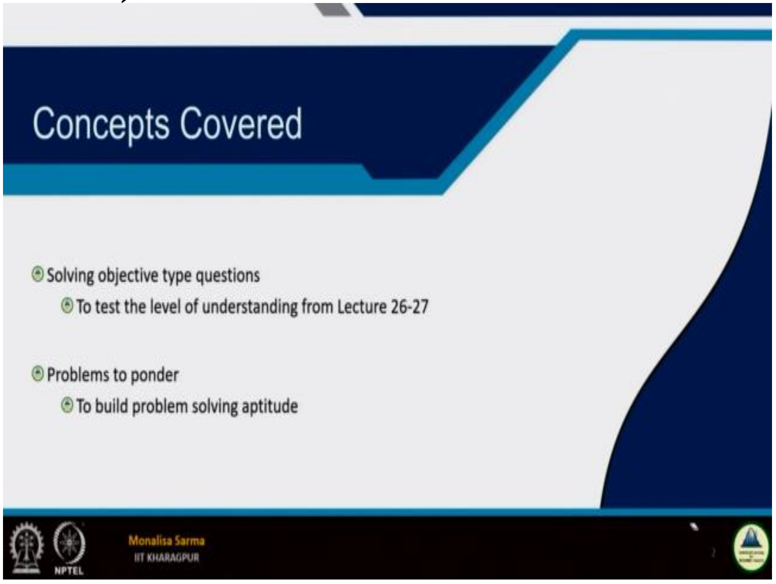
Statistical Learning for Reliability Analysis
Prof. Monalisa Sarma
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology - Kharagpur

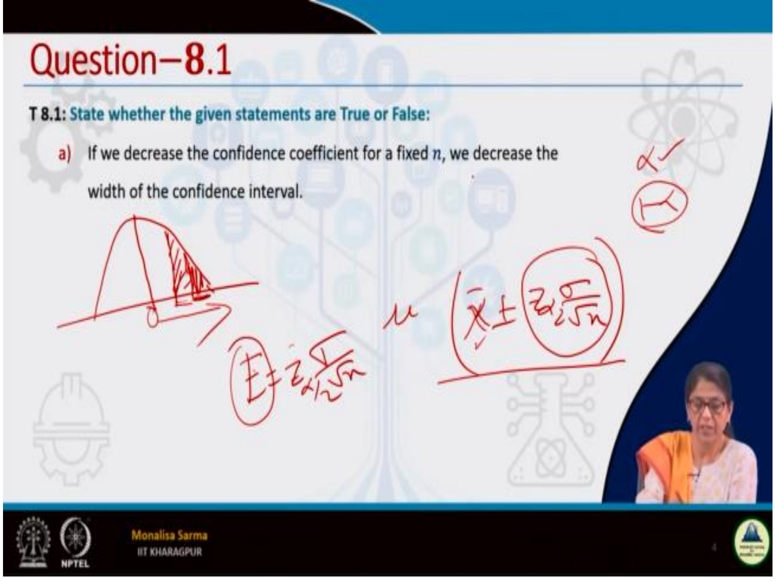
Lecture - 28
Tutorial on Confidence Interval

(Refer Slide Time: 00:30)



Hello guys, so we will do a tutorial today based on the topics what we have learned. So, last 2 lectures basically lecture 26 and 27. So, we are doing a couple of tutorials like because this things we will understand it more better using more and more problem solving. So that is the reason why I am giving so many tutorials.

(Refer Slide Time: 00:48)



So, now, first we will start with objective type question which basically will help us to keep making a quick recap of whatever we studied. So, first is if we decrease the confidence coefficient for a fixed n we have decreased the confidence coefficient. We have decreased the confidence coefficient meaning what? Means we have increased α we have decreased the confidence coefficient that means, we have increased α we have decreased $1 - \alpha$ this $1 - \alpha$ is divisible.

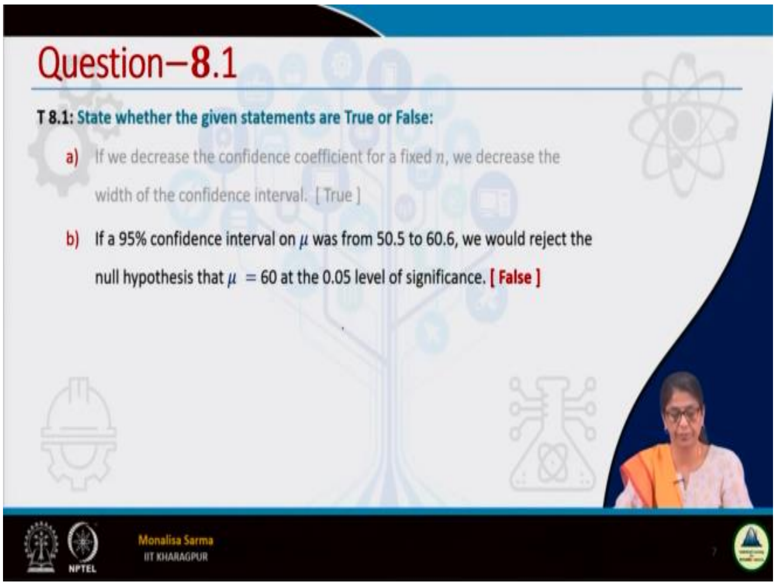
So, we have increased α if we decrease the confidence coefficient for a fixed n we decrease the width of the confidence interval, what is the width of the confidence interval is basically the error of estimation remember my what to say μ^2 value will lie within what range x bar plus minus what does that remember $z \alpha / 2 \sigma / \sqrt{n}$. So, my μ^2 will lie within this range I sorry my μ not μ^2 my μ will lie it within this range.

Similarly, if we talk of variance proportion whatever it is, accordingly it is value will get changed, so, this is my total range. So, this is my total width, my width is basically E, what is this? This value is E, so, I can write $E = z$ of $\alpha / 2 \sigma / \sqrt{n}$. So, here when I am increasing α, you see the diagram here from the table, increasing α means what? I am increasing the critical region, increase it; suppose if my α was here.

