Understanding Variance, Standard Deviation, Population, Sample, and the Importance of Dividing by (n-1) i sample varience

What is Variance?
Variance measures how far a set of numbers are spread out from their mean. It quantifies data dispersion but is expressed in squared units, making it less interpretable.

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

Variance

What is Standard Deviation?
The standard deviation is the square root of the variance. It provides a measure of data dispersion while being in the same units as the original dataset, making it easier to interpret.

<u>Population</u>

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

Standard Deviation

Population vs. Sample in the Context of Variance and Standard Deviation:

Population:

The entire set of data or all possible observations that could be studied. For example, if studying the heights of all adult men in a country, the population includes every adult male's height.

Sample:

A subset of the population selected for analysis. Since studying an entire population is often impractical, researchers analyze a sample to make inferences about the population.

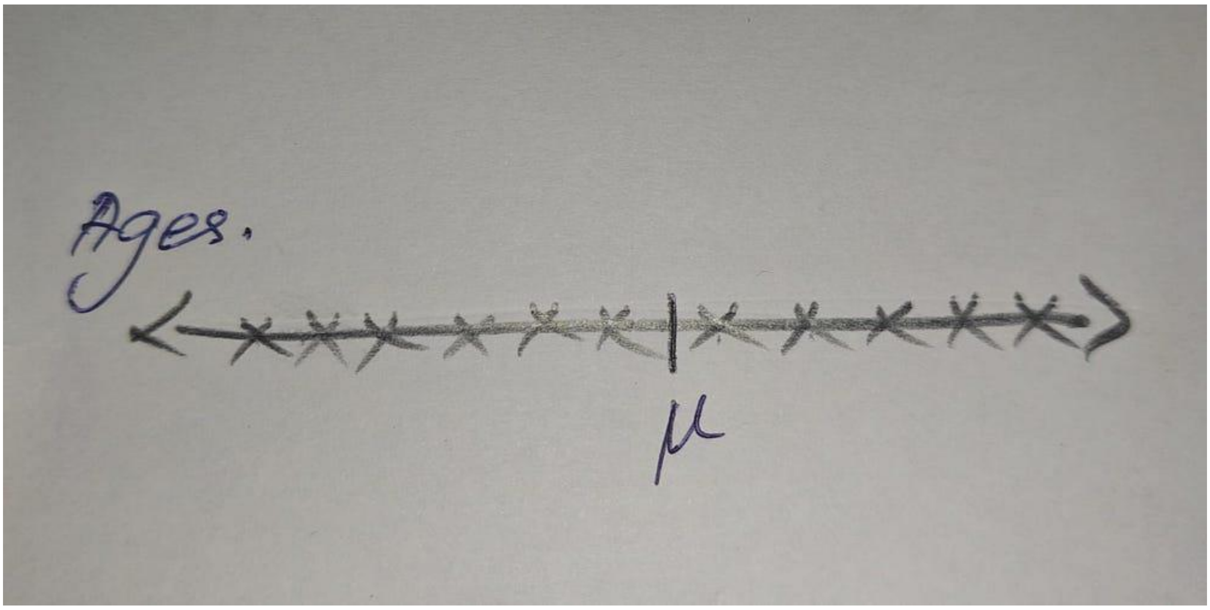
	Population	Sample
Variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$	$s = \sqrt{\frac{1}{n-1}\sum_{x=1}^{n}(x_i - \bar{x})^2}$

Population vs. Sample Variance and Standard Deviation: Key Differences

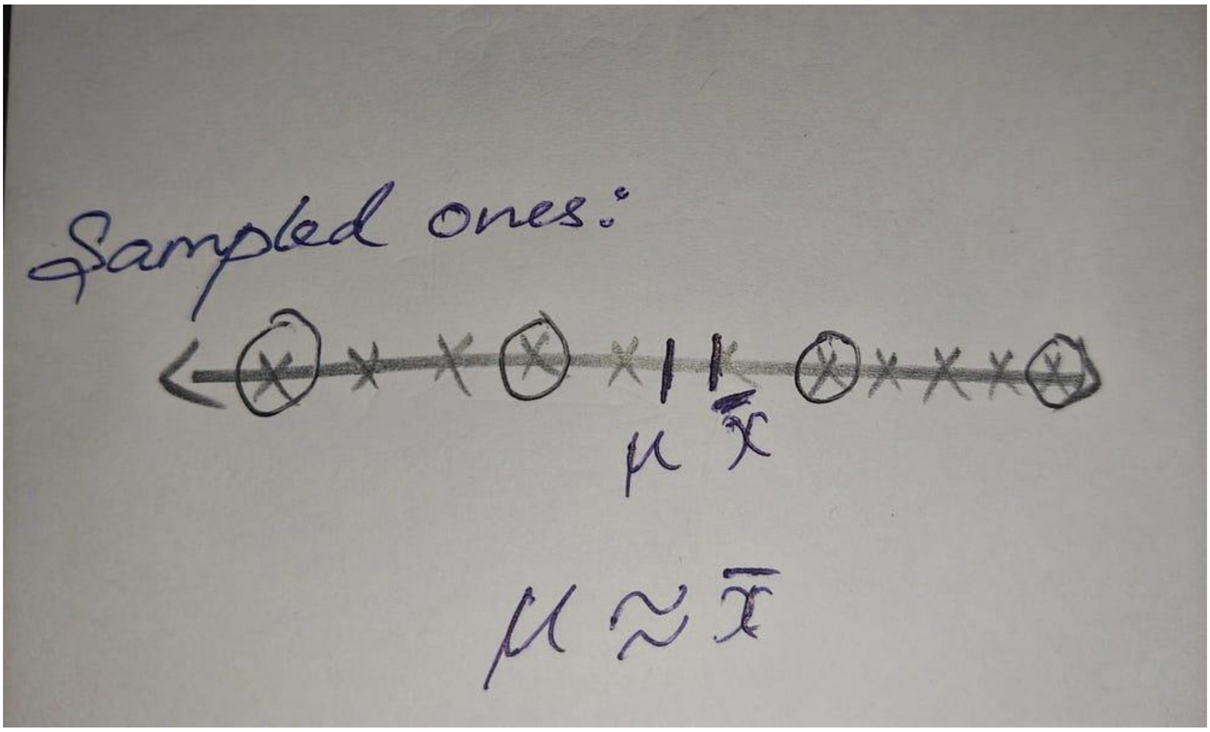
Why is Sample Variance Divided by (n-1)?
When calculating the variance of a sample, we divide by (n-1) instead of n. This adjustment, known as Bessel's correction, ensures an unbiased estimate of the population variance.

