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

Lecture - 30
Statistical Inference (Part - 7)

(Refer Slide Time: 00:31)

Hello guys. So, again today we will also learn some other topics on statistical inference. So, in the last class, we have seen inference on difference between variance or difference between means of 2 populations remember, so, for that we have considered different cases. So, in today's lecture, we will see inference on difference between variances for 2 populations. So, that means, we want to compare the variance of 2 populations.

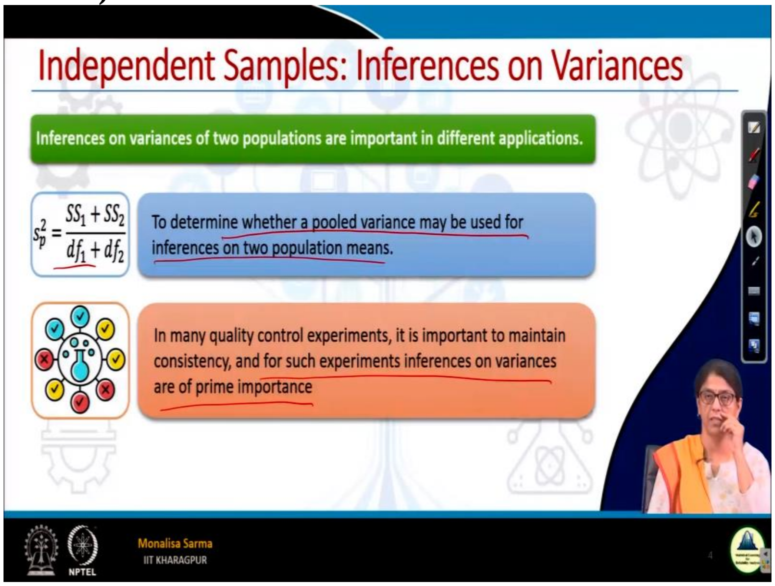
Now, can you from the last lecture, can you just find out where there is some application of this in the last lecture also, can you just think for a moment. And if you guys got the answer very good, if you did not get the answer, like when we discussed statistical inference on the difference between 2 population mean so, we have considered 3 different scenarios remember, for independent samples, we have only discussed in our independent samples, for independent samples, we have discussed 3 differences, 1 is variance known that was not an issue at all.

Very equally we could use this z distribution, another is variance is unknown, when variance is unknown, then what we have done when the for variance unknown, again there are 2 cases for variance unknown 1 is that we have assumed that the variances are equal and accordingly we have assumed that the variances are equal and from that we tried to find out the pool variance estimate remember.

And so, that test that we use for that we call it pool t test. And the statistics was pool t statistics. So, there what is our assumption was that both the variances are equal, but directly can we assume that both the variances are equal it will be because if the variances are not equal and if you assume the variances are equal by just when we are assuming the variance are equal, we are pooling the 2 variance by finding out a weight and giving out a pooled variance estimate.

If the pool estimated variances are not equal and you were assuming it is equals the result which you will get the result may be very incorrect. So, our total statistical analysis will go wrong in that case. So, we will have to find out whether the 2 population variances are equal or not, is not it? So, that is here we see inference on the difference between variances for 2 populations, this is how we will try to find out in this class we will see that.

(Refer Slide Time: 02:52)



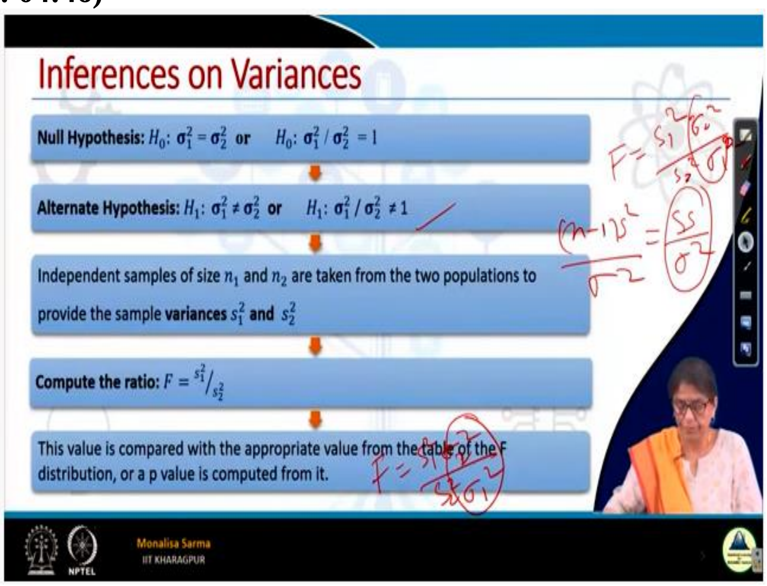
So, as I mentioned just now, what I mentioned when I am trying to find out a pooled variance estimate. So, the way to determine whether the pool variance may be used for the inference of 2 population means. So, this is 1 region why we may have to infer the variance and 2 populations to determine whether we can use the pool variance. Another is that you know lesser variance means that any system or equipment is of good quality any material if you see if there are lots of variances to here.

