Six Sigma

Prof. Jithesh J Thakkar
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

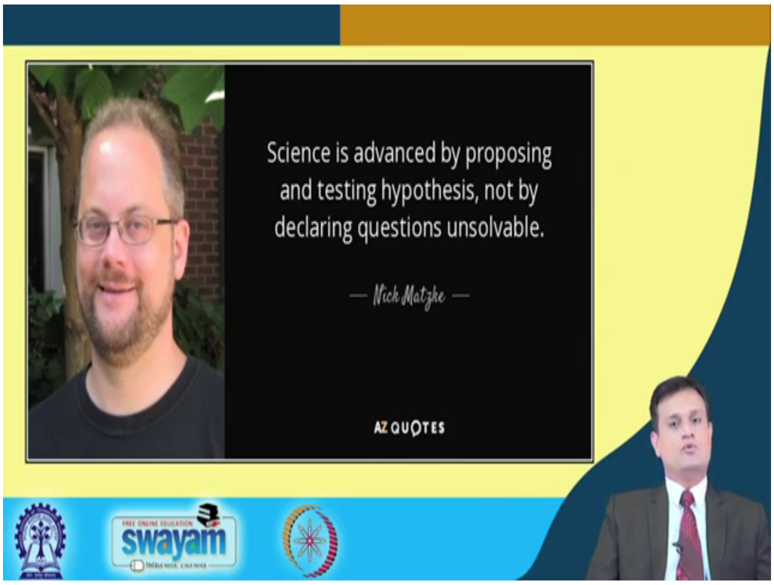
Lecture - 30
Hypothesis Testing: Two Population Test

Hello friends, we are going through our Six Sigma journey. And typically we are discussing the analysed phase of DMAIC cycle. And we have already advanced in our discussion on hypothesis testing. We had seen the concepts; we had seen the hypothesis testing for single population. And this lecture 30; we will basically try to help to appreciate Hypothesis Testing for Two Population Test.

So, the concept will remain same and the general steps that we have follow in the hypothesis testing will remain same. But it would be done for two populations; I am interested to compare or to check the fact for two different population. And this could be let us say you are receiving the material from two different vendor, and you want to check that whether the percentage defective from vendor 1 vendor 2 are the same or they are different.

Or let us say you want to check some fact about the people who are less than 30 years of age, more than 30 years of age, or let us say you want to check some gender specific fact then male versus female. So, you have two different population of interest. And I want to test the hypothesis for the comparison of this two different population and this is called my hypothesis testing for two population test.

(Refer Slide Time: 01:53)



So, once again you are reminded that scientific science is advanced by proposing and testing hypothesis, not by declaring questions unsolvable.

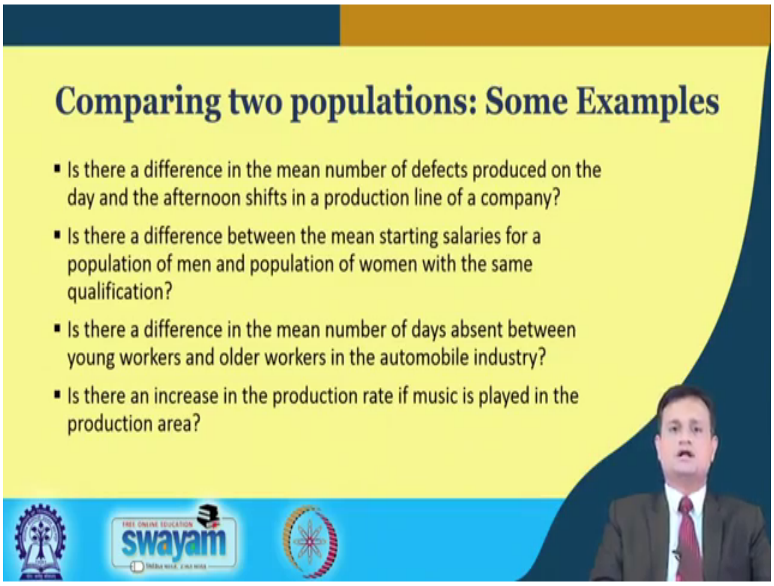
(Refer Slide Time: 02:07)

So, we had seen the z statistic, t statistic, and for proportion in the previous lecture.

(Refer Slide Time: 02:15)

Now, this lecture we would basically like to focus on two population test where we have four different cases. Case number 1: the means of two independent populations I want to compare. Number 2- the proportion of two independent populations. Number 3- the variance of two independent populations by testing the ratio of two variances. And number 4- dependent related samples same group before versus after treatment.

(Refer Slide Time: 02:51)



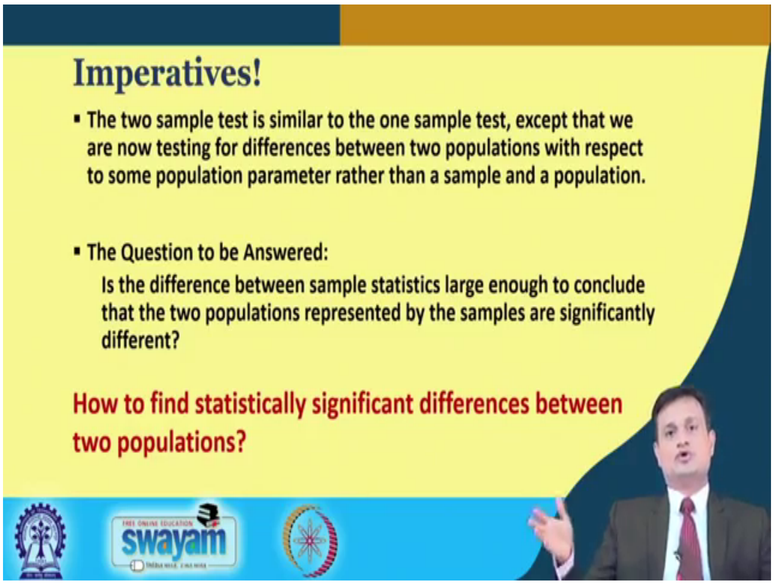
So, just try to create your interest in this particular topic. Just see that what could be the questions that may prompt you to go for such kind of hypothesis testing for two population. Say is there a difference in mean number of defects produced on the day and the afternoon shift in a production line of a company. You want to investigate; your production line is same manufacturing setup is same, but there could be influence of workers skill or some other parameter issues.

And is there really a difference so for the defects are concerned in the morning shift and in the afternoon shift. Is there a difference between the mean starting salaries for a population of man and population of women with the same qualification? You want to check is there a gender bias? Or is there a preference for one particular gender and is there a difference in the average salary. Then is there a difference in mean number of days absent between young workers and the older workers in automobile industry it is also interesting.

That you are the young worker and older worker when you just go by the number you will say fine young worker they were say 27 days absent on an average and the old worker let us say 35 days. But this is just the number descriptive statistics mean I am not making any inferential analysis this could be statistically different this could be same. So, I want to check is there really a statistical evidence available that; average number of absenteeism young worker and the old worker there is a difference.

Is there an increase in the production rate if music is played in the production area very interesting can that have a soothing effect a positive effect on the behaviour skill of the worker and that can help them to improve the say production or the productivity. So, these are some of the interesting questions that you would like to investigate through two population test.

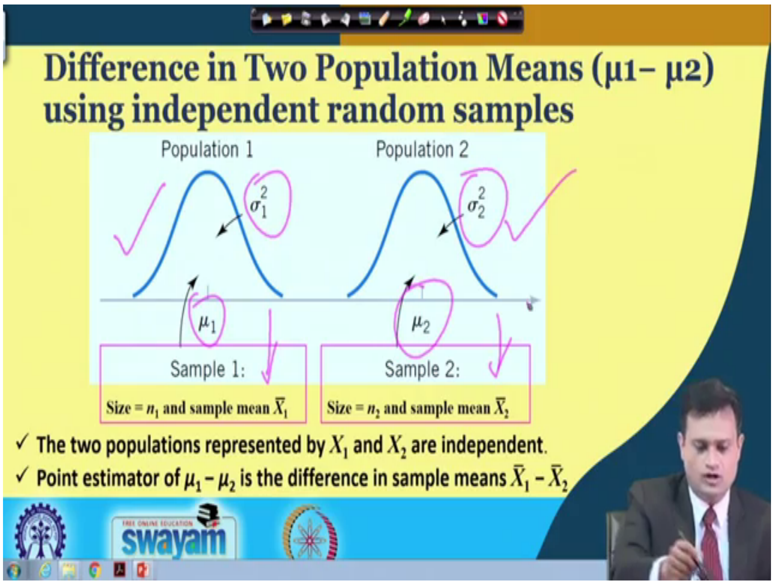
(Refer Slide Time: 05:09)



So, the imperatives are there two sample test is similar to one sample test. As I said there would not be any difference in the procedure and hence I would not spend too much time on explaining the critical value p value concept. I would more focus on the real life application and the interpretation part in this lecture.

So, the question to be answered; is there a difference between sample statistics large enough to conclude that two populations represented by the samples are significantly different. So, we have already read some of the interesting questions that can really help you to appreciate the importance of hypothesis testing for two population.

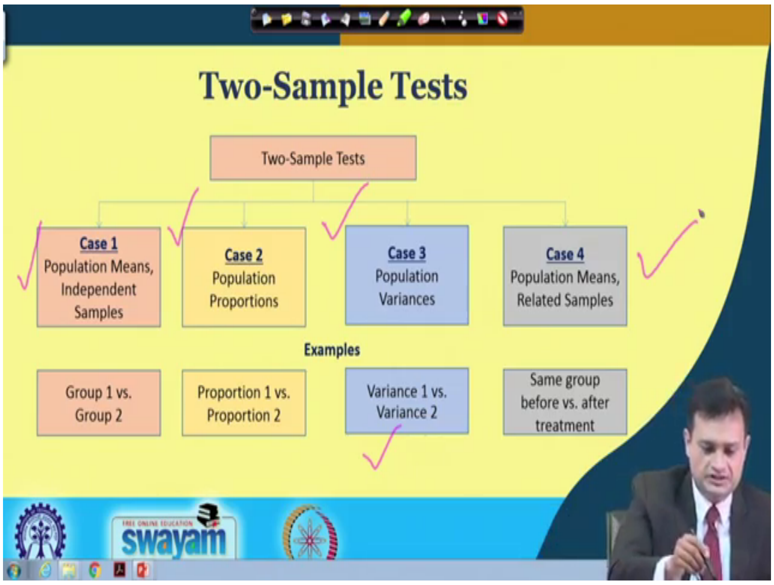
(Refer Slide Time: 06:01)



Just see this to get a better idea, you have basically difference in two population mean mu 1 minus mu 2. And using independent random sample let us say; we have population 1 and population 2. You have sigma 1 square for population 1 sigma 2 square for population 2. This is my mu 1, this is my mu 2 and I am drawing this sample 1 for population 1 sample 2 for population 2.

So, two populations are represented by X 1 X 2 which are independent and mu 1 minus mu 2 is the difference in the mean X 1 bar and X 2 bar. So, I want to check compare my claim for two different populations which are independent. And this is what exactly we would do in this particular lecture.

(Refer Slide Time: 06:53)



Let me give you the frame work of this particular lecture and what are the different cases which I would like to discuss as a part of hypothesis testing for two population test. So, case 1 we have population means and independent samples group 1 versus group 2. Suppose average performance of the boys in the school may be higher secondary and average performance of the girls in sports in the higher secondary is there a really significance difference or they are same?

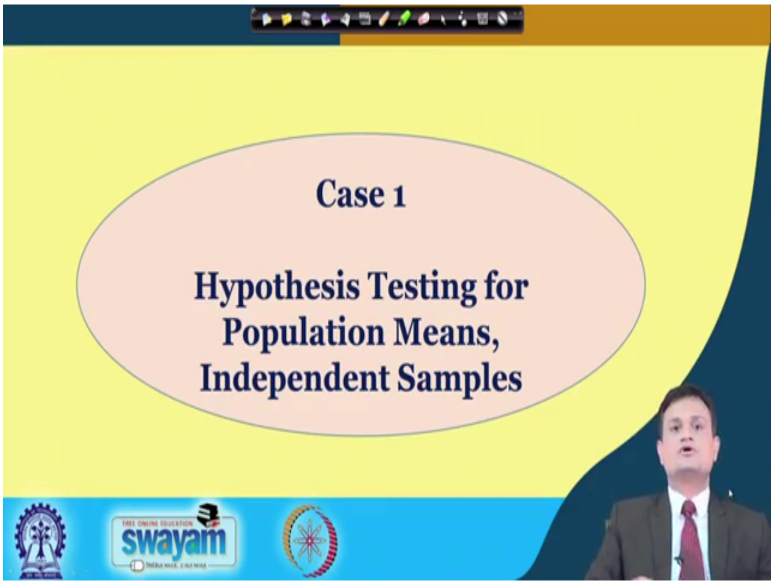
We have seen average salary for male average salary for female production in the morning shift, production in the afternoon shift. So, this would fall typically in the first case that is population means independent samples. Now just see the case 2 population proportion as I have explained previously number of success and what is that proportion that something is happening some number of time and I would like to compare proportion 1 with proportion 2 and not the two means as we are doing in the case 1. Now you have case 3.

So, many a times the variance is of interest and I want to check the variability present in the population 1 and the variability in the population 2. So, I want to compare variance one versus variance 2 by taking an appropriate sample and this is my case 3. When I say case 4 population means related samples same group before or after treatment. Now this would be really interesting suppose you are implementing a six sigma program for a particular function or process.

Now you want to see that is there really a reduction in the defect rate after implementing a six sigma program. So, you have the before data you have the after data and then you would like to check it. Similar way you can think that there are patients suffering from a particular viral disease they are given the antibiotics and you want to see the improvement in their symptom.

