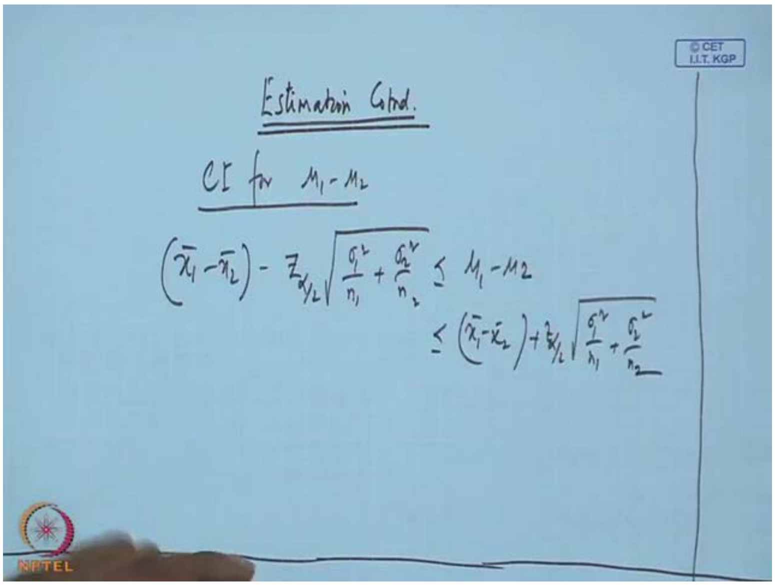
Applied Multivariate Statistical Modeling

Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

Lecture - 06
Estimation (Contd.)

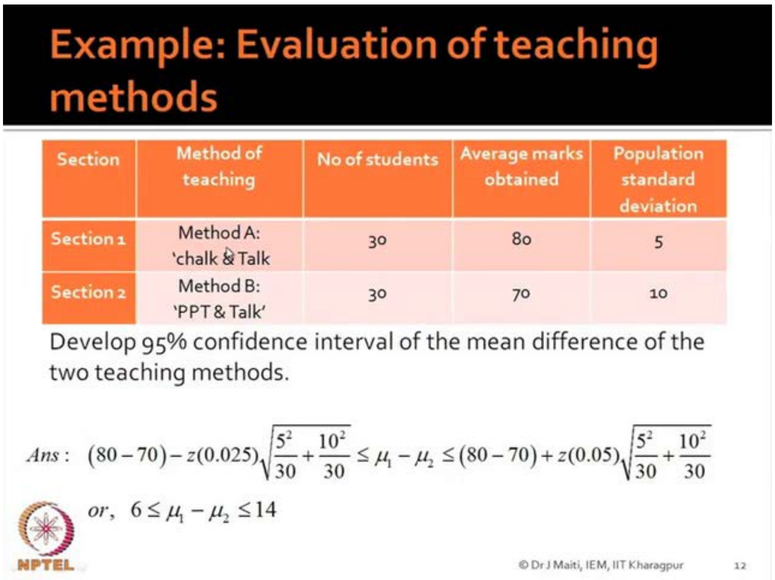
So, we will continue estimation last class what we have seen that we have seen the confidence interval for…

(Refer Slide Time: 00:28)



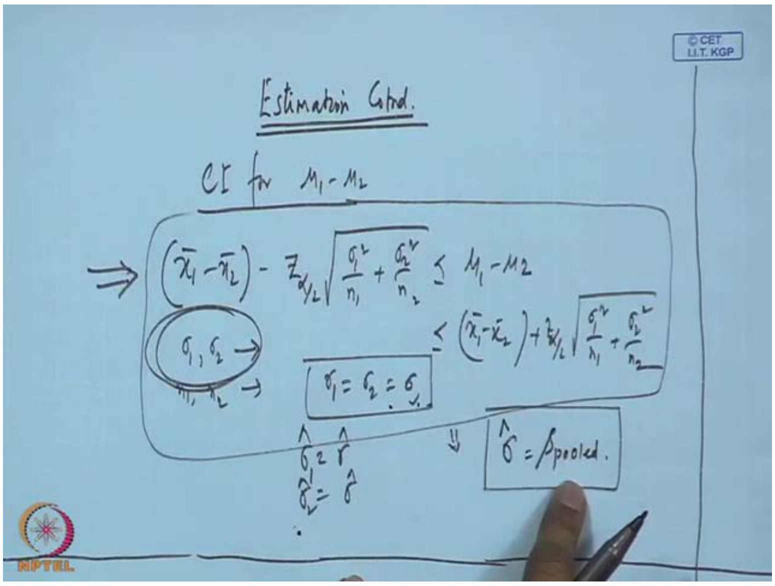
mu 1 minus mu 2, where mu 1 is the mean population 1 mu 2 is a mean population 2. And we have used this formula that x 1 bar minus x 2 bar minus z alpha by 2, then our square root of sigma 1 square by n 1 plus sigma 2 square by into less then equal to mu 1 minus mu 2 less than equal to x 1 bar minus x 2 bar plus z alpha by 2 square root of sigma 1 square by n 1 plus sigma 2 square by n 2, this is a case and we have seen one example also.

