Robert_S2_L13

Wed, 1/12 1:33PM 11:45

SUMMARY KEYWORDS

summary statistic, covariance, calculate, series, numerator, divided, variables, deviations, central tendency, lower socio economic, square root, variability, expression, equal, standard deviation, higher, multiply, denominator, variance, values

SPEAKERS

Robert McKeown

R

Robert McKeown 00:05

Hello, everyone and welcome back. Previously you learned about summary statistics or descriptive statistics for one series. What was the central tendency? What was the average for a series? How can we measure its variability, or also called dispersion or spread and you learned about the mean and the median, you learned about standard deviation and variance? Now, we want to extend our analysis from one variable to two variables. How can we summarize and describe the way that two variables move or don't move together? Now, when we talk about two variables, or two series of values and how they move together, we talk we're talking about co-movements. common expression is to just call them correlation. So let's get started. And we'll talk about how two different series move with each other. Our measures of central tendency and variability, we're very good at summarizing us single individual series, like the height of a soccer player, or how many children were in a family. Remember, a family could have eight children, or could have one child, and having one child was much more common than having eight children, variability, or measures of variability, try and capture that in a single summary statistic. Of course, social scientists are interested in more than just understanding a single series they'd like to study and they want to understand how series move together. Social scientists are interested how two variables move together. Specifically, that means if a value in series A is higher, would we expect the corresponding value in series B to be higher or lower or the same? Well, I can show you some examples of what we mean. We might be interested in answering the question, Do people with more education have higher income doesn't necessarily mean that education causes higher income. But do people with more education tend to have higher income? Do people with lower socio economic status have a higher likelihood of committing a crime or a lower likelihood of committing crime or the same do businesses with better corporate governance have higher value is the company worth more when it has very good corporate governance? Now we're ready to look at how we would calculate or come up with a summary statistic that would summarize the co-movement between two series. Let's consider two quantitative variables or two series values, X and Y. And there are n observations. So we have n, n as the number of values in each of the series. And for simplicity, we'll assume that both series have the same number of observations. Now a measure of how two variables, two series of values move together is the Pearson correlation. And that's what we're going to look at calculating next. The formula is a little intimidating. But we're going to approach this methodically. And we're going to break it into parts. So that'll be much less intimidating, easier to remember. And you'll have a little bit of intuition behind how to calculate it, but also how to interpret it. So we're going to rewrite our very intimidating equation here in parts to better understand it. And to make it easier to understand, I want to show this expression to you right here. As in parts that you've already seen before, or divided into parts, some of which you've already seen. We can do that by dividing or I should say by multiplying by one over n, and we're going to multiply both the numerator and the denominator by one over n. And of course, one over n divided by one over n is just equal to one. And so when we perform this operation, we're not changing the actual value of the correlation coefficient at all. So we can do this and not worry about actually changing that value. Now when we multiply the numerator and denominator by one over n, we get this expression right here. Now the numerator is straightforward. The n Just divides the numerator that was here. The denominator is a little trickier. We've got one over n, outside this square root, we've got a square root here, and one over n is equal to the square root of one over n times one over n. To get rid of that dots, it's not confusing. And understanding that if we move the n into the square root operator, we get this expression right here where we've got this thing being divided by n and this thing here, also being divided by n. That's okay. We showed earlier that we can do this algebra without changing the correlation coefficient on the left hand side. Now come back a little bit later and talk more about the denominator here. Let's focus our attention on the numerator. So let's focus on this thing up here. It's actually new to you. This is new. That's not something you've seen before. And it is called the covariance. And so the name hence, that idea of a co movement, it's a covariance How To series are related to each other, or at least, they're related to each other through their co-movements. Here's our covariance formula. If we look at the numerator, here, we've got deviations from the mean. And we've got the product of the deviations in x times the deviations and Y. All that is going to be divided by n. So it's pretty similar to variance, except instead of squaring the deviations of X, we're going to take those deviations of X and multiply them by the deviations in y. And of course, we're going to sum up those deviations. So let's look at an example. We've got two series. Each of them has three values in it, we've got X, and we've got Y. So our first step is going to be to calculate those deviations from the mean. So we're going to need to calculate the mean of X and the mean of Y. The mean of X is equal to zero plus three, plus six divided by three, which gives us three nine divided by three. And the mean of y is two minus four, minus four, also divided by three, which is going to give us negative six over three, which is equal to negative two. That's let's calculate the deviations and X. So for X, we have zero minus three, three minus three, and six minus three, we have negative 303. And if we want to calculate the deviations in Y, we're going to have two minus minus two equals four. Negative four minus minus two is negative two. And the same thing again, negative four minus negative two is going to be equal to negative two. Now we want to multiply X and Y together, so we're going to have negative three times four is equal to negative 12. zero times negative two is equal to zero. And three, multiplied by negative two is going to give us negative six, one, add them all up, we've got negative 18. And now we're going to divide negative 18 by three. So we have here we have the sum of X minus X bar times Y minus Y bar is equal to negative 18. And if we divide it by n, and we're going to have negative 18 divided by three, and so the covariance is equal to negative six. Here often see it written like that. The covariance is illustrated by the sigma, much like the standard deviation, and the variance. And like the standard deviation, there's a subscript that tells us about that summary statistic we've calculated about this, in this case, the covariance. And we've got two variables in the subscript. So we know that this is a covariance, and it's the covariance between X and Y. What is its value? Well, in this case, the example we went through, it's equal to negative six