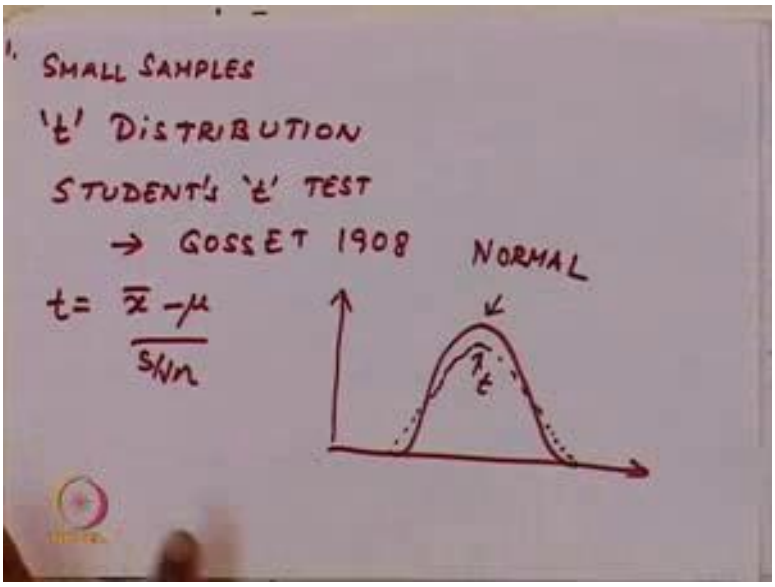
Introduction to Biostatistics
Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture - 36
1 tailed and 2 tailed T-distribution, Chi-square test

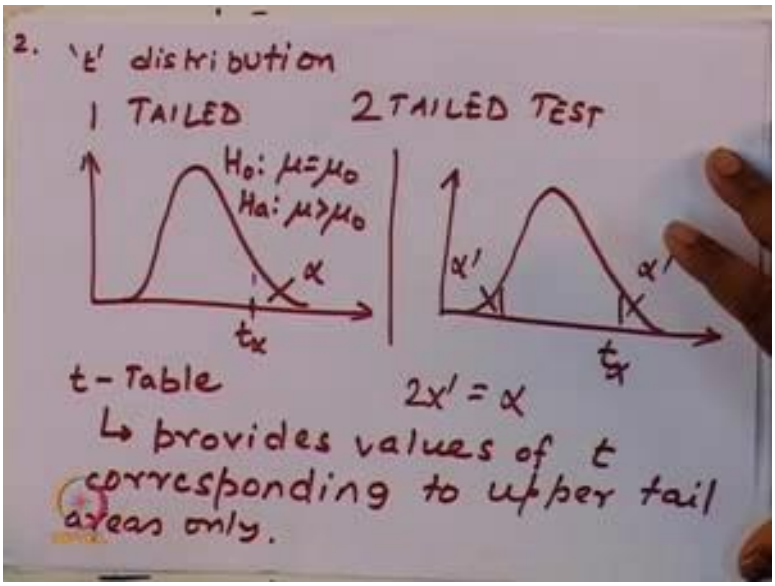
Hello and welcome to today's lecture. So, in the last class we had introduced the student's t test for small samples.

(Refer Slide Time: 00:26)



So, for small samples you can no longer use the z statistic, but you use the t distribution. This is called students was also referred to as students t test. This was introduced by Gosset in 1908. So, what you do is you calculate the t statistics given by this expression, and what you will find compared to the normal distribution, the t distribution is much flatter, but as your n increases the t distribution approaches the normal distribution. So, this is your t and this is your normal distribution.