(Refer Slide Time: 01:21)



What I have that method a and method b that these are two type to teaching methods and this is a formulation only thing that z 0.025 you write, it this one this is not 0.05 this will be written as 0.025.

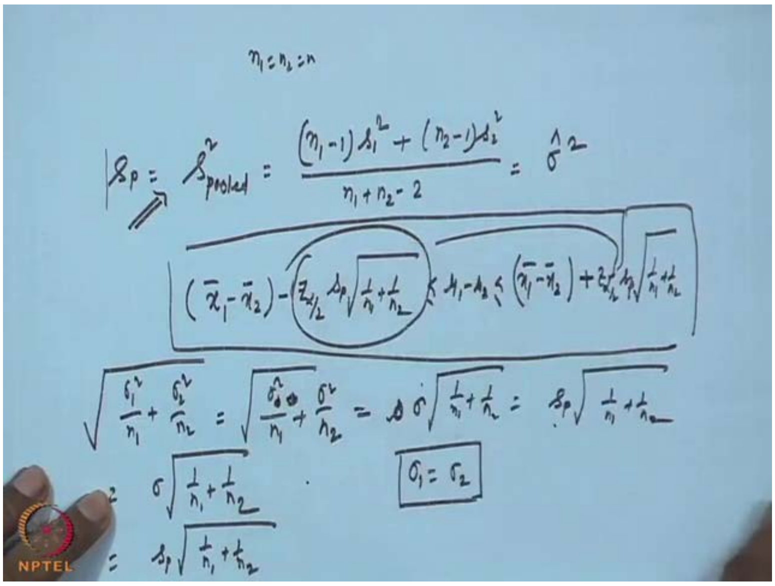
(Refer Slide Time: 01:41)



So, using this formula like this one, what are the prerequisite. So, the conditions, conditions are many first a fall sigma 1 and sigma 2 must be known and will be collecting in n 1 and n 2 in to this step sample size, and it is for normal populations and but there is difficulty arises here knowing that sigma 1, sigma 2. Another thing it may also happen that that sigma 1 equal to sigma 2 maybe that maybe the case sigma 1 equal to sigma 2 equal to sigma, what does it mean? The population rinses are same equal if population variance are same, then we will not use this equation we will go for because of this special condition, we will go for another type of estimation, why? We will use first we find out that what is the estimate of sigma.

The reason is as sigma 1 equal to sigma 2 equal to sigma we have n 1 n 2, two set sample are available from the different population. So, we will instead of using one that means I you can go for a sample 1 and calculate sigma estimate also you can that is basically sigma 1 estimate equal to sigma estimate because they are same, sigma 2 estimate equal to sigma estimate, but in that case what will happen that find out the sigma 1, sigma 2 will be differentially different because they are coming from a two population. So, n sample only that's why what happen you say that we will calculate a spooled is that basically that two sample will be spooled gather held then you will be calculating the estimate of that estimate of sigma cap, which is basically spooled. Now, what is this s spooled, s spooled your s spooled will be like this…

(Refer Slide Time: 04:20)



s square spooled you write this will be suppose the sample 1 of the size 1 minus 1 is 1 square plus population 2 k is n 2 minus 1 s 2 square divided by n 1 plus n 2 minus 2. Getting me? What you are doing here, what is this? n1 minus 1 s 1 square that is the total variability in the sample from population one. Second one is total variable in the sample from the population 2 point to mu and this n 1 minus n 2 minus 2, this is a degree of available.

The reason is the calculated x 1 bar 1 degree is the last x 2 bar is calculated under degree is lost, this is a resultant this is what we are saying that s spooled are variance combined variance. Now, this one we are saying that this is a estimate of sigma square, if this is a case and if we consider that things are coming from normal population and sample size is large in that case, what will be your interval x 1 bar minus x 2 bar minus z alpha by 2. Then the variance what you have written in the earlier case what we have written, we have written like this that sigma 1 square by n 1 plus sigma 2 square by n 2 then square root.

