

Mean, Median, Mode and Standard Deviation

mean, median, mode, and standard deviation are valuable tools for summarizing and analyzing data in various fields, providing insights into central tendency, variability, and distribution of data. They help decision-makers make informed choices, researchers draw conclusions, and businesses understand trends and patterns in their operations.

1. Mean:

- The mean, often referred to as the average, is the sum of all values in a dataset divided by the total number of values.
- Formula: Mean = (Sum of all values) / (Number of values)
- Example: For the dataset [1, 2, 3, 4, 5], the mean is $(1 + 2 + 3 + 4 + 5) / 5 = 3$.

2. Median:

- The median is the middle value of a sorted dataset. If there is an odd number of values, it is the value at the center position. If there is an even number of values, it is the average of the two middle values.
- Example: For the dataset [1, 2, 3, 4, 5], the median is 3. For the dataset [1, 2, 3, 4], the median is $(2 + 3) / 2 = 2.5$.

3. Mode:

- The mode is the value that appears most frequently in a dataset. A dataset may have one mode (unimodal), multiple modes (multimodal), or no mode if all values occur with the same frequency.
- Example: For the dataset [1, 2, 2, 3, 4], the mode is 2.

4. Standard Deviation:

- The standard deviation measures the amount of variation or dispersion of a dataset from the mean. A low standard deviation indicates that the data points tend to be close to the mean, while a high standard deviation indicates that the data points are spread out over a wider range.
- Formula: Standard Deviation = $\sqrt{(1 / N) * \sum (x_i - \mu)^2}$, where N is the number of values, x_i is each value in the dataset, and μ is the mean of the dataset.
- Example: For the dataset [1, 2, 3, 4, 5], with a mean of 3, the standard deviation is calculated as $\sqrt{((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2) / 5} \approx 1.41$.

Differences:

- **Mean** represents the average value of a dataset and is sensitive to outliers.
- **Median** represents the middle value of a dataset and is less affected by extreme values compared to the mean.
- **Mode** represents the most frequently occurring value(s) in a dataset.
- **Standard deviation** measures the dispersion of data points around the mean and provides insight into the variability of the dataset.