Then definitely it is not meeting the specification I suppose, I bought a material some cloth material I expected certain density of it and I got an next slot again suppose I liked the material and I bought a same cloth but then that cloth is totally different thing different density. So, that is not expected, is not it? So, this there in different again I have given the cool drinks example also, if I am buying a cool drink can of 1 litre I expect it to be 1 litre only I do not expect that I am buying it 1 litre.

And I am getting 900ml or 950ml I do not expect that if I get more that is merrier, but for me it is merrier, but for the manufacturer, it is a loss forget about the manufacture, look the consumer point of view also, Smaller variance will okay do 1 litre means slight variation I can agree but more variance definitely it is not acceptable. So, for quality experiment also it is important to maintain consistency and for such experiments inference and variance are of prime importance.

So, inference on variances, it has many applications we have seen these 2 applications these are the main applications.

(Refer Slide Time: 04:46)



So, how we go about same case how we have inferred on variance of a single population same way we will go, but when we have discussed inference on a single population, remember which distribution we have used when we have tried to infer on the variance of a single population we have used chi square distribution and we try to find out when the test statistics that we calculated it remember n - 1 S^2 by σ^2, n- 1 S^2 this is also we call it also sum of square by σ^2, this is my test statistics.

And if this test statistics falls within the acceptable region, acceptable region means within the 95% of the whole chi square distribution, then I say that is whatever is the null hypothesis is accepted remember, so, now here when we are trying to compare 2 different populations. Then 2 different population in the way and we are comparing the 2 different population variance, we cannot use chi square distribution chi square distribution is found for only 1 population.

Then we can we will be using f distribution we have already discussed this when we have discussed sampling distribution as I told you statistical inferences backbone is the sampling distribution only. So, what are my hypothesis my null hypothesis is $\mu_1 = \mu_2$ or I can read write $\mu_1^2 / \mu_2^2 = 1$, it is 1 the same thing then my alternate hypothesis $\mu_1^2 \neq \mu_2^2$ or $\mu_1^2 / \mu_2^2 \neq 1$.

But usually and most of the cases we will find when we are trying to infer on variance we always try to check whether my particular variance is greater than a particular value. Always in variances, most of the cases it is a single tailed and we try to compare the greatness of the variance because if the variance is less that is good, why and this is some in my null hypothesis is equal that is something which I want and less is also something which I want. So, why will again go and test it for less. So, usually when we are testing we are testing it for greater than.

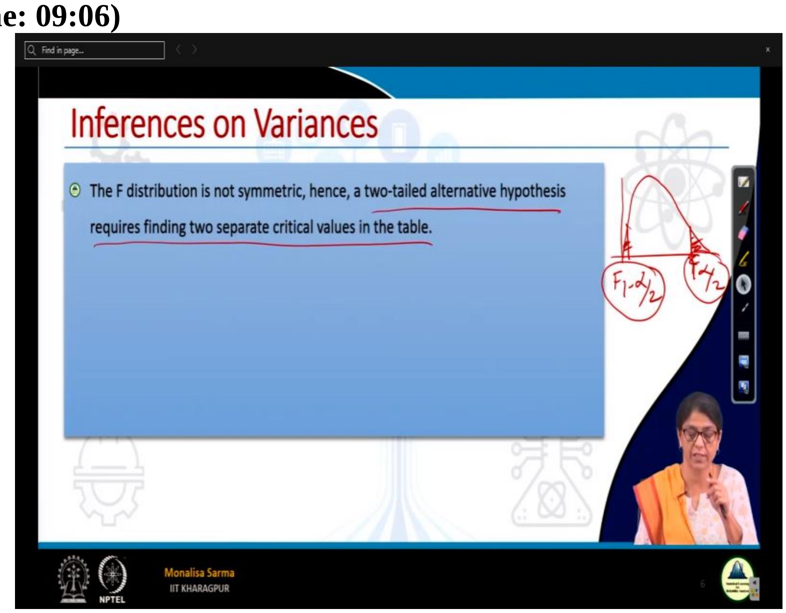
So, when you are trying to find out any inference of the variance at a single population or double population, usually we go for 1 tailed alternate hypothesis that is greater than but that does not mean we go for 2 tailed hypothesis. So, here is an example of 2 tailed hypothesis. So, independent sample of size n_1 and n_2 are taken from the 2 populations to provide the sample variance s_1^2 and s_2^2.

So, from that we will compute the ratio what is my F remember, $F = s_1^2 \sigma_2^2 s_2^2 \sigma_1^2$ is not it? This is my F ratio when I have taken $\sigma^2 \sigma_2^2$ equals to 1 that means, this book I am writing it properly my handwriting is not very good actually $s_1^2 \sigma_2^2$, $s_2^2 \sigma_1^2$. So, this portion is equal to 1 that means, what is remaining is my test statistics is s_1^2 / s_2^2. So, this is my test statistics.

