**Measures of Central Tendency**

**Introduction:**

Measures of central tendency help summarize a dataset by identifying its center. These measures are essential in statistical analysis as they provide insights into the general distribution of data. The three main measures are:

1. **Mean**: The arithmetic average of all values in a dataset.
2. **Median**: The middle value when data is arranged in ascending order.
3. **Mode**: The most frequently occurring value in the dataset.

Each of these measures has its significance and is used depending on the nature of the dataset.

**Given Dataset: [3, 5, 5, 6, 7, 100]:**

We will calculate the mean, median, and mode step by step.

**1. Mean (Arithmetic Average):**

The mean is calculated by summing all values and dividing by the number of values:

Mean=3+5+5+6+7+1006=1266=21\text{Mean} = \frac{3 + 5 + 5 + 6 + 7 + 100}{6} = \frac{126}{6} = 21

The mean represents the overall average but can be significantly affected by extreme values (outliers).

**2. Median (Middle Value):**

To find the median, we first sort the dataset:

[3,5,5,6,7,100] [3, 5, 5, 6, 7, 100]

Since we have an even number of values (6), the median is the average of the two middle numbers:

Median=5+62=112=5.5\text {Median} = \frac{5 + 6}{2} = \frac{11}{2} = 5.5

The median is **less sensitive to outliers** and represents the central value better in skewed datasets.

**3. Mode (Most Frequent Value):**

The mode is the number that appears most frequently. In this dataset, **5** appears twice, making it the mode:

Mode=5\text {Mode} = 5

A dataset can have **one mode (unimodal), multiple modes (multimodal), or no mode at all** if no value repeats.

**Comparison of Measures of Central Tendency;**

|  |  |  |
| --- | --- | --- |
| **Measure** | **Value** | **Characteristics** |
| **Mean** | 21 | Affected by extreme values (outlier: 100) |
| **Median** | 5.5 | Less influenced by outliers, good for skewed data |
| **Mode** | 5 | Represents the most frequently occurring value |

**Interpretation of Results:**

* The **mean (21)** is much higher than most values in the dataset due to the presence of an extreme outlier (**100**). This shows that the mean is **sensitive to outliers** and may not accurately represent the dataset’s central value.
* The **median (5.5)** provides a better central value for this dataset, as it is unaffected by extreme values. It is more reliable in skewed distributions.
* The **mode (5)** represents the most common value, giving another measure of central tendency that aligns closely with most of the data points.

**Why Do We Use Different Measures?**

Each measure of central tendency serves different purposes in data analysis:

* **Mean** is useful when data is **normally distributed** without extreme values.
* **Median** is preferable when data is **skewed or contains outliers**.
* **Mode** is helpful in **categorical data** where numerical averages are not meaningful (e.g., survey responses: “Agree,” “Neutral,” “Disagree”).

**Real-World Applications:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Mean Example** | **Median Example** | **Mode Example** |
| **Education** | Average exam score in a school | The middle score of a class | Most common grade received |
| **Business** | Average sales per day | Median income of employees | Most sold product in a store |
| **Health** | Average body temperature | Median age of patients | Most common blood type |

**Conclusion:**

* The **mean** is useful for general analysis but can be misleading when outliers are present.
* The **median** provides a more **accurate measure** of central tendency in **skewed datasets**.
* The **mode** is helpful in identifying the **most frequent value** in a dataset.

In this dataset, the **median and mode provide a better summary of the central tendency** than the mean, which is skewed by the extreme value **100**. Understanding these differences is essential for effective data interpretation in various real-world applications.