But basically here, when say 0.1, now I am increasing to 0.5, I am increasing this area, α I am increasing means from here I am bringing to here before it was this portion, now increasing means I added this portion as well. So, when I am increasing α, what is happening my z value is decreasing because this is 0 from 0, I am going this way it is increasing z value is increasing, when I am increasing α my z value is decreasing.

So, in this expression, when my z value is decreasing, what is happening? z value is decreasing my means my E will be smaller, my error of estimation will be smaller when my error of estimation will be smaller that means my width will become less, is not it? So, we decrease the width of the confidence interval that is true.

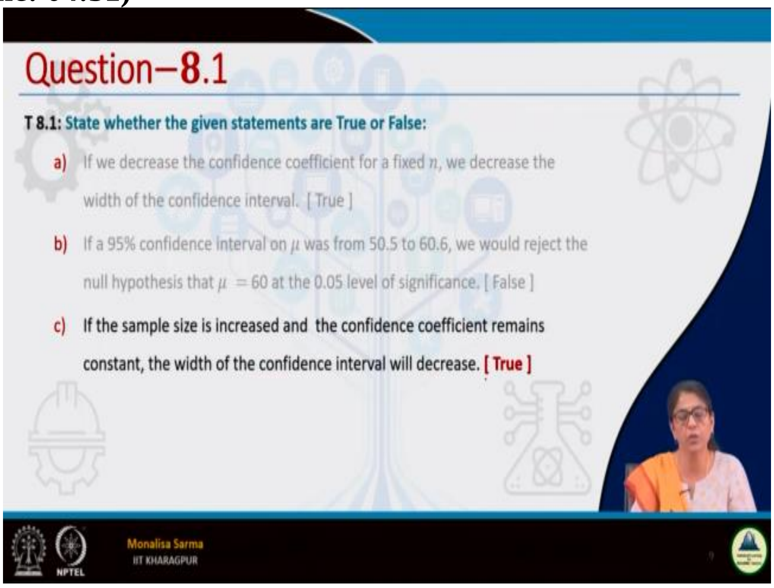
(Refer Slide Time: 03:47)



So, next question, if a 95% confidence interval on μ was from 50.5 to 60.6, this is a confidence interval given we would reject the null hypothesis that μ = 60 at 0.05 level of significance so, this is my confidence interval. So, μ very well lies within this region, because this ranges to 50.5 to 60.6, 60 lies in this region. So, any value that lie in this region, we will not reject the null hypothesis that is what we have seen.

If the value lies within the confidence interval, then the null hypothesis is not rejected. So, this value is lying within this region. So, we will not reject so this is false.

(Refer Slide Time: 04:31)

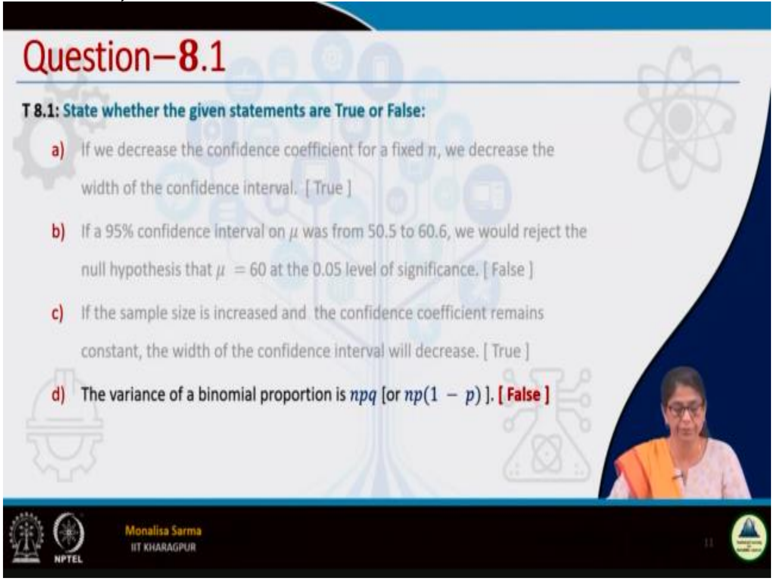


So, if the sample size is increased, and the confidence coefficient remains constant, the width of the confidence interval will decrease. Again same E what is my E? $E = z_{\alpha/2} \sigma / \sqrt{n}$, $\sqrt{n}$ is in the denominator. If denominator is bigger what happens our value becomes smaller. So, what happens if the sample size is increased? If I increase the sample size, my E will decrease E means the width will decrease.

Then and the confidence coefficient remains constant when α remains same, but I am the increase in the sample size, the width of the confidence interval will decrease, the width will decrease. That is one of the advantages, if I increase the sample size, then what happens? My precision is also increasing. At the same time, I am not losing on the confidence also. Usually, what happens when I change the confidence coefficients.

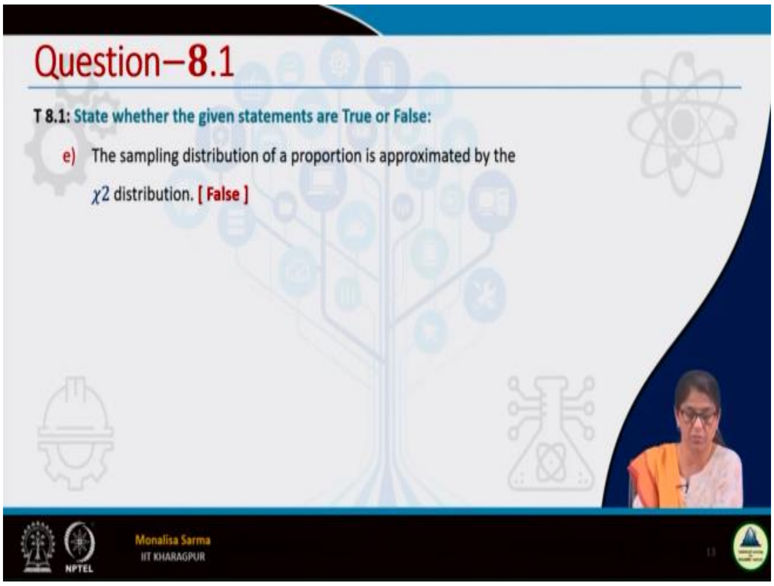
Change the confidence coefficients means, when I reduce the confidence coefficient my precision is increasing, but my confidence is going down is not it? But if I increase the sample size, still my precision is increased, but my confidence does not change my confidence still remains the same. So that is the advantage of increasing having a bigger sample size so, this is true.