So, this value is compared with the appropriate value from the table of the F distribution or a p value is calculated anything we can do either from whatever we can either we can find out the p value corresponding to whatever F we get or we can find out the rejection region for a particular specific a significance level we find that clear rejection region, here F distribution is not symmetric remember F distribution and then chi square distribution is not symmetric.

So, we will have to find out a rejection region for both the side both lower tail and the upper tail. So, we find a critical region and if the value of F falls within the critical regions, that means not means what to say if it is less than the critical regions, then we say that the variances are equal that means, the null hypothesis is accepted if it is false in the critical region, then we say that null hypothesis rejected.

(Refer Slide Time: 09:06)

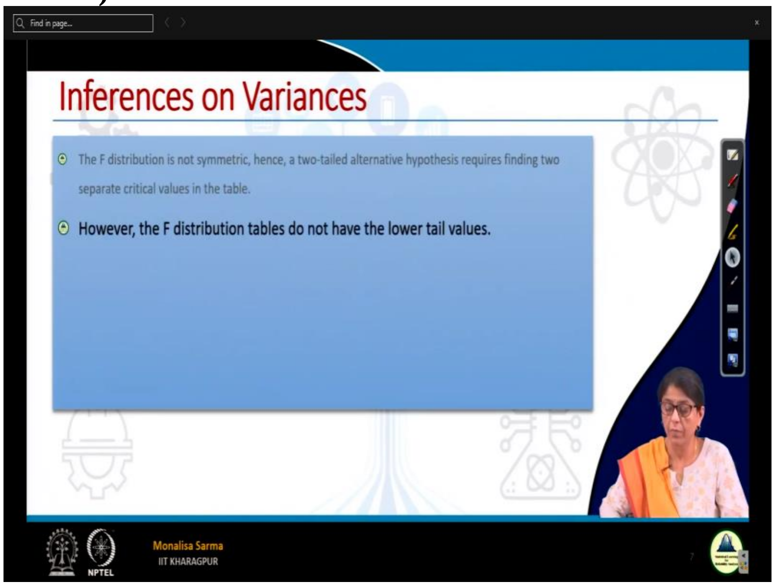


Now, the F Distribution is not symmetric hence a 2 tailed alternative hypothesis requires finding 2 separate categorical values in the table. So, if we see the F distribution it is something some this is not an exact figure of F distribution you can see a table my drawing is very bad actually.

So, it will be something this sort of this is 1 critical region, this is 1 critical region this critical region is probability of α to this.

So, it is F of α the value corresponding to this F of α. So; any value greater than F of α will fall in the critical region if this is F of α. So, this is what will be F of α / 2 1 - α / 2. So, if it falls in the left of this then also this falling in the critical region. So, if my value is less than this value F of 1- α / 2, or if my value is greater than F of α / 2, then it falls into the critical region, then we reject the null hypothesis.

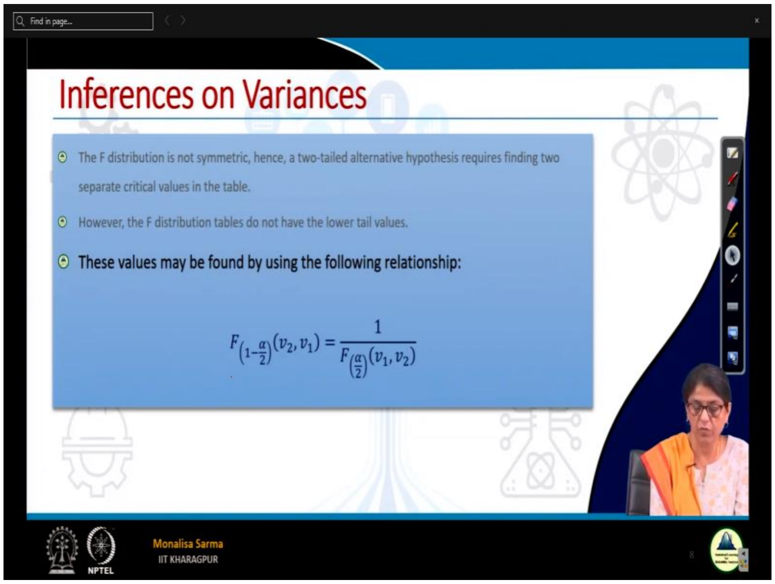
(Refer Slide Time: 10:12)



However, F distribution do not have the lower tail values, in the F distribution table, you see you do not have the lower tail value, lower tail means my α is a 5%. So, I will go to this figure this is α is 5% so, this is α / 2 is what will be the α / 2 0.025 is not it? F of 0.025 if this is 0.025, this is my lower tail value will be 1- 0.025 this value, this F of 1 - 0.025 this we do not have in that table.

