Descriptive Statistics:

1. Purpose of Descriptive Statistics:

Descriptive statistics provide a concise overview of a dataset through measures like mean, median, mode, range, and standard deviation, simplifying large amounts of data into understandable formats.

By using descriptive statistics, one can easily understand the central tendency, dispersion, and general pattern in the data, making it easier to communicate and interpret information.

2. Difference Between Mean, Median, and Mode:

The average of all data points. It's calculated by summing all values and dividing by the number of observations. Sensitive to outliers.

The middle value when data points are arranged in ascending or descending order. It divides the dataset into two equal parts and is less affected by outliers.

The most frequently occurring value in a dataset. There can be more than one mode, and it is the only measure of central tendency that can be used with nominal data.

3. Interpreting Standard Deviation:

Standard deviation indicates how spread out the numbers in a dataset are. A low standard deviation means that the data points are close to the mean.

When combined with the mean, it provides a comprehensive picture of the data distribution. A higher standard deviation implies greater variability in the data points.

4. Concept of Skewness:

Skewness quantifies how much a dataset deviates from a normal distribution in terms of asymmetry.

Positive skew (tail on the right) indicates that the right tail is longer or fatter, while negative skew (tail on the left) suggests the opposite. Symmetric distributions have a skewness close to zero.

Inferential Statistics:

5. Main Goal of Inferential Statistics:

Inferential statistics are used to draw conclusions or make predictions about a larger population based on a sample of data.

It allows us to infer trends about a larger population without having to collect data from everyone in that population.

6. Difference Between a Population and a Sample:

Refers to the entire group that is the subject of a study. It includes all members of a defined group.

A subset of the population, selected for the actual study. It is used to draw conclusions about the entire population.

7. Confidence Interval:

A confidence interval is a range of values, derived from the sample data, that is likely to contain the value of an unknown population parameter.

It expresses the degree of certainty or confidence about the interval. For example, a 95% confidence interval suggests that if the sampling were repeated many times, 95% of the intervals would contain the population parameter.

8. Definition of p-value:

The p-value measures the probability of obtaining a result at least as extreme as the one observed, assuming that the null hypothesis is true.

A low p-value (typically ≤ 0.05) indicates that the observed data are unlikely under the null hypothesis, leading to the rejection of the null hypothesis in favor of the alternative hypothesis.