satisfy $m = wv$ where m, w, v are the natural number. It is also assumed that the population is arranged into $q/2$ clusters with the size of $2m$. In each cluster, a further sample by using SSC is derived in order to obtain two samples of size m with only one sample randomly selected from all the clusters. In order to adopt SSC for obtaining a sample of size m from a population of N units, we can proceed as follows:

Step 1

After the population (N) and the required sample size (m) are specified, next is to determine the values of q, c, w and v. $N/m = q$ where $q \in 2\mathbb{N}$ (meaning q can always be divided by 2), so $q/2 = c$ (c is always an integer or natural number). Then, determine w where $m/w = v$ (w and v are always an integer or natural number). Thus, the value of w is a common divisor for m.

Step 2

The population will be divided into $q/2$ clusters, namely: $c = (c_1, c_2, ..., c_{q/2})$. The sample size in each cluster will be $2m$ and let $(y_1, y_2, ..., y_{2m})$ be the corresponding values of the study variable y.

Step 3

Within each cluster, the selection will begin by obtaining the first sample and then followed by the next w^{th} sample $(y_1, y_2, ..., y_l)$ in a consecutive manner. This will be followed by a skip of the w^{th} samples. After that, the next batch of w^{th} $(y_{1+2w}, y_{2+2w}, ..., y_{w+2w})$ consecutive samples will be obtained. The arrangement of sample unit in the first cluster need to be completed first before arrange

the sample unit for the subsequent cluster. By doing so, it is possible to adopt SSC in generating the total number of q samples, with each of size m, and of equal probability which is presented in Table 2.

Table 2 Schematic diagram showing the systematic sampling with consecutive approach

	Cluster number (c_{a+1}) where $a = 0, 1, ..., (q-2)/2$ and sample number (q_j) where $j = 1, 2, ..., q$			
	c_1		c_{a+1}	
	q_1	q_2	q_{1+2a}	q_{2+2a}
Individual unit values (y_i) where $i = 1, 2, ..., m$ and $l = 0, 1, ..., v-1$	y_1	y_{w+1}	$y_{2w(av+0)+1}$	$y_{w\lceil 2(av+0)+1\rceil+1}$
	y_2	y_{w+2}	$y_{2w(av+0)+2}$	$y_{w\lceil 2(av+0)+1\rceil+2}$
	•	•	•	•
	y_w	y_{w+w}	$y_{2w(av+0)+w}$	$y_{w\lceil 2(av+0)+1\rceil+w}$
	$y_{2w(1)+1}$	$y_{w\lceil 2(1)+1\rceil+1}$	$y_{2w(av+1)+1}$	$y_{w\lceil 2(av+1)+1\rceil+1}$
	$y_{2w(1)+2}$	$y_{w\lceil 2(1)+1\rceil+2}$	$y_{2w(av+1)+2}$	$y_{w\lceil 2(av+1)+1\rceil+2}$
	•	•	•	•
	$y_{2w(1)+w}$	$y_{w\lceil 2(1)+1\rceil+w}$	$y_{2w(av+1)+w}$	$y_{w\lceil 2(av+1)+1\rceil+w}$
	•	•	•	•
	•	•	•	•
	•	•	•	•
	$y_{2w(v-1)+1}$	$y_{w\lceil 2(v-1)+1\rceil+1}$	$y_{2w\lceil av+(v-1)\rceil+1}$	$y_{w\lceil 2(av+(v-1))+1\rceil+1}$
	$y_{2w(v-1)+2}$	$y_{w\lceil 2(v-1)+1\rceil+2}$	$y_{2w\lceil av+(v-1)\rceil+2}$	$y_{w\lceil 2(av+(v-1))+1\rceil+2}$
	•	•	•	•
	$y_{2w(v-1)+w}$	$y_{w\lceil 2(v-1)+1\rceil+w}$	$y_{2w\lceil av+(v-1)\rceil+w}$	$y_{w\lceil 2(av+(v-1))+1\rceil+w}$
Sample mean	$\overline{y}_1$	$\overline{y}_2$	$\overline{y}_{(1+2a)}$	$\overline{y}_{(2+2a)}$

To better understand the selection of SSC is by using a numerical example from an ordered population, $y_1, y_2, y_3 ..., y_{47}, y_{48}$. For instance, let us consider $N = 48,\ m = 12$. The similar three steps were applied to generate the sample selection for SCC. The steps are as follow:

