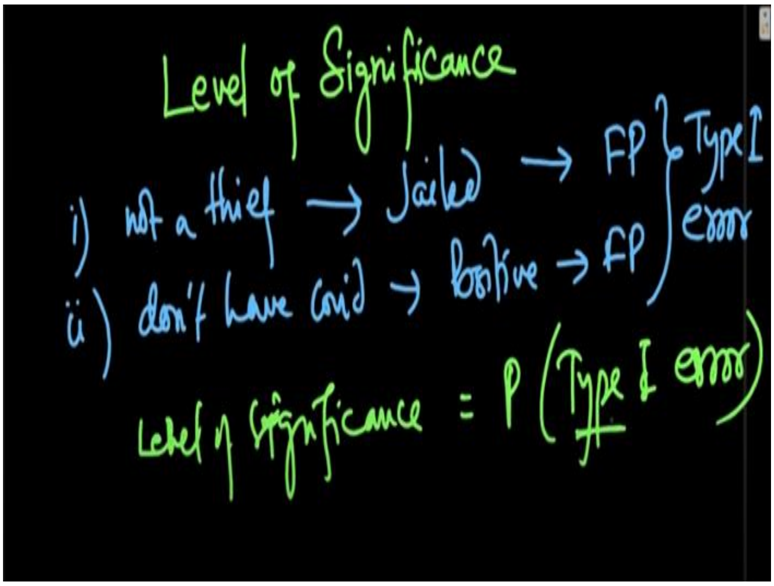
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology-Kharagpur

Lecture - 34
Level of Significance

Hello and welcome back to the lecture on Applied Econometrics. So we are talking about hypothesis testing. This is the third lecture on hypothesis testing. So in the first two lectures, we explained what is hypothesis testing. We also spoke in detail about the types of errors that we have to encounter. We also introduced some of the concepts that we need to learn to make a decision for inferencing.

We have to make a decision based on the data. So we need to use some concepts to actually do the decision making. And one such concept is level of significance. Now before I get into that, let me again recap the concept of type I error.

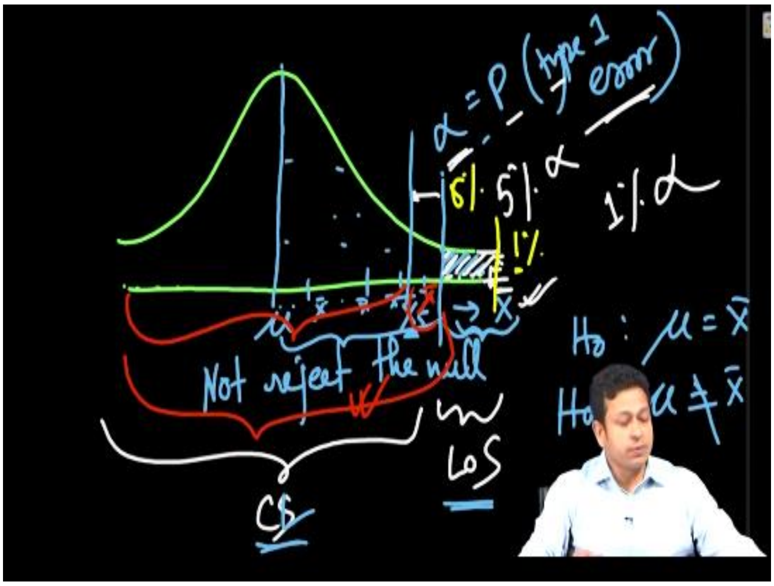
(Refer Slide Time: 01:07)



So the type I error means, let us say we gave many examples. So one example was let us say, I am not a criminal, not a thief, but jailed, okay. Or I do not have COVID, do not have COVID, but diagnosed as positive right, and then I am treated. So neither of these situations are desirable, right? Now both these cases are essentially false positive. We know type I error is essentially the false positive cases, false positive or you can also write type I error right, type I error.

Now we claim that the level of significance is nothing but level of significance is a probability of type I error and I will explain type I error, alright? So essentially it is telling us how much tolerance you have for type I error. The level of significance essentially telling us how much tolerance we have for type I error.

(Refer Slide Time: 02:26)



So if I draw distribution, let us say I have let us say some one-tailed test and I have the level of significance alpha and that is basically there is nothing but the probability is the probability of type I error, type I error. So it means the error part, so basically I will be I will not reject the null, I will not reject the null. So this is one of the very important criteria of decision making, not reject the null if my x bar is in this side of the alpha.

Whereas, I will reject the null if my x bar or the sample statistic is actually on the right side of the alpha, the level of significance. Now why I do that? There is an intuition behind it. And the intuition here is that well, so these are the errors, these are the errors. Now how much error I am allowing, okay? So essentially try to feel this problem. So essentially I have my mu here, okay.

So this is my null hypothesis that my null hypothesis is that my mu and x bar they are basically the same, okay? Whereas my alternative hypothesis is saying mu is not equal to not equal to x bar, okay. Now I claim that as long as my x bar is falling anywhere here, it could be here, here, here, here anywhere I am so I am allowing so much of space not to change the status quo.

