

module_stat1_lecture2

📅 Fri, 1/21 12:22PM ⌚ 14:00

SUMMARY KEYWORDS

experiments, data, surveys, similarly, social sciences, differ, people, impact, economics, called, organizations, aggregate, suppose, lots, talk, outcomes, country, electoral, teacher, generate

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

As we've talked about in the previous video, statistics is all about data. So the question is, in the social sciences, where does data come from? What are the different sources of data that you may encounter in the social sciences? So let me talk about that in this video. So one of the commonest sources is the data that's already out there, and somebody collects it for you. So this is what I mean by data, which comes from collection or observation. So what are examples of this? So, for example, what's the web pages that someone is viewing. So this has become big data and an important business model for a lot of companies these days, like Facebook, Google, etc. So what they're doing is they're collecting that data, and then probably selling it or using it for some other purpose. Other sorts of data are, for example, crime data, police have records of crimes, right. And then they aggregate them to produce crime records. Similarly, if you look at economics data, there's a lot of economics data out there. And either the, every country collects it, or there are organizations like the World Bank, or the OECD, who aggregated and give you the data across lots of different countries. Similarly, there are others like, for example, TV, you would look at Nielsen's for other types of data, you'd look at other organizations. So some of this data is proprietary. That means you have to pay for it if you want to access it, or maybe they don't want to sell it to anyone else. And some of it is very commonly available. So for example, the World Bank data, or Statistics Canada data on crimes, these are easily available on the web. So this is one important source and common source of data. And what's a second source? The second source is where you have to generate the data because the data doesn't automatically exist. So a common form of this is what's called surveys. So, for example, like many of you have probably taken opinion surveys about something. Anytime you call customer service, or most major companies at the end, they ask you whether you'd like to take a survey, right, so that generates data for them. Now more common types of surveys are things like what's called Canadian Labor Force Survey, or the US equivalent, which is the current population survey. So these are surveys, which are done monthly basis. And what they do, for example, in the Canadian Labor Force Survey, every month, they survey about 100,000 individuals, they asked them about various things are the are you employed? Are you looking for a job? Are you not looking for a job? So things like that. And that's where, like, anytime we we talk about unemployment numbers this month, or the unemployment numbers decreased or increased from last month, most of that data comes from these labor for service. So these are data which is generated through service. These days,

you will see that there are lots and lots of COVID related surveys. How has COVID related employment, how has COVID Related affected businesses? How is COVID Related affected students, parents, like there are lots and lots of service, all of them are generating data, trying to understand how COVID has impacted different segments of the population. Now, there are also a lot of electoral polls, for example, right? And these you see much more of these before in the elections. So if you go to this website, 338canada.com What it does is that it aggregates a lot of this electoral polls. The Electoral polls are run by newspapers, TV organizations, and lots of different by the parties themselves. Right. And websites like this. They aggregate those those polls from many different sources. So the name 338 Canada, I don't know if you if you recognize why it's called that way. So it's because there are 338 seats in the House of Commons in Canada, right? So. So hence the name 338. So if you want to get a majority government, you want to get at least half of that many seats. Now, as I said, there are also lots of opinion surveys which are done by firms and by different organizations, right, even at the end of any course that you take in university, there is usually a student survey about things like the effectiveness of the course, effectiveness of the teacher, how much have you learned, things like that. And the Pew Research Center, this is a very interesting organization, which does a lot of opinion surveys about a very wide number of things in the US. So if you go to their website, you will see they have done things like recently do things like what's attitudes and users of Twitter among us adults? Right? So for example, they found, like they asked a question in that survey, do you think Twitter is good for demand for democracy? And what they found is that there was a difference between how Republican people who support Republicans differ in their answer to that question, most of them think that Twitter is bad for democracy. On the other hand, if you look at people who support the democracy, there's a much bigger chunk, who think that Twitter is good for democracy. So they do lots of interesting surveys like this, where they go out and maybe serve 1000, maybe 5000 people about their opinions on different things. They cover religion, attitudes towards immigrants towards religion, or towards the economy, towards government, and very interesting stuff. Right. So that's another source of serving of data, which is service. Now, there is a third source of data that you may encounter in the social sciences. And this is experiments, then, you know, in science experiments are very common. But these days, you will see that experiments are also very common in the social sciences, especially, in things like economics, sociology, psychology has always been big on, on experiments, but these days, it has percolated through lots of different social science areas. Now, if you look at experiments, right, there, essentially three types of experiments. One is what's called lab experiments. Now, lab experiments are literally conducted in a lab that, so, instead of mice, right, the the subjects here are human. So it could be things like you show someone a video of a particular situation, and then ask how they react to it. So for example, you may show the same sort of video. This kind of experiments, there are some celebrated experiments like this, maybe you show the same sort of video of some conflict. And you showed maybe two people from one ethnicity versus people of another ethnicity, and then see how the reactions differ between the two, although they're watching the same video. In economics, there are experiments where they ask people that okay, here is \$10, you decide how you want to split it between yourself and another player. And then we look at how that decision differs depending on what the other player is whether the other player is a male or a female, whether the other player belongs to the same ethnicity, same religion as you or not. So these are kind of experiments which are run in the social sciences. So obviously, they generate a lot of data. Because you run this experiment across, let's say, 20 people, and then see how each of them has differed in terms of how they want to split that \$10. Now, other type of interesting experiments are what's called field experiments. So these are experiments which are run in real life settings. So, if you've been following the Nobel Prize, the recent Nobel Prize in Economics went to three economists who are known for doing field experiments, Avijit Bannerjee, Esther Duflo, and Michael Kramer. And they have done very interesting experiments

in a lot of Asian countries and a lot of African countries. For example, they said, Okay, teacher attendance in schools is a big problem in some parts of India. So they say, okay, so suppose we give bonuses to teachers based on their attendance? How much impact does that have? Similarly, they said, Okay, mosquitoes, suppose we give mosquito nets to some families in Africa. And we don't give mosquito nets to some other group of families, let's see how these two impact their health and their other outcomes. Similarly, since it was that two groups of individuals, one group, they made it easier to access a loan, the other didn't, right? How does this impact their economical outcome? Again, this generates a lot of data. And you can analyze this. The third type of experiments are what are called natural experiments. So these take advantage of some natural or policy occurrences. So what's examples of this? So for example, suppose you want to know, this scholarship, help, education outcomes. So what do we look at this, that maybe to avail of a scholarship, you need to get marks above a certain cutoff, let's come up with a cut off of, let's say, 90%. So what they would like to see this for those people who are just slightly above 90%, and that's why they got the scholarship, versus those who just slightly below 90%, let's say they got 89%. And so they couldn't avail of the scholarship. So how did this impact right, this cut off policy rule? How did this impact the outcomes of people who got above 90, and therefore were able to avail of the scholarship and those who just slightly below and therefore couldn't avail of the scholarship. Similarly, in terms of a natural occurrence, suppose you look at a country and one part got affected by an earthquake, another part didn't. And let's look at the growth of the part of the country which got affected by the earthquake, and the part which did. So this is a natural experiment. Similarly, maybe oil was discovered in one part of the country or in one country and not in the neighboring country. Let's look at what's the consequences of that for the country, which discovered oil and the other which did not. So these are natural experiments. And again, you can get a lot of data from that and use and use that data to analyze whatever is the question that you want to analyze. So these are the three main sources of data that you commonly see in social sciences, one is collection or observation, secondly, surveys and third, experiments. So let me stop here with this. And in the next video module, I'll be talking about data variance.