**Descriptive Statistics:**

1.What is the purpose of descriptive statistics?

Ans:

The purpose of descriptive statistics is to provide the summary and describe the different features of the dataset in a meaningful way.

2.Can you explain the difference between mean, median, and mode?

Ans:

Mean: This is used to find the average of all values.

Median: is used to find the middle value or average of two middle values arranged in ascending or descending order.

Mode: is used to find the most frequently occurring values in the dataset.

3. How do you interpret the standard deviation of a dataset?

Ans:

interpret the standard deviation of a dataset describes how the data is spread out from the average.

If a deviation is high means: the values of the dataset are spread out in a larger range. whereas the deviation is lower means values are closer to average.

4. Describe the concept of skewness in statistics.

Ans:

Right skewness: distribution is shifted to the right and the mean is greater than the median.

Left skewness: distribution is shifted and the mean is less than the median

Normal skewness: data is distributed evenly, thus giving a bell curve.

**Inferential Statistics:**

1.What is the main goal of inferential statistics?

Ans:

Goal is to make a prediction of data in the population.

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

Ans:

Population refers to the entire set of data whereas sample is the subset of population which is selected to study the entire group.

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

Ans:

It is a range of values that indicate the likelihood of getting a result. It is useful in inferential statistics that provide an informative way to understand and communicate the uncertainty associated with the population parameter.

4. Define p-value

Ans:

The p-value is the probability of observing data as extreme, more extreme than the observed results.

5. Techniques of inferential statistics

Ans:

1. Population inference: This is used to check the performance of the model and evaluate model performance by taking the subset, hence making predictions of the data.

2. Hypothesis testing: It takes some assumptions based on the assumption it will predict the data.

3. Cross Validation: Used to evaluate the performance of the model by splitting into subsets, Training is done on one subset and testing on other data to see how well the model performed on unseen data.