

module_stat1_lecture1

Fri, 1/21 12:23PM 10:55

SUMMARY KEYWORDS

data, social sciences, statistics, question, justice system, theories, policy, opponent, deduce, module, prosecutors, predictions, equal, professional leagues, scholarships, access, market, economics, normative, problem

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So welcome to our modules on statistics. In this first module, I'll introduce you to the idea of statistics. So what I'm going to cover in this module, I'll tell you why statistics is important in the social sciences. And statistics is all about data. So the important question of where does data come from? And what are the different types of data? What are the different types of variables. So that's what we'll begin this module with. And then in the latter half of the module, I'll show you that once you have data, how to visually represent those. So the first thing is that if you think about the social sciences, there, in terms of the broad picture, you will see that there are a couple of big questions. So one question is what's called the normative in the sense as what ought to be? So what's an example of that? So for example, we, we cannot but all agree that the justice system should be fair and impartial to everyone. So that's what ought to be an ideal justice system. But the other side of the coin is what's called a positive question that what is it? What it is, right? Is the justice system actually fair and impartial? Due to people who have committed, let's say, the same similar sort of crime? Do they get the same punishment? Okay, or is it different? So this is a sort of fundamental question in the social sciences. And similarly, if you look at economics, but so a question could be that how should the markets look like? And one of the one of the things that is often discussed in economics is that the markets should be perfectly competitive, no one should have an extra or monopoly power. But that, so that's the normative side of things. Now, the positive side of things is that how is the market actually, let's look at the market for a particular good, maybe real estate, maybe oil, maybe sugar? Right? The question is, is it actually perfectly competitive? Right, so this latter question is what's called the positive questions of social sciences. The first the former is called the normative questions of social sciences. Right? So most social sciences subjects have these two types of questions, what it ought to be, and what it is actually. Now, if we look at the positive, the second question, the answering the what is question? So there are essentially two ways of answering this. One is question. One is based on what's called theory. And so for example, if you want to continue with the example that I was talking about before, in terms of the justice system, right, so the theory would be that, how to actually courts function in practice, right. And there are theories of what drives prosecutors, what drives judges, what drives investigating agencies. So these are different theories. Based on that this is the objective function of prosecutors, this is what they would like to do. This is the objective of judges, this is

what they would like to do, and so on. But the second way of answering these what is questions is imperative that is using data. And this will. So if we want to answer the question does actually, to be two people who have committed the same sort of crime get equal the same punishment? So we'll try and use data to answer that question. So that's the empirical way of answering this question. The what is question and that's where statistics comes in. And this sort of what is questions occur all the time, for example, in economics, like, one of the questions is that do everyone does everyone have an equal access to the job market? Does everyone have an equal access to education. So again, these are questions which could be answered by theory, but which could also be answered by data by looking at data and seeing whether where- the different people all have equal access to opportunity or jobs. So this is where statistics comes in. And that's why, because these what is questions are so important in the social sciences. That's why statistics is so important. And why again, other what is questions important, because we are often looking for trying to figure out where the problem lies, right? Before you try and implement a policy to fix a problem. You want to know if there is a problem at all? And question. Secondly, where is the problem? Again, going back to my justice example, is there a problem that not everyone has equal access to the justice system? Is there some degree of impartiality or partiality somewhere? Is there a discrimination? And what what so where is that discrimination coming from? Is it at the court? Is it at the prosecutors? Is it at the investigation? Is it at the police law? Right, so once we get a hang of that, we can get a better idea of what type of policies may be used to correct those imperfections. So again, to identify those we need data. And that's why we need statistics. So, again, so why statistics? One is that it tells us about the world or tries to answer the what is questions? What else does it do? It also enables us to test theories, because people may have different theories about why the justice system is flawed, or why not everyone has an equal access to education or the job market. Right? Now, you can test those theories using data. You can also test the impact of policies and actions. So for example, suppose you do increase access to education, you say, oh, a good policy could be to give a widespread scholarships, right? Is it that then what you check for is does scholarships really improve access to education? Similarly, how can we reduce health care? Wait times, which is a big problem plaguing a lot of health care facilities in Ontario. And suppose someone says that we should do everything online? Right. Now, of course, that's a policy. And you can test it using data, whether implementing an online system does reduce wait time or not. Another use of statistics is of course to make predictions. So in the current pandemic, people use models and data to try and predict where what's going to be the part of the pandemic going forward. In other contexts firms always try and make predictions about their business in the future. What are the products which are likely to increase in demand are likely to decrease in demand in the future. And once you can make better predictions, you can gain a competitive advantage over your over your competitors. And in fact, this is the age of big data. And you see, companies like Google, Facebook, they're making millions and billions of dollars, just based on the data that they have trying to predict you looking at the web pages that a consumer visit from that try and deduce which products are more likely to appeal to particular consumers. You don't even have to think that big. You can even think in your own life. So say for example, your hockey team or your soccer team has a match against another opponent. You may want to look at previous videos of this opponent and try and deduce how do they play which are the main players who are the main defendants. So that's not rigorous statistics perhaps, right but you are trying to use data And the more professional leagues these days, all teams, they have data analysts who are trying to look at opponents trying to break down what's the strength and weaknesses of opponents and trying to give their own team a leg up for a stakeout or whatever your game may be. Okay, so statistics is that's why super important and even more so in this age of the big data. So in this module, I'll try and introduce you to what type of data where does data come from, how do we visually represent them? And then in the, in the other modules, we'll go a little deeper into statistics

