Why $n - 1$ in the Formula for the Sample Standard Deviation?

Stephen A. Book

Stephen A. Book is Associate Professor of Mathematics at California State College, Dominguez Hills. He received his Ph.D. in Mathematics from the University of Oregon in 1970 and has authored two introductory textbooks, "Statistics: Basic Techniques for Solving Applied Problems" and "Essentials of Statistics", published by McGraw-Hill.

Perhaps the single most important lesson for students of elementary statistics to learn is how to use a random sample of n data points $x_1, x_2, \ldots, x_n$ to estimate the mean μ of a population. Generally the students have no difficulty understanding that the best estimate of μ is the "sample mean" $\bar{x} = (\sum_{k=1}^{n} x_k)/n$, and they are very receptive to the law of averages, which asserts that, as n increases, $\bar{x}$ tends to μ as a limit.

The question then arises of how accurate $\bar{x}$ is as an estimate of μ. To answer this question, the concept of a confidence interval is introduced. The first step in the confidence interval approach, namely, the **central limit theorem**, is willingly accepted by the students. It says:

> **For large values of n, the set of all possible sample means of samples consisting of n data points has approximately a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$. Here μ and σ are the population mean and standard deviation of the population from which the samples were chosen.**

The formula for confidence intervals, namely, the statement that we can be $(1 - \alpha)100\%$ sure that $\mu = \bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$, is a simple algebraic consequence of the central limit theorem.

In most applications, however, the above formula cannot be used as it stands, because it contains the (generally unknown) population standard deviation σ. The usual procedure to get around this difficulty is to replace σ by the "sample standard deviation"

$$s = \sqrt{\frac{\sum_{k=1}^{n} (x_k - \bar{x})^2}{n - 1}}.$$

While the idea of a sample standard deviation makes sense to the students, the formula for s encounters a considerable amount of resistance due to the term $n - 1$ in the denominator. The students find the $n - 1$ illogical, since they have previously been taught that the population standard deviation, the square root of the average squared deviation from the mean, is given by the formula

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

where $\tilde{\sum}$ denotes summation over the entire population, which we denote as $w_1, w_2, \ldots, w_N$, and N is the number of points in the entire population. (Of course, N might be ∞, and then $\tilde{\sum}$ would theoretically have to be replaced by an integral.) It should be pointed out that the sample of data, the x's, is a subset of the population, the w's.

One common justification used at the elementary level for this situation is that it is not possible to obtain an estimate of σ from a sample of 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. Another popular viewpoint holds that, when choosing a random sample, one is not likely to come up with too many of the extreme values in the population and therefore the sample standard deviation will tend to underestimate the true standard deviation σ. To account for this underestimation, the argument goes, we should divide by $n - 1$ instead of n.

Neither of these approaches provides a fully satisfactory account of why we use $n - 1$ rather than some other factor in computing the sample standard deviation. The objective of this article is to present, using only elementary algebra combined with results that the students have already accepted, a reasonable explanation of why the denominator in s should be exactly $n - 1$.

The underlying reason for the switch from N to $n - 1$ in going from σ to s is related to the distinction between μ and $\bar{x}$. Note that μ appears in the formula for σ, while $\bar{x}$ appears in the formula for s. The result can be presented formally by way of the following theorem and proof:

Theorem. *If, for a population having mean μ and standard deviation σ, we were to calculate the sample "variance" (the variance is simply the square of the standard deviation)*

$$s^2 = \frac{\sum_{k=1}^{n} (x_k - \bar{x})^2}{n - 1}$$

for every possible sample $x_1, x_2, \ldots, x_n$ consisting of n data points, then the average value (mean) of all these sample variances (of all possible samples of size n) would be the population variance

$$\sigma^2 = \frac{\tilde{\sum} (w_j - \mu)^2}{N} .$$

Proof. Using the binomial theorem, together with commutativity and associativity, we can write

$$\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}(x_k-\mu+\mu-\bar{x})^2$$

$$=\sum_{k=1}^{n}(x_k-\mu)^2+2\sum_{k=1}^{n}(x_k-\mu)(\mu-\bar{x})+\sum_{k=1}^{n}(\mu-\bar{x})^2$$

$$=\sum_{k=1}^{n}(x_k-\mu)^2+2(\mu-\bar{x})\sum_{k=1}^{n}(x_k-\mu)+n(\mu-\bar{x})^2.$$

