

module_stat1_lecture3

Fri, 1/21 12:23PM 13:27

SUMMARY KEYWORDS

variable, data, qualitative, examples, number, quantitative variables, discrete, numeric, rainfall, talking, observation, similarly, called, measure, natural, categorical, measurement, fraction, categorical variables, job vacancies

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So in the last clip, I talked about different sources of data in social sciences. So in this clip, I'll be talking about the actual data and the different types of variables and data that there is. So first thing for talking about data, we need to talk about what the data is about. So this introduces the term of variables. We've talked about variables earlier also in probability theory. And the same idea carries over here as well. So when we talk about variable X, so this means some kind of characteristic which we measure. So let me give you some examples. So it could be the daily value of a stock, it could be the monthly income of an individual, it could be the amount of rainfall or the water level in the lake, the number of protests, the number of votes. So these are all things which can be measured. So all of these are examples of variables. Now, for any variable, there are three very crucial elements that one needs to keep in mind. So first thing, what is it? What is this variable measure? Right? So what's the description? So if we're talking about here, right, so this is, if we talk about X being the daily value for stock, what does it measure? It's the value of a particular stock. The second thing is, what's the unit of measurement? So for, for this example, it will be in dollars, right? But then the question is, is this in Canadian dollars? Or is this in US dollars? So you need to specify that thirdly, what's the unit of observation? Is it? So here? It's the D, right? It's per day. But now, suppose I say, okay, here is the variable, which is the amount of rainfall. Now, I'm not become-not been very specific about this. Because, okay, it-we have the description, it's the amount of rainfall. What's the unit of measurement? Are we measuring it in litres? In feet? In centimeters? What's the unit of measure? And thirdly, most crucially, what's the unit of observation? Is it the amount of rainfall per day? Is it per month? per year? Is it for a whole country? Is it for a particular city? Is it for a particular district? What is that? So for any variable, you need to have these three elements? Its description, what is the unit of measurement? And what's the unit of observation. So let's do a specific example. So suppose the variable is the number of new COVID cases in Ontario per day. So here is the description is the number of new COVID cases. Again, the unit of measurement is is the number, right? And and this is, it's in Ontario per day. Right, so those are the three elements. Now, once you have the variable, and you go out and collect observations on it, so that's what generates your data. So datum is the singular, and data is the plural. So if you have a single observation on that variable, then that would be called datum. If you have multiple observations on that variable than that would be called data. So if you're now go out

and look at a try and measure this particular variable, X , which is the number of new COVID cases in Ontario for day, so I'm recording this in this in on December 8. So I went to Ontario's COVID-19 website and looked at the data over the past seven days. So I got this as my data, right. So these were the number of new COVID cases in each day, over the past seven days, that gave me these seven observations. So that constituted my data. Now we've talked about variables in general. Now, let's think a little bit more systematically about the different types of variables that there are out there. And you'll see that the two main types of variables, one is numeric or quantitative variables, and the second is non numeric or qualitative. So what do I mean by that? So the first type is what's called quantitative or numeric variables. So this is where the data is numeric or in numbers. So examples of this are most common, right things like income, rainfall, the number of COVID cases, the number of votes, but this is all where the variable takes numeric values. On the other hand, data could, the variable could also be qualitative or categorical. So this is where the data is not numeric. So what are examples of this is it could be the religion of a person, it could be the major, the undergraduate major, that the subject that you're majoring in, it could be your opinion about something. So this is where the data is what's called qualitative. And it's also called categorical. So why categorical, because for example, if you look at religion, right, it could either be, if you ask a person's religion, he or she could either be Christian, Muslim, Hindu, Buddhist, Jewish, or could be unaffiliated with any religion. So you check mark one of these categories. So that's why this type of data is also called categorical. Similarly, if you ask someone, their, their major what he or she is majoring in, it's either biology or geology or sociology, economics straight, so it's, it's a particular category. So that's why these are called qualitative or categorical variables. Now, within these, you will see that there are some sub classes. So if you look at quantitative variables, there are two sub classes. One is where the variable is discrete. So it takes some specific particular values. And nothing in between. So let me give you some examples. And that'll clarify this. So if we look at the number of children, or the number of protests, so the number of job vacancies, these can either be 01234 and so on, it cannot be something in between, it cannot be 1.6 Children, it cannot be 2.72 job vacancies. So these are examples of discrete quantitative variables. We dealt with this term earlier also in probability, and it's the same meaning there and here as well. The other type of quantitative variables are what what's called continuous, so they can take any value, including all in between values, including fractions and deserts. So what are examples of this? So examples of this are things like if you look at the GDP growth rate, it could be point 6%. Or it could be 1.35%. We look at the fraction of the population vaccinated, it's a fraction.



Similarly, if we

S

Sumon Majumdar 09:01

look at the amount of rainfall, it could be, technically, it could be 10.67 centimeters, right? We typically round that off to level. But if we look at the scientific measurement, it could be any number between not just integer numbers. So these are examples of continuous quantitative variables. And it's also easy to sometimes move from one to the other. For example, if we look at the number of crimes, so this will be some discrete now, that will be can 10, 12, 15 or depending on like it's- now say the number of crimes per day. But if we look at the crime rate, so things like the number of crimes per 100,000 of the population. So what you will do is take the number of crimes and divided by the by the population Right, so per 100,000, so that will

immediately be a fraction, and it will be a decimal and a continuous variable. Right. So in terms of quantitative variables, they're either discrete quantitative variables or continuous quantitative variables. Similarly, if you look at the other type of variables, which is the non numeric, the qualitative variables. So remember, these are categories, right. So they, they are in particular the categorical variables. Now, this could also be what's called ordinal that is, there is some natural ordering of the values that these variables can take. So, what are examples of this? So, an example of this is for example, if we look at opinion, right, so, people could either say okay, what's your opinion about this? You could say I strongly disagree somewhat disagree somewhat agree strongly agree. So, there is a natural ordering, this is the most negative to this as the most positive. Similarly, suppose I ask someone's wealth level, right? And it could either be poor, lower middle income, upper middle income or rich. So there's a natural order of things. So, that's why this type of variables are called ordinal. Qualitative. On the other hand, they could also be qualitative variables, which don't have any natural order. So, these are called nominal. So, what are examples of this? So, examples of this are things like religion, right. So, as I said, it could be Capitole Muslim, Jewish, Hindu, so, there is no natural order of these are your major biology economics, geography, social sociology, right, there is no particular order. So, these are called nominal qualitative.



So, in this



Sumon Majumdar 12:17

lecture, so, I introduced you to the different types of data variables that there are two broad categories quantitative, which are numeric and qualitative, which are categorical. So, under quantitative, they could be either a discrete set takes some particular discrete values, or it could be continuous that means it takes all values in between. And on the other side, in the qualitative side, there could be either ordinal, in which case there's a particular order to those variables or it could be nominal in which case there is no particular order. So, remember this and in any particular variable, try to identify which category or which type it is. Okay? I don't want to use the word category because you know, then we're talking about categorical variable. So which type of variable it is.