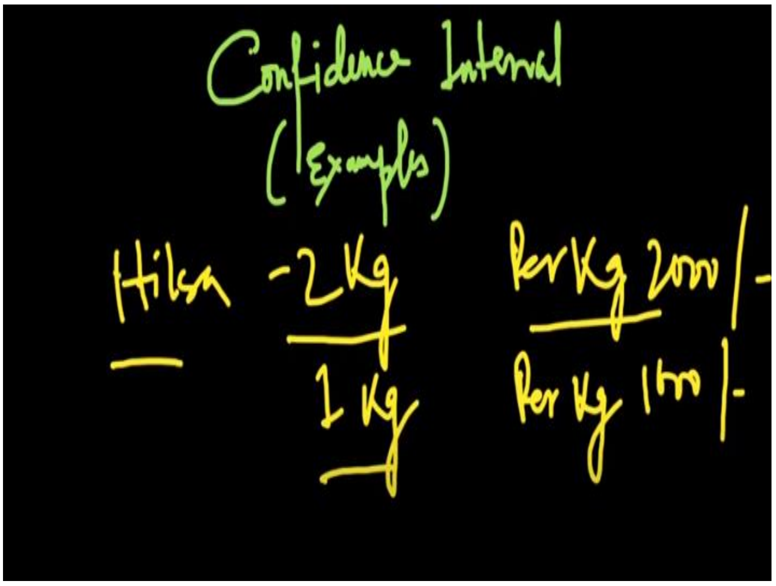
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology-Kharagpur

Lecture - 38
Confidence Interval Example

Hello and welcome back to the lectures on Applied Econometrics.

(Refer Slide Time: 00:30)



In this lecture we are going to do a couple of more examples on creation of confidence interval. So here as a sample statistic we are not going to use the popular sample mean but instead of that we will be using sample proportion. Let us start with the problem. Let us say we will deal with the same problem of hilsa population. This hilsa is really a nice fish and you know it is really a commercial fish and we have enough reason to (()) (00:55).

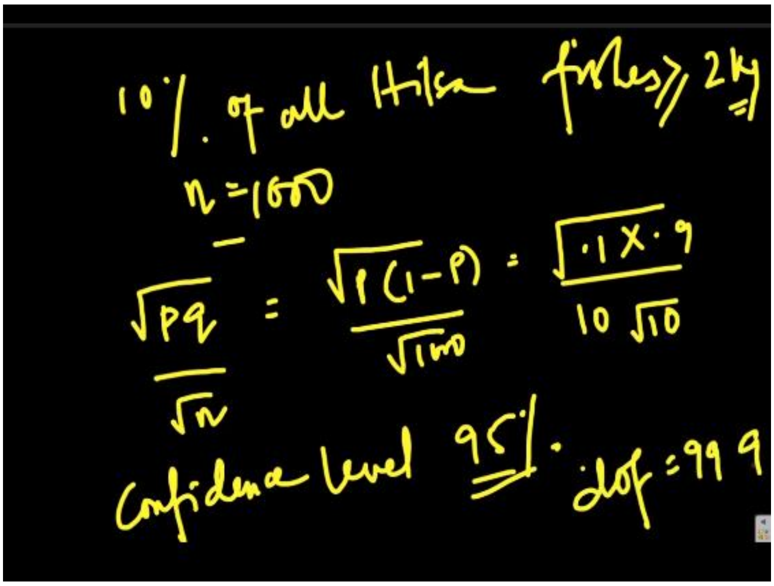
People in the Padma River Basin they depend, a lot of people their livelihood depends on the hilsa. Now the thing is actually there is something very interesting about hilsa and the size like if the size of hilsa fish is like let us say you know basically if it is a big hilsa, let us say 2 kg okay. So the price you will get perhaps per kg would be maybe let us say 2000 Rs. okay, per kg Rs. 2000, okay.

Whereas if you get a hilsa of let us say 1 kg, then the you know you will get in the market say per kg, I do not know let us say 1000 Rs. So depending on the size, basically the taste of hilsa fish depends on the size. So if the hilsa fish size is big usually the price is very high because we consider it to be really tasty, okay.

