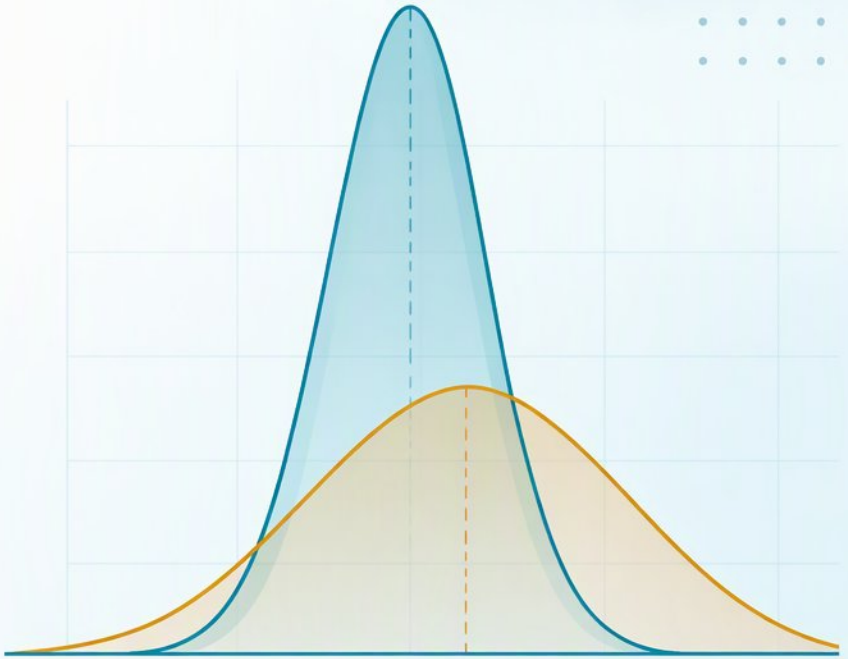
# Variance in Statistics:

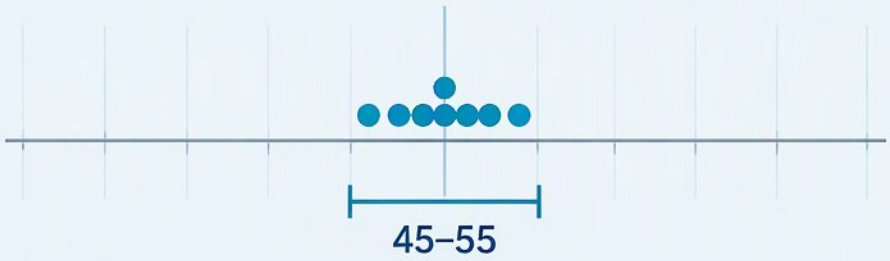
## Concepts, Purpose, and Examples

- Averages hide spread: same mean, very different variability
- Learn to compute variance by hand and in spreadsheets
- Choose population versus sample variance with confidence
- Interpret variance and standard deviation in real terms
- Scope: descriptive variance for students and early-career analysts

