

module_stat1_lecture11

Fri, 1/21 12:23PM 11:48

SUMMARY KEYWORDS

life expectancy, index, education, income, plot, graph, excel, variable, country, scatter diagram, point, relationship, title, scatterplot, upward sloping, horizontal axis, scatter plot, chart, legend, richer countries

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So in the previous clip, I showed you how you can look at the relationship between two variables X and Y by plotting a graph of Y versus X. And we did that by using the scatterplot of the scatter diagram function in Excel to generate a graph between the X variable and the Y variable. Now, the scatter plot is a very important tool and a very widely used tool. And it's one of the first things that people often make use of, to try and see if there is a relationship between between two variables. So in this clip, I'll show you another example of doing the scatter diagram, using Excel just to make sure that you understand the procedure. So what I'm going to do now is do an example which we covered before and asked you to do as an exercise in a different context. So this is the three components of the human development index. Now, the human development index, this is published every year by the United Nations. And it's often used to rank countries in terms of the development. And this has three components. One is the income index, which gives an idea about the the richness or poorness, just in terms of wealth or income of a country. And this income index, it ranks from one to zero. So what are the richer countries, so closer to one, it's the country is rich, closer to zero, the country's similarly, it has an education index, which tries to measure the total amount of education of adults, the expected education among children and so on. And again, this education index runs from one which has very high education to zero, which has very low depression. And the third component is a health component, which is the life expectancy index. Again, this runs from zero to one. So question that social scientists may be interested in is that isn't that richer countries have higher levels of education. So or do richer countries also have higher life expectancy? So you may be interested in the relationship between this variable which is income, and this variable, which is education? So how do we plot so suppose we select this data. And now we want to plot a scatter plot between these two. So I'm going to go to insert, I'm going to go to my charts, and pick the first one, which is the scatterplot. And this gives you the relationship between the X variable which is income, and the Y variable, which is education. So first thing, let's put in some axes on them, right, so let's go to the chart element, and put an axis titles. So here on the on the X axis, so this is, this is going to be my income variable. So one way to do I can either type out that it is the income index, or if I go to the tab here and say equal to income, right? So Excel that knows that whatever is in this cell should be the title of the horizontal axis. So if I hit this, see, it changes the horizontal in title into income index. And

you can do the same thing with the axis title with the vertical axis title as well. Second thing is that if you want to bet, let's give it the chart title, right? So this is education versus income index, across countries. So the set this gives you the title of this chart, the third thing, if you look at this chart, there's a lot of whitespace. Here, there are no points in this part of the graph, right? And the reason is because here Excel is plotting from zero to 1.2. And this way, it's plotting from zero to one, but there are no observations where the X value is less than point four. And they know observations with a Y value is less than point four. So we can adjust that. So if I hit, if I double click on the vertical axis, it opens up a tab on the side panel on the side. And here, you can choose the axis options, you have the bounds, the minimum and the maximum. So what I'm going to put in the minimum is point four, so that it starts from point four. So if I hit that, now you see, so the vertical the Y's, they only start from point four, I can do the same thing for the horizontal axis as well. Right, so I've double click, and again, I want the minimum to be point four, the maximum goes to about one, so let me just do up to it, right. And this gives you that this axis, you know, runs from point four to point. And what that does is that it eliminates some of the extraneous parts of the graph, and lets you focus on, which is the more important portion, which is this. And you can do other stuff, like change the color of the dots, change the style, if you want to change it to a darker hue to something like this, right, you can take care of would be me show you one other thing. So the second thing is we have plotted income and education, right? But there's also the life expectancy index. Right? So this is a third variable. What if we wanted to plot this against income? Along with education, now, Excel allows you to do that pretty easily, what do you just have to do is you go to the corner of this last cell, and drag it out to here. And then it it automatically knows to include this life expectancy index, also, in the plot, right, so those are the orange dots. So now, what you want is actually, you want a legend that which one does the orange dots correspond to which one do the blue dots correspond to? So what you can do is go in here and put in the legend, that will tell you that the blue one is the education index, the orange one is the life expectancy index. Now, if you don't like this, so what this is done is that these two legends come in on the side, right, and this is squished the graph a little bit, it's if you don't like that style, you can choose a different style like this, or like this. Where where you have the legends at the bottom, and the graph is not squashed. So let me change the access title as well. Right, so let me go in there. And put in this is education, coma life expectancy index. Now, if you don't like the vertical look of this, you can again, click here, and and there'll be text options, sorry, title options here. Say it will give you the vertical alignment if you want and the text direction. If you want this to be horizontal, you can do that as well. So this is the very end, you have lots of other options, you can if you want to do it instead of horizontal, if you wanted to do it at an angle, you can do that as well by setting customer angle. So there's lots of things to play around that Excel gives you. Now before ending this clip, I just want to talk about what comes out of this right. So at the end of the day, what do you want to do is analyze this right? And what do you see is that if you look at the education index, right, so if you see the relationship between income and education, you see that this is mostly an upward sloping relationship. So Most of the time, countries with a higher level of income, they also have a higher level of education. Same for life expectancy, mostly this is an upward sloping relation ship, where most of the time where countries which have a higher level of income, right income is increasing in this direction, and life expectancy and this, right, so countries which have a higher level of income, they also have a higher level of life expectancy. So, it's not always true. So, for example, if you compare this country and this country here, right, so, this is a country which has a higher level of income, right, but it has a lower level of education. Right, similarly for life expectancy. So, there are some countries here, right, where all the income is higher life expectancy is lower than this country here, right. But overall, because statistics is, is not just looking at one point or two points, but you take a bunch of points and try and see, on average, what's the relationship? On average, here, the

relationships seems to be a positive one. So, this is how you create a scatterplot in Excel. And as I said before, this is one of the first things that people look at when you're they're trying to figure out if there is a positive, negative or no relationship between two variables.