Understanding the Need for (n-1):

Dataset Well disttibuted:

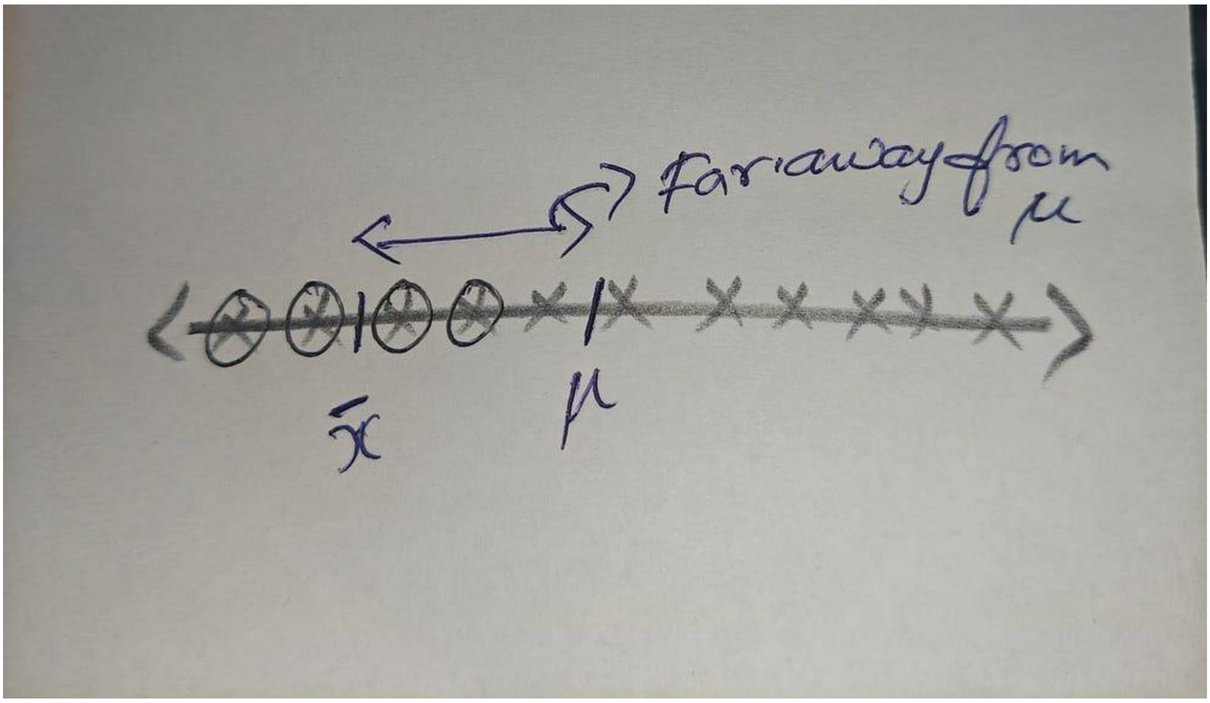


Sampling from a Well-Distributed Population :



When randomly selecting a sample from a population, the sample mean (x̄) is usually close to the population mean (μ), making it a good estimate.

Sampling from a Skewed Population:



If the sample is not representative and comes from a specific cluster within the population, the sample mean (x̄) may be significantly different from (μ), leading to an underestimated variance.

Correction for Bias:

Since a sample tends to underestimate the true population variance, dividing by (n-1) instead of n inflates the variance slightly, compensating for this bias.
This adjustment ensures that the sample variance provides a better estimate of the true population variance.

Conclusion
The use of (n-1) in sample variance calculations corrects for the natural bias that occurs when estimating population variance from a sample. By making this adjustment, we ensure that our statistical estimates are more accurate and reliable, bringing the sample variance closer to the true population variance.