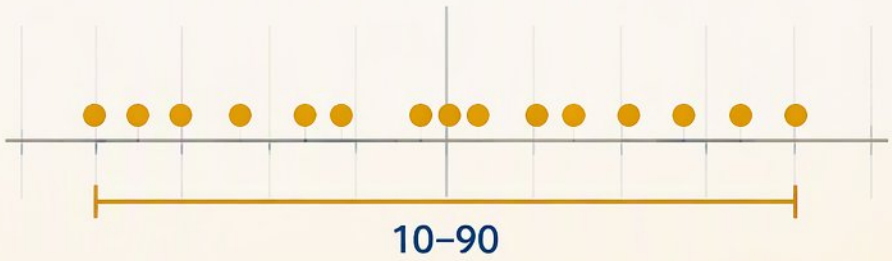


# Why Averages Alone Are Not Enough

### Same Average, Smaller Spread



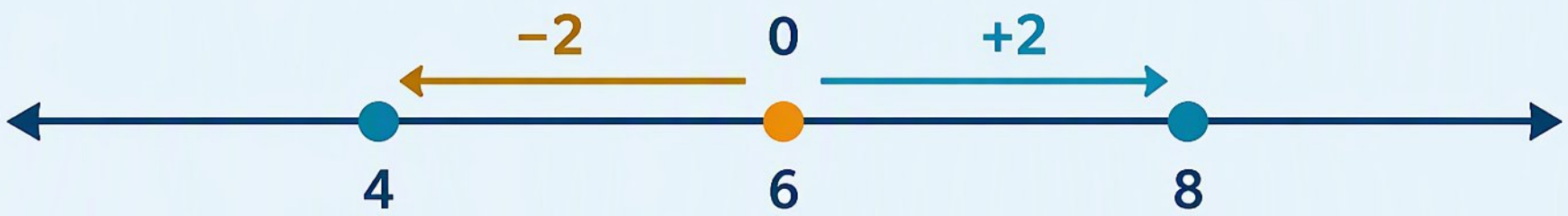

### Same Average, Larger Spread




- Center, shape, and dispersion are separate dataset properties; a mean shows only center
- Two datasets can both average 50: one ranges 45–55, the other 10–90
- Consistency matters: delivery reliability, quality control, performance stability, financial risk
- Reporting only an average can mislead when values vary widely

# Deviations from the Mean

- Deviation = one value minus the mean: $x - \bar{x}$
- Deviations always sum to zero in every dataset
- Example (4, 6, 8): mean 6, deviations −2, 0, +2
- Positives and negatives cancel, so average deviation is always zero
- Need a way to remove signs before averaging

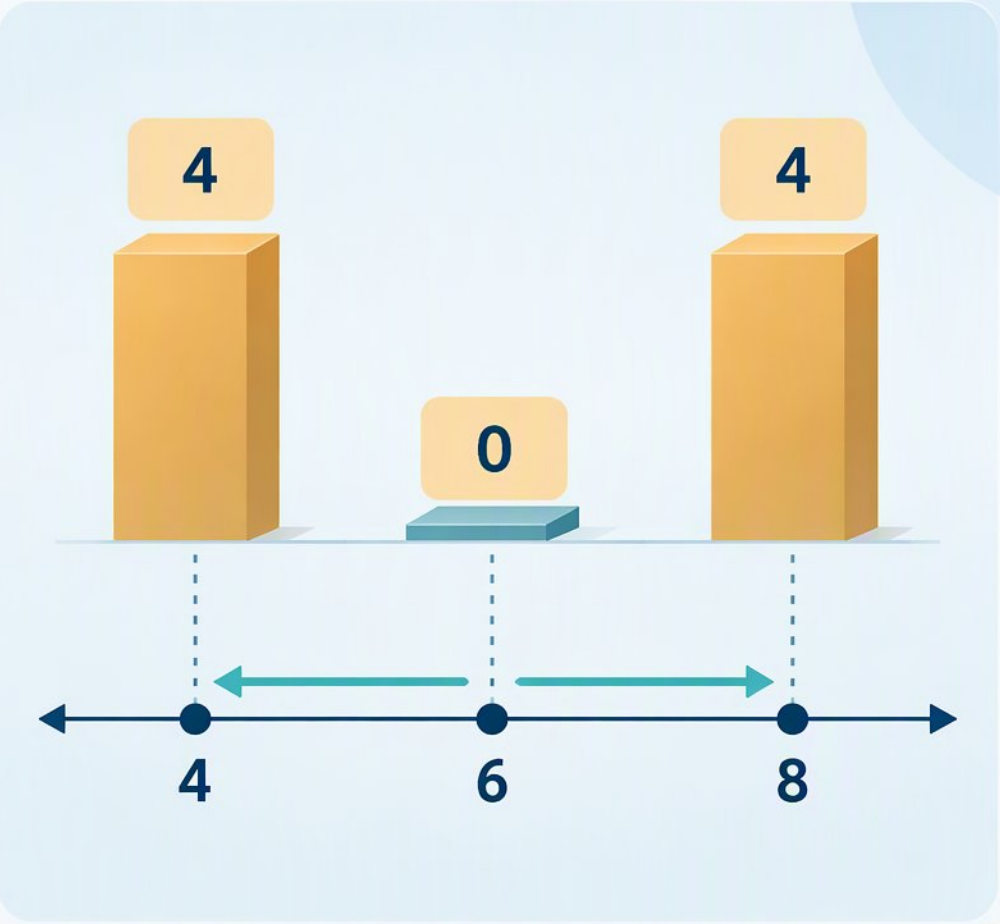



**Why it matters:**

A plain average deviation can never measure spread.

