

Descriptive statistics:

1.What is the purpose of descriptive statistics?

The purpose of descriptive statistics is to summarize and describe the main features of a dataset, providing insights into its key characteristics. This helps in better understanding and interpreting the underlying patterns and trends within the large data.

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

The mean, also known as the average, is calculated by adding up all the values in a dataset and then dividing the sum by the total number of values. The mean is sensitive to extreme values (outliers) in the dataset. If there are extreme values, they can significantly influence the mean. The median is the middle value of 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. The median is not affected by extreme values (outliers) and is a measure of the central position that divides the dataset into two equal halves.

The mode is the value that appears most frequently in a dataset. The mode is not affected by extreme values (outliers).

3.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 set of values. It provides a way to quantify how much individual data points in a dataset deviate from the mean (average) of the dataset. If the standard deviation is small, it indicates that the data points tend to be close to the mean. In other words, there is little variability or spread in the data. A large standard deviation suggests that the data points are spread out over a wider range from the mean. This indicates greater variability or dispersion in the data.

4.Describe the concept of skewness in statistics.

A distribution is said to be skewed when the mean and the median fall at different points in the distribution and the data points are more towards one side or the other(left or right).

There are two types of skewness:

1. Right or positively skewed ($\text{mean} - \text{mode} > 0$)
2. Left or negatively skewed ($\text{mean} - \text{mode} < 0$)

Inferential Statistics:

5.What is the main goal of inferential statistics?

The main goal of inferential statistics is to draw conclusions or make inferences about a population based on a sample of data from that population. Various techniques of inferential statistics are population inference, hypothesis testing, cross validation.

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

The population is the entire group of individuals, items, or observations about which the researcher wants to draw conclusions. It includes all possible subjects that share a common characteristic. It includes all the elements from the data set.

A sample is a subset of the population that is selected for a detailed study. It consists of a smaller group of individuals or observations chosen to represent the larger population.

7.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. It provides a level of confidence regarding the precision of the estimate or assumption

8.Define p-value

The P-value is known as the probability value. It is defined as the probability of getting a result that is either the same or more extreme than the actual observations. P-value tells you how likely it is to get the observed data if the null hypothesis is true. The null hypothesis typically states that there is no effect, no difference, or no relationship in the population being studied.

A low p-value ($p < 0.05$) suggests that the observed results are unlikely to occur by random chance alone. This often leads to rejecting the null hypothesis and concluding that there is evidence for a real effect or difference.

A high p-value ($p > 0.05$) suggests that the observed results are likely to happen by random chance, and there is not enough evidence to reject the null hypothesis.