

---

Hi, this tutorial gives an introduction to the measures of variation. So let's start with a couple distributions. So the boxplot below shows the distributions of January temperatures and degrees Fahrenheit of two fictional cities.

So we have Smallville here and Gotham City. So we have boxplots here. So remember, these are based on the five-number summaries. The boxplot shows the minimum, Q1. The median, Q3. And the maximum for both cities.

OK, so we can see Gotham City has both a colder minimum temperature as well as a warmer maximum temperature. Their medians are pretty similar. Q3 are quite a bit different, and Q1's seem to be pretty similar.

So remember, when we're comparing distributions, there are three things to look at. We want to think about the similarities and differences in-- remember the center, the spread, and the shape. Spread, remember, another term for that is variation.

OK, so let's start by comparing the center and the shape. So the center-- the distributions have similar centers. Smallville has median temperature of 23 degrees. And Gotham City has a median temperature of 25 degrees. We can see they're pretty similar just by looking at those boxplots. 23 is here, 25 is here. So the center of the distributions are pretty similar.

OK. The shape. The distributions have similar shapes. Neither distribution has a heavy skew. And both appear roughly symmetric. So you can always tell skew on a boxplot by looking at the lengths of the whiskers. OK, so if we look at the whiskers here and here. Those aren't too different in lengths.

So I wouldn't say that this has to heavy of a skew. It is a little more skewed inside of the box, just because there's a lot less distance between the median and Q3 than the median and Q1. But the whiskers don't show a lot of skew.

Gotham City, again, not a lot of skew. The lower whisker is about the same length as the upper whisker. Not a lot of variation in the box. OK. So we can say that both of these appear probably roughly symmetric. So the centers are similar, and the shapes are similar.

The difference-- the big difference that we're going to see is in their spread, or in their variation. Gotham City's boxplot is much longer than Smallville's, and the same could be said for just the box portion of the boxplot. So again, looking back, we can definitely see that, for Gotham City, the length

of the box is much longer than the Smallville box.

The whole boxplot is longer, As well as the box itself being longer. So this box is certainly longer than that. So this box is about 15 units long. This box is about 21 units long. So we do have considerable more variation in temperatures in Gotham City.

So instead of just saying more variation, less variation, or more spread or less spread, wouldn't it be nice to be able to quantify this difference in spread? So assign a number to it. And we can by using the measures of variation.

They're also known as the measures of spread. So what a measure of variation is basically what we just said. A way of measuring how spread out or scattered a data set is. OK. And really, we're going to look at three common measures of variation in future study.

But the three are the range, the interquartile range, abbreviated as the IQR, and the standard deviation. So we're not going to go into a lot of detail on how to calculate these, but you should just know these as measures of variation.

Remember that the measures of center were mean, median, and mode. The measures of variation, range, IQR, standard deviation. Now for all three of these measures, high values mean that the data is spread out away from the center. So it's a lot of spread away from the center.

Low values mean that the data is bunched up or clustered around the center of the distribution. So high values mean lots of variation. Low values mean little variation. And you're going to see-- you'll be able to see that in all three of those measures of variation. So this has been the tutorial on the measures of variation Thanks for watching.