

## Biased vs Unbiased Estimators;

### Why do we divide with $n-1$ the sample standard deviation?

**Simple Explanation:** it is not possible to obtain an estimate of  $\sigma$  from a sample size  $n=1$ , because there is no internal variation of *any* degree within such a sample. Having  $n-1$  in the denominator reflects this impossibility, and therefore at least  $n=2$  data points are needed if we want to make the formula work

**In more detail:** The different denominators arise from the theory of parameter estimation in statistics, whereby some random variable has a probability distribution with some parameter unknown, and it is calculated from a series of observations of outcomes of the random variable (the sample).

When estimating the standard deviation or variance when the mean is unknown, using  $n$  in the denominator results in a biased estimator (that is, one such that its expected value is not precisely the parameter that is being estimated); in fact, one can show that the expectation of the estimator for the variance ends up as  $(n-1)/n * \sigma^2$ , and thus by multiplying the original estimator by  $n/(n-1)$  -- that is, using  $(n-1)$  as the denominator instead of  $n$  -- we obtain an unbiased estimator, i.e. with expectation  $\sigma^2$ , at the cost of a higher mean squared error.