# Squaring Deviations and the Sum of Squares

- Squaring removes signs, so every deviation contributes positively
- Large deviations are weighted more heavily, so outliers dominate
- Example (4, 6, 8): squared deviations are 4, 0, 4
- Sum of squared deviations equals 8
- This sum of squares is the raw material for every variance formula



# The Variance Formula: Population vs. Sample

## Population Variance

Population variance: sum of squared deviations divided by $N$

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

## Sample Variance

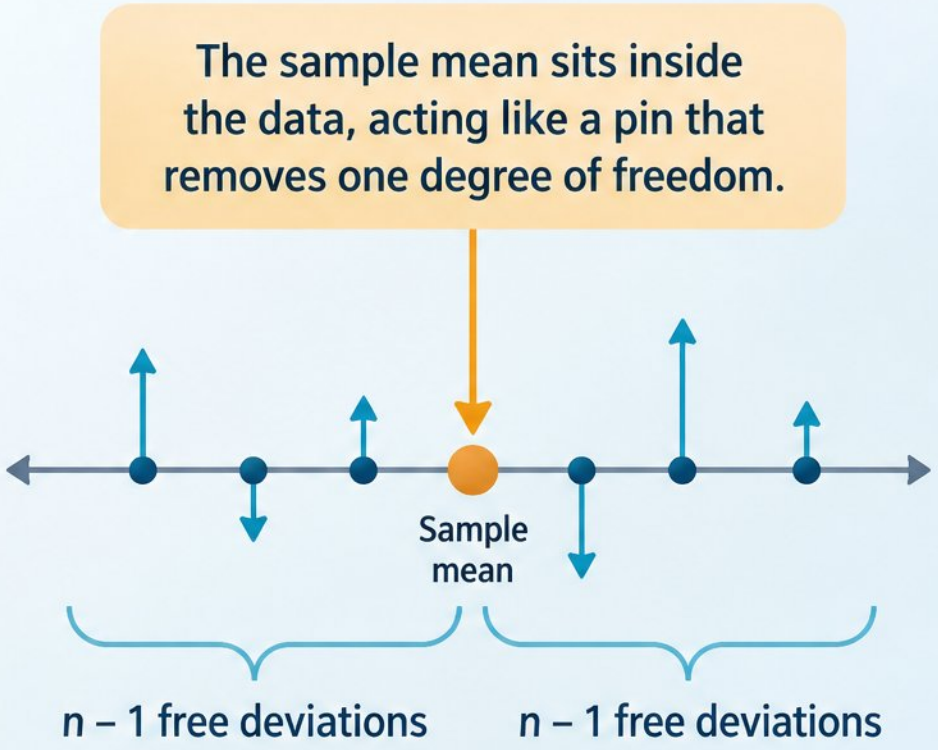
Sample variance: sum of squared deviations divided by $n - 1$

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

- Entire population → divide by $N$
- Sample from a larger group → divide by $n - 1$
- Common error: dividing by $n$ when data are only a sample

# Why Divide by n Minus 1?

- Deviations from the sample mean tend to be too small
- The sample mean sits inside the data, pulling squared deviations down
- Deviations from the mean always sum to zero: only n – 1 vary freely
- Degrees of freedom = n – 1
- Dividing by n – 1 (Bessel's correction) makes the estimator unbiased
- Large samples: dividing by n or n – 1 gives nearly the same result



# Worked Example: Variance Step by Step

| Value (minutes) | Deviation from mean (25) | Squared deviation |
|---|---|---|
| 22 | −3 | 9 |
| 26 | 1 | 1 |
| 25 | 0 | 0 |
| 30 | 5 | 25 |
| 27 | 2 | 4 |
| 20 | −5 | 25 |
| **Sum of squares** | | **52** |

