Robert_S2_L15

Wed, 1/12 1:33PM 8:33

SUMMARY KEYWORDS

negative, correlation coefficient, correlation, equal, variance, series, covariance, deviations, squared, x minus, divided, factor, numerator, multiplied, bar, correlated, left, calculate, survives, summation

SPEAKERS

Robert McKeown

R Robert McKeown 00:05

Welcome. In this video, we're going to be looking at some of the characteristics of correlation, we'll look at why correlation is bounded between negative one and positive one. And we'll look at how to interpret correlation coefficients. As the slide says, It's no accident that our correlation in the previous example was less than one, we can illustrate this phenomenon by letting our Y series of variables here and their means be equal to the series X values and the series X mean. This means our correl-correlation coefficient is going to become an X there x with itself. All the Y's have been replaced with Xâ€™s. And we can see that we're multiplying the deviations from the mean and the numerator by themselves. Similarly, in the denominator, we have the variance of X twice the variance of X multiplied by itself. What does this look like? Well, we're going to end up with an expression that looks like this. The numerator here, this just turns out to be equivalent to the sum of the X minus X bars squared divided by n. And that, of course, is equal to the variance of X. And our correlation coefficient of X with itself is just going to be equal to one, there is no series of values, that moves more closely with another series of values than itself. If a series has a correlation of one with itself, a correlation of one must be the maximum correlation possible. A series with itself has a correlation equal to one, therefore, correlation cannot be higher than one, we can do a similar procedure, this time, we're going to let Y be equal to negative X. If we let Y be equal to negative X, we're going to have our X and negative X. So you can think of negative X as being if X value one says it's two, then negative X is going to be negative two. Or if you want to think of it in the real world, if you take two steps forward, the negative of your steps is going to be two steps back. When we make this change, and we let Y be equal to negative X, we get a change here. And we can actually factor out some of the negative signs. And we'll get and if we do that, we're going to get the sum of X minus X bar times X minus X bar with a negative and we can take that negative all the way to in front of the summation sign. And we did this by factoring out a negative one in the denominator, we might have had negative X plus X bar squared. But again, we could factor out a negative one. And we're just going to be left with x minus x bar squared, because we're going to have negative one squared times X minus X bar squared. The important thing for our analysis is this negative survives out here in front of the entire term. And so when that happens in there, you can see it again right here. That negative is going to make this factor right here, a negative one. And so the correlation can be no less than negative one because if we have a series and the negative event series, and no series can be less correlated with itself, or more

negatively correlated with itself, we'll talk about the term terminology in a moment. Then negative X, X and negative X. And so our correlation coefficient is always going to be greater than negative one and less than one. Let's work through an example together, let's show that the series Y has a correlation of one with itself to do that, and there's a few ways to do that. But the first thing we need is to know what the mean is. And the mean is going to be equal to 16 over four, which is just equal to four. Now we can calculate the deviations from the mean, and the among the Y values. One minus four, one minus four, seven minus four, and seven minus four. So we're going to have negative three, negative three, positive three and positive three. And if we want to know the covariance of Y with itself guess I can write it like that. We're gonna get negative three times negative 336 divided by four is going to give us a covariance of nine. What about the variance the variance of Y is going to be equal to I'm looking over here at the deviations from the mean, go negative three squared, plus negative three squared plus three squared plus three squared. That's going to be and all that it's going to be divided by four. The answer to this is again, 36 divided by four, which is going to be nine. And so if we want to calculate the covariance, or excuse me the correlation of y with itself, it's going to be equal to the covariance divided by the square root of nine times nine the variance the variance times itself within that square root and so we're just left with the correlation of Y with itself is equal to one