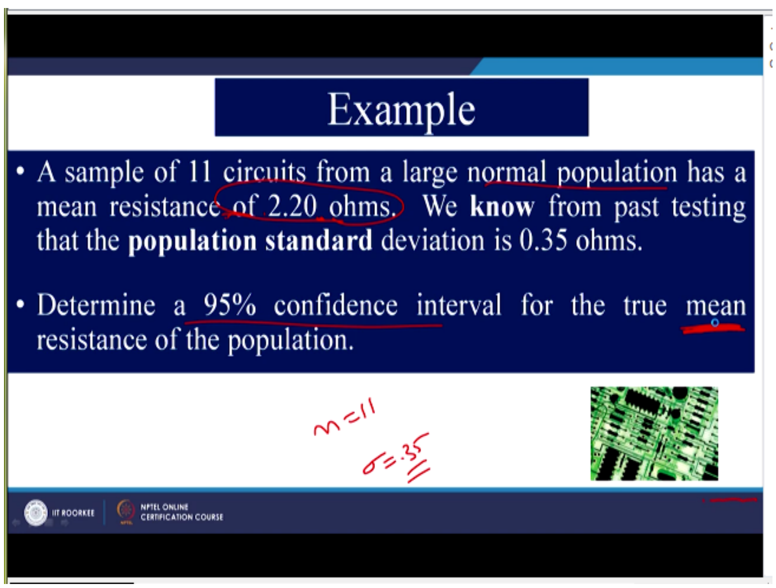
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-27
Interval Estimation

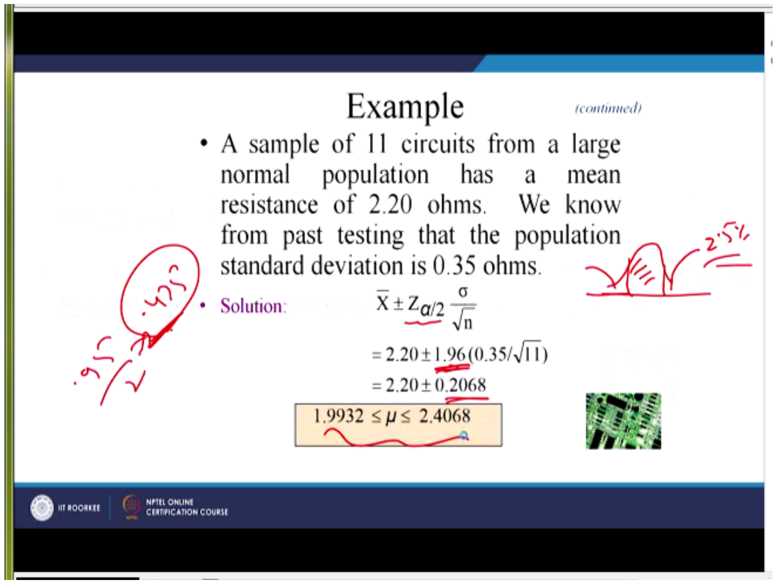
Good afternoon friends, I welcome you all in this session as you are aware in previous session we were discussing about estimation and we have seen point estimation and interval estimation. It is always good to have interval estimation because it gives you range, so there is higher probability that your estimation would be within a given range. So let us workout couple of examples related to estimation, so let us take a sample of a 11 circuits from large normal population.

(Refer Slide Time: 00:59)



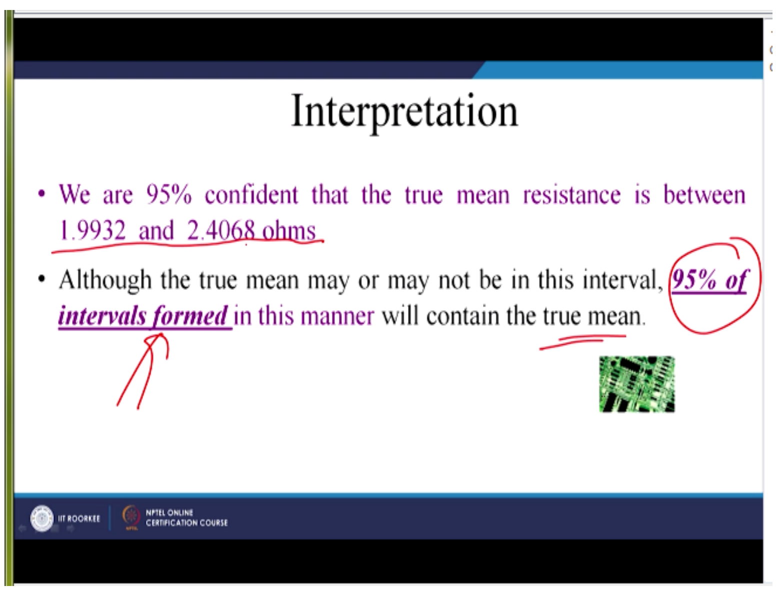
And has a mean resistance of 2.2 ohms, so mean is given, so n is 11 right and mean is 2.20 right. We know from the past testing that the population standard deviation, so sigma is also given right is 0.3 and you have been given this is nothing but their population mean right. Determine a 95% confidence interval for the true mean resistance of the population, so we are looking for population mean ok and in fact this is your sample mean right, this not population mean this sample mean. We have to estimate population mean right.