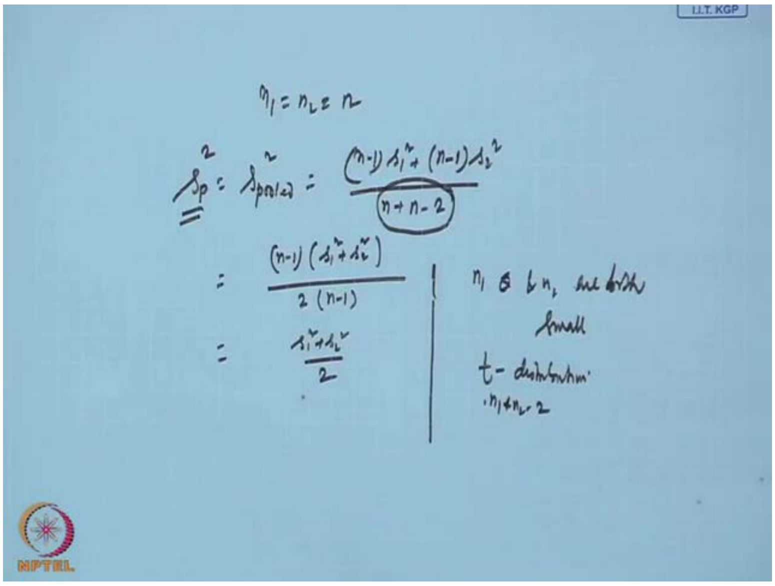
Now, what is happening here your sigma 1 equal to sigma 2 so I can write this one like this sigma 1 sigma square by n 1 plus sigma square by n 2 this is nothing but sigma will come out and 1 by n 1 plus 1 by n 2. Now, you see that s spooled is the estimate of the sigma. So, that mean I can write s p if I write s p equal to s spooled then s p square root of 1 by n 1 plus 1 by n 2.

So, you will be writing s p that this portion is now, I am writing again sigma square root of 1 by n 1 plus 1 by n 2 equal to s p 1 by n 1 plus 1 by n 2, why s p? S p is a estimate of sigma, then you result now your earlier that derivation like this x 1 bar minus x 2 bar minus z alpha by 2 sigma n square by n 1 plus less than equal to this less than this. Now, this quantity will be replaced now, it will be replaced by s p into this.

So, I will write it here now s p 1 by n 1 plus n by n 2 square root this less than equal to mu 1 minus mu 2 less than equal to x 1 bar minus x 2 bar, correct? Then minus you write not minus plus I think you can write the place of the portion plus this one, this will come here plus this. And most of the cases will be sigma 1 knowing sigma 1 sigma 2 is not possible population variance. And then what will happen you will be using this s 1 and s 2, but if s 1 and s 2 and the formulation like this s 1 square by n 1 plus s 1 square by n 2 that variable variability formula if you use, then the sample size is larger that is better.

But if the sample size is small there is doubt in this formulation. So, our one of the assumption will be that sigma 1 square equal to sigma 2 square sigma square. That means population variance standard deviations are equal that assumptions we will do, and because of these assumption that will be using a spool. So, this is a equation you find out the most of the time this equation is used z alpha by 2 s p square root of n 1 plus 1 by n 2. And it is also cost to many that will go for the equal sample size so n 1 plus equal to n 1 equal to n 2 equal to n also, if you use n 1 equal n 2 then what will happen ultimately, this is n minus 1, this 1 is also n minus 1 then it is 2 n minus 2 minus n minus, what will happen if n 1 equal n 1 equal to n 2 equal n.

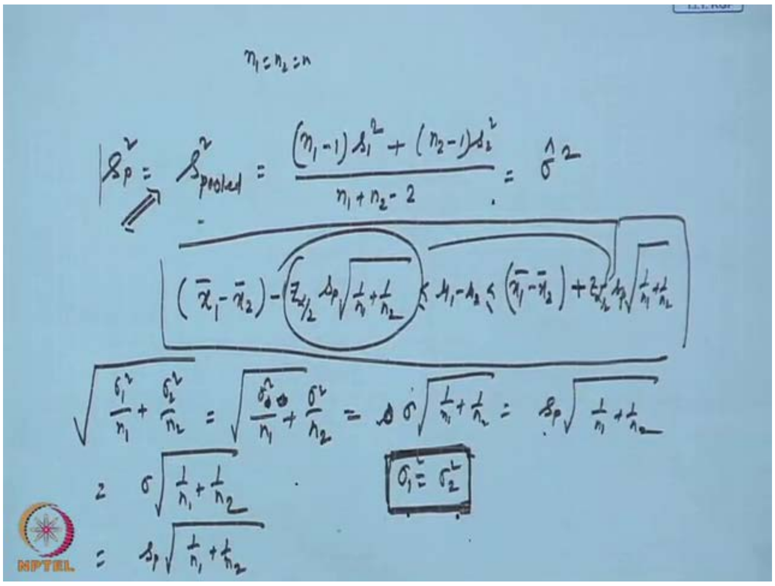
(Refer Slide Time: 10:35)



What will happen then you s p square equal to or s p spooled equal to n minus 1, s 1 square plus n minus 1 is 2 square by n minus n plus 2. So, this will become n minus 1 s 1 square plus s 2 square by 2 into n minus 1. So, resultant will become s 1 square plus s 2 square by 2, and in the same equation you use. And when another condition is suppose if I go to the I think next one is not there what will happen if n 1 less than an n into and n 1 into are both small, we will using t distribution, what will be the degree of the degree distribution when you use s spool minus 2 getting that is the that is a.