Step 1

The values of N, m and w must satisfy two requirements as depicted by these two equations $N/m = q$ and $m/w = v$, where $q \in 2\mathbb{N}$ and $v \in \mathbb{N}$. Since $N = 48$ and $m = 12$, so $N/m = 48/12 = q = 4$. This also satisfies one of the requirements as depicted by $q \in 2\mathbb{N}$ since $4/2 = 2$ (where N is a natural number). In the other hand, w is a common divisor for m. Since $m = 12$, there are 6 possible values for w (i.e., $w = 1, 2, 3, 4, 6, 12$) which also satisfy the other requirement, $v \in \mathbb{N}$. Say, if we let $w = 4$. In this case, we shall obtain 4 samples in each interval

selection and the sample will be taken at every 4^{th} selection. To satisfy $m/w = v$ or $m = wv$, there should be v = three times for sample selection while obtaining each of four samples, which shall therefore satisfy $m = 12 = 4(3) = wv$.

Step 2

The population will be divided into $q/2$ clusters, $c = (c_1, c_2, ..., c_{q/2})$. The sample size in each cluster shall be $2m$ and let $(y_1, y_2, ..., y_{2m})$ be the corresponding ordered values of the study variable y. Since $q = 4$, thus there are 2 clusters, and there will be 2 samples in each cluster. Again, since $N = 48$, therefore cluster 1 has a sample from 1 to 24 and cluster 2 has a sample from 25 to 48.

Step 3

In each of the two clusters, the first sample will be obtained followed by the next w^{th} sample $(y_1, y_2, ..., y_w)$ in a consecutive manner. Therefore, there will be a skipped of the w^{th} samples. Thus, $c = 2$ (2 clusters), $q = 4$ (4 sets of samples), $v = 3$ (3 times of selection) and $w = 4$ (4 samples will be taken and skipped consecutively). The arrangement of sample unit in the first cluster need to be completed first before arrange the sample unit for the subsequent cluster and these can be simplified and illustrated in Figure 1 and Table 3.

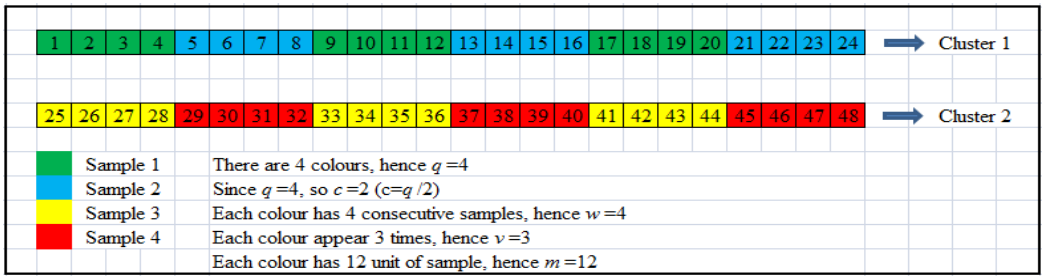


Figure 1 Illustration of selection of systematic sampling with consecutive approach when $N = 48$, $m = 12$, $q = 4$, $w = 4$ and $v = 3$

To determine the sample, a random number from q (either 1, 2, 3 or 4) can be determined to select the sample unit with $m = 12$. Say a random number of 2 is chosen, so sample unit of $y_5, y_6, y_7, y_8, y_{13}, y_{14}, y_{15}, y_{16}, y_{21}, y_{22}, y_{23}$ and y_{24} will be selected as the sample for the population.

Here, the sample selection by adopting this newly modified SSC is more flexible than that by adopting the conventional systematic sampling, especially in the selection of w. In this example, the simplest form of SSC is when $w = m$ since this will only require $v = 1$ which is the number of times for a batch of consecutive selection. The selection is presented in Table 4.

Table 3 Schematic diagram that showing one of the sets of SSC when $N = 48$ and $m = 12$

v	w	c_1		c_2	
		q_1	q_2	q_3	q_4
1	1	y_1	y_5	y_{25}	y_{29}
	2	y_2	y_6	y_{26}	y_{30}
	3	y_3	y_7	y_{27}	y_{31}
	4	y_4	y_8	y_{28}	y_{32}
2	1	y_9	y_{13}	y_{33}	y_{37}
	2	y_{10}	y_{14}	y_{34}	y_{38}
	3	y_{11}	y_{15}	y_{35}	y_{39}
	4	y_{12}	y_{16}	y_{36}	y_{40}
3	1	y_{17}	y_{21}	y_{41}	y_{45}
	2	y_{18}	y_{22}	y_{42}	y_{46}
	3	y_{19}	y_{23}	y_{43}	y_{47}
	4	y_{20}	y_{24}	y_{44}	y_{48}

