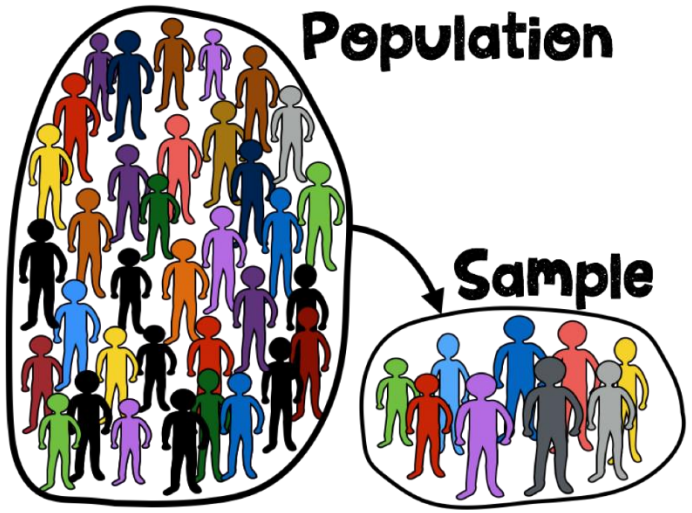
Siddhardhan

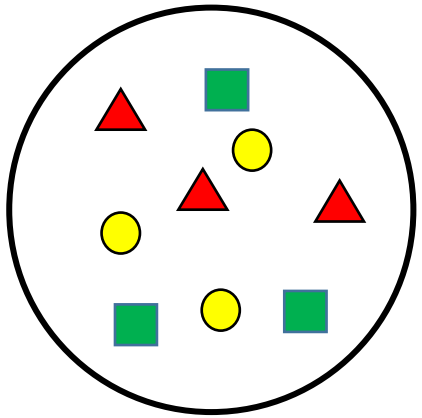
Population & Sample - Sampling Techniques

Math for Machine Learning

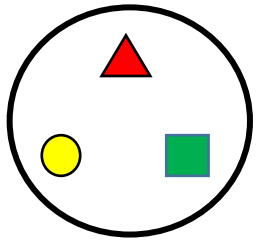


1. Sample Study

*A **sample study** is a study which is carried out on a sample which represents the total population.*



Population



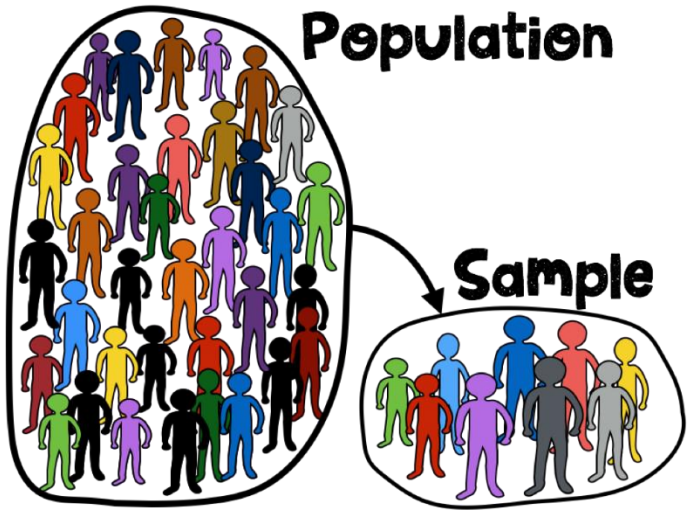
Sample

Average Blood Sugar Level = ?

Types of Sampling Techniques

Sampling Techniques:

- Simple Random Sampling
- Systematic Sampling
- Stratified Random Sampling
- Cluster Sampling



(Probability Sampling Techniques)

(Non-Probability Sampling Techniques)

Simple Random Sampling

In **Simple Random Sampling**, the sample is randomly picked from a larger population. Hence, all the individual datapoints has an equal probability to be selected as sample data.

Example: Employee survey in a company

Pros:

1. No sample Bias
2. Balanced Sample
3. Simple Method of sampling
4. Requires less domain knowledge

Cons:

1. Population size should be high
2. Cannot represent the population well sometimes

Systematic Sampling

In **Systematic Sampling**, the sample is picked from the population at regular intervals. This type of sampling is carried out if the population is homogeneous and the data points are uniformly distributed

Example: Selecting every 10^{th} member from a population of 10,000

Pros:

1. Quick & easy
2. Less bias
3. Even distribution of data

Cons:

1. Data manipulation risk
2. Requires randomness in data
3. Population should not have patterns.

Stratified Random Sampling

In **Stratified Random Sampling**, the population is subdivided into smaller groups called **Strata.** Samples are obtained randomly from all these strata.

Example: Smartphone sales in all the states

Pros:

1. Finds important characteristics in the population
2. High precision can be obtained if the differences in the strata is high

Cons:

1. Cannot be performed on populations that cannot be classified into groups.
2. Overlapping data points

Cluster Sampling

Cluster Sampling is carried out on population that has inherent groups. This population is subdivided into **clusters** and then random clusters are taken as sample.

Example: Smartphone sales in randomly selected states

Pros:

1. Requires only fewer resources
2. Reduced Variability
3. Advantages of both Random sampling and Stratified Sampling

Cons:

1. Cannot be performed on populations without natural groups
2. Overlapping data points
3. Can’t provide a general insight for the entire population