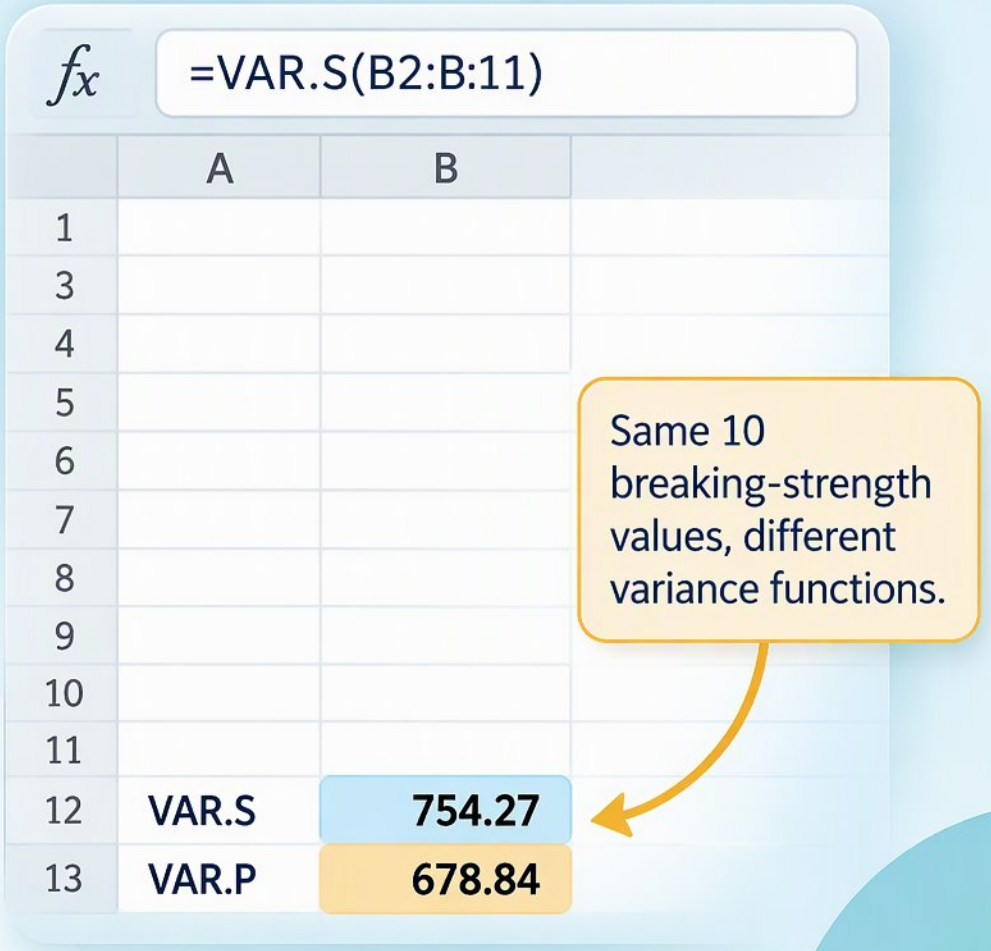
- Dataset: 6 delivery times (minutes) — 22, 26, 25, 30, 27, 20
- Mean = 25; sum of squared deviations = 52
- Population variance: 52 ÷ 6 = **8.67** square minutes
- Sample variance: 52 ÷ 5 = **10.40** square minutes
- Verify with VAR.P and VAR.S; sample variance is larger

SAME DATA.
DIFFERENT INSIGHTS.



