04. Measure Of Central Tendency

Mahir Faysal Tusher

June 2025

Overview

Measures of central tendency are statistical measures used to determine the center or typical value of a dataset. The three main types are:

1. Mean

2. Median

3. Mode

1 Mean (Average)

Population Mean

Given a population of size N, the mean (denoted by μ) is calculated as:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Sample Mean

Given a sample of size n, the sample mean (denoted by $\bar{x}$) is:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Example

Let the values be:
$$x = \{1, 3, 4, 5\}$$

This is a sample of size $n = 4$.

$$\bar{x} = \frac{1 + 3 + 4 + 5}{4} = \frac{13}{4} = 3.25$$

Note: Mean is sensitive to outliers. For example, adding 100 to the above dataset:
$$x = \{1, 3, 4, 5, 100\} \Rightarrow \bar{x} = \frac{113}{5} = 22.6$$

This drastically shifts the mean due to the presence of an outlier.

2 Median

The median is the middle value of a sorted dataset. It is robust to outliers.

Steps to Calculate Median

1. Sort the dataset.

2. If number of elements (n) is **odd**, median is the middle element:

$$\text{Median} = x_{\left(\frac{n+1}{2}\right)}$$

3. If n is **even**, median is the average of the two middle elements:

$$\text{Median} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Examples

- Odd-sized dataset:

$$x = \{1, 3, 4, 5, 100\} \Rightarrow \text{Median} = 4$$

- Even-sized dataset:

$$x = \{1, 3, 4, 5, 100, 200\} \Rightarrow \text{Median} = \frac{4 + 5}{2} = 4.5$$

3   Mode

The mode is the value that appears most frequently in a dataset.

Example

Let the dataset be:

$$x = \{4, 3, 2, 1, 1, 4, 4, 5, 2, 100\}$$

The value 4 appears the most (3 times), so:

$$\text{Mode} = 4$$

Properties

- The mode is useful in categorical data.

- It is not affected by outliers.

- A dataset can be unimodal (1 mode), bimodal (2 modes), or multi-modal.

Conclusion

- **Mean** is a good general-purpose measure but sensitive to outliers.

- **Median** is robust to outliers and better represents skewed distributions.

- **Mode** captures the most frequent value and is especially useful in non-numeric data.

Understanding these three measures provides a strong foundation for further statistical topics like dispersion and data distribution.