Transcriber's Name Maria Priya

Applied Econometrics

Prof. Tutan Ahmed

Vinod Gupta School of Management

Indian Institute of Technology, Kharagpur

Module No. # 02

Lecture No. # 14

Double Structure of Variable

(refer time: 00:26) Hello and welcome back to the lecture on Applied Econometrics. So, in the previous lecture, we spoke about the difference between probability distribution and frequency distribution. In this lecture, we are going to see that in details. And we will, we are going to see something very important concept that I sort of spoke few times, is the double structure of variable. We are going to come to that concept. But before we talk about all these things, let us actually do an exercise, okay, and let us try to sort of, you know, forget about all these theories; let us just try to see from the number, you know, and try to interpret things. Okay.

So, let us say we are actually trying to do a problem uh one of the very interesting problem in economics, uh and that is, let us say, the participation of females in the labour force. And you will know that in India, this is exceptionally, in this exceptionally different beside other countries, because here the participation of female in the labour force is really, really low. And you can read about it. uh It is like, as good as the number of entire population of Brazil; I mean, if the female (popula), female population participated in the labour force, then another country like Brazil would have been in the workforce in India itself.

So, they are like a huge uh number of people and not in the labour force, not in the workforce. Now, let us say you try to understand sort of, you try to get some insight about why the participation of labour force is so low in India. Okay. And how you do that? You, as an economist, you try to see, you try to have some idea about like, uh some variables which might be influencing it, and you try to get data of course. uh From data, you try to sort of interpret results. Okay.

So, let us say, let us say I deploy a survey team; uh we we really do not have data; and that survey team has collected some data from district let us say, a sample survey. They have

collected data, okay. And let us say uh they are collecting data on these 2 variables. One variable is that; I as an economist, I told them that you need to collect data on their family income. And I want to see how female workforce participation is actually changing with family income. Okay. And that is a a pretty sort of interesting question.

And let us say I get these uh results. I will actually, you know, create a fictitious sort of, you know, number and tip, you know, number. And let us say I have this family income. And let us say, I have 100; I have, let us say 500; I have, let us say 1000; then I have 2000. And this is (mon) uh this is weekly income, let us say. So, 100 rupees a week means very, very low for a family. uh Some families might earn 2000 rupees. Let us say some families are earning 20,000 rupees a week; some families are earning 50,000 rupees a week; some families are earning 75,000 rupees a week; some are earning, let us say 1,00,000 rupees a week; some might be earning 5,00,000 rupees a week, they are pretty rich; some families might might be earning 10,00,000 rupees a week.

I hope my number of zeros are correct. Now, this is the family incomes, okay. And you sort of get the sort of, you know, like uh a number of females. So, you sort of create some uh ratio in that uh family, the adult females who can participate in the workforce, and how many of them are actually participating in the workforce. And let us say, you got this number, the actual numbers, okay. You got the actual numbers. Let me write down here. Actual participation:

You will get something like, say here it is 0.5; here, let us say it is 0.38; let us say it is 0.34; here it is 0.3; oops, sorry; point say 3 1 or 0.3; here, let us say it is 0.28 or 0.25; here, you 0.25, 0.25; here you have, say 0.21; here you see, suddenly it has spiked up, let us say 0.75; here you see something like 0.79; here you see, let us say 0.82; here you see 0.83. So, let us say these are the numbers I got, okay. So, this, I just got these numbers from my mind, okay. Now, these numbers are not without any basis, okay.

So, you can actually see this kind of number. And I will explain why these numbers would make sense. So, you see, when the family income is really low, like 100 rupees a week, you actually see, the ax actual participation is 0.5. Then, as the income is increasing, you see the participation is decreasing; and it is happening till your earning is, let us say 50,000 rupees a month, a week. And then, suddenly it is increasing when your earning is becoming 75,000 rupees a week, okay. And then, and then it keeps on increasing.

Now, there is a, there is an explanation behind it. So, you know, it it is sort of, uh you know, if you see, it is sort of an U-curve, okay. So, if I actually draw this, (refer time: 05:53) if I actually draw this, it will look like, let us say, I have workforce participation, workforce

participation rate, and here I have family income. And if I plot these two, what I see is something like this. So, the workforce participation rate is, let us say, like this. So, here it is 0.5, and here it is a 0.8.

And there is a theory behind it, I mean, why why we may see something like this in India, and that is because, when we have the the income of a family is very low, the females need to participate in the workforce, because they are in the need to sort of add to their family income, because the family income is very low. So, if I have weekly 100 rupees income, so, perhaps it is a good idea for the females to actually work, because, you know, that will actually add up to the total income.