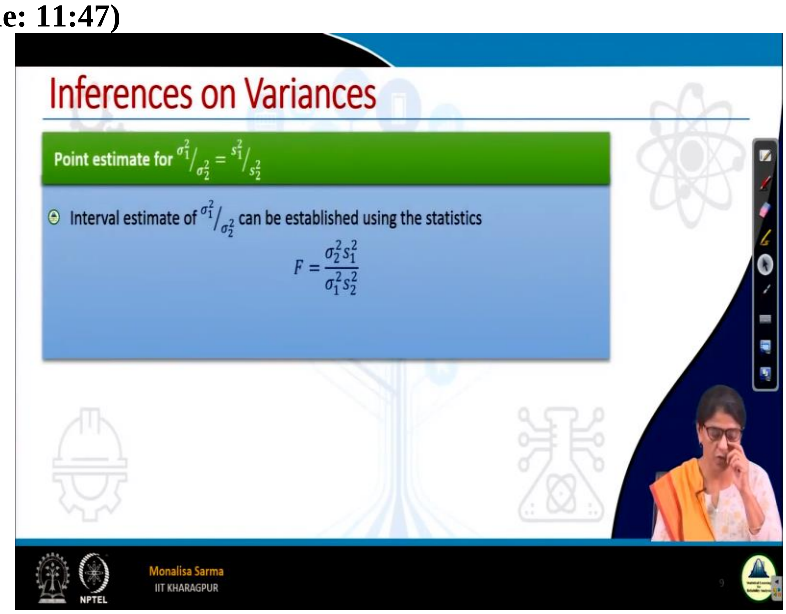
In there for F table we have values for 0.01 0.05 0.025 0.005 something this for very smaller α value we have values, but then for bigger α value, we do not have the values in the F table. But then we have one theorem we have seen that theorem when discussing the sampling distribution; we can use that theorem for our rescue.

(Refer Slide Time: 11:08)



So, what was the theorem remember, so F of 1- α / 2, but here the degrees of freedom changes, if F of 1- α / 2 degrees of freedom F of v 2, v 1 = 1- F of α / 2 v 1, v 2. What does v 1, v 2 means that the numerator I have the sample size first what to say variance of the first population that is s 1 ^{2} and the denominator I have the second variance of the second populations. So, the degrees of freedom just when I try to find out the upper tailed value, we will see with an example.

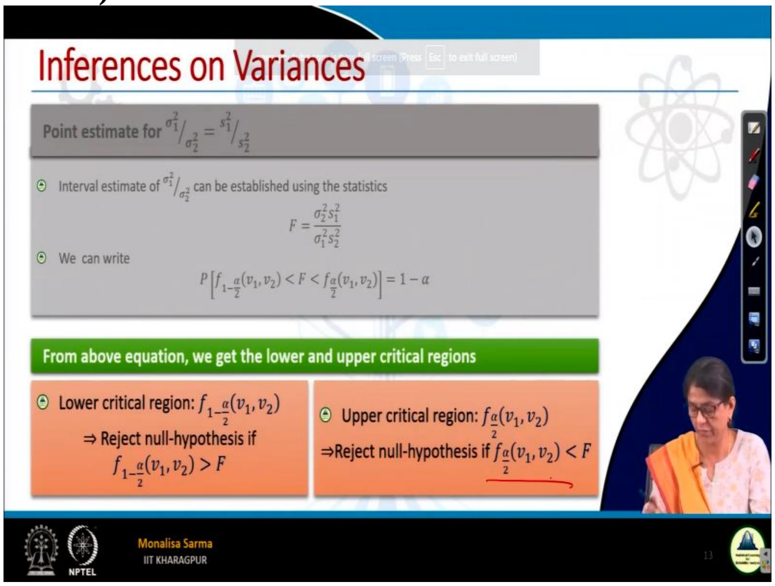
(Refer Slide Time: 11:47)



So, this is so, interval when I tried to find out the interval estimate, so, what is my corresponding point estimate, my corresponding point estimate is that this is my corresponding sorry my corresponding point estimate is s 1 ^{2} / s 2 ^{2} so, we will use this F statistics.

