VARIANCE vs STANDARD DEVIATION

Why do we even name variance since it's just the square of the standard deviation? Is it used alone as variance other than to add standard deviations?

The variance is actually the basic measure - the sum of squares of deviations shows up in many places - for example, the usual regression line is the "line of best fit in the sense of least squares" (sum of squares of the residuals - serving a similar role to the deviations) and many of the formulas for the next set of statistical tests are based on the sum of squares (not the squares of the data - but the squares of some sort of deviation measure). The variance is, in essence, the mean of the squares of the deviations.

What may be more useful to tell students is that the variance (rather than the standard deviation) is the form that shows up in theory and in development of formulas (for example, variances add when we add or subtract independent random variables, standard deviations don't) while standard deviation is a bit more intuitively meaningful as a description of distribution, so we keep both ideas and names in use for different situations.

Charlie Peltier
Saint Mary's College

The main reason for using SDs for variability ... to me anyway ... is that a SD is in the units of the metric you are working with. In that sense, it has more intuitive meaning. If we are talking about length in inches ... a variance might have meaning in that a SQUARE inch has meaning. But, if you were working with data on pounds ... using the variance and having to think in terms of SQUARED pounds ... is a bit confusing. The SD makes perfectly good sense.

As a descriptive tool ... using the SD most of the time makes more sense because of the underlying connection between the statistic and the metric being used. For theoretical purposes, such a link between the statistic and the underlying metric ... is really not that important.

Standard deviation and variance are measures of the variation of the values in a data set around the mean of that data set. Loosely speaking, the standard deviation is a measure of the average distance of the values in the data set from their mean. Interestingly enough, question 1 on this year's AP Statistics exam required students to interpret the standard deviation of the discoloration ratings of strawberries in the control group of a randomized experiment. The mean of a data set gives you information about the "balance point" of the distribution. The standard deviation adds information about how the values in that data set are spread around the mean.

In the special case of data that are approximately normally distributed, we can use the 68-95-99.7 rule to say more about the approximate % of values in the data set that are within 1, 2, and 3 standard deviations of the mean.

So the standard deviation has a useful interpretation on its own as a measure of spread/variability around the mean in any data set. In fact, Chebyshev's inequality gives more general information about the % of data values within 2, 3, 4, and so forth standard deviations of the mean in any distribution.

Daren Starnes