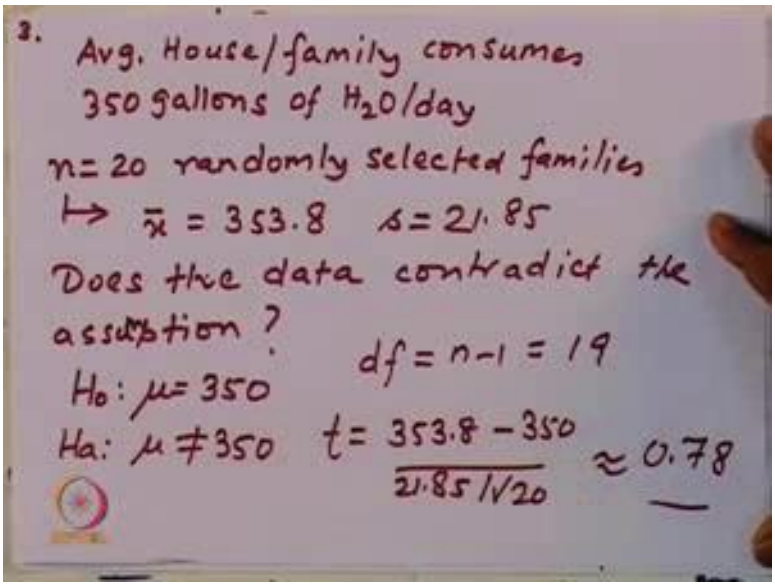
(Refer Slide Time: 01:42)



And just like for the normal distribution you had one tailed or 2 tailed test. For the t distribution also you have either one tailed or 2 tailed test, here as before you calculate. So, this is your t of alpha where alpha is the area under this curve and your hypothesis is mu equal to mu naught and H a is mu is greater than mu naught. For the 2 tailed test you have 2 tails; however, please note for the t table it only gives for the upper half, which means if I want to calculate the this alpha if this is my alpha prime, let us say I have to set 2 alpha prime equal to alpha.

So, let us take a sample example. So, let me again assert. So, your t table t table only provides the values. It only provides the values of t corresponding to upper tail areas only. So, let us take an example to see it.

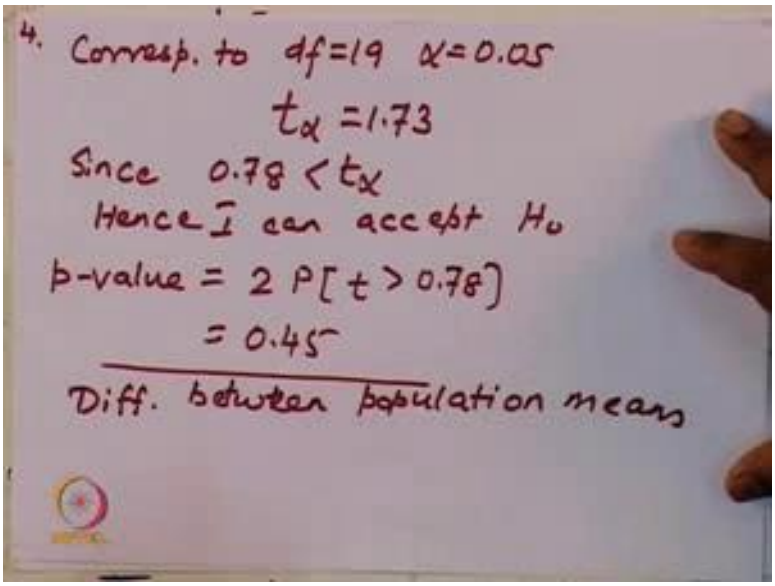
(Refer Slide Time: 03:41)



So, imagine, you it is estimated the average house or family consumes 350 gallons of water per day of water per day. So, you have taken a small sample of 20 randomly selected families. And for these 20 you have calculated a mean x bar of 353.8 and s of 21.85.

So, the question is does the data contradict the assumption. So, you have your H naught is mu equal to 350 your H a is mu naught equal to 350. So, I can calculate my d f is n minus 1 is a degrees of freedom equal to 19 and the value of t is 353.8 minus 350 by 21.85 divided by root of 20 is approximately 0.78.

