---

**TOOL:** The sample variance, $S^2$, is an unbiased estimate of variance $\sigma^2$, (when the samples, $X_i$, are independent and identically distributed). (This accounts for why the formula employs a multiplicative factor of $\frac{1}{n-1}$ instead of $\frac{1}{n}$.)

**DERIV:** $S^2$ is an unbiased estimate of variance $\sigma^2$ means $E(S^2) = \sigma^2$.

We use tools for linear combinations of random variables from probability to compute $E(S^2)$, starting from the definition of $S^2$.

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}E\left((X_i - \overline{X})^2\right)$$

By symmetry and independence of the $X_i$, we can argue that each term of the summation must have the same expected value as the first term:

$$E(S^2) = \frac{n}{n-1}E\left((X_1 - \overline{X})^2\right)$$

If we expand the sample mean, we have the following:

$$E(S^2) = \frac{n}{n-1}E\left(\left(X_1 - \frac{1}{n}\sum_{i=1}^{n}X_i\right)^2\right) = \frac{n}{n-1}E\left(\left(X_1 - \frac{1}{n}X_1 - \frac{1}{n}\sum_{i=2}^{n}X_i\right)^2\right)$$

or

$$E(S^2) = \frac{n}{n-1}E\left(\left(X_1 - \frac{1}{n}\sum_{i=1}^{n}X_i\right)^2\right) = \frac{n}{n-1}E\left(\left(\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{i=2}^{n}X_i\right)^2\right)$$

The above expression reveals that using $X_1$ in the sample mean as well as in the distance from the sample mean reduces the effective value of $X_1$ by a factor of $(n-1)/n$. The remainder of the derivation consists of manipulations that ultimately demonstrate that the estimated sample variance we would obtain by taking the average of squared distances of samples from the sample mean is reduced by this same factor. In other words, using the data to compute a sample mean results in a mean that is closer to the sample values than it should be. It is closer by a factor of $(n-1)/n$.

We now expand the term in the expected value.

$$E(S^2) = \frac{n}{n-1} E\left[\left(\frac{n-1}{n} X_1\right)^2 - 2\left(\frac{n-1}{n} X_1\right)\left(\frac{1}{n}\sum_{i=2}^{n} X_i\right) + \left(\frac{1}{n}\sum_{i=2}^{n} X_i\right)^2\right]$$

or

$$E(S^2) = \frac{n-1}{n}\left\{E\left(X_1\right)^2 - 2\frac{n}{n-1}\cdot\frac{1}{n}E\left(X_1\sum_{i=2}^{n} X_i\right) + \frac{1}{(n-1)^2}E\left[\left(\sum_{i=2}^{n} X_i\right)^2\right]\right\}$$

**NOTE:** The summations in the above equation start at $i = 2$, and the factor in front has been inverted.

Exploiting the independence of the $X_i$, we have the following identity for the middle term:

$$E\left(X_1\sum_{i=2}^{n} X_i\right) = E(X_1)\cdot(n-1)E(X_{i\neq 1}) = \mu(n-1)\mu = (n-1)\mu^2$$

For the third term, we have the following expansion:

$$E\left(\left(X_2 + \ldots + X_n\right)\sum_{i=2}^{n} X_i\right) = E\left(X_2\sum_{i=2}^{n} X_i + \ldots + X_n\sum_{i=2}^{n} X_i\right)$$

Again exploiting the independence and identical distributions of the $X_i$, we have the following:

$$E\left(\left(X_2 + \ldots + X_n\right)\sum_{i=2}^{n} X_i\right) = (n-1)E\left(X_2\sum_{i=2}^{n} X_i\right)$$

In the expected value, we get an $E(X_2^2)$ term plus $n-2$ terms of the form $E(X_i)E(X_{j\neq i})$ that yield a value of $\mu^2$:

$$E\left(\left(X_2 + \ldots + X_n\right)\sum_{i=2}^{n} X_i\right) = (n-1)\left[E(X_2^2) + (n-2)\mu^2\right]$$

Making substitutions based on these identities, we have the following:

$$E(S^2) = \frac{n-1}{n}\left\{E(X_1)^2 - 2\frac{n}{n-1}\frac{(n-1)}{n}\mu^2 + \frac{1}{(n-1)^2}(n-1)[E(X_2^2) + (n-2)\mu^2]\right\}$$

Since all $X_i$ are identically distributed, we may use a generic $X$ in place of $X_1$ or $X_2$:

$$E(S^2) = \frac{n-1}{n}\left\{E(X)^2 - 2\mu^2 + \frac{1}{(n-1)}[E(X^2) + (n-2)\mu^2]\right\}$$

or

$$E(S^2) = \frac{n-1}{n}\left\{E(X)^2 - \mu^2 - \frac{(n-1)}{(n-1)}\mu^2 + \frac{1}{(n-1)}[E(X^2) + (n-2)\mu^2]\right\}$$

or

$$E(S^2) = \frac{n-1}{n}\left\{E(X)^2 - \mu^2 + \frac{1}{(n-1)}[E(X^2) - \mu^2]\right\}$$

or

$$E(S^2) = \frac{n-1}{n}\left\{\frac{n}{n-1}\left[E(X)^2 - \mu^2\right]\right\}$$

or

$$E(S^2) = E(X^2) - \mu^2 = \sigma^2$$

Thus, $S^2$ is verified to be an unbiased estimate of the variance, $\sigma^2$.