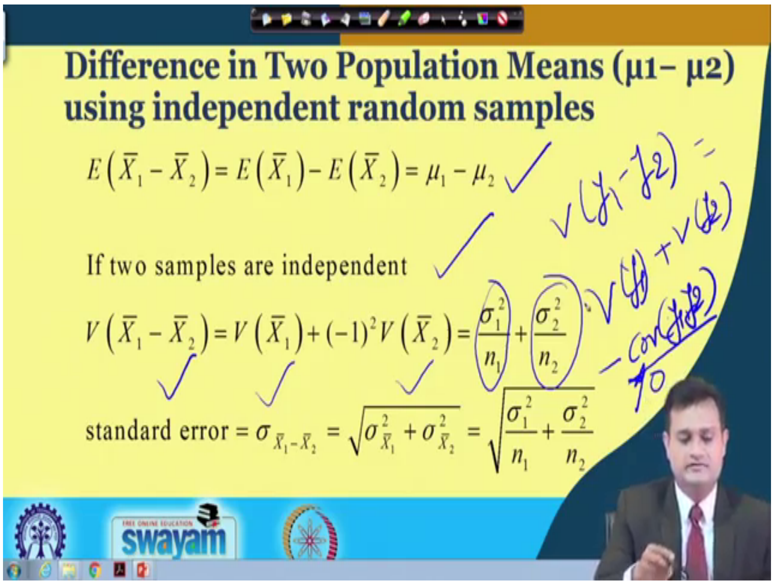
So, you have the same group before and after treatment and you are interested to compare this two. And here these two before and after are treated as two different population. So, with this broad classification let us try to discuss each particular case with some real life example.

(Refer Slide Time: 09:39)



So, case 1- hypothesis testing for population means independent samples.

(Refer Slide Time: 09:45)

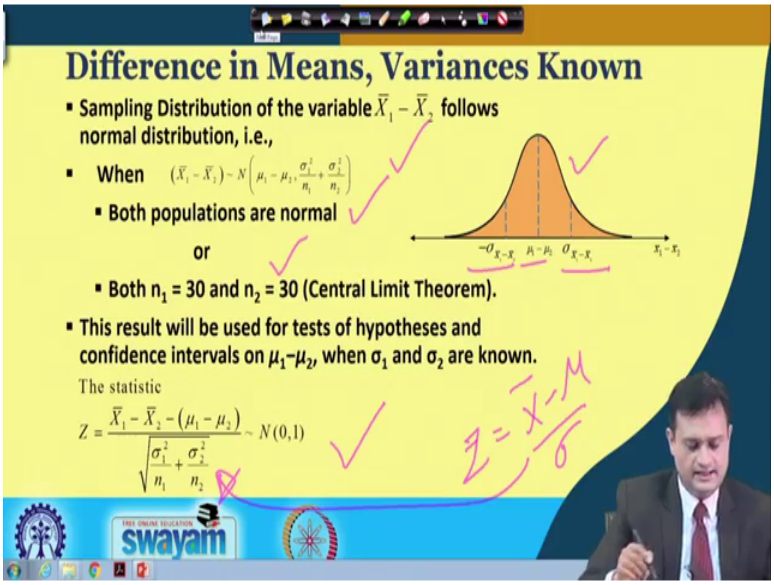


So, just I would like to present little bit mathematics. So, you have expectation E stands for expectation X 1 bar minus X 2 bar is expressed as expectation of X 1 bar minus expectation of X 2 bar. And in statistics we say that expectation of X 1 bar is mu 1. Means expectation of x bar as per central limit theorem is close to mu. So, I have mu 1 minus mu 2 as expectation of X 1 bar minus X 2 bar so this is mu 1 minus mu 2.

Now if two samples are independent; variance of X 1 bar minus X 2 bar would be expressed as; variance of X 1 bar plus minus 1 square variance of X 2 bar sigma 1 square divided by n 1 sigma 2 square divided by n 2 as per central limit theorem. Typically suppose you have two variable then variance of y 1 minus y 2 is typically expressed as variance of y 1 plus variance of y 2 minus co variance of y 1 into y 2. Because you have these two events y 1 and y 2 independent my covariance is basically 0.

So, it is expected that you have some basic knowledge of statistics, you should revise the suggested book. And on this logic I have variance of X 1 bar minus X 2 bar is equal to sigma 1 square divided by n 1 plus sigma 2 square divided by n 2. And typically you can express your standard error as sigma X 1 bar minus X 2 bar which is equal to square root of sigma X 1 bar plus sigma X 2 bar. And this is basically the square root of sigma 1 square divided n 1 plus sigma 2 square divided by n 2. So, this is what we say.

(Refer Slide Time: 11:57)

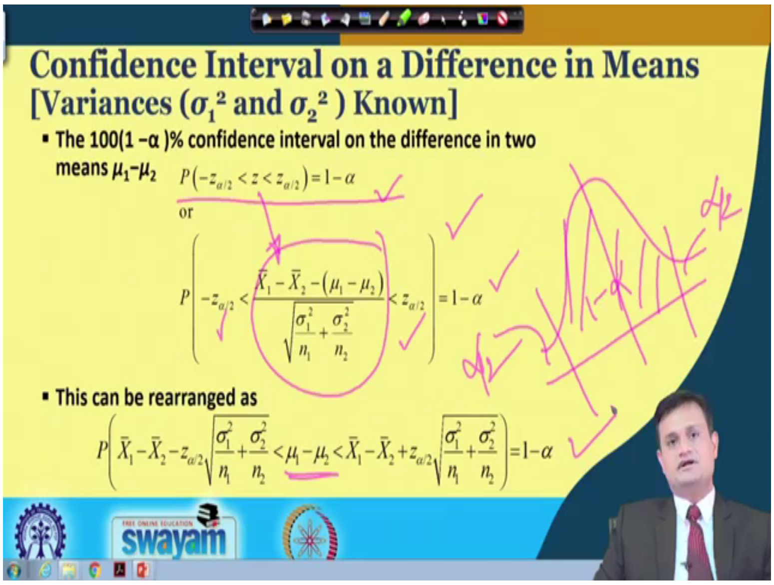


Now, you have two populations. So, I want to check the difference in mean and suppose variances are known. So, variances are known means I am aware about the population variance and I want to say that my sampling distribution of the variable X 1 bar minus X 2 bar follows the normal distribution. So, please understand that a distribution typically represents the behaviour of a particular random variable studied over a period of time.

And here I measure the random variable as X 1 bar minus X 2 bar. So, this is my sampling distribution for X 1 bar minus X 2 bar. And that is why here it is expressed as mu minus mu 2 that is population mean. And sigma X 1 bar minus X 2 bar minus sigma X 1 bar minus X 2 bar. So, both population are normal both n 1 is equal to 30 n 2 is equal to 30 central limit theorem can be applied.

And your typically standard normal z statistics can be computed as basically z is equal to x bar minus mu divided by sigma. Same thing is basically transcribed here as X 1 bar minus X 2 bar. Because I am considering X 1 bar minus X 2 bar as my random variable so X bar minus mu 1 minus mu 2 bar equivalent to minus mu divided by my standard deviation. So, this approaches to standard normal. And this is my test statistics.

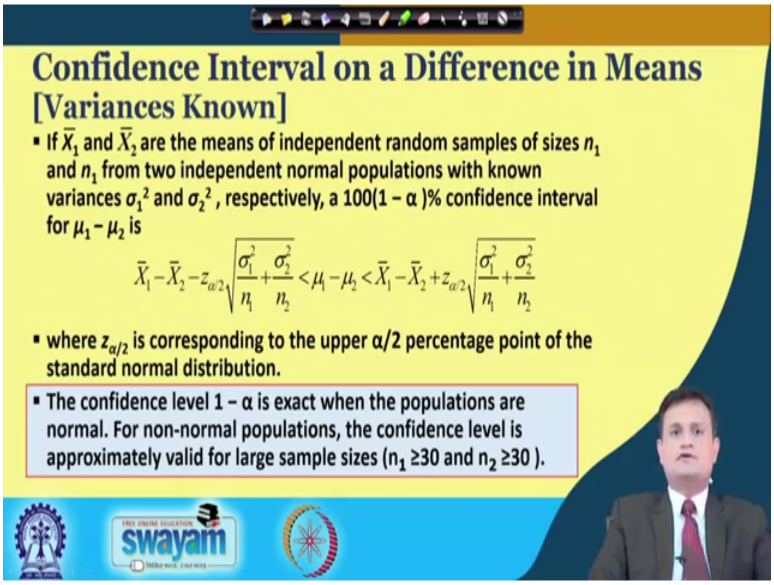
(Refer Slide Time: 13:37)



So, now I can also create the confidence interval for plus or minus z sigma or k sigma. And you can see here that I want to find the probability which is probability of minus z alpha by 2 less than z less than z alpha by 2 it should be 1 minus alpha. So, I want to typically find a region in this particular normal distribution. So, I have this alpha by 2 I have this alpha by 2 I am interested to find 1 minus alpha region this particular region. And say probability when I convert this into standard normal you can see this is what we had.

So, this particular quantity is my standard normal z statistics minus z alpha by 2, z alpha by 2 is equal to 1 minus alpha. So, you can just do little bit iteration to convert this expression simply having mu 1 minus mu 2 in the centre. And you will find that this is the expression after rearranging gives me the probability that what is that particular confidence interval which gives me 1 or minus alpha region. So, this is what we can do as confidence interval.

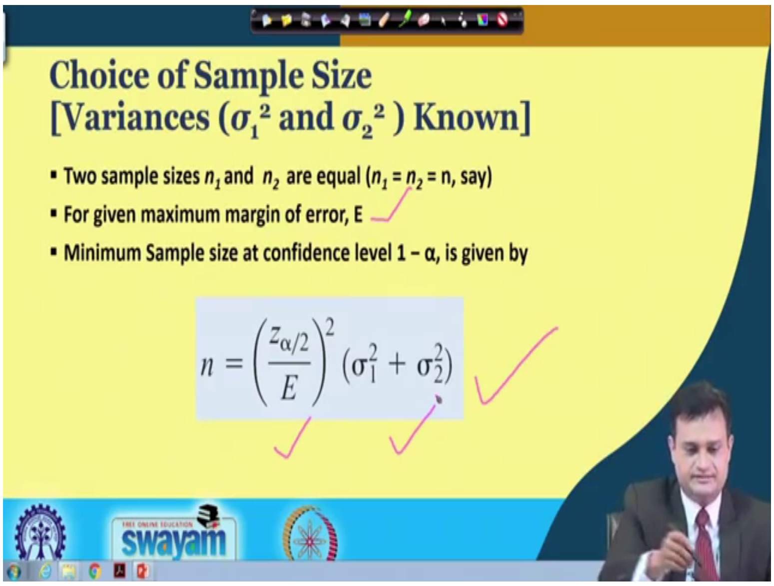
(Refer Slide Time: 15:01)



Now, you can have confidence interval in mean for variances known. Now here I assume that my variances are known about the population and if X 1 bar and X 2 bar are the means of independent random sample of size n 1 and typically n 1 from two independent normal population with known variance.

Then 100 1 minus alpha percent confidence interval for mu 1 minus mu 2 can be constructed like this. So, you only need to follow the same procedure in order to get the confidence interval for the given situation.

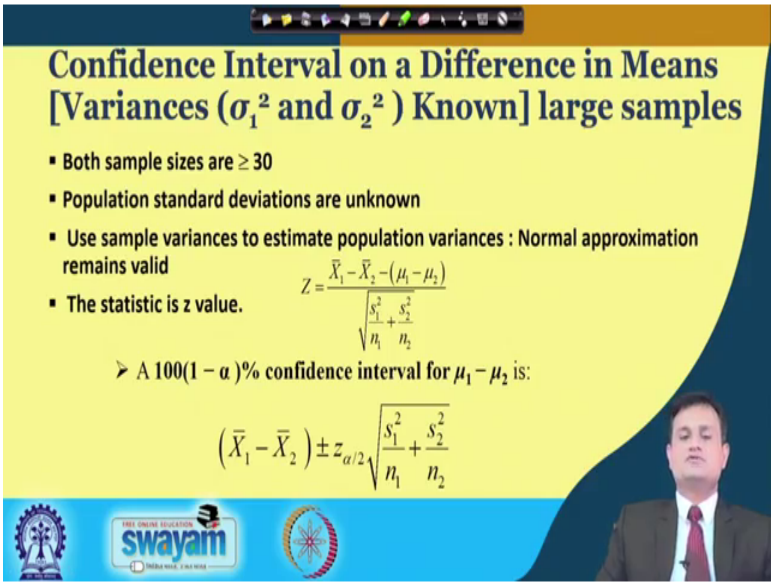
(Refer Slide Time: 15:51)



Now, you have the sample size to be determined and for choice of sample size. Let us say if I am aware about the variance sigma 1 square and sigma 2 square of two different population. And two sample n 1 and n 2 let us say are equal and equal to n 1 is equal to n 2 is equal to n. So, for given maximum margin of error E this can be expressed as n is equal to z alpha by 2 divided by E square plus sigma 1 square plus sigma 2 square.

We are not going into the details of derivation part. We accept some of the expressions as it is in the standard form just in order to facilitate our computation and the investigation for a typical real life problem through hypothesis testing.

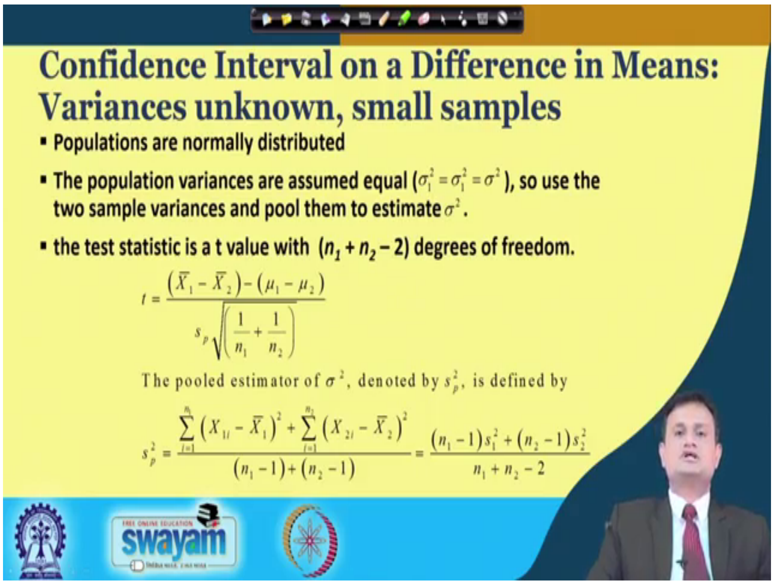
(Refer Slide Time: 16:49)



So, you have let us say confidence interval on a difference in means variance sigma 1 square and sigma 2 square are known in your large samples. So, as we have seen that your typical say expression would follow X 1 bar minus X 2 bar minus mu 1 bar minus mu 2 divided by s 1 square divided by n 1 plus s 2 square divided by n 2.

