

Descriptive stats

Wednesday, February 25, 2026 11:40 AM

Definition

Methods used to summarize or organize the data using numbers or graphs

Types of Descriptive Statistics

1. Where the data centers (Measure of Central Tendency)
2. How spread out the data is (Measure of Variability)
3. How the data is distributed (Measures of Frequency Distribution)

Measures of central tendency

Wednesday, 25 February 2026 03:00

1- Mean

Definition:

The mean can be calculated by summing all values present in the sample divided by total number of values present in the sample or population

Formula:

Sum of values / number of values

Example:

Values : 10, 50, 60, 40

Mean = 40

The mean works well for symmetric data because values balance around the center but it becomes misleading with outliers because extreme values pull the average toward them.

2- Median

Definition:

The median is the middle of a dataset when arranged from lowest to highest or highest to lowest in order to find the median, the data must be sorted. For an odd number of data points the median is the middle value and for an even number of data points median is the average of the two middle values.

Example:

Values : 10, 50, 60, 40

Median = 55

The median is reliable when data contains outliers because it depends on position rather than magnitude, but it ignores how far values are from the center

3- Mode

Definition:

The most frequently occurring value in the Sample or Population

Example:

Values : 55, 60, 65, 55, 70, 75, 55

Mode = 55

The mode is useful for identifying the most frequent value, especially in categorical data but it may not represent the overall dataset if values vary widely or multiple modes exist

Measure of Variability

Wednesday, February 25, 2026 11:50 AM

1- Range

Definition:

Range is the difference between the maximum and minimum values of the Sample.

Example:

Values: 44,55,66,77,88

$$\text{Range} = 88 - 44 = 44$$

The range gives a quick idea of total spread

but it is misleading because it depends only on the minimum and maximum values and ignores the rest of the data

2- Variance

Definition:

Variance is a measure of how spread-out values from the mean by measuring the dispersion around the Mean

Formula:

$$\sigma^2 = \sum(X-\mu)^2/n$$

Example:

Values: 2,4,6

Mean = 4

$$\text{Variance} = 2.67$$

Variance measures overall variability mathematically precisely but it is difficult to interpret because its unit is squared

3- Standard Deviation

Definition:

Standard Deviation is the square root of variance. The measuring unit of S.D. is same as the Sample values' unit. It indicates the average distance of data points from the mean and is widely used due to its intuitive interpretation

Formula:

$$\sigma = (\sigma^2)^{0.5}$$

From previous example:

Variance = 2.67

$$\text{Standard deviation} = (2.67)^{0.5} = 1.63$$

Standard deviation clearly shows the average distance from the mean in the original unit
but it is sensitive to outliers and skewed distributions

4- Quartiles

Definition:

divides the dataset into four equal parts

Q1 (First Quartile): Median of the lower 50% of the dataset (25th percentile).

Q2 (Second Quartile / Median): Median of the entire dataset (50th percentile).

Q3 (Third Quartile): Median of the upper 50% of the dataset (75th percentile).

Example:

Values: 10, 20, 30, 40, 50, 60, 70

$Q2 = 40$

$Q1 = 20$

$Q3 = 60$

Quartiles divide data into four equal parts and help detect spread and outliers
but results may vary slightly depending on the calculation method especially in small datasets.

5- Interquartile Range (IQR):

Definition:

The range between the first quartile (Q1) and the third quartile (Q3)

From the previous Example:

$$\text{IQR} = Q3 - Q1 = 60 - 20 = 40$$

- Outliers

[$Q1 - 1.5\text{IQR}$, $Q3 + 1.5\text{IQR}$] -- any number outside this range is outlier