06. Why Sample Variance Is Divided By n-1

Mahir Faysal Tusher

June 2025

Sample Variance and Population Variance

The formula for **sample variance** is:

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

Where:

- s^2: Sample variance
- $\bar{x}$: Sample mean
- n: Sample size

In contrast, the formula for **population variance** is:

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

Where:

- σ^2: Population variance
- μ: Population mean
- N: Population size

Why Divide by $n - 1$ in Sample Variance?

You may wonder why we divide by $n - 1$ instead of n. This is because:

- A sample is used to estimate the population parameters.
- If we divide by n, we **underestimate** the population variance on average.
- Dividing by $n-1$ corrects this bias and provides an **unbiased estimator**.

This correction is known as **Bessel's correction**.

Graphical Intuition

Imagine a population of age data distributed along a line. Selecting a random sample from this population and calculating its mean ($\bar{x}$) and variance (s^2) should give values approximately close to the population parameters (μ, σ^2).

However, if the sample is not representative (e.g., it contains only younger or only older individuals), then:

$$\bar{x} \ll \mu \quad \text{and} \quad s^2 \ll \sigma^2$$

Using $n-1$ helps reduce this underestimation by increasing the calculated sample variance.

Degrees of Freedom

- When we compute the sample variance, one degree of freedom is lost due to the estimation of the mean from the same data.
- Hence, degrees of freedom (DOF) for variance in a sample is:

$$\text{DOF} = n - 1$$

Summary

- Use $n-1$ when calculating sample variance to obtain an unbiased estimate of population variance.
- This is known as Bessel's correction.
- Degrees of freedom play a critical role in this correction.