

June 25th Assignment

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} \quad \text{Population Variance}$$
$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \text{Sample Variance}$$

Why n-1?

Question: Why the denominator "n-1" is used in a sample variance formula unlike population variance which is just N.

Answer: When we draw samples from a population, we assume that the samples are drawn in such a way that their mean "s" is a true estimate for the population mean "μ". But there might be situations where we end up drawing samples from the extreme end of the population and their mean is either way less or way more from the actual population mean thereby resulting in a significant deviation of the sample variance from the actual population variance. To counter that, we use "n-1" in the denominator which provides an unbiased estimate for the true population variance.

There are arguments as to why only "n-1" is used and not "n-2" or "n-3" "n-x" etc. but researchers after doing so many studies came to the conclusion that "n-1" tends to bring the sample variance much closer to the population variance as compared to "n-2", "n-3" "n-x" and so on.