So, you have this sample standard sample variance for population 1 and as well as for population 2. And you can create the 100 1 minus alpha percent confidence for mu 1 minus mu 2. So, we have different cases where we tried to.

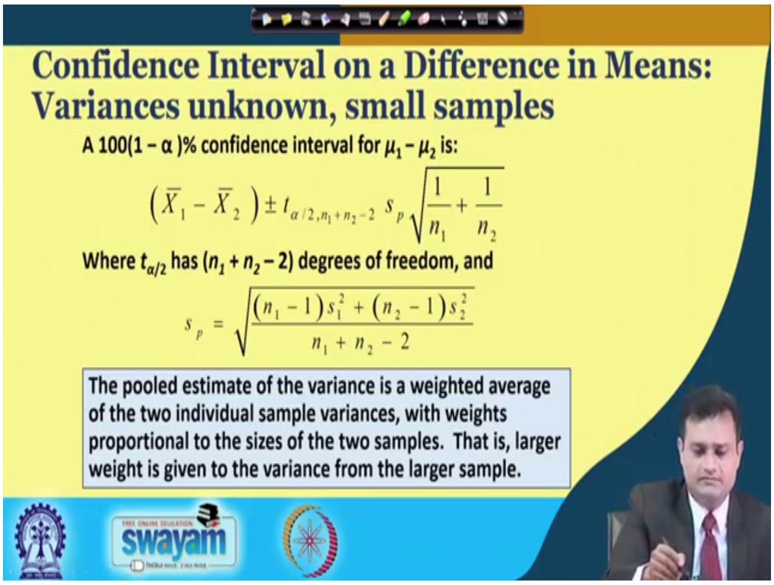
(Refer Slide Time: 17:37)



Say compute the z statistics t statistics and create the confidence interval. Now you want to create the confidence interval on let us say mean for a small sample. So, when I talk about small sample usually we refer to the student t distribution and your expression would become something like this. So, if this is your t statistics X 1 bar minus X 2 bar minus mu 1 minus mu 2 divided by s p which is called pooled variance.

So, s p square root of 1 divided by n 1 plus 1 divided by n 2 a standard expression of t statistics we are accepting for the checking or testing of the hypothesis. And your pooled variance is given by this particular expression n 1 minus 1 s 1 square plus n 2 minus 1 s 2 square divided by n 1 plus n 2 minus 2. Because I am considering 2 sample so I have to consider the n 1 plus n 2 minus 2 as the denominator.

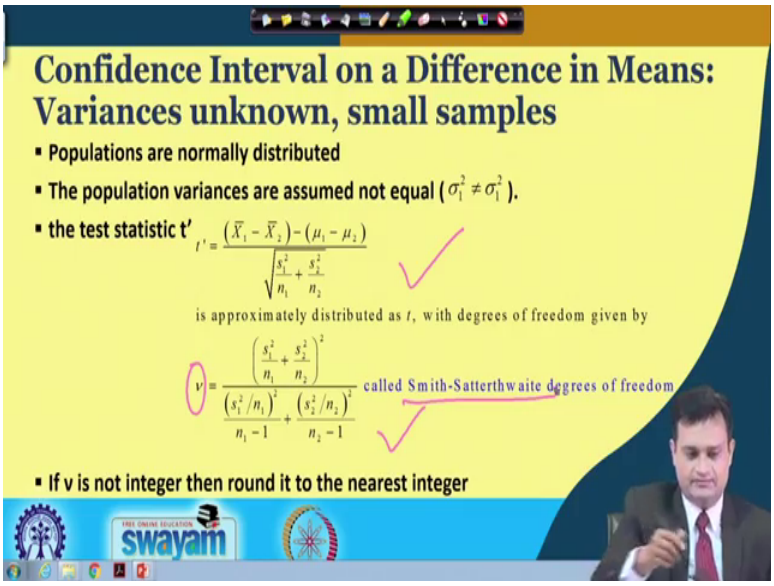
(Refer Slide Time: 18:53)



So, you have 100 1 minus alpha percent confidence interval for mu 1 minus mu 2. And basically it is your t because we are referring the small sample case t distribution.

So, basically X bar or mu plus or minus k sigma z sigma in case of t it is t sigma and pooled variance is given by this. So, this is the standard procedure we follow.

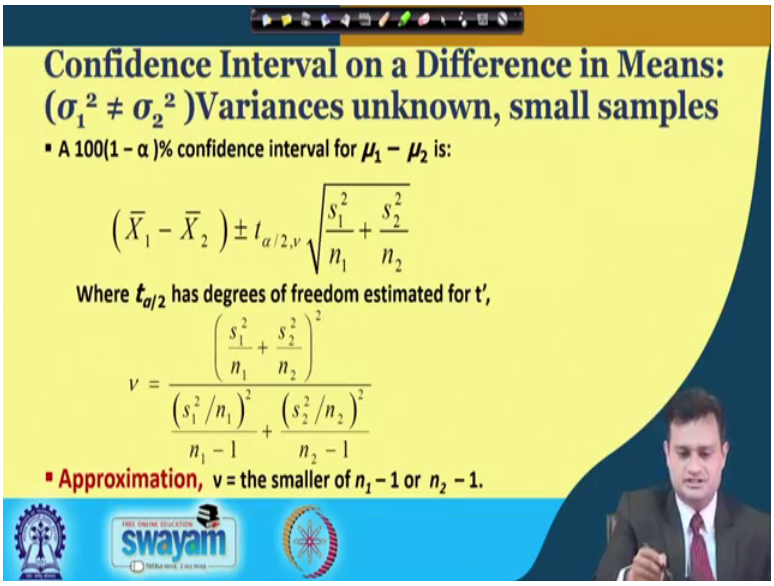
(Refer Slide Time: 19:21)



And you have another situation where; confidence interval on difference in means variances are unknown. So, so far we assumed that I am aware of the variance of the population. But many a times say particular phenomenon you are encountering first time or little evidences is available little pass data is available. Then you are not very much aware of the population variance in this case you will little bit modify the expression.

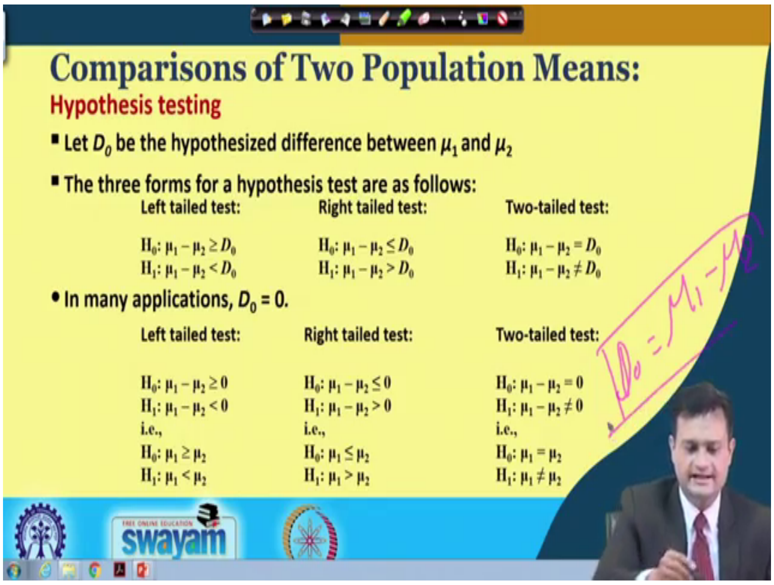
And your test statistic will become t dash express like this and you will consider nu typically nu something called s 1 square divided by n 1 plus s 2 square divided by n 2. And you have denominator s 1 square divided by n 1 whole square divided by n 1 minus n 2 and this is also called Smith-Satterthwaite say degree of freedom. And this degree of freedom you will use to find the critical value or the p value from your say t table statistical table.

(Refer Slide Time: 20:35)



So, we have this particular 100 1 minus alpha percent confidence for my t when variances are not known.

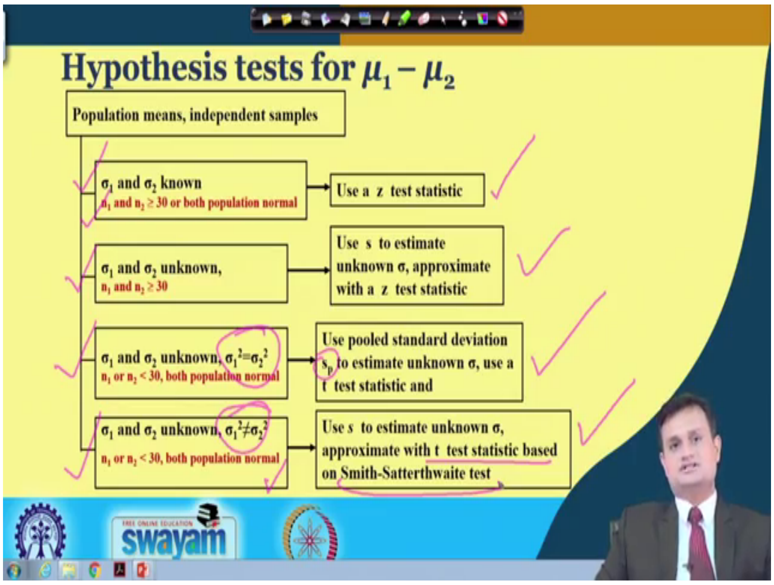
(Refer Slide Time: 20:47)



And I have the situation like this that let D 0 be hypothesise difference between mu 1 and minus mu 1 and mu 2 so typically mu 1 minus mu 2. So, D 0 is mu 1 minus mu 2 and I can have different forms of hypothesis to be tested. I can say that X 0 is mu 1 minus mu 2 greater than or equal to D 0 or less than equal to D 0.

I can have right tailed test I can have two tailed test I can have say many applications suppose D 0 is equal to 0 then; left tailed test right tailed test and two tailed test. So, you just need to appreciate that depending upon your interest right tailed test two tailed test you would like to set the hypothesis for checking your claim. So, we will see couple of examples; so the idea would be better clear.

(Refer Slide Time: 21:49)

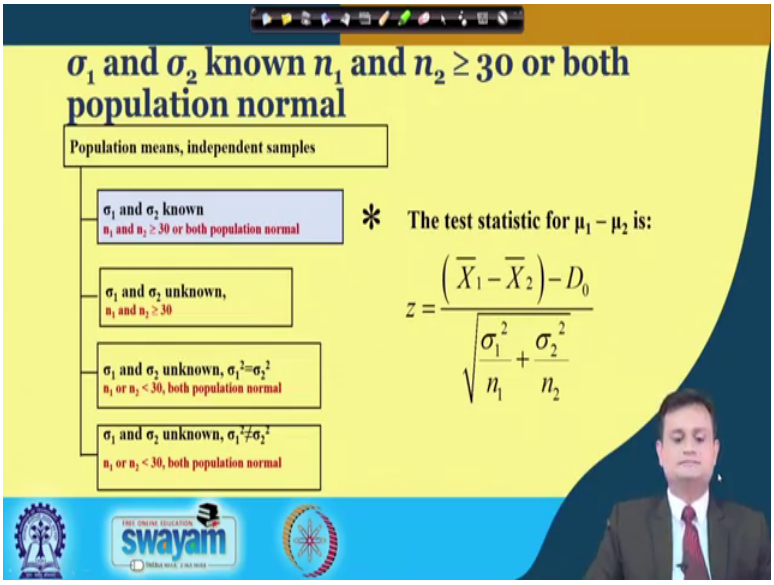


Now, just see that what could be the different situations in which we can refer our case 1. We are with case 1 and we have population mean independent sample this is my case 1. So, I can have a situation where sigma 1 and sigma 2 known means population standard deviations are known and n 1 and n 2 greater than equal to 30. So, both the populations are normal. You compute z statistic the guiding rule is that you compute z statistic.

You have sigma 1 and sigma 2 unknown so n 1 and n 2 greater than equal to 30. Use s to estimate unknown sigma approximate with z test statistic; sigma 1 and sigma 2 unknown sigma 1 square is equal to sigma 2 square, n 1 or n 2 less than 30 both population be assumed to be normal. You go for t statistic because your sample size is less than 30 you have sigma 1 and sigma 2 unknown sigma 1 square is not equal to sigma 2 square n 1 or n 2 they less than 30 both population normal. You go for s estimate of unknown sigma and approximate with t statistic based on Smith-Satterthwaite test.