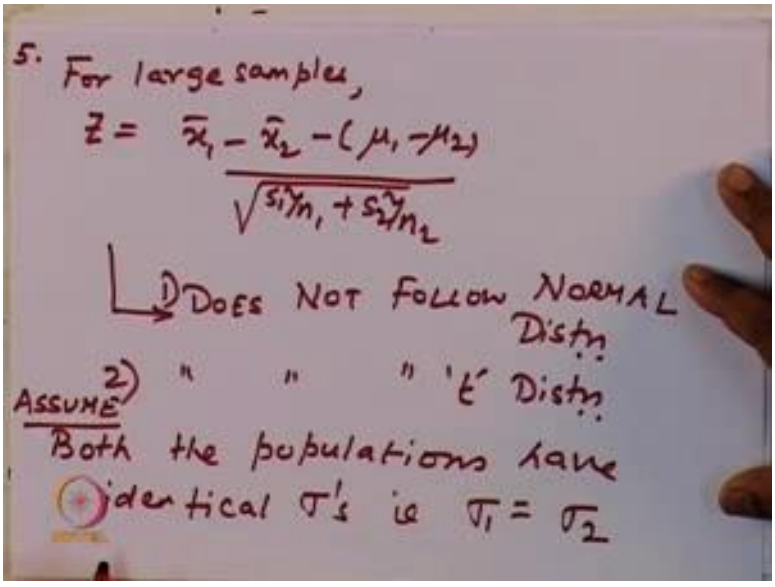
(Refer Slide Time: 05:40)



So, corresponding to d f equal to 19 and alpha equal to 0.05, you can find out t of alpha is 1.73. So, since 0.78 is less than t of alpha, hence I can accept H naught. So, for calculating the p value, I have to write it as 2 of probability of t greater than 0.78. Why this 2? Because I only get the upper tail information from a t table, this comes out to be 0.45 this again asserts that this difference is not statistically significant.

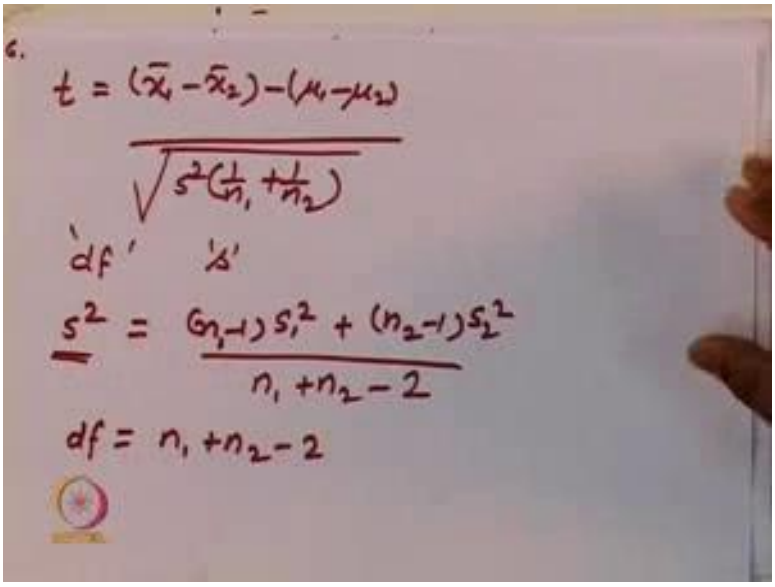
So, using a t test I can do population difference between population means. So, how do I do it you know that I can calculate?

(Refer Slide Time: 07:11)



For large samples, you calculate the test statistic as x 1 bar minus x 2 bar minus of mu 1 minus mu 2 by root of s 1 square by n 1 plus s 2 square by n 2. Now what has been what is known is this test statistic for small samples does not follow normal distribution one, and Secondly, it does not follow t distribution. So, the only way to make use of t distribution is to make one particular assumption that both the populations have identical. So, this is the assumption. So, if you assume that both of identical sigma's that is sigma 1 equal to sigma 2.