Table 4 Schematic diagram that showing one of the sets of SSC when $N = 48$ and $w = m$ where both w and m are equal to 12

v	w	c_1		c_2	
		q_1	q_2	q_3	q_4
1	1	y_1	y_{13}	y_{25}	y_{37}
	2	y_2	y_{14}	y_{26}	y_{38}
	3	y_3	y_{15}	y_{27}	y_{39}
	4	y_4	y_{16}	y_{28}	y_{40}
	5	y_5	y_{17}	y_{29}	y_{41}
	6	y_6	y_{18}	y_{30}	y_{42}
	7	y_7	y_{19}	y_{31}	y_{43}
	8	y_8	y_{20}	y_{32}	y_{44}
	9	y_9	y_{21}	y_{33}	y_{45}
	10	y_{10}	y_{22}	y_{34}	y_{46}
	11	y_{11}	y_{23}	y_{35}	y_{47}
	12	y_{12}	y_{24}	y_{36}	y_{48}

To determine the sample, say a random number of 2 (out of sample 1, 2, 3 or 4) is chosen, so sample unit of $y_{13}, y_{14}, y_{15}, y_{16}, y_{17}, y_{18}, y_{19}, y_{20}, y_{21}, y_{22}, y_{23}$ and y_{24} will be selected as the sample for the population.

2. Unbiased Estimator of the Population Mean Based on SSC

When $N = nk$, the sample mean from the systematic sampling will be an unbiased estimator of the population mean. This occurs when systematic sampling technique is used. Similar to systematic sampling, the sample mean derived from the newly-modified SSC technique is also found to be unbiased estimators of the population mean when the terms and selection are complied. To illustrate that the sample mean is an unbiased estimator of the population mean, let

$$\overline{y}_i = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right),$$

be the sample mean for the q^{th} SSC sample. Then,

$$E(\overline{y}) = \frac{1}{q}\left(\overline{y}_1 + \overline{y}_2 + \cdots + \overline{y}_q\right). \tag{1}$$

Since there are only q possible samples that we can select, and the probability of selecting a particular sample by using SSC will be $1/q$. Based on our example, since we have $N = 48$, $m = 12$, $q = 4$, $w = 4$ and $v = 3$; therefore, by using the modified formula for mean, we find that:

$$\overline{y}_{1+2b} = \frac{1}{vw}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{2w(bv+l)+j}\right), \tag{2}$$

$$\overline{y}_{2+2b} = \frac{1}{vw}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{w\left[2(bv+l)+1\right]+j}\right), \tag{3}$$

where $b = 0,1,\ldots,\frac{q-2}{2}$; q is defined as the number of samples (e.g., sample 1, sample 2, etc.) of size m that can be obtained from the population where q is always an even number; v is defined as the number of batch of consecutive sample until all samples unit have been obtained (e.g., $v = 4$ indicates that the selection needs 4 batches until all the desired sample size is achieved); and w is defined as the total number of consecutive subjects that will be obtained and also the total number of subjects to be skipped (e.g., $w = 4$ indicates that 4 consecutive subjects unit will be taken, then skip another 4 subjects unit before select another 4 consecutive subjects. The selection of consecutive subjects continues until v times);

Equation (2) is applied to estimate mean estimator for sample with odd number such as q_1 and q_3. While equation (3) is applied to estimate mean estimator for sample with even number such as q_2 and q_4. Here, sample for SSC is label with symbol q. Hence, for $\overline{y}_1$, we will use (2) where $b = 0$,

$$\begin{aligned}
\overline{y}_1 &= \frac{1}{vw}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{2w(bv+l)+j}\right) = \frac{1}{3(4)}\left(\sum_{l=0}^{2}\sum_{j=1}^{4} y_{2(4)\left[0(3)+l\right]+j}\right) = \frac{1}{12}\left(\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j}\right), \\
&= \frac{1}{12}\left(\sum_{l=0}^{2}\left(y_{8l+1} + y_{8l+2} + y_{8l+3} + y_{8l+4}\right)\right), \\
&= \frac{1}{12}\left(y_1 + y_2 + y_3 + y_4 + y_9 + y_{10} + y_{11} + y_{12} + y_{17} + y_{18} + y_{19} + y_{20}\right).
\end{aligned}$$

Using equation (3) for $\overline{y}_2$ where $b = 0$,

$$\bar{y}_2 = \frac{1}{vw}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{w\lfloor 2(bv+l)+1\rfloor+j}\right) = \frac{1}{3(4)}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{4\lfloor 2(0(3)+l)+1\rfloor+j}\right) = \frac{1}{12}\left(\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j+4}\right),$$
$$= \frac{1}{12}\left(y_5 + y_6 + y_7 + y_8 + y_{13} + y_{14} + y_{15} + y_{16} + y_{21} + y_{22} + y_{23} + y_{24}\right).$$