(Refer Slide Time: 01:47)



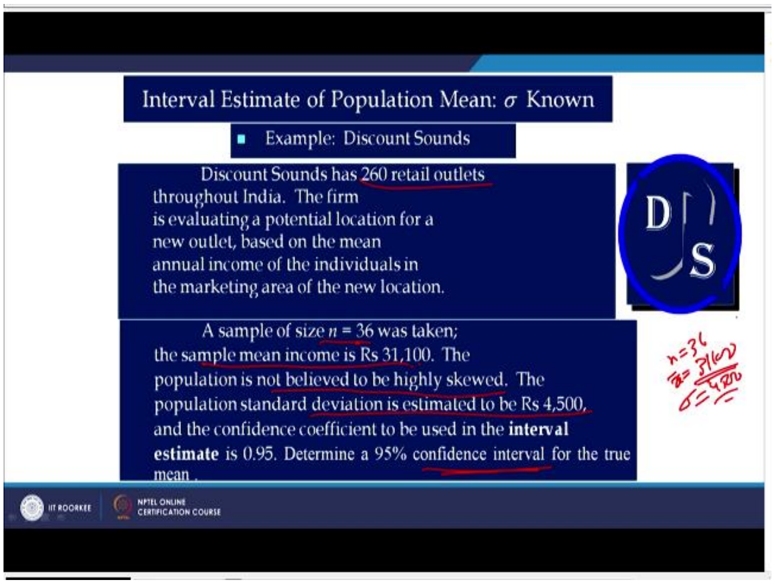
So, this how you should be proceeding, this X bar ± Z alpha/2 right, so this is your distribution right and this is in fact this is 95% right this is point 2.5% and this is also 2.5 %. So, look at the value of Z, from Z table and at 95% as I said it is 0.95/2, so 0.475. So, look at the Z table where probabilities 0.475 and we will find that it is a 1.96 ok. So, 2.20 ± 0.20 right, so this your interval right ok.

(Refer Slide Time: 02:44)



So, what does it mean, so we are 95% confident that the true mean resistance is between 1.99 ohms to 2.40 ohms, again the true population mean may not be in this range. But what we are saying is that 95% of the intervals formed in this manner will contain true mean. In this manner let us take one more sample of sample size 11 right calculate its mean and again come up with this confidence interval. So, when you frame these confidence intervals, so 95% of the intervals will have population mean.

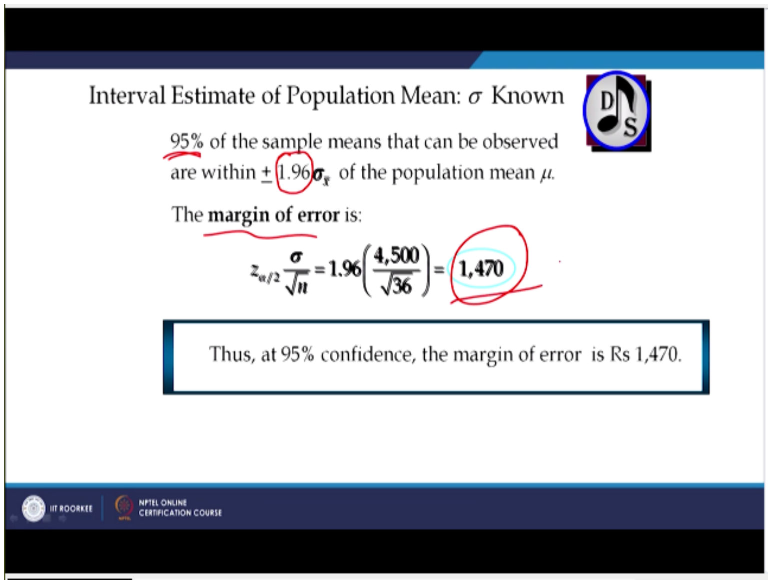
(Refer Slide Time: 03:32)



Let us look at one more question, so discount sounds has 260 retail outlets throughout India. The firm is evaluating to open up a new outlet, so based on the income of the people the firm is decided to go for study. When the firm selected 36 households and it was found that the sample mean income was 31100 rupees. The population mean is not believed to be rightly skewed the population standard deviation is estimated to be 4500.

