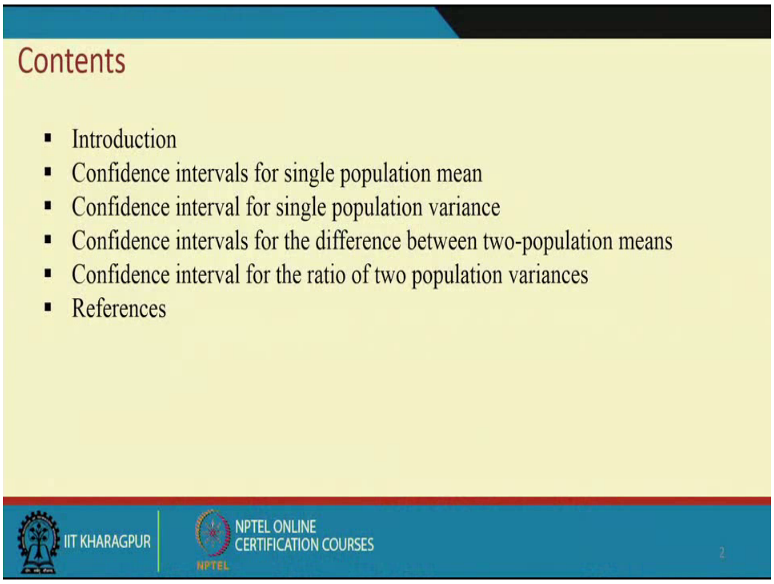
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

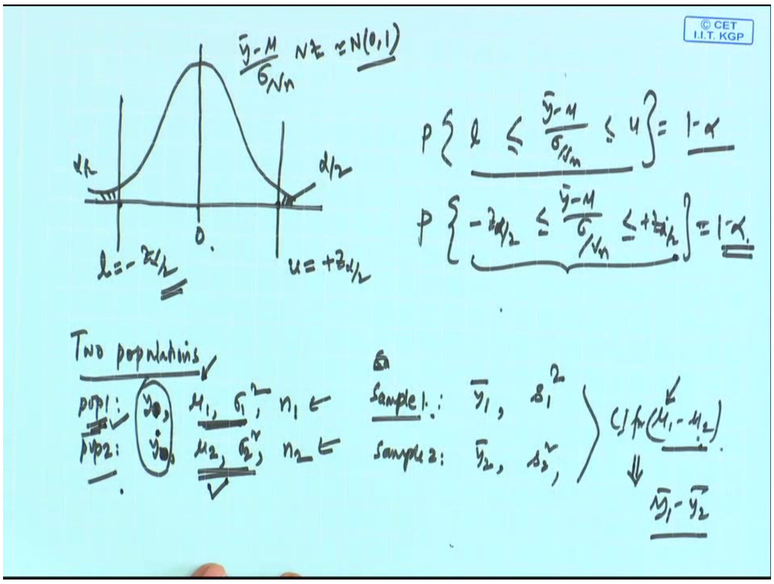
Lecture – 10
Estimation (Contd.)

(Refer Slide Time: 00:21)



Welcome, I will continue estimation. In last class we could cover the confidence interval of single population mean, and confidence interval for single population variance. Confidence interval for the difference between two population means, and confidence interval for the ratio of two population means, or a ratio of two population variances. This is in not covered, and now within to diminish of time, we will try to cover this two concepts that confidence interval for the difference between two population means and ratio of two population variances.

(Refer Slide Time: 01:09)



So, if we quickly recapitulate what we have discussed from the confidence interval point of view with reference to a single population means. I said that the quantity y bar minus mu by sigma by root n. This follows z distribution. So, where it will be z and t, those kinds of discussions I have already made.

You see the first few slides of my previous lecture you will know, and we say that you will create a interval which is known as low and low limit and high limit with reference to, let us say the normal distribution this will be plus z alpha by 2, and this will be minus z alpha by 2, where alpha by 2 is the area under the curve right to this point upper point limit. Similarly this one is also alpha by 2 which is left to the lower limit, and there is the mean of this quantity is, as it is normally distributed unit normal and 0 1. So, this value is 0. Our sole purpose is basically to find out what is the interval of the different destination.

