

1. Differentiate between Inferential Statistics and Descriptive Statistics.  
 Descriptive Statistics.
  - It is a summary statistic that quantitatively describes or summarizes the feature of a collection of information. It helps in knowing the data better.
  - It gives information about raw data which describes the data in some manner.
  - It helps in organizing, analyzing, and to present data in a meaningful manner.
  - It is used to describe a situation.
  - It explains already known data and is limited to a sample or population having a small size.
  - It can be achieved with the help of charts, graphs, tables, etc.
 Inferential Statistics
  - It is defined as inferring population with the help of sample(true representation or a subset of population)
  - It makes inferences about the population using data drawn from the population.
  - It allows us to compare data, and make hypotheses and predictions.
  - It is used to explain the chance of occurrence of an event.
  - It attempts to reach the conclusion about the population.
  - It can be achieved by probability.
  
2. Differentiate between population and sample.  
 A population is the entire group that you want to draw conclusions about. Universe of elements to be studied. It can be classified according to the number of individuals that make it up. It has statistical variables. To analyze the data collected regarding the common characteristics shared by the elements for various purposes.  
 A sample is the specific group that you will collect data from. The size of the sample is always less than the total size of the population. Selection of a part of the population. It is part of the population: it should comprise between 5% and 10% to be most effective. Variable could be random. To study the behavior, characteristics, tastes, or properties of a representative part of the population.
  
3. What is a hypothesis? Differentiate between null and alternative hypothesis.  
 A hypothesis is a testable prediction which is expected to occur. It can be true or false based on the underlying information in the data provided for the testing.  
 Alternate Hypothesis: In the Alternate hypothesis, there is no relationship between the two variables. Generally, researchers and scientists try to reject or disprove the null hypothesis. If the null hypothesis is accepted researchers have to make changes in their opinions and statements. This hypothesis is denoted by  $H_1$  or  $H_a$ . It is generally used when we reject the null hypothesis. In this hypothesis, the p-value is smaller than the significance level.  
 Null Hypothesis: In the null hypothesis, there is some relationship between the two variables i.e. They are dependent upon each other. Generally, researchers and scientists try to accept or approve the null hypothesis. If the alternative hypothesis gets accepted researchers do not have to make changes in their opinions and statements. This hypothesis is denoted by  $H_0$ . It gets accepted if we fail to reject the null hypothesis. In this hypothesis, the p-value is greater than the significance level.
  
4. What is the central limit theorem?  
 Central limit theorem is a statistical theory which states that when the large sample size has a finite variance, the samples will be normally distributed and the mean of samples will be approximately equal to the mean of the whole population.  
 In other words, the central limit theorem states that for any population with mean and

standard deviation, the distribution of the sample mean for sample size  $N$  has mean  $\mu$  and standard deviation  $\sigma / \sqrt{n}$ .

5. Differentiate between type I and type II errors.  
Type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population
  - It is also known as a false-positive.
  - It occurs if the researcher rejects a correct null hypothesis in the population.
  - i.e., incorrect rejection of the null hypothesis.
  - Measured by alpha (significance level).
  - If the significance level is fixed at 5%,
  - It means there are about five chances of type – 1 error out of 100.Type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.
  - It is also known as a false negative.
  - It occurs if a researcher fails to reject a null hypothesis that is actually a false hypothesis.
  - Measured by beta (the power of test).
  - The probability of committing a type -2 error is calculated by  $1 - \text{beta}$  (the power of test).
  
6. What is linear regression?  
Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value. It mathematically models the unknown or dependent variable and the known or independent variable as a linear equation.  
For instance, suppose that you have data about your expenses and income for last year. Linear regression techniques analyze this data and determine that your expenses are half your income. They then calculate an unknown future expense by halving a future known income.
  
7. What are the assumptions required for linear regression?
  1. Linear relationship: There exists a linear relationship between the independent variable,  $x$ , and the dependent variable,  $y$ .
  2. Independence: The residuals are independent. In particular, there is no correlation between consecutive residuals in time series data.
  3. Homoscedasticity: The residuals have constant variance at every level of  $x$ .
  4. Normality: The residuals of the model are normally distributed.