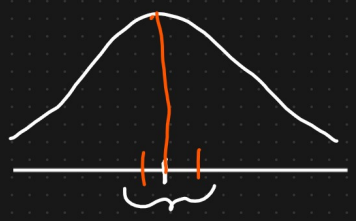
MEAN MEDIAN AND MODE

① Measure of Central Tendency

① Mean or Average
② Median
③ Mode



① Mean

The mean is commonly used to represent the central tendency of a dataset. It provides a single value that summarizes the entire dataset and is useful for making comparisons or generalizations.

Population (N) | Sample (n)

$X = \{1, 1, 2, 2, 3, 3, 4, 5, 5, 6\}$

Population mean $(\mu) = \frac{\sum_{i=1}^{n} x_i}{N}$ | Sample mean $(\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n}$

$= \frac{[1+1+2+2+3+3+4+5+5+6]}{10}$ =

$= \frac{32}{10} = 3.2$ //.

② Median

The median is the middle value in a dataset when the values are arranged in ascending or descending order.

$X = \{4, 5, 2, 3, 2, 1\}$.

The median is less sensitive to outliers compared to the mean and is therefore often used when the dataset contains extreme values or is skewed.

Steps

① Sort the Random Variable $\{1, 2, 2, 3, 4, 5\}$.

② No. of elements Count = 6

③ if Count == even

{ 1, 2, 2, 3, 4, 5 }

⇓

Median = $\frac{2+3}{2} = 2.5$

④ if Count is Odd

{ 1, 2, 2, 3, 4, 5, 6 }

Median = 3

If 'n' is odd: $\left(\frac{n+1}{2}\right)^{th}$ term

If 'n' is even: $\frac{\left(\frac{n}{2}\right)^{th}\text{term} + \left(\frac{n}{2}+1\right)^{th}\text{term}}{2}$

Why Median?

Outlier

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

$\bar{x} = \frac{1+2+3+4+5}{5}$ $\bar{x} = \frac{1+2+3+4+5+100}{6}$

$\bar{x} = \frac{15}{5} = 3$ → $\frac{115}{6} \approx 19.$

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

Median = $\frac{3+4}{2} = 3.5$

Median is used to find the central Tendency when outliers is present.

③ Mode : Frequency Maximum

{ 2, 1, 1, 1, 4, 5, 7, 8, 9, 9, 10 }

The mode is the value that appears most frequently in a dataset. A dataset can have one mode (unimodal), multiple modes (multimodal), or no mode (no repeated values).

Use Case: The mode is useful for identifying the most common or popular value in a dataset. It is commonly used in categorical data analysis, such as identifying the most common answer in a survey.

Mode = 1

Use In EDA AND Feature Engineering

Missing values

Use Cases:

1.In financial analysis, the mean is used to calculate the average return on investment.

2. In healthcare, the median household income is often used to assess the socioeconomic status of a community.

3. In marketing, the mode is used to identify the most popular product or service among consumers.

In summary, mean, median, and mode are statistical measures used to describe the central tendency or typical value of a dataset. Each measure has its own use cases and applications, depending on the characteristics of the data and the goals of the analysis.