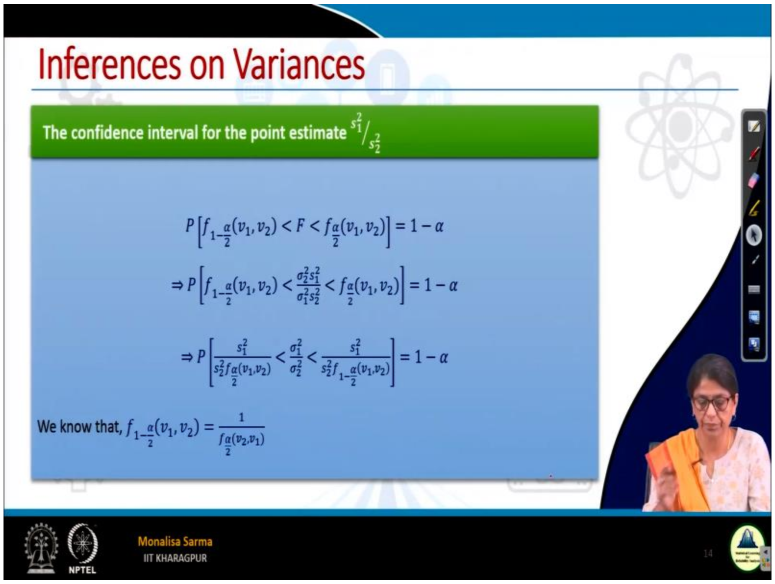
So, we can write if α is the significance level what is my confidence coefficient? Confidence coefficient is 1- α. So, to fall within my confidence interval F value is this value, this should fall within this range, it should fall within this means it should be less than this value and it should be greater than this value it should fall in this, this is my confidence interval, is not it? So, this is my confidence interval. So, now, it is a replace F by this value.

(Refer Slide Time: 12:53)



And then do some simplification, after simplifying we can find out what is the boundary for σ_1^2 / σ_2^2 simple simplification if we do simple simplification, we will see the lower critical region this is the of course, we have seen the lower critical region will reject the null hypothesis if it is less than this we have already seen, we will reject the null hypothesis it might add value is greater than this value we have already seen.

(Refer Slide Time: 13:18)

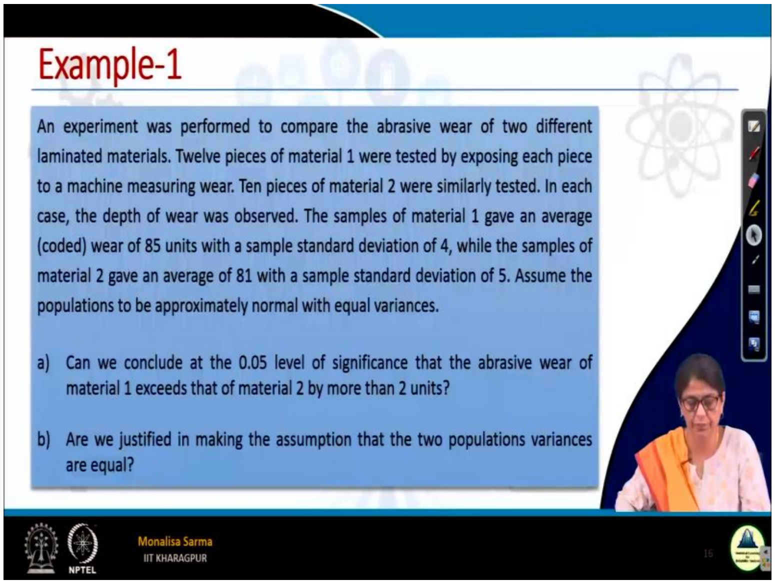


Now, to find out the significance level, so, we have simplified F with this value and we will do simple simplification and this is the confidence interval, this is the upper confidence interval, this is the lower confidence interval. Now, here in the upper confidence interval what I have in the denominator I have F of 1- α / 2 v 1 v 2 instead of F of 1- α / 2 v 1 v 2 cannot I replaced it because this value I will not get it from the F table I will need some value which will get it from the F table.

So, this value I can replace it F of 1- α v 1 v 2 I can replace it with this value is not it? Simple simplification what I have done as a simple simplifications dividing first step what I have done first step I have divided all the expression by s 1 ^{2} / 2 ^{2}. That is the first step I have done then second step my denominator was σ 2 ^{2} and numerator σ 1 ^{2}, I have made σ 1 ^{2} numerator σ 2 ^{2} denominator then accordingly my inequality also getting changed.

So, accordingly I got this value. Now, what is my upper confidence by sending this like this is the value I got. So, this is my upper confidence limit, this is my lower confidence limit.

(Refer Slide Time: 14:49)



So, now a simple example so, in the last class, remember we were trying to find out the average ^{2} of 2 different laminated material where we are trying to compare the means that the 2 different materials for which we are trying to find out the weight of we have taken from 2 different population we have taken a sample 2 different types of samples of size if I remember correctly 1 was of size then another was of size 12 something like that.

And then the variance of the population was not given. So, what we have we assumed? We have assumed the there is we have assumed it to be equal and that way we found out the pool variance estimator we found out S p ^{2}. Now in this question, but we will try to everything else remains same. Here what we try to do is that are we firstly let us read the question, an experiment was performed to compare the abrasive wear for the 2 different laminated materials, 12 piece of material 1 were tested by exposing each piece to a machine measuring wear fine.