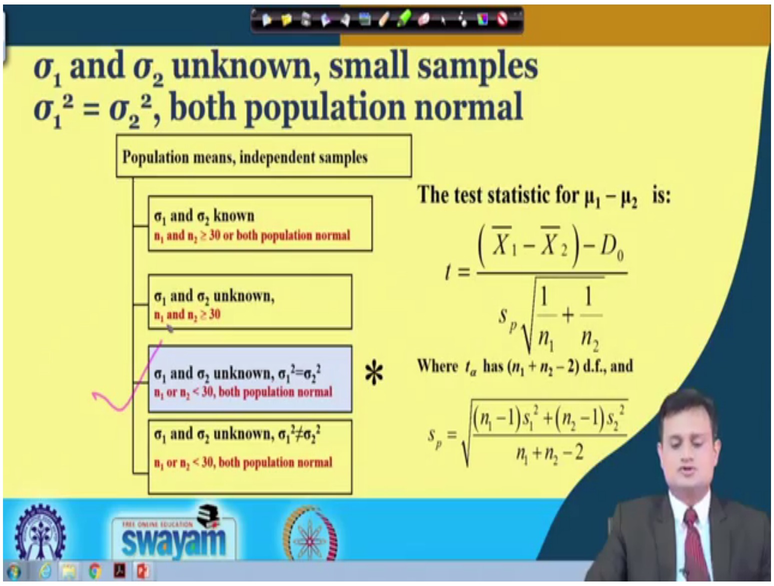
So, this are the variance things you can see in this last two case that here; this two are equal I would just like to draw your attention. Here this two are not equal so in one case we are using the pooled standard deviation in another case you go with the t test statistics based on Smith-Satterthwaite test. So, these are the standard guidelines to follow for testing the hypothesis for two population independent of each other

(Refer Slide Time: 23:45)



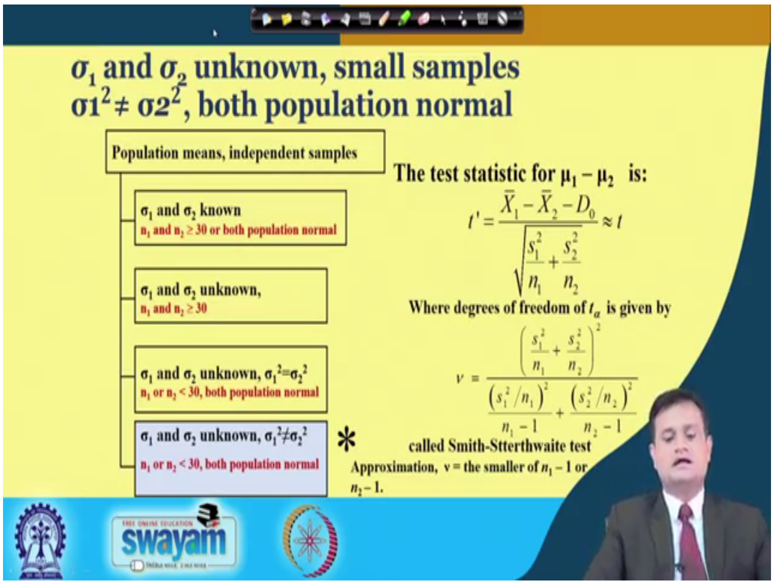
And the variance test statistics which I explained before I have just summarized here if you have a case 1; this is the test statistic which needs to be computed for finding the critical value, as well as finding the p value.

(Refer Slide Time: 24:23)



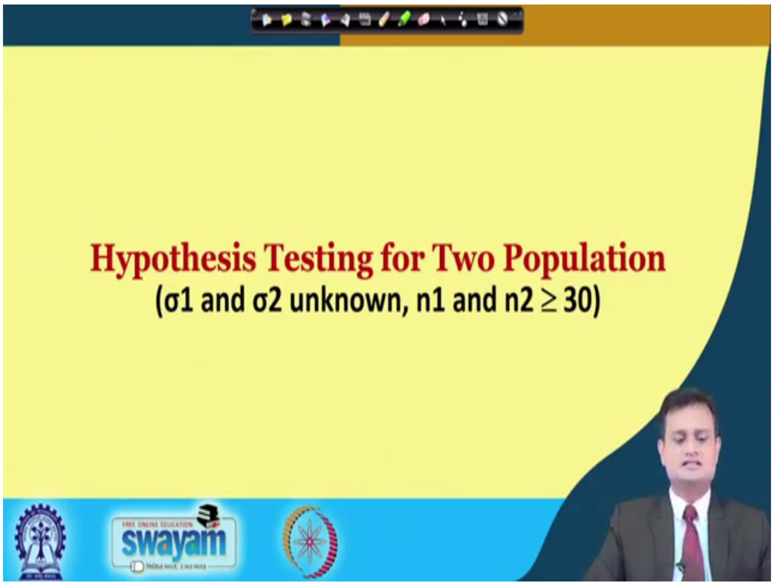
Now second one is this one so, you have the case sigma 1 sigma 2 known this is your z statistic you will make use of pooled variance. The third one; is the case where you have sigma 1 square is equal to sigma 2 square you will use this as the test statistic.

(Refer Slide Time: 24:39)



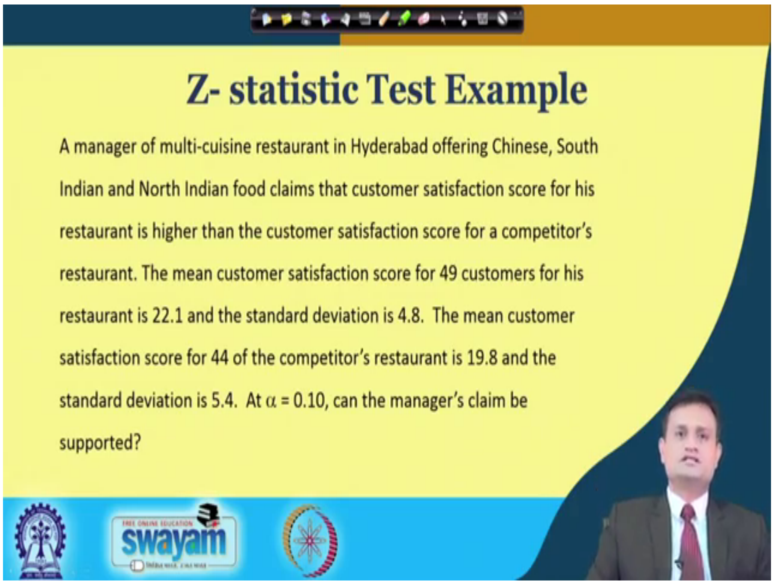
And similar way you have fourth case; where sigma 1 and sigma 1 square and sigma 2 square are basically not equal. And you will go by this Smith-Satterthwaite test to check your hypothesis to test your hypothesis.

(Refer Slide Time: 24:55)



So, we have some examples to discuss hypothesis testing for two population sigma 1 and sigma 2 unknown n 1 and n 2 greater than or equal to 30.

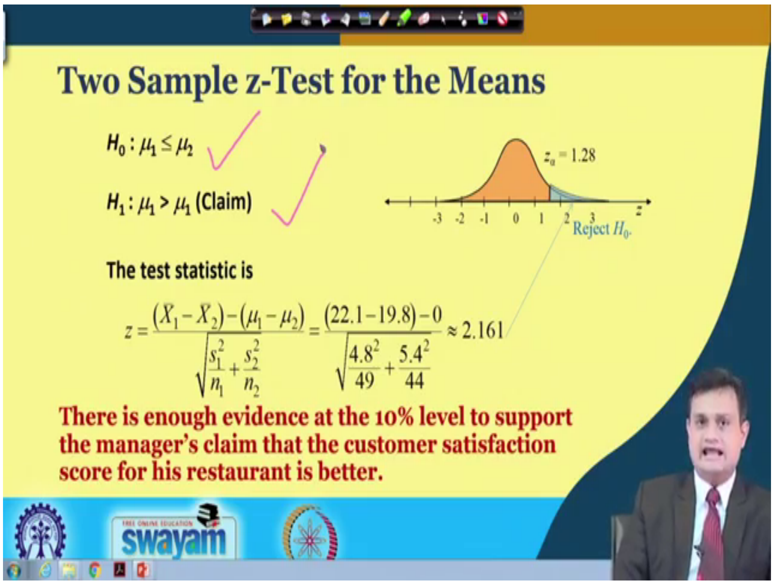
(Refer Slide Time: 25:05)



Let us say my situation is like this a manager of multi cuisine restaurant; in Hyderabad offering Chinese, South Indian, North Indian food. And he claims that the customer satisfaction score for his restaurant is higher than the customer satisfaction score for a competitors restaurant.

The mean customer satisfaction score for 49 customer for his restaurant is 22.1 in the standard deviation he found is 4.8. The mean customer satisfaction score for 44 of the competitors restaurant is 19.8 and the standard deviation is 5.4 he wants to check his claim that alpha is equal to 0.1 and he want to see that whether his claim is supported or not.

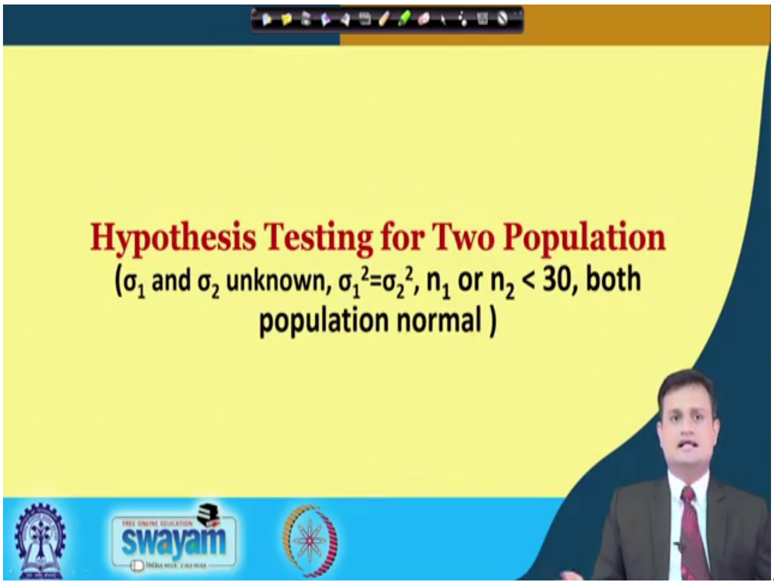
(Refer Slide Time: 25:59)



So, I will not go too much into detail of explaining; how to find critical value and refer it all this is well explained you have the hypothesis, mu 1 is less than equal to mu 2 it is about customer satisfaction score mu 2 is greater than mu 1 this is the claim of manager. And I want to check whether I am with H 0 or H 1. So, you will compute z statistic and 2.16 is the value. So, you will say that 2.1 and your z alpha is 1.28.

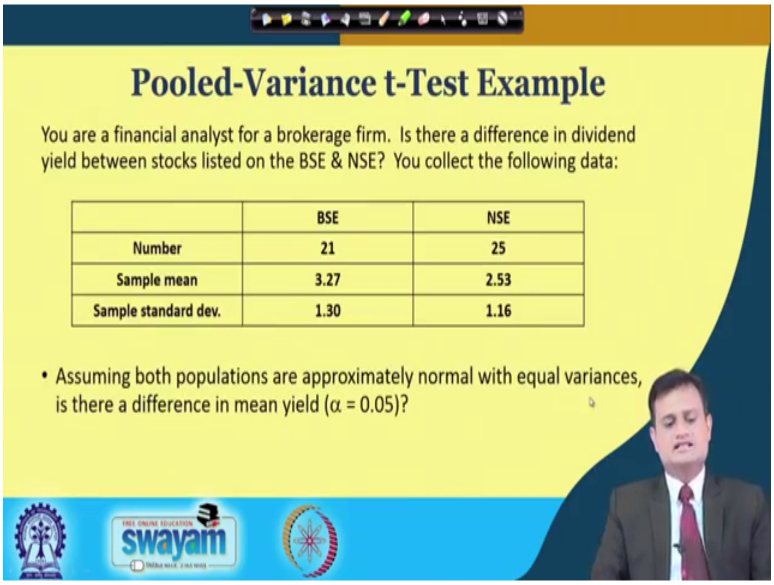
So, when you start from 0 your 2.161 is somewhere here and this falls in the blue region; which is the rejection region. So, you will say there is enough evidence that ten percent level alpha is equal to 0.1. Support the claim of the manager that customer satisfaction score for his restaurant is better. So, he can feel satisfied that yes customers they are more happy with my strategy, my offerings compared to the competitor. So, this is something that can really give the confidence to the business and the managers.

(Refer Slide Time: 27:13)



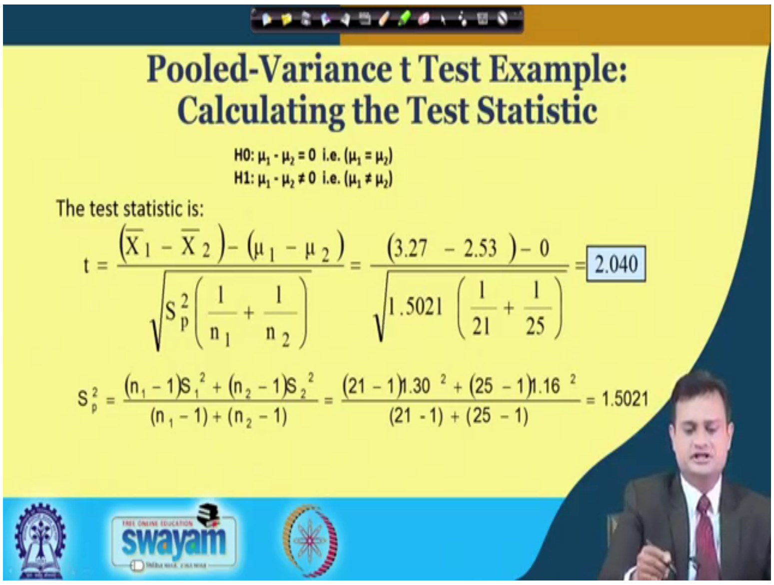
We can have another case; sigma 1 and sigma 2 unknown sigma 1 square is equal to sigma 2 square n 1 or n 2 is less than 30 both population normal.

(Refer Slide Time: 27:23)



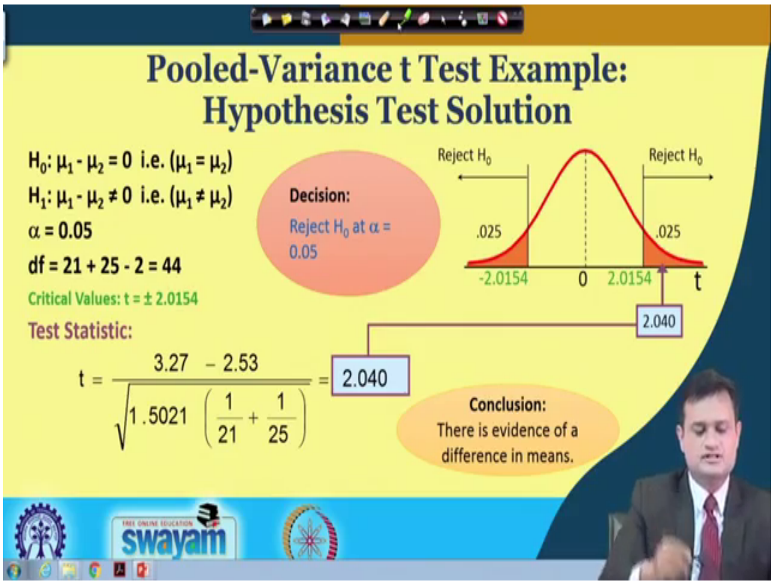
So, let us say another interesting case you have the stock exchange BSE and NSE. Suppose I am interested to see or an analyst is interested to see that is there a difference in dividend yield between stocks listed on BSE and NSE. So, he just said some data crunching found let us say; 21 BSE and 25 say NSE share. Sample mean is 3.27 for dividend and standard deviation is 1.3, 1.16 wants to check at alpha is equal to 0.05.

