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

   Ans. The central limit theorem in statistics states that, given a sufficiently large sample size, the sampling distribution of the mean for a variable will approximate a normal distribution regardless of that variable's distribution in the population.

The central limit theorem is vital in statistics for two main reasons—the normality assumption and the precision of the estimates. Central limit theorem and the normality assumption

The fact that sampling distributions can approximate a normal distribution has critical implications. In statistics, the normality assumption is vital for parametric hypothesis tests of the mean, such as the t-test. Consequently, you might think that these tests are not valid when the data are nonnormally distributed. However, if your sample size is large enough, the central limit theorem kicks in and produces sampling distributions that approximate a normal distribution. This fact allows you to use these hypothesis tests even when your data are nonnormally distributed—as long as **your sample size is large enough.**

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

   Ans. Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate the characteristics Method of Sampling.
   1. Simple random sampling · 2. Systematic sampling · 3. Stratified sampling · 4. Clustered sampling · 1. Convenience sampling.

3. **What is the difference between error i and error ii ?**

     Ans- Type – 1 error is known as false positive, i.e., when we reject the correct null hypothesis,

whereas type -2 error is also known as a false negative, i.e., when we fail to reject the false null hypothesis.

In this article, we will discuss difference between type 1 and type 2 error.

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

   Ans. The normal distribution is a continuous probability distribution that is symmetrical around its mean with most values near the central peak.

5. What is correlation and covariance in statistics?

Ans. Covariance and correlation are two terms that are opposed and are both used in statistics and regression analysis. Covariance shows you how the two variables differ, whereas correlation shows you how the two variables are related

6. Differentiate between univariate Biavariate, and multivariate analysis.

   Ans. Univariate Analysis- Univariate analysis is the simplest of the three analyses where the data you are analyzing is only one variable. There are many different ways people use univariate analysis. The most common univariate analysis is checking the central tendency (mean, median and mode), the range, the maximum and minimum values, and standard deviation of a variable.

   Biavariate: Bivariate analysis is where you are comparing two variables to study their relationships. These variables could be dependent or independent to each other. In Bivariate analysis is that there is always a Y-value for each X-value.

   Multivariate: Multivariate analysis is similar to Bivariate analysis but you are comparing more than two variables. For three variables, you can create a 3-D model to study the relationship (also known as Trivariate Analysis). However, since we cannot visualize anything above the third dimension, we often rely on other softwares and techniques for us to be able to grasp the relationship in the data.

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

Ans. A sensitivity analysis is a financial model that allows you to understand the effect of fluctuations in selected variables on your business' profitability.

Gross profit = total revenue - cost of goods sold

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

Ans. Null hypothesis (H0): The null hypothesis here is what currently stated to be true about the population. In our case it will be the average height of students in the batch is 100. Alternate hypothesis (H1): The alternate hypothesis is always what is being claimed.

let's say you were running a z test with an alpha level of 5% (0.05). In a one tailed test, the entire 5% would be in a single tail. But with a two tailed test, that **5% is split between the two tails, giving you 2.5% (0.025) in each tail**.

9. What is quantitative data and qualitative data?

Ans. Quantitative data refers to any information that can be quantified, counted or measured, and given a numerical value. Qualitative data is descriptive in nature, expressed in terms of language rather than numerical values. Quantitative research is based on numeric data.

10. How to calculate range and interquartile range?

Ans. To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1

11. What do you understand by bell curve distribution ?

Ans for statistical experiments like when coins are flipped several time. The bell curve graph is useful for repeated measurements of equipment. The bell curve graph is also useful in measuring characteristics in biology. The bell curve also has its relevance s.

12. Mention one method to find outliers.

Ans. An outlier is a piece of data that is an abnormal distance from other points. In other words, it's data that lies **outside the other values** in the set. If you had Pinocchio in a class of children, the length of his nose compared to the other children would be an outlier.
In this set of random numbers, 1 and 201 are outliers:
1, 99, 100, 101, 103, 109, 110, 201
"1" is an extremely low value and "201" is an extremely high value.

13. What is p-value in hypothesis testing?

Ans- A p value is used in hypothesis testing to help you support or reject the null hypothesis. The p value is the evidence **against** a null hypothesis. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

14. What is the Binomial Probability Formula?

Ans. if we toss a coin, there could be only two possible outcomes: heads or tails, and if any test is taken, then there could be only two results: pass or fail. This distribution is also called a binomial probability distribution.

15. Explain ANOVA and it's applications.

    Ans. ANOVA stands for "analysis of variance" and is used in statistics when you are testing a hypothesis to understand how different groups respond to each other

    ,