So as long as x bar is in any in this entire zone, in this entire zone, I am saying that I still consider the null hypothesis may not be rejected. I still consider null hypothesis may not be rejected because the evidence that I have gathered that is still in this zone, okay. I only reject the null hypothesis if my evidence is so strong that it falls here, if it is here, if it is in this zone, if it is right to the level of significance, okay.

Only then only in that particular scenario where my evidence is really very strong that my x bar, my statistic is really far away from my true population parameter, only then I say that we may reject the null hypothesis or reject the null hypothesis. Till that point my x bar, I do not reject the null hypothesis, okay. So I have to, I really have to have a very strong evidence that my sample statistic is away, is really away from the population parameter.

Now if I look at from the other perspective, let us say I have my x bar here, anywhere here, I need to write again. So my x bar here let me just use a different color, my x bar is here in this point. Now it can also happen, if you remember when we said that in the previous lecture that you know like we can have extreme values, right? We can actually see some extreme values of x bar.

We can see that they are really far away from the population parameter, but still it might be just because of an error. It might you know, actually the sample might be actually drawn from the true population, right? Now considering that I will say that if the x bar is here, and if it is wrongly done, so that is basically the type I error. So essentially that would mean that this is the area that represents the type I error.

I repeat, it can also happen that the x bar are just the extreme observations whereas the sample is actually representing the population. Now when I have extreme observation, the x bar is going to be here, but still x bar is representing mu, okay. So this part that all these points where x bar can actually you know lie and yet x bar becomes greater than alpha or yeah the value of x bar is right side of the alpha.

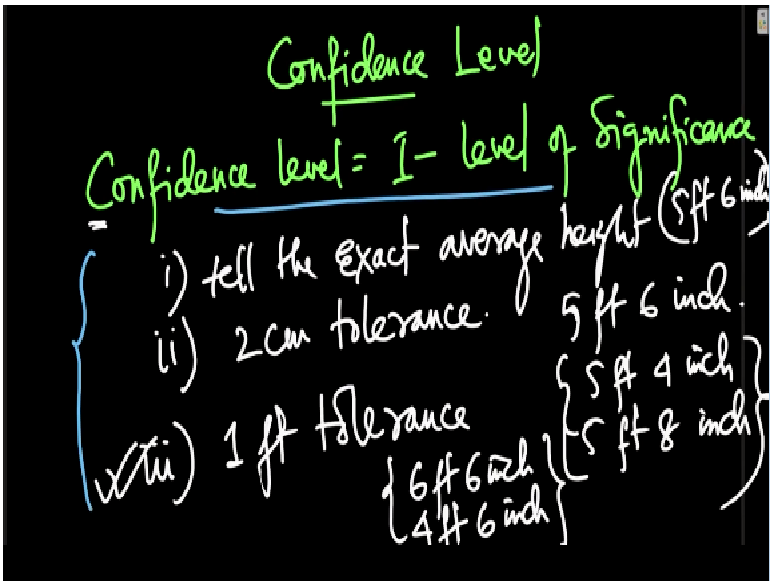
So then we can actually say that it is wrongly done. And when it is wrongly done, so that is nothing but the type I error, okay. So that is why we call the level of significance is the probability of type I error. So this is basically the concept of level of significance. And this is how we connect the level of, the concept of level of significance with hypothesis testing, because we are trying to reduce the probability of type I error.

So depending on the value of alpha I take, I decide my tolerance to type I error. So I can have my alpha here. Let us say I have a 5, we usually have 5% level of significance, or I can also have 1% level of significance, 1% alpha. So if I have 1% alpha, so what I will do? I will actually push my bar even farther, okay. So this is my 1% alpha. This is my 5% alpha.

So that means, if I have 1% alpha, I am really not tolerant towards type I error. And depends, it depends, test to test or context to context where you want to keep a 5% alpha, where you want to keep a 1% alpha. Sometimes you really want to reduce type I error. Sometimes you actually want to reduce type II error, okay. So we will see that. Depending on the context, we decide this alpha value.

And as a thought exercise the previous three examples we have sort of illustrated, you can actually think in which cases you actually need to have very low alpha or in which cases you can actually allow some tolerances. So that is one thought exercise that you can do. So that is basically the concept of level of significance. Now coming forward, going forward, we will explain another concept and that is called confidence level, confidence level.

(Refer Slide Time: 09:02)



Now confidence level, mathematically speaking, is very simple. Confidence level is nothing but 1 minus level of significance. But we need to understand, we need to understand intuitively what is confidence level. So before I actually explain that mathematically or graphically, let me actually explain it intuitively with a different example.

Let us say the professor in your class is, you know he just wants to have fun in one day in a statistics class, okay. And he asks you that he actually is doing a gamble, okay. So it is like he has a reward amount of let us say rupees, maybe 1000 rupees okay and if you have to participate in that gamble you have to pay rupees let us say 100 rupees. Now the gamble is this.