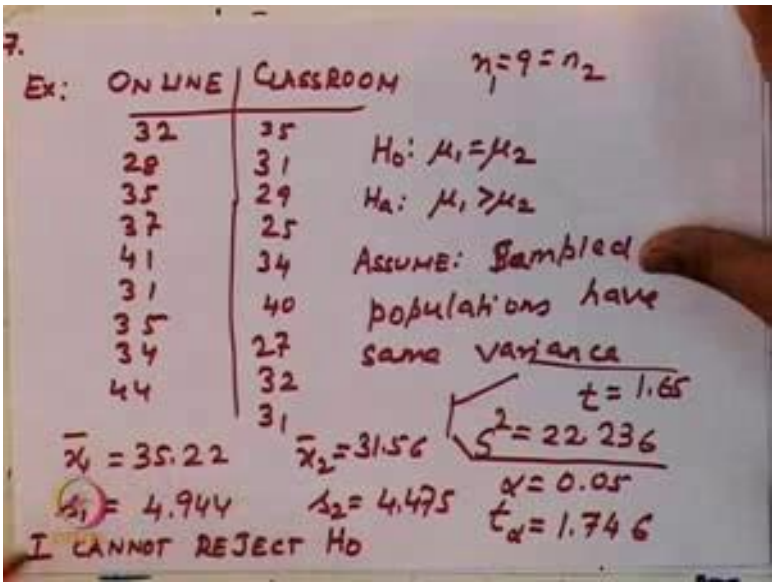
(Refer Slide Time: 08:47)



Then you can calculate the t statistic, in that case you can calculate the t statistic as x 1 bar minus x 2 bar minus of mu 1 minus mu 2, by root of s square into 1 by n 1 plus 1 by n 2. So, because your population standard deviations are same, you assume similar standard deviations for both the samples.

So, this for calculating the t statistic, I need to calculate d f and I need to calculate s. So, how do I do it? So, you estimate the sample standard deviation using the following expression. And you set d f equal to n 1 plus n 2 minus 2. So, n 1 minus 1 into s 1 square will give you the population the sample variance, same for the sample variance for this, and then you average it by n 1 plus n 2 minus 2 to estimate the average sample variance for both the samples.

(Refer Slide Time: 10:14)

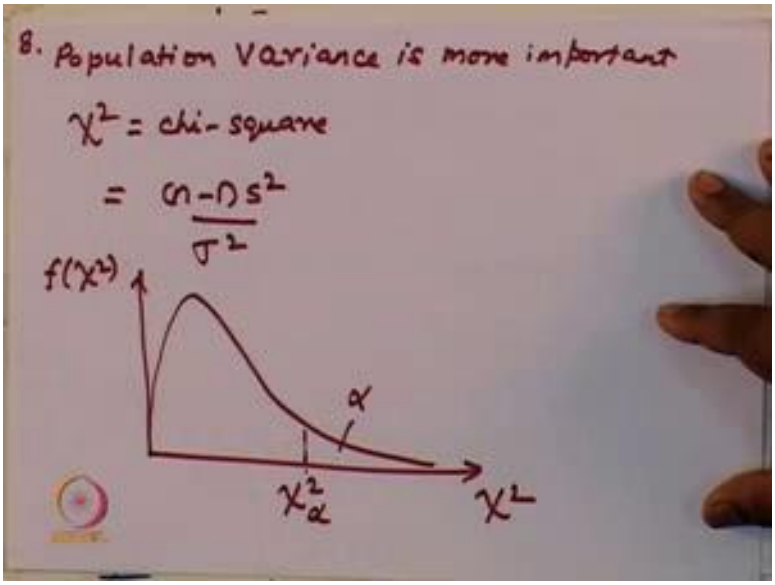


So let us take an example. So, imagine an examiner wants to assess whether students would perform similarly odd differently if the exam was online versus if the exam was classroom. So, here the exam is held in the classroom, here the exam is online. And he has chosen n equal to 9 samples for both of them for he has taken 9 students, who have taken online exam and 9 students have taken classroom exam. So, their marks as are as follows and we want to know whether the performance of students taking the online exam is identical or different than those when they give the classroom exam.

So, my null hypothesis is mu 1 equal to mu 2. And my alternative hypothesis is mu 1 let us say greater than mu 2. So, we have to what do we assume? You have to assume that populations that the sampled populations have same variance. So, with this assumption, I can calculate. So, for this I can calculate x 1 bar as 35.22 and s 1 as 4.944. For this population accordingly I can calculate x 2 bar as 31.56 and s 2 as 4.475. Now using this expression, you know n 1 and n 2 both of them are 9, as with this expression you have calculated s 1 and s 2. So, you want to calculate s. So, using this expression you will get s square equal to 22.236, and for at significance level of 0.5 for alpha equal to 0.5 you will calculate t of. So, s square is 2 22.236 and t is 1.65. So, this is what you calculate and for 5 percent, you know 0.5 percent is alpha equal to 0.5 you can calculate t of alpha as 1.746.