So then we have a reason to actually have you know or actually you know create such an environment that we get big hilsa fish so they get enough time to grow up. Now then we will try to know, okay let us say if that is if this is the price and weight you know relationship so we will want as many as possible hilsa fish which are more than 2 kg, okay.

Now so with that interest, let us say we do actually, you know get some you know we do some fishing, we go in the fishing boat, and we actually do the fishing. And let us say we actually get this hilsa fish.

(Refer Slide Time: 02:33)

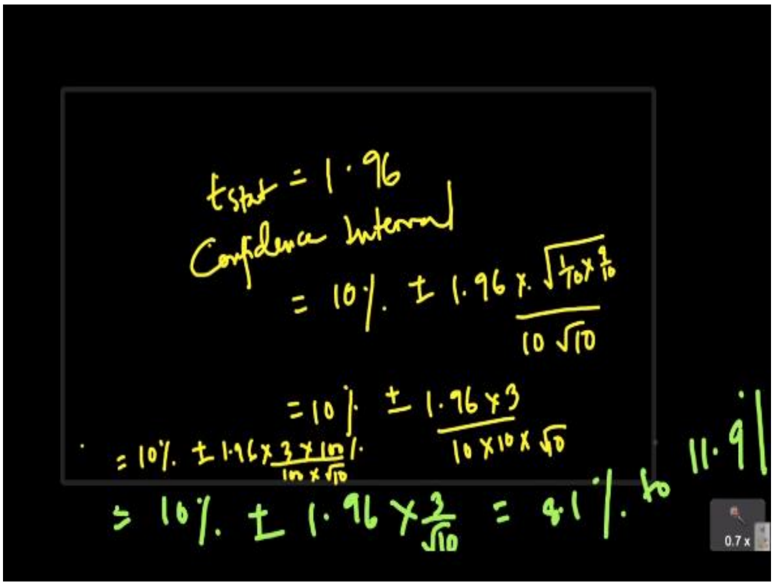


And we actually see that let us say there are 10% of all the fishes of all hilsa fishes, we have above or equal to or above 2 kg, okay. So that is the proportion we have, which are equal to or above 2 kg, all right? Now let us see we have managed to get 1000 hilsa fishes or n is equal to 1000. And now we have to create the confidence interval here. So how do we do that for proportion?

We get the standard deviation is, standard square root of pq that is square root of P into 1 – P. So that is 0.1 into 0.9, right. So that is our square root of pq. But then we have a correction factor, which is basically our square root n right? Square root n and if I do that square root n is going to be 1000. So basically it is going to be 10 square root of 10, okay. So that is basically the number that we will get.

Again, we will have from the previous distribution, let us say with a, we have a confidence level of let us say given confidence level of 95%. And we have seen previously for confidence level of 95% degrees of freedom, of course, if n is equal to 1000 for a t distribution, that dof is going to be 999. So we have already seen for the dof is equal to 999.

(Refer Slide Time: 04:24)



My t stat value is going to be 1.96. So I am not just going back again to sort of you know show you how to explore the t table. So now we get the t stat. Now we get the confidence, now we get the standard deviation. And then we have our you know the sample proportion 10%. So basically, how we will construct the confidence interval is, the confidence interval is equal to so 10%.

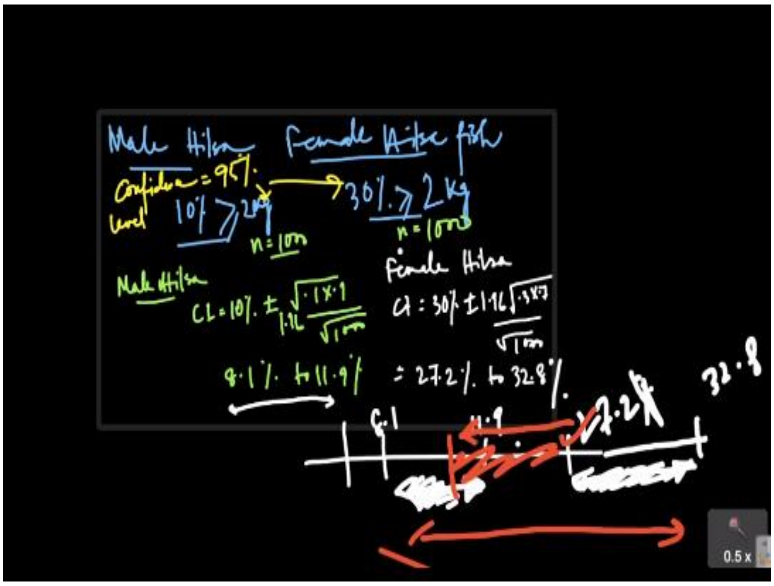
Then we have a plus minus sign of course here. And then we will have 1.96. And then we will have, so essentially 0.1 into 0.9. So let me actually do this math here. So 0.1 into 0.9 is nothing but 1 by 10 into 9 by 10, right. And then I have in the denominator 10 root 10. So that means 10% plus minus 1.96 and then 10, 10 denominator, so I can just take the 10 out.