(Refer Slide Time: 05:52)



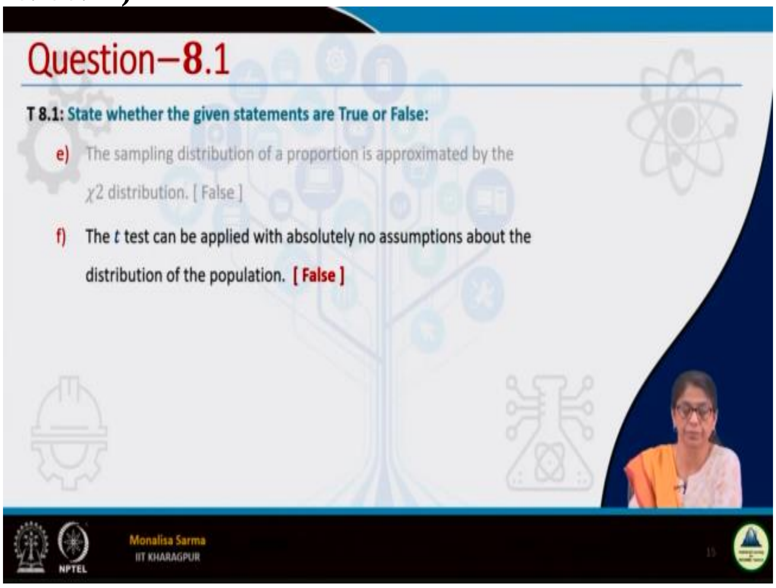
The variance of a binomial proportion is npq or np 1 - p, is it true? Just now, we have seen it I mean sorry, not just now, I mean, in the last lecture, we have seen, so, the variance of a binomial proportion is p x 1 - p, p x q, that is the variance of a binomial proportion so, this is false.

(Refer Slide Time: 06:17)



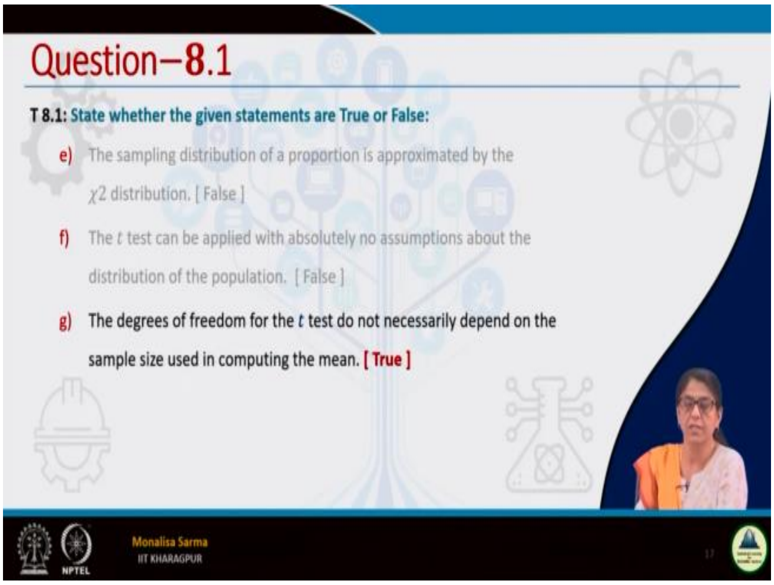
Again, there are some more questions, the sampling distribution of a proportion is approximated by chi square distribution as a true we have seen it in the last lecture, the sampling distribution of proportion is not approximated by chi square distribution it is approximated by the normal distribution considering this as the binomial population and considering a bigger sample size it is approximated by normal distribution it is chi square distribution.

(Refer Slide Time: 06:44)



The t test can be applied with absolutely no assumption about the distribution of the population completely false t test is very much sensitivity to normality assumption of the population.

(Refer Slide Time: 06:59)



The degree of freedom for a t test do not necessarily depend on the sample size used in computing the mean this thing I did not take this in the class so, we will explain it here usually what happens whatever we have seen in the t test or chi square it has one parameter that is the degrees of freedom and degrees of freedom how do we take? It is the sample size is not it?

Sample size - 1 is out in degrees of freedom for t test chi square test f square test or f test for all these days we have seen the they have just 1 parameter that parameter is the degrees of freedom what is the degrees of freedom? Degrees of freedom is the sample size - 1 but this question, but you see there are some situation like let me take an example suppose I am interested suppose a factory is producing stone chips.

You know this small shown stone chips so, a factory is producing that and I want to know the mean width of the stone chips. So, factory is producing huge number of stone chips and I am interested in me not be stones is because I need a particular size of stones for building something some I want to build something somehow some whatever some hotel whatever it is. So, now, how do I take for that?

Definitely, I should to know the mean of the population, I will have to take a sample. So, suppose taking a small sample will definitely make no sense in such cases. We will take a bigger sample suppose we took 100 a sample of say around 100 stones, definitely we will not count 1, 2 on an estimate we have taken around 100 stones and now will have to take the mean of this 100 stones how we will have to take the mean of the each.

