1.Introduction

Welcome to the series of E-learning modules on Systematic Sampling: Advantages and Limitations. In this module, we are going to cover the basic principle of systematic sampling, types of the technique, advantages and limitations of systematic sampling.

By the end of this session, you will be able to:

- Explain systematic sampling
- Explain the principal stages of the method
- Explain various types of the technique
- Explain the advantages and limitations of systematic sampling

Systematic sampling is a probability sampling procedure, in which first element of the sample is randomly selected and then the subsequent elements are selected using a fixed or systematic interval until the desired sample size is reached.
Systematic sampling is also called as interval random sampling.

Systematic sampling is a technique, which has a nice feature of selecting a whole sample with just one random start.
Systematic random sampling is a technique in which first unit is selected with a help of random numbers and the others are automatically selected according to some pre-designed pattern until the desired sample size is reached.

Systematic sampling is often used instead of random sampling. It is also called as an n^{th} name selection technique. After the required sample size has been calculated, every n^{th} record is selected from a list of population members. As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method.

Only advantage of Systematic sampling over the random sampling technique is simplicity. Systematic sampling is frequently used to select a specified number of records from a computer file. Systematic sampling is a modification of random sampling. To arrive at a systematic sample, we simply calculate the desired sampling fraction.

For example, if there are hundred distributors of a particular product in which we are interested and our budget allows us to sample say twenty of them, then we divide hundred by twenty and get the sampling fraction as five.

Thereafter, we go through our sampling frame selecting every fifth distributor. In the purest sense, this does not give rise to a true random sample since some systematic arrangement is used in listing and not every distributor has a chance of being selected once the sampling fraction is calculated.

Generally, systematic sampling is easier, simpler, less time-consuming, and more economical than simple random sampling. If the ordering is unrelated to the study variables, but randomized, systematic sampling will yield results similar to simple random sampling. On the other hand, periodicity in the sampling frame is a constant concern in

systematic sampling.

The random start distinguishes this sampling procedure from its non-probability counterpart, non-probability systematic sampling.

In some instances, a sampling frame is not used. The target population need not be numbered and a sampling frame is compiled, if there is a physical presentation such as a continuous flow of population elements at specific locations.

For example,

- After a random start, one may systematically select every i^{th} patient visiting an emergency room in a hospital
- Store customers standing in line
- Records in file drawers

2. Steps Involved in Selecting a Systematic Sample

What are the steps involved in selecting a systematic sample?
Generally, there are eight major steps involved in selecting a systematic sample:

1. Define the target population.
2. Determine the desired sample size (n).
3. Identify an existing sampling frame or develop a sampling frame of the target population.
4. Evaluate the sampling frame for under coverage, over coverage, multiple coverage, clustering, and periodicity, and make adjustments wherever necessary.
   Ideally, the list will be in a random order with respect to study variable or ordered in terms of the variable of interest or its correlate, thereby creating implicit stratification. If the sampling frame is randomized, systematic sampling is considered to be a good approximation of simple random sampling.
5. Determine the number of elements in the sampling frame (N).
6. Calculate the sampling interval (k) by dividing the number of elements in the sampling frame (N) by the targeted sample size (n).
   One should ignore the remainder and round down or truncate to the nearest whole number. Rounding down and truncating may cause the sample size to be larger than desired. If so, one may randomly delete the extra selections. If the exact size of the population is not known and it is impractical to determine, then one may fix the sampling fraction.
7. Randomly select a number r from “1” through k.
8. Select for the sample, r, r plus k, r plus 2k, r plus 3k, and so forth, until the frame is exhausted.
   At a technical level, systematic sampling does not create a true random sample. It is often referred to as “pseudo random sampling” or “quasi-random sampling.”

Only the selection of the first element in systematic sampling is a probability selection. Once the first element is selected, some of the elements will have a zero probability of selection. Moreover, certain combinations of elements, such as elements that are adjacent to each other in the sampling frame are not likely to be selected.

The Subtypes of Systematic Sampling:
Systematic sampling may be classified into three major types.

1. Linear Systematic Sampling:

Suppose the population is linear in order, such that the units can be referred to by number. Further, let N be expressible in the form of N is equal to n into k and let the selected random number be r (less than k), k being called as a sampling interval.

In this procedure, a sample comprises of the unit r, r plus k, r plus 2k up to r plus (n minus 1) k. This technique will generate k systematic samples with equal probability, which is equal to 1 by k. Under systematic sampling, let yri be the ith unit corresponding to the rth sample, where r is equal to 1, 2 up to k and i is equal to 1 up to n.