(Refer Slide Time: 28:07)



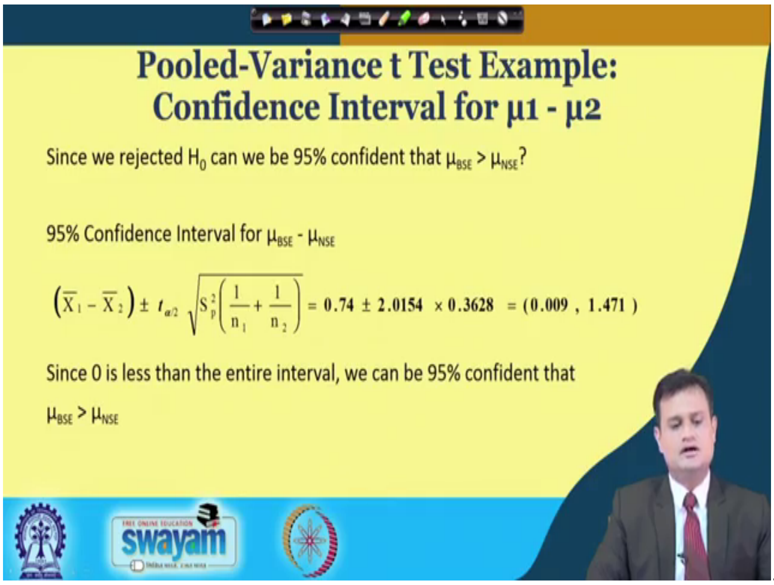
So, now we can just compute the t statistic. So, this is my null hypothesis mu 1 minus mu 2 is equal to 0; mu 1 minus mu 2 is not equal to 0 it means there is not much difference in the dividend yield by the BSE or NSE, null hypothesis says this and alternate says there is a difference. So, I will compute here t for this case and comes out to be 2.040. My pooled variance is computed with this expression which is already given to you and I can take the decision.

(Refer Slide Time: 28:47)



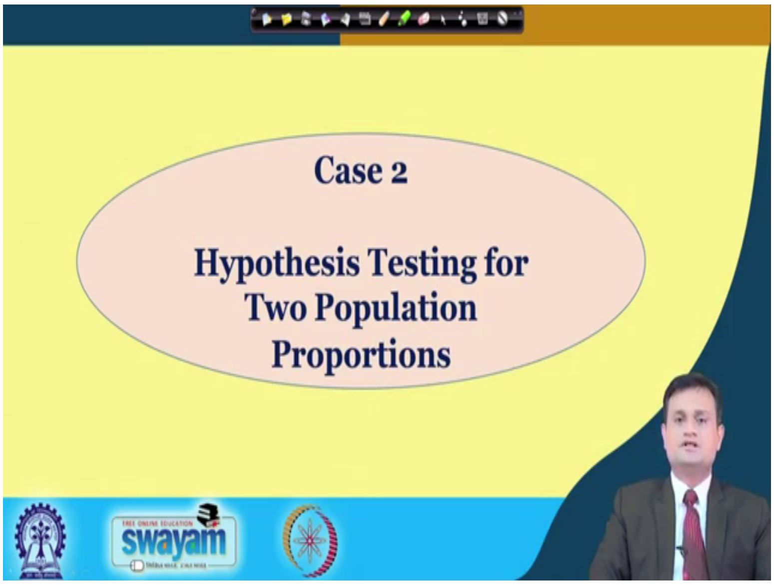
So, here you will see that alpha is equal to 0.05 degree of freedom is 21 plus 25 n 1 plus n 2 minus 2 is 44. So, my computed statistics 2.40 falls in this there is a evidence of difference in mean. So, yes the dividend is different when you look at the stocks listed on the BSE and NSE that is what my financial analysis says. And I reject the null hypothesis so my alternate claim is true that is mu 1 is not equal to mu 2. It means there is a difference in the dividend if I refer BSE and if I refer a stock on the NSE. So, this is something really interesting.

(Refer Slide Time: 29:37)



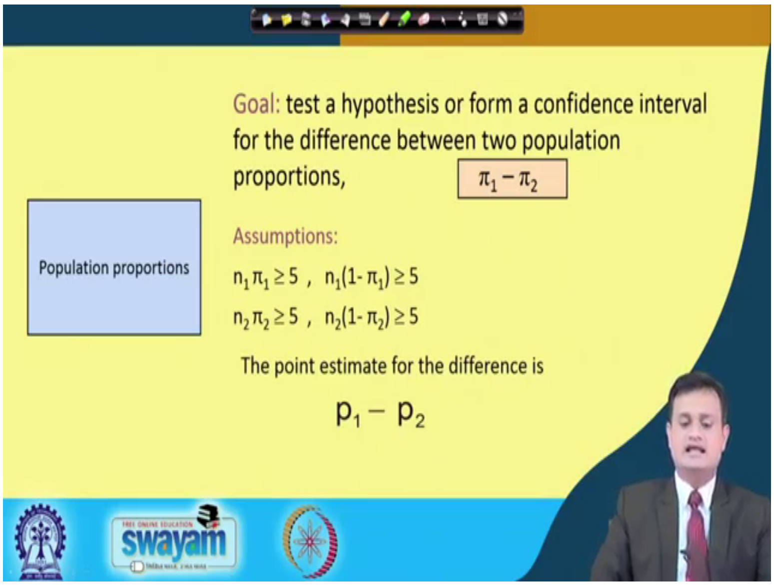
And I can also have confidence interval on my X 1 bar minus X 2 bar for BSE and NSE. So, you can use this particular expression X 1 bar minus X 2 bar plus or minus t alpha by 2. And you can see that whether my fall value falls within this will again yield the same result for 95 percent confidence interval. That mu of BSE is greater than mu of NSE.

(Refer Slide Time: 30:09)



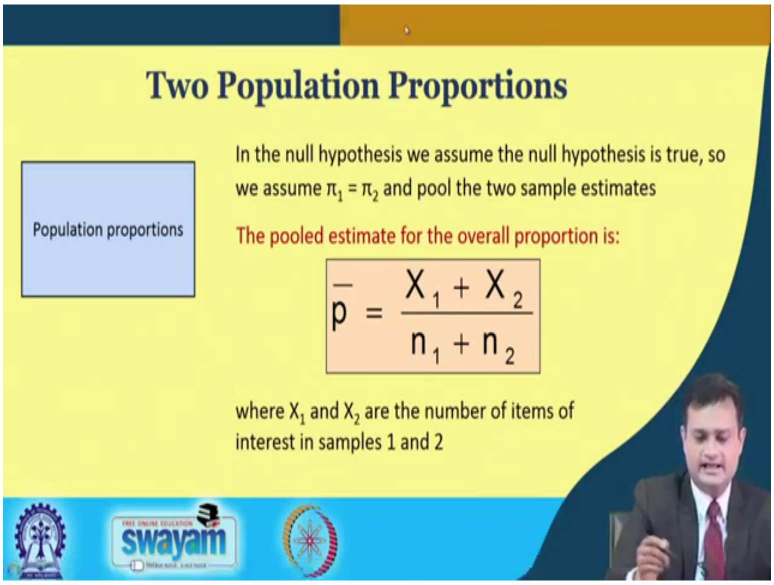
I have case 2 so this is about two population proportion we have already discussed; the case for single population proportion case here there is only a difference that it would be two population proportion.

(Refer Slide Time: 30:25)



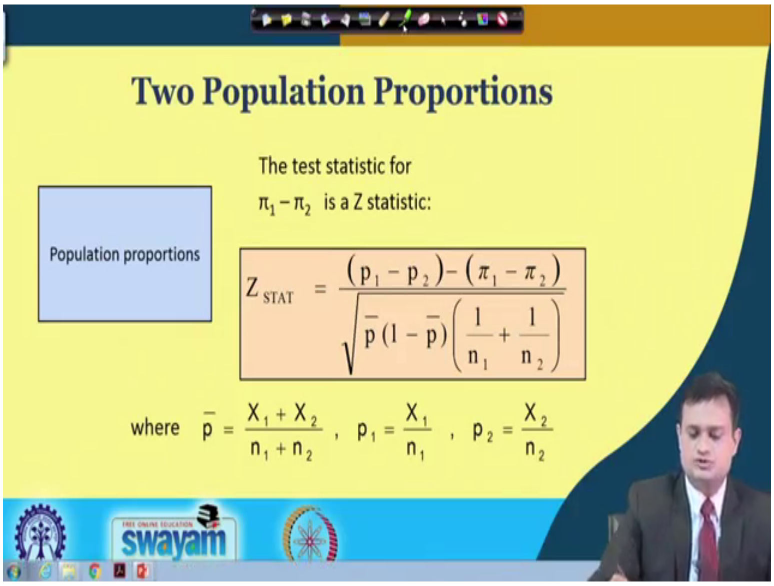
So, we will just little bit do it quickly. So, I have the proportion pi one minus pi 2 and I have n 1 pi 1 greater than equal to 5. Let us say n 2 pi 2 greater than equal to 5. The point estimate for the difference is p 1 minus p 2 and I want to check the difference in population proportions.

(Refer Slide Time: 30:49)



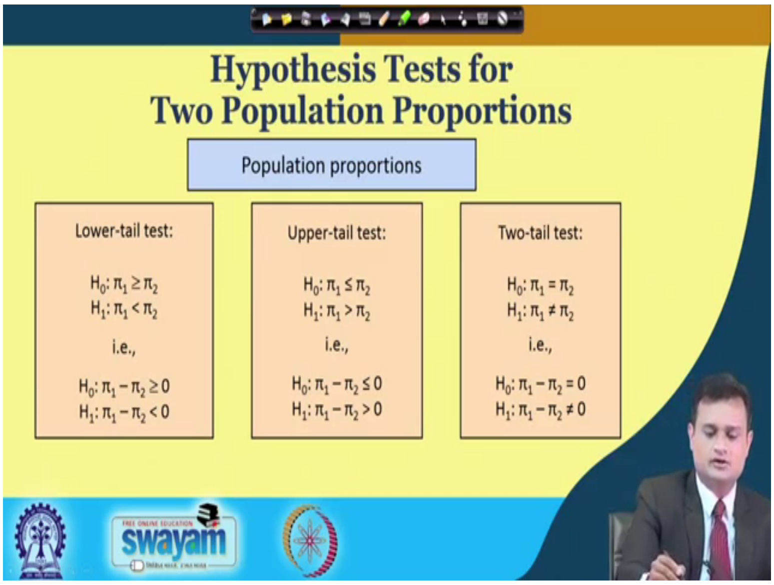
So, this is how I can easily compute my p bar X 1 plus X 2 divided by n 1 plus n 2. And X 1 and X 2 are the numbers of items of interest in sample 1 and sample 2. And I have the sample size n 1 and n 2 so you can easily compute this.

(Refer Slide Time: 31:09)



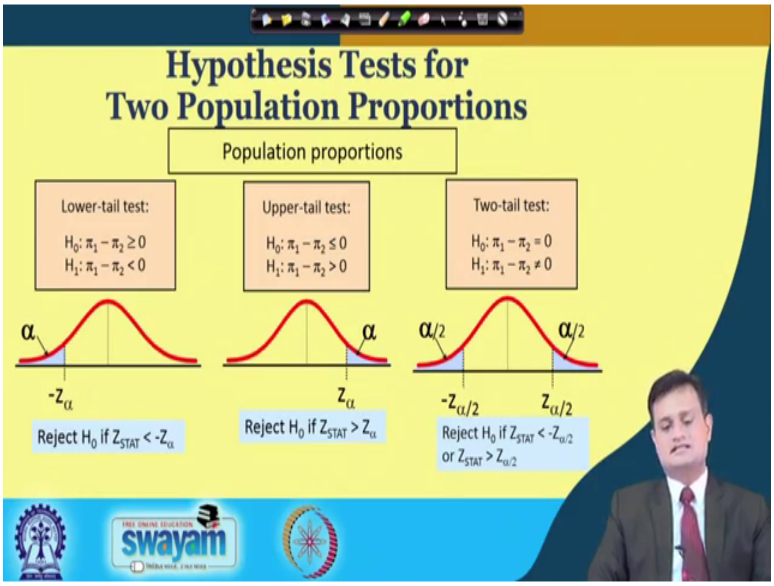
And here I would follow the Z distribution Z statistics for this and various values.

(Refer Slide Time: 31:23)



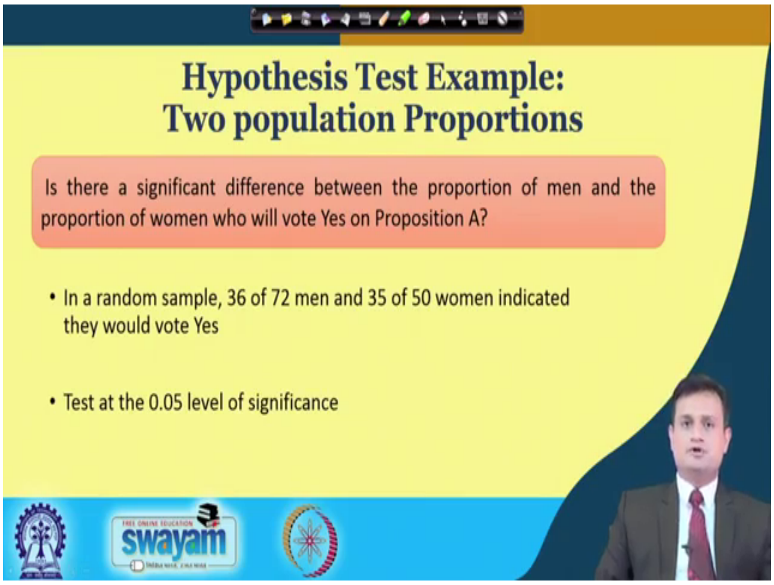
You can compute using these expressions. So, for a typical case maybe you can compute the lower tailed test, you can compute the upper tailed test, you can compute the two tailed test and would like to verify your hypothesis.

