

Robert_S3_L14

Thu, 2/17 4:01PM 12:05

SUMMARY KEYWORDS

ice cream, crime, temperature, correlation, causation, click, cell, statistical techniques, copy, commit, press, sold, column, parentheses, high temperature, equal sign, highlight, decimal places, correlated, correlation matrix

SPEAKERS

Robert McKeown



Robert McKeown 00:07

We're looking at questions three and four. And we want to create a correlation matrix using the data on the crime spreadsheet. It's called a sheet in Excel. And the data we're going to look at are crimes, ice cream sold and daily temperature highest. So let's go ahead and click on the sheet labeled crime down in the bottom left corner. When we do that, we can see the title, which says there's daily temperature highs, quantity of ice cream sold and crimes committed. These are on specific days of the year. And our first column is the temperatures, then we have the amount of ice cream sold, and then finally crimes. So let's create a correlation matrix, we're going to highlight temp ice cream and crime. And I'm going to press Ctrl C, so we've copied it. And then I'm going to click on this cell right here, F, F 3- F 3, and I'm going to right click on it. And I'm going to click on paste special. And I'm going to choose transpose, I'm going to choose transpose and then I'm going to press OK, the other options, we can just leave as it is. And notice that this row has been copied as a an error, this row, row number two has been copied as a column in column F, then I'm going to copy the same area again, and I'm going to paste it in cell G 2. And so notice we've kind of created a matrix for ourselves, and the matrix is a four by four area, all matrices are the same. While no I shouldn't say that. It's a it is a matrix with equal number of rows as columns. And we can go a little bit further we know that the correlation of temperature with itself is perfect, perfectly correlated temperature with ice cream with itself is perfectly correlated, and the temperature or sorry, crime is perfectly correlated with itself. Now if we want to make this maybe a little prettier, we could highlight these cells click on Home and choose the centering options. So the the values and the titles appear in the center center of each cell. Next, I want to calculate the correlation between ice cream and temperature. And so I'm going to I'm going to select cell G 4, I press the equal sign, I'm going to type in C O R R E L for correlation. There is our Excel function CORREL, you can double click on it or just keep going. And it wants us to select two arrays. So we want to find the correlation of ice cream and temperature. So let's start with temperature. Remember, the correlation of temperature and ice cream is equal to the correlation of ice cream and temperature. So the order doesn't particularly matter in this situation. I'm going to scroll back up. Now I'll choose ice cream. And I press Ctrl Shift and down arrow and a highlight all the observations in the B column, close my parentheses, can see up here, what I've created as a formula. And then I'm going to press Enter. When I do that, I get a correlation. Now, I want to

change the format of the cell so that it's a little bit easier on the eyes. And maybe I will go to pay special and I will copy only the format into these cells, not the equations. So the next time when I now when I populate them now when we populate together, you'll see that they also have just the two decimal places. Okay, so let's click an equal sign in cell G for now we want the correlation between temperature and crime. Now we are going to press the equal sign, type in C, O, R, R, E, L, open parentheses. We know that temperature is in the A column so I'll go A 2 colon A 103. I guess it was A 3, A 102. And we want the array of crimes which is C 3 to notice the C 3 is highlighted once I type it in to C 102 close the parentheses and make sure there's a comma between the two arrays. And we see that the correlation between crime and temperature is 0.68. Lastly, we're going to look, we're going to calculate the correlation between crime and ice cream. So I'll press the equal sign in cell H 5 CORREL open parentheses, B 3 to, B 102, comma, C 3 to C 102. Press Enter. And now we've got the correlation between ice cream, and crime. If we wanted to make this area pretty say you were trying to present something to your boss, there's a few things you can do, you can go to Format Cells, you could add a border. So we might want to add a border outline in it. And we might want a heavy border outline in it. Let's pick this line here and click on the outline button. That'll make it heavy. And we might want a lighter, a lighter series of lines on the inside maybe once with some dots in it. Let's see how that looks. Not very pretty. I don't like it. If I press Ctrl Z. It's undone. Maybe.. maybe we were okay with just a heavier line along the outside line, like so. So that was our question three. Now we're under question four. And the question for us is to analyze the result. There's a few things we could do. The question didn't ask us to. But we could we could calculate the correlation of determination. And we would do that by taking the correlations and squaring them. And I don't think we want we would like to have just two decimal places, I think. And we can just copy and paste that formula into each of the cells. And we can see that we get our correlation determination, which tells us how much variation in one variable is explained by the other. Now we noticed that when there is a high temperature, much ice cream is sold more ice cream is sold. Right we see more ice cream sold on warm days, which makes sense, we would expect ice cream is a cool treat on a hot day. And the warmer it is the more ice cream is sold. Living in Canada, I can say that not as much ice cream is sold in the winter, as is sold in the summer. We also see that crime is correlated with higher temperatures, more crimes are committed when temperatures are higher. It's more comfortable to be out committing crimes, or maybe there are more people out and about that criminals can prey upon. But there seems to be a connection between crime and temperature. Of course, we don't know the direction of causation. But in this case, we can infer that committing crimes will not increase the daily temperature. Same with eating more ice cream will not increase the temperature the temperature is determined by the Earth and how much ice cream is sold. How many crimes are committed, those are human decisions so that causation could only be going one way. That said, we want to remember correlation is not causation, especially because if we look at this field right here, ice cream and crime. It seems like there is a high correlation between ice cream and crime. That's kind of an unusual result. And what is happening here? Is it possible that criminals after committing a crime like to have an ice cream? I suppose that's possible, or is it possible that when people eat ice cream, they're more likely to commit crimes? I suppose it's possible but neither of those situations seems very likely. What's likely happening here is that both crime and ice cream are to some degree cause by temperature Maybe cause is kind of an usual word encouraged by warmer temperatures. Crime and ice cream consumption are caused or they're being encouraged by warmer temperatures. Both are both are related to the same underlying phenomenon. So we see that these two otherwise unrelated human decisions or behaviors, eating ice cream and committing crimes, breaking laws are correlated. And so you have to be very careful about interpreting interpreting correlation. The golden rule is correlation is not causation. That is the golden rule that we really want. Anyone who takes a course in statistics to walk away with in the back of

their mind correlation is not causation. When we have a problem like the ice cream and the crime situation here, there are more sophisticated statistical techniques that we could use to disentangle the impact of temperature on ice cream and crime, such that we can try to isolate and see exactly whether ice cream eating ice cream is causing people to to commit crimes or if people who commit crimes like to have an ice cream on the same day. There are statistical techniques available. But here I wanted to show you this is a classic example of statistics where we see correlation is not causation.