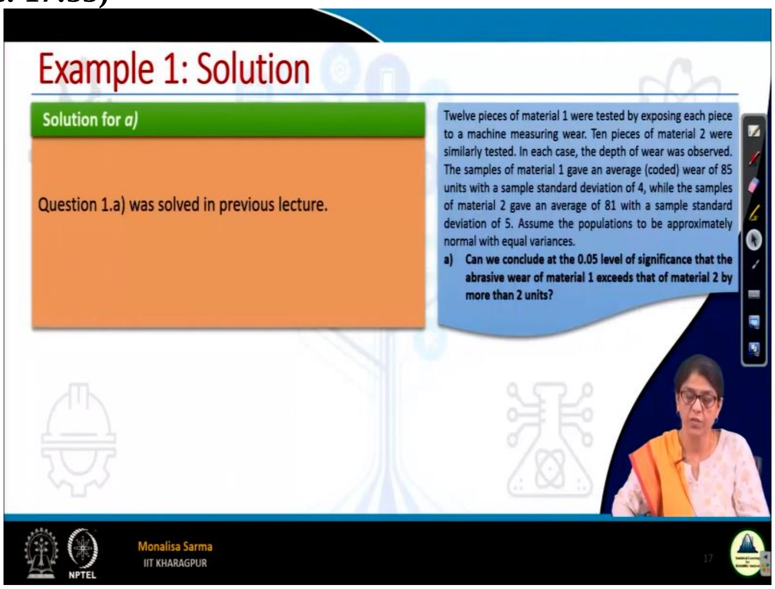
10 piece of material 2 were similarly tested. In each case, the depth of the wear was observed the sample of material 1 gave an average wear of 85 units with a sample standard of 4, standard deviation of 4 while the sample of material 2 it gives an average of 81 with a sample standard deviation of 5. So, n 1 is given n 2 is given for sample 1 size, standard deviation is given, mean is given, x bar is given, s 1 ^{2} is given, x 2 is given, x 2 ^{2} is given, n 1 is given, n 2 is given.

So, from that we found out the S p ^{2}, what is S p ^{2}? Pooled variance estimate that means, we try to find out the weighted variance of both the sample so, can we conclude that 0.05 level of significance that average wear of material 1 exceeds that of material 2 by more than 2 units. This problem we have already solved is not it? For this what was our hypothesis testing remember how hypothesis testing null hypothesis x 1 bar - x 2 bar = 2 is not it?

Sorry μ 1 bar- μ 2 bar = 2 and an alternate hypothesis is μ 1 bar - μ 2 bar is greater than 2 that was my alternate hypothesis and we have tested it and when we have found it that my null hypothesis is not rejected. So, in there we have assumed here it was then we have assumed that the population is has equal variances. Now, are we justified in making the assumption that the 2 population variances are equal.

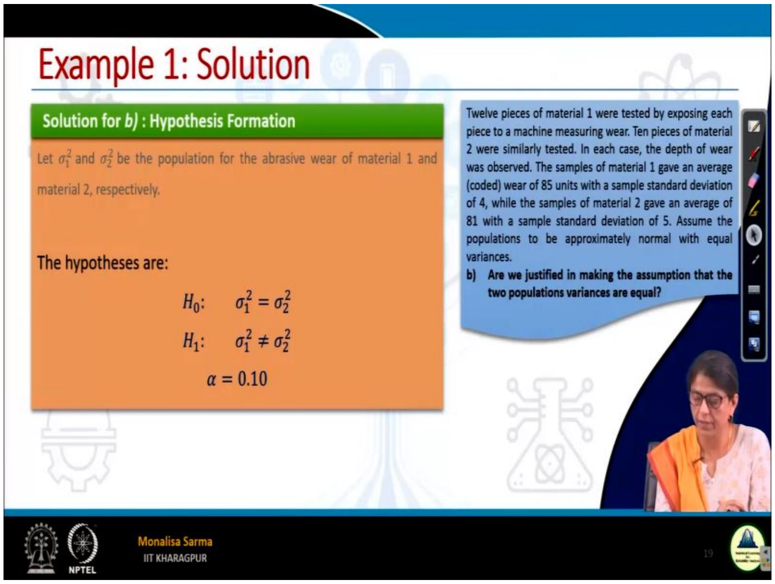
So, that is the same question first we have assumed it is equal actually we will do other way around before assuming first we will do this test whether the 2 population variances can be considered equal, first we will do this and then if it is can be considered equal then we will assume then we know it is equal then we will find out a pool variance estimate. First this is done then that is done now, in this first since we have introduced that so, first we have done the mean there we have done this assumption now, we are trying to prove that whether the 2 population variances are equal.

