

## Descriptive Statistics(04-07-2024)

Descriptive statistics summarize and organize the characteristics of a data set. They provide simple summaries about the sample and the measures. Descriptive statistics are divided into measures of central tendency and measures of variability (spread).

### Measures of Central Tendency

1. **Mean:** The average of the data set.

Mean=sum of observation/No. of observation

data= 10,20,30,40,50

Mean=  $(10+20+30+40+50)/5=30$

2. **Median:** The middle value when the data set is ordered. If the data set has an even number of observations, the median is the average of the two middle numbers.

a)data=10,20,30,40,50

median=30

b)data=10,20,30,40,50,60

median= $(30+40)/2=35$

3. **Mode:** The most frequently occurring value in the data set.

data=1,3,4,6,7,3,3,5,10

mode=3

### Measures of Variability (Spread)

1. **Range:** The difference between the highest and lowest values in the data set.

Range=Maximum Value–Minimum Value

2. **Variance:** The average of the squared differences from the mean. It measures how far each number in the set is from the mean and thus from every other number in the set.

$$\text{Variance}(\sigma^2) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

3. **Standard Deviation:** The square root of the variance. It provides a measure of the average distance from the mean.

$$\text{Standard Deviation}(\sigma) = \sqrt{\text{Variance}}$$

Q) Find standard deviation and variance of following:

5,9,8,12,16,10,6,8

$$\text{Mean} = (5+9+8++12+16+10+6+8)/8=8$$

$$\begin{aligned} \text{Variance} &= ((5-8)^2 + (9-8)^2 + (8-8)^2 + (12-8)^2 + (16-8)^2 + (10-8)^2 + (6-8)^2 + (8-8)^2)/8 \\ &= 4.75 \end{aligned}$$

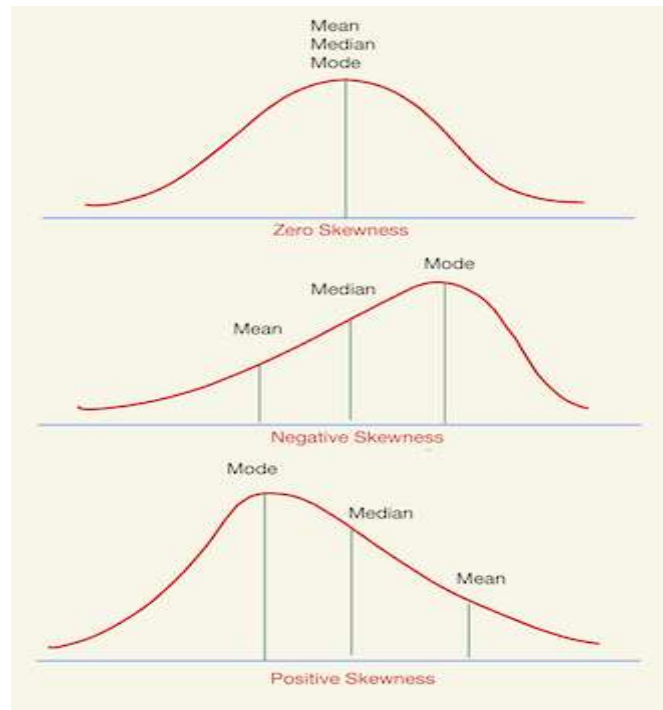
$$\text{Standard deviation} = \sqrt{4.75} = 2.18$$

4. **Interquartile Range (IQR):** The difference between the first quartile (Q1) and the third quartile (Q3). It measures the spread of the middle 50% of the data.

$$\text{IQR} = \text{Q3} - \text{Q1}$$

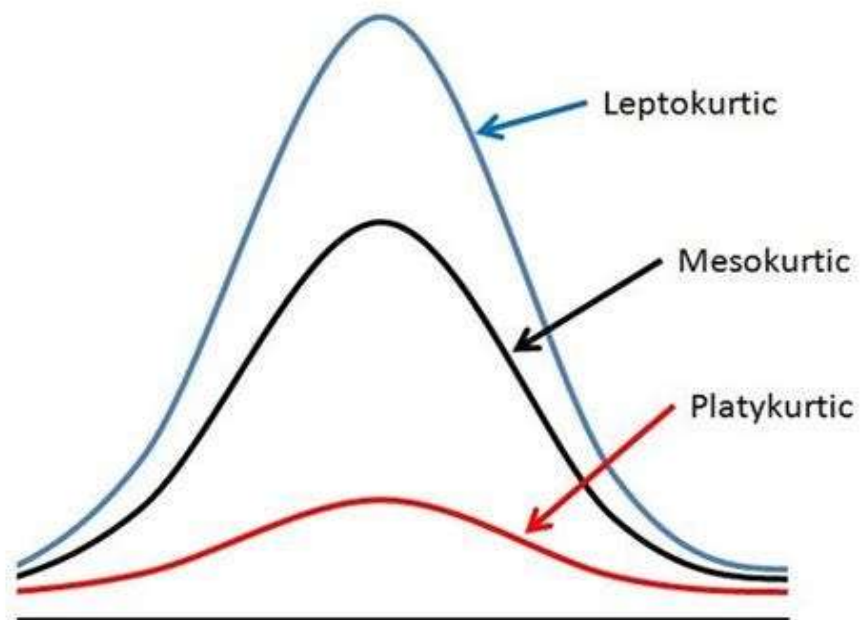
### Other Descriptive Statistics

1. **Skewness:** Measures the asymmetry of the data distribution.
  - **Positive Skew:** Tail on the right side is longer or fatter.
  - **Negative Skew:** Tail on the left side is longer or fatter.



2. **Kurtosis:** Measures the "tailedness" of the data distribution.

- **Leptokurtic:** Higher peak and fatter tails.
- **Platykurtic:** Lower peak and thinner tails.
- **Mesokurtic:** Similar to a normal distribution.



## Graphical Representations

1. **Histogram:** A graphical representation showing the frequency distribution of the data.
2. **Box Plot:** Summarizes data using the median, quartiles, and extremes.
3. **Bar Chart:** Represents categorical data with rectangular bars.
4. **Pie Chart:** Circular chart divided into sectors, illustrating numerical proportions.
5. **Scatter Plot:** Plots individual data points on a two-dimensional graph.
6. **Line Graph:** Displays data points connected by straight lines, showing trends over time.
7. **Correlation Matrix:** A table showing correlation coefficients between variables, indicating the strength and direction of relationships.

## Importance of Descriptive Statistics

- Simplifies large amounts of data in a sensible way.
- Identifies patterns and trends in the data.
- Provides a basis for inferential statistics, which further analyze the data.