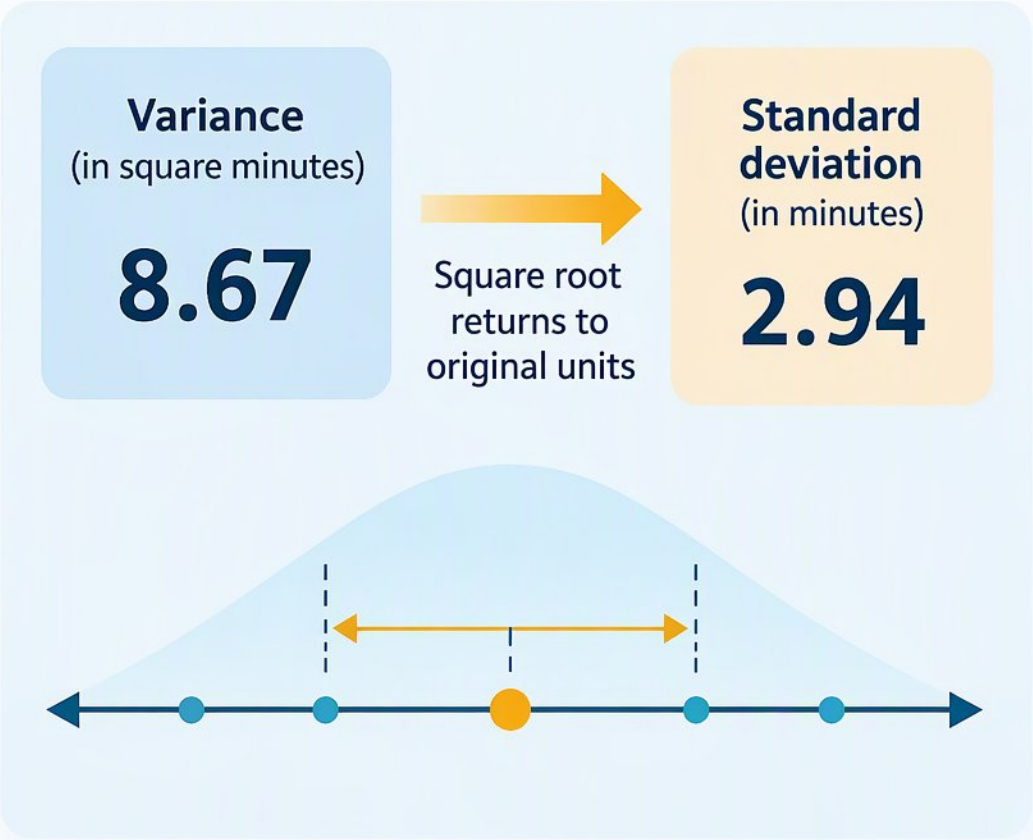
# Verifying Results in a Spreadsheet

- Use VAR.S for sample data; use VAR.P for a full population
- Same 10 breaking-strength values: VAR.S = 754.27, VAR.P = 678.84
- The gap is expected: n minus 1 versus n denominators
- Legacy VAR and VARP still work, but VAR.S and VAR.P are modern
- Always confirm hand calculations with the matching spreadsheet function



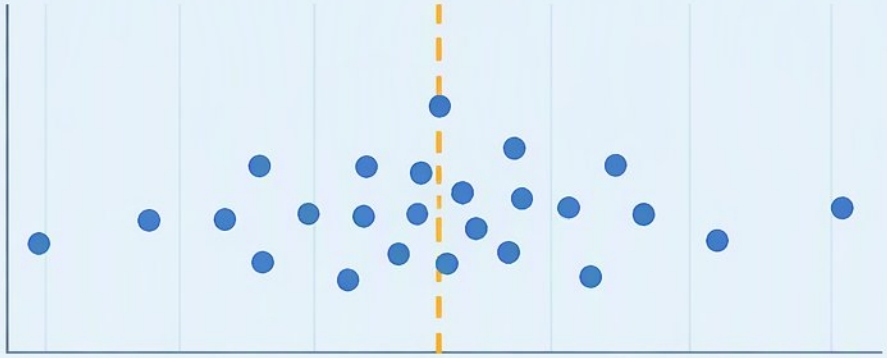
# Interpreting Variance and Standard Deviation

- Variance uses squared units (e.g., square minutes), making direct interpretation difficult
- Standard deviation is the square root of variance, back in original units: about 2.94 minutes
- Variance of zero means no spread; larger variance means greater dispersion
- Many datasets have most values within one to two standard deviations of the mean
- Do not assume a normal distribution without checking

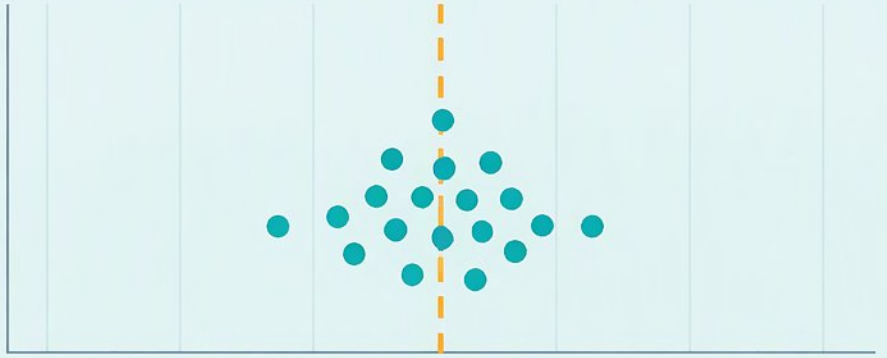


# Comparing Variability Across Groups

**Group A**
(larger variance)



**Group B**
(smaller variance)



- Same mean, larger variance: the less consistent group
- Squaring makes extreme values disproportionately influential
- Rescaling units changes the number, not the underlying spread
- Stakeholder phrasing: "Group A varies about twice as much as group B"

