Sampling Techniques

Introduction to Samples:

a) Sample is an important concept in Statistics and forms an integral part of Inferential Statistics.

b) Samples are taken from the Population in order to conduct Sampling Studies

c) In day-to-day life or under normal circumstances, it is rather difficult and tedious to conduct studies on the Population due to various constraints.

d) In comparison, collection of data for the samples is comparatively convenient and less tedious for collection as well as for conducting Statistical Studies &/or Statistical Analysis.
   Advantages of incorporating Sampling Studies include the following:
   i)   Easy to collect
   ii)   Less time consuming
   iii)   Less expensive
   iv)   Lesser constraints involved vis-à-vis resources involved

e) Hence, in order to overcome these shortcomings, Sampling Studies are conducted in order to gain insights about the chosen population.

f) In order to conduct these Sampling Studies, it is essential & prudent to take Random Samples from the Population.

g) Random Samples thereby taken or collected are used for calculating the Sample Statistics viz. Mean, Proportion or Variance.
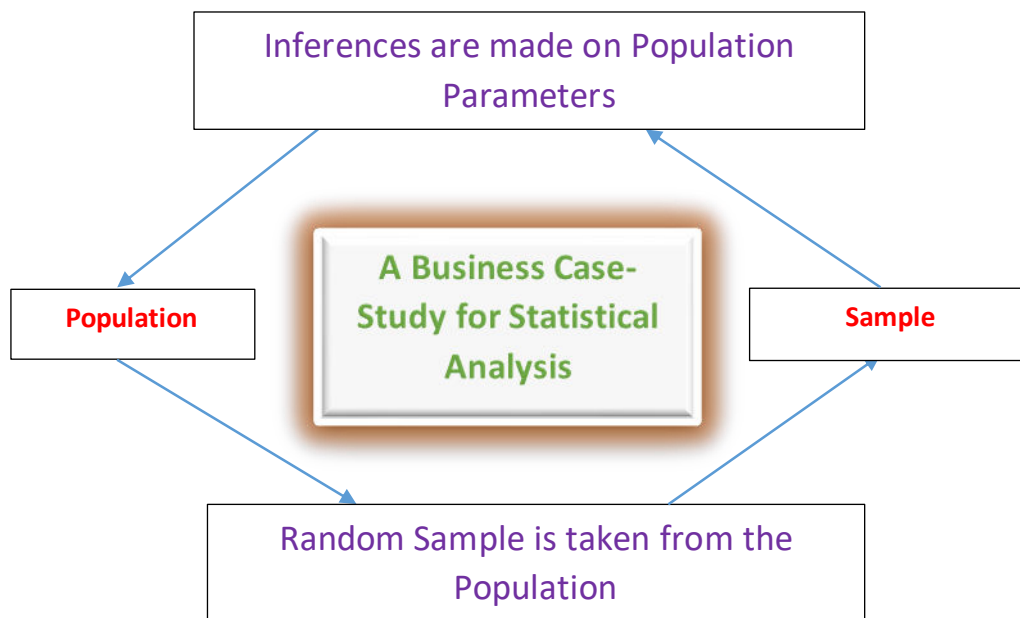
   Notations for calculated Sample Statistics:

   ·   Sample mean: $\bar{x}$ (or x bar)

   ·   Sample Proportion: $\hat{p}$ (or p-hat)

   ·   Sample Variance: s^2

h) These calculated Sampling Statistics are used for making inferences on Population Parameters viz. Mean, Proportion or Variance.
   Notations for Population Parameters:
   ▪ Population Mean: μ
   ▪ Population Proportion: P
   ▪ Population Variance: σ^2

i) Every business-case study presents a unique scenario(s) with numerous challenges or problems accompanying it. Based on the interest of the Business Researcher or the Statistician, in order to find the solution for the chosen case study, these Sampling Statistics are carefully chosen & calculated in order to make Inferences on the Population Parameters.

j) Thus in order to conduct such Sampling studies, there are various techniques associated in collecting such Samples. Such techniques are called as **Sampling Techniques**, thus forming an important part of Sampling Studies.

k) Sampling Techniques must be appropriately collected or carried out) in order to arrive at a meaningful solution & insights for the chosen business case-study.

l) It is of utmost importance that Samples should not be biased.

m) The Sample should be a Representative of the Entire Population.