Now, it keeps on decreasing because, as an when your family income is increasing, so your family members, you know, think perhaps that females, you know, the female folks in the household, they do not need to do outside work, they can just stay back in the home, they can do their work, you know, whatever household work, but they do not really need to go out and do the, you know, do the income. Now, when it keeps on increasing, uh it happens that; usually, uh you know, the families whi which are uh economically well off, they are also educationally sa somehow, you know, they are they are probably living in the city, they have better education and they probably do not have much of social taboo uh towards letting female to work, right?

And here, this this phenomena actually, it is our research, it is from a research paper. uh And, you know, when females actually; be the the phenomena is sort of U-curve we see because, uh because the social taboo, when my family income is increasing, so I usually try try the female folk in the house; not, you know, I do not really want them to work because it it, you know, to a family, it might appear that if the females are staying back in home, so that is actually good, okay.

So, because females going out and working does not give a good signal to for the household, because uh probably men folks are not working, you know, here, the family is really not uh, family is deprived or something. So, that is the reason usually the females are not; uh you know, it is not a very common practice for females to work. And that is why, as when your income is increasing, you see this kind of trend where your female participation in the workforce is decreasing.

But here, your uh family income ah, with the increase in family income and possibly with increase in, you know, with a better education in this or probably because of the fact that you are living in the city, what is happening is, you are more liberal to allow females to work, okay, or females are also educated enough to actually to or to make their own decision to

work, right. And that is the reason uh we can see this kind of picture, and we can actually see these kind of numbers. So, that is the economics part.

So, we are not really dealing a lot uh uh of economics here, we are rather more interested to look at the data, and then go back to the definitions that we have been talking about. So, let us go back to the table. Now, I have seen that; I have seen the actual participation. Now, as an economist, your job is to model a things, okay, from observation, you want to model, okay. Now, how do I model that? Now, I am going to, I am going to actually write a formula.

And that formula is, let me write down, let us say; uh this pro, I am basically trying to create a relationship between family income and actual female labour force participation. And I am just writing it down first. So, this is 1 by log this 10. I provide this number here. And that is the, that is the probability, that is the probability of female workforce participation. (refer time: 09:55) And I will explain that. So, let us see the numbers as I go down up to 50 let us say. This is the function it is sort of following, okay.

Now, here I change the functional form. So, you see, the numbers are pretty consistent here, the numbers are pretty consistent here, almost same. Now, the second function I define, 1 by log 10, the same 10 base, with this number. (refer time: 10:37) And I see, the numbers are 0.79, 0.8, 0.82; oops; and here I see 0.83. Okay. So, you can see both the tables are actually, you know, sort of both the num both the columns are actually representing the same number. So, uh of course, I actually did it before, and that is why I know what the numbers are going to be.

And that is why I have sort of, you know, did a back calculation when I put the actual participation; no magic here. So, the thing is, uh on the third column, where I see these numbers, where I where I actually plotted these uh, you know, this uh mathematical form; and let me actually write it down, how I have written down the math math mathematic form. So, basically, uh so, let us say probability probability of workforce participation rate is 1 by log, say X base 10, when X is less than equal to 50K. Okay.

And it is equal to 1 minus 1 by log base 10 X when X is above 50K. Okay. So, this is the functional form I defined to explain the workforce participation rate. Now, what have I done here? So, what have I done here is that, I actually got the actual numbers from; so, I actually got the; so, these are my frequencies. I mean, from the frequency, I, what I saw actually, from where I calculated this percentages, I actually predicted, I actually predicted this workforce participation rate with, I sort of created a functional form. Okay.

This functional form, we are we are just going to see, we are going to talk about this functional form a little later. So, with this functional form, I sort of converted that into a probability. Now, going forward, when I am, I am not familiar about certain outcome, certain results, so, what I do is, I sometimes use those uh functional forms to sort of talk about what could be the, uh you know, workforce participation rate, given a certain income, right. So, I do not; so, I can sample only a small part, I cannot sample a lot.

But then, uh when I cannot, you know, like uh uh sample a lot, so, what I do for the other samples, I can only estimate, I can only predict. And that prediction part comes from this formula. So, I have created a formula. So, I am going from known to unknown. So, this is a journey from known to unknown, okay. So, how I would write; (refer time: 13:17) so, let me actually write down these terms, these are important. So, it is basically from known to unknown, okay, from let us say frequency to some probability with some functional form, right, with functional form. All right.