(Refer Slide Time: 17:55)



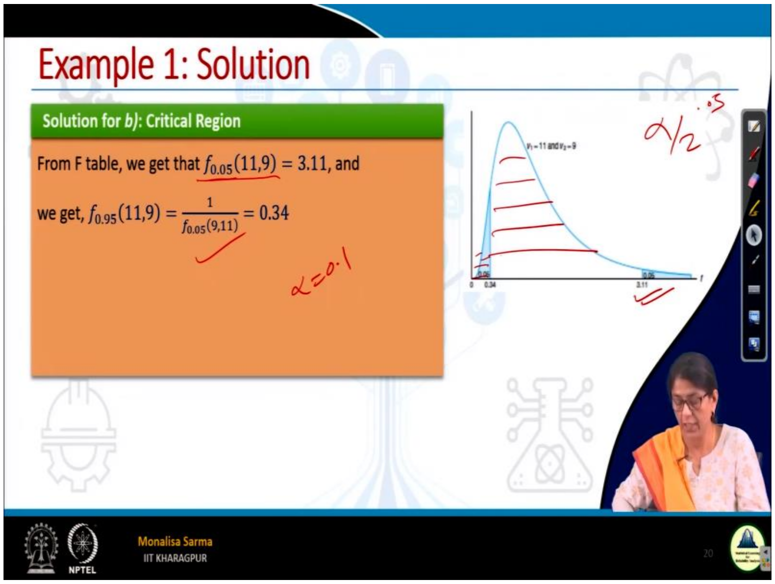
So, that is why I have taken given the same question here. So, question 1 was solved in my previous lecture.

(Refer Slide Time: 18:01)



So, here let σ 1 [2] σ 2 [2] be the population for the population variance for the abrasive wear of material 1 and material 2 respectively. The hypothesis are, this is my hypothesis that is the significance level is given in the question, can we conclude that 0.05 level of significance that abrasive wear material, significance level is here in this question significance level is not there and we have taken his α = 0.10.

(Refer Slide Time: 18:36)



So, there is a mistake anyway, I will correct it in the slides, but you will get it basically. So, here we are taking the significant levels 0.05. So, it is 0.05, α is 0.05. So, now, what we will do is that we will from the table we will try to find out from the F table we will try to find out what is the F of 0.05 corresponding to 11, 9. So, this is what we will do is that for 11 9 this is the value 3.11.

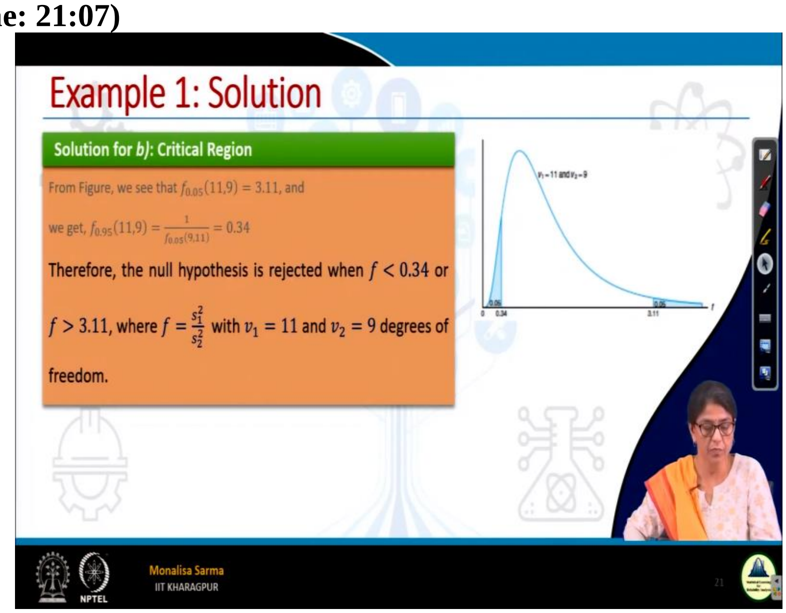
So, if this is 3.11 then how we can find out the corresponding value for this? What to say this value this critical region is nothing but 1 / 0.05 9 11 will give me 0.95 11 9.

So, it is 0.34, see here, here we have taken a significance level of because this is 2 tailed, 2 tailed means it is α / 2 is not it? So, if my α is 0.05, then my α should be 0.025 so, here I have considered 0.05 I am sorry for the mistake I will correct it in my slide this that means I have considered a significant level of α = 0.1 only this slide was right 0.1 but the question is wrong a equation I have put it is 0.1 I will correct it no, that means my significance level is 0.1.

So, it is to tail it is α / 2 these are minor things you can understand that is no, you already know this concept. So, it will be different, it will be easier for you to understand. So, when it is 2 tailed and we consider it α / 2, so, it is 0.1 my α / 2 is 0.05. So, I found out F of 0.05 11, 9 remember we took the some on the numerator, but we took numerator we always take the sample with the larger variance.

