

# Descriptive Statistics:

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

One of the fundamental aspects of statistics is descriptive statistics, which is used to describe and summarize a set of data. The purpose of descriptive statistics is to provide an overview of the data under study, and this is achieved through various measures such as measures of central tendency, measures of variability, and graphical displays.

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

The difference between mean, median and mode are:

Mean is the average value of the given observations Median is the middle value of the given observations Mode is the most repeated value in the given observation

```
In [1]: #Mean(It is the sum of all given numbers(data) divide by the total number of  
import numpy as np  
a=[1,2,3,4,5,6,76,53,7,4,5,6,6]  
print(np.mean(a))  
#median(It is the middlest term or number when the given numbers are arranged)  
import statistics  
print(statistics.median(a))  
#Mode(It is the number which is repeated most times in the given data)  
print(statistics.mode(a))
```

13.692307692307692

5

6

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

A standard deviation (or  $\sigma$ ) is a measure of how dispersed the data is in relation to the mean. Low, or small, standard deviation indicates data are clustered tightly around the mean, and high, or large, standard deviation indicates data are more spread out. A standard deviation close to zero indicates that data points are very close to the mean, whereas a larger standard deviation indicates data points are spread further away from the mean.

```
In [2]: #standard deviation(square root of variance)
a=[1,2,3,4,5,6,7,6,5,3,7,4,5,6,6]
print(np.std(a))
```

22.18934122287828

## 4. Describe the concept of skewness in statistics.

Skewness is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side are not mirror images. Zero Skew: When a distribution has zero skew, it is symmetrical. Its left and right sides are mirror images.  $\text{mean} = \text{median}$

Right Skewed: A right-skewed distribution is longer on the right side of its peak than on its left. Right skew is also referred to as positive skew.  $\text{mean} > \text{median}$

Left Skewed (negative skewed): A left-skewed distribution is longer on the left side of its peak than on its right. In other words, a left-skewed distribution has a long tail on its left side. Left skew is also referred to as negative skew.  $\text{mean} < \text{median}$ . Calculate Skewness: Pearson's median skewness =  $3 * ((\text{Mean} - \text{Median}) / \text{standard deviation})$

## Inferential statistics

## 5. What is the main goal of inferential statistics?

Inferential statistics can be defined as a field of statistics that uses analytical tools for drawing conclusions about a population by examining random samples. The goal of inferential statistics is to make generalizations about a population. In inferential statistics, a statistic is taken from the sample data (e.g., the sample mean) that used to make inferences about the population parameter (e.g., the population mean).

Inferential statistics can be classified into hypothesis testing and regression analysis. Hypothesis testing also includes the use of confidence intervals to test the parameters of a population. Given below are the different types of inferential statistics.

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

A population is the entire group that you want to draw conclusions about. Example: Total citizens in India.

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. Example: Taking a group of 500 citizen in India.

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

A Confidence Interval is a range of reasonable values for a population statistic based on sample data. It measures uncertainty around a Point Estimate of a population statistic.

Confidence Interval uses the level of confidence, generally expressed as a percentage, and it represents the probability that the confidence interval will contain the true population parameter/statistic. For example, a Confidence Interval with a 95% confidence level indicates a 95% probability that the actual population means will be within our confidence interval

Confidence Interval can be computed using the below formula -

## **8. Define p-value**

In statistics, the p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. The p-value serves as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected. A smaller p-value means that there is stronger evidence in favor of the alternative hypothesis.

The lower the p-value, the greater the statistical significance of the observed difference.

A p-value of 0.05 or lower is generally considered statistically significant.