Introduction to Systematic Sampling, Cluster Sampling and Two Stage Sampling

Systematic Sampling:

Systematic sampling is a commonly used technique if the complete and up-to-date list of sampling units is available. This sampling technique is more convenient than the simple random sampling. It also ensures that each unit has equal probability of inclusion in the sample. This consists in selecting only the first unit at random, the rest being automatically selected according to some predetermined pattern involving regular spacing units.

Method of selecting Systematic Sample:

Suppose the N units in the population are numbered 1 to N in some order. Suppose further that N is expressible as a product of two integers n and k such that N = nk.

To select a sample of 'n' units, we take an unit at random from the first k units. Suppose i^{th} ($1 \leq i \leq k$) unit , then select every k^{th} unit thereafter. So the sample will contain i, i+k, i+2k,…….,i+(n-1)k serial number units. So first unit is selected at random and other units are selected systematically. Here k is termed as **sampling interval**. This systematic technique is also known as **linear systematic sampling.**

Note: When the population size is N = nk, then k systematic samples each of size n can be drawn. For example: if N = 40 and k=8 , n=5 and 7^{th} unit is drawn first, then the subsequent units in the sample corresponds to the population units 15, 23, 31,39 and so the systematic sample contains units numbered 7,15,23,31,39.

Merits and demerits of Systematic Sampling:

Merits:

(1) It is easier to draw a sample and often easier to execute it without mistakes. This is more convenient than Simple Random Sampling and Stratified Sampling. Time and work involved is also relatively less.
(2) The cost is low and the selection of units is simple. Much less training is needed for surveyors to collect units through systematic sampling.
(3) The systematic sampling is efficient if the frame (list of population units) is arranged completely at random.

Demerits:

(1) If N ǂ nk, then it is difficult to draw a systematic sample. The sample size may not be n.
(2) Systematic sample may yield biased estimate if the population frame is not arranged randomly.
(3) If periodic variations are observed in population units, then k is equal to or a multiple of a period then the estimate will be highly biased.
(4) The systematic sample is spread more evenly over the population. So no large part will fail to be represented in the sample. The sample is evenly spread and cross section is better.
(5) Systematic sampling fails in case of too many blanks.

Note: To overcome demerit 1 Circular Systematic Sampling can be used. Circular systematic sampling is introduced by Indian Statistician Prof. D. B. Lahiri.

Applications of systematic sampling:

Systematic sampling is to be applied only if the given population is logically homogeneous, because systematic sample units are uniformly distributed over the population.

This method is especially applicable when the population to be studied is arranged in time. For example, a supermarket , which wishes to investigate customer attitudes, may interview each 3rd or 10th customer arriving to the shop.

The same principle is applied to many phenomenon. For example: the temperatures of Mumbai are registered each hour in order to determine average temperature, etc.

Cluster Sampling:

The method of cluster sampling or area sampling is used when a list of elements of the population is not available and so the use of an element is not feasible.

In cluster sampling we split the population into groups (clusters), randomly choose a sample of clusters, then we measure each individual from each selected cluster.

Hence, cluster sampling can be interpreted as random sampling of groups. In this case too we have the choice of selecting the clusters with or without replacement. Cluster sampling usually leads to a lower precision than simple random sampling. Cluster sampling is therefore applied only when required by the practical situation, or when the loss of precision is compensated by a substantial reduction in the data collection costs.

Conditions for using cluster sampling:

Cluster sampling is used when:

(i) No reliable listing of elements is available and it is expensive to prepare it.

(ii) Even if the list of elements is available, the location or identification of the units may be difficult.

(iii) A necessary condition for the validity of this procedure is that every unit of the population under study must correspond to one and only one unit of the cluster so that the total number of sampling units in the frame may cover all the units of the population under study without any omission or duplication. When this condition is not satisfied, bias is introduced.

For example: the list of all agricultural farms in a village or a district may not be easily available but the list of village or districts are generally available. In this case, every farm is a sampling unit and every village or a district is a cluster.

Advantages:

(1) **Consumes less time and cost:** Sampling of geographically divided groups require less work, time and cost. It is highly economical method to observe clusters instead of randomly doing it throughout a particular region by allocating a limited number of resources to those selected clusters.

(2) **Convenient access:** Large samples can be chosen with this sampling technique and that will increase accessibility to various clusters.

(3) **Least loss in accuracy of data:** Since there can be large samples in each cluster, loss of accuracy in information per individual can be compensated.

(4) **Ease of implementation:** Since cluster sampling facilitates information from various areas and groups, it can be easily implemented in practical situations in comparison to other probability sampling methods such as simple random sampling, systematic sampling and stratified sampling or non-probability sampling methods such as convenience sampling.

Note: In comparison to simple random sampling, cluster sampling can be effective in deciding the characteristics of a group such as population and it can also be implemented without having a sampling frame for all the elements for the entire population.

Cluster sampling is **more economical and more practical** than stratified sampling or simple random sampling.

Applications of Cluster Sampling:

(i) This method is used in statistics when natural groups are present in a population. The whole population is subdivided into clusters or groups and random samples are the selected from each group.

(ii) This sampling technique is used in an area or geographical cluster sampling for market research.

(iii) This method of sampling is used in situations like wars and natural calamities.

(iv) This method is typically used in market research. It is used when a researcher cannot get information

about the population as a whole, but they can get information about the clusters.

Two-Stage Sampling:

When the clusters in the cluster sampling method are homogeneous then this method is not very efficient. If the elements within a cluster very closely resemble each other, it may actually be a waste of time and money to investigate all elements. The same information could have been obtained by examining just a couple of elements. In such cases cluster sampling will perform fairly poorly, and it may be worthwhile to take a sample from the selected clusters. This is referred to as **two-stage sampling.** The first stage of two-stage sampling involves selecting clusters or primary sampling units at random. The second stage of two-stage sampling involves randomly selecting elements which are now referred to as secondary sampling units, from each selected primary sampling unit.

Assuming that the clusters are homogeneous, two-stage sampling generally allows more clusters to be selected in the first sampling stage which may increase precision. This method gives greater control of sample size where the sizes of primary sample units vary greatly. In this case too a sampling frame is required only for the selected primary sample units. The costs and precision of two-stage sampling fall roughly between those of cluster sampling and stratified sampling. Two-stage sampling is generally

more expensive than cluster sampling with the same sample size, but less expensive than stratified sampling. On the other hand, two-stage sampling is generally more precise than cluster sampling and less precise than stratified sampling. For example: in a crop survey, villages are the first stage units, fields within the villages are second stage units.

Two stage sampling selection procedure:

Let the population consists of N groups such that the ith group contains Mi units. Here N are number of first stage units. All the sample of n first stage units are selected (i.e., choose n clusters). Then from each selected cluster or group, units are selected at random.

The units within a group or cluster are known as second stage units. If ith cluster containing Miunits is selected at the first stage then mi units are selected at random from Mi units at the second stage.

Advantages:

(i) It is more flexible than one stage sampling.

(ii) It reduces to one stage sampling when M = m but unless this is the best choice of m, we have the opportunity of taking some smaller value that appears more efficient.

(iii) This method reduces to a balance between statistical precision and cost.
It is easy to extend two-stage sampling to multi-stage sampling(with three, four or more stages).

For example: selecting a sample of secondary school students could involve selecting a sample of schools, then selecting a number of classes within a selected school, and finally selecting a number of students within a selected class.

TCSC