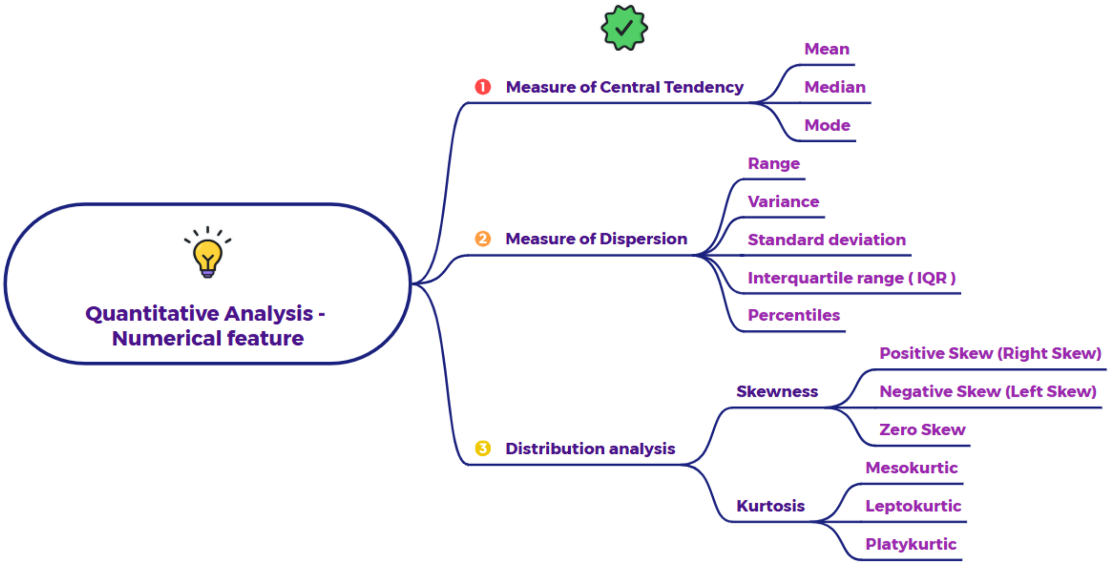
Explain measure of central tendency



Imagine we have the following set of scores obtained by 10 students on a recent math quiz (out of 20):

12, 15, 18, 16, 15, 10, 15, 19, 14, 15

1. Mean: The Arithmetic Average

The **mean** is the most commonly used measure of central tendency. It's calculated by summing all the values in the dataset and then dividing by the total number of values.

Calculation:

Sum of scores = 12+15+18+16+15+10+15+19+14+15=149 Number of scores = 10

$$\text{Mean} = \frac{\text{Sum of scores}}{\text{Number of scores}} = \frac{149}{10} = 14.9$$

Interpretation:

The mean score on the math quiz is 14.9. This value represents the average performance of the students. It's a good overall summary of the dataset, assuming the scores are relatively evenly distributed without extreme outliers.

When to Use:

The mean is most appropriate when the data is numerical, and the distribution is roughly symmetrical without significant outliers. It uses all the data points in its calculation, making it sensitive to every value.

Potential Drawback:

The mean can be heavily influenced by extreme values (outliers). For example, if one student had scored a 0 due to illness, the mean would drop significantly, potentially not representing the typical performance of the other students.

2. Median: The Middle Value

The **median** is the middle value in a dataset that has been ordered from the smallest to the largest. It divides the data into two equal halves – half the values are below the median, and half are above it.

Calculation:

First, we need to arrange the scores in ascending order:

10, 12, 14, 15, 15, 15, 15, 16, 18, 19

Since we have an even number of scores (10), the median is the average of the two middle values. The two middle values are the 5th and 6th values in the ordered list, which are both 15.

$$\text{Median} = \frac{15+15}{2} = \frac{30}{2} = 15$$

Interpretation:

The median score on the math quiz is 15. This means that half of the students scored below 15, and half scored above 15.

When to Use:

The median is particularly useful when the dataset contains outliers or is skewed. Because it only considers the position of the values, extreme high or low scores don't affect it as much as they affect the mean. It provides a more robust measure of central tendency in such cases.

Advantage over Mean in Certain Situations:

If, in our example, one student had scored a very low score, say 2, the ordered list would be:

2, 10, 12, 14, 15, 15, 15, 15, 16, 18, 19

(Assuming we had 11 students now for illustration). The median would still be 15 (the 6th value), but the mean would be significantly lower due to the outlier.

3. Mode: The Most Frequent Value

The **mode** is the value that appears most frequently in the dataset. A dataset can have no mode (if all values are unique), one mode (unimodal), or more than one mode (bimodal, trimodal, etc.).

Calculation:

Looking at our original set of scores:

12, 15, 18, 16, 15, 10, 15, 19, 14, 15

We can see that the score of 15 appears four times, which is more than any other score.

Mode = 15

Interpretation:

The mode score on the math quiz is 15. This indicates that the score of 15 was the most common score achieved by the students.

When to Use:

The mode is most useful for categorical data, but it can also be informative for numerical data, especially when you want to know the most typical or frequent value. It's easy to understand and identify.

Considerations:

- A dataset might not have a mode if all values occur only once.
- A dataset can have multiple modes if several values have the same highest frequency.

In Summary:

- The **mean** provides the average value, sensitive to all data points.
- The **median** gives the middle value, robust to outliers and skewness.
- The **mode** identifies the most frequent value.

Choosing the most appropriate measure of central tendency depends on the nature of your data and what you want to highlight. For a symmetrical distribution without outliers, the mean is often a good choice. However, in the presence of skewness or outliers, the median can provide a more representative central value. The mode is useful for identifying the most common observation.