

Variance - Measuring Spread

Detailed Study Notes

Variance is a measure of dispersion that quantifies how much the values in a dataset differ from the mean. It's the foundation for many statistical concepts.

Definition:

Variance measures the average squared deviation from the mean. It shows how spread out the data is.

Formula:

Population Variance: $\sigma^2 = \Sigma(x - \mu)^2 / N$

Sample Variance: $s^2 = \Sigma(x - \bar{x})^2 / (n - 1)$

Detailed Example:

Dataset: 10, 20, 30, 40, 50

Step 1: Calculate the mean

Mean (μ) = $(10 + 20 + 30 + 40 + 50) / 5 = 30$

Step 2: Calculate deviations from mean

$(10 - 30) = -20$

$(20 - 30) = -10$

$(30 - 30) = 0$

$(40 - 30) = 10$

$(50 - 30) = 20$

Step 3: Square the deviations

$(-20)^2 = 400$

$(-10)^2 = 100$

$(0)^2 = 0$

$(10)^2 = 100$

$(20)^2 = 400$

Step 4: Sum squared deviations

Sum = $400 + 100 + 0 + 100 + 400 = 1000$

Step 5: Divide by n (for population)

Variance = $1000 / 5 = 200$

Properties of Variance:

1. Always non-negative (≥ 0)
2. Variance of 0 means all values are identical
3. Sensitive to outliers
4. Units are squared (e.g., if data is in meters, variance is in square meters)

Why Use $n-1$ for Sample Variance?

Using $n-1$ (Bessel's correction) provides an unbiased estimate of population variance from a sample.

Applications:

- Quality control: measuring consistency in manufacturing
- Finance: assessing risk and volatility
- Scientific research: evaluating experimental precision
- Engineering: analyzing signal noise