Why $n-1$ in variance formula for sample?

Guide: This explanation is for those who are interested in the question. It is not mandatory for everyone.

In order to see the difference between the sample variance calculated by the denominator n versus $n-1$ consider a population with known mean μ and variance σ^2. Now imagine we take a sample of size n from this population and estimate the sample mean $\bar{x}$. This sample mean $\bar{x}$ is not necessarily equal to the population mean μ.

Suppose we use n in the denominator to estimate the variance which gives

$$\tilde{s}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - n\bar{x}^2\right) \tag{@}$$

For the population, on the other hand, if we use population mean instead of $\bar{x}$ we get

$$\begin{aligned} \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \\ \Rightarrow \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - n\mu^2\right) \\ \Rightarrow \frac{1}{n}\sum_{i=1}^{n}x_i^2 &= \sigma^2+\mu^2 \end{aligned} \tag{*}$$

In order to complete the calculation we will also need the value of 'variance of the sample mean'. This is because the sample mean is going to change with each new sample. We can calculate it as

$$\begin{aligned} \mathrm{Var}(\bar{x}) &= \mathrm{Var}\left(\frac{\sum x_i}{n}\right) \\ &= \mathrm{Var}\left(\sum \frac{x_i}{n}\right) \\ &= \frac{1}{n^2}\left(\sum \mathrm{Var}(x_i)\right) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n} \end{aligned}$$

This gives

$$\frac{1}{n}\sum \bar{x}^2 = \mu^2 + \frac{\sigma^2}{n}. \tag{**}$$

Plugging (*) and (**) in equation (@) gives

$$\tilde{s}_x^2 = \sigma^2+\mu^2-\mu^2-\frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

It is clear that this does not give us the correct variance σ^2. However, if we normalize this sample variance by $\frac{n}{n-1}$ then we are going to get the exact population variance.

5. **Standard deviation** is the positive square root of the variance. Similar to the variance, here too, we will use N and $n-1$ in the denominator for population and sample, respectively. The standard deviation for population, therefore, is given by

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

and for sample