(Refer Slide Time: 31:37)



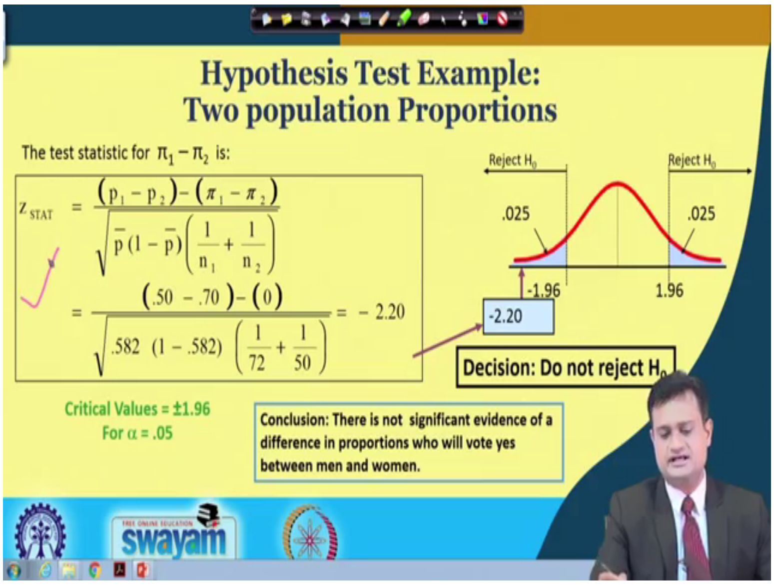
So, the decision rule we already discussed this is just kind of revisit. That if it is a lower tailed test this will be your alpha and this would be your rejection region. This would be alpha this would be rejection region this is divided by alpha by 2 and alpha by 2 and same way the rule follows for proportion population hypothesis testing.

(Refer Slide Time: 31:55)



So, let us see the example is there a significant difference between the proportion of men and the proportion of women who will vote yes on proposition a. Suppose there is some policy and you want to check that will there be a difference on the proportion of yes and proportion of no from men and women two different gender. I am taking a random sample of 36 of 72 men and 35 of 52 women indicated that they would say yes. I want to check my claim at level of significance 0.05.

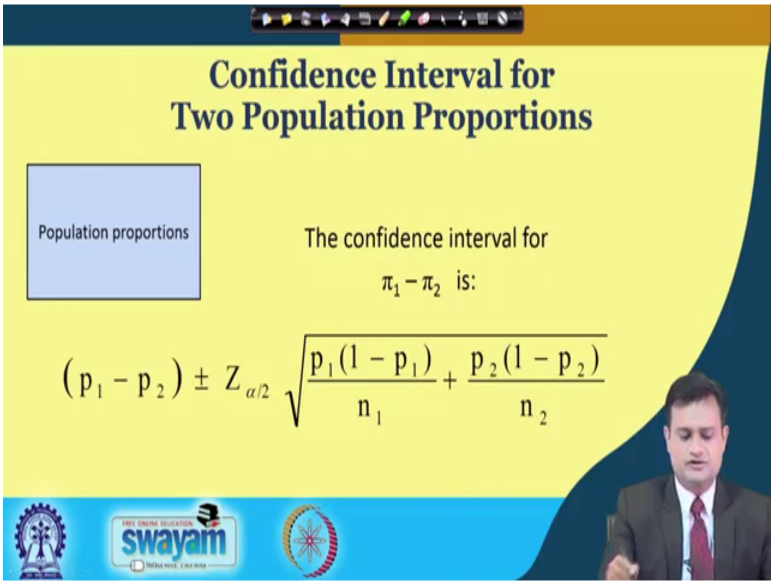
(Refer Slide Time: 32:45)



So, just compute the various values and then when you plug in these values into z statistic; then to check it. And you will see that for this minus point; 2.20 which is my computed value and 1.96 minus and plus it is a two tailed test it is my critical value. So, this particular value minus 2.20 is less than minus 1.96.

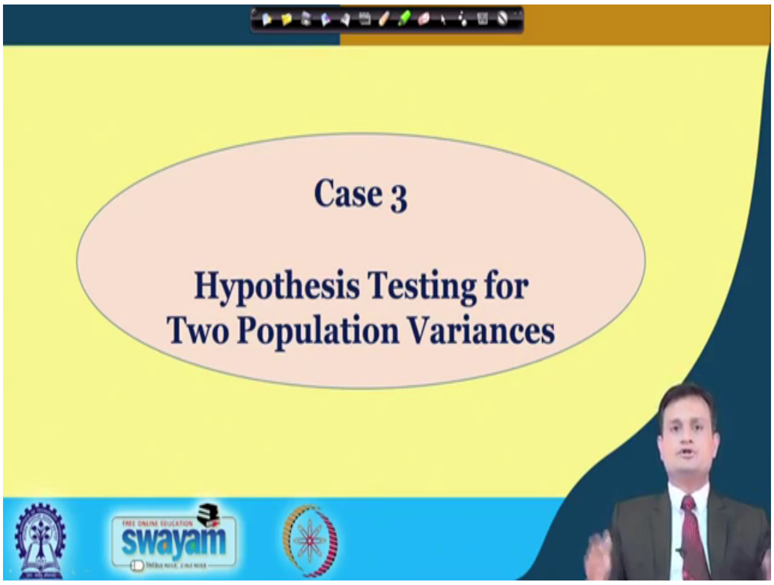
So, it is in the rejection region so there is no significant evidence of a difference in proportion who will vote yes between men and women. So, my null hypothesis issay basically there is not significant difference and I would say my null hypothesis is basically not supported. So, I will say there is not significant evidence of a difference in proportion.

(Refer Slide Time: 33:43)



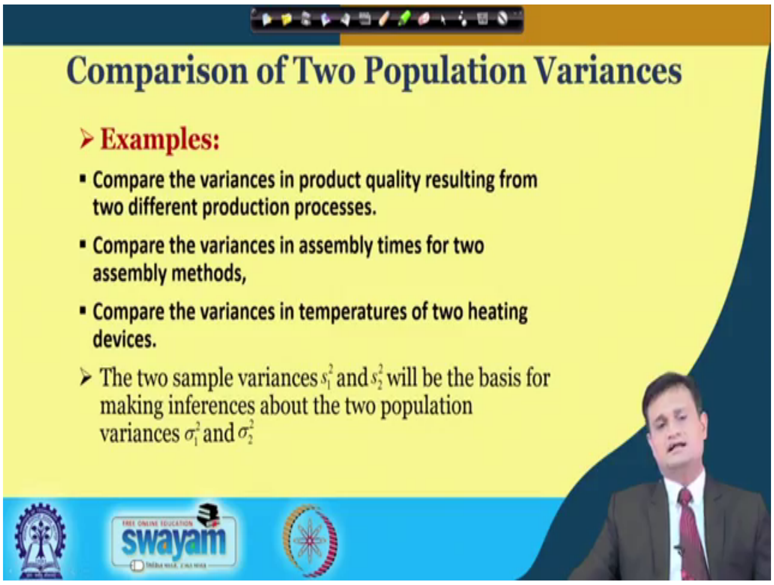
This is the confidence interval as usual you can compute it with Z alpha by 2 here and use this expression. So, this will help you to compute the confidence interval.

(Refer Slide Time: 33:55)



Now, I have another case 3 this is hypothesis testing for two population variances.

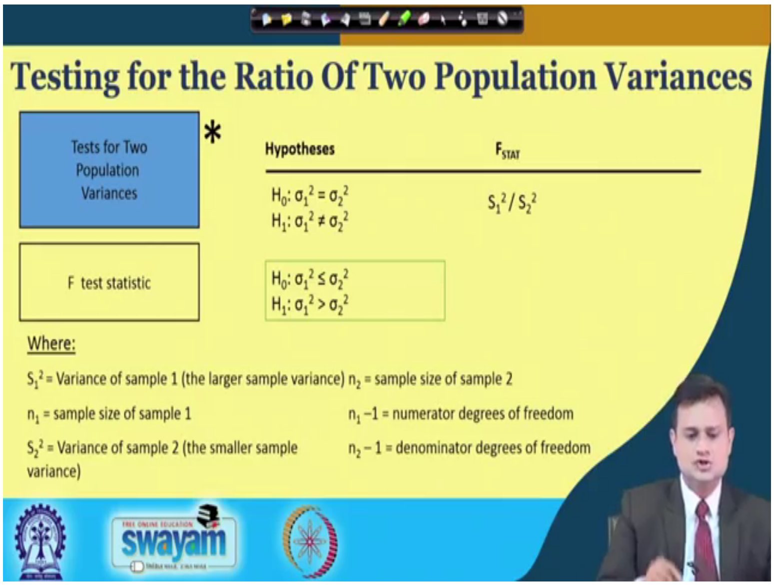
(Refer Slide Time: 34:03)



So, let us see the practical part that what could be the situations in which I would like to go for this. Suppose compare the variance in product quality; so the important parameter for quality is to check the variability and this is resulting from two different production process. I want to see that the variability in quality from process 1 and the process 2 is there really a significant difference or they are same in terms of variability.

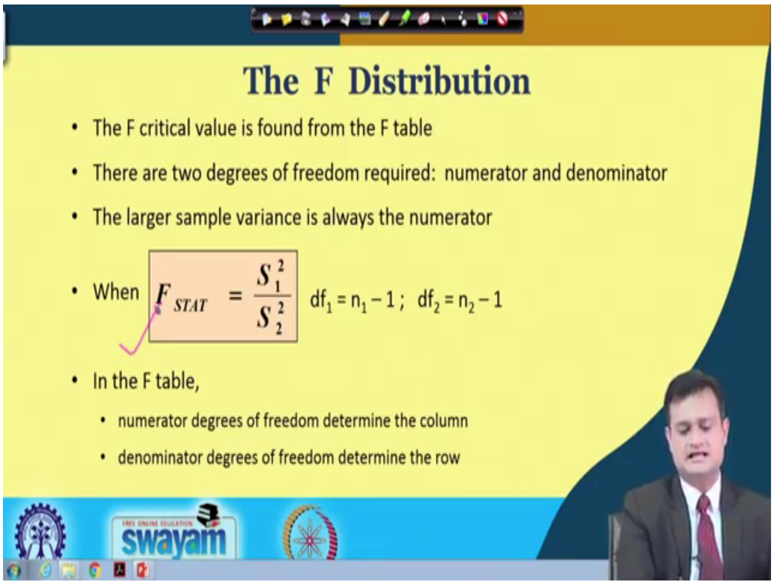
Compare the variance in assembly times for two assembly method compare the variances in temperature of 218 devices. And these are some of the phenomena that could be of interest as a part of case 3. So, two sample variance S 1 square and S 2 square will be the basis for making inferences about two population variance sigma 1 square and sigma 2 square.

(Refer Slide Time: 34:59)



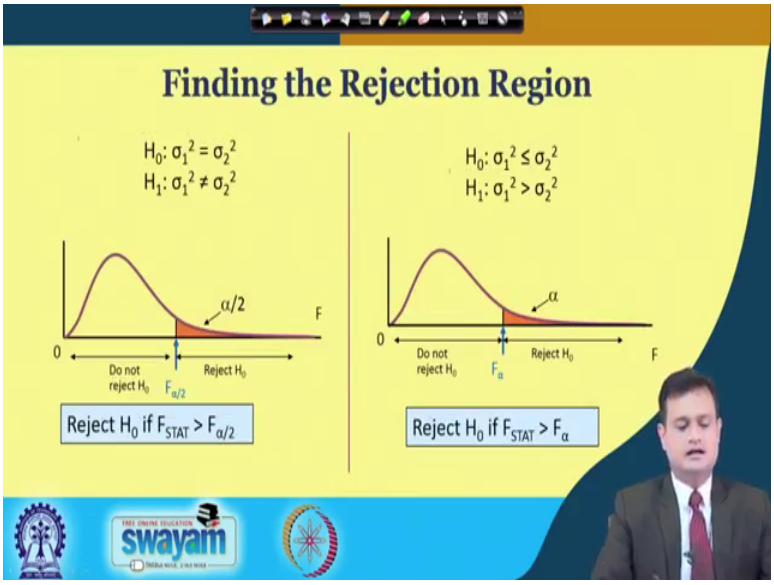
So, this is something explained like this yes the variability is same or it is different express as null hypothesis alternate. And this is my F statistic which is basically the ratio of my sample variance S 1 square divided by S 2 square and this is your numbers of degrees of freedom that you need to refer the particular table F table and find the F statistics. So, you have S 1 square n 1 S 2 square n 1 minus 1 and n 2 minus 1.

(Refer Slide Time: 35:37)



Now, this is my F statistic and these are my degrees of freedom.

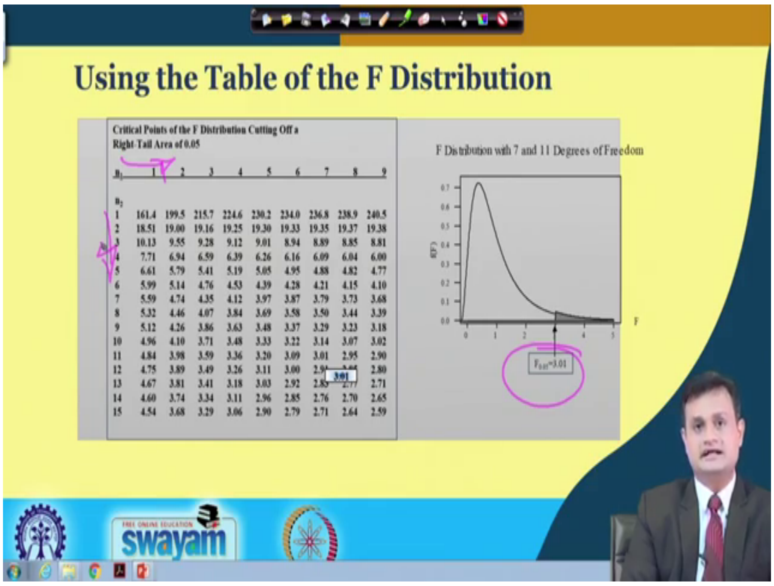
(Refer Slide Time: 35:47)



So, I will have a very simple rule to follow this is my alpha by two this is my alpha and this is my do not reject null hypothesis region this is reject null hypothesis. So, if F stat is greater than F alpha by 2. So, calculated value is greater than F alpha by 2 reject null hypothesis only if stat F stat calculated is greater than F alpha by 2.

