Notation Here is a summary of notation for the standard deviation and variance:

s = *sample* standard deviation
s^2 = *sample* variance
σ = *population* standard deviation
σ^2 = *population* variance

Note: Articles in professional journals and reports often use SD for standard deviation and VAR for variance.

Important Properties of Variance

- The units of the variance are the *squares* of the units of the original data values. (If the original data values are in feet, the variance will have units of ft^2; if the original data values are in seconds, the variance will have units of sec^2.)
- The value of the variance can increase dramatically with the inclusion of one or more outliers (data values that are very far away from all of the others).
- The value of the variance is usually positive. It is zero only when all of the data values are the same number. (It is never negative.)
- The sample variance s^2 is an **unbiased estimator** of the population variance σ^2, as described in Part 2 of this section.

The variance is a statistic used in some statistical methods, but for our present purposes, the variance has the serious disadvantage of using units that are *different than the units of the original data set.* This makes it difficult to understand variance as it relates to the original data set. Because of this property, it is better to focus on the standard deviation when trying to develop an understanding of variation, as we do in this section.

Part 2: Beyond the Basics of Variation

In this subsection we focus on making sense of the standard deviation so that it is not some mysterious number devoid of any practical significance. We begin by addressing common questions that relate to the standard deviation.

Why is Standard Deviation Defined as in Formula 3-4?

Why do we measure variation using Formula 3-4? In measuring variation in a set of sample data, it makes sense to begin with the individual amounts by which values deviate from the mean. For a particular data value x, the amount of **deviation** is $x - \bar{x}$, which is the difference between the individual x value and the mean. It makes sense to somehow combine those deviations into one number that can serve as a measure of the variation. Simply adding the deviations doesn't work, because the sum will always be zero. To get a statistic that measures variation (instead of always being zero), we need to avoid the canceling out of negative and positive numbers. One simple and natural approach is to add absolute values, as in $\Sigma|x - \bar{x}|$. If we find the mean of that sum, we get the **mean absolute deviation** (or **MAD**), which is the mean distance of the data from the mean:

$$\text{mean absolute deviation} = \frac{\Sigma|x - \bar{x}|}{n}$$

Why Not Use the Mean Absolute Deviation Instead of the Standard Deviation? Computation of the mean absolute deviation uses absolute values, so it uses an operation that is not "algebraic." (The algebraic operations include addition, multiplication, extracting roots, and raising to powers that are integers or fractions, but absolute