**Sampling Techniques**

**Introduction:**

Sampling is concerned with the selection of a subset/sample from population. The aim of any sample is to represent the characteristics of the whole population. It is nothing but a shortcut method for investigating a whole population. Using these techniques, data is gathered on a small part of the whole parent population or sampling frame, and used to inform what the whole picture is like. The three main advantages of sampling are that the cost is lower, data collection is faster, and the accuracy and quality of the data can be easily improved.

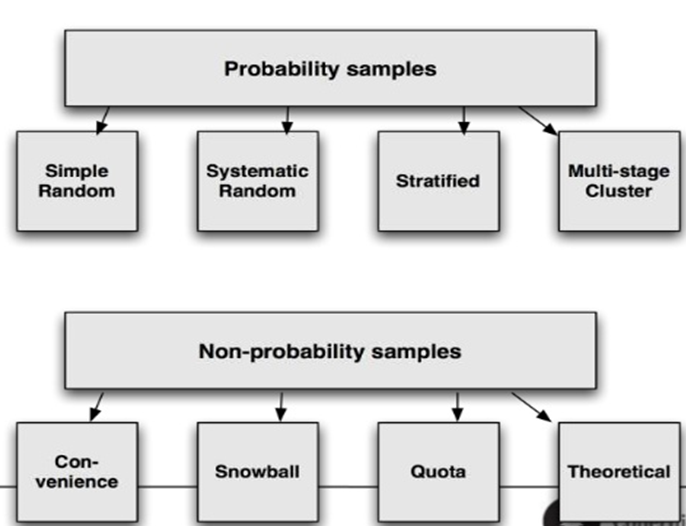
* We generally use sampling in our day to day life. When we go to purchase grains from market, we take handful of grains from bag to assess the quality of grains of whole bag
* If you visit doctor, he will take small sample of blood to know the constitute of blood in your whole body.

## Why sample?

In reality there is simply not enough; time, energy, money, labor/man power, equipment, access to suitable sites to measure every single item or site within the parent population or whole sampling frame. Therefore an appropriate sampling strategy is adopted to obtain a representative, and statistically valid sample of the whole.

**Types of sampling Techniques:**

There are different methods to generate a sample. As a researcher you will have to select the most appropriate method meet the requirements of your research.



## Few points need to consider before selecting any Sampling technique:

* Larger sample sizes are more accurate representations of the whole
* The sample size chosen is a balance between obtaining a statistically valid representation, and the time, energy, money, labor, equipment and access available
* A sampling strategy made with the minimum of bias is the most statistically valid
* Most approaches assume that the parent population has a normal distribution where most items or individuals clustered close to the mean, with few extremes
* A 95% probability or confidence level is usually assumed, for example 95% of items or individuals will be within plus or minus two standard deviations from the mean
* This also means that up to five per cent may lie outside of this - sampling, no matter how good can only ever be claimed to be a very close estimate

Let us discuss few important sampling methods in detail.

1. **Simple Random Sampling**: Random sampling is the purest form of probability sampling. In SRS of a given size, all such subsets of the frame are given an equal probability. Each element has an equal probability of selection. Any given pair of elements has the same chance of selection as any other pair. This minimizes bias and simplifies analysis of results. Example
   * The lottery method, e.g. picking numbers out of a hat or bag
   * The use of a table of random numbers.

Disadvantages:

* Can lead to poor representation of the overall parent population or area if large areas are not hit by the random numbers generated. This is made worse if the study area is very large
* There may be practical constraints in terms of time available and access to certain parts of the study area

## Systematic sampling: Systematic sampling relies on arranging the target population according to some ordering scheme, a random start, and then selecting elements at regular intervals through that ordered list. It is also called a Kth name selection technique.

## Algorithm:

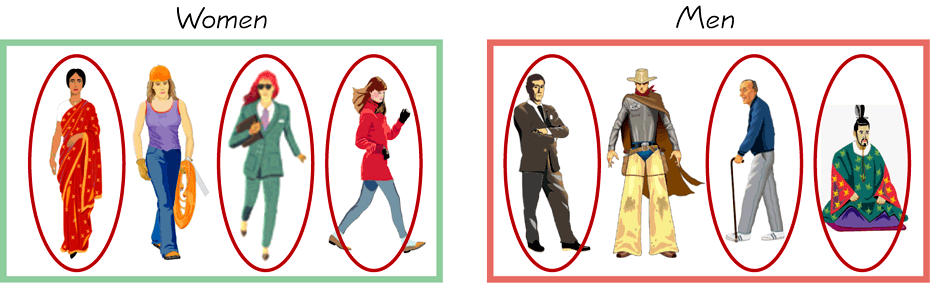
* Randomly select one data point from the population and create sample
* Calculate K value using (K = N/n, N = number of observations in whole population, n = number of observation we want in the sample)
* After the required sample size has been calculated, every Kth record is selected from a list of population members and add in the sample subset.
* As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method.
  + If every fifth person in list is rich or senior or other consistent pattern then avoid this method



## Stratified sampling: Stratified sampling is commonly used probability method that is superior to random sampling because it reduces sampling error.

Algorithm:

* Select a categorical variable and divide the data based on the categories present in that variable, also called stratums.
* A stratum is a subset of the population that share at least one common characteristic; such as males and females.
* Identify relevant stratums and their actual representation in the population.
* Random sampling is then used to select a sufficient number of subjects from each stratum.



Stratified sampling is often used when one or more of the stratums in the population have a low incidence relative to the other stratums.

Additionally, since each stratum is treated as an independent population, different sampling approaches can be applied to different strata, potentially enabling researchers to use the approach best suited for each identified subgroup. Stratified sampling can increase the cost and complicate the research design.

**Additional Information:**

Sampling schemes may be without replacement ('WOR'—no element can be selected more than once in the same sample) or with replacement ('WR'—an element may appear multiple times in the one sample). For example, if we catch fish, measure them, and immediately return them to the water before continuing with the sample, this is a WR design, because we might end up catching and measuring the same fish more than once. However, if we do not return the fish to the water (e.g., if we eat the fish), this becomes a WOR design.

**Interview Questions:**

1. Why do we go for sampling techniques?
2. Which technique should we prefer?
3. **What is the difference between stratified and Systematic Sampling?**
4. **Can we use sampling for infinite population?**