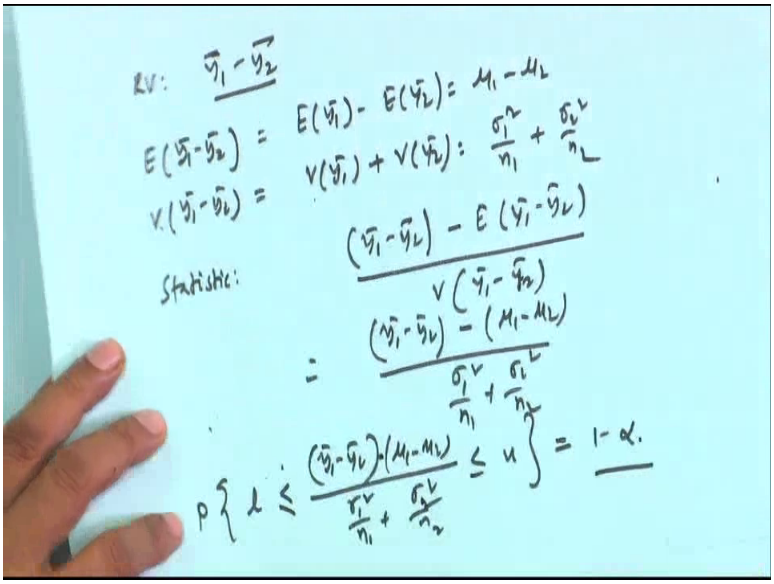
So, accordingly we have created the probability low, less than equal to y bar minus mu by sigma by root n less than equal to u, and this is 1 by alpha, and accordingly this l and for z distribution l and u are replaced by minus z alpha by 2 less than equal to y bar minus mu sigma by root n 10, and plus z alpha by 2. This is 1 by 1 1 minus alpha, and our inter confident interval is this. This we have computed, and we got the interval for mu.

No longer we have one population will have two different populations. So, two population. So, two populations case. Population 1; suppose the variable of interest is y 1 and its mean is mu 1 and variance is sigma 1 square and you have collected n 1 data points. Population 2; variable of interest y 2 mean is mu 2 and you have this one sigma 2 square variance, and you collect a data n 2. So, what we want, basically there is the, that slight difference is there.

A variable of interest is same y from the two population point, point of view, but as population 1 is having certain distribution one with these, and population 2 with these fine. So, you collect a two different sample of size n 1 and n 2, and this samples are independent samples. Now you, from the sample, we are denoting that sample 1 from population 1, and sample 2 from population 2. So, sample 1 point estimation is y 1 bar, sample 2 point estimation is y 2 bar from mean point of view. From standard deviation point of view, a variance point of view s 1 square and s 2 square. So, s 1 square and s 2 square fine.

So, what is our objective here? We want to find the confidence interval for mu 1 minus mu 2 C I for mu 1 minus mu 2. So, should what is mu 1. Mu 1 is the population mean for the variable of interest y and mu 2 a from population 1 mu 2 is for population 2. So, what if, what is the, then say sample equivalence definitely y 1 bar minus y 2 bar. So, you collect a sample from first population, and another one from second population. Compute these, you were getting these. You repeat this process; you will be getting several such y 1 bar minus y 2 bar values.

(Refer Slide Time: 05:54)

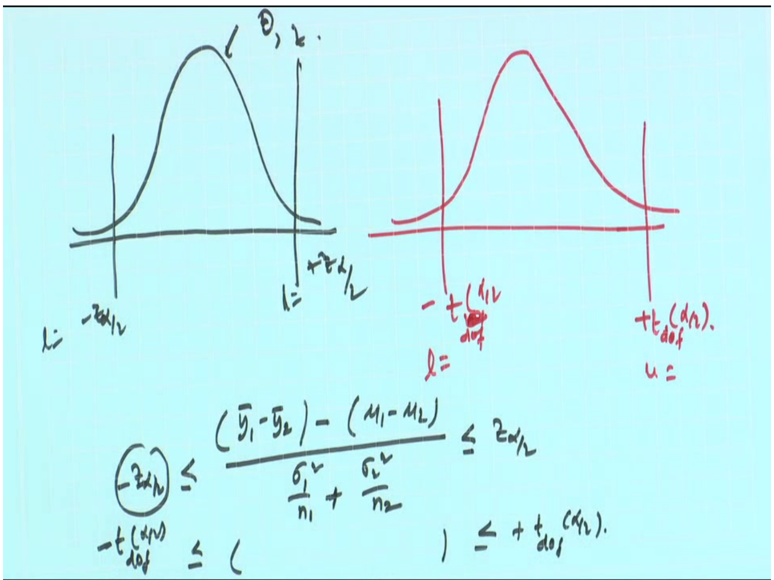


So, y 1 bar y 2 bar is the random variable of interest y 1 bar minus y 2 bar. So, what do you require? Now you require to know what is the expected value of y 1 bar minus y 2 bar. This will be expected value of y 1 bar minus expected value of y 2 bar, and this will be mu 1 minus mu 2

You know this earlier, we have discussed. Similarly variance of y 1 bar minus y 2 bar is important. This will be variance of y 1 bar plus variance of y 2 bar, and why this. You know I have explained earlier, and there is no covariance for, because while the two samples are collected from two different population, and they are independent in nature. So, this will be sigma 1 square by n 1 plus sigma 2 square by n 2. Now we will create a statistic here, that statistics that is y 1 bar y 2 bar that is the random variable minus its expected value y 1 bar minus y 2 bar by its variance y 1 bar minus y 2 bar.

