1. What is central limit theorem and why is it important?

Ans. Instead of analyzing entire population data, we always take out a sample for analysis. The problem with sampling is that "sample means is a random variable – varies for different samples". And random sample we draw can never be an exact representation of the population. This phenomenon is called sample variation. To nullify the sample variation, we use the central limit theorem. And according to the Central Limit Theorem:

1. The distribution of sample means follows a normal distribution if the population is normal.

2. the distribution of sample means follows a normal distribution even though the population is not normal. But the sample size should be large enough.

3. The grand average of all the sample mean values give us the population mean.

2. What is sampling? How many sampling methods do you know?

Ans. Sampling, in simple terms, means selecting a group (a sample) from a population from which we will collect data for our research. Sampling is an important aspect of a research study as the results of the study majorly depend on the sampling technique used. So, in order to get accurate results or the results that can estimate the population well, the sampling technique should be chosen wisely.

Sampling techniques:

Simple Random Sampling (SRS)

Cluster sampling

3. What is the difference between type1 and typeII error?

Ans. Type 1 error, often referred to as a false positive, happens when a researcher rejects a real null hypothesis incorrectly. this suggests you're claiming your findings are noteworthy after they actually happened by coincidence. Your alpha level (), which is that the p-value below which you reject the null hypothesis, represents the likelihood of constructing a sort I error. When rejecting the null hypothesis, a p-value of 0.05 suggests that you simply are willing to tolerate a 5% probability of being mistaken. By setting p to a lesser value, you'll lessen your chances of constructing a kind I error.

Type II error: A type II error commonly said as a false negative happens when a researcher fails to reject a null hypothesis that's actually true. during this case, a researcher finds that there's no significant influence when, in fact, there is. Beta () is that the probability of creating a sort II error, and it's proportional to the statistical test's power (power = 1-). By ensuring that your test has enough power, you'll reduce your chances of constructing a sort II error. This can be accomplished by ensuring that your sample size is sufficient to spot a practical difference when one exists.

4. What do you understand by the term Normal distribution?

Ans. The normal distribution, also known as the Gaussian distribution, is a symmetric probability distribution

centred on the mean, indicating that data around the mean occur more frequently than data far from it. The normal distribution will show as a bell curve on a graph. A probability bell curve is referred to as a normal distribution. · The mean of a normal distribution is 0 and the standard deviation is 1. It has a kurtosis of 3 and zero skew. · Although all symmetrical distributions are normal, not all normal distributions are symmetrical. · Most pricing distributions aren't totally typical.

5. What is correlation and covariance in statistics?

Covariance is a measure of the relationship between two random variables in mathematics and statistics. The statistic assesses how much – and how far – the variables change in tandem. To put it another way, it's a measure of the variance between two variables. The metric, on the other hand, does not consider the interdependence of factors. Any positive or negative value can be used for the variance.

Correlation is a statistical method for determining whether or not two quantitative or categorical variables are related. To put it another way, it's a measure of how things are connected. Correlation analysis is the study of how variables are connected.

6. Differentiate between univariate ,Biavariate,and multivariate analysis.

Ans. Univariate statistics summarize only one variable at a time.

Bivariate statistics compare two variables.

Multivariate statistics compare more than two variables.

7. What do you understand by sensitivity and how would you calculate it?

Ans. The technique used to determine how independent variable values will impact a particular dependent variable under a given set of assumptions is defined as sensitive analysis. It's usage will depend on one or more input variables within the specific boundaries, such as the effect that changes in interest rates will have on a bond's price.

8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?

Hypothesis testing may be a method within which an analyst verifies a hypothesis a couple of population parameters. The analyst's approach is set by the kind of the info and also the purpose of the study. the utilization of sample data to assess the plausibility of a hypothesis is thought of as hypothesis testing.

Null Hypothesis (H0) A population parameter (such as the mean, standard deviation, and so on) is equal to a hypothesised value, according to the null hypothesis. The null hypothesis is a claim that is frequently made based on previous research or specialised expertise.

Alternative hypothesis (H1) The alternative hypothesis says that a population parameter is less, more, or different than the null hypothesis's hypothesised value. The alternative hypothesis is what you believe or want to prove to be correct.

9. What is quantitative data and qualitative data?

Quantitative data: The data collected on the grounds of the numerical variables are quantitative data. Quantitative data are more objective and conclusive in nature. It measures the values and is expressed in numbers. The data collection is based on "how much" is the quantity. The data in quantitative analysis is expressed in numbers so it can be counted or measured. The data is extracted from experiments, surveys, market reports, matrices, etc.

Qualitative data: The data collected on grounds of categorical variables are qualitative data. Qualitative data are more descriptive and conceptual in nature. It measures the data on basis of the type of data, collection, or category. The data collection is based on what type of quality is given. Qualitative data is categorized into different groups based on characteristics. The data obtained from these kinds of analysis or research is used in theorization, perceptions, and developing hypothetical theories. These data are collected from texts, documents, transcripts, audio and video recordings, etc.

10. How to calculate range and interquartile range?

Ans. The difference between the third and first quartiles is defined by the interquartile range. The partitioned values that divide the entire series into four equal parts are known as quartiles. So, there are three quartiles. The first quartile, known as the lower quartile, is denoted by Q1, the second quartile by Q2, and the third quartile by Q3, known as the upper quartile. As a result, the interquartile range equals the upper quartile minus the lower quartile.

IQR = Upper Quartile – Lower Quartile

11. What do you understand by bell curve distribution ?

A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side. Bell curves are visual representations of normal distribution, also called Gaussian distribution.

A normal distribution curve, when graphed out, typically follows a bell-shaped curve

12. Mention one method to find outliers

Z-scores can quantify the unusualness of an observation when your data follow the normal distribution. Z-scores are the number of standard deviations above and below the mean that each value falls. For example, a Z-score of 2 indicates that an observation is two standard deviations above the average while a Z-score of -2 signifies it is two standard deviations below the mean. A Z-score of zero represents a value that equals the mean.

To calculate the Z-score for an observation, take the raw measurement, subtract the mean, and divide by the standard deviation. Mathematically, the formula for that process is the following:

$$Z = \frac{X - \mu}{\sigma}$$

13. What is p-value in hypothesis testing?

Ans. The p-value in statistics is that the likelihood of getting outcomes a minimum of as extreme because the observed results of a statistical hypothesis test, given the null hypothesis is valid. The p-value, instead of rejection points, is employed to work out the smallest amount level of significance at which the null hypothesis is rejected. A lower p-value indicates that the choice hypothesis has more evidence supporting it.

14. What is the Binomial Probability Formula?

A discrete distribution is a binomial distribution. It's a well-known probability distribution. The model is then used to depict a variety of discrete phenomena seen in business, social science, natural science, and medical research. Because of its relationship with a binomial distribution, the binomial distribution is commonly employed. For binomial distribution to be used, the following conditions must be met:

There are n identical trials in the experiment, with n being a limited number.

Each trial has only two possible outcomes, i.e., each trial is a Bernoulli's trial.

One outcome is denoted by the letter S (for success) and the other by the letter F (for failure) (for failure).

From trial to trial, the chance of S remains the same. The chance of success is represented by p, and the likelihood of failure is represented by q (where p+q=1).

Each trial is conducted independently.

The number of successful trials in n trials is the binomial random variable

15. Explain ANOVA and it's applications.

Ans. ANOVA is the way to find out if experimental results are significant. One-way ANOVA compares two means from two independent groups using only one independent variable. Two-way ANOVA is the extension of one-way ANOVA using two independent variables to calculate the main effect and interaction effect.