**Descriptive Statistics**

1. **What is the purpose of descriptive statistics?**

Descriptive statistics refers to a set of methods used to summarize and describe the main features of a dataset, such as its central tendency, variability, and distribution. These methods provide an overview of the data and help identify patterns and relationships.

It's about organizing, presenting, and summarizing data in a meaningful and informative way.

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

The mean is the average of a set of numbers.The median is the middle value in a dataset when the values are arranged in ascending or descending order.The mode is the value that appears most frequently in a dataset.

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

The standard deviation is a measure of the amount of variation or dispersion in a dataset. It quantifies the average distance of data points from the mean of the dataset. A larger standard deviation indicates greater variability, while a smaller standard deviation suggests that the data points are closer to the mean.

1. **Describe the concept of skewness in statistics.**

Skewness is the measure of the asymmetry of the distribution of data. The data is not symmetrical (i.e.) it is skewed towards one side. Skewness is classified into two types: positive skew and negative skew.

1. **Positively skewed:** In a Positively skewed distribution, the data values are clustered around the left side of the distribution, and the right side is longer. The mean and median will be greater than the mode in the positive skew.
2. **Negatively skewed:** In a Negatively skewed distribution, the data values are clustered around the right side of the distribution, and the left side is longer. The mean and median will be less than the mode.

**Inferential Statistics**

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

The goal of inferential statistics is to make inferences or conclusions about a population based on a sample taken from that population. It involves using sample data to draw conclusions, make predictions, test hypotheses, and generalize findings to a larger population.

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

A population is the entire group that you want to draw conclusions about. 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.

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

The confidence interval is a type of interval estimate from the sampling distribution which gives a range of values in which the population statistic may lie. It's a statistical tool used in inferential statistics to estimate the unknown population parameter, such as a population mean or proportion, based on sample statistics.

1. **Define p-value**

A p-value is a metric that expresses the likelihood that an observed difference could have occurred by chance. As the p-value decreases the statistical significance of the observed difference increases. If the p-value is too low, you reject the null hypothesis.

1. **Techniques of Inferential Statistics**

**Population inference:** It is the process of making conclusions or inferences about an entire population based on information gathered from a sample of that population. The goal is to use data collected from a sample to draw generalizations or make predictions about the population.

**Hypothesis Testing:** Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test. It is used to estimate the relationship between 2 statistical variables.

**Cross validation:** Cross validation is a technique for evaluating a machine learning model and testing its performance. It helps to compare and select an appropriate model for the specific predictive modelling problem.