This is nothing, but y 1 bar minus y 2 bar minus mu 1 minus mu 2 divided by sigma 1 square by n 1 plus sigma 2 square by n 2. So, this is the statistics. So, what will be its distribution, if sigma 1 and sigma 2 known, and coming from the normal population. Then we can assume that this will be z distribution unit normal distribution. So, in that case, we can write this, that probability l. We were interested to know probability l is less than equal to y 1 bar minus y 2 bar minus mu 1 minus mu 2 divided by sigma 1 square by n 1 plus sigma 2 square by n 2 u equal to 1 minus alpha that is inter of interest. So, we want this. So, essentially what you require. You require l and u.

(Refer Slide Time: 08:21)

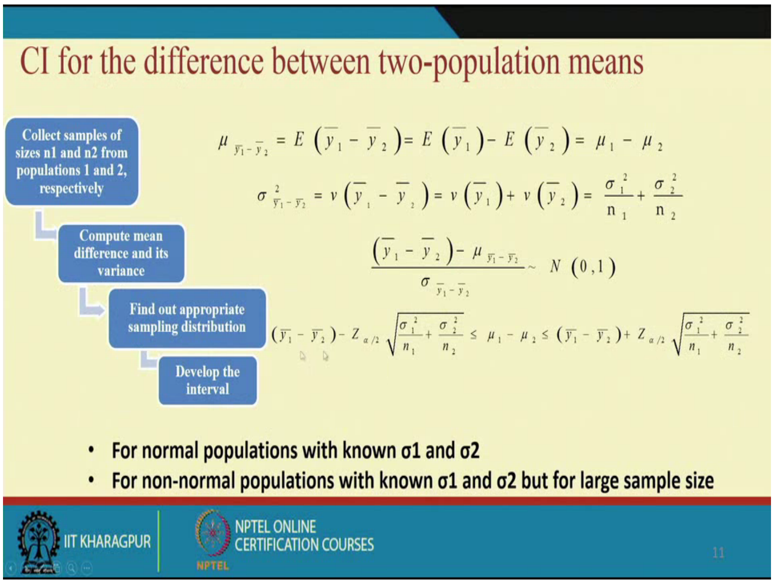


So, if you see the last class; first, second or third slides, where we have given that, when you go for t n, you dist z distribution.

So, depending on this suppose it can be t, it can be z distributed. If it is suppose, if it is z distributed, then already I have explained this minus z alpha by 2 plus z alpha by 2. If it is t distribution, then this will be minus t n minus t its degrees of freedom. Suppose d I write, let me write alpha by 2 and this side mi plus t d is d d o f degree of freedom into alpha by 2. So, this is l here, this is u here, and for z this one is l and this one is u.

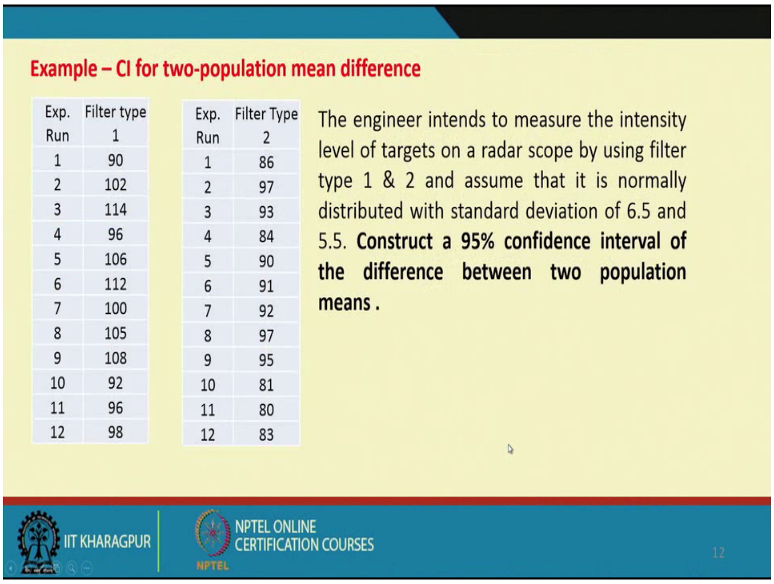
So, you, what do you do. Confidence interval, when you try to find out y 1 bar minus y 2 bar minus mu 1 minus mu 2 divided by sigma 1 square by n 1 plus sigma 2 square by n 2, then this less than equal to you write minus z alpha by 2 plus z alpha by 2 or you write t minus t alpha by 2 d o f less than equal to this quantity less than equal to plus t d o f alpha by 2. So, we have to understand that what is the, what is your. I can say that situation, whether this t or z situation accordingly. So, ultimately what happened, you just see the some other slide. The concept is given to you. Now, what happened? You see that if it is z distribution, this is the equation.