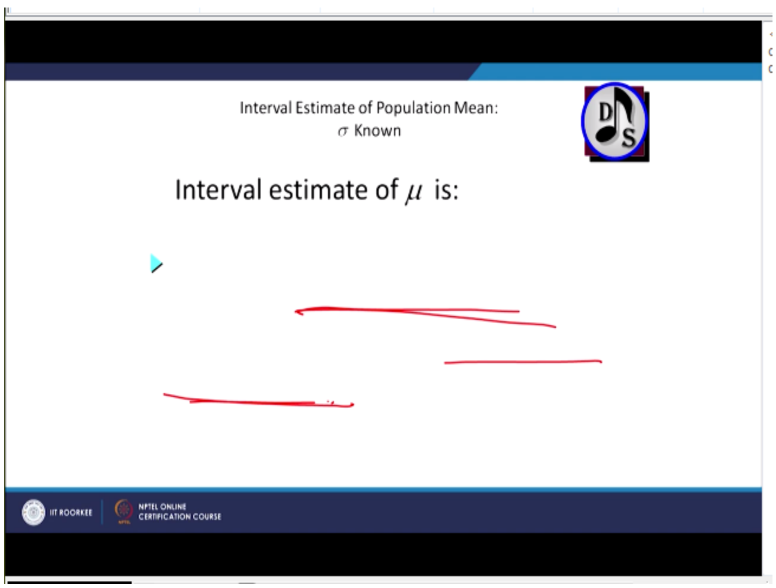
So, sample 36, population mean sorry sample mean is 31100 and standard deviation is 4500 right. So we have to find out confidence interval at 95% level, so it is similar to what we have done in previous example this is there. So this X bar ± Z alpha/2 under root of sigma/n right.

(Refer Slide Time: 05:00)



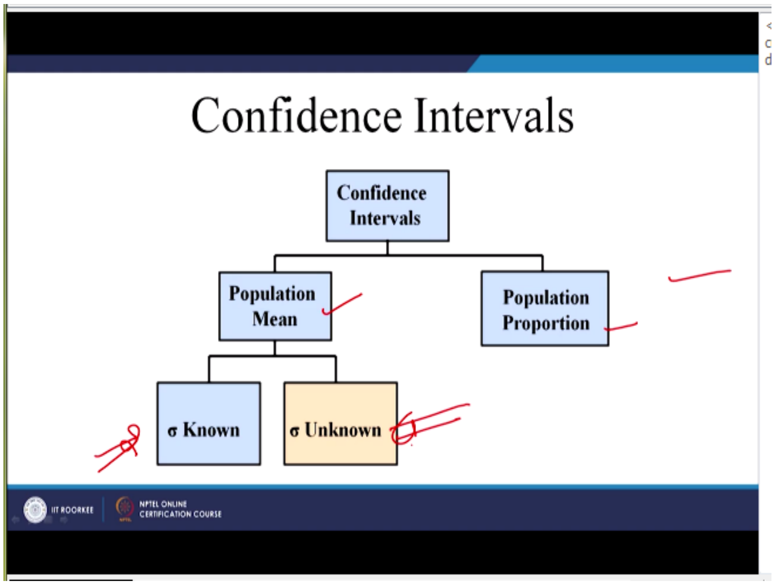
So, this what is the value of this is this Z table value for 95% significance level right. So, this is margin of error and you got mean, so mean ± this right. So, what was mean if you look at previous slide, so mean is 31100 right, so 31100 ± this.

(Refer Slide Time: 05:34)



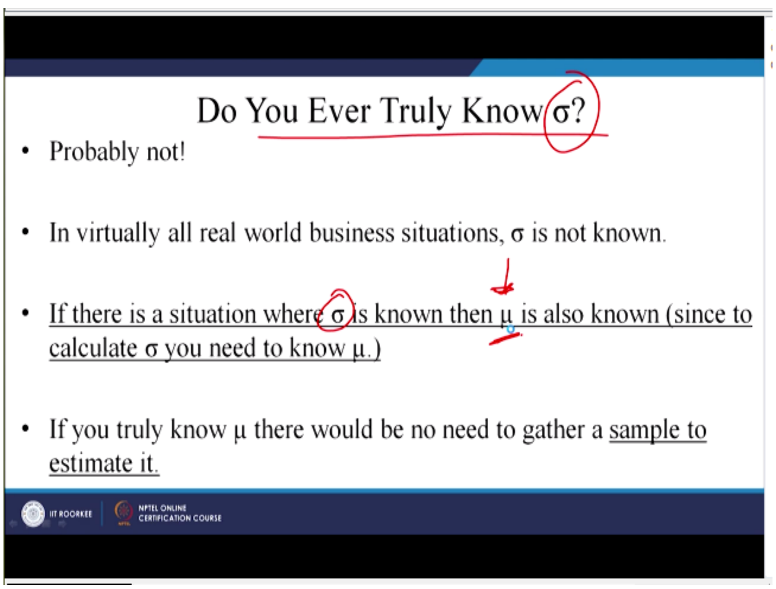
So, this is the interval estimate of income, so we are saying that 95% confident we are 95% confident that the interval contains the population mean right. And it is it has got the same meaning if we take let say one more sample of sample size let say 36 