

module_stat1_lecture6

Fri, 1/21 12:23PM 12:02

SUMMARY KEYWORDS

frequency distribution, cedar, chart, data, pie chart, column chart, trees, hemlock, suppose, forest, pine trees, spruce, bar chart, excel, count, type, range, construct, distribution, called

SPEAKERS

Sumon Majumdar

Sumon Majumdar 00:05

So in the previous clip, I showed you how to construct a bar chart or column chart from a given data on with the frequency distribution. So here in this clip, I'm going to show you that what happens if you don't have a frequency distribution. Suppose you just have some data, and you want to construct that frequency distribution. Because once you have the frequency distribution, you know how to create these charts, right. But what if you didn't have the frequency distribution, because in the previous example, we already had the frequency distribution, for example, on the COVID cases, or on the violent crimes. So this time, let's do the example from geography. So suppose you're interested. So let me go to the next sheet on my Excel sheet. And now suppose you're interested in the type of trees that are in a particular forest. So obviously, you cannot go out and collect data on every tree in that entire forest. So what typically people do is that they look at a small part of the forest a small lot. And look at the distribution of type of trees, the numbers, the distribution of trees in that lot, and from that make an inference about the rest of the forest. So for example, here, I've collected this data here, or rather made up this data. So say you have sampled a lot of 30 trees, right. And you have written down the different trees, that there are in this lot. So if you look down, you have fine for Hemlock, spruce setter, etc. Right? And now suppose you want to construct the frequency distribution. So first, let's write down the different types of trees that there are their spine. There is far what else is there? There's hemlock. There's also spruce, spruce. And what else? Can I see the cedar? Is there anything else? Right? So if I go down this column A rights if I keep going? I don't notice anything else. So now, if I want to get the frequencies of these, right, so suppose I want to get the frequency, the number of pine trees that there are in the slot? So one way would be to count it, right? One, this is two, three, and so on. And that will be fine. But if if you have a much larger here, there are only 30 data points. So that's easy to do. But suppose you had 300. Right, then obviously, it will be tedious to count. And Excel actually allows you to do that in a pretty easy way. So let me show that to you. So what you use is an inbuilt function called count if. So this tells you the counts the number of cells within a given range that meet a certain condition. So firstly, right, so we the count if the first part of it is the range, right? Where are we looking, so we're looking from A to B, and up to which point. And if you keep going down, the last point is a 31. So that's the range over which we are looking. And secondly, we have to specify the criteria. And the way it works in Excel is that you specify the criteria within inverted quotes. Okay, so I haven't put in the criteria here, the criteria here is

fine, because that's what we're looking for here. And suppose I hit enter, this tells you that the number of pine trees in the slot is 11. So that saves you a good deal of effort. Now, similarly, if you want to go to four, right, so what we can do is that we can copy and paste this, just expand this here. The problem is now let's let's go to the four right, so see when I copy it from the first cell down to the second cell, but it also changed my my range, it went from a three. So I want to go back to a two and the last one is a 31 and this time I'm looking for four so let me change 4.24 So that means there are five fourths in the sample. Similarly, now if we go to Hemlock, right, again, notice that the range has changed, I want to change it back to a two to a 31. And this time looking for hemlock. So the only three hemlocks in this lot. Now if I go down to spruce, a spruce, then it's again, a two to a 31. And this time I'm looking for spruce that gives me eight and four cedar again, let me go and change that range a two to a 31. Hit this, this will give you the number of cedar in this sample, and see this is three. And if you go down this, you can verify. So this is one cedar, this is two cedar, and there's a third CEDR that we check. Look, there's nothing more, right. So Excel, you can use the count if function to figure out the frequency distribution of the data. And now once you have this frequency distribution, now you can do lots of stuff with it. So for example, if you want to create a bar chart here, again, we go back, go to insert, and do exactly the procedure that we did in the last clip, right. So if we want to create a bar chart, let's just hit that, that'll give you the bar chart of that sorry, there's the column chart of pine for Hemlock, spruce cedar. If you want to change the color, for example, if you want to make this green, as these are trees, you can do right, you can change the chart title, again, by writing this the the frequency distribution of tree types. So this is how you can create actually a frequency distribution from a given data. And then once you have that, you can use it to construct a bar chart column chart, whatever you wished. Now, there's another type of chart, which is also often very useful, is because see, what you're probably interested in is what's the distribution of these trees in the forest. Right. So what fraction of the forest consists of pine trees? What fraction is for Hemlock, spruce, ce-and cedar. So there's a different type of chart, which is often useful to show you the composition, or the percentage distribution of, of different things. And that's called a pie chart. So let me show you how you can do that in Excel. Again, I'm going to select this data. I'm going to go to Insert. And see there are charts here you see this round circle one, where it says Insert pie or donut chart. So if you hit this, this gives you the pie chart. Let me change the type, right, so. And so see this is the pie chart. So what it does is that it shows you the breakup, the percentage break up of the different types of trees that there are in this forest, right so 36% pine 70% is fir 10% is hemlock. 27% is spruce, and then a cedar. So it shows you pretty nicely and pretty immediately that the biggest fraction of trees in this forest is pine. Then comes the yellow which is spruce, then comes this one the orange which is fir, and then spruce and cedar, sorry, this is ce- gray is Hemlock and dark blue cedar. These two are in equal proportion. So this is called a pie chart. And we can again right So this is the percentage percentage composition of tree types. Again, as with the bar chart, you can change lots of things, you can change the chart styles. There are lots of different styles here. You can pick according to your aesthetic, your your sense of aesthetics. And in fact, you can also there are two interesting ones, which you may see there's a 3d version of it. And some people like this. And if you often see this, sometimes portrayed in the media, but But essentially, it's the same data just presented in a 3d fast fashion. So this is called a pie chart. And let me sorry, as you saw that, once I expanded it, you can break it up like this, or if you want to go back, and I can, I can move this chart anywhere that I want, right, so I'm going to move it so that it doesn't interact. interrupt my column chart here. So this is a second type of chart that I'm presenting to you here. It's called a pie chart. So what I've shown you so far is three ways of presenting data for a qualitative variable, either a bar chart, column chart, and also a pie chart. So let's stop here. And I'm going to continue with this in the next clip