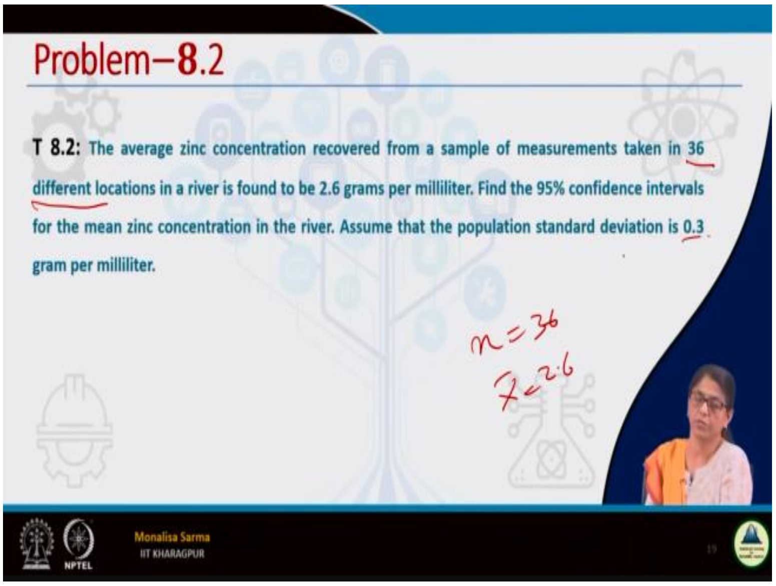
And every how will have to weigh it what is the weight and then we will have to find out the mean no that we will not do that we are not so jobless. So, what we will do is that we will take this whole 100 samples in a weighing machine we will just weigh the weight of this 100 stones and divided by 100 that becomes my mean weight of the stones, that is my mean weight of the stones now, what is the variance of the stones?

So, variance definitely I cannot just weigh it together and can find out the variance and it is also not possible for me to find out the variance of this 100 stones my sample size after 100. So, it is not possible for me to find a variance of these 100 stone also from each stone I will subtract from the weight of each stone I will subtract it from the mean whatever mean I got that is also not a very interesting job.

So, what I will do, I will take suppose I took 10 stones, and from for the 10 stones, I have sincerely calculated a variance. For each stone variance you know how to calculate the I will have to calculate from the mean I will subtract this value of weight of each stones. So, I have taken a from this sample I have taken a sub sample of say 10 and from there I found out the variance. So, in this case, then my degrees of freedom will not be 100 - 99. Later my degrees of freedom will be 10 - 1 9.

So that is very much correct. And in this sort of situations, where it is really not feasible to take the find out the variance of whole sample, this sort of thing this process is also taken up. So, the degrees of freedom for t tests do not necessarily see the word depend on the sample size using computing the mean yes, it is very true it depends on the sample size use for computing the variance so, this statement is true.

(Refer Slide Time: 10:31)

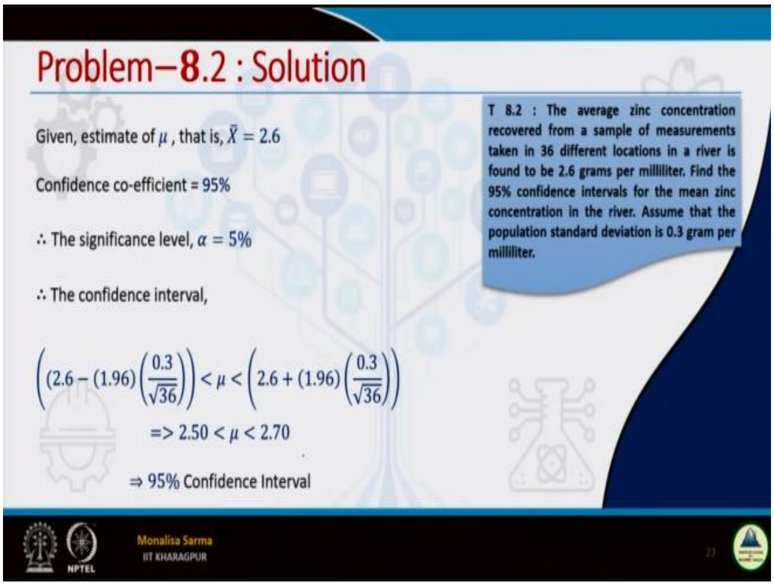


Now, we will see some problems. So, last 2 lectures whatever what we have learned, we try to find out what to say confidence interval basically confidence interval and how it changes with the sample size, what is the error of estimation and all those things we have learned and how to infer about the population how to infer about the variance, population proportion how to infer the variance of population we have learned in the last 2 lectures we have learned all those. So, based on that we; will be seeing some problems.

The first question you see the average zinc concentration recovered from a sample measurements taken in 36 different location in the river. So, 36 different locations that means, my n is 36 in a river is found to be 2.6 grams per millilitre. So, every is that means is this μ or x bar? This is x bar because this is we are talking about the sample. So, what is my x bar? x bar is 2.6.

Find a 95% confidence interval for the zinc concentration in the river we have to find a confidence interval assuming the population standard deviation is 0.3 good it is they have given the population standard deviation they have assumed the population standard deviations 0.3. And we have to find the confidence interval we can very well find out a confidence interval because we will be using z α / 2 instead of t α / 2.

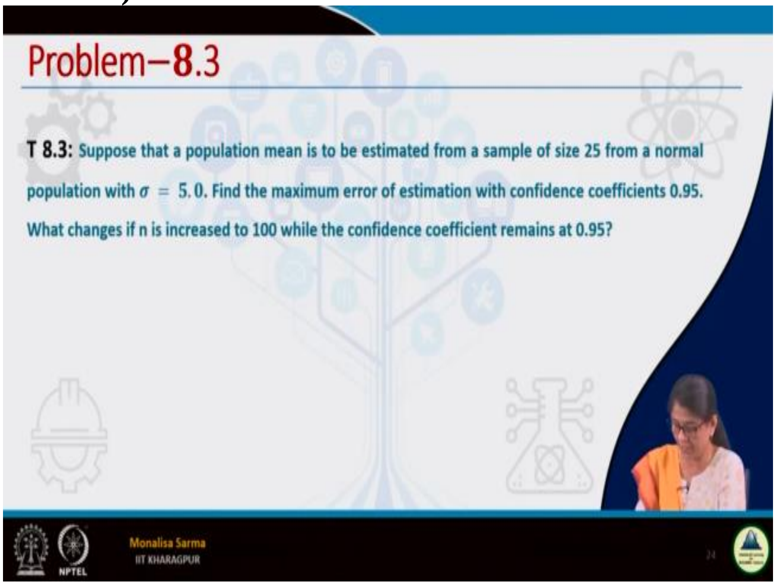
(Refer Slide Time: 11:58)



So, this is given significance level so, this is how we calculate the confidence interval is not it? Confidence interval how we calculate x bar - z α / 2 σ / √ n this is less than equals to μ this is less than or equal to x bar + z α / 2 what to say σ / √ n is not it? So, that is all we have all the value what is the 5% means? 0.025 corresponding to 0.025 z table we will find our value is 1.96.