Ultimately you are calculating this an you have computed you have this much degree of freedom available. So, you have to go for n 1 plus n 2 minus 2 degree of freedom, then I ask when you use this formulation are this other formulation first one that is this formulation z, this formulation form the normal population it is coming sigma 1 sigma 2 is known, irrespective of this sample size sigma 1 sigma 2 is not known sample size is large, you have assumed this distribution. If sample size is small sigma 1 sigma 2 is not known.

(Refer Slide Time: 13:14)

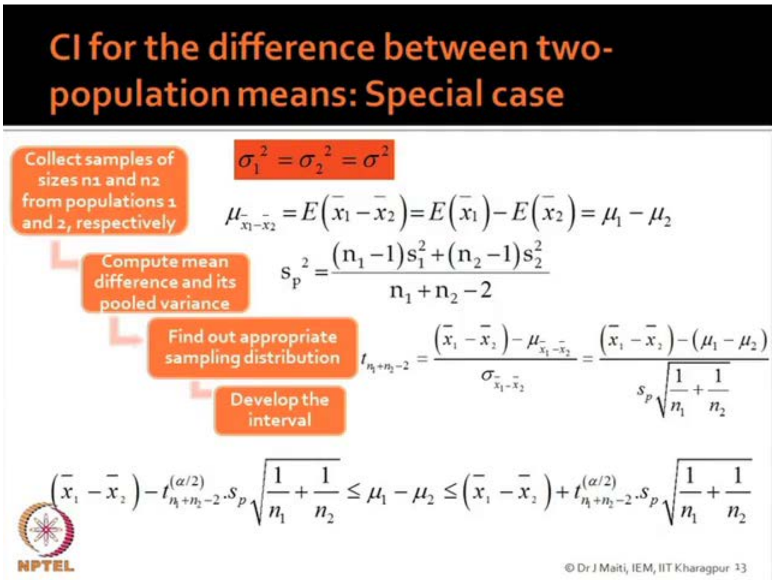


Then you have to first find out that the assumption is sigma 1 equal to sigma 2 that answer is found out, that assumption to be checked if that assumption is true, then you have to calculate s p square. And then you formulate like this provocation and go for t distribution getting me. If how do you know that sigma 1 square sigma 1 equal to sigma 2 or sigma 2 square 1 square equal to sigma 2 square, they are equality of variance.

How do you know that is the equal that is to be tested because if you want to use this formulation, we require to test also that the two population variances are equal. If two population are variances are not equal then you cannot go for this. Suppose population variances are not equal sample size are different you cannot and also population is not equal sample size are small also, and then you cannot use z distribution. And another issue will taken place that because of this differentiation, so ultimately there are some other type of derivation is that we are not considering here.

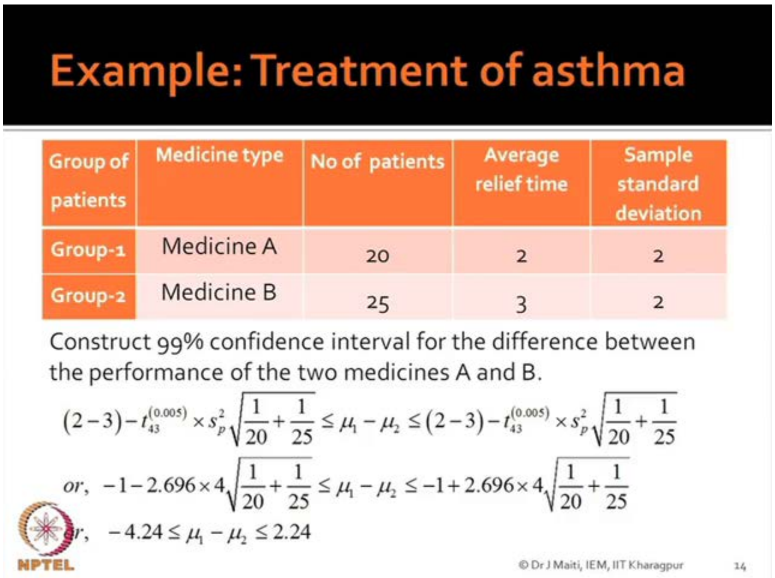
So, essentially we are considering things sampling from the normal population either the variance component are known. And either sample in variances are not known that is a large sample and when variances are not known also small sample, we are assuming that the variances are equal and then we are using a spool then we are going for a t distribution under this case. If they are equal and some sample size is large then you can go for z distribution no problem. So, j and t distribution for the equality of means of the two population all these things mostly we are used, but where are we ended now we ended that if we want to use this, we require to know that the population variance are equivalent.

