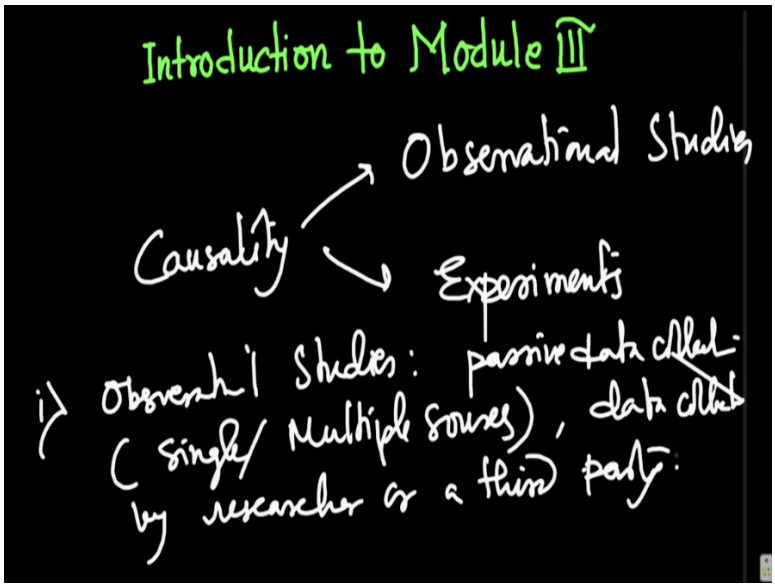
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Lecture – 93
Introduction to Module III

Hello and welcome back to the lecture on applied econometrics. Now we are in module 3. This is a much coveted module, where you want to talk about causality. And in the previous two modules, we actually talked about correlational studies. We have built a regression model, we have done different diagnostics, but we have remained in the domain of correlation. At the beginning of our lecture if you remember, we very clearly outlined the difference between correlation and causality.

And we said that causality is an elusive concept and we usually do not talk about causality as much, but you all want to actually achieve causality and causality is something that is not just in economics, but any other discipline; physics, science, any other scientific discipline we all are actually looking for causality. So this is what we are going to talk about in this module. So it is going to be of 2 weeks, this week and the coming week, and we will talk about causality mostly in two types of research.

(Refer Slide Time: 01:17)



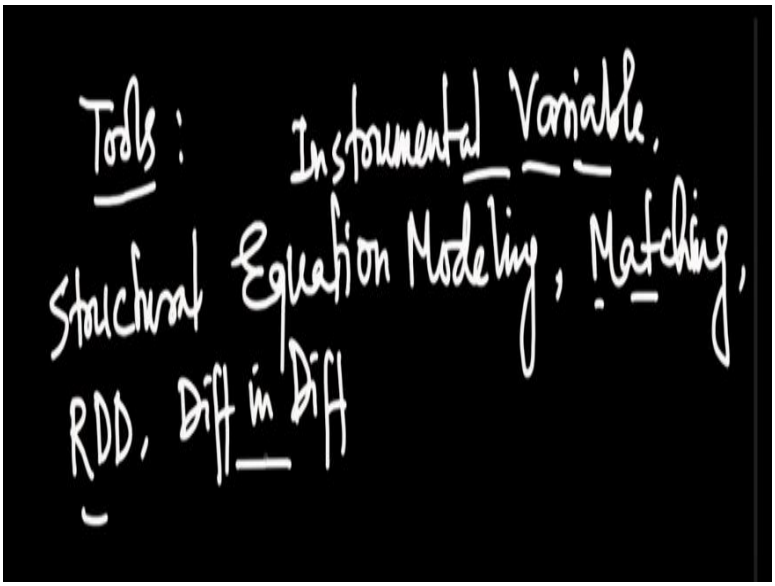
So we will search causality in observational studies and in experiments. So, how these two concepts are different? So in observational studies what happen is that the data is collected by maybe a third party or you can view the researcher himself or herself and collect the data, but

essentially you do not do anything actively. You basically passively collect data, you just go to the field, you ask some questions, you collect data.

And you can collect data for different time periods also. You can collect data using multiple sources also. But essentially, the whole mode of data collection remains very passive. So let us say this observational studies is a passive data collection where data is collected from single or multiple sources or data may be collected by the researcher or by a third party. But the whole point remains is that it is passive data collection.

I do not design any experiment here. And for example we can even think about the data collected by a National Sample Survey in India and then we use the data and we use our research question, use the data to ensure that their compatibility like the data we use is actually addressing the research question we have. Now, what are the tools and techniques that we are going to talk about when we talk about causality?

(Refer Slide Time: 03:15)



So the tools that we're going to talk about, we will be talking about things like instrumental variable which is actually a very interesting concept, but it is very difficult to find an instrumental variable. So essentially, you will see some; it is kind of a natural experiment, we try to mimic or experiment that we can perhaps do it is very difficult to do natural experiments. So if we get a good instrument, we can actually mimic that experiment.

