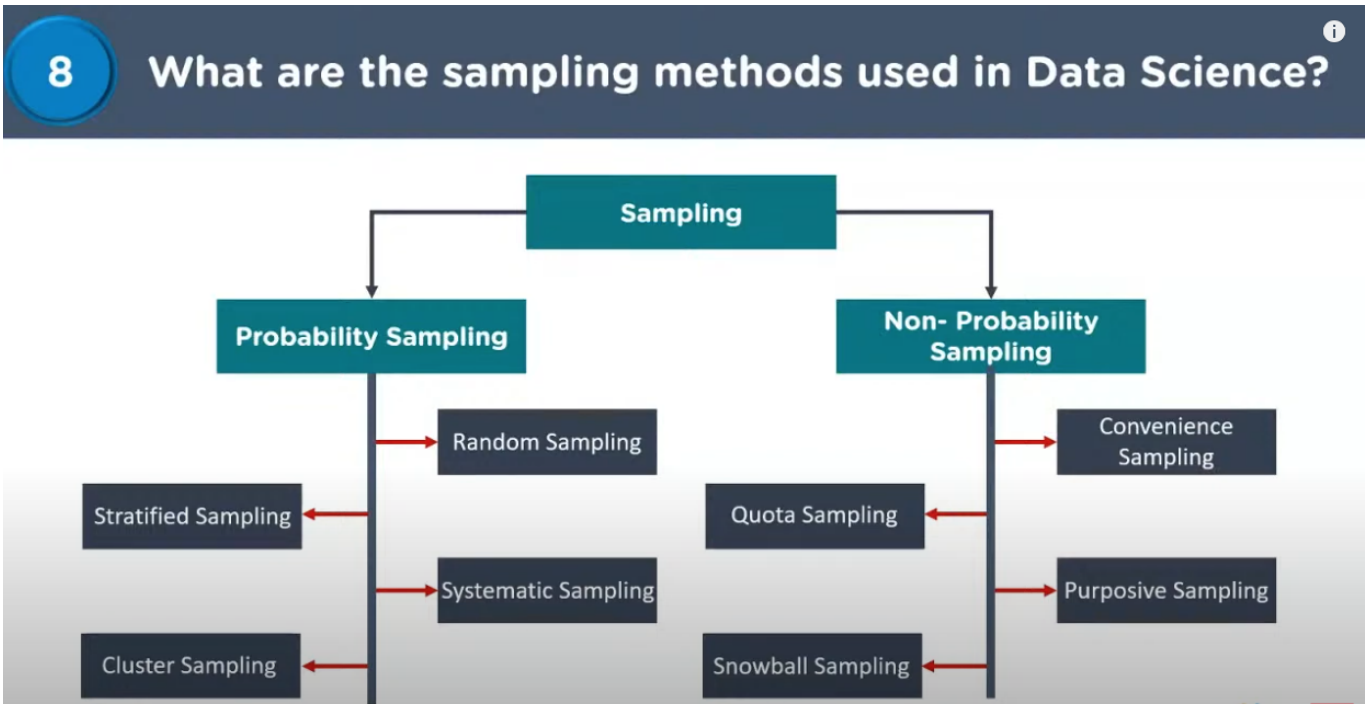
Sampling methods



Questions

Q474. What is the difference between population and sample in data?

Sample is the set of people who participated in your study whereas the population is the set of people to whom you want to generalize the results. For example – If you want to study the obesity among the children in India and you study 1000 children then those 1000 became sample whereas the all the children in the country is the population.

Sample is the subset of population.

Q475. What is the difference sample and sample frame?

Sample frame is the number of people who wanted to study whereas sample is the actual number of people who participated in your study. Ex – If you sent a marketing survey link to 300 people through email and only 100 participated in the survey then 300 is the sample survey and 100 is the sample.

Sample is the subset of sample frame. Both Sample and Sample Frame are subset of population.