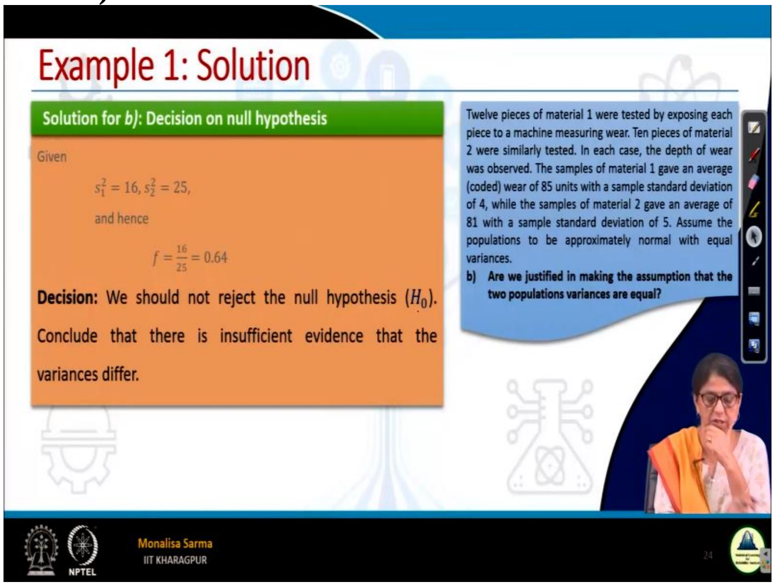
So, this we got it 3.11. So, once we get to 3.11 corresponding the value for what to say lower tailed, we can find out is in this formula and we got 0.34. Now, if my F value lies within this range, then my null hypothesis is accepted what is accepted that the both the population variances are equal.

(Refer Slide Time: 21:07)



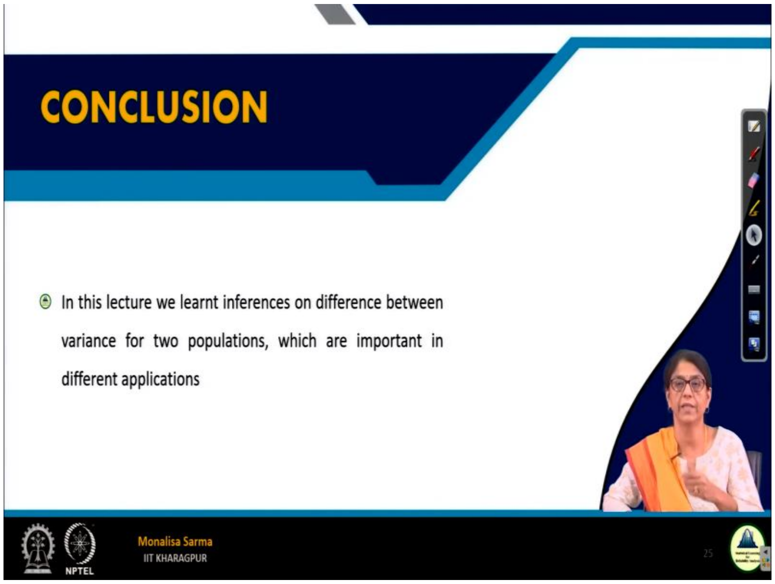
So, I calculated my F value. So, what is my F value here? F value is 0.64, null hypothesis will be rejected if it is less than 0.34 and if it is greater than 3.11 if it is greater than 3.11 or if it is less than 0.34 it will be rejected with v 1 = 11 and v 2 is also 9 degrees of freedom.

(Refer Slide Time: 21:36)



Now, for to calculate the F value I need to find out s 1 ^{2} and s 2 ^{2} from it is given here. So, I found my F value is 0.64 very well lies within this range. So, my null hypothesis is accepted what is my null hypothesis? That is both the variances are equal. So, we should not reject the null hypothesis H 0 conclude that there is insufficient evidence that the variances differ because it is asking are we justified in making the assumption that the 2 population variances are equal yes. We are justified because we are not being able to reject the null hypothesis.

(Refer Slide Time: 22:13)



So, in this lecture, we learn inference on difference between variance of 2 population we have seen what are the usefulness of it, mainly, I have specified 2 different cases where it is useful once for finding other inference and population means wherever variances are not known. In that case, we will have to see the variances are equal then I our life become cool, then we can directly use the t distribution.

And however, after finding out the variance in 2 population, we found that the variance are not equal, then we cannot use t distribution then, what we will have to use them first 1 way is that we will see if we can transform the data if transformation data is at all feasible, then we can transform the data and we can find out the variances are estimated if the variances are equal. Another way is that if the sample size is bigger that means, if it is more than 30 say then we can very well use the normal distribution.

If we can use the normal distribution, there is no issue at all of the degrees of freedom then we can very quickly use that. Then other is that if the value that if the parent population, if it is known to us, that the parent population is very closer to the normal population. Then we can also use the t distribution where we will use the degrees of freedom of the smaller variances as the smallest sample as the degrees of freedom. So, that is what we have seen.

(Refer Slide Time: 23:35)

So, these are the references and thank you guys. Thank you.