Inferences are made on Population Parameters

Population

A Business Case-Study for Statistical Analysis

Sample

Random Sample is taken from the Population

Hence in order to conduct Statistical Studies, Sampling techniques are considered to be an important aspect. If the Sampling techniques aren't adhered to in principle, the outcomes & data interpretations could greatly vary & horribly go wrong thereby impacting the entire Business Case Study of interest.

Examples of Randomness

Some examples/illustrations of Randomness include:

a) Drops of water during a rainy day.
b) Picking up a ticket from Raffle draw
c) Playing the game of Roulette in the Casinos
d) Picking up random numbers while playing the game of Lotto or Housie
e) Randomly sprinkling food grains for the birds
f) Movement of the Sprinkler used for watering the plants

g) Taking a pinch of salt while cooking
h) Dispersion of light-rays from a transparent glass prism
i) Pouring oil in the frying-pan directly from the oil-can
j) Randomly plucking grass from the garden

& many more illustrations on Randomness.

Properties of a Random Sample

a) Sample size (n) – Should be reasonably large containing sufficient data points or observations in order to arrive at a conclusion or for decision-making process.
b) Representative of the Population
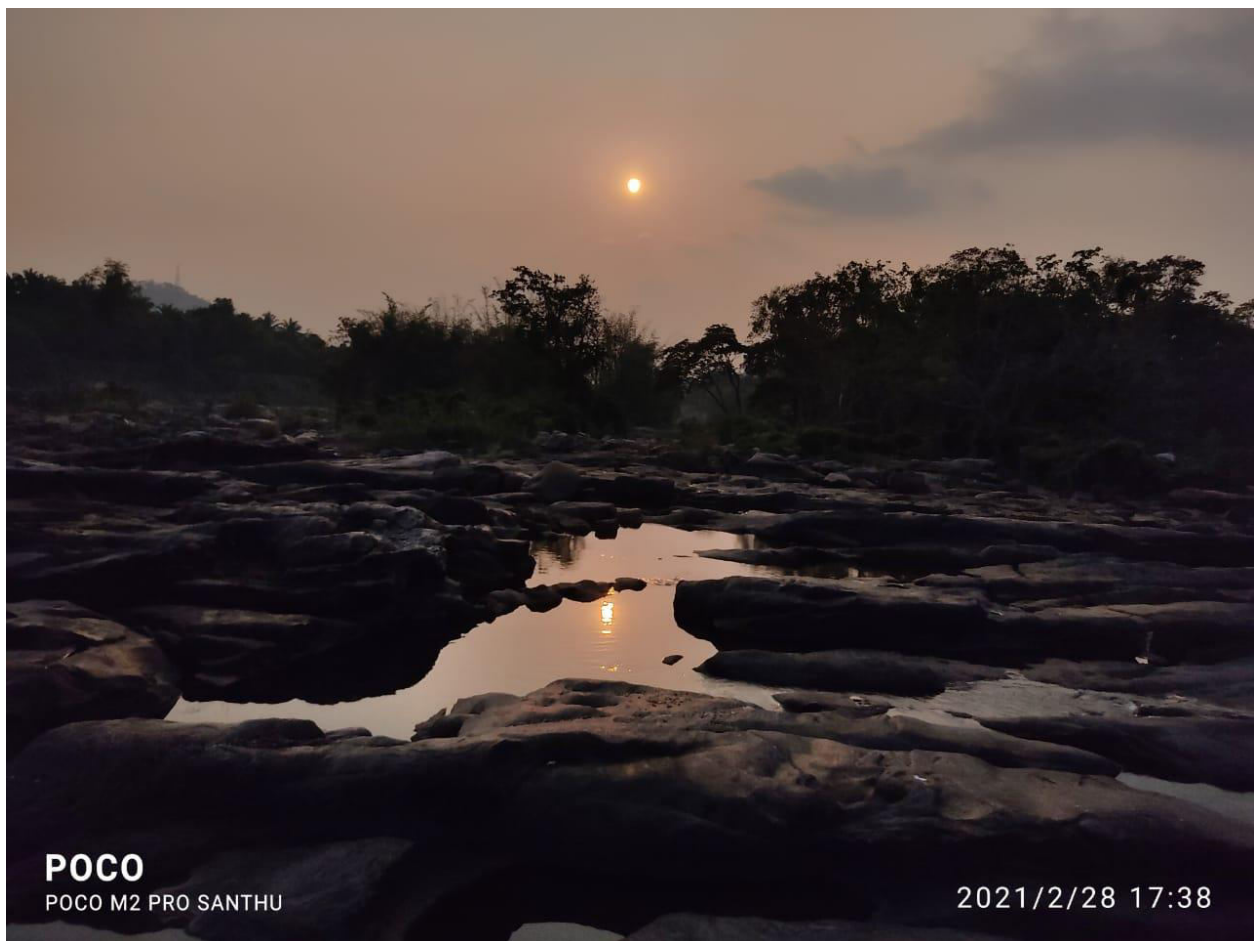
Salient & Distinguishing Features - Random Samples