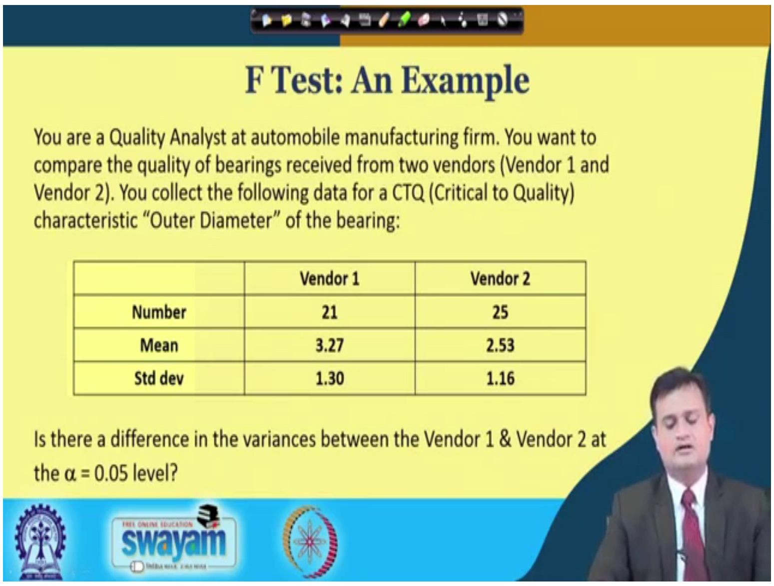
So, it will fall basically in the rejection region. And same way this is reject H 0 this is do not reject H 0 and this is my reject H 0 if F stat is greater than F alpha. Only difference is that here I am using alpha by 2 here I am using alpha. So, this is something that we can do for the for checking the variability.

(Refer Slide Time: 36:39)



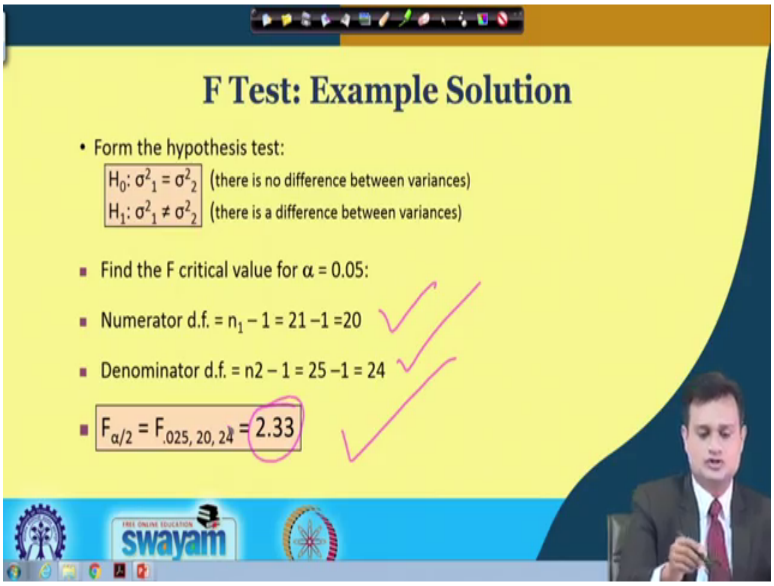
This is my F table you can see that you can find the critical value from the F table. Let us say F 0.05 this is the 3.01 and you have n 1 and you have n 2. So, with this n 1 and n 2 you can figure out that what could be the critical value for your F test.

(Refer Slide Time: 37:07)



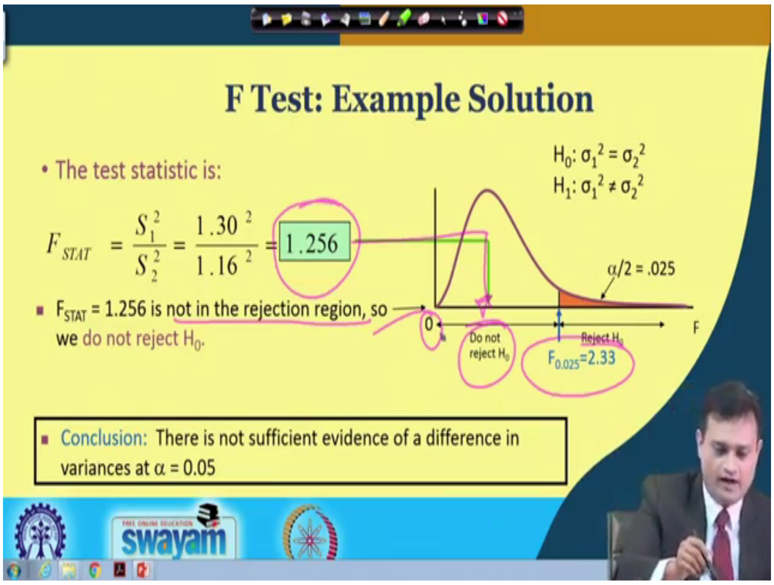
Now, let us say I want to see the example that there are two vendors and quality analyst want to check. That quality of bearing coming from vendor one and coming from vendor two are they same or there is a huge variability in terms of the quality. So, data is vendor 1 vendor 2 number is 21 that is my sample, 25 mean is 3.27, 2.53, 1.30, 1.16 I want to check it at alpha 0.05.

(Refer Slide Time: 37:43)



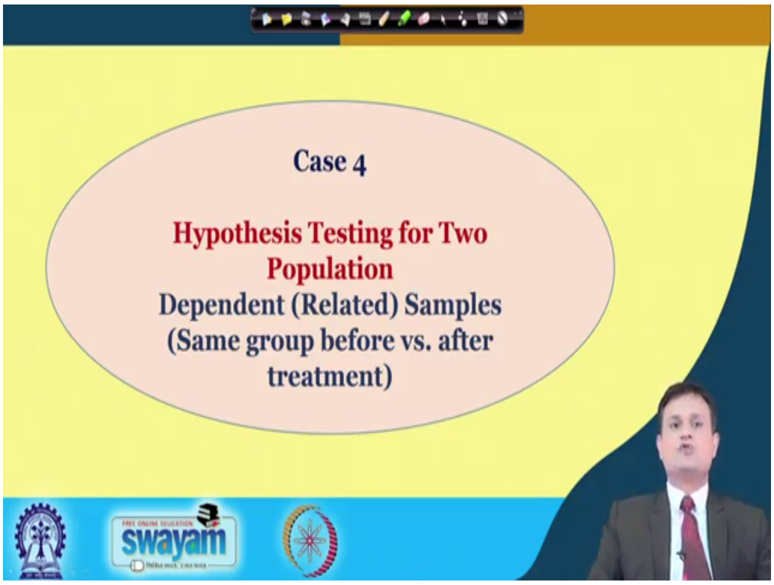
So, we can just compute couple of statistics and we can see that my F alpha by 2 at F 0.025 which is my level of significance alpha by 2. And 20 24 which is my degree of freedom mu 1 and mu 2 it comes out to be 2.33. So, now, what does it mean ?

(Refer Slide Time: 38:05)



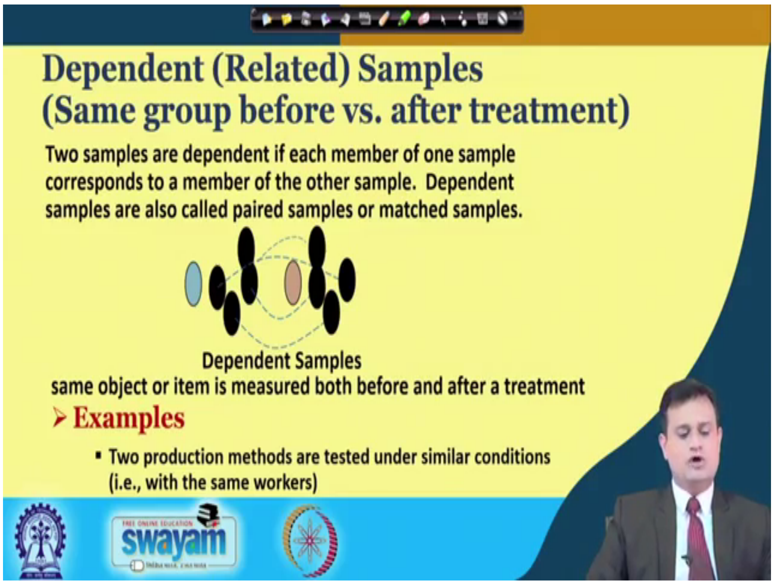
It means that this particular value 2.33 this is my F 0.025. And F stat when I compute it is 1.256. So, 1.256 is not in the rejection region so we do not reject H 0. Because 1.256 will fall somewhere here this is the 0. So, I do not reject null hypothesis it means I do not have sufficient evidence to say that quality coming from or variability in quality coming from vendor 1 and the variability quality coming from vendor 2 they are different more or less they are same. So, I do not have sufficient evidence at alpha 0.05 to say that the quality of vendor 1 and quality of vendor 2 is different. So, fine we have seen the case 3.

(Refer Slide Time: 39:09)



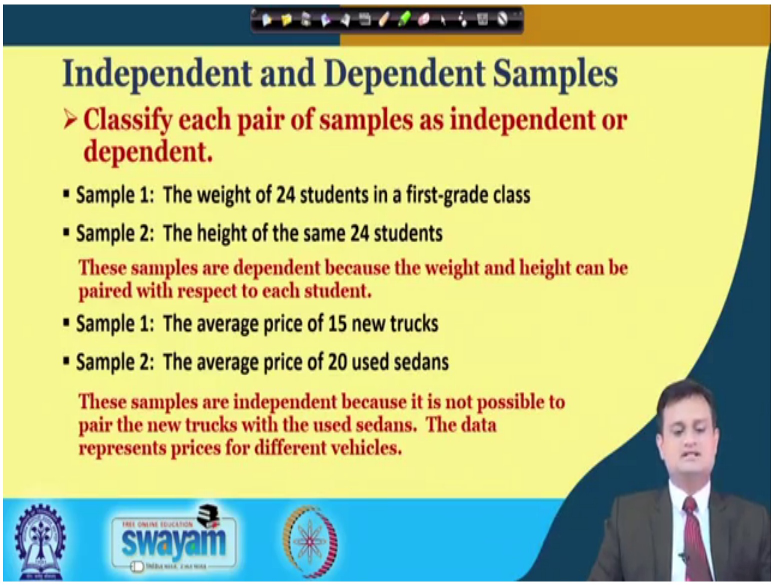
And let us try to end up with one more case that is the case four. So, hypothesis testing for two population dependent related sample. You have the sample patients give them a treatment you are interested to see what is the effect of the treatment before there where a symptoms after treatment what is the reduction in the symptom and you want to check this two situation.

(Refer Slide Time: 39:35)



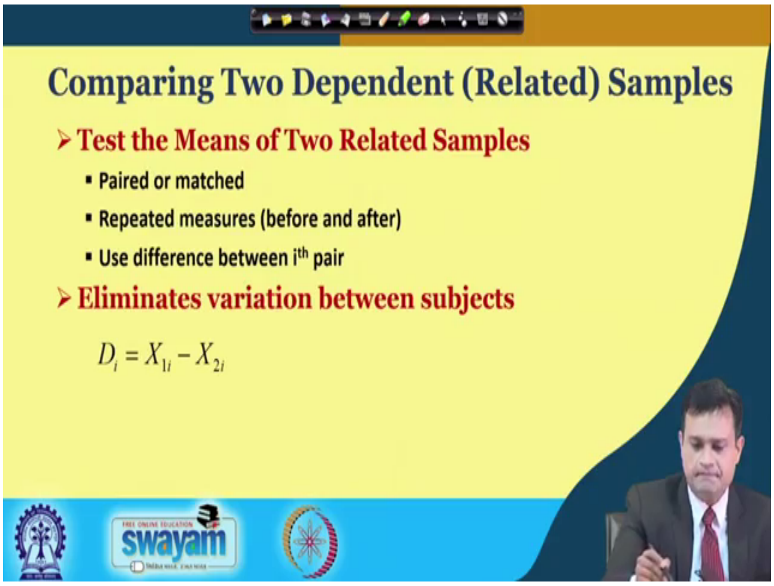
So, this is what it is that two production methods may be are tested under similar condition are they same or they are different.

(Refer Slide Time: 39:43)



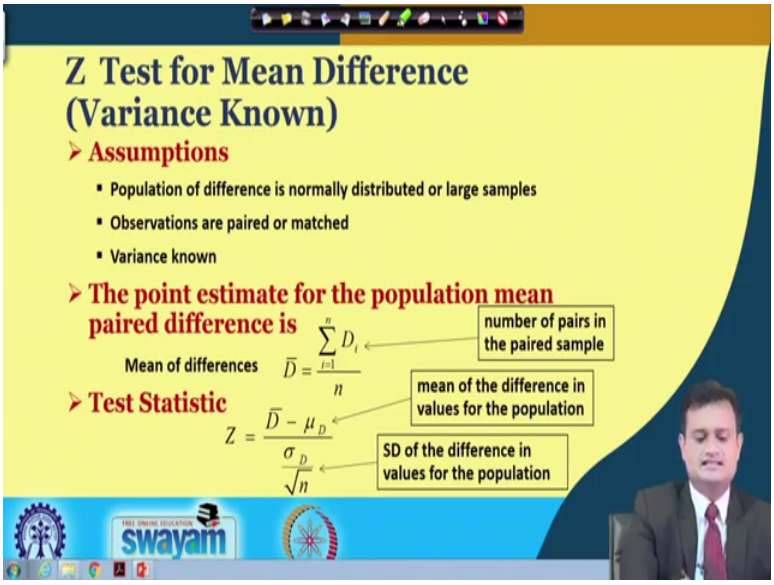
We may have many examples that independent and dependent sample. Suppose sample 1 the weight of 24 students in a first grade class student through the height of the same 24 students. The sample are dependent because weight and height can be paired there could be some relationship. But if you see the second one sample 1 the average price of 15 new trucks sample 2 average price of 20 used sedan these are independent there could not be any dependency or relationship.