So, what you can clearly see is because this value is less than this value I cannot reject H naught. So, I cannot reject H naught. So, I need to increase the sample size in order to say anything about whether the performance remains same or it changes. So, that brings us to a close of our discussions on population mean and doing test of hypothesis for population mean, but in many cases population variance might be more important.

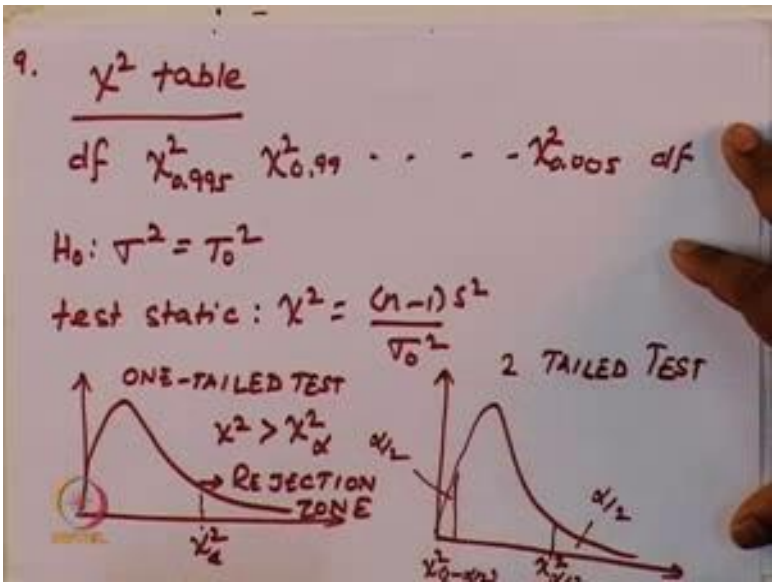
(Refer Slide Time: 14:00)



So, for doing for saying anything about whether 2 samples have same population variance or different population variance, you use the test statistics as something called a chi square statistics.

So, you this is chi square and chi square is given by the expression n minus 1 into s square by sigma square. So, s is your sample variance sigma is your sigma square is your population variance. And if you plot this the chi square curve is asymmetric. So, chi square curve looks something like. So, just like any other distribution you can calculate chi square according to alpha, where alpha is the area under this curve. So, how does a chi square table look like?

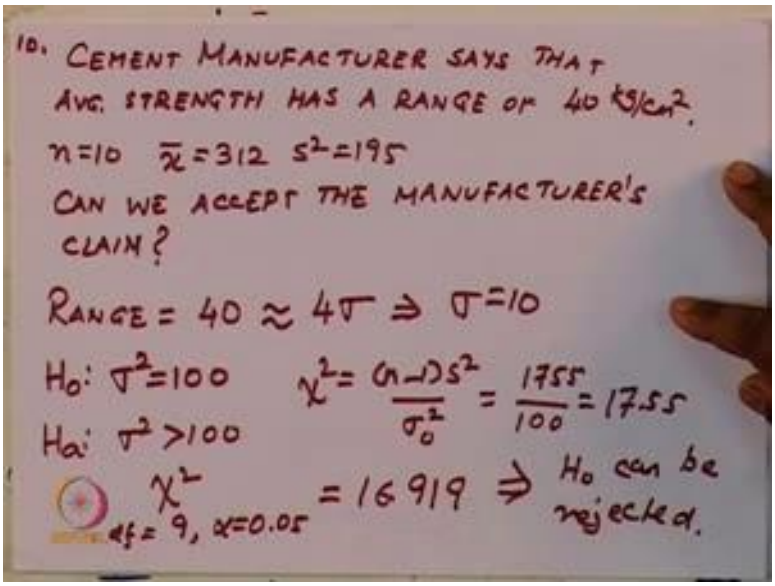
(Refer Slide Time: 15:26)



So, in a chi square table, you have the following values. You have d f which is your degrees of freedom n minus 1. And you have these values provided. So, if you want to test your population variance is certain value. So, you can have your null hypothesis as sigma square equal to sigma naught square, and you calculate the statistics. So, your test statistic is chi square equal to n minus 1 into s square by sigma naught square. For a one tailed test, so you have to accept. So, your hypothesis is not true if your chi square is greater than chi square of alpha.