In another practical situation, N is not expressible in the form of N is equal to nk. In this case, the present sampling scheme will give rise to the samples of unequal size and k is taken as an integer nearest to N by n. Then, the random number is chosen from 1 to k and every k^{th} unit is drawn in the sample.

Under this case, sample size is not necessarily n and in some cases, it may be n minus 1. To overcome the difficulty of varying sample size under N is equal to nk, circular systematic sampling procedure is adopted.

2. Circular Systematic Sampling:

This scheme is slightly modified to get a sample of constant size. The procedure consists of selecting a unit by a random start, from 1 to N. Thereafter, selecting every k^{th} unit, for k being an integer nearest to N by n in a circular manner until a sample of size n units is obtained.

This technique is generally known as circular systematic sampling.

In this method, it can be easily verified that every unit has equal probability of selection 1 by N. As an illustration N is equal to 11, n is equal to 4 and k is equal to 3.

Therefore, the possible samples are: (1, 4, 7, 10), (2, 5, 8, 11), (3, 6, 9, 1), (4, 7, 10, 2), (5, 8, 11, 3), (6, 9, 4), (7, 10, 2, 5), (8, 11, 3, 6), (9, 1 4, 7), (10, 2, 5, 8), (11, 3, 6, 9).

3. Subtypes of Systematic Sampling-Repeated Systematic Sampling

3. Repeated (or replicated) Systematic Sampling:
Repeated systematic sampling involves the selection of multiple samples from the target population and then combining them into a single sample. Instead of only one random start, several smaller systematic samples are selected using multiple random starts. This makes the process more time-consuming compared to linear systematic sampling.

However, repeated sampling minimizes the effect of bias due to periodicity, which is a regularly occurring pattern in the sampling frame. Moreover, linear systematic sampling generates only one “cluster” of elements (although the cluster may contain multiple elements). Technically, an unbiased estimate of sampling error cannot be obtained without making certain assumptions.Linear systematic sampling is the most frequently used form of systematic sampling.

The steps in selecting a linear systematic sample are those listed above. Linear systematic sampling and circular systematic sampling create a single sample.

Circular systematic sampling may be viewed as a subtype of linear systematic sampling. In using this procedure, in Step 7, instead of selecting a random number between “1” and “k,” the size of the interval, a random number is selected between “1” and “N”.

When one gets to the end of the list in selecting the sample, one would continue from the beginning of the list. This creates a circular pattern in selecting the sample.

Systematic sampling is a method of selecting sample members from a larger population according to a random starting point and a fixed periodic interval.
Typically, every "n^{th}" member is selected from the total population for inclusion in the sample population. Systematic sampling is still considered as being random, as long as the periodic interval is determined beforehand and the starting point is random.

A common way of selecting members for a sample population using systematic sampling is simply to divide the total number of units in the general population by the desired number of units for the sample population. The result of the division serves as the marker for selecting sample units from within the general population.

For example, if you want to select a random group of one thousand people from a population of fifty thousand using systematic sampling, you can simply select every 50th person. This is because fifty thousand by one thousand is equal to 50.

Systematic Sampling involves selecting items using a constant interval between selections, the first interval having a random start. The interval might be based on a certain number of items (for example every twentieth voucher) or monetary total.
When using systematic selection, the auditor should determine that the population is not structured in such a manner that the sampling interval corresponds with a particular pattern in the population.

For example, in a population of branch sales, particular branch sales occur only as every hundredth item and the sampling interval selected is hundred. The result would be that the auditor would have selected either all or none of the sales of that particular branch.

To minimize the effect of the possible known buyers through a pattern in the population, more than one starting point may be taken. The multiple random starting points are taken because it minimizes the risk of interval sampling pattern with that of the population being sampled.
The process of obtaining the systematic sample is much like an arithmetic progression.

Starting number:

The researcher selects an integer that must be less than the total number of individuals in the population. This integer will correspond to the first subject. The number r is known as the random start. If r is a random start, then a systematic sample of size n consists of the units r, r plus k, r plus 2k, r plus 3k, up to r plus (n minus 1)k.
Interval**:**
The researcher picks another integer, which will serve as the constant difference between any two consecutive numbers in the progression.
The integer is selected, so that the researcher obtains the correct sample size.

