

Transcriber's Name Priya Prashanth

Applied Econometrics

Prof. Tutan Ahmed

Vinod Gupta School of Management

Indian Institute of Technology-Kharagpur

Lecture No. #04

Correlation and Causality (Contd.)

Hello and welcome back to the lecture on Applied Econometrics and we are talking about the difference in the concept of correlation and causality and this is the third part of the lecture. (refer time: 00:43)

Now we have seen when you talk about causality, it is pretty intricately related to the question why and we have also seen the why is coming from a hierarchy of whats, okay. So we will just say what 1 what 2 and so forth. So all these what questions are sort of giving us an answer to the why. Now for example, we have seen the example, initially we have seen the example of scurvy.

In the previous lecture, we talked about the, you know the owners decision whether to kill a chicken or not. And we also talked about the cigarette example. So let us say in the cigarette example, let us say, you know we actually end up proving that cigarette is causing cancer. You know everywhere we see, you know like cigarette is causing cancer so it is still a very strong conviction, it is a string strong correlation.

But we really do not know for a fact why cigarette is causing cancer. Now let us say we discovered that there is something called nicotine in cigarette that is causing cancer. Now what is nicotine and why it is causing cancer? So that also you know like, we do not know, okay? So perhaps we know that okay, there is something called let us say, polycyclic hydrocarbon that is causing cancer, okay.

So again, what in nicotine is causing cancer is polycyclic hydrocarbon. Now then, we can further ask, okay, what is there in the polycyclic hydrocarbon that is causing cancer? So you can further actually get into it and actually discover the molecular structure of polycyclic hydrocarbon and that actually, you can see, you know what sort of chemical reaction that happens you know that that causes when it enters inside your lungs right?

And then you see that some xyz factor is actually you know happening because of the polycyclic hydrocarbon and that is actually causing cancer. So that is the thing like as you sort of go deep into this, you know layers of what and what and what you actually end up reaching to the why and your why is becoming more and more strong, right? So that is basically the structure of causality.

That is basically how you see the how causality is different from correlation, okay. Now why it is so important? (refer time: 02:47)

Now the next question I would ask, why is it so important? Why is it so important to understand to understand causality? I understand, okay. So there is a series of what questions that is leading to the why question, but my point is, why do you even bother? Why do we care? Okay? And to answer this question, I will actually tell you another basically it is a it is a beautiful book called Sapiens by Yuval Noah Harari.

And I will sort of, you know bring some lessons from that book. And in that book author is actually asking, so it is like, you know the animal kingdom, let us go back to the animal kingdom, maybe you know some million years back. And if we see that humans like us, we are if we consider our physical structure we are rather frail if you compare us with the lions and tigers.

And that time if we had mammoths and you like this dinosaurs, so we are rather frail animal, right who could have like, you know lived on the, on the, on the trees and branches. And we were sort of scared of the animals because we are really frail. Now fast forward million years, we see that we are basically the masters of the animal kingdom, the humans and that frail animal, million years back has become the master of the animal kingdom.

Now why that has happened? What you know we had, what this human being had that made us the master of the animal kingdom and that is rather an interesting question, right? And this superiority you know when you when you try to uncover the superiority you will say that, okay, so the superiority is because of, we have learned to sort of, you know how to use fire or how to create fire, or let us say how to do agriculture, right?

Instead of gathering and hunting, we actually learn how to do agriculture. So you can give me give us you know give me many facts on let us say, humans actually build dams, okay. So so many other factors, you can actually uncover. We build the house, right? So all these things,

we did over a period of time, we learned how to do that. Now you can say that you know why humans are superior?

Why humans if I ask you why humans are superior to other animals? You know how we have become superior to other animals? And you can give me so to the to why answer this fire, agriculture, dams and house and so forth. So you can say we have done that. Now I will ask why or how humans have learned this whereas other animals could not learn this, right?

So other animals could have learned the same way like the fire and agriculture and dams and house and so many things. But why other animals could not learn it, whereas humans can learn it? Now to answer that Yuval is actually giving an answer that you know it is imagination. So that it is humans we could imagine certain things, okay.