(Refer Slide Time: 40:19)



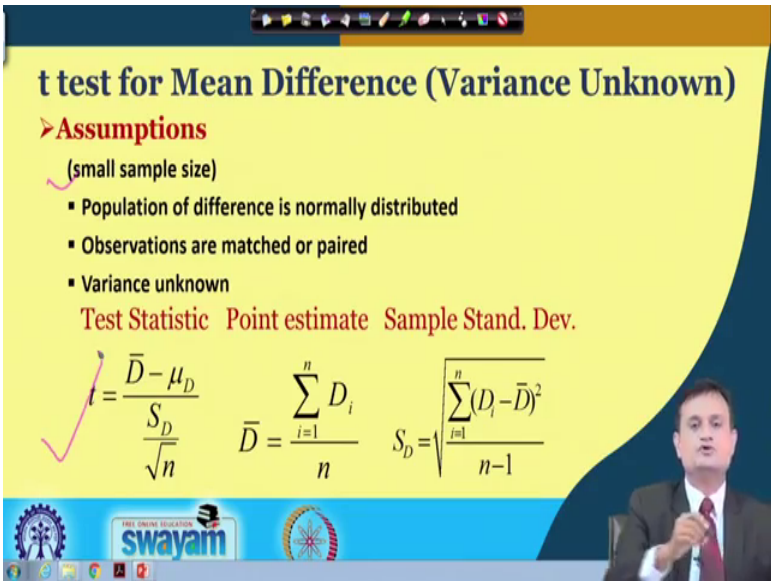
So, basically you have paired or matched two related samples repeated measures. And use difference between i th pair; D i is equal to X 1 i minus X 2 i. So, observation specific to sample 1 or before treatment and X 2 i after treatment and you will ; obviously, take number of sample to test your claim.

(Refer Slide Time: 40:45)



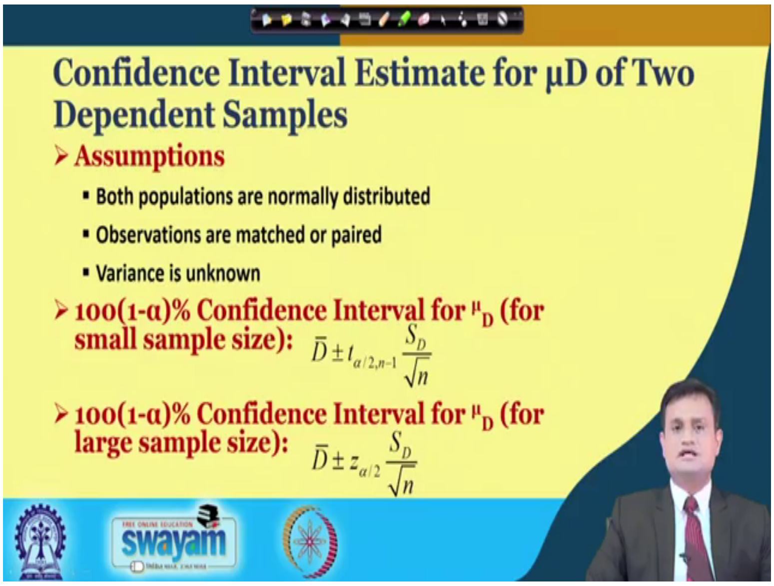
So, you can compute couple of things like D bar this is basically D i. So, number of pairs in the paired sample then D bar minus mu D divided by this. Same procedure same computation as we do for z statistics.

(Refer Slide Time: 41:11)



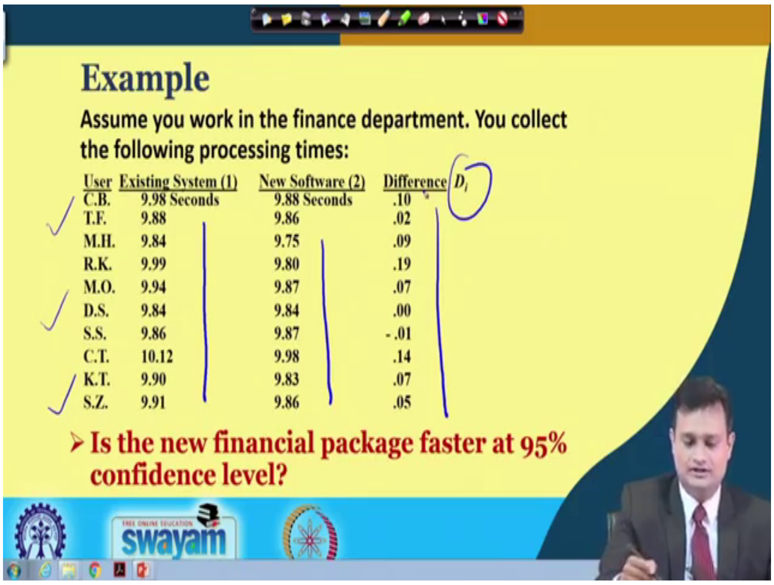
So, you will suppose you have small sample. Then definitely you will use the logic that you will use student t distribution by plugging in the values of X bar minus mu divided by sigma. Here it is D bar minus mu D divided by sigma that is S D divided by square root of n and D bar is this your S D is this. So, these are the standard expressions.

(Refer Slide Time: 41:37)



We have the assumptions that both populations are normally distributed observations are matched or paired variance is unknown. In this case you can again do the confidence interval analysis and construct it and you can find 100 1 minus alpha percent confidence interval.

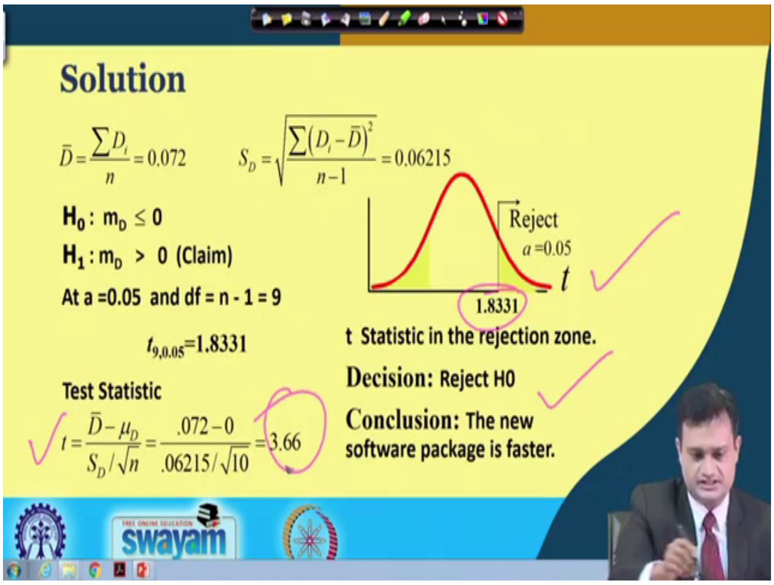
(Refer Slide Time: 42:03)



Now, let us say I have the data that you work in the finance department and collect some data regarding the existing system and the new software. And suppose I have let us say some user preference some user rating and let us say this is the existing system in terms of seconds experience or the running time.

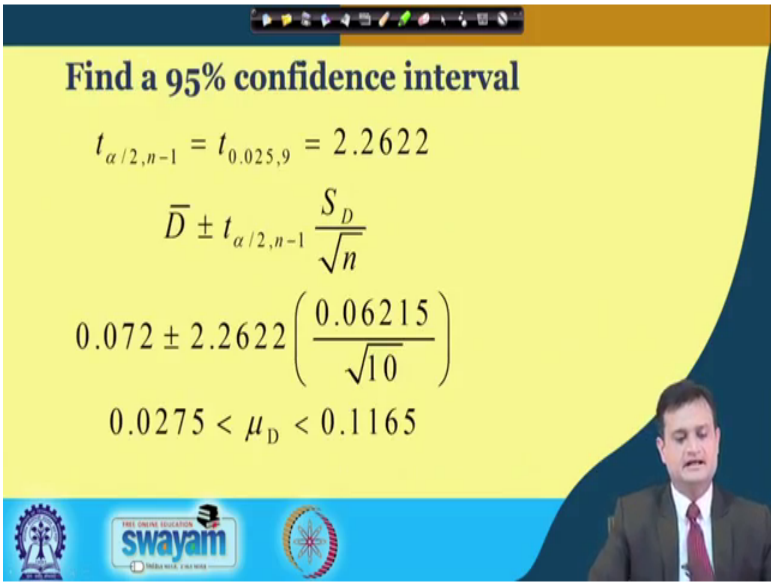
This is the new software running time I am finding the D i value that is my difference. I want to check that is the new financial package faster that 95 percent confidence level. It is a huge investment in software you would like to see that whether the new system purchased is really worthy or not.

(Refer Slide Time: 42:55)



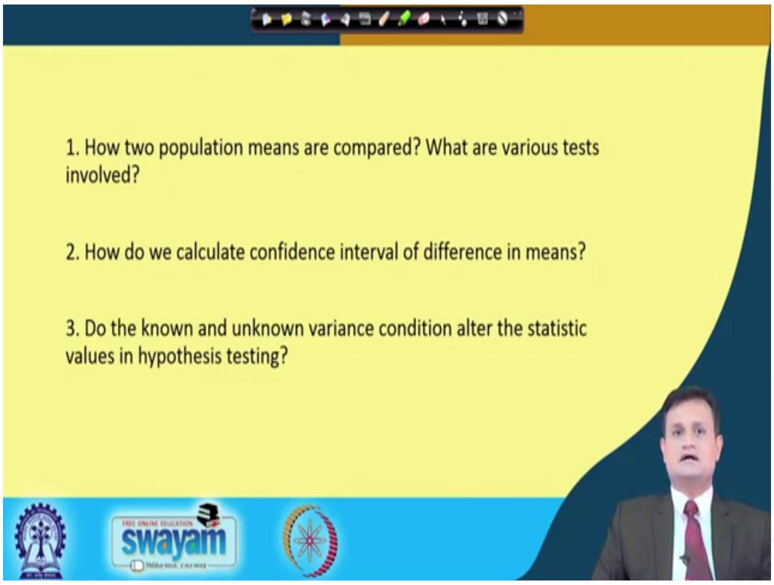
So, I would test the hypothesis find the t distribution t statistics. And for this t statistics I will compute the value and this would be compared with the critical value the procedure remains same. And I will say reject H 0 because this is 3.66. So, this falls in the rejection region and the new software package is faster.

(Refer Slide Time: 43:27)



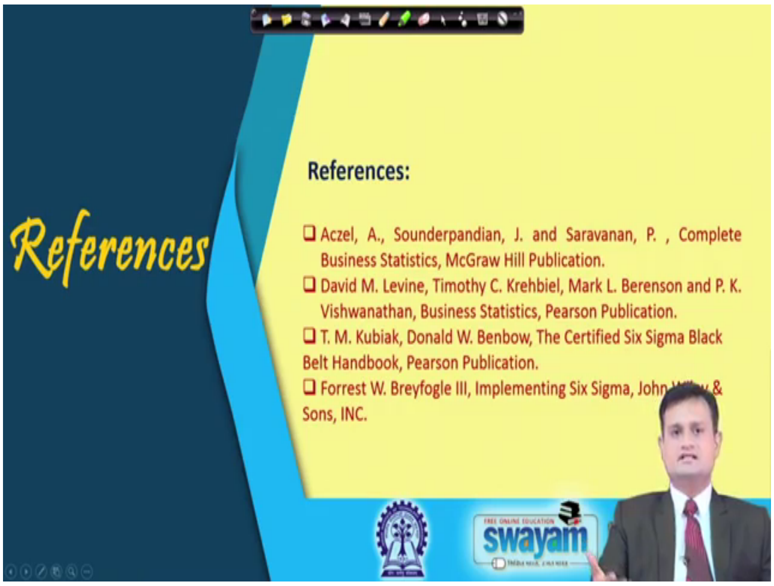
So, with this is what we basically try to do. and you can construct the confidence interval same way by considering plus or minus t alpha by 2 at n minus 1 degree of freedom.

(Refer Slide Time: 43:37)



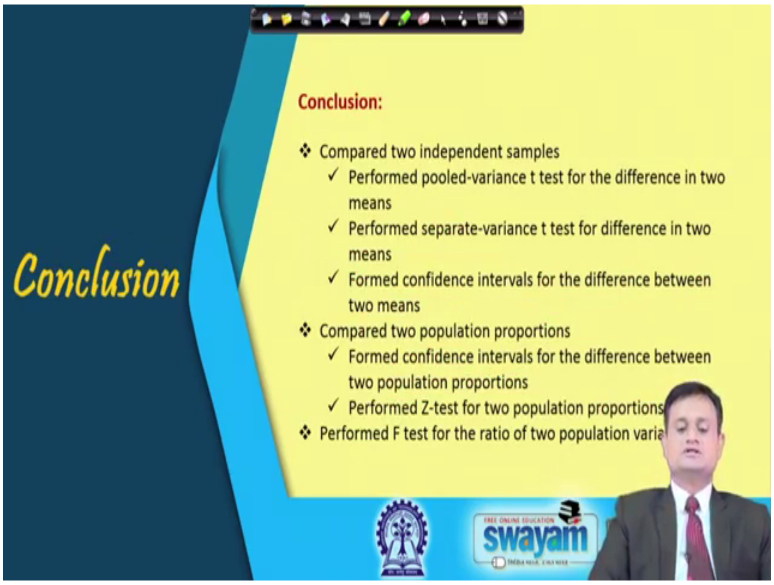
Let us end this section with couple of think it question. How two population means are compared? What are various tests involved? How do we calculate confidence interval for difference in means and do the known and the unknown variance condition alter the statistics value in hypothesis testing?

(Refer Slide Time: 43:57)



Please refer this reference if you find this session little bit difficult to understand. And I hope this session would provide a good conceptual understanding to digest the concept and see the importance of two population test.

(Refer Slide Time: 44:13)



So, these are the cases basically we have discussed for two population test, two independent sample, compared two population proportion variances and before treatment and after treatment. So, thank you very much for your interest in learning two population hypothesis testing; keep revising, be with me, enjoy.