Descriptive Statistics

What is the purpose of descriptive statistics?
Descriptive statistics primary goal is to provide a clear and concise summary of the data, enabling us to gain insights and understand patterns, trends and distribution within the dataset

Can you explain the difference between mean, median, and mode?
Mean - gives us the average value of the data
Mean = sum of all values / total number of values
Median - median gives us the center of the data points
Mode - gives the most occurring value in the dataset

How do you interpret the standard deviation of a dataset?

Standard Deviation shows us how far each value lies from the mean of the data. If the standard deviation is low then it means that most of the data points are close to the mean. If the standard deviation is large then most of the points are spread out from the mean.

Describe the concept of skewness in statistics.
Skewness is the measure of asymmetry in probability distribution of a dataset.
Positive skewness - The distribution will be skewed to the right. The mean will be greater than the median and the distribution is skewed toward the higher values.
Negative skewness - The distribution will be skewed to the left. The mean will be less than the median and the distribution will be skewed towards the lower values.

Inferential Statistics

What is the main goal of inferential statistics?
Inferential statistics uses data from samples to make inferences or predictions about a larger population.

Explain the difference between a population and a sample.
Every possible element on which the analysis is done
A sample is a subset or a smaller group selected from the larger population. It's chosen to represent the characteristics or properties of the entire population.

What is a confidence interval, and how is it useful in inferential statistics?
A confidence interval is a range of values that is calculated from sample data and is believed to contain the true parameter with a certain level of confidence. It's a way of quantifying the uncertainty or variability associated with an estimate.

Define p-value.
The p-value is a measure used to determine the strength of evidence against the null hypothesis in hypothesis testing. It quantifies the level of statistical significance of an observed result.

Techniques of Inferential Statistics (three techniques)
Population Inference - it refers to the process of drawing conclusions or making predictions about an entire population based on the information gathered from a sample of the population.

Hypothesis testing - uses sample data to evaluate a hypothesis about a population. It tells us whether any observation or relationships within a sample are likely to represent the relationships or characteristics of the entire population.
Cross-Validation - it assesses how well the model trained on a certain dataset will perform on new unseen data.