So in this class, the professor asks you that you have to tell the average height of all the students inside the class, okay? Now there are certain categories of the gamble, okay. It is not just one category, so there are different options the professor is giving for you to play. So one is that, I will write it down, one is that you have to tell the exact average height. You have to tell the exact average height, okay?

In other case, he will say well, I am fine with let us say you know 2 cm tolerance, okay. So if the average height of the class is 5.5 feet 6 inch 5 feet 6 inch, so the professor is okay, he will still give you the reward, if you can predict, let us say 5 feet 4 inch to 5 feet, if you say any number between in this range 5 feet 4 inch to 5 feet 8 inch. So professor will reward you here if you have to say 5 feet 6 inch to get the reward, okay?

If you say 5 feet 5 inch, the professor will not reward you, okay? And in this case, there is some sort of tolerance. Let us say the professor is okay with let us say, 1 feet tolerance okay, 1 feet tolerance, which means that as long as you are saying anything like 6 feet to 6 feet 6 inch and 4 feet 6 inch, as long as you are saying any number between this range, the professor is going to give you award, okay?

Now looking at these three possible scenarios, it is very common sensical, anybody if I ask you which category of game you want to play okay and given that the entry fee remains the same and the reward money remains the same. So I am sure 100 you know 100% of you will say that I want to play the third game because you have more tolerance, you have more tolerance to do error right?

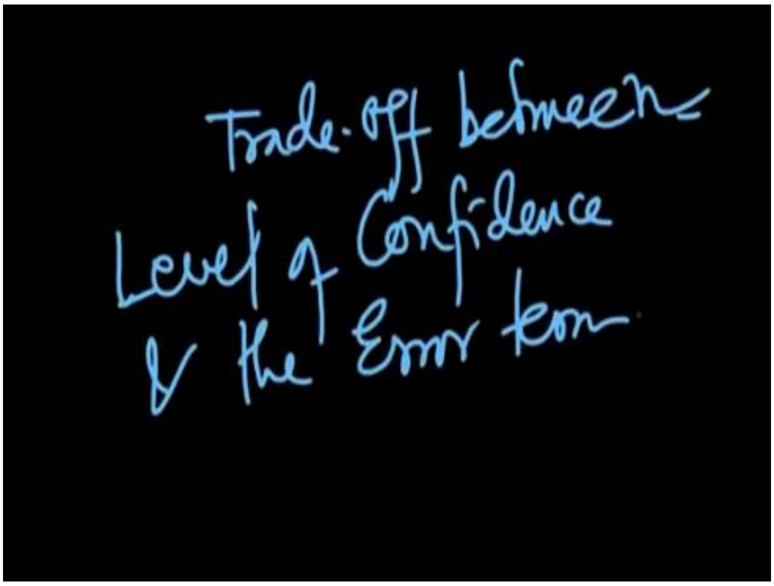
You can, you have that tolerance and yet you are getting the reward money. So essentially for the third game you are more confident to play, you are more confident to play the third game. And that is where the concept confidence level is coming, okay? So the more you widen, the more you widen your tolerance, the more confidence you have okay and that is what is the confidence level.

So if I go back to my previous diagram, here it is my level of significance and this part is my confidence level, this part is my confidence level, okay. So essentially, the more I have level of, if I push this line here, if I push this line here, so essentially, my confidence level is narrowing, my confidence level is narrowing.

So essentially, that is how I say this relationship comes from here, confidence level is 1 minus level of significance, because the area under the curve is always 1. So these two things, level of significance and confidence level, when I add them, they are giving me a number of 1, right? So that is why I have written down that confidence level is nothing but 1 minus level of significance.

All right. Now the example that I have given, now this is a very interesting example because it is, what do you understand intuitively is that the moment I allow a wider level of confidence or confidence level, I am actually also allowing more error to be committed because the moment my you know my range for which I say that the x bar the sample statistic is actually can represent the population parameter, if I widen it so that means I am also allowing more room for the error term, right? So this is a tradeoff we have to remember.

(Refer Slide Time: 14:26)



The tradeoff is tradeoff between level of confidence and the error term, okay. So if I go back to the previous diagram if I, let me see which one I can show, let us say this one. If I allow more sort of room more room here, if I basically have, it is a bit dirty, let me use a different color. If I allow let us say, if I allow this room here and then this room here.

So in this case, I am allowing more room, okay. This one is I am allowing more room and when I am allowing more room, so that means I you know the x bar can be here and even then I would say that I cannot reject the null hypothesis. But this x bar can actually come erroneously, right? It is, there is the chances are you know increasing as we sort of you know move towards the extreme part of the distribution.

So this is why we, this is something we need to remember that the confidence interval or confidence level and the error term there is a tradeoff, okay. So with this we sort of end this lecture. We have talked about level of significance. We have talked about confidence level. We will learn the other concepts in the next lectures. Thank you.