So when you see that the, you know the branches of tree are, you know sort of rubbing against each other and the fire is being created the same way human think that okay, so if we probably rub the stones with each other, we can actually create fire. Or if it sees that the seeds are falling from the tree, and it is actually creating new trees, so perhaps if we plant the seed somewhere else, we can get, you know a lot of fruits or lot of vegetables from the other other part.

So that way human could actually imagine things, right? Human could actually imagine things and which is actually not the case with monkeys or chimpanzees, or you know like for the other animals for that matter. So when human you know we created dams, for example. So dams or pyramids let us say.

So we are told stories, like lot of laborers, you know they toil very hard, and because they are told that if you do this kind of work, you can go to the heaven and so forth. And those stories they believe, because they could imagine, they could believe you know that, okay, this is a possibility. Now this imagination is not there, for you know monkeys or chimpanzee or other animals.

You can tell a monkey to actually, you know you can tell the story of heaven to a monkey, and you can try to see if they are actually, you know bonding together to actually coming together to actually build your dam. That perhaps will not happen. So it happened for humans, because we could imagine. So this power of imagination that actually lead us to where we are the master of the animal kingdom, right?

Now why we talk about imaginations, when we talk about causality. So that is that is quite intricately related. (refer time: 07:30)

Because the moment we can imagine, the moment we can imagine, so imaginations actually giving a plausible, plausible causal structure. So it is not the causal structure, but a plausible causal structure, right? So it is basically telling that you know why one factor is leading to the other. So be it a story, be it our understanding of fire, be it the understanding of our you know how the agriculture can happen.

So all of these are actually giving human mind many possibilities that, okay, this is a possibility it can happen, right? And the imagination is basically nothing but dealing with all these plausible structures, plausible causal structures, right? So that is the beginning of it. So the moment we have this plausible causal structures, we then try to understand what for a fact will you know lead x from y, x to y.

So I know if I plant the seed somewhere else for a fact, that will you know I can see the plant somewhere else, right? It is not that the seed has to be planted under the tree. It can go somewhere else and still I can get the same tree. Or for a fact, I know that I do not have to rub the branches against each other, but I can actually rub the stones, and I can see the fire, right?

But you know like, and also for the same thing, like you know when it does not matter, you know like uh whether that causal link is correct or not, but the moment human has this plausible structure in their mind, you know we do not know whether we will go to heaven or not, but still we work, we toil hard to build the dam or to build the pyramid, right? So that is the thing. So the plausible structures are really, really important.

And from this plausible causal structures, we end up getting the causality. So out of many explanations, we actually end up getting the causal causality. So in the same vein, in the same vein, the way humans started imagining, in the same vein, when we fast forward and see how, let us say, you know the scientists discovered gravity, right?

The scientists discovered gravity, the apple has fallen on the ground, and they have they have understood, you know why the apple has not, you know it did not fly to the sky, right? So there are many plausible structure and they had come to the fact that this is a fact that it is gravity, which is actually explaining why the apple has fallen down.

Or for example, if you read the story of discovery of penicillin, this Alexander Fleming's this the fascinating story, you know he just found out that you know some, you know some bacteria, some some uh some sort of mushrooms, you know that that grew on on his bread is actually good.

And because there was no harmful, you know other bacterias or germs, you know are there on his bread, you know after, after his wife kept the bread few days for, you know for his lunch and he forgot to eat that. So essentially, the point is that, you know the, we see that, you know certain things are causing certain things, or certain things and certain things are related, which is correlation, and then we try to find out why.

And the moment we actually uncover the why we become very powerful. Because, you see because we knew the concept called gravity, we we are, you know we are doing so many things, including the exploration of space, or if since we knew the the thing about penicillin, or, or the, you know the the discovery of Alexander Fleming, we could actually, you know advance so much into medical science.

So the causality, the moment we know A for a fact is causing B, for a fact, there is no doubt about it, A for a fact is causing B, then we become really powerful. That is the power, the causality gives you power. You can actually, you know do a lot into that. You you sort of advance the science and technologies and medicines and you know everywhere. So this is the power, this is the power of causality, which has advanced human being where we are today.

