Robert_S2_L11

Wed, 1/12 12:25PM 10:32

SUMMARY KEYWORDS

standard deviation, variance, sample standard deviation, formula, population, larger, sigma, variability, values, population standard deviation, sample, smaller, square root, sample size, deviation, calculate, squared, denominator, canadians, data

SPEAKERS

Robert McKeown

R Robert McKeown 00:05

Hello, everyone. In this video, now that you've mastered standard deviation, a measure of variability, all the measures of average, we're going to look at statistical bias. So you've learned a formula for calculating standard deviation that you see in front of you on the slide. This is really the way to calculate the standard deviation when we know the population. And when we say the population, what we really mean is we we know what the true standard deviation is, or we have all possible values in a series. When we're working in statistics, often we only have a sample. So for example, the population of Canada is the population of Canada. But if you want to do a survey on what Canadians think about the say, politics, you can only you would only realistically be able to answer a certain sample, you might be able to answer ask questions to 1000 randomly selected Canadians, and get a reasonably good estimate of the population. But whether it's a reasonably good estimate of the population depends on a number of factors. And one thing that you can do, and one thing you'll be required to do is to adjust our formula for standard deviation. So let's go ahead and look at that together. Right now. Statistics is particularly useful when we use a smaller sample to make a prediction about a larger population. Now, if we already if we know the population, we have data on the entire population, then this is the formula that we want to use. However, we often do not. And when we don't know what the population variance actually is, and we want to try and estimate it using what data we have available, what values we have available, we need to make an adjustment to our formula up here. And we're going to do that by adjusting the denominator. And I'll point out a few things that have now appeared on the slide to you. One is if we use this formula here, turns out that it will underestimate the variability and will underestimate. Now, the population standard deviation is our sigma, our Greek letter, sigma, lowercase sigma, the sample standard deviation is the s, the letter S, and we're going to use a different denominator, it's going to be the number of observations minus one. Now ask yourself, what effect is that going to have on the variance it's going to make the variance or I should say the sample standard deviation, larger and minus one is smaller, it's in the denominator. So as is going to be greater than sigma. And it turns out that the, if we've got a smaller sample size, and we're trying to make an inference about a population based on that sample size, the smaller the sample is, the larger this bias is going to be. Allowing for a larger standard deviation is what we consider to be a conservative estimate. And that's not a political position. That is to err..err-err on the side of caution and assume variability is higher, at least higher in the population than it might appear based on a small sample Let's go through an exercise and we can see here using our two formulas, one our standard deviation where we're dividing by and our sample standard deviation while we're dividing by n minus one, and let's compare how the calculations differ. So suppose we have a sample size of two 10. So n is equal to 10. And are the sum of our deviations from the mean squared is equal to 600. That means we're going to have standard deviation 600 divided by 10. And take the square root of that, compared with the sample standard deviation, which is going to be 600 divided by 10 minus one, the square root of that. And if we make those calculations, you can see that the sample standard deviation is larger than than the standard deviation or the population standard deviation. And the sample standard deviation is 0.419. Larger. And so this is our the difference between the two. So we need to adjust our standard deviation to take into account basically, what we don't know about the population. Now as the sample size gets larger, and we're just going to hold the numerator constant, we won't allow it to change. So 600, and 600, it's not changing. Now we see that our sample and population standard deviations are still different core samples always going to be larger than the population standard deviation. But the size of that difference, the magnitude is much smaller. And by the time we get up to 1000, there's really not much difference at all. So you can imagine that as the sample size increases, the difference between the standard deviation and the sample standard deviation are going to become smaller and smaller. To summarize what we learned about the standard deviation, the standard deviation is larger, the values are going to be more spread out, they're going to be more dispersed, they're gonna have higher variability. They are, in some sense, the values differ more from each other, the magnitude of the distance between them is going to be larger, all else being equal, the standard deviation is sensitive to extremely large and small values, much like the mean. Generally, we consider that to be a good thing. Remember that an extremely large or small value is one that is far away from the mean. That is our we're not defining far away, quantitatively. We haven't defined far away. But we do have this concept that if something is extremely large or extremely small, it is far away from the mean. Eventually, we're going to use the standard deviation to help us put a measure on just how far away from the mean it is, we could talk about the variance a little bit, you've already calculated the variance, you calculated the variance, in order to find the standard deviation, you can think of the variance as being equal to the standard deviation squared. Like so if you're studying social sciences, you can become very familiar with the variance. Here's our formula for the variance. It's defined as sigma squared. And it's essentially equal to the standard deviation formula minus that square root at the end. So I could write the standard deviation being the square root of that thing, square root of the variance and of course, comes from simple algebra. So when we did that example, with the distance between from Calgary to various other cities, our variance was here, this was our variance, our standard deviation. The variance can be a little more difficult to interpret than the standard deviation. Remember, the standard deviation is stated in the same units as the values in the series itself. So if the values are in kilometers, the standard deviation is in kilometers. If the values are stated in dollars, then the standard deviation is also stated in dollars. The variance itself would be stated and variate deviations from the mean and dollar squared, so not quite as easy to work with as a Stan deviation. However, I'll warn you in future studies, the variance comes up very, very much comes up frequently. You can see that it's actually a little bit easier to calculate the variance in the standard deviation. You don't have to do the square root at the end. And the variance shows up in various different formulas and equations and, and different theories as well as in statistics.