So which would mean 10 into 10, square root of 10 and square root of 9, which is 3, okay. Of course, if I take a percentage, so if I have taken here in decimal, so I need to convert that into percentage. So if I do that, so I will get 10% plus minus 1.96 into 3 into 100%, of course by 100 into root 10, okay. Now 100% and 100 will cancel out. And let me use a different color so that we do not have any confusion.

10% plus minus 1.96 into 3 by the root 10. So root 10 is almost equal to 3, but it is not 3. But if I just calculate that, I did it separately. I am not doing it here, the whole calculation. So it is going to be actually 8.1% to 11.9%, okay. So basically the true population mean, the true or the true population proportion of you know the hilsa fish being above 2 kg is going to be in the interval of 8.1% to 11.9% with 95% confidence level.

So that is how we will interpret this result, okay. Now this is a simple problem. But actually there is a good use of confidence interval when you try to see if the two, let us say, you know two samples are drawn from the same population or different population. So let us say I am keen to understand, let us say I again go back to my hilsa example, okay. Hilsa is a favorite fish, hilsa is a favorite fish. So I will again go back to the hilsa example.

(Refer Slide Time: 08:32)



And here, I will actually try to see let us say I, in my sort of leisure time I started thinking, okay, let us see there could be male hilsa fish and there could be female hilsa fish, okay. Now I really do not know let us say if the size of male hilsa and females hilsa actually differ. So let us say hypothetically, male hilsas are usually bigger than the female hilsa.

So then is there a possibility that in the whole population I have, like more, you know sort of bigger fishes are basically coming from the male population vis-à-vis the female population. So let us say that is a problem. And I want to see the difference between the male population who are above 2 kg and the female population who are above 2 kg if there is any significant difference, okay.

So let us try to construct this problem. Let us say I get, again I do the sampling. And for male hilsa, I let us say I get the, you know all the male fishes. And let us say I get a 10% male hilsa, which are above 2 kg, okay, more than 2 kg, more than equal to 2 kg. Whereas I actually ended up getting 30% of the female hilsa which are above or equal to 2 kg, okay. So that is intriguing.

So my, it is actually sort of counteract, basically contradicting my hypothesis and our hypothesis. So let us say I want to see from this information if I can create the confidence interval for male hilsa and for female hilsa and if we can actually summarize something or if we can actually conclude something about whether the two populations, whether the two samples are coming from the same population or different population.

Or to say in other words, if there is a significant difference in the proportion of male hilsa being above 2 kg and the proportion of female hilsa being above 2 kg, okay. That is what we want to see. Now let us take, you know let us take the problem, step by step, okay. The first step, let us say we first deal with the male hilsa. Let me use a different color, okay. Male hilsa, first we will talk about male hilsa.

So for male hilsa we will sort of use this. We know how to by now how to do how to deal with the proportions. So we know it is a confidence interval. I am not doing the whole exercise here. But we know how to get the pq. So it is going to be plus minus and let us say okay, I forgot to mention the n. So n, let us say here 1000.

Let us say I have taken 1000 male and 1000 female, just to make life easy for us, same sample size. And here, I will take the proportion. So basically 10 and 90. So basically the same way we do the math here 0.1, 0.9. And then a for n is equal to 1000

I know the t stat is going to be 1.96. And I should have mentioned previously, the confidence level confidence level for both the cases, confidence level for both the cases is equal to 95% let us say for this one and for this one.