So it is the power of imagination, which are giving us plausible causal structure. And from there, our capacity to discover what for a fact is causing A to B. And that is, that is why causality is so important. Now the last part of the lecture, the last part of the lecture is the role of causality in econometrics. (refer time: 11:39)

Causality in let us say statistics and econometrics. Now causality, uh so in the early days, the statisticians were not very clear how causality and correlation are different. Initially people used to think that perhaps correlation and causality are same. And in fact, uh you know correlations are easy to understand. You can just see two things are covarying, x and y are covarying, and you say that they are correlated.

And that was kind of a celebrated concept at that time and that is why we had Pearson's correlation coefficient, but we will never hear something like Pearson's causal coefficient. You know people used to talk about correlation and correlation used to be celebrated. As I

said, there is Pearson, then his student, his name was Galton, who is I think cousin brother of Charles Darwin.

So they all have like you know they they used to talk talk in the language of correlation. Now it was realized that, of course correlation is an indication, indication of causality, right? But correlation never proves causality, right? So uh people in this discipline or even in other disciplines started becoming a little sort of, you know aware of or you know they are rather cautious when they use the term causality.

They would rather use the term correlation, because you know there is no dispute. But when you use the term causality, it is rather a challenging thing to for you to prove that it is causally related. And that is why the term causality becomes somewhat sacrosanct, okay. So only later on uh, we started talking about causality in in statistics and in econometrics.

And particularly in the context of econometrics, we will see something called structural equation modeling, where we sort of, sort of try to see a sequence of events, something like that, when A causes B, B causes C and C causes D. So like, it is basically understanding, you know the multiple, you know you will see the correlations, but how if the, you know all the correlations are holding true all the time.

So that is much more robust approach than just to see if A and B only are correlated. So if I see there are three correlations, which are happening time and again, so I have more conviction about if one thing is leading to the other thing. Also uh in the same vein in in computer science, you know like people started talking about the graph network theory.

So essentially, the fundamentals remain same to understand how this one factor is leading to other and then when we see that there are multiple factors leading, you know are related to each other then we have more conviction about saying the this causal path, this path, okay. I will end this lecture with one more concept that is uh, you know pretty modern and like, it is a concept that has been used for for many, many years.

But in economics, in the domain of economics or in econometrics, that is, you know that is rather new. And that is randomized control trial. We will talk about these, in the third module, the randomized control trial. And essentially what it does is it uh tries to control like, all so like it has a treatment group. And it has a control group. And it tries to sort of mimic uh, sort of an experimental setting.

I am talking about in the domain of economics. So randomized control trial, as I said, it is not an old concept in in medicine. You know it has been in use for quite some time. You know it is like at least 100 years old concept. Or people like Fisher spoke about this. Now what happens is be it in the experimental setup in in, in the domain of medicine or be it in the domain of economics uh, where again we try to create a experimental setup in the natural environment.

And what we do is we create a treatment group, a control group, and what I mean by a treatment group is we create some sort of intervention designed for the treatment group. And we try to, you know sort of create an environment where you can sort of say that all the other, you know variables are controlled, like they are not changing. So only the treatment variable is the only thing that is actually influencing the treatment group.

Now then we say if certain results are happening, that is because of the intervention, because everything else is controlled, okay. So we will talk in more detail about randomized control trial and how it explains causality. So to end this lecture, so what we understood is that the correlation and causality both are important concept. Correlations are important because correlations are actually indicative of causality.

It gives a plausible structure, plausible causal path. But then you have to find out for a fact, what is actually making happen, you know what is actually you know making this A and B linkages to be together, the causal path, basically. And causality as we see that it is really a very powerful concept.

And this is the concept that is that has advanced different streams of science and medicines, and you know every aspect where human has made progress. So human progress is very intricately related with our discovery of causality of different domains. So we will talk a lot about causality in the third module. So with this, we sort of end this lecture. So this is sort of an idea to give you uh what we are going to deal with in the next lectures. So we will talk about correlation first.