(Refer Slide Time: 15:52)



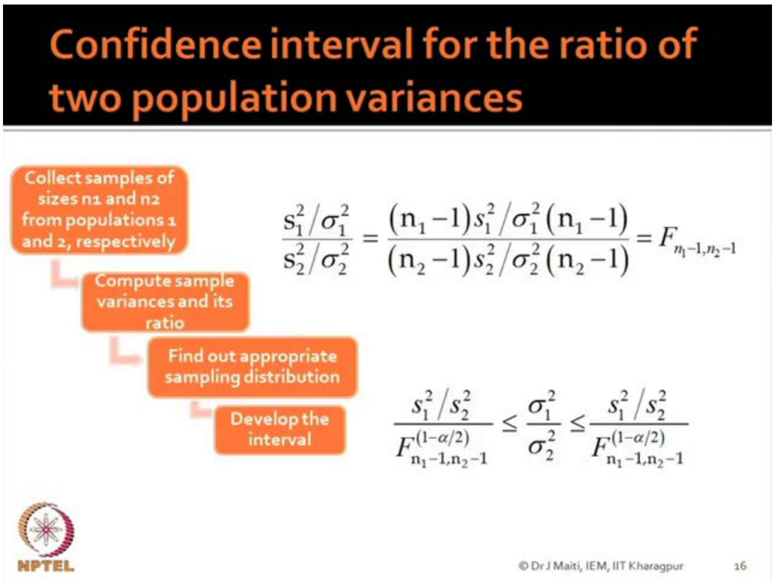
So, can we not find out the taste by which we can do that this is a this case.

(Refer Slide Time: 15:58)



Can we not find out the other one before going to the other one, I think you can simply find out that how we can calculate…

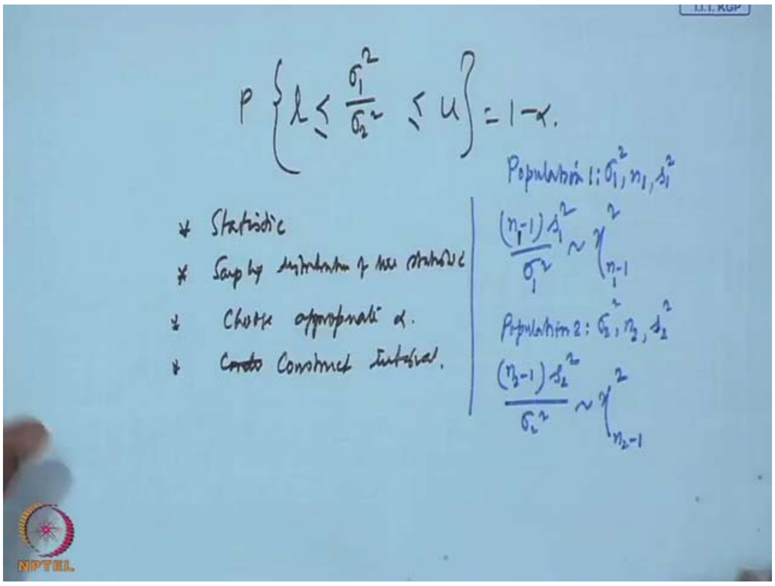
(Refer Slide Time: 16:09)



The construction interval always the we have discussed now confidence interval all the things we discussed, now confidence interval for the ratio of two population variances. How do you test the two population variances are equal in any static any test. So, eleven test is there in innova, innova analysis of inova that is this what is innova in innova what will happen will find out that one of the important test is there eleven test, we want to do that the population. The level variances is equal generally go for mean test, eleven test is one that we will see later on if time permits.

