SAMPLING

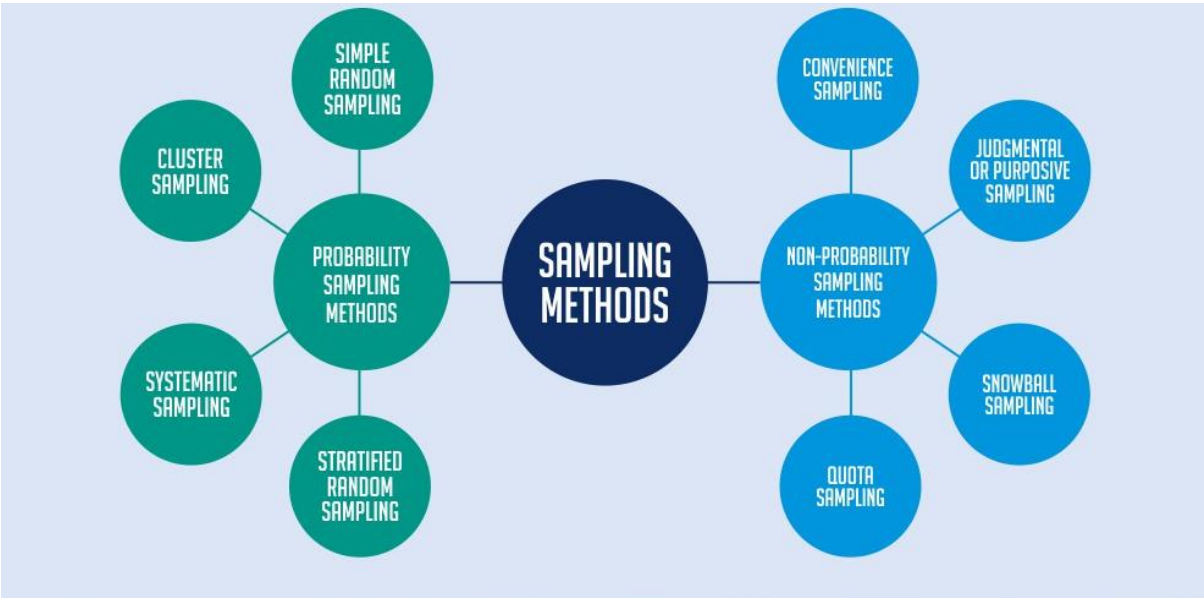
Sampling is the statistical process of selecting a subset (called a "sample") of a population of interest for purposes of making observations and statistical inferences about that population. Social science research is generally about inferring patterns of behaviours within specific populations. We cannot study entire populations because of feasibility and cost constraints, and hence, we must select a representative sample from the population of interest for observation and analysis. It is extremely important to choose a sample that is truly representative of the population so that the inferences derived from the sample can be generalized back to the population of interest.

The sampling process comprises of several stage.

- The first stage is defining the target population. A population can be defined as all people or items (unit of analysis) with the characteristics that one wishes to study.
- The second step in the sampling process is to choose a sampling frame. This is an accessible section of the target population (usually a list with contact information) from where a sample can be drawn. If your target population is professional employees at work, because you cannot access all professional employees around the world, a more realistic sampling frame will be employee lists of one or two local companies that are willing to participate in your study.
- The last step in sampling is choosing a sample from the sampling frame using a well-defined sampling technique.

Sampling Techniques and Types

Sampling techniques can be grouped into two broad categories: **Probability (random) sampling and non-probability sampling**. Probability sampling is ideal if generalizability of results is important for your study, but there may be unique circumstances where non-probability sampling can also be justified. These techniques are discussed in the next two sections.



Probability Sampling

Probability sampling is a technique in which every unit in the population has a chance (non-zero probability) of being selected in the sample, and this chance can be accurately determined. Sample statistics thus produced, such as sample mean or standard deviation, are unbiased estimates of population parameters, as long as the sampled units are weighted according to their probability of selection. All probability sampling have two attributes in common:

1. Every unit in the population has a known non-zero probability of being sampled, and
2. The sampling procedure involves random selection at some point.

The different types of probability sampling techniques include:

- **Simple random sampling**: In this technique, all possible subsets of a population (more accurately, of a sampling frame) are given an equal probability of being selected. Simple random sampling involves randomly selecting respondents from a sampling frame, but with large sampling frames, usually a table of random numbers or a computerized random number generator is used. This is the simplest of all probability sampling techniques; However, the simplicity is also the strength of this technique. Because the sampling frame is not subdivided or partitioned, the sample is unbiased and the inferences are most generalizable amongst all probability sampling techniques.
- **Systematic sampling**: In this technique, the sampling frame is ordered according to some criteria and elements are selected at regular intervals through that ordered list. This process will ensure that there is no overrepresentation of large or small institutions in your sample, but rather that institutions of all sizes are generally uniformly represented, as it is in your sampling frame. In other words, the sample is representative of the population, at least on the basis of the sorting criterion.
- **Stratified sampling**: In stratified sampling, the sampling frame is divided into homogeneous and non-overlapping subgroups (called “strata”), and a simple random sample is drawn within each subgroup. However, since there are many more small institutions in a sampling frame than large institutions, having an equal number of small, medium, and large institutions will make the sample less representative of the population (i.e., biased in favour of large institutions that are fewer in number in the target population). This is called non-proportional stratified sampling because the proportion of sample within each subgroup does not reflect the proportions in the sampling frame (or the population of interest), and the smaller subgroup (large-sized institutions) is over-sampled . An alternative technique will be to select subgroup samples in proportion to their size in the population. In this case, the proportional distribution of institutions in the population is retained in the sample, and hence this technique is called proportional stratified sampling.
- **Cluster sampling:** If you have a population dispersed over a wide geographic region, it may not be feasible to conduct a simple random sampling of the entire population. In such case, it may be reasonable to divide the population into “clusters” (usually along geographic boundaries), randomly sample a few clusters, and measure all units within that cluster. However, depending on between- cluster differences, the variability of sample estimates in a cluster sample will generally be higher than that of a simple random sample, and hence the results are less generalizable to the population than those obtained from simple random samples.
- **Matched-pairs sampling:** In this technique, researchers may want to compare two subgroups within one population based on a specific criterion. Matched-pairs sampling technique is often an ideal way of understanding bipolar differences between different subgroups within a given population.
- **Multi-stage sampling:** The probability sampling techniques described previously are all examples of single-stage sampling techniques. Depending on your sampling needs, you may combine these single-stage techniques to conduct multi-stage sampling. This is a two-stage combination of stratified and systematic sampling. For instance, you can start with a cluster of schools in the state of UP, and within each cluster, select a simple random sample of schools; within each school, select a simple random sample of

different classes in a school; and within each class group, select a simple random sample of students for study. In this case, you have a four-stage sampling process consisting of cluster and simple random sampling.

Non-Probability Sampling

Non-probability sampling is a sampling technique in which some units of the population have zero chance of selection or where the probability of selection cannot be accurately determined. Typically, units are selected based on certain non-random criteria, such as quota or convenience. Because selection is non-random, non-probability sampling does not allow the estimation of sampling errors, and may be subjected to a sampling bias. Therefore, information from a sample cannot be generalized back to the population. Types of non-probability sampling techniques include:

- **Convenience sampling:** Also called accidental or opportunity sampling, this is a technique in which a sample is drawn from that part of the population that is close to hand, readily available, or convenient. For instance, if you stand outside a shopping centre and hand out questionnaire surveys to people or interview them as they walk in, the sample of respondents you will obtain will be a convenience sample. This is a non-probability sample because you are systematically excluding all people who shop at other shopping centres. The opinions that you would get from your chosen sample may reflect the unique characteristics of this shopping centre such as the nature of its stores (e.g., high end-stores will attract a more affluent demographic), the demographic profile of its patrons, or its location (e.g., a shopping centre close to a university will attract primarily university students with unique purchase habits), and therefore may not be representative of the opinions of the shopper population at large. Hence, the scientific generalizability of such observations will be very limited. Other examples of convenience sampling are sampling students registered in a certain class or sampling patients arriving at a certain hospital. This type of sampling is most useful for pilot testing, where the goal is instrument testing or measurement validation rather than obtaining generalizable inferences.
- **Quota sampling:** In this technique, the population is segmented into mutually-exclusive subgroups (just as in stratified sampling), and then a non-random set of observations is chosen from each subgroup to meet a predefined quota. In proportional quota sampling , the proportion of respondents in each subgroup should match that of the population. For instance, if the population in a particular state consists of 55% Hindus, 20% Muslims,13% Buddhists, 5% Christians, 5% Jains and you wish to understand their voting preferences in an sample of 98 people, you can stand outside a shopping mall and ask people their voting preferences. But you will have to stop asking Hindus when you have 55 responses from that subgroup (or Buddhists when you have 13 responses) even as you continue sampling other religious groups, so that the religious composition of your sample matches that of the general population in that particular state. Non-proportional quota sampling is less restrictive in that you don't have to achieve a proportional representation, but perhaps meet a minimum size in each subgroup. In this case, you may decide to have 50 respondents from each of the five religious subgroups and stop when your quota for each subgroup is reached. Neither type of quota sampling will be representative of the state population, since depending on whether your study was conducted in a shopping mall near a university or near a corporation or near a residential society, your results may be entirely different. The non-proportional technique is even less representative of the population but may be useful in that it allows capturing the opinions of small and underrepresented groups through oversampling.
- **Expert sampling**: Also known as **Judgemental Sampling**, this is a technique where respondents are chosen in a non-random manner based on their expertise on the phenomenon being studied. The advantage of this approach is that since experts tend to be more familiar with the subject matter than non-experts, opinions from a sample of

experts are more credible than a sample that includes both experts and non-experts, although the findings are still not generalizable to the overall population at large.

- **Snowball sampling**: In snowball sampling, you start by identifying a few respondents that match the criteria for inclusion in your study, and then ask them to recommend others they know who also meet your selection criteria. For instance, if you wish to survey computer network administrators and you know of only one or two such people, you can start with them and ask them to recommend others who also do network administration. Although this method hardly leads to representative samples, it may sometimes be the only way to reach hard-to-reach populations or when no sampling frame is available.

.....