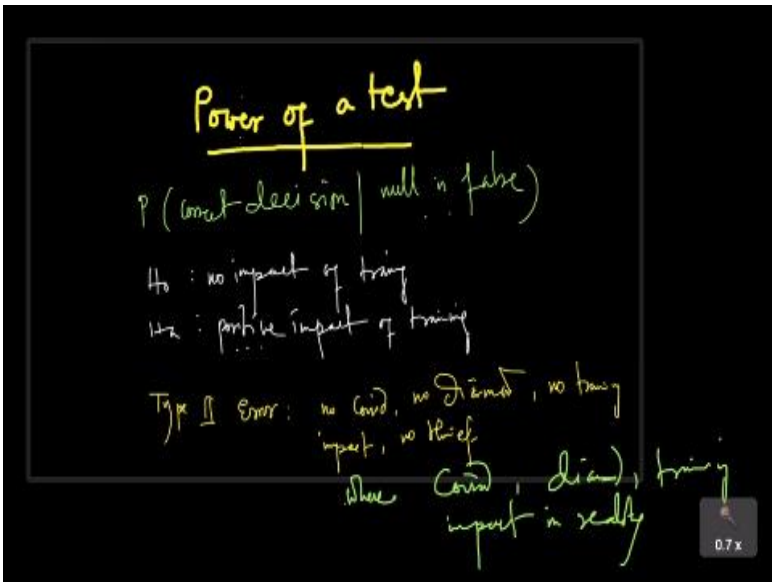
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology-Kharagpur

Lecture - 36
Power of a Test

Hello and welcome back to the lecture on Applied Econometrics.

(Refer Slide Time: 00:30)



In this lecture we will cover the concept of power of a test. Now power of a test is a concept that is very important when it comes to hypothesis testing and we need to decide whether to accept or reject the alternative hypothesis. Now power of a test could be sort of conceived as the probability of making a correct decision, probability of correct decision, of a correct decision when given your null is false, null is false.

So it means, let us say we actually impart training to the students of VGSoM of Python training and the null hypothesis would be that there is no status quo change. So H naught is no impact of trading impact of training and H1 is impact of training, there is some positive impact of training. Let us say it is a one-sided test, no impact of training. And H 1 is or H a is a, that is a positive impact of training.

Now if I think it in terms of the power of a test, I would say probability of correct decision, probability that I could identify the positive impact of training given that there is a positive impact of training, right? So essentially, we also say power is one

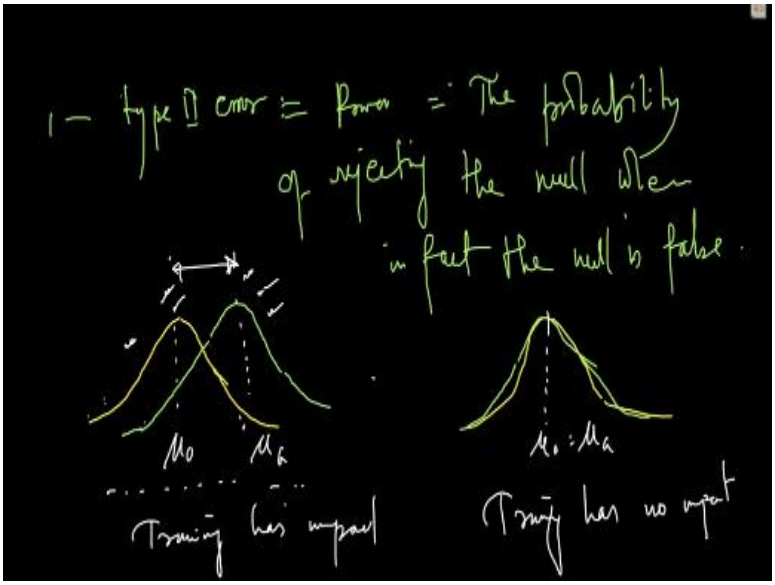
minus type II errors. If we go back to the definition of type II error, type II error would be, type II error is if you recall it correctly, type II error is that type II we basically deal with the false negatives, right?

So essentially, we wrongly identify something as negative when in reality that is positive, right? So for example, we wrongly identified that there is no impact of training. So we have wrongly identified it as negative. Whereas there is a true impact, right? So we have basically found that there is no status quo change whereas there is status quo change, right?

So we have given example of you know like for example, say we say no COVID, no COVID or say no diamond, no diamond where there is really COVID or say, we say that no training impact where there is really a training impact, right? Or say no theft or people are not thief where there is people are really thieves. Yeah, basically no thief.