1) Each & Every element of the population **has the same probability** of getting selected for Sampling purposes or for the Sampling Studies. For eg. In a game of Ludo or Snakes and Ladder, all the participants have the same probability of winning the game. Or in a Horse-Race, all the participant horses have equal probability of winning the race.
2) Also called as Probability Sampling because not every element is likely to be selected. For eg. In a football tournament with 16 countries participating, only 2 countries are likely to make it into the Semi-Finals from amongst the participants.
3) Used for conducting Inferential Statistical Studies.
4) Application of Quantitative techniques (viz. Hypothesis Testing, Confidence Intervals, Estimation of Standard Errors – Type I & Type II) on Random Samples gives better knowledge & insights regarding the Population.

Non Random Samples aren't used for Statistical Analysis. If at all, such samples are used, it can lead to high inaccuracy and thereby give a false picture of the chosen scenario.

Types of Random Sampling Techniques

Random Sampling Techniques can be categorized into 4 types viz.:

1) Simple Random Sampling
2) Stratified Random Sampling
3) Systematic Sampling
4) Cluster Sampling

1) Simple Random Sampling
   Following are its characteristics**---**
   ➢ A Sampling technique which is Basic or Elementary in nature, thereby it is most commonly used for conducting Statistical Studies.
   ➢ The outcomes from such Sampling Studies is used for Data interpretation & thereafter Statistical Analysis by using suitable Statistical techniques.
   ➢ ***Combination as well as Permutation techniques*** can be used ***for selecting the elements*** that constitutes the sample. Thus, in brief, the selection of elements for the Sample can be in any manner and in any direction.
   ➢ It includes ***Sampling techniques with Replacement*** as well ***as Sampling Techniques without Replacement***.
   ➢ **Example**: Random selection of customers to seek Customer feedback regarding a Product launch or a lucky-draw.
   ➢ In Real time, it is easy to conduct this technique on small population. As the Population size i.e. N increases, the steps involved for executing this technique becomes lengthy, tedious, cumbersome and unviable.
   ➢ Discrete data as well as Continuous data constitute such samples.
   ➢ This sampling technique is widely used in Machine Learning for dividing the Dataset into – Train & Test datasets.

2) Stratified Random Sampling
   Following are its characteristics**---**
   ➢ In this technique, the Population is further divided into several smaller segments or categories. These segments or categories also called as Sub-population ***do not*** overlap each other. In brief, these non-overlapping segments or categories are called as Strata.
   ➢ Thereafter, a random sample is created or extracted ***from each of these Strata or Subpopulation***. This is how Samples are chosen from the Population under this methodology or technique.
   ➢ In order to conduct this technique, the Business Researcher or Statistician must have must possess the requisite knowledge regarding the characteristics of the population.

➢ By invoking this technique i.e. Stratified Random Sampling process, the potentiality of the chosen sample matching with the given population remains high.

➢ Thus, in brief, portions of the samples are chosen or taken from different Strata or Population subgroups.

➢ The costs involved for conducting this Sampling technique can escalate given the fact that each unit or element in the population must be assigned to a Stratum before conducting the Random Selection Process.

➢ Strata selection is usually based on available data that is gathered while conducting the Census or Survey studies.

