This tutorial is going to explain to you the concept of standard deviation. It's a measure of variation that we use quite often in statistics. Standard deviation measures spread. And for as ugly as the formula looks, it's actually the one we prefer to use most, provided that the distribution that we're looking for is roughly symmetric with no outliers.

If the data isn't roughly symmetric or has outliers, we might not want to use this one. We'll use a different measure of spread, the interquartile range. The formula looks pretty nasty. But we'll walk through it and show you how to calculate it.

So taking a look, these are the heights of the Chicago Bulls basketball team. We're going to take those values and place them in a list. These are the x subscript i's. OK? And what we're going to do with each of them is subtract the mean, which means we have to calculate the mean first.

Calculating the mean was done in another tutorial. But we can calculate the mean, which was around 78.33 inches. So we're going to subtract 78.33 from each of these values. And so that is subtracting the mean.

The next thing we're going to do with it is square those values. So we end up with x minus the mean squared. And we get these values. It's starting to look very unwieldy, I understand. But we're going to keep going with it.

The next thing we're going to do is use this sigma notation, which is the same as summation notation, to add these values up. They sum up to 205 and 1/3. We're going to then divide that sum by n minus 1. n in this case is 15. Because there were 15 players.

It's almost like we're averaging by dividing. But we're not dividing exactly by n. We're dividing by n minus 1. So dividing that by 14 gets you 14 and 2/3. Now, if we stopped here, this 14.6666 number would measure some kind of variation. This is called "variance." We don't use it too horribly often, mainly because it is, in fact, still a squared value.

The units on variance are not the same as the units for the ones that we measured in. So this measurement here is 14.66667 inches squared, not inches. And then finally, we take the square root of that number. And you get 3.83.

Now, the standard deviation is almost always found on a calculator or a spreadsheet or some kind of applet on the internet that you find. Typically, we do not solve for it by hand. So if you're frustrated with that last series of events, don't be. Because it's going to be OK. You can use your calculator. Or you can use technology, like we're going to do here.

Here's that same list of data. And you won't believe how easy this is. You're going to say, =stdev. You're going to select the list that you want to find the standard deviation for and hit Enter. And sure enough, it's 3.83, just like we decided it was before.

It couldn't be easier. So don't worry too much about it. If you have Excel or some spreadsheet program, there should be a standard deviation formula that you can use.

So we interpret the standard deviation as the typical amount that we would expect data to be within the mean. The full name for standard deviation is "standard deviation from the mean." And if we break that down, "standard" just means it's typical. "Deviation" means that we expect it to be off from the mean, just by chance. That's fine.

And "standard deviation from the mean" is we're going to be away from the mean. So we would expect a good portion of the heights to be within 3.83 inches of 78.33. Let's check those heights again to see how many of the heights were, in fact, within 3.83 inches of the mean.

This 84 wasn't. This 74 wasn't. But you'll notice about 2/3 of the players had heights between 78.33 minus 3.83 or 78.33 plus 3.83 Within that range, about 2/3 of the players were in there. And so that's how we would interpret the standard deviation. It's a typical amount by which we would expect values to vary around the mean.

And so to recap, the standard deviation is kind of difficult to calculate. I'm not going to deny that. But it's typically done using technology. It's a measure of how far we would expect a typical data point to be from the mean.

Standard deviation is the square root of a value called variance. Variance is a little bit easier to calculate. But it's really not as useful as a standard deviation number. And since the standard deviation is based on the mean-- if you recall, we subtracted the mean right off the bat-- the standard deviation should only be reported as the measure of spread when you're reporting the measure of center to be the mean.

You shouldn't mix and match the standard deviation with the median. And you shouldn't mix and match the mean with the interquartile range, either. Terms we introduced this tutorial were standard deviation and variance, which is the square of standard deviation. Again, 'not a horribly useful value for all intents and purposes. Good luck. And we'll see you next time.