**1.What is the main goal of inferential statistics?**

The goal of inferential statistics is to make predictions based on the population by looking at the sample of the data.

It involves sample data to make predictions about a larger group.

Eg: There is a box of candies and from that we take out a few candies and based on the few selected candies we can predict the behavior/nature of the other candies.

**2.Explain the difference between a population and a sample.**

The difference between the population and sample are:

Population:

The population is the entire set of dataset which we are going to use for making predictions or generalizations.

Eg: Calculating the average height of males in a country. Then the population would be all the males staying in the country.

Sample:

The sample is a subset of the population. It is the group of individuals or observations selected from a larger population.

Eg: Indian cricket team 11 Players are sample and the entire country males are the population.

**3. What is a confidence interval, and how is it useful in inferential statistics?**

A confidence interval (CI) is a statistical tool used in inferential statistics to estimate the range within which a population parameter is likely to fall.

The confidence interval is often expressed as a percentage, such as a 95% confidence interval.

Eg Calculating the scores of the students in a class by taking there mean and standard deviation.

**4.Techniques of Inferential Statistics (three techniques)**

There are three techniques in Inferential Statistics

* Population Inference
* Hypothesis testing
* Cross Validation

Population Inference:

In population inference we check the performance of model the performance of matrix and how we evaluate the Matrix by taking the subset of data and based on that subset how we can make a prediction

Hypothesis testing:

Hypothesis means assumption from which we can make a conclusion.

It is used to make inferences about a population based on a sample of data.

Cross Validation:

The basic idea behind cross-validation is to split the dataset into multiple subsets, train the model on some of these subsets, and evaluate its performance on the remaining data.

**5. Difference between mean median and mode?**

Mean

The mean, also known as the average, is calculated by summing up all the values in a dataset and then dividing the sum by the number of values.

Median

The median is the middle value in a dataset when it is ordered from least to greatest. If there is an even number of values, the median is the average of the two middle values.

Mode

The mode is the value that appears most frequently in a dataset.

The main difference between the three are:

* Mean is sensitive to extreme values.
* Median is less affected by outliers.
* Mode is not influenced by outliers

**6. What is the purpose of Descriptive Statistics?**

Descriptive statistics serve the purpose of summarizing and presenting key features of a dataset. These statistical measures provide a clear and concise summary of the main characteristics of a set of data.

**7. What is skewness in statistics?**

Skewness is a measure of skew of a probability distribution.It quantifies the extent and direction of skew in a dataset. Skewness provides information about the shape of the distribution and the location of the tail relative to the center of the distribution.

**8. What is p value?**

p-value is a measure that helps researchers assess the evidence against a null hypothesis. The p-value is used in hypothesis testing to determine the statistical significance of an observed effect or result. It quantifies the probability of obtaining results as extreme as, or more extreme than, the ones observed in the data if the null hypothesis is true.

**9. How do you interpret standard deviation?**

Standard deviation is a measure of the amount of variation or dispersion in a set of values.

It provides a way to quantify how spread out the values in a dataset are from the mean (average).