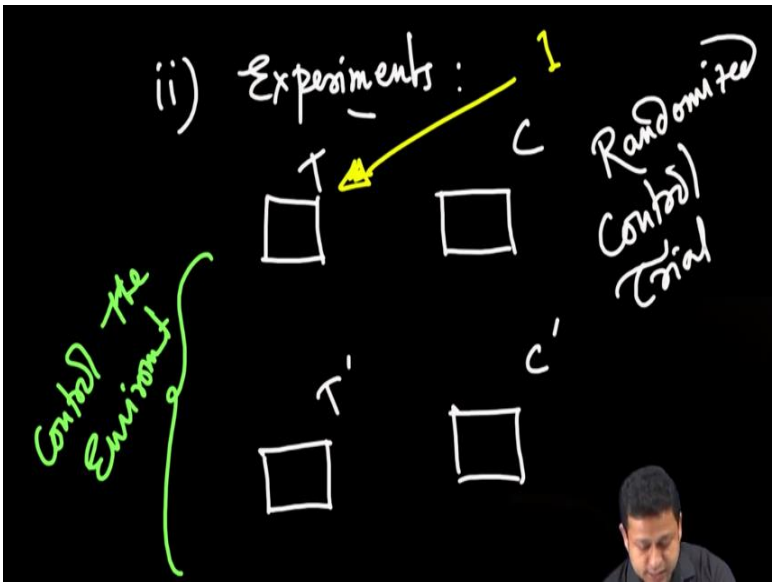
So instrumental variable is very important. We will also talk about structural equation modeling where we see a series of events, when X leads to Y, Y leads to Z and something

and we try to find out how strongly these different events are actually related. And then we will actually talk about causality, a lot of what is X leading to Y. Then we will talk about things like matching.

Matching is another important concept we try to match those observations which are very close, which looks similar, which are identical, compare these two groups and we will ensure that all conditions applied on two groups are almost equal and we will see the impact of some intervention at different time points and that is how we use matching. So there are a lot of techniques involved to ensure that the two groups that were are studying are almost similar.

We can also use regression discontinuity designing or Diff in Diff. So these are basically different tools and techniques that we will be using.

(Refer Slide Time: 05:08)



The other concept that I talked about is experiment. Experiment is very interesting in economics. So we know clinical trial, right. This is a very old concept. In clinical trial what happens is that you take subjects like two groups and perhaps you just have a new medicine you have discovered and what you do is both the groups are subjected to that same medicine. And at the beginning of the study, you ensure the two groups that you have selected are almost identical or if possible they are just identical.

Now, once you have these two groups, what you do is you put them in the same condition like you do not vary any other condition except the treatment that you are doing that is except the drug or anything that you are actually giving to them. So one group gets that drug and

another group does not get the drug. And then you sort of allow certain time period to elapse to see if that drug is actually; what is the impact of the drug on the group of people who is subjected to that treatment.

And that is how you try to understand the real the true impact of some intervention. Now, you need to take care of many other things. Factors that may actually influence your experiment. For example the control group actually knows that treatment is happening, so there might be some sort of they might actually adopt some practice and suppose so many other things. So those external conditions you need to or the conditions coming out of the experiment you need to control.

So this is something that is really very powerful concept and that actually we can convincingly argue about causality in experiments. So let us say this is a treatment group, this is my control group, and I do an intervention to the treatment group. And then I sort of control the environment, I am using different colors so that it looks good, and then I allow some time period to elapse.

And then I see the treatment group at some T 1 time control group at T 1 time and I basically look at the difference. So the the previous concepts also that I talked about like matching or regression, discontinuity design, diff in diff also we use in in this kind of controlled trial experiment. So, this is basically called randomized control trial, which is a very powerful concept. So, we will talk about these concepts in this third module.

(Refer Slide Time: 07:49)

So along with this, we also talk about some topics that we wanted to cover, maybe we will be brief on this concept. So one is that we wanted to talk about logistic regression. We have not yet familiarized you with idea of logistic regression. And second, we will also talk about the concept of parametric and nonparametric. So far, we have been talking all the different tools, techniques; these are all parametric because they are coming from some known distribution.

And nonparametric is basically where you do not know the distribution. So we will talk about the differences a little bit. Finally, in this module we are also going to talk about a couple of things which I promised that I will cover, but I really did not cover it yet. One is that minimum detectable effect. So this is something very important when we actually do some experiment design and this is actually going to be relevant for our third module.

And finally, we will also talk about something that is really intriguing and that is the difference in the field, econometrics and machine learning. Machine learning is an emerging field and people are often trying to see what machine learning can do. So it is important that we should also have some idea between how the boundaries of econometrics and machine learning is actually if there are intersections, how much you can do with each other, so we try to see the possibilities.

So this is going to be broadly the ambit of model 3. And I am really excited to actually take you in the journey in module 3. There are a lot of exciting concepts and topics we are going to talk about and I cannot wait to start module 3. Let us begin.