You can check it that is 1.96 this is x bar value this is the variance σ and we have taken from 36 different locations so, √ n is 36. So, this is the confidence interval.

(Refer Slide Time: 13:10)

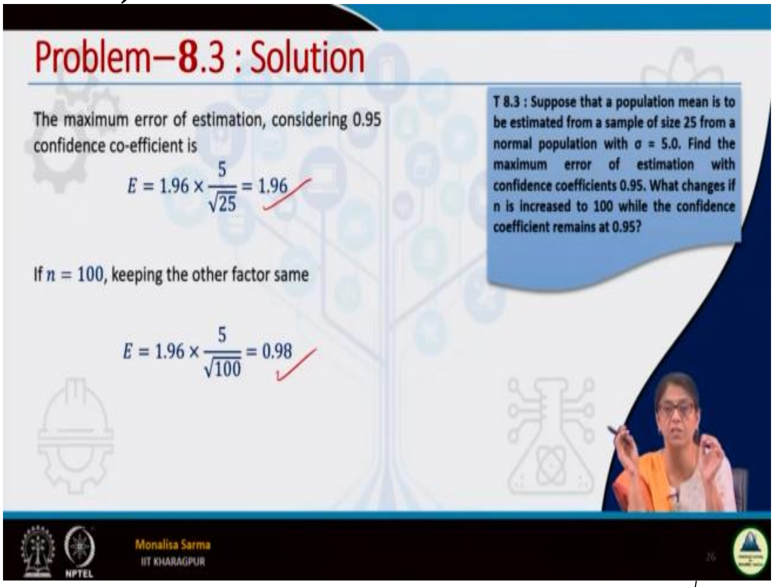


Similar type of question suppose that the population mean is to be estimated from a sample size of 25 from a normal population which σ = 5.0 and previous problem also average problem also population mean is not we do not have any hypothesized value. So, we just have to estimate from the sample find maximum error estimation with confidence coefficient 0.95.

So, here we have to find out maximum error estimation basically we have to find out a width what changes if n is increased to 100 the confidence coefficient remains 0.95.

We have already seen when we increase the n what happens our width decreases when we decrease our confidence coefficient our width decreases when we increase our significance level our width decreases we have seen that.

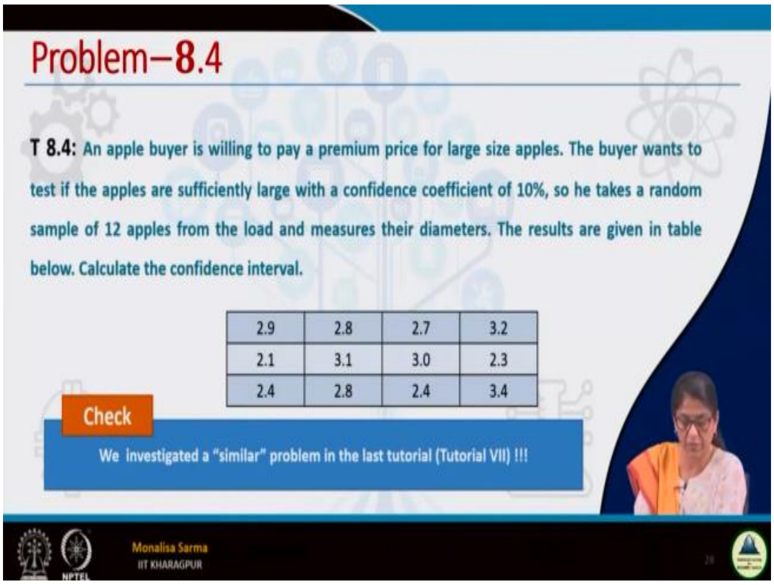
(Refer Slide Time: 14:07)



So, this is the formula for E what is E? E is nothing but z α / 2 x σ / √ n. So, this is z α / 2 is 9.95 is a significance coefficient that means 5% is a significance level α / 2 will be 0.025. So, 0.025% to 0.025 values 1.96. So, this is the E value considering the sample size of 25. Now, if we take a sample size of 100 see here the width was 1.96. Now, the width has become 0.98 our width has become reduced means our precision has increased.

Remember the concept of precision I am telling my x value lie between 2 to 5. I am telling my x value lies between 2 to 5 or if I tell my x value lies from 1 to 10 which is more precise my first statement is more precise. I am telling my x value is lying from 2 to 5. And other when I am telling my x value line from 1 to 10, I am a bit imprecise precision is going down. So, here when the width decreases, precision increases.

(Refer Slide Time: 15:16)

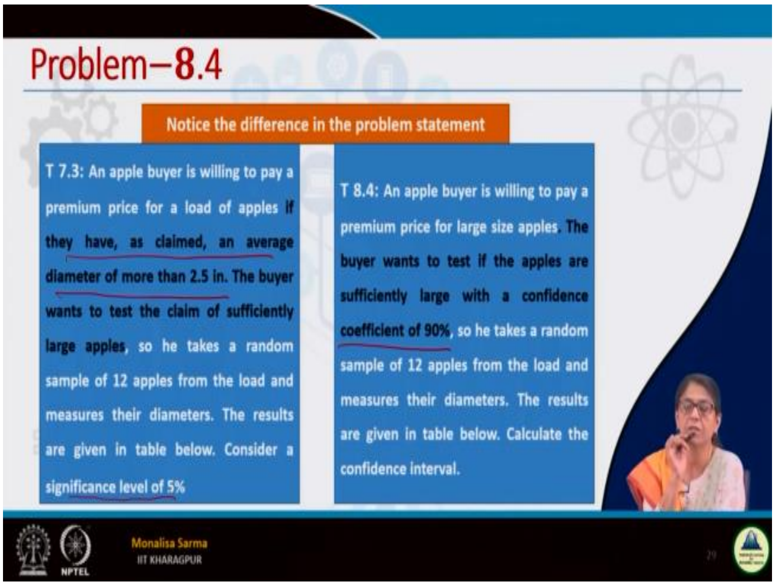


An apple buyer is willing to pay a premium, this question we have solved it, remember this question we have solved it, but we will see there is a difference in this the question what we have solved and what we will be doing now, an apple buyer is willing to pay a premium price for larger size apple there in that question, our size of apple was hypothesized size was given the buyer wants to buy an apple which is better than some 2.5 diameter.