Using equation (2) for $\bar{y}_3$ where $b = 1$,

$$\bar{y}_3 = \frac{1}{vw}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{2w(bv+l)+j)}\right) = \frac{1}{3(4)}\left(\sum_{l=0}^{2}\sum_{j=1}^{4} y_{2(4)\lfloor 1(3)+l\rfloor+j}\right) = \frac{1}{12}\left(\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j+24}\right),$$
$$= \frac{1}{12}\left(y_{25} + y_{26} + y_{27} + y_{28} + y_{33} + y_{34} + y_{35} + y_{36} + y_{41} + y_{42} + y_{43} + y_{44}\right).$$

Using equation (3) for $\bar{y}_4$ where $b = 1$,

$$\bar{y}_4 = \frac{1}{vw}\left(\sum_{l=0}^{v-1}\sum_{j=1}^{w} y_{w\lfloor 2(bv+l)+1\rfloor+j}\right) = \frac{1}{12}\left(\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j+28}\right),$$
$$= \frac{1}{12}\left(y_{29} + y_{30} + y_{31} + y_{32} + y_{37} + y_{38} + y_{39} + y_{40} + y_{45} + y_{46} + y_{47} + y_{48}\right).$$

So, the estimator of the population mean:

$$E(\bar{y}) = \frac{1}{q}\left(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_q\right) = \frac{1}{4}\left(\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4\right),$$
$$= \frac{1}{4}\left[\left(\frac{1}{12}\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j}\right) + \left(\frac{1}{12}\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j+4}\right) + \left(\frac{1}{12}\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j+24}\right) + \left(\frac{1}{12}\sum_{l=0}^{2}\sum_{j=1}^{4} y_{8l+j+28}\right)\right],$$
$$= \frac{1}{4}\left(\frac{1}{12}\right)\left[\sum_{l=0}^{2}\sum_{j=1}^{4}\left(y_{8l+j} + y_{8l+j+4} + y_{8l+j+24} + y_{8l+j+28}\right)\right] = \frac{1}{48}\left(y_1 + y_2 + \cdots + y_{48}\right),$$
$$= \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right) = \bar{Y}.$$

Hence, this illustrates that the sample mean of a SSC sample is an unbiased estimator of the population mean.

The evaluation of efficiency and consistency is now presented in this section. In previous section, only efficiency is evaluated based on random datasets. Now, both efficiency and consistency were evaluated by using simulated data which are then analysed by using R software. This R software is a statistical programming software which randomly selects a sample from the desired population by adopting both the new SSC technique and the SS technique.

The population is set in few distributions such as normal distribution with mean = 80 and various standard deviations (i.e., $\sigma = 5$, $\sigma = 15$ and $\sigma = 25$). Besides that, this study also assesses the efficiency within exponential distribution with $\lambda = 0.0125$ and Poisson distribution with $\lambda = 4$. The population parameters will then be computed repeatedly (i.e., approximately 1,000 times) in an iterative process. Various sub-samples (i.e., 25, 30, 50, 75 and 150) are tested from the simulation analysis and the mean and the standard deviation are calculated accordingly. The rule of selection for both SS and SSC is shown in Table 5.

A particular sampling technique is said to be more efficient when the estimator such as the mean which is derived from a particular sampling technique yields the least variance compared with another sampling technique. The relative efficiency of B to A is given by (Scheaffer et al. 2012),

$$\frac{V(A)}{V(B)},$$

or in other words, variance of the sample means derived from sampling technique A divided by the variance of the sample means derived from sampling technique B. If the relative efficiency is greater than 1, then $V(A)$ is larger than $V(B)$, which is favourable to sampling technique B. The findings obtained from analysis of simulation data are presented in Tables 6-8.