4. Advantages of Systematic Sampling

Let us discuss the advantages and limitations of systematic sampling.
Systematic sampling has the strengths and weaknesses associated with most probability sampling procedures when compared to non-probability sampling Procedures.
In highlighting the strengths and weaknesses of systematic sampling, we may compare it to simple random sampling. Systematic sampling is often used when it is impractical or impossible to use simple random sampling.
When compared to simple random sampling, in some instances, it is a stronger sampling procedure, and in other instances, it is a weaker sampling procedure. Compared to simple random sampling, the strengths of systematic sampling include the following essentials:

- If the selection process is manual, systematic sampling is easier, simpler, less time-consuming, and more economical than simple random sampling. One needs to use a random process to select only the first element.
  On the other hand, if the selection process is computerized, the ease in the selection process of systematic sampling and simple random sampling may be comparable to each other.

- If the sampling frame has a monotonic ordering that is related to a study variable (e.g. ordering of stores by dollar value, listing of employees by number of years employed, and listings of schools by graduation rates), implicit stratification may result in the statistical efficiency. This is equivalent to that of proportionate stratified sampling and is thereby more efficient than simple random sampling. If the ordering is randomized, systematic sampling may yield results similar to simple random sampling.

- Systematic sampling ensures that the sample is more spread across the population
- An advantage of systematic random sampling over simple random sampling is the assurance that the population will be evenly sampled. There exists a chance in simple random sampling that allows a clustered selection of subjects. This is systematically eliminated in systematic sampling

Unlike simple random sampling, if the selection process is manual, systematic sampling is easier, simpler, less time-consuming, and more economical. If the sampling interval is related to periodic ordering of the elements in the sampling frame, increased variability may result.

The target population need not be numbered and a sampling frame compiled if there is physical representation. Combinations of elements have different probabilities of being selected.
If the ordering of the elements in the sampling frame is randomized, systematic sampling may yield results similar to simple random sampling.

Technically, only the selection of the first element is a probability selection since for subsequent selections, there will be elements of the target population that will have a zero chance of being selected.

If the ordering of the elements in the sampling frame is related to a study variable creating implicit stratification, systematic sampling is more efficient than simple random sampling. Principle of independence is violated, for the selection of the first element determines the selection of all others. Systematic sampling eliminates the possibility of autocorrelation.

Estimating variances is more complex than simple random sampling.
Systematic sampling ensures that the sample is spread across the population. Similarity of adjacent elements in a list makes for autocorrelation, the correlation among elements in the population. Although rare, this may occur in simple random sampling. Spatial autocorrelation is likely to exist in a listing of addresses.

People, who live at addresses that are close to each other are likely to be more similar to each other, say in terms of socioeconomic status, than they are to people living at addresses that are not as close.
A positive autocorrelation creates lower precision, and a negative autocorrelation creates higher precision when compared to simple random sampling. However, systematic sampling eliminates the possibility of autocorrelation.

For example, in using a voter's list for the selection of a sample for a study of voter preferences, it is possible that members of the same family are selected using simple random sampling, but this is not possible using systematic sampling.

5. Limitations of Systematic Sampling

Limitations:

Compared to simple random sampling, systematic sampling has more number of weaknesses. Some of them include:

- Periodicity in the sampling frame is a constant concern in systematic sampling. A biased sample could result if a periodic or cyclical pattern in the sampling frame corresponds to the sampling fraction.

This problem will exist, if the sampling fraction is equal to or a multiple of a periodic interval in the list.

For example, a systematic sample of students would be biased if students were listed by class and within each class ranked by performance on an achievement test. If the classes have approximately the same number of students, periodic bias will result.

- In simple random sampling, every combination of n elements has an equal chance of being selected. However, this is not the same in case of systematic sampling.
- Only the first element is a probability selection and the rest is predetermined. Consequently, there are members in the population that will have a zero chance of being selected.
- As the first element is completely at random, one can observe the violation of independence as it determines the selection of all the others.
- Process of estimation of variances is more complicated than simple random sampling.
- Systematic sampling may not represent the whole population.
- There is a chance of personal bias of the investigators.
- The process of selection can interact with a hidden periodic trait within the population. If the sampling technique coincides with the periodicity of the trait, the sampling technique will no longer be random and representativeness of the sample is compromised.
- The procedure involved in systematic random sampling is very easy and can be done manually.
- The results are representative of the population unless certain characteristics of the population are repeated for every nth individual, which is highly unlikely to happen.

Hence, Systematic Sampling is a method of selecting sample observations from a population of large size with just one random number, which is known as a random start and a fixed periodic interval, known as sampling interval.However, systematic sampling can be considered as a random sampling technique as long as sampling interval is determined before hand and a random start is a random number.

Here's a summary of our learning in this session, where we have understood:

- The concepts of systematic sampling

- The principal stages of the technique
- The types of systematic sampling
- The addressed advantages and limitations of the systematic sampling