Then only, the buyer will buy the apple in the premium price that was the question. Now, he just he does not know the size of the apple he just know that it is the size is bigger than he will pay the premium price. So, he just wants to know what is the size of apples for that, since he has he has no idea so he is just finds out the confidence interval same question, but the hypothesis value is not given the buyer wants to test if the apples are sufficiently large with a confidence coefficient of 90% that means a significance level 10%.

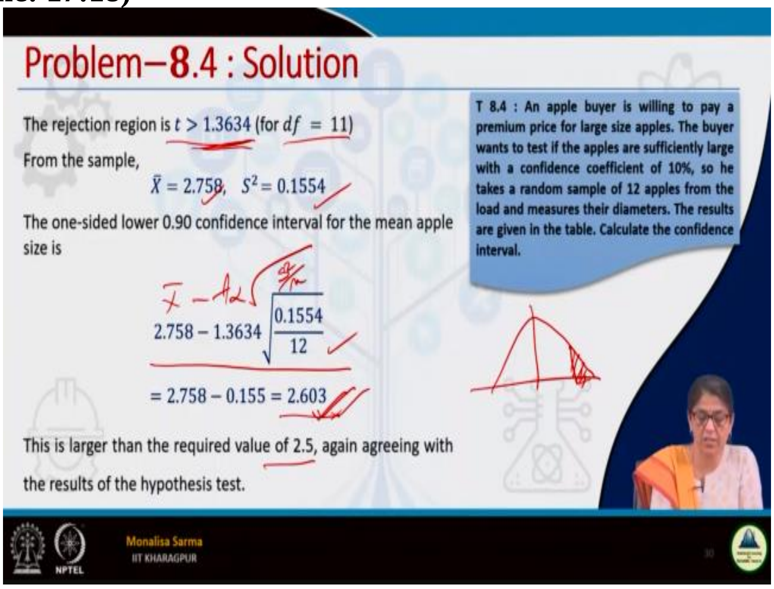
So, he takes a random sample of 12 apples from the load and measure the damages, the results are given below in this table. So, from here what you will find out from this table, so, he is what he wants to find out the diameter of the apple. So, these are the diameter of the apples are given. So, basically he will find out x bar he will find out S^2 is not it? So, and we investigate a similar problem in the last tutorial, you can check it.

(Refer Slide Time: 16:48)



So, that what was the difference? You will see the difference in the 2 questions here it was given if they have as claimed and average diameter more than 2.5 inch the buyer wants to then the buyer will be paying a premium price, but now nothing of that sort is given an earlier the significance level of 5% is given now, it is given a confidence coefficient of 90% when the significance level of 5% and what is the confidence coefficient confidence coefficient was 10% I mean sorry, confidence coefficient is 95%.

(Refer Slide Time: 17:18)



So, here from this region see buyer wants to find out if he is not interested in finding out he is basically is interested in finding out if the size of the apple is greater if the size of the apples and more is not it? So, what he will do is that in this case I say try to understand very carefully. So, he wants to find out the; if the size of the apple is more than he will pay the premium price. So, in this case, he will definitely not want to find out what is the size of the apples lies in what he wants bigger apple more the bigger more the better.

So, in that case, he will just find out what is the lower confidence interval because the apple can be how small from the sample apple can be how small. So, more the bigger more the better. So, he do not want to consider the lower upper confidence interval he will just find out the lower confidence interval. And if this question would have been a hypothesis testing; then remember we have here also we have checked for greater.

And condition μ is was to 2.5 μ greater than 2.5 when μ was greater than 2 my alternate hypothesis was greater than 2.5 then our rejection region was in the upper tail see the difference our rejection region was in the upper tail when we want to check it for greater at the same time for the same we want to look for greater but when we find a confidence interval for when our interest in finding out a greater.

So, when we look for a confidence interval confidence interval will always go for lower confidence interval why? Because we want to see how less it can go how small it can go greater means it can be any value we do not want to interval for that. So, t value for 1 degree of freedom since we are interested in finding out greater is 1 portion. So, we will not consider is the $\alpha / 2$ it will be just t of α t of α means t of 10 0.10.

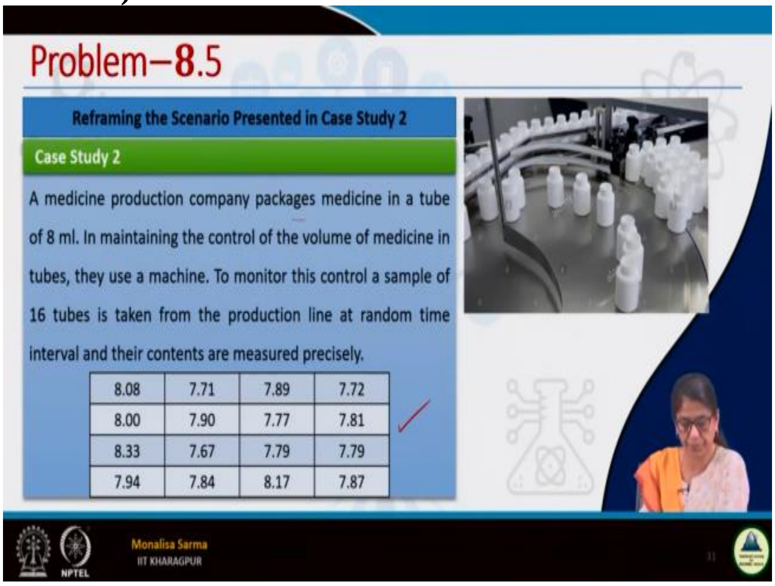
Sorry t of 0.10 for 11 degrees of freedom in the table if you will see you will get this value in the last problem in the last tutorial I have sold the t table here we could see this value so this is my t α. So, from the sample I got this is my x bar sample this is my from the sample I got x bar value I got a square value. So, now, I just have to find out a lower confidence interval. What is the lower confidence interval?