So, this is your rejection zone or rejection criteria for a 2 tailed test for. So, this area is alpha by 2 and this area is alpha by 2 also. So, for 2 for a 2 tail chi square test, this is 2 tailed test. So, I have these 2 zones which fall within the rejection zone. Let us take an example.

(Refer Slide Time: 17:45)

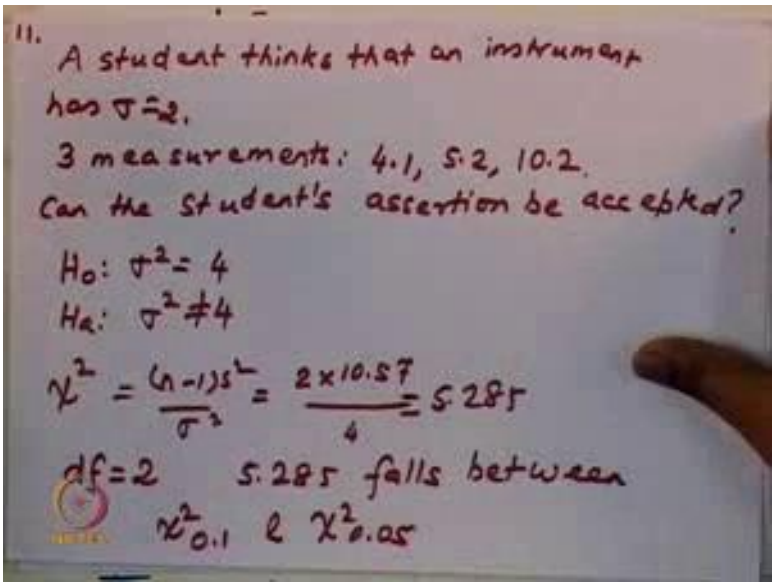


So, let us say a cement manufacturer says that average strength of the cement that he makes has a range of 40 k g per centimeter square. So, to test it n equal to 10 samples size has been drawn and based on this you have obtained an x bar value of 312 and an a square value of 195. So, can we accept the manufacturers claim?

So, what you are given is a range is 40. And you can use the approximation range is approximately equal to 4 times sigma. So, this gives you sigma is equal to 10. So, your null hypothesis is sigma square equal to hundred and your alternate hypothesis is sigma square. Let us see greater than hundred. So, you can calculate the chi square statistics which is n minus 1 into s square by sigma square sigma naught square, which is 1755 by 100 you get a value of 17.55. Or the chi square for a degree of freedom of 9, for d f equal to 9 and significance of alpha equal to 0.5.

If you look up the table, you get a value of 16.919. So, what you clearly see is this value chi squared value is greater than this threshold. So, this would mean that H naught can be rejected.

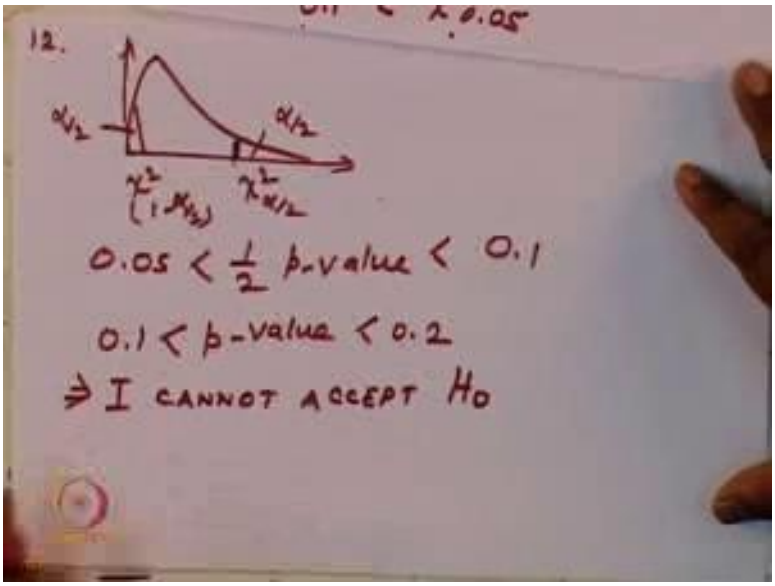
(Refer Slide Time: 20:36)



Let us take one more example. So, a student thinks that an instrument has sigma equal to 2. So, and she has done 3 experiments 3 measurements, which is led to the values of 4.1 5.2 and 10.2 respectively. So, can the students' assertion be accepted?

