Statistics

Descriptive Statistics

- The primary goal of descriptive statistics is to organize, summarize, and present data in a meaningful way to facilitate understanding
- Descriptive statistics consists of three basic categories of measures: measures of central tendency, measures of variability (or spread), and frequency distribution.
- Measures of central tendency describe the center of the data set (mean, median, mode).
- Measures of variability describe the dispersion of the data set (variance, standard deviation).
- Measures of frequency distribution describe the occurrence of data within the data set (count).

Inferential Statistics:

- Inferential statistics helps to develop a good understanding of the population data by analyzing the samples obtained from it.
- It helps in making generalizations about the population by using various analytical tests and tools.
- It classified into hypothesis testing and regression analysis

- Measures of Central Tendency

Mean:

The mean (or average) is the most popular and well known measure of central tendency.The mean is equal to the sum of all the values in the data set divided by the number of values in the dataset

$$\overline{X} = \sum X \div n$$

Median

The median is the middle score for a set of data that has been arranged in order of magnitude . The median is less affected by outliers and skewed data

Mode

The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram

Range:

Range is a measure of dispersion, not central tendency. It represents the difference between the maximum and minimum values in the dataset.

-Measures of Dispersion or Variability

Variance

It measures the spread or dispersion of a set of data points from the mean value.

Steps to Calculate Variance:

1. Calculate mean
2. Subtract each value in set from mean
3. Square each number from 2)
4. Sum the values from 3)
5. Divide by the number of values in the set

$$S^2=(1/n)\sum (Xi-\bar{X})^2$$

The standard deviation

is a measure of the amount of variation or dispersion in a set of values. It tells you how much individual values in a dataset differ from the mean (average) value of the dataset

A low standard deviation indicates that the data points tend to be close to the mean of the dataset.

A high standard deviation indicates that the data points are spread out over a wider range of values.

It is the square root of the variance

Quartiles

Quartiles are like dividing that list into four equal parts.

1. **First Quartile (Q1)**:

   Q1 is the value that's right in the middle of the first half of the list. It's the point where 25% of the numbers are below it, and 75% are above it.

   It's kind of like finding the cutoff point for the bottom 25% of the list.

2. **Second Quartile (Q2)**:

   Q2 is the median of the whole list. It's the value right in the middle, where half of the numbers are below it and half are above it.

   It's like finding the middle point of the list.

3. **Third Quartile (Q3)**:

   Q3 is the value that's right in the middle of the second half of the list. It's the point where 75% of the numbers are below it, and only 25% are above it.

   It's similar to finding the cutoff point for the top 25% of the list.