This is my lower confidence and x bar - t α this is x bar this is t α x bar - t α, then what is that $S^2 / n \sqrt{S^2}$ the n. If I would have interested in finding out the upper confidence, then it will be x bar plus, if I am interested in both finding out the lower and upper, then my α will be α / 2, is not it? So, since I am interested here, in finding out the greater, I do not want the lower regions, so it will be α x bar - t of $\alpha \sqrt{S^2 / n}$.

And so putting this value I got this is my value, my lower bound, that my apple can be as small as 2.60, the size of the apple will start from 2.60 to any value that I have not checked, I am not interested because more the bigger, it is better. So, this is my lower value. So, this is larger than the required size of 2.5. And if you see if you remember, if you do not remember I request you to please go back to the tutorial there in that for the same problem same all the data were same there.

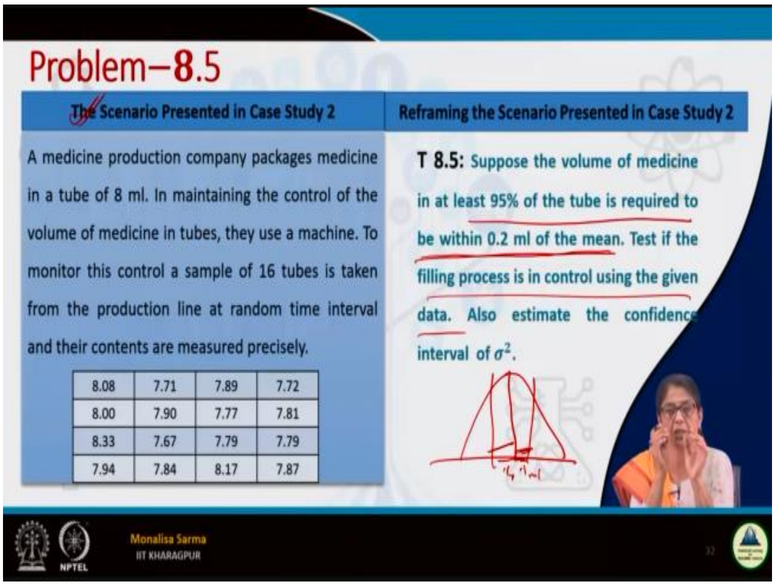
We have checked for hypothesis μ = 2.5 μ greater than 2.5 and we have rejected the null hypothesis μ = 2.5. So, here the confidence interval also says so, μ = 2.5 it is definitely it would have been rejected because my lower bound is only 2.6. So, μ = 2.5 will definitely be rejected. So, this is the larger than required value of 2.5. Again, agreeing with the results of the hypothesis tests that we have done in the last tutorial.

(Refer Slide Time: 21:35)



So, again, I am coming back to the case study to which I have used in my first lecture of statistical inference, and here I am reframing it a bit how I am reframing it here like so, in medicine production company packages medicine a tube of 8ml in maintaining the control of the volume of medicine in tubes they use a machine to monitor this control a sample of 16 tube is taken from the production line at random time interval and their contents are measured besides these are the volume of the different volume of the samples that we have taken.

(Refer Slide Time: 22:15)

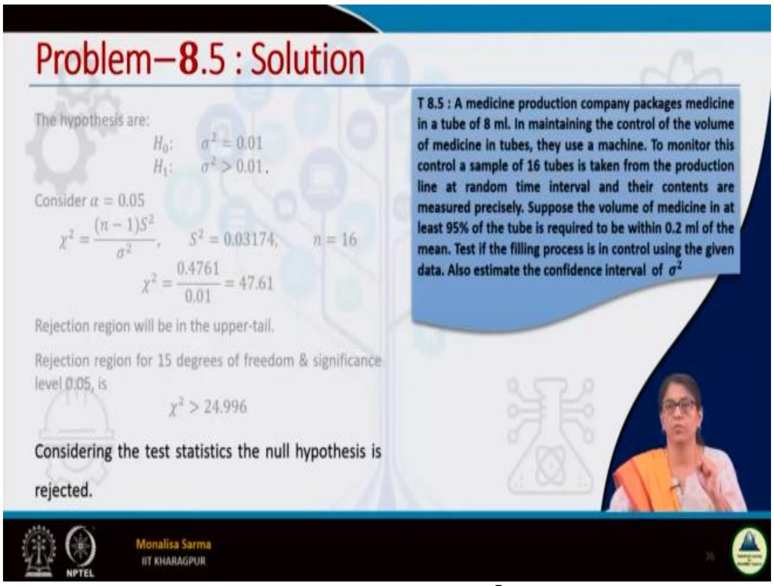


Now, this was my scenario which was presented my previous lecture, I have used this as case study 2 have used in many of my discussion the same example now I am changing a bit here, what is my change I am reframing it suppose the volume of medicine in at least 95% of the tube be within 0.2ml of the mean volume should be within 0.2ml of the mean. So, whatever the mean volume, it should be within 0.2ml of the mean that means, if this is the mean.

So, 0.2ml means 0.1 is this side 0.1 is this side 0.2ml of the mean this is 0.1ml this is 0.1, this side is 0.1 this side is 0.1 that is what we want to check is it the filling processes in control using the given data this is what our we want our variance to be I mean our standard division it is when it is given 0.1 it is not variance, when it is given within 0.2ml of the mean I think you can understand it is not talking about a variance it is talking about the standard deviation is not it? So, how much it is changing from the mean it is changing 0.1.

So, we need to check whether our volume is within 0.2ml of the mean that means maximum we can what we can bear is a standard deviation of 0.1 that we have to check. So, if the standard deviation is 0.1, so, whatever the variance will be 0.1^2.

(Refer Slide Time: 23:55)



So, what will be the hypothesis we have achieved σ^2 as a variance is equals to 0.01 here again now, alternate hypothesis we will check σ^2 not equal to 0.01 or greater than I am claiming that we will not check for not equal we will check for greater than why? Because when it is given within 0.1 of the thing so, anything which is less is covered there and null hypothesis.