68. Discuss how to randomly select a sample from a product user population.

The sampling techniques to select a sample from a product user population can be divided into two categories:

Probability sampling methods

- Simple Random Sampling
- Stratified Sampling
- Clustered Sampling
- Systematic Sampling

Non-Probability sampling methods

- Convenience Sampling
- Snowball Sampling
- Quota Sampling
- Judgement Sampling

43. What is Cluster Sampling?

Answer: Cluster sampling is a technique used when it becomes difficult to study the target population spread across a wide area and simple random sampling cannot be applied. Cluster Sample is a probability sample where each sampling unit is a collection or cluster of elements.

For eg., A researcher wants to survey the academic performance of high school students in Japan. He can divide the entire population of Japan into different clusters (cities). Then the researcher selects a number of clusters depending on his research through simple or systematic random sampling.

Q32. What is Cluster Sampling?

Cluster sampling is a technique used when it becomes difficult to study the target population spread across a wide area and simple random sampling cannot be applied. Cluster Sample is a probability sample where each sampling unit is a collection or cluster of elements.

For eg., A researcher wants to survey the academic performance of high school students in Japan. He can divide the entire population of Japan into different clusters (cities). Then the researcher selects a number of clusters depending on his research through simple or systematic random sampling.

Let's continue our Data Science Interview Questions blog with some more statistics questions.

Q33. What is Systematic Sampling?

Systematic sampling is a statistical technique where elements are selected from an ordered sampling frame. In systematic sampling, the list is progressed in a circular manner so once you reach the end of the list, it is progressed from the top again. The best example of systematic sampling is equal probability method.

Question 377. What Are Sampling Methods?

Answer:

There are four sampling methods:

Simple Random (purely random),
Systematic(every kth member of population),
Cluster (population divided into groups or clusters)
Stratified (divided by exclusive groups or strata, sample from each group) samplings.

Q32. What is Cluster Sampling?

Cluster sampling is a technique used when it becomes difficult to study the target population spread across a wide area and simple random sampling cannot be applied. Cluster Sample is a probability sample where each sampling unit is a collection or cluster of elements.

For eg., A researcher wants to survey the academic performance of high school students in Japan. He can divide the entire population of Japan into different clusters (cities). Then the researcher selects a number of clusters depending on his research through simple or systematic random sampling.

Let's continue our Data Science Interview Questions blog with some more statistics questions.

Q33. What is Systematic Sampling?

Systematic sampling is a statistical technique where elements are selected from an ordered sampling frame. In systematic sampling, the list is progressed in a circular manner so once you reach the end of the list, it is progressed from the top again. The best example of systematic sampling is equal probability method.

36. What is systematic sampling?

Answer: Systematic sampling is a technique, and the name resembles that it follows some systematic way, and the samples are selected from an ordered sampling frame. In systematic sampling, the list is actually in a circular manner and the selection starts from one end and reaches the final, and the cycle goes on. Equal probability method would be the best example for the systematic sampling.

20. Why is resampling done?

Answer: Resampling is done in any of these cases:

Estimating the accuracy of sample statistics by using subsets of accessible data or drawing randomly with replacement from a set of data points

Substituting labels on data points when performing significance tests

Validating models by using random subsets (bootstrapping, cross-validation)

9. Explain what resampling methods are and why they are useful. Also explain their limitations.

Answer by Gregory Piatetsky:

Classical statistical parametric tests compare observed statistics to theoretical sampling distributions. Resampling a data-driven, not theory-driven methodology which is based upon repeated sampling within the same sample.

Resampling refers to methods for doing one of these

- Estimating the precision of sample statistics (medians, variances, percentiles) by using subsets of available data (jackknifing) or drawing randomly with replacement from a set of data points (bootstrapping)
- Exchanging labels on data points when performing significance tests (permutation tests, also called exact tests, randomization tests, or re-randomization tests)
- Validating models by using random subsets (bootstrapping, cross validation)