Standard Deviation - Understanding Variability

Detailed Study Notes

Standard deviation is the square root of variance and is one of the most important measures of spread. It's expressed in the same units as the original data.

Definition:
Standard deviation measures the typical distance of data points from the mean.

Formula:
Population SD: $\sigma = \sqrt{[\Sigma(x - \mu)^2 / N]}$
Sample SD: $s = \sqrt{[\Sigma(x - \bar{x}\blacksquare)^2 / (n - 1)]}$

Detailed Example:
Dataset: 10, 20, 30, 40, 50

From previous variance calculation:
Variance = 200

Standard Deviation = $\sqrt{200} \approx 14.14$

Interpretation:
On average, values deviate from the mean (30) by about 14.14 units.

The Empirical Rule (68-95-99.7 Rule):
For normally distributed data:
- 68% of values fall within 1 SD of the mean
- 95% of values fall within 2 SD of the mean
- 99.7% of values fall within 3 SD of the mean

Example Application:
If test scores have mean = 75 and SD = 10:
- 68% of students score between 65 and 85
- 95% of students score between 55 and 95
- 99.7% of students score between 45 and 105

Properties of Standard Deviation:
1. Same units as the original data

2. Always non-negative
3. More interpretable than variance
4. Commonly used in statistical tests

When to Use Standard Deviation:
- Describing spread in normally distributed data
- Comparing variability between datasets
- Identifying outliers (values > 2 or 3 SD from mean)
- Quality control (control charts)

Low vs High Standard Deviation:
- Low SD: Data points cluster closely around the mean
- High SD: Data points are spread out widely

Applications:
- Investment analysis: measuring portfolio risk
- Quality assurance: monitoring process consistency
- Education: comparing test score distributions
- Healthcare: evaluating treatment consistency