
Hi. This tutorial will help you answer the question, which measure of center? So we're going to look at a couple different measures of center and see which one gives us the best way of summarizing a data set. So we're going to take a look at this data set.

So the picture shows the estimated values of the houses on the block in St. Paul, Minnesota. And this is, by no means, a representative neighborhood in St. Paul-- pretty affluent. So there are nine houses on the block. And each of the numbers next to the house represents the estimated value that is placed on that house. So we can see that the house on the corner here is valued at \$429,000, the k meaning thousand.

So down here in the data set, I have 429, because I'm measuring this in thousands of dollars. \$681,000 is here, \$623,000 is here. If you go to the end of the block, the corner house here, very big house valued at \$2.08 million. So down here, I have that listed as 2,080, because in thousands of dollars, that would give you \$2.08 million.

So that's the data set we're going to use. And what we're going to do with that data set first is to find all three of the measures of center. So we need to find the mean, the median, and the mode. So we'll start with the mean. So recall that with the mean, you need to add up all of the values and divide by how many there are.

So we already said there was nine houses on this block. So we're going to divide by 9. But let's do the sum first. So I'm just adding all of these up, making sure I don't type anything in wrong, and let's see if I got them all. It looks like I've missed one. Oh, I missed the 776. So I'm going to put that in there.

So it's always a good idea to go back and double check to make sure you have all the numbers so you don't make a mistake like I just did. So the sum ends up being 7,122. So really, what that means is that all of the houses on the block added together end up being valued at about \$7 million.

But now we want to spread that out over all nine houses. So we need to divide that by 9, the sample size. And this will give me something-- or it'll be approximately equal to about 791. And I'll just round that to the nearest whole number. So the mean is about \$791,000.

So we'll come back to that number, and we'll interpret all of them at the end. The next thing I want is the median. So with the median, what I need is, I need all of my data in order from-- usually, I do least to greatest. So I'm going to reorder the numbers, least to greatest. So 429, my next largest number is

568. Then I have 623, then 637, then 664, then 681, 776, and the big one, 2080.

So again, I am looking for the median. So with the median, the median is the middle number. If you have an odd number of numbers, it is the middle number. If you have an even number of numbers, it's the average of the two middle numbers.

So in this case, we have an odd number of numbers. So if I have nine values, that means that the median will be the fifth value. So we just count in, 1, 2, 3, 4, 5. This ends up being 664. So that means that since that's the median, I have four values on this side, four values on this side. So my median here is 664.

And now the mode-- the mode is the most frequently occurring number. And really in this case, there's only one number that's repeated-- those two houses that were next to each other. So your mode is also 664. And it's just a coincidence that these two numbers are the same, but still interesting.

So now if we take a look at all of these numbers, let's start with the mean. The mean notice is quite a bit higher. It's actually higher than all of the values, except for the big mansion at the end of the block. So what happened here is that remember that the mean is very sensitive to outliers. So it's sensitive to outliers, so that means that this \$2.08 million house is going to have a big effect on this mean. So it pulled the mean way up.

Now, the median, that number is right in the middle. So the median is resistant to outliers. So actually, a lot of times when they're quoting house values, they're going to quote you the median house value. They're not going to give you the mean, because if there is a mansion on the block, or let's say that there's a house worth very little, that's really going to change that mean housing value. So a lot of times, they're going to quote you the median value.

So a lot of times, if there are outliers, a median is generally a better value that will help you summarize the data. That mean is usually good too. And that's good, again, when you don't have outliers. But then again, the mean does, it tells you well, there's that big house at the end of the block. That it does have an effect on the mean. So sometimes, it is good to quote both values.

Now the mode-- for this type of data set, the mode is not very meaningful. Yeah, it does tell you that there's two houses of the same price. But other than that, it doesn't really tell you much. Generally, the mode is more effective when you have a large list of data. Or maybe if you have a frequency table, again, with many values, and you can get a range of values that occurs the most.

I will note that the mean is the most commonly used measure of center. And usually, it does do a pretty good job summarizing the data. When describing a data set, sometimes the mean does not provide a representative value due to outliers. In the house example, the \$2.08 million house pulls the mean up. The median provides a more representative measure of center. And although there was a mode in the house value example, the mode is more relevant when data is grouped into numerical intervals or categories.

I'll also note that when you're dealing with qualitative data, the mode is really the only measure of center that you can use. You're not going to be able to calculate a mean or a median for qualitative data. And we use the mode because then we can talk about the most frequent category.

So again, to recap, mean is sensitive to outliers, median is resistant to outliers. And the mode is generally more valuable in a larger data set, especially when it's grouped, when we're dealing with group data.

So that has been the tutorial on choosing which measure of center. Thanks for watching.