Robert_S2_L16

Wed, 1/12 1:33PM 8:18

SUMMARY KEYWORDS

correlation coefficient, correlation, negative correlation, negative, positive correlation, causation, coefficient, series, calculate, decrease, correlated, variability, variables, equator, lower income neighborhoods, gestation period, types, temperature, calories, magnitude

SPEAKERS

Robert McKeown

R Robert McKeown 00:06

Here we have a table for different types of correlation. Generally, or broadly, there are two types of correlation, there's positive correlation which we see right here. When we say positive correlation, we mean that the correlation coefficient is between zero and one. How to interpret a positive correlation, that means that if there is an increase in X, you can expect to see a increase in Y. Or if there's an increase in Y, you can expect to see an increase in X. Similarly, if there's a decrease in X, you expect to see a decrease in Y. Or if there's a decrease in Y, you expect to see a decrease in X. And with correlation, there's no causation. One way or the other economics is very interested and causation. Lots of econometric statistical techniques look at trying to determine causation. Correlation does not speak towards causation. It says that when some two series are correlated, X could be causing Y or Y could be causing X. Or neither could be causing the other they're just correlated. The other type of correlation is negative correlation. A negative correlation coefficient is between negative one and zero. If X increases, you would expect to see a decrease in Y. Similarly, if Y decreases, you expect to see an increase in X. If X were to decrease, you'd expect to see an increase in Y. Or if Y were to increase, you'd expect to see a decrease in X. So these are the two types, we've got positive and negative correlation. What are some examples of positive and negative correlation? Well, a positive correlation exists between a child's brain weight at birth and the gestation period, the amount of time that the child develops in the womb before birth. Another example would be calories burned, and how long you've been running. How long you went for a run or how long you went for a walk for. That's going to be positively correlated with calories burned. A negative correlation could be something like income, neighborhood income and the crime rate. Lower income neighborhoods tend to have higher crime rates, while high income neighborhoods tend to have lower crime rates. Another negative correlation is temperature and altitude. So many cities close to the equator, we think of being close to the equator as being hot. And so many of the cities that are close to the equator are built at higher altitudes where the temperature is cooler. Now the magnitude of the correlation is reflected and the correlation coefficient or if you instead of magnitude, we could call it the strength of correlation. So for example, if two series have a correlation of 0.4, and two other series have a correlation of point two, the series with a correlation of 0.4 is stronger than this series that has a correlation of 0.2. Similarly, but not identically, a correlation of negative eight and negative point eight, excuse me, of course, it

can never have a correlation, negative eight, a correlation of negative 0.8 is stronger than a correlation of 0.4. Even though negative point eight is less than 0.4. The correlation of negative 0.8 is stronger than a correlation of 0.4. Remember that, to calculate a correlation coefficient, you need to have at least two values in a series, you need to have two data points or two observations. Otherwise, you won't have enough numbers to calculate the correlation. We've already seen this but to remind you are with the underscore x y is the general notation for the Pearson correlation coefficient and correlation coefficients more generally. So for example, if you were looking at students scores on an IQ test, and their vocabulary and theirs and how they scored on a vocabulary test, you're good to follow Probably there's a positive correlation between the two. But also you might see a written like this correlation is described as R subscript IQ. Now, correlation measures the variability that is shared between two variables. And a little bit later, I'll show you how we can use how we can use the correlation coefficient to measure how much variability is shared between two variables or two series of values. Something to keep in mind if a variable is constant, then it has no variance, and its correlation coefficient will be zero with any other series. Something to keep in mind in practice when you're using correlation coefficients, if you want to calculate one for a research project that you might be working on, anytime you restrict the range of a series, all else the same, the correlation coefficient will be lower. Remember, the range is a measure of variability. If you are decreasing the variability in a series, then the numerator of the correlation coefficient the covariance is going to become smaller, and the correlation coefficient is also going to become smaller. The following table comes from Neil Sal Klein's book statistics for people who think they hate statistics. And it is a guideline to interpreting a correlation coefficient. So according to Professor Sal kind, a correlation coefficient between 0.8 and 1.0 is very strong. If we decrease the interval from point six 2.8, we could define that as strong. While point four 2.6 is moderate, point two, point four is weak and zero to 0.2 is weak, or none, you can also add a negative sign in front of those correlation coefficients and interpret them. Similarly, if we have a negative point eight to negative 1.0, or negative 1.0 to negative 0.8 correlation coefficient that would be very strong. And it would descend at intervals of 0.2. From very strong down to weak which would be between negative 0.2 and zero. In my discipline of economics, if we have a correlation coefficient equal to one, we sometimes say that this is perfectly correlated. And if we have a correlation coefficient that is equal to negative one, we say that it is and if we have a correlation coefficient equal to zero, we say that it's, well you could say I guess you could say it's uncorrelated or no correlation.