So these are the cases where there, where people are actually, where there is COVID or diamond or there is training impact and so forth in reality let us say, in reality let us say. So these are the cases of type II error and I claim that type II error a power of a test is one minus type II error.

(Refer Slide Time: 04:08)



So if I trying to actually see that, so type II error is essentially. a type II error is essentially, we wrongly identify something as negative when in reality that is positive.

So power we can say type II error one minus type II error is power. So it is going to be the probability of rejecting the null when in fact the null is false, right? Null is false, okay.

So that is basically the idea of power, okay. So you are basically correctly identifying the cases when the null is false. So these are the two things. You are correctly identifying. And basically we are identifying those cases where the null is false. So essentially unlike the type I error when you are talking about type II error and we are basically talking about two different distributions, because we are talking about the alternative hypothesis here.

So null hypothesis, it has a distribution. And the moment I bring in another hypothesis, which is alternative hypothesis, it means that I am considering the population, the population, the null and alternative hypothesis population, they are different. So basically, the samples drawn from these two distributions, their characteristics are going to be different.

So basically, when I infer the population characteristics using those samples, we are going to get different population characteristics. So this one, the null hypothesis one, and the alternative hypothesis one, they have, they are actually in a sense, they are actually drawn from different population. So to give a better example, like to kind of explain this, so the students who get Python training, so they constitute a different population now, okay.

And the students who do not get a Python training, they are here, right? So that is another population or the diamond is a different population than stone. Stone is this population, whereas diamond is this population. Or the people who are not thieves, this is this population, and people who are thieves, this is this population, right? So they are basically different populations.

The moment you identify these populations, so you have an alternative hypothesis, so we are basically talking about two distributions. So that is what you need to remember. And if let us say it has a mean, mu 0 and mu a, for example. Now I have to, what I have to understand from here is that is after the training or after say, you

know let us say after the training, those who have gone through the training actually constitute a different population or not.

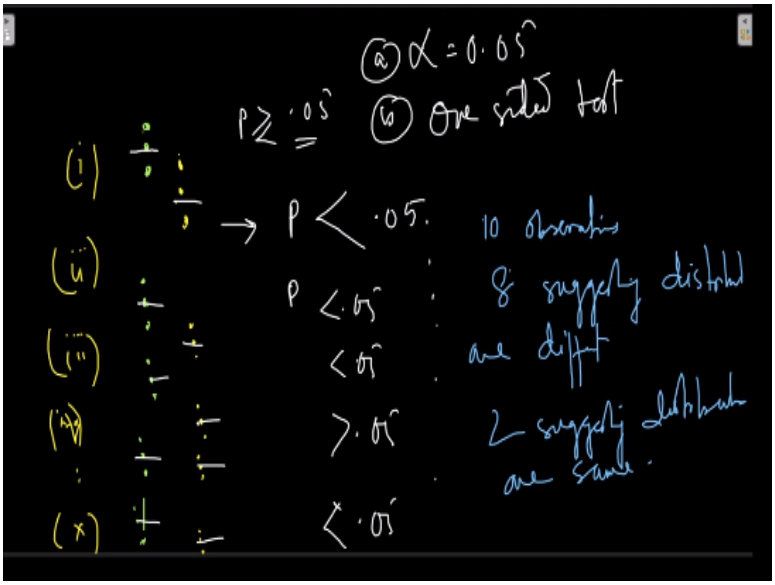
So that is what I have to understand. So if the training has no impact, so let us say we just take the training example, if the training does not have an impact, so let us say this training has an impact has impact, but if the training does not have any impact, what will happen is that my yellow distribution and the green distribution they will sort of they will sort of coincide.

Because you know the population are basically going to remain same. The every characteristics are going to remain same. So let me write down training say, so I will write mu 0 equal to mu a. What I will write is that the training has no impact, has no impact right? Now what we do when we analyze the power of a test is simply we try to see if the difference between this two distribution is actually significant or not.

And that we do with a certain procedure and I am going to show you that. So what we will do here is say we take the let us say from the previous one I will take certain samples from each distribution okay say people who have gone through training I take this information, the samples, and people who does not have gone through training I take this the yellow distribution and people who have gone through training let us say this is my green distribution.

So I take these observations here, okay. And I what I do is I know that what I have to do I have to calculate the P value, okay.

(Refer Slide Time: 08:54)



And I take this observations between from this two distributions, and I see if the mean, if the mean of these two distributions, the difference between the two means, if the difference between the two means if we take and the corresponding P value is actually, you know whatever is greater or lower than 0.05. Let us say that is our decision criteria, okay.