So, what we have to check whether it is greater than 0.01 within 0.1 means anything it is less than between that will be here only is not it? Null hypothesis so, what we have to check whether it is greater than 0.01. So, it is again a 1 tailed hypothesis that is where σ^2 we have to check for 0.01 and our α is 0.05. So, what is the statistics for here we will be using because we are in have to infer about the population will be using chi square distribution.

The statistics that is used is n - 1 S^2 / σ as far as σ^2 from the data that is given we found as this is their square value, our sample size is 16 we got our chi square value 47.619. Now, we will have to first we should have calculated a rejection region according to hypothesis testing this rejection region should have been calculated here and before calculating the statistics rejection region first we calculate a rejection region; then we try to find out a statistic here anyway you understand the thing that is okay.

So, the how rejection region see now when we are interested in finding that it is better than 0.01. So, what will be where we will be our rejection region? It will be in the lower tail or in the upper tail? This is the lower tail this is the upper tail this is χ^2 $\alpha / 2$ this is χ^2 $1 - \alpha / 2$ if it is two tailed, if it is single tail, first let me delete this since we are interested in here single tail is something of this sort, but it is not a normal distribution chi square is very much a skewed distribution.

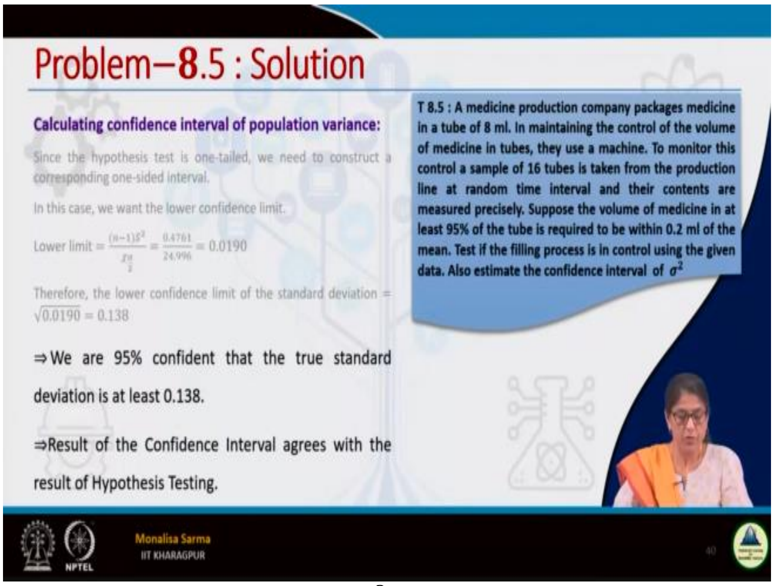
So, we will be interested in the upper tail or lower tail, this is the upper tail this is the lower tail. So, when we are interested in finding out than greater than 0.01. Our detection our null hypothesis is less than or equal to so, if we get a value greater that means, we are rejecting the null hypothesis is not it? So, it will be our rejection result it will be the upper tail. So, basically we have to find out what is the rejection region taking the significance level of 0.05.

Because it is 1 tailed we will not do α / 2, but it will be 0.05 what is the chi square value for the α 0.05 with degrees of freedom how much is the degrees of freedom 15 for 15 degrees of freedom and α = 0.05 we will find out what is the chi square value that is the rejection region. So, if rejection region if this value is greater than the rejection value the we reject the null hypothesis. So, what is the rejection region?

Rejection region is 24.996 rejection region for 15 degrees of freedom significance level of 0.5 on the chi square table, if you see the chi square table, this table is available in all standard textbook. Even if you do not consult a standard textbook you just Google in the Google is just write chi square table z table whatever it is you will get the table. So, this rejection region is for value greater into 24.96 what is our value?

Our value is 47.61 so, definitely reject the null hypothesis. So, considering the test statistic the null hypothesis is rejected. Now null hypothesis is rejected that means my σ square value is greater than 0.01. Now greater means how much data let us find it out.

(Refer Slide Time: 28:20)



How will you find out? We found that our σ^2 value is greater than 0.01. Now, actually how much value is σ so, it is better understood is greater than 0.01. But how much is that so, we will find the cause confidence interval. So, this is how we can find out from here again here which one confidence interval we will find it is greater than 0.01. And I accepted the hypothesis that it is greater than 0.01.

So, when I want to find out the confidence interval, so what confidence interval I will find I will find a lower confidence interval an upper confidence interval, definitely I will find a lower confidence interval greater means it can be any value, but it will start from what value I am interested in the lower value always remember when we are testing for an alternate hypothesis which is greater than my confidence interval will be always the lower.

And the rejection region will be to the right and my confidence interval will be the lower confidence intervals. If I am checking for something which is less than then my rejection region will be towards the left that means the upper tail and my confidence interval I will check for my upper confidence interval. And if it is not equal to definitely I will look for both lower and upper and my rejection region will be both left and right both the lower tail and the upper tail.

So, here it is greater so my confidence interval will be the upper one or sorry greater means my confidence interval will be the lower confidence limit. So, this is the formula for lower confidence in limit we have seen n - 1 $S^2 / \chi (\alpha / 2)$. So, here $\alpha / 2$ means we will not consider

1.5 / 2 we will consider because it is a single tail always α / 2 becomes α. So, it is χ (α) if α we have seen it is 24.996.

So, this is the value this is my lower limit. So, my variance is at least 0.01 variants in this tubes it at least 0.01. It is it can be much more than done but it is at least 0.01 that means the system really needs some change. So, if this variance is 0.01 say this question has asked for standard deviation, standard deviation should be within 0.2ml. So, I will have to give the accordingly I will have to find out the answer standard deviation only if this is my variance so what is my standard deviation?

It is 0.138 so it should be within 0.2 is not it? So that means my standard deviation should be 0.1 but I got standard deviation as 0.138 greater than that. So, you why the null hypothesis is rejected my lower interval is only greater than 0.1 that is 0.138 so results of confidence interval agree with the results of hypothesis testing.

(Refer Slide Time: 31:29)

So, with this I end this lecture here, these are the references and thank you guys.