Now, confidence interval for the ratio of two population variances, how do you go around it, what is that say usual of that our, our total when your structure what is usual states we follow, we will follow the steps like this, I have two population both population either here you are interested to test the ratio the sigma 1 square by…

(Refer Slide Time: 17:45)



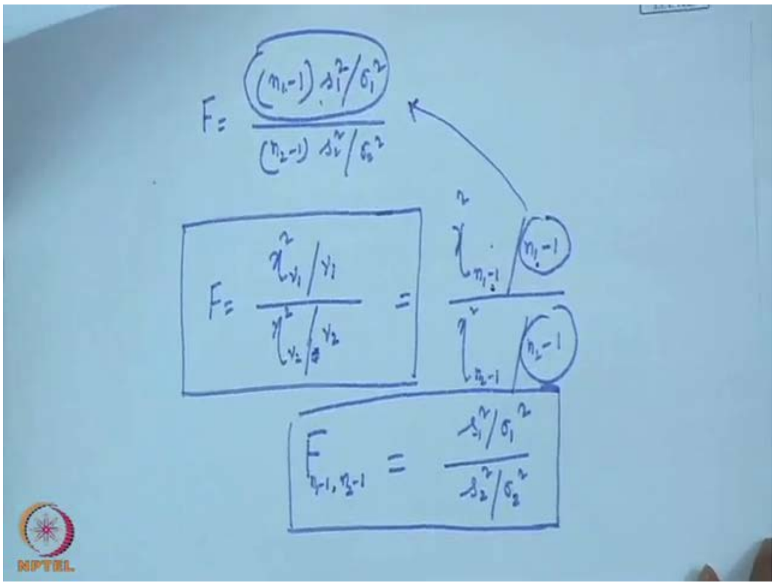
That you want to get something like this sigma 2 square you want to get something like this. Something like this what will be the l and u and you are interested to know this is the case. Now, if you want to get this interval you must know that, what is statistic you will generate, and that based on that statistic you will go for sampling distribution. So, the few things are important you must know the appropriate statistic, then distribution of sampling distribution of statistic.

And then you basically choose the alpha value appropriate alpha that what do you want then you construct the interval? These are these are the steps and every where you are doing like this only, if it is mean the statistic is mean minus expected value or mean by variance it will be z or t then if it is variance you are using n minus 1, s square by sigma square that is statistic square distribution.

Now, if it is mean difference again you are finding out the x 1 minus x 2 bar expected value by its standard deviation, that also follow z distribution depending on the conditions. Now, we are talking about ratio of two and I think you all know because we have discussed this that n minus s square by sigma square follows distribution with n minus 1 degree of freedom yes or no?

Now, let us see that it is our population one that sigma 1 is the variance, sigma 1 square is the variance you collected simple n 1 and you have computed also s 1 square, then your statics here is n 1 minus 1, s 1 square by sigma 1 square will follow chi square distribution with n 1 minus 1 degree of freedom, yes or no? You go for population 2. That the variation is sigma 2 square you collected n 2 simple, size is basically n 2 then you have calculated variance simple variance. Can I not say that n 2 minus 1 is 2 square by sigma 2 square, this is chi square distributed with n 2 minus 1 degrees of freedom, we can say, if we can say this we also can say that…

(Refer Slide Time: 21:37)

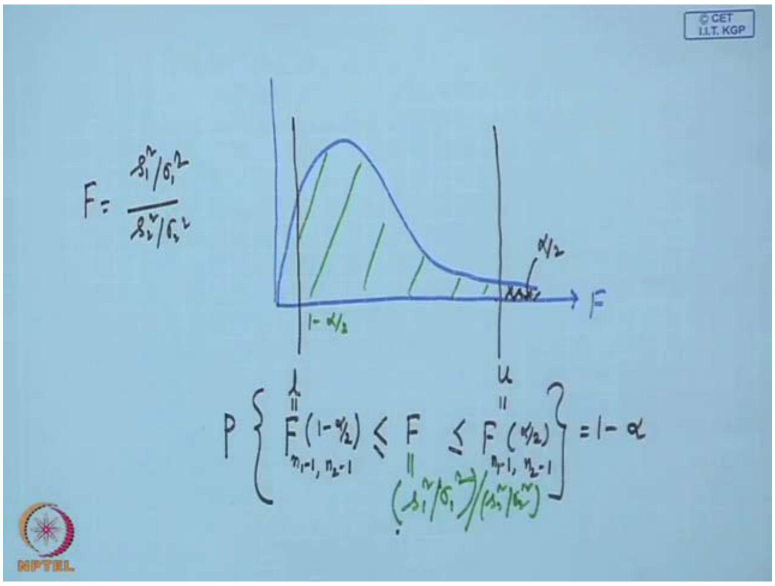


F equal to n 1 minus 1 s 1 square by sigma 1 square divided by n 2 minus 1 s 2 square by sigma 2 square. We can clear like this, but what we say the earlier that that if these quantity is divided, what do you say the earlier for a F distribution, F distribution we say F is the ratio of 2 chi square variable, chi square. Suppose nu 1 degrees of freedom divided by nu 1 followed by the denominator case chi square in nu two degrees of freedom divided by nu two.

