

module_stat1_lecture10

Fri, 1/21 12:23PM 12:06

SUMMARY KEYWORDS

gdp, data, co2 emissions, chart, axis titles, plotted, relationship, capita, variable, scatter plot, interested, excel, axis, change, graph, onwards, trend, india, colder winters, insert

SPEAKERS

Sumon Majumdar

Sumon Majumdar 00:04

So in the previous clip, I talk to you about time trends, right how to plot trends over time of particular data, whether it's GDP COVID cases or anything else. Now, if you go back to the concept of a time trend, so what it basically is, is a relationship between two things on one axis is the time variable. The other is whatever is the variable of interest. Now, you can take that idea more generally, to examining relationships between any two variables. And oftentimes, in the social sciences, we're very interested in this idea of what's the relationship or a connection between one variable and another. So so an example could be that we could be interested in is that as an economy grows faster and faster, does it also produce more air pollution? So you could be interested in the relationship between a country's GDP per capita and its CO2 emissions per capita? What's the relationship between the two? Does it as one goes up as X goes up? Does also Y go up? Right? As yearly GDP goes up, does it also mean that yearly CO2 emissions? Now the lots of this sort of relationships, which are of interest to social scientists, for example, the study more increased marks? Or, for example, if you spent if a city spends more on its police force, does it reduce crime? As do hot, how to hotter summers are colder winters have an impact on a city's budget. So you could be interested in in this sort of relationships as a social scientist. And the first thing in studying any such relationship is people try and plot a graph to see if they can visually see some connection between the two variables before digging deeper. Right. So if you wanted to ask this question that is there a relationship between yearly GDP per capita and CO2 emissions per capita? What would you do? Right? What do you might do is gather data on this X. So year by year, gather data on the GDP per capita, year by year, gather data on CO2 emissions per capita, right? And then do a graph of Y versus X. So such a graph in statistics is called a scatter plot. So here is an example. So what I've done is for Canada, I've collected data on from 1970 onwards, right? So for each year, I've looked at what is its GDP, and what is the CO2 emissions per capita, then the same for 1971? Same for 1972. so on up to 2018, right? So I have, these are my data, right? So for each year, I have value of variable X and a variable one for variable x and variable Y. So that means what I have is data for XT, and YT. And then I've gone into Excel and plotted this. So what does this mean? So if we look at the data point, like here, so this is the value for a particular year, where the were on this axis is the GDP per capita. So the Canadian GDP per capita was 15,000. And the CO2 emissions per capita was 15.5. Right? So this was some other data point where this was the GDP, and this was the co2 emissions. And now having plotted this, so what are interested in is that can we

see any relationship between the two? Can we see if this is upwards, downwards, or anything else? And just looking at this plot, per se, right? At least for Canada, we don't see an obvious connection between GDP and CO2 emissions. Now, of course, this doesn't mean that we shouldn't dig deeper and look more closely at the data. But right off the bat, this is a good visual tool, which can give us an first glance at how the data looks like. So the scatter plot is a very, very important tool in the arsenal of more social scientists. Now, I did the same thing for another country, India. And here, so again, from 1970 onwards, I looked at the Indian GDP per capita, and its CO2 emissions, right and plotted this on the X axis and the CO2 emissions on the Y axis. And now if you look at the Indian case, right, what they does seem to be, there's an there's a positive relationship or an upward relationship between, like as GDP per capita is increasing CO2 emissions also seem to be going up. Now, whether this is actually a causal effect, whether it's because GDP went up, that's why CO2 emissions went up, or was there some other reason, one has to dig deeper to understand that, but right off the bat, this gives you a good visual representation of this underlying data. So what I want to show you is that how you can create the sort of charts in Excel pretty easily, just using one command. So let's do that. So back now in my Excel sheet, and under this sheet, which says GDP and emissions, I have here data for Canada. And here I have data for India. So first, let's look at the Canadian data. So have data from 1970 onwards, and going down up to 2018. So what I'm going to do is let me select these two GDP per capita and co2 emissions. And then, I'm going to go to Insert. And I'm going to go to charts. So here see it says Insert scatter or bubble chart. So I'm going to insert a scatter chart. And this gives you this GDP per capita is on the X axis and CO2 emissions is on the Y axis. Now see, if you look at this graph here, right, it doesn't give you the axis labels, it doesn't give you it doesn't tell you this is GDP, this is CO2 emissions. So what I can go in and change those sorts of things, right. So if I go to chart elements, here, if I asked you to put in axis titles. So here if I put in axis titles and going here, right, and I can type in my axis title, this is GDP per capita. Same thing here, if I go into the the Y axis and go into axis titles, I can put in CO2 emissions. Furthermore, see, if you look at the data, there's a lot of blank space here, where there are no data points here. So what I can do is, see it has plotted this, the scale from zero up to 20. But see there are no data points below 14. So if I want to get a closer look at the data, what I can do is I can change this, right? So see, it says the minimum of this axis goes from zero to 20. I don't need to start from zero. I can start from let's say 14. So I put in 14 in the minimum and hit this. Now you see it helps you get the data more upfront, right more closer to you. Right and you and you can more clearly see if there is a trend or not. As with others, right You can also change a lot of stuff with this, you can change the chart design, if you wish, you can make it look like this, or you can leave it as I had, you can change the colors, you can change the colors, change it to something else. That's your choice. So this is how you would produce a scatterplot trying to look at the relationship between two variables. So let me quickly do that also for India, right, so here I have for India. So I'm going to select the data. Again, go to insert, and put in a scatter chart. So this will be the chart for me. And let me move it to this side so that we can look at it. Right. And here you see for India, there seems to be a positive upward relationship between GDP on this axis and co2 emissions on this axis. So what I would like you to do, I'm going to stop the clip here. But what I would like you to do is, is going to use this Excel spreadsheet, try and use this data to generate these charts. Put in axis titles, change the minimum and maximum point on the axis, put in a chart title, and change the style of the chart if you wished. So again, as I keep on emphasizing the way to learn a software like Excel, is to actually do it hands on. So I do hope that you're not just watching the videos, but actually trying your hands in Excel.