

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

Ans :The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution. Sample sizes equal to or greater than 30 are often considered sufficient for the CLT to hold.

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

Ans :When you conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, you select a sample. The sample is the group of individuals who will actually participate in the research.

There are two types of sampling methods:

- Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group.
- Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

3. What is the difference between type I and type II error?

Ans : A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population. Although type I and type II errors can never be avoided entirely, the investigator can reduce their likelihood by increasing the sample size (the larger the sample, the lesser is the likelihood that it will differ substantially from the population).

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

Ans : A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

5. What is correlation and covariance in statistics?

Ans : Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency.

Correlation is a statistical measure that indicates how strongly two variables are related.

6. Differentiate between univariate, Bivariate, 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 **sensitivity** of a test is the proportion of people who **test positive** among all those who actually **have the disease**.

8. What is hypothesis testing? What is H_0 and H_1 ? What is H_0 and H_1 for two-tail test? Ans :

A statistical hypothesis is an assertion or conjecture concerning one or more populations. To prove that a hypothesis is true, or false, with absolute certainty, we would need absolute knowledge. That is, we would have to examine the entire population. Instead, hypothesis testing concerns on how to use a random sample to judge if it is evidence that supports or not the hypothesis.

- H_0 : the null hypothesis;
- H_1 : the alternate hypothesis

9. What is quantitative data and qualitative data?

Ans : Quantitative data is anything that can be counted or measured; it refers to numerical data. Qualitative data is descriptive, referring to things that can be observed but not measured

10. How to calculate range and interquartile range?

Ans : Range = Max-Min

Interquartile range = $Q2 = Q3 - Q1$

$Q1 = 1/4(n+1)$ th term

$Q3 = 3/4(n+1)$ th term

11. What do you understand by bell curve distribution ?

Ans : A bell curve is a graph depicting the normal distribution, which has a shape reminiscent of a bell. The top of the curve shows the mean, mode, and median of the data collected. Its standard deviation depicts the bell curve's relative width around the mean.

12. Mention one method to find outliers.

Ans : IQR method

13. What is p-value in hypothesis testing?

Ans : The p-value is defined as the probability of obtaining the result at least as extreme as the observed result of a statistical hypothesis test, assuming that the null hypothesis is true.

14. What is the Binomial Probability Formula?

Ans : $P(r) = {}^nC_r \cdot p^r (1 - p)^{n-r}$.

15. Explain ANOVA and it's applications.

Ans : Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.