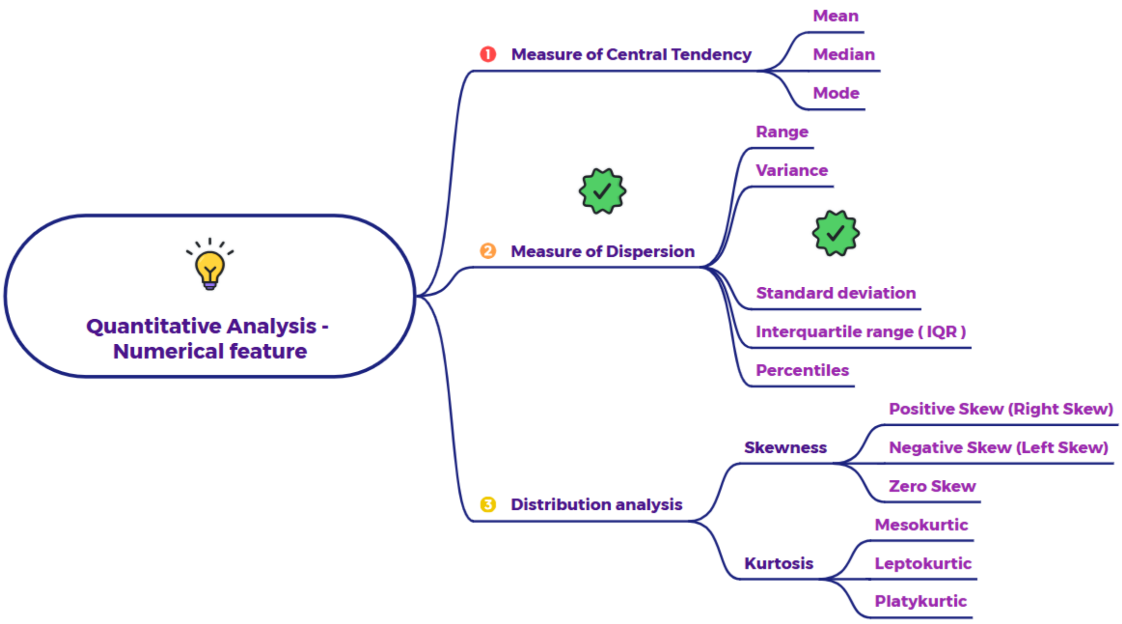
Explain measure of dispersion – Standard Deviation



Concept:

The **standard deviation** is simply the square root of the variance. By taking the square root, we convert the measure of spread back into the original units of the data, making it much easier to understand and compare. A lower standard deviation indicates that data points tend to be close to the mean, while a higher standard deviation indicates that the data points are more spread out over a wider range of values.

Formulas:

Corresponding to population variance and sample variance, we have population standard deviation (σ) and sample standard deviation (s):

- **Population Standard Deviation (σ)**: The square root of the population variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

- **Sample Standard Deviation (s)**: The square root of the sample variance.
  Detailed Examples:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Let's revisit our math quiz scores to calculate the standard deviation.

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

We previously calculated the sample variance (s2) for this data to be approximately 6.77.

Step 1: Take the Square Root of the Sample Variance

Sample Standard Deviation (s) = 6.77 ≈2.60

Interpretation:

The sample standard deviation of the math quiz scores is approximately 2.60. This means that, on average, the students' scores deviate from the mean score of 14.9 by about 2.60 points. This value gives us a more intuitive sense of the typical spread of the scores compared to the variance (6.77 squared points).

Example 2: Comparing the Spread of Two Datasets (Revisiting)

Recall the two sets of test scores with a mean of 70:

- **Set A:** 65, 75, 70, 68, 72 (Sample variance sA2=14.5)
- **Set B:** 40, 100, 70, 55, 85 (Sample variance sB2=562.5)

Let's calculate the sample standard deviation for each:

Set A:

Sample Standard Deviation (sA) = 14.5 ≈3.81

Set B:

Sample Standard Deviation (s_B) = 562.5 $\approx$23.72

Interpretation:

- For Set A, the scores typically deviate from the mean of 70 by about 3.81 points. This indicates a relatively tight clustering of scores around the average.
- For Set B, the scores typically deviate from the mean of 70 by a much larger margin of about 23.72 points. This confirms our earlier observation from the variance that the scores in Set B are much more spread out.

The standard deviation values (3.81 and 23.72) are easier to relate back to the original test score units, making the comparison of spread more intuitive than comparing the variance values (14.5 and 562.5).

Example 3: Real-World Application - Heights of Individuals

Suppose we have the heights (in cm) of two groups of adult males:

- **Group 1:** 175, 178, 180, 172, 175 (Mean = 176 cm, Sample Standard Deviation ≈ 2.92 cm)
- **Group 2:** 165, 190, 170, 185, 170 (Mean = 176 cm, Sample Standard Deviation ≈ 9.62 cm)

Interpretation:

Both groups have the same average height (176 cm). However, Group 1 has a much smaller standard deviation (2.92 cm) compared to Group 2 (9.62 cm). This tells us that the heights in Group 1 are more consistent and clustered around the average height, while the heights in Group 2 are more variable and spread out over a wider range.

Key Advantages of Standard Deviation:

- **Interpretable Units:** Standard deviation is expressed in the same units as the original data, making it easier to understand the typical spread.
- **Widely Used:** It is a fundamental measure in statistics and is used extensively in various analyses and models.

- **Basis for Other Statistics:** Standard deviation is a key component in calculating other important statistical measures like the coefficient of variation and standard error.

Relationship to Variance:

Standard deviation is simply the square root of the variance. While variance is a useful concept in the mathematical derivation of standard deviation and in certain statistical tests, the standard deviation is often preferred for descriptive purposes due to its intuitive units.

In summary, the standard deviation is a powerful and widely used measure of dispersion that quantifies the typical deviation of data points from the mean, expressed in the original units of the data. It provides a clear and interpretable understanding of the spread or variability within a dataset.