Table 5 Rules of selection for systematic sampling and systematic sampling with consecutive approach for the simulation analysis

n or m	SS		SSC	
	k	w	v	q
25	12	5	5	12
30	10	10	3	10
50	6	5	10	6
75	4	5	15	4
150	2	2	75	2

Note: SS refers to systematic sampling; SSC refers to systematic sampling with consecutive approach

Table 6 Measures of efficiency and consistency based on normal distribution derived from systematic sampling and systematic sampling with consecutive approach through simulation analysis with 1,000 iterations

Population ($N = 300$)	Sample Size (n)	$\hat{\mu}[V(\bar{y})]^{\#}$ from SS	$\hat{\mu}[V(\bar{y})]^{\#}$ from SSC	Summary results of relative efficiency
Normal Distribution $\mu = 80$ $\sigma = 5$	25	80.0228 (0.9221)	80.0209 (0.9107)	49.9%
	30	80.0117 (0.7349)	80.0348 (0.7284)	51.2%
	50	79.9965 (0.4222)	79.9920 (0.4085)	51.8%
	75	80.0042 (0.2439)	80.0009 (0.2460)	51.0%
	150	79.9934 (0.0798)	79.9960 (0.0826)	50.4%
	25	80.0685 (8.2989)	80.0626 (8.1967)	49.9%
	30	80.0350 (6.6138)	80.1044 (6.5552)	51.2%
	50	79.9895 (3.8002)	79.9760 (3.6765)	51.8%
	75	80.0126 (2.1949)	80.0027 (2.2142)	51.0%
	150	79.9802 (0.7183)	79.9880 (0.7527)	50.4%

Note: The summary results of relative efficiency were computed based on relative efficiency > 1 out of 1,000 iterations in terms of percentage that favours SSC, $\left[v(\bar{y})\right]^{\#}$ refers to average variance of the sample means for 1,000 iterations.

Table 6 (Continued)

Population ($N = 300$)	Sample Size (n)	$\hat{\mu}\left[V(\bar{y})\right]^{\#}$ from SS	$\hat{\mu}\left[V(\bar{y})\right]^{\#}$ from SSC	Summary results of relative efficiency
Normal Distribution $\mu = 80$ $\sigma = 15$	25	80.1141 (23.0524)	80.1043 (22.7686)	49.9%
	30	80.0583 (18.3717)	80.1740 (18.2088)	51.2%
	50	79.9826 (10.5561)	79.9600 (10.2126)	51.8%
	75	80.0210 (6.0969)	80.0045 (6.1506)	51.0%
	150	79.9671 (1.9953)	79.9800 (2.0909)	50.4%

Table 7 Measures of efficiency and consistency based on exponential and Poisson distributions derived from systematic sampling and systematic sampling with consecutive approach through simulation analysis with 1000 iterations

Population ($N = 300$)	Sample Size (n)	$\hat{\mu}\left[V(\bar{y})\right]^{\#}$ from SS	$\hat{\mu}\left[V(\bar{y})\right]^{\#}$ from SSC	Summary results of relative efficiency
Exponential Distribution $\lambda = 0.0125$	25	81.1166 (235.6960)	79.3009 (227.9922)	53.0%
	30	79.3659 (197.9049)	79.5679 (185.0346)	53.8%
	50	79.8332 (107.9407)	80.4560 (103.0512)	52.4%
	75	79.7323 (65.7368)	80.1133 (61.9246)	52.3%
	150	80.1094 (22.0447)	80.1200 (21.4917)	51.8%
Poisson Distribution $\lambda = 4$	25	4.0110 (0.1491)	4.0086 (0.1455)	51.1%
	30	4.0153 (0.1169)	4.0077 (0.1180)	47.7%
	50	4.0153 (0.0669)	4.0112 (0.0651)	52.8%
	75	4.0029 (0.0401)	3.9943 (0.0384)	50.8%

	150	4.0044 (0.0125)	4.0072 (0.0137)	48.1%

Note: The summary results of relative efficiency were computed based on relative efficiency > 1 out of 1,000 iterations in terms of percentage that favours SSC, $\left[v(\bar{y})\right]^{\#}$ refers to average variance of the sample means for 1,000 iterations

Table 7 (Continued)

Population ($N = 300$)	Sample Size (n)	$\hat{\mu}\left[V(\bar{y})\right]^{\#}$ from SS	$\hat{\mu}\left[V(\bar{y})\right]^{\#}$ from SSC	Summary results of relative efficiency
Exponential Distribution $\lambda = 0.0125$	25	81.1166 (235.6960)	79.3009 (227.9922)	53.0%
	30	79.3659 (197.9049)	79.5679 (185.0346)	53.8%
	50	79.8332 (107.9407)	80.4560 (103.0512)	52.4%
	75	79.7323 (65.7368)	80.1133 (61.9246)	52.3%
	150	80.1094 (22.0447)	80.1200 (21.4917)	51.8%
Poisson Distribution $\lambda = 4$	25	4.0110 (0.1491)	4.0086 (0.1455)	51.1%
	30	4.0153 (0.1169)	4.0077 (0.1180)	47.7%
	50	4.0153 (0.0669)	4.0112 (0.0651)	52.8%
	75	4.0029 (0.0401)	3.9943 (0.0384)	50.8%
	150	4.0044 (0.0125)	4.0072 (0.0137)	48.1%