(Refer Slide Time: 10:25)



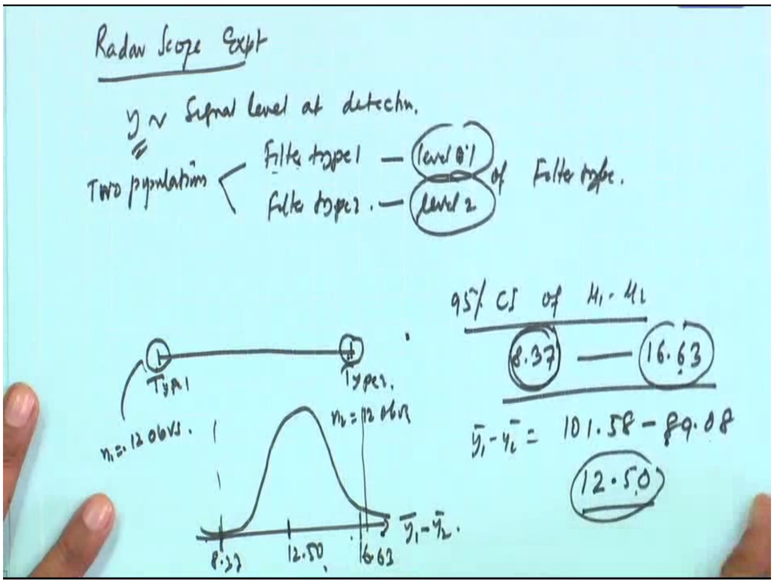
And from these we can very easily, can come to this. This equation y 1 bar minus y 2 bar minus z alpha by 2 into this variance, but and there this will be this, for normal population with known sigma 1, sigma 2. For non normal population with known sigma 1 sigma 2, but for large sample size, the situation when it will be is, we use this formula for computation of confidence interval for difference between two population. Means it is that for all normal population, when sigma 1 and sigma 2 the variances are known for all other population, when variance sigma 1 square sigma 2 square or standard deviation is, are known on large sample size, you have taken into consideration ok.

(Refer Slide Time: 11:20)



So, what will happen, that is one example is given here. So, the example is, we have chosen two filter; filter type 1 and filter type 2. A filter is a factor, it is two levels; one is filter type 1, and type 2, and you have run the experiment 12 times using filter type 1, and 12 times using filter type 2; obviously, randomized designed. It is not that the insequence only with filter type 1. So, then what is required to know. We are interested to know what is the 95 percent confident interval for the difference between two population? Means; that mean the response time; see that, that is label and detection signal at level at detection, whether they are different, if we use different filter types.

(Refer Slide Time: 12:17)