Now, because $\bar{x}=n^{-1}\sum_{k=1}^{n}x_k$, the middle term becomes

$$2(\mu-\bar{x})\sum_{k=1}^{n}(x_k-\mu)=2(\mu-\bar{x})\left(\sum_{k=1}^{n}x_k-\sum_{k=1}^{n}\mu\right)$$

$$=2(\mu-\bar{x})(n\bar{x}-n\mu)=-2n(\bar{x}-\mu)^2.$$

Therefore

$$\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}(x_k-\mu)^2-2n(\bar{x}-\mu)^2+n(\bar{x}-\mu)^2$$

$$=\sum_{k=1}^{n}(x_k-\mu)^2-n(\bar{x}-\mu)^2.$$

Now, when we say that the population variance is σ^2, we mean that σ^2 is the average (mean) of the squared deviations of the individual members w_j of the population from μ. Rewording this, we can say that $(w_j-\mu)^2$ is "on the average" equal to σ^2, because σ^2 is the average of all the numbers $(w_j-\mu)^2$. Therefore $\sum_{k=1}^{n}(x_k-\mu)^2$ is "on the average" equal to $n\sigma^2$, since each number x_k of our random sample is going to be some member w_j of the population and the average $(x_k-\mu)^2$ is the same as the average $(w_j-\mu)^2$.

What about $(\bar{x}-\mu)^2$? Recall from the statement of the central limit theorem that the set of all possible sample means of samples consisting of n data points has sample standard deviation $\sigma/\sqrt{n}$, and therefore sample variance σ^2/n. (As it turns out, this holds for *all* values of n; large n's are needed only to assert that the sample means are approximately normally distributed.) This means that $(\bar{x}-\mu)^2$ is "on the average" equal to σ^2/n. That is to say, if we compute $\bar{x}$ and then $(\bar{x}-\mu)^2$ for *every* possible sample of size n, the numbers $(\bar{x}-\mu)^2$ will average out to σ^2/n. Therefore $n(\bar{x}-\mu)^2$ is "on the average" equal to σ^2.

It then follows that

$$\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}(x_k-\mu)^2-n(\bar{x}-\mu)^2$$

is "on the average" equal to $n\sigma^2 - \sigma^2 = (n-1)\sigma^2$, the average being taken over *all* possible samples of size n. We are therefore justified in asserting that $s^2 = (n-1)^{-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$ is "on the average" equal to σ^2. This completes the proof.

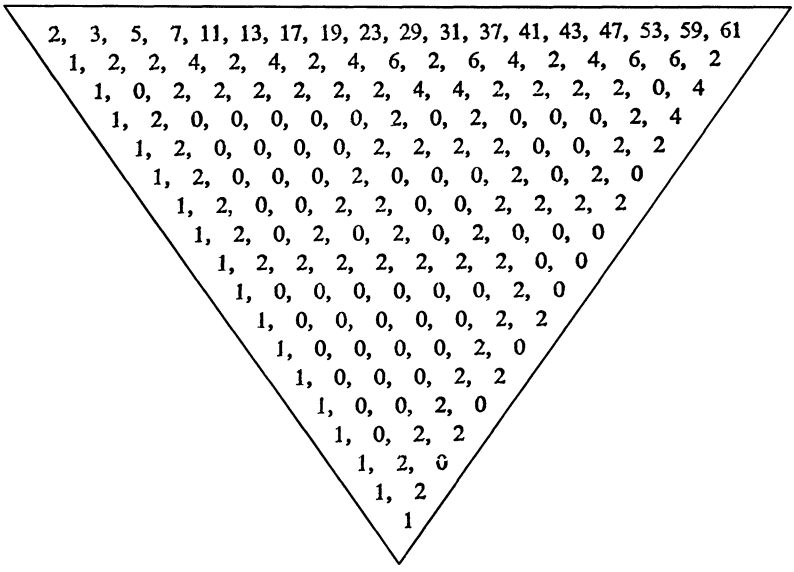
Now that we have shown that s^2 is, on the average, equal to σ^2, we can remark that the estimator the students would like to use, namely,

$$\mathbb{S}^2 = \frac{\sum_{k=1}^{n}(x_k - \bar{x})^2}{n}$$

is, on the average, equal to $((n-1)/n)\sigma^2$. Therefore $\mathbb{S}^2$ tends to *underestimate* σ^2 slightly.

To summarize: we use s rather than $\mathbb{S}$ as a sample estimate of the population standard deviation σ, because s tends "on the average" to give us a more accurate estimate of σ.

Note: The author would like to thank an anonymous reviewer for several valuable comments serving to improve the exposition.



Do you believe in Patterns?