There are no obvious differences in the parameter estimates (mean) between the estimates derived from SS and SSC. This finding supports evidence of consistency of SSC. The estimates from either normal distribution, exponential distribution or Poisson distribution provide consistent measure when sample size increases. Most importantly, the parameter estimates were found to be very close approximations of the true parameters. In general, SSC is more efficient compare with SS. Based on normal distribution, SSC has relative efficiency more than one with 12 over 15 simulation analysis. While exponential distribution provides SSC has relative efficiency more than one with 5 over 5 simulation analysis and Poisson distribution provides 3 over 5 with relative efficiency more than one for SSC. Based on all these findings, it is concluded that the estimators derived from SSC are at least equally efficient and consistent as those estimators derived from SS. Based on another indicator such as error or bias which is measured using mean square error (MSE), there are no obvious different in terms of bias between SS and SSC (Table 9).

Table 8 Comparison of mean square error between systematic sampling and systematic sampling with consecutive approach based on normal distribution and different sample sizes

Population ($N = 300$)	Sample Size (n)	MSE from SS	MSE from SSC
Normal Distribution	25	0.880	0.855
$\mu = 80, \sigma = 5$	30	0.743	0.718
	50	0.418	0.440
	75	0.253	0.255
	150	0.080	0.084
Normal Distribution	25	7.921	7.699
$\mu = 80, \sigma = 15$	30	6.687	6.463
	50	3.758	3.957
	75	2.275	2.291
	150	0.718	0.753
Normal Distribution	25	22.003	21.385
$\mu = 80, \sigma = 25$	30	18.576	17.954
	50	10.440	10.991
	75	6.319	6.364
	150	1.995	2.091

Table 9 Comparison of mean square error between systematic sampling and systematic sampling with consecutive approach different distributions (exponential and Poisson) and different sample sizes

Population ($N = 300$)	Sample Size (n)	MSE from SS	MSE from SSC
Exponential Distribution	25	243.044	218.255
$\lambda = 0.0125$	30	209.752	185.121
	50	112.263	101.255
	75	59.563	61.751
	150	22.045	21.492
Poisson Distribution	25	0.151	0.145
$\lambda = 4$	30	0.123	0.117
	50	0.069	0.071
	75	0.040	0.041
	150	0.012	0.014

3. Sampling Efficiency between SSC and SS using Real-Life Dataset

The subsequent evaluation of the sampling efficiency was conducted based on a real-life datasets and the data was derived from handbook on sampling methods. The datasets are described as in Table 10. Descriptive analysis for the two datasets were presented in Table 11, sample selection for the datasets were presented in Tables 12 and 13, while sampling efficiencies were evaluated and presented in Table 14. The mean (SD) parameter for both “area under grove crop” and “clean content of wool” are 54.9 (53.8) and 57.7 (1.8), respectively. The distribution of both data are approximately

normal. Data for area under grove crop is slightly skewed with the value of skewness of 1.067 and -0.341 for "clean content of wool". Based on the two real-life datasets, result showed that SSC has almost 80 times efficient more than SS based on data from "area under grove crop" and SS has slight advantage based on data from "clean content of wool". This could be explained by huge differences of standard deviations from the two datasets where area under grove crop has larger standard deviation as compared with clean content of wool. However, the aim of the analysis is not to prove SSC is always better than SS. The main points of the whole discussion are both SS and SSC are unbiased and result of the efficiencies are almost comparable as proven by the findings from simulation analysis.

Table 10 Data for area under grove crop $(N = 16)$ and clean content of wool $(N = 24)$

No.	Area under grove crop	Clean content of wool
1	166.15	54.3
2	24.73	56.2
3	100.77	58.9
4	10.52	55.5
5	116.28	57.0
6	12.92	58.7
7	41.70	58.2
8	13.59	57.4
9	87.14	54.6
10	13.85	57.5
11	60.22	59.3
12	10.73	57.5
13	38.64	54.9
14	15.92	60.1
15	155.51	58.7
16	9.09	55.6
17		59.9
18		57.8
19		60.9
20		57.5
21		57.8
22		59.7
23		59.6
24		58.1

Source: Kumar P. Handbook on Sampling Methods. New Delhi, IASRI

Table 11 The descriptive parameter of area under grove crop $(N = 16)$ and clean content of wool $(N = 24)$