Now with this information, we already have all these values calculated for us, so we really do not have to bother so much. And here we have, just to maintain the consistency of color, let us say here it is going to be square root of 1000. And we know the value of it, okay. So and that is 8.1 to 11.9%, 8.1% to 11.9%; the same value that we got previously. Now comes the female hilsa, the numbers a little different here.

And let us do that separately. Female hilsa, and here at my CL is going to be 30%. Everything remains same, just the value under the square root is going to be little different, 1.96. And here it is going to be 0.3 into 0.7. And then it is going to be square root of 1000, all right. And if we do that it is actually going to get, I already have done the math. So it is going to be 27.2% to 32.8%.

You do the math on your own. So minimize the size a little bit, 32.8%, all right? Now how do I come to any conclusion from here? So I got this numbers, all right, I got sort of two integral CL. Now how do I really use this values to actually say anything about you know this two distribution, okay. One thing we can actually, you know we have to see couple of things here.

So one thing is that we can see if this confidence interval, in both the cases, If they are going to contain 0, okay. Oh, let me actually, so this is something I got for you know for male hilsa population and female hilsa population. Now for both the population, so one thing I can see is that I can see the sort of the interval, okay? So it is let us say the number line, in a number line it is going to be somewhere here, 8.1 to 11.9.

And then we have something like here very far from the first one. And that is going to be 27.2% to 32.8. And let us not use the percent symbol here, 27.2 to 32.8, okay. So we clearly see that there is no overlap between these two confidence intervals.

So basically the value, the population parameter, let us say the mean value in the proportion here actually, proportion of the hilsa, which are above 2 kg for males, the proportion is lying here 8.1 to 11.9. Whereas, for this one for the second case, the proportion is lying in between this, okay. So we can clearly see that there is an overlap, right there is an overlap.

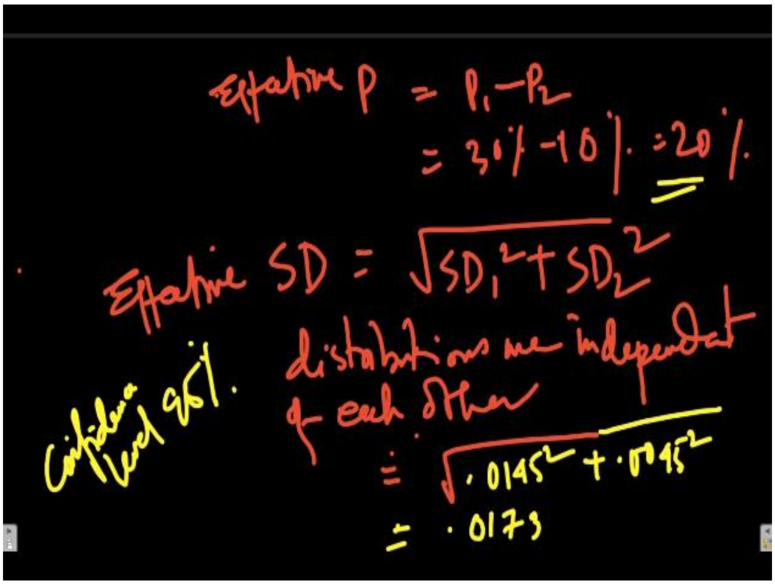
So the mean the proportion here in the male case and the female is going to be quite, they are going to be absolutely different. So that is the question of, you know these two things being same. So basically, they are different. So you can say that they have actually come from a different distribution, or actually the female, you know the female hilsa are actually going to have more, proportion.

Or the female hilsa is going to be more than 2 kg as compared to the male hilsa, okay for the same criteria. Now, so that is, you know this is kind of straightforward because we, the way we have created the confidence interval, they are going to be actually, you know there is an overlap. So they are going to be apart from each other. So we do not have much doubt here, okay.

But what if, what if let us say a part of the you know part of the confidence interval, if there are some overlap let us say, if there are some overlap let us say. Okay, instead of this one being here, if that one shifted to here. So then there would have been overlap and our life would have been a little difficult because nothing was clear here, right.

So another technique that we use here, we actually take the difference of the two different you know samples. So how we do that? So let me actually do the math here. So how we do that is we actually take the difference of the sample proportion. So we basically take the difference of these two distribution. So there are some formula and I am just going to do that. So basically, it goes like this.