So whenever we do calculate this power of a test, we always have to define the alpha. Let us say we will take alpha equal to 0.05 and we say this is a one-sided test okay, one-sided test, alright. Now that is one case. So let us say in here we have seen that let us say in this particular case, I have seen P to be less than 0.05. So we are we will think that okay, this two distributions are different.

Now that is one observation, right? So you have to take n number of observation to come to this conclusion. So let us say if I just keep on plotting, so let us say this is my case (i) and I will take case (ii), case (iii), let us say case (iv). And let us say I take up to case (x). And what I find is that in each case, I, you know take these observations here, let me just draw it like this randomly.

And this is here, let us say I have this yellow dots here. Let us say this is really far, this is really far, this is really far, this is really away, let us say this one is close. And this one is let us say again very far, okay. So if I take the mean of all this samples, so I get this, I get this, I get this, I get this, I get this, I get this, these are very close, I get this, I get somewhere this, okay.

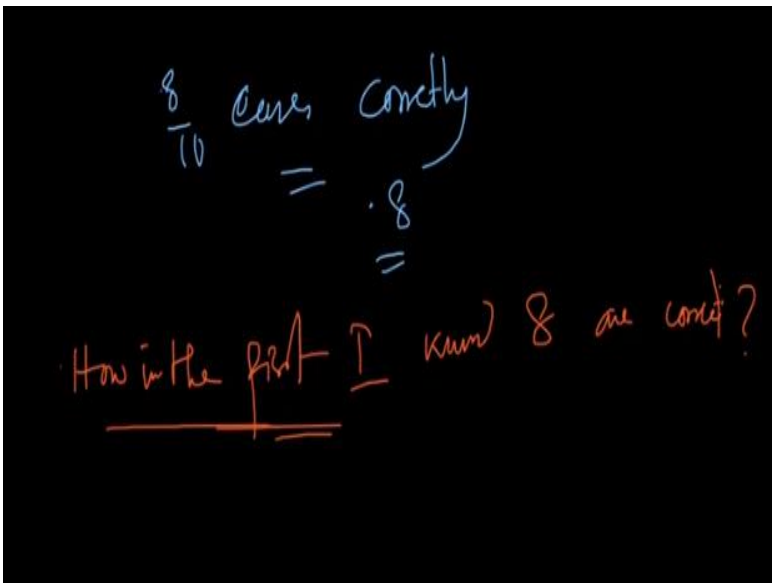
And let us say the P value corresponding to this is less than 0.05, this is less than 0.05, but this one is actually greater than 0.05, this one is actually less than 0.05. Now since there are n number of cases, since there are so many cases, I will see this one is actually violating, this P value is actually violating, but the rest of the P values are suggesting sorry, the rest of the P values are suggesting that, let me not do that.

So the rest of the P value is suggesting that these two distributions are different. Now the power of it is, what it does is, is actually if I have say, let us say there are 10 such observations and 8, so let us say out of the 10 observations 10 observations and 8 suggesting difference, 8 suggesting that the 8 suggesting difference 2 suggesting distributions are same.

8 suggesting distributions are different 2 suggesting distributions are same. Now what statistical power would do is that, so in 8 cases, we are able to reject. So we are able to reject the null hypothesis in 8 cases, okay. And in 2 cases, we are unable to reject the null hypothesis. So it is the essentially we are able to, we are assuming that this two distributions are different. We are assuming that the two basically there is an impact of the training.

And if that is so then we are able to identify 8 out of 10 cases correctly, okay.

(Refer Slide Time: 13:01)



Correctly we can identify. And the power of test is going to be 0.8 because we are able to identify 8 out of 10 cases. Now a very important question will come is that well, how did I know in the first place, how in the first place I know 8 are correct? Yes, my you know sort of specification the way I have you know my samples are showing that 8 correct out of 10.

But how would I know that these 8 are correct whereas these 2 are false? I mean it could have, it might have so happened that these 2 are correct and 8 are false, right? So how would I know that the 8 are correct? So what is where is the sanctity in this claim? Now to actually talk about that, it actually comes from the design, how we are actually designing the whole power thing and how we are actually designing the sample size and some of the other criteria, which I am going to talk about.

So those are kind of giving sanctity to the whole thing called power of a test, right? Now we will talk about that. So couple of things about power, but in this lecture let me just finish here and I will talk about couple of characteristics about the power and how actually I can design the power of a test. Thank you.