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

When calculating **sample variance**, we use $n - 1$ in the denominator instead of n to correct for bias in estimating the **population variance** from a sample. Here's the proper explanation:

---

1. **Population Variance vs. Sample Variance**

- **Population variance** (σ^2) is defined as:

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

  where μ is the population mean and N is the population size.
- When we don't have access to the entire population, we calculate the **sample variance** as an estimate of the population variance.

---

2. **Bias in Using n**

- In a sample, the mean $\bar{x}$ is calculated from the data, so it is **closer to the data points than the true population mean** (μ).
- As a result, using n in the denominator systematically **underestimates the population variance** because the deviations from $\bar{x}$ are smaller than the deviations from μ.

---

3. **Bessel's Correction: Why $n - 1$?**

- To adjust for this bias, we divide by $n - 1$ instead of n.
- This accounts for the fact that one degree of freedom is lost when calculating the sample mean $\bar{x}$, as the sample data are constrained to sum up to $n \cdot \bar{x}$.
- Mathematically, using $n - 1$ makes the sample variance an **unbiased estimator** of the population variance:

  $$\mathrm{E}[s^2] = \sigma^2$$

---

4. **Key Concept**

- Dividing by $n-1$ increases the sample variance slightly, reducing the gap between the sample variance and the true population variance.
- It ensures the sample variance is neither too small nor systematically biased when used to estimate the population variance.

Summary

Using $n-1$ in the denominator for sample variance:

1. Adjusts for the bias caused by using the sample mean ($\bar{x}$).
2. Ensures the sample variance is an **unbiased estimate** of the population variance.
3. Reflects the loss of 1 degree of freedom in the sample.