(Refer Slide Time: 17:00)



So the effective sample proportion, effective let us say effective x bar or let us say effective proportion, effective P is going to be P 1 – P 2, okay. So that means 30% - 10%. So that is relatively simple. And that is equal to 20%. Whereas the effective sigma, whereas, if I want to have effective SD, now when I compare the SDs, standard deviations, there is a formula okay, effective SD.

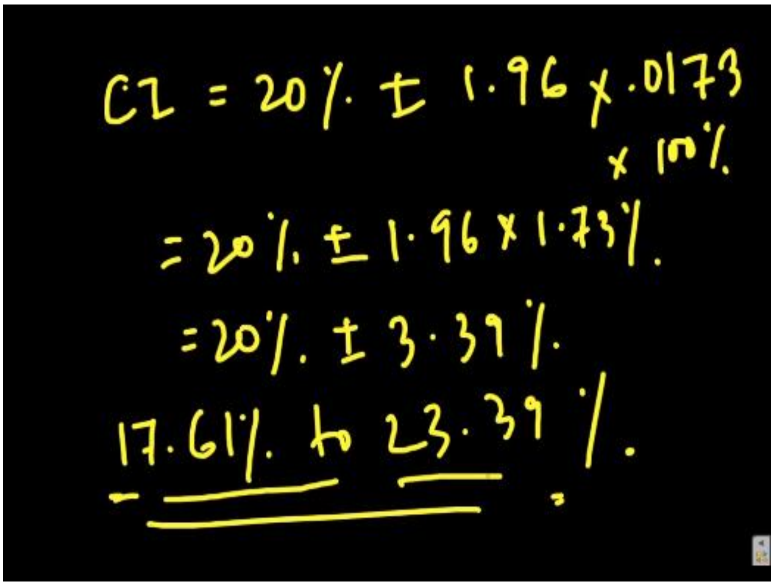
So that is basically, you know we take the difference and basically that is going to be SD 1 square plus SD 2 square. Obviously, that is the formula and one criteria which we remember that the distributions, that two distributions we are talking about, distributions are independent of each other, distributions are independent of each other, alright.

So that is what we basically keep in mind here. Now we assume that they are independent, the male female hilsa population, they are basically independent. And we get the sample, the standard deviation 1, standard deviation 2 from here, you get the value. And I already have done the calculation. So it is going to be, please put the value there, math there.

I will change the color, it is going to be 0.0145 square plus 0.0045 square, okay. Now if we do that, the effective value that we get, the effective SD that we get is equal to 0.0173 right, 0.0173. Now we got the confidence or we got the standard deviation, right? Now we have the effective proportion, we have the standard deviation.

And of course it is a confidence interval where we have, let us say the, of course I did not mention that, but confidence level here is 95%, okay.

(Refer Slide Time: 19:43)



So since our sample size is same, so it is going to be the value the CI that we are going to construct is 20%, which is 30% minus 10% plus minus 1.96 into, so value was 0.0173, okay. Now if we do that, we also do the necessary percentage adjustment. So that mean 100%. And what I will get here is 20% plus minus 1.96 into 1.773%, right? And this is going to give you something like 3.39%. So 20% plus minus 3.39%.

So do the math yourself. And if I do this, I will essentially get my confidence interval to be something like 17.61 I think. And that seems correct here and to 23.39%, right? So this is basically the confidence interval I get. Now look carefully, so the confidence interval that I get for the, when I have taken the difference of the two samples, so ideally, if the sample means were equal, so I would have a zero value in the confidence interval.

So it should have ranged from something negative to something positive, okay. So only then if the confidence interval is ranged from a negative value to a positive value, we can have the zero within that confidence interval. And we need to have the zero because if the difference between the two sample are not significant, then there has to be a zero value.

So essentially that means, when I take a difference between two samples, it will yield a value is equal to zero, but unfortunately or you know what you see here is that that is not the case here. So the essentially what you see is that all the values are positive 17.61 to 23.39%. So it means that the differences between these two distributions are actually significant, okay.

So essentially, we would, so that is the way of actually taking a difference of two distributions, two samples and we sort of predict whether there is a significant difference between the two distributions or not. So with this we will end the lecture on confidence interval. And in the next lecture, we are going to again talk about the statistical power and in detail I will, we will see the properties of statistical power. So thank you.