

Measures of central tendency

help you find the middle, or the average, of a dataset. The 3 most common measures of central tendency are the mode, median, and mean.

- **Mode:** the most frequent value.
- **Median:** the middle number in an ordered dataset.
- **Mean:** the sum of all values divided by the total number of values.

1- Mean

The mean, also commonly referred to as the average, is a statistical measure that represents the central tendency of a set of data. It essentially tells you what the "typical" value in your data is.

Formula for the Mean:

The mean is calculated by summing all the values in your data set and then dividing that sum by the total number of values (n). Here's the mathematical formula:

$$Mean = (\sum x_i) / n$$

Where:

- Σ (sigma) represents the sum of all the values (x_i) in your data set.
- x_i represents each individual value in your data set (i goes from 1 to n).
- n represents the total number of values in your data set.

For example, if you have a data set of {2, 4, 6, 8}, the mean would be:

$$Mean = (2 + 4 + 6 + 8) / 4 = 20 / 4 = 5$$

Hint : Suppose this data represents the number of students coming to school each day. The number 5 indicates that 5 students come to school daily.

2. Median

Here's how to find the median:

1. **Order your data:** Arrange all the data points in your set from least to greatest (ascending order).
2. **Identify the middle:**
 - If you have an **odd number of values**, the median is the **middle value** in the ordered list.
 - If you have an **even number of values**, the median is the **average** of the two middle values.

For example

consider the data set {2, 4, 6, 8, 10}:

1. Ordered data: {2, 4, 6, 8, 10}
2. Median: Since we have an odd number (5) of values, the middle value (6) is the median.

Why Use the Median?

The median offers some advantages over the mean, especially when dealing with skewed data or outliers.

- **Resistance to outliers:** The median is less influenced by extreme values in the data set. Imagine a data set about incomes: {20000, 30000, 40000, 1000000}. The very high value (exceptionally high earner) would significantly inflate the mean, but the median (around 30000) would be a better representation of typical incomes.

3. Mode

Finding the Mode:

- There's no specific formula for the mode, but you can find it by simply examining your data and identifying the value that repeats the most.

Here are some things to consider about the mode:

- **Multiple Modes:** A data set can have one mode (unimodal), two modes (bimodal), or even more (multimodal) if there are multiple values with the highest frequency.
- **No Mode:** In some cases, a data set might not have a mode at all if all the values appear with the same frequency (no repetitions).

Example:

Consider the data set: {2, 4, 5, 5, 6, 8, 5}. In this case, the number 5 appears three times, more than any other value. Therefore, 5 is the mode of this data set.

Why Use the Mode?

The mode can be useful in certain situations:

- **Identifying the most popular choice:** Imagine a survey on favorite colors. The mode would tell you the color chosen by the most people.
- **Understanding common categories:** In data on clothing sizes, the mode could indicate the most frequent size (e.g., medium).