EX: An engineer measures the following beta values for bipolar transistors with the aim of finding nominal values of gain, (i.e., beta), to list on a datasheet. The sample variance is also of interest, as it serves as a guide to min and max values to list on the datasheet.

$\beta_1 = 111$ $\beta_2 = 136$ $\beta_3 = 159$ $\beta_4 = 141$ $\beta_5 = 109$ $\beta_6 = 121$

$\beta_7 = 117$ $\beta_8 = 105$ $\beta_9 = 99$ $\beta_{10} = 102$

Find the sample variance and sample standard deviation of the data.

SOL'N: The sample variance, S^2, is *almost* the average of the squared differences between data values and the sample mean, (which is the average data value):

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \quad \text{where } n \equiv \text{number of data values}$$

Notice, however, that the sum of the squared differences is divided by $n - 1$ rather than n. The reason for this curious feature of the sample variance is that it makes S^2 an unbiased estimator. That is to say, the expected value of S^2 equals the true variance of the data.

$$E(S^2) = \sigma_X^2$$

Although we forego the detailed proof here, the proof that $n - 1$ gives an unbiased estimator begins with a substitution of $(X_i - \mu_X) + (\mu_X - \overline{X})$ for $X_i - \overline{X}$. After squaring and canceling cross terms, we find that the first term in parentheses has a variance of σ^2 and the second term in parentheses has a variance of σ^2/n. Thus, the second term is responsible for the –1 in the $n - 1$. What has happened is that we have used our data twice when we use $\overline{X}$ in place of μ. This overuse of the data increases the variance of S^2.

Using a spreadsheet to compute the sample mean of the data, we find the following:

$$\overline{X} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1200}{10} = 120$$

Calculation of S^2 gives the following value:

$$S^2 = \frac{3420}{10-1} = 380$$

The sample standard deviation is $S = \sqrt{S^2}$.

$S \approx 14.5$