If this is the case then we will cleared like this what is your chi square variable here chi square instead of write these I am writing now chi square, you please write see that new 1 is n 1 minus 1 divided by n 1 minus 1 then chi square n 2 minus 1 divided by n 2 minus 1. Can you find out what will be these from these? We say this one is chi square n minus 1. So, these divided by n 1 minus 1 means, this is s 1 square by sigma 1 square. Now, but denominator will be s 2 square by sigma 2 square, so this will this definitely follow chi square F distribution. That is the F what to will be the numerated degrees of freedom follow from here what you have divided here n 1 minus 1.

So, n 1 minus 1 now what is the degree of n that this one n 2 minus 1, so you have tell like the n 2 minus 1 so this is the distribution, getting me? So you know the distribution, you very well know s 1 square by sigma 1 square divided s 2 square by sigma 2 square follow F distribution with n 1 minus 1 and n 2 minus 1, that numerator and denominator degrees of freedom. Can you now develop? You cannot develop you know that so what is our what is our statistic your statistics is s 1 square by sigma 1 square s 2 square by sigma 2 square that is a divided by this, this follows F distribution.

(Refer Slide Time: 25:03)



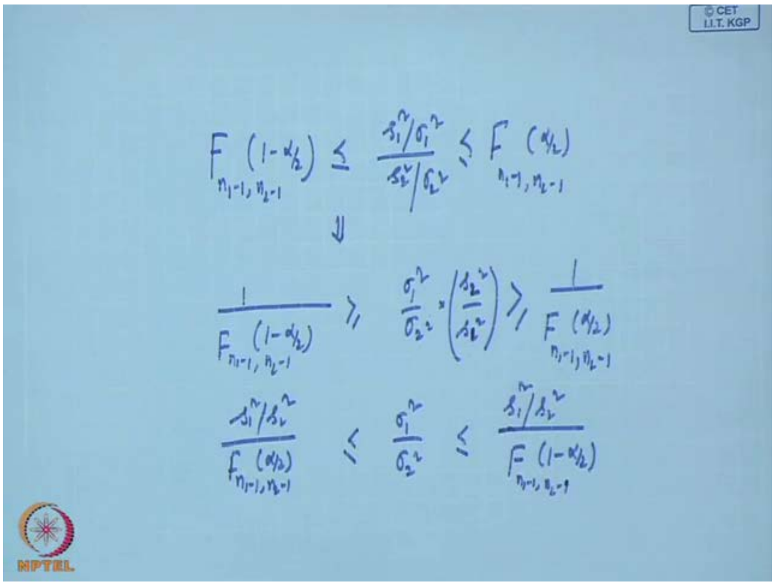
So, here F distribution is like this F you are saying that this one follows F distribution. What do we want to create? We want to create the interval you create like this earlier also you have seen this, this is my u and let this is my l. So, this one is if I consider alpha then this side only this will be alpha by 2 this side wall it will be alpha by 2, but the right hand side probability if I consider the integrity here, this total probability here this is 1 minus alpha by 2. So, 1 minus alpha by 2 will be this probability then corresponding f values you will be getting, what will be the F value here?

F is our n 1 minus 1 n 2 minus 1 that is the degrees of freedom numerator and denominator degrees of freedom, but this corresponding 1 minus alpha that probability value that probability value for that probability value, what is the value of this. Then that means l equal to this what is your u f n 1 minus 1 n 2 minus 1, this is alpha by 2. Now, again you know that F equal to s 1 square by sigma 1 square s 2 square by sigma 2 square.

Also we say that probability that this quantity less than equal to F less than equal to this, this equal to 1 minus alpha, it is now instead of s what we required to prove it you write down this one s 1 square by sigma 1 square divided by whole divided by s 2 square by sigma 2 square. So, I want what you want ultimately from here, you want a expression where in the middle portion there will be sigma 1 square by sigma 1 square by sigma 2 square, then shake the left and right hand side. Can you recap now, you take 2 minutes take 2 minutes and do it.

Essentially why I am putting so much of effort here because you see the all the cases, procedure remains same you have to know the statistics after the distribution, lower bound and upper bound that based on alpha value and probability of this will be this then your interval is same. Now, not necessarily that this quantity suppose some other quantity you are deriving you are able to approve, find out the proof in statistics as well as its distribution. You can create the interval getting me it is very simple. So, what we have done then we found out that…

(Refer Slide Time: 28:58)



F n 1 minus 1 into minus 1, 1 minus alpha by 2 less than equal to s 1 square by sigma 1 square s 2 square by sigma 2 square less than equal to F n 1 minus 1 n 2 minus 1 alpha by 2. You want sigma 1 by sigma 2 that whole square. So, you have to what you are required to do now? You require first take the inverse in the sense that a reciprocal, if you do the reciprocal part what will happen that this will be greater than greater than.