Variable	Parameter	Statistics
Area under grove crop	Mean	54.860
	Median	31.685
	Variance	2898.637
	Std. Deviation	53.839
	Minimum	9.090
	Maximum	166.150
	Skewness	1.067
	Kurtosis	–0.153
Clean content of wool	Mean	57.738
	Median	57.800
	Variance	3.307
	Std. Deviation	1.818
	Minimum	54.300
	Maximum	60.900
	Skewness	–0.341
	Kurtosis	–0.633

Table 12 Sample selection $(n = 8)$ from systematic sampling and systematic sampling with consecutive approach for area under grove crop

No.	SS ($k = 2$)		SSC ($q = 2,\ w = 2$ and $v = 4$)	
	S_1	S_2	S_1	S_2
1	166.15	24.73	166.15	100.77
2	100.77	10.52	24.73	10.52
3	116.28	12.92	116.28	41.70
4	41.70	13.59	12.92	13.59
5	87.14	13.85	87.14	60.22
6	60.22	10.73	13.85	10.73
7	38.64	15.92	38.64	155.51
8	155.51	9.09	15.92	9.09

Note: SS refers to systematic sampling; SSC refers to systematic sampling with consecutive approach

Table 13 Sample selection $(n = 12)$ from systematic sampling and systematic sampling with consecutive approach for clean content of wool

No.	SS ($k = 2$)		SSC ($q = 2,\ w = 2$ and $v = 6$)	
	S_1	S_2	S_1	S_2
1	54.3	56.2	54.3	58.9
2	58.9	53.5	56.2	53.5
3	57.0	58.7	57.0	58.2
4	58.2	57.4	58.7	57.4
5	54.6	57.5	54.6	59.3
6	59.3	57.5	57.5	57.5
7	54.9	60.1	54.9	58.7
8	58.7	55.6	60.1	55.6
9	59.9	57.8	59.9	60.9
10	60.9	57.5	57.8	57.5
11	57.8	59.7	57.8	59.6
12	59.6	58.1	59.7	58.1

Note: SS refers to systematic sampling; SSC refers to systematic sampling with consecutive approach

Table 14 Relative efficiency from systematic sampling and systematic sampling with consecutive approach for population from area under grove crop and clean content of wool

Statistics	Area under grove crop		Clean content of wool	
	SS	SSC	SS	SSC
$E\left(\bar{y}\right)$	54.9	54.9	57.7	57.7
$V\left(\bar{y}\right)$	1676.2	21.2	0.04	0.08
Efficiency	79.4		0.5	

Note: SS refers to systematic sampling; SSC refers to systematic sampling with consecutive approach

4. Discussions

The array of modified systematic sampling methods has captured interest among statisticians after its early development in the 1940s. In spite of these modifications, a strong similarity still exists with the previous modified systematic sampling methods whereby the selection process will obtain only one-unit subject in each interval selection.

This study has introduced a newly modified systematic sampling which is different from all the previously modified systematic sampling methods. This difference stems from the uniqueness of the SSC, which allows more than one-unit subject to be recruited consecutively in each interval selection. Besides achieving an unbiased estimator, this study is also presenting an initial evaluation of sampling efficiency between SSC and SS based on a simulation analysis and published datasets. Based on the evaluations, most findings especially result in simulation analysis have shown that, SSC is better than SS in terms of sampling efficiency.

This comparison based on published dataset were made only based on two datasets. We postulate that when more real-life datasets are tested, the sampling efficiency between SSC and SS will be

likely to be equally efficient. Ideally, the idea of developing this newly modified SSC is not merely for it to become a substitute for conventional systematic sampling. However, it is an improvised alternative for probability sampling where it offers greater levels of convenience and flexibility during recruitment of samples. The most important finding is SSC is able to obtain an unbiased estimator of the population mean.

Similar to SS, SSC is not suitable to be adopted if there is a unique consecutive pattern in the population. Therefore, researchers will need to confirm that the population from which the sample will be taken does not have a unique consecutive pattern. If there is a unique consecutive pattern of sample selection, then it can lead to bias for a particular characteristic of subjects or respondents. A simple analysis such as "Run test" can be used to examine the randomness of the data sequence (Bujang and Sapri 2018). Despite the above recent progress, the variance of the sample mean for SSC is still remains biased. This situation is the same with SS since it was developed in early 1940s. Therefore, many research works have been done to minimize the variance of the sample mean (Zinger 1980, Yves 2005).

Besides that, one of limitations in SS is the expect of inefficiency especially for data distributed with linear or parabolic distributions (Leu and Kao 2006). Since SSC is the mixture between consecutive sampling and SS, therefore SSC may have same limitation. However, the SSC is a newly modified systematic sampling and indeed an innovation in sampling by the virtue of combining two sampling techniques, consecutive sampling and systematic sampling. the beauty of SSC is besides able to produce an unbiased estimator for the mean population, the concept in recruiting the subject is more user friendly as compared with the SS.