# Limits of Variance and Complementary Measures

Different views. A clearer answer.

Variance is sensitive to outliers; pair it with other measures

## Standard deviation

Same units, easier to communicate



What it helps answer:
How much do values typically vary?

## Range

Quick but driven by the extremes



What it helps answer:
What is the total spread from lowest to highest?

## IQR

Resistant to outliers, ignores the middle detail



What it helps answer:
How much do the middle values vary?

## Coefficient of variation

Coefficient of variation = SD ÷ mean, compares across units and scales

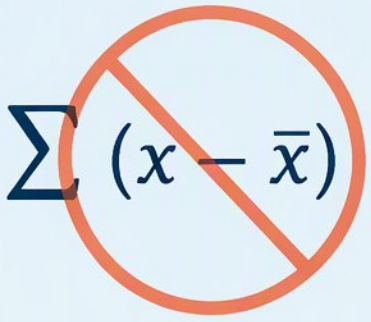


What it helps answer:
How does variability compare across different units or scales?

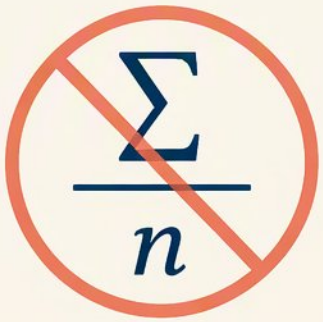
? Choose based on the question: total, typical, or relative dispersion

A broader view leads to better insights.

# Common Mistakes to Avoid



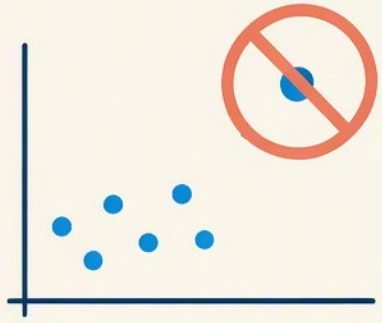

Averaging deviations without squaring — the result is always zero




Dividing by $n$ instead of $n$ minus 1 when your data are a sample




Reading variance as if it were in the original units, not squared units



Forgetting that a few extreme values can dominate the variance

**Quick check:** verify the denominator, the units, and the effect of outliers

| Value | Deviation | Squared Deviation |
| --- | --- | --- |
| | | |
| | | |
| | | |

# Quick Practice and Self-Check

PRACTICE BUILDS CONFIDENCE

Work through the dataset, answer the questions, and check your understanding.

**1** **Practice dataset:** quiz scores 12, 18, 15, 20, 15; mean = 16, sum of squared deviations = 34

| Value | Deviation (from mean) | Squared deviation |
|---|---|---|
| 12 | | |
| 18 | | |
| 15 | | |
| 20 | | |
| 15 | | |

**2** **Compute population variance (6.8) and sample variance (8.5); state which fits and why.**

**Population variance:** _____

**Sample variance:** _____

**Which fits and why?**

_____

_____

**3** **Interpretation task:** explain the spread to a manager without using squared units.

_____

_____

**4** Self-check: name the denominator, verify in a spreadsheet, convert back to original units.

- [ ] Name the denominator.
- [ ] Verify in a spreadsheet.
- [ ] Convert back to original units.

SAME DATA. DEEPER UNDERSTANDING.

# Key Takeaways and Next Steps

- Variance: average of squared deviations from the mean
- Populasi? No: use N for full data, n – 1 for samples
- Standard deviation returns variance to original units
- Practice: compute VAR.S and VAR.P on your own data
- Always state which denominator you used and why



PersonWise.

PersonWise • AI digital-human course platform

Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/15c65253/public