Now, look at look at this. Again, let us go back to the, to the Excel sheet. Now, here, if you look at the numbers again; so, you had, when we had this actual participation; so, that is something you knew. These numbers you get, these numbers you get, in this part, this is something you already, you you knew about it, right? But this part, when you are talking about this, you, this is something you are predicting, right. So, this part is already realised. In your sample, you actually see that these numbers are there, but when I am creating this log function or 1 minus 1 by log function; this is 1 by log and bo, these 2 functions; so, when I create these 2 functions, I am actually predicting; I, no longer I am actually getting this data from uh this uh actual samples, right.

uh So, this is, this part is called realised. This part is called realised sample. So, this sample have already happened, so, it is realised. And this part is called unrealised. Okay. So, I will, I will write this. I will write this. So, basically, realised, and this is unrealised. Okay. This part you observe, right? This is observational. And this is theoretical, because; I will explain. So, in the (prev) in when when we talk about a frequency distribution, so, we actually observe that we got the numbers, and that is why, from there, we sort of uh constituted the frequency distribution.

But when you talk about the probability distribution, we are actually using a formula, right. We are using a formula to actually get the probabilities. Also in the, just uh in case you are having any question here, you might actually think about this one. (refer time: 15:41) here also you have used a formula. We used a classical probability formula, how to (constit) how to uh calculate the probability, you know. So, basically, you divided by the, you know, all possible outcomes.

So, here also uh you are using this formula to constitute the probability, probability distribution. And that is where uh basically these two, frequency distribution and probability distribution are differing. Okay. So, you are sort of using a formula to create a uh basically a new distribution, okay, or basically a theoretical distribution; and that is what you are calling the probability distribution. It is a factual thing, so, you actually see the facts; that is factual. Whereas this one, what you saw, it is more like, you know, so, uncertain, right, or here, the chances are involved or chances are involved.

So, this is what you have been talking about. you There has to be some chances, you know, which are needed to be there. And also, here, when I use this formula, this formula to create a new distribution, so, this is becoming my potential distribution. Okay. This is the potential distribution; this is a potential distribution, potential distribution. Right. So, this phenomena, this is sort of, this is something, you know, like this frequency to probability is known e as a double structure of variables.

So, one way you have the actual numbers, and the other way, you do not have actual numbers, but you have a potential sort of distribution. Okay. And quite obviously, quite obviously, we can say that, when I am talking about uhm the potential distribution, I am using a random variable, right, (refer time: 17:24) I am using a random variable. But in case I am talking about this uh frequency; so, when I have already realised the observations, I have already got the numbers, so, they are no longer random variable, they are no longer a random variable, okay.

So, this is something we should sort of keep in mind when we talk about frequency distribution, probability distribution, and from there, how we are actually deriving the concept of double structure of variable. Now, we can also think it uh this way that uh you have a formula, right, you have a; so, so, here I have, I have explained how from a set of (freq) how from a set of numbers you are coming down to a formula, right. Now, it can so happen that you have a, you have a formula already; now, perhaps, this time you have this formula that you have just derived and you take this formula to sort of uh predict the the observations uh about an unknown sample or unknown population, right, which you which you do not know, from where you have not drawn any sample.

Now, in this cases also, this is a double; so, basically, again, this is a double structure of variable, because I have a known thing; uh I have an unknown thing first, and then, from there, you are going to predict something known, okay. So, that is also called a double structure of variable. So, usually how we represent? So, a random variable X is usually

represented as, you know, capital X, and all the observations that you may, the distributions that you can constitute, you can, you know, represent these numbers like this and so forth.

Whereas, for actual x, when you actually, you know, get the values of the variable; so, usually, that is represented by small x, and all the observations are represented as small x's, okay. So, this is a no notational thing which you are going to follow. uh So, that is something we need to remember in case of double structure of variable. So, if I am writing a capital X, so, that is a random variable; whether, whereas, I am writing a small x, so, that means actual values of the variable; so, that is already realised. Okay.

So, uh okay, so, this is it. uhm So, this is basically the idea of double structure of variable. And we are going to see, uh you know, time and again, we are going to use the concept of double structure of variable in uh in, you know, in various cases. We are going to see how the formula, the, you know, which from, using which we are actually converting the frequencies into probabilities and so forth. uh We will we will learn these, you know, these these, uh these functions, what we call this; you know, you know, ha how do we really explain these functions; we will learn about estimator. So, these are some of the things we are going to cover in the next few lectures. So, with this, we end the uh discussion about probability distribution and frequency distribution and double structure of variable. Thank you.