It is important to take note that SSC as a form of probability sampling will be superior to consecutive sampling because it will eliminate bias in the selection. The main advantage of SSC is that it enables researcher to recruit more than one-unit subject during each interval selection. This will increase the speed for recruitment of samples. Hence, SSC has the advantage in terms of a probability sampling and also has the element of convenience in the selection.

There is an impending need of a thorough review of SSC's potential applications in survey for an ordered population. In this study, only two real-life datasets were tested with the aim to introduce and familiarize the audience with the concept of SSC. Hence for future studies, we recommended SSC to be evaluated with more real-life datasets especially for larger sample size. Besides that, researchers may also be interested to compare sampling efficiency of SSC with various sampling methods. Since SSC is a new sampling method, therefore it is timely that greater promotion needs to be exercised to encourage researchers to apply SSC in survey research.

5. Conclusions

This study has successfully developed the terms and condition for selection of the newly modified systematic sampling and most importantly, the estimator that is derived from this sampling technique has been proved to be unbiased and efficient. This newly modified systematic sampling is named as "Systematic Sampling with Consecutive Approach" or in short SSC in indeed an innovation in sampling by the virtue of combining two sampling techniques, consecutive sampling and systematic sampling.

References

Bujang MA, Ghani PA, Zolkepali NA, Adnan TH, Ali MM, Selvarajah S, Haniff J. A comparison between convenience sampling versus systematic sampling in getting the true parameter in a population: explore from a clinical database: the Audit Diabetes Control Management (ADCM)

registry in 2009. ICSSBE 2012: Proceedings of 2012 International Conference on Statistics in Science, Business and Engineering; 2012 Sep 10-12; Langkawi, Malaysia. IEEE; 2012. pp. 1-5.

Bujang MA, Sapri FE. An application of the runs test to test for randomness of observations obtained from a clinical survey in an ordered population. Malays J Med Sci. 2018; 25(4): 146-151.

Leu CH, Kao FF. Modified balanced circular systematic sampling. Stat Prob Lett. 2006; 76(4): 373-383.

Cochran WG. Relative accuracy of systematic and stratified random samples for a certain class of populations. Ann Math Stat. 1946; 17(2): 164-177.

Hartley HO, Rao JNK. Sampling with unequal probabilities and without replacement. Ann Math Stat. 1962; 33(2): 350-374.

Khan L, Shabbir J, Khalil U. A new systematic ranked set-sampling scheme for symmetric distributions. Life Cycle Reliab Saf Eng. 2019; 8(8): 1-6.

Khan Z, Shabbir J, Gupta S. A new sampling design for systematic sampling. Commun Stat Theory Methods. 2013; 42(18): 3359-3370.

Madow WG, Madow LH. On the theory of systematic sampling, I. Ann Math Stat. 1944; 15(1): 1-24.

Mclntyre G. A method for unbiased selective sampling, using ranked sets. Aust J Agric Res. 1952; 3(4): 385-390.

Shiue CJ. Systematic sampling with multiple random starts. Forest Sci. 1960; 6(1): 42-50.

Singh D, Jindal KK, Garg JN. On modified systematic sampling. Biometrika. 1968; 55(3): 541-546.

Singh D, Singh P. New systematic sampling. J Stat Plan Inference. 1977; 1(2), 163-177.

Sampath S, Varalakshmi V. Diagonal circular systematic sampling. Model Assist Stat Appl. 2008; 3(4): 345-352.

Scheaffer RC, Mendenhall W, Ott RC, Gerow K. Elementary survey sampling. Boston, USA: Brooks/Cole; 2012.

Subramani J. Diagonal systematic sampling scheme for finite population. J Indian Soc Agric Stat. 2000; 53(2): 187-195.

Subramani J. Generalization of diagonal systematic sampling scheme for finite population. Model Assist Stat Appl. 2010; 5(2): 117-128.

Yates F. Systematic Sampling. Philos Trans Phys Sci Eng. 1948; 241(834): 345-377.

Yves GB. A Variance estimator for systematic sampling from a deliberately ordered population. Commun Stat Theory Methods. 2005; 34(7): 1533-1541.

Uthayakumaran N. Additional circular systematic sampling methods. Biom J. 1998; 40(4): 467-474.

Zinger A. Variance estimation in partially systematic sampling. J Am Stat Assoc. 1980; 75(369): 206-211.