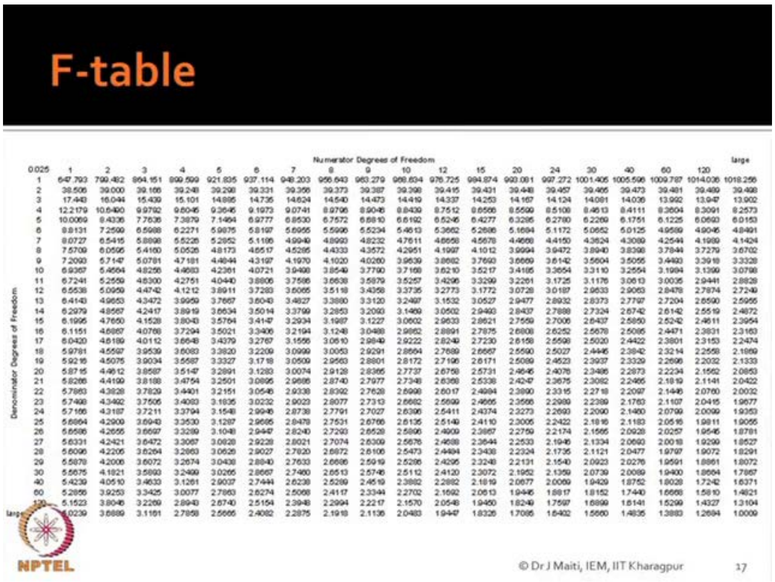
So, if I write like this ultimately what will happen, the resultant part if you write like this can you not write like this then I am writing, it is going up sigma 1 square by sigma 2 square into what is after remaining here, that is s 1 square by s 2 square that is basically this one I want to give a inverse sign, so that is why so s 2 square is there it will go up s 1 square will come down here sigma 1 square, where in that denominator its goes up sigma 2 square come down then this is 1 by f n 1 minus 1 n 2 minus 1 alpha by 2.

Now, I am writing just in other way, so what we are doing we are again going to the standard format the way we write sigma 1 square by sigma 2 square, I am writing like this so what will happen this side this side portion will come here and s 1 square if I take out of this to this side as well as this side, then what will happen here s 1 square by s 2 square divided by f n 1 minus 1 n 2 minus 1 1 minus alpha by 2.

And this side will be s 1 square by s 2 square by f n 1 minus 1, n 2 minus 1 alpha by 2. That is what is now see sigma 1 and sigma 2, which one is greater so there are again they are change combination changes are there and if you go to statistical any basics statistical book, you may find out some other type of derivation, presentation not derivation same way you derive.

You depending on that that possibility they are because we will not go into that there because it is not required for all of us we have to proceed to this are all prerequisites for multi variant statistical modeling. So, we have to finish as early as possible I am planning to finish by another lecture, the prerequisite so that will straightly go to the start the multi variant one that what is the ultimate aim for, when you are looking for if this is the case.

(Refer Slide Time: 32:40)



So, can you not find out some suppose this is the one of the example say two medicine that is treatment of asthma, there are two different type medicine may be you are taking some medicine or you are using inhaler. Or let it be that two different type of medicinal on only you are consuming, when they have the manufacturers, manufacturer is claiming that may be medicine both medicine is working as usual or not, but they are saying that they are competitive, competitive in the sense. Then when we talk of comparative then one of the issue is mean value, you have to find out the response variable we are saying that response variable is that once I take the medicine, how long it will take the up may to get relief from the asthmatic problem? That is we are saying that average is relief time.

So, if I take medicine a and this are all emergency medicine in the sense that when the asthma is problem will occur that time you are taking for relief purpose, let it be like this then if you take medicine a then what we have found out that average relief time is 2 hours. If I take medicine b average relief time is 3 hours and we have checked with 20 patient for medicine a and here 25 patient using medicine b, you got this average time and also you got the sample standard deviation both are two and two from that 20 percent from that 25 percent from medicine a and medicine b.

Now, you may be interested to see that using this information you may be interested to the confidence interval for the variability part. It may so happen why variability is important? It may so happen that the same medicine for the same medicine one patient that let mister x is getting cure relief by 1 hour, for mister y it may be take 4 hours so that variability because as we have already discussed variability is very, very important issue. So, I we want to using this we want to test that what is the that interval, it will vary. So, you can very easily calculate then you can may not calculate definitely you will be able to calculate because s 1 square and s 2 square are what are those things s 1 in this example what is our s 1 square 4 that is s 1.