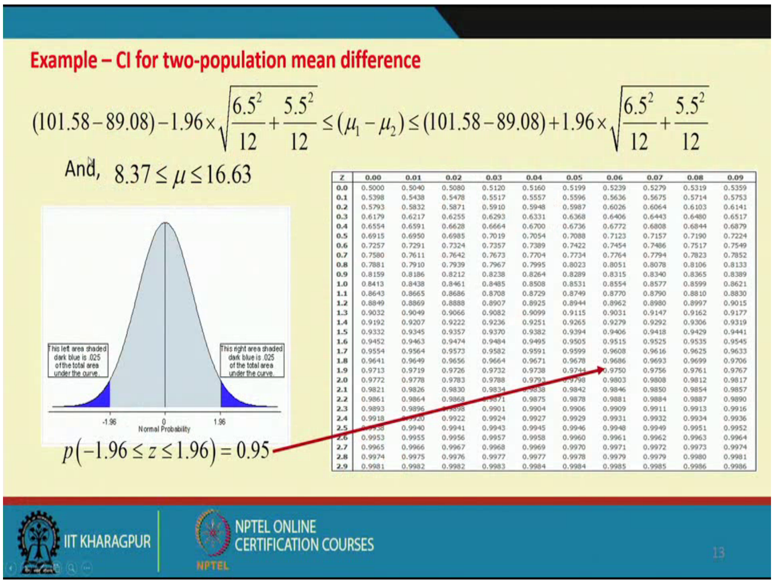


So, essentially what I mean to say, you have conducted experiment; that is the radar scope experiment. So, your y variable of interest is signal level at detection. So, here your two populations, are two populations are two levels of that filter type; one is filter type 1 is one first population. Filter type 2 is second population. Please remember this is nothing, but the two levels of these are the two levels; level 1 and level 2. Level 1 and level 2 of filter type.

When I explained the experiment, and in first few lectures this explained. They are two levels and that is way. So, that will factor levels are treated as population. So, other way I can say that you are doing the experiment with type 1, type 1 filter and type 2 filter. This is level 1, this you, when you are doing experiment, whatever results you are getting; that is from we are seeing that from population 1. Whatever results we are getting that coming from population 2.

So; that means, this 12 observations here. So, n 1 equal to 12 observations again n 2 equal to 12 observations here you have collected in your sole interest to see that whether the mean y y mean will differ for the two or not important, because when you have a factor, a factor has two levels you can do this. So, now then what we have computed.

(Refer Slide Time: 14:13)



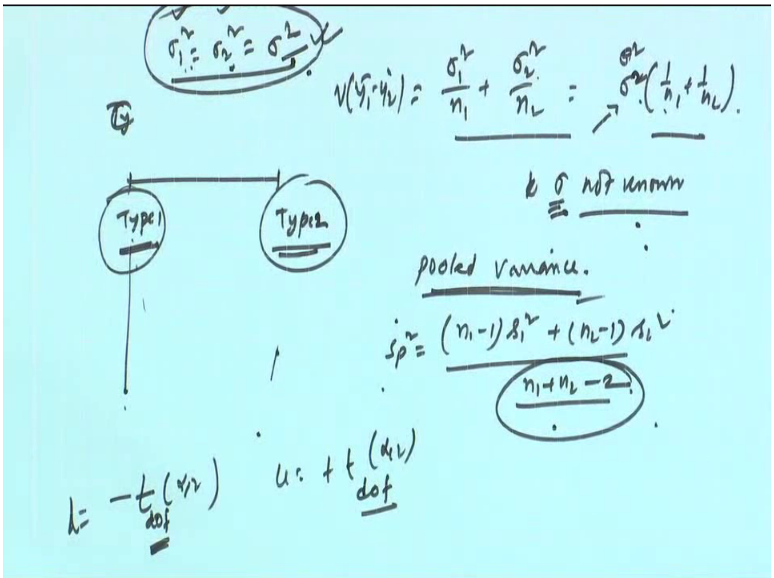
We have computed mean y for population 1 that is levta filter type 1 and filter type 2, and the variance 6.5 and 5.5, is it given here 6.5 and 5.5 standard deviation which is the square root of variance is given ok.

N 1 equal to n 2 equal to 12 and we have used mean, this random variable minus z into the variance component a standard deviation and then this plus z and standard deviation part. So, the result end quantity 8.37, this will be this, this mu. In the sense it is basically we try to write theta; that is mu 1 minus mu 2. So, 8.37 to 16.63; that means, the confidence interval 95 percent confidence interval of mu 1 minus mu 2 in this case, is 8.37 to 16.63 .

So, what is your interpretation? Interpretation is that, what is the y, what is the point estimate y 1 bar minus y 2 bar 101.58 minus 89.08. So, this value is 0 then 5. Then this is 92891. So, 12.50, if you go by 12.50, then if I write this side that this my y 1 bar, this the difference between two population means. So, y 1 bar minus y 2 bar. This is the axis point estimates somewhere here; 12.50 and interval estimate will be 8.37 to 16.63. This is 16.63, and this is 8.37.

So, 95 percent of the data will fall under these. So, this is my distribution, and 95 percent of the data points will be within this range. So, when you take decision, you also should consider that it can be as low as this, as high as this, but average value is this. This is the decision. Now go for the second one. Confidence interval between two population mean the special case, very important one. Most of the time we assume that the two population variances are equal that means sigma 1 square equal to sigma 2 square equal to sigma square.