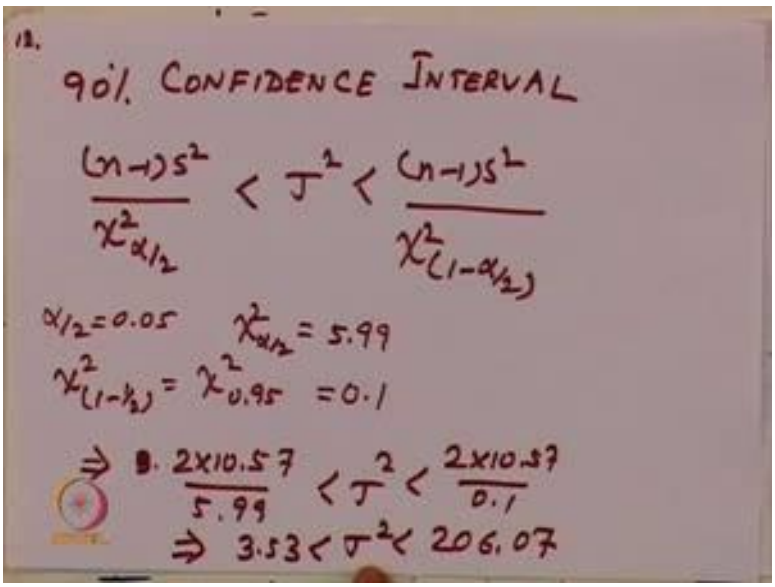
So, what do we do our H naught is sigma square equal to 4. Alternative hypothesis is sigma square not equal to 4. So, my chi square value is n minus 1 into s square, by sigma square. This gives you 2 into 10.57 is what you calculate the s square s by 4 is 5.285. You get a value of 5.285. So, for d f equal to 2, this value 5.285 falls, so 5.285 falls between chi square 0.1 and chi square 0.5.

(Refer Slide Time: 22:54)



Now, since it is a 2 tailed test, now remember for a 2 tailed test, so this area is alpha by 2, this area is alpha by 2. So, chi square 1 minus alpha by 2 and chi square alpha by 2. So, I can say that this p value that I calculated that 0.5 less than half of p value, less than point of 1. Why? Because this value falls within pi square of 0.1 and chi square of 0.25, but this range, this half is only giving you half of the total p value. So, the total p value ranges between 0.1 and 0.2. This implies that I cannot accept. So, by p value is greater than 0.1, implying I cannot accept.

(Refer Slide Time: 24:16)



You can also calculate the 90 percent confidence interval, which is given by, so for alpha by 2 equal to 0.5 chi square of 1 minus alpha by 2 equal to chi square of 0.95. So, for alpha by 2 I can have chi square of alpha by 2 as 5.99. And chi square 1 minus half is chi square of 0.99 equal to 0.1. So, this if I plug in the values here, what I get is a range between you can plug in the values 2 into 10.57 by 5.99 less than sigma square less than 2 into 10.57 by 0.1. This gives me a range 3.53 less than sigma square less than 206.07. So, what you clearly see is this variance sigma square has a wide range. This tells you that based on 3 measurements. You really cannot say whether your standard deviation is 2 or not.

With that I conclude our class for today. And today we had briefly discussed about how you can do the students t test. Briefly you taken a call on student's t test using this test at 60 which we found that for small sample sizes. It is deviates from a normal distribution, but for large sample sizes you will get the same result. And then towards the latter half we discussed about how we can calculate chi square statistics to measure population variance with that.

I thank you for your attention.