➢ These Stratifications are done on the basis of similarities i.e. same properties, same characteristics, same functionalities etc.

➢ In brief, each Strata is **Homogeneous from within or internally** whereas two or more strata are **Heterogeneous from outside** thereby constituting different characteristics, properties and functionalities.

➢ **Demographic data** is chiefly used while carrying out Stratified Random Sampling technique.

➢ **Examples of Stratified Random Sampling** include:
   a) Indian Election Analysis based on Gender, Age Groups, Geographic Region etc.
   b) Cricket Match Analysis – Constitution of two cricket teams, Weather, Pitch conditions etc.

➢ High preference is accorded to Stratified Random Sample so as to **reduce or minimize the Sampling Errors (both – Type I Error & Type II Error)** & in order to **closely resemble the Population**.

➢ Furthermore, Stratified Random Sampling can be further classified into two categories viz.:
   a) **Proportionate Stratified Random Sampling** – When the % each stratum from the sample **matches or is approximately proportionate** with the % of each stratum within the population.
   b) **Disproportionate Stratified Random Sampling** – When the % (or proportions) of each strata taken from the sample **differs** from the % (or proportions) of each strata taken from the population.

   In brief, Stratified Random Sampling technique is carried out to **minimize the Sampling Errors**.

➢ This methodology can be expensive & time-consuming too.

3) Systematic Random Sampling

Following are its characteristics**---**

➢ This technique is carried out mainly because of its convenience and the ease associated in order to conduct this technique.

- In this technique, occurrence of Sampling Errors cannot be controlled unlike in Stratified Random Sampling techniques.
- Say for example there are N elements in a Frame. From this framework, **every nth element** (say for eg. every 5th element) is selected from the given frame as decided at the beginning of an experiment, in order to constitute a Sample.
- It helps in determining whether a Sampling plan has been adhered to or whether there is a violation in the Sampling strategy or plan.
- However, there is a major drawback associated with this Systematic Random Sampling technique. In case, if there is periodicity associated with the data, it can lead to Non-randomness. For eg. Sampling data collated on the Agricultural Output, collated data regarding temperature of a particular region, Performance of Stock Markets etc. that are associated with Periodicity can lead to Non-Randomness thereby affecting the Statistical outcomes.

  Also if there are underlying patterns in the data, it could lead to High Bias.
- Systematic Random Sampling technique is widely practiced across Industries (i.e both – Product & Service Providers) as part **of QC (Quality Control)** measures & conducting the **Sample Audits**.

4) **Cluster Sampling** –

Following are its characteristics**---**

- Also called as Area Sampling technique.
- Similar to Stratified Random Sampling wherein each strata of the population is divided into Non-overlapping areas or clusters.
- Subgroups within the Population are used for this technique
- However, in contrast to Stratified Sampling technique wherein clusters are homogenous from within, rather it is the opposite in this case. In cluster Sampling technique, the clusters are heterogeneous from within, thereby constituting diverse characteristics, properties &/or functionalities.
- Thus, in brief, each cluster has variety of elements from within.
- In general, clusters are naturally occurring groups. Thus Cluster Sampling is invoked or preferred wherein focus lies on particular geographical region or area for conducting the Business Case Study.
- In order to conduct or invoke Cluster Sampling technique, the Statistician or the Business Researcher can adopt either of the two techniques viz.:
  a) Either to select all the elements from the chosen cluster or
  b) Randomly select individual elements from the chosen cluster
- Most importantly, just like Stratified Random Sampling, Discrete Data is used for constituting a Cluster Sampling Technique.
- Examples:

a) A geographical cluster can constitute human population that conduct various economic activities in order to earn his/her livelihood.

a. A geographical cluster can occur based on Topography & Natural Vegetation in that area for eg. Rivers, lakes, mountains, desert, forest, swamp & marshy lands etc.

➢ If the chosen cluster is quite large, this cluster can be broken down into two parts thereby leading to 2-stage Cluster Sampling technique.

➢ Advantages of Cluster Sampling includes:
  ▪ Convenience
  ▪ Cost of resources is comparatively less
  ▪ Data collection is easy to obtain
  ▪ Simplified approach

➢ Cluster Sampling techniques are incorporated while doing launch of products, services or facilities in a Test Market.

.