(Refer Slide Time: 17:22)



So, what I have say, we are saying that whether you use type 1 filter or type 2 filter. Type 2 filter in the radar scope experiment the mean value will differ, but the standard deviation will not differ. So, it may be, it may not be a correct assumption for this case, but many a time we will consider this and it also true.

So, when you know; that means sigma, your things are becoming little better, because you, when you calculate the population variance for the mean from filter 1 or filter 2, you take the data point; that is the one difference you take these two. So, now, what happen for the time being, you think like this that the, this is, this is this is equal sigma square. Later this other concept would, I suppose to tell you, i will not tell now, i will tell later on, because it require something more.

So, under such situation what you will do. You will use that the variance part is like this; that means, variance of y 1 bar minus y 2 bar is this. So, then it will be what sigma 1 square sigma square, sigma square into 1 by n 1 plus 1 by n 2 correct. So, suppose a, suppose what happens that you do not know the sigma square. This sigma square is not known, sigma not known. If sigma is known, sample size is large or sigma known, and coming from normal population, the z distribution fantastic, no problem.

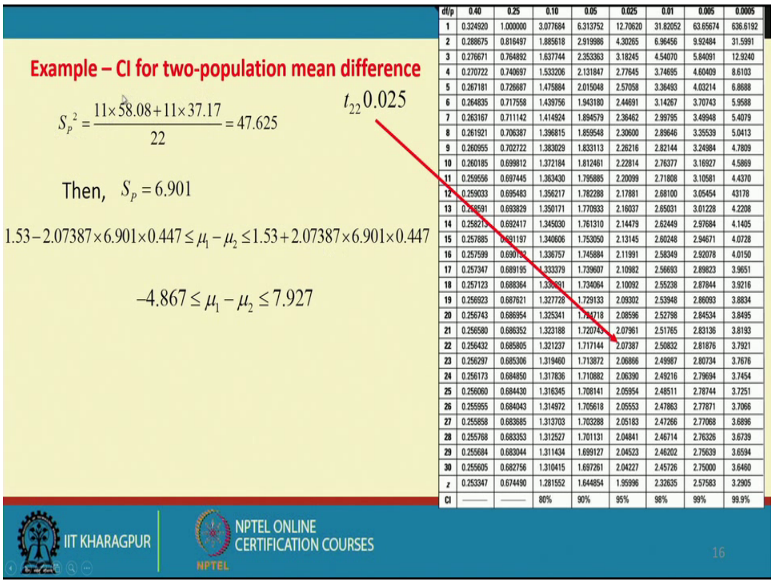
Suppose sigma equal part not known, under this situation you can find out the sigma. Estimate the sigma not from its one population or from only data, from type 1 filter or data for type 2 filter, whether you mix them and get pooled variance; that is what want to tell you the pooled variance is very important, because if sigma 1 square equal to sigma 2 square sigma square, and then you may expect that their sample variance will be equal, but that will never happen.

So, it may happen it will bit. So, it will not happen actually, as a now again as this is the assumption and it is assumed to be true. So, why cannot will take all this n 1 plus n 2 observations together, and calculate the variance; that is known as pooled variance. So, we will not take, will not merge them, and then finally, after if it is n 1 a plus n 2 data points, and we find out the s square using the formula, rather we use just different formula, which is known as s p square is the pooled variance. This is n 1 minus 1 s 1 square plus n 2 minus 1 s 2 square by n 1 plus n 2 minus 2

Now, this you, what you will do not, that mean your this value is known. One is, this value is known, and as you as you know a the sigma is not known this is know, this is known, and many a times if the sample size will be not be large also. So, whatever, but if we need sample size is large, we will use the distribution no difference, because for last sample size z and t distribution coincides; now, see this slide. So, it is a t distribution case. So, you know that minus t alpha by 2; that is a l and here d o f u will be plus t alpha by 2 d o f. Now what will be the d o p in this case, d o p in this case is n 1 plus n 2 minus 2. Now that is what is written here.

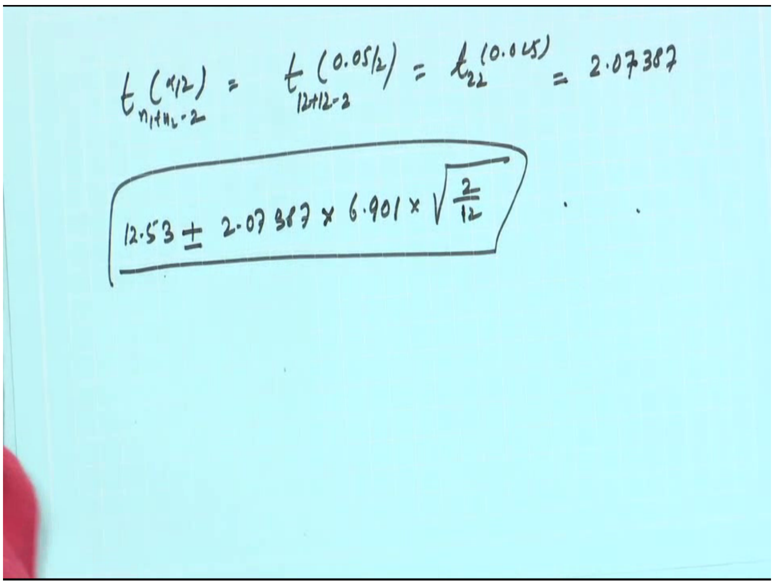
What we have written that our random variable of the interest is the average difference y 1 bar minus y 2 bar, and we are interested to find out the confidence interval mu 1 minus mu 2. Now this is this S p will be used in place of sigma, and S p is computed using this formula. So, it follows t distribution with n 1 plus n 2 minus 2 degrees of freedom n 1 plus n 2 2 degrees of freedom. So, now, see one example. Yes example is given that same example as the number of, because here a sigma 1 sigma 2 not known, and the sample size is 12 for population, it is not a last sample also, and what we will do. We will use, we assume that the population variances are equal under such situation, we will use the formula what is given.

(Refer Slide Time: 23:02)



What is this S p square formula n 1 minus 1 means 12 minus 1, it is 11 into the s 1 square 11 into s 2 square by 12 plus 12 minus 2; that is 22. So, S p square is coming at 47.625 S p is square root of the, this now the formula y 1 bar minus y 2 bar that value is 1.56 minus t 20 n 1 plus n 2 minus 2; that is 22 alpha by 20.05. You see that in this, this is the theta distribution table. So, degree of freedom is 22 alpha is 0.025 and the t value.

(Refer Slide Time: 23:55)

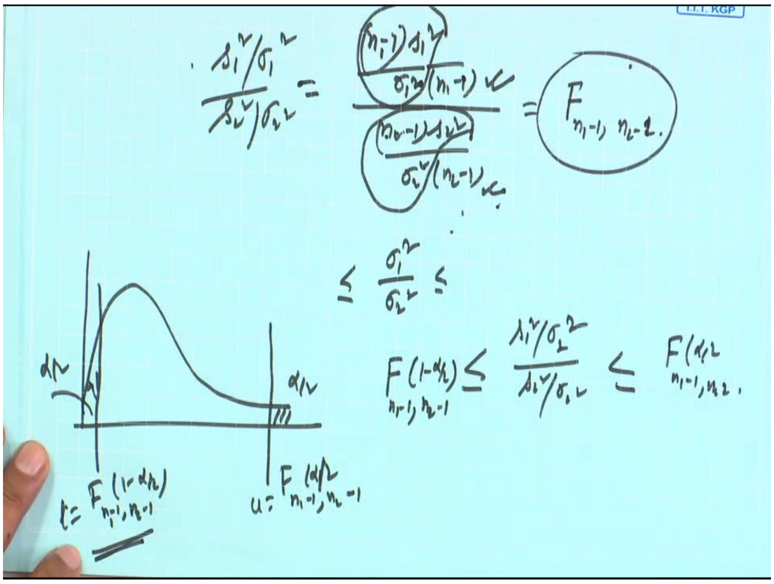


So, what I mean to say that t n 1 plus n 2 minus 2 alpha by 2 equal to t 12 plus 12 minus 2 into 0.05 by 2. This is t 22 0.025. And from table you get this value is 07 2.07387 2.07387.

So, from the slide what you will see that, you will see that t 22 0.025 equal to 2.07387, and you put this formula and you are getting this interval. So, it is coming minus. So, S p is this, and 1 by 1 minute formula 1 by n 1 minute 2 by 12; that is 1 by 6 t is this value y 1 bar minus y 2 bar y 1 bar minus y 2 bar. Same example we have taken y 1 bar minus 12.56 ok.

So, there is calculation mistake I think, it will be 12.53 12.53, and its 12.53 I think anyhow. So, what I mean to say that we use this formula, formula y 1 bar minus y 2 bar. And here y 1 bar minus y 2 bar is 12.5. So, if I consider this formula that 12.53 minus 2.07387 into S p value is our. What is our S p value 6.901 into root over 2 by, and that is 1 by 12 plus 1 by 12 2 by 12 this plus minus, this is the interval. So, you calculate this.

(Refer Slide Time: 26:29)



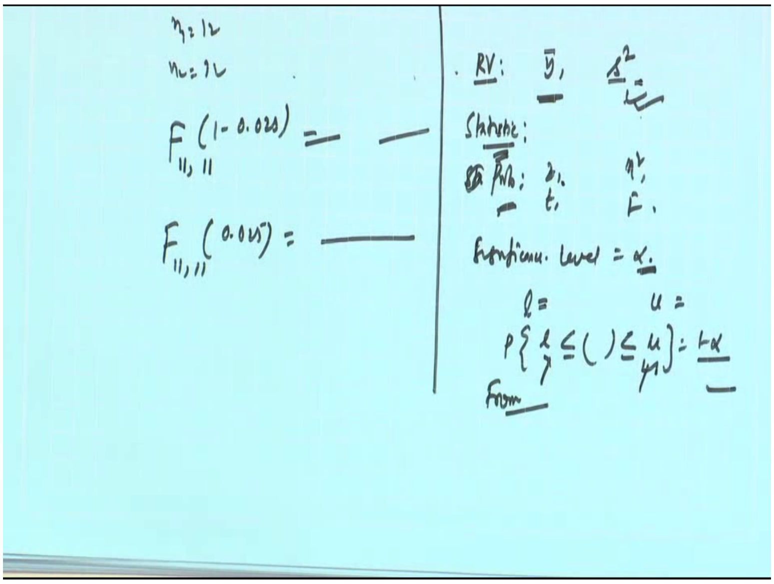
Now, we will discuss another concept called confidence interval for the ratio of two population variances. So, here we will use s 1 square by sigma 1 square and s 2 square by sigma 2 square. So, this, because we are considering n 1 minus that this quantity will statistics will create, and a. We will see that this can be written as n 1 minus 1 s 1 square by sigma 1 square into n 1 minus 1 divided by n 2 minus 1 s 2 square by sigma 2 square into n 2 minus 1. So, this quantity is chi square distribution, following chi square distribution, this follows chi square distribution divided by their respected degree of freedom.

So, chi square variable by degree of freedom rush by chi square, another chi square by degree of freedom. This will be f distribution and with n 1 minus 1 n 2 minus n 2 minus 1 n 1 minus 1 n 2 minus 1 degrees of freedom. So, that mean this quantity follows F distribution. So, what you have data you, what you want to know. You want to know the ratio. So, you are interested to find out interval something like this. So, you create what you will do. Suppose how you do create. You think that this is F distribution.

So, this one is alpha by 2. here also you take another area alpha by 2, and this is F n 1 minus 1 n 2 minus 1 alpha by 2, and this value will be F n 1 minus 1 n 2 minus 1 1 minus alpha by 2. So, this is my low value, this is my upper value. So, then what happen S 1 square by sigma 1 square, and S 2 square by sigma 2 square will lie in between F. This lower value n 1 minus 1 n 2 minus 1 1 minus alpha by 2 F alpha by 2 n 1 minus 1 n 2 minus 1. So, a now you manipulate this and you will be getting this.

This slide the S slide S 1 square by S 2 square divided by F n 1 plus minus 1 n 2 minus 1 alpha by 2 sigma 1 squared by sigma 2 and this. So, this is the formula. Now see the; however, same example filter type 1 and filter type 2 you want to construct the ratio of the variances. So, that S 1 square, what is the formula here. Formula is S 1 square by S 2 square. So, S 1 square by S 2 square, this is 1.563, then what is the F value F value what will be the F value F value will be. So, F a what is n 1 n 1 equal to 12 n 2 equal to 12. So, what you require F of 11 11 and 1 minus 0.025 this value, you find out and another one F 11 11 0.025 you find out and put their, put in into this equations and then calculate this one.

(Refer Slide Time: 30:45)



So, as in just in order, just to you know do this one. What is this that conclude that. Please remember that you first required to know, in the confidence interval you required to know the random variable of interest, when it is single population, then the either mean or your standard variance will be variable of interest, and you create a statistic. Particularly using the center limit theorem here, and here also that based on the distribution, what it will follow, it need the statistic find out the statistic that appropriate a probability distribution for the statistic.

So, some minute and z distribution t distribution for y bar chi square and F distribution for variance. So, appropriate statistics here will be z or t here will be chi square or F, then you choose that alp significance level. Usually we say it is alpha and this will give you the lower limit and upper limit, and you choose in such a manner that l this a statistics. So, it will u it will be 1 minus alpha, and then pooled take from table from appropriate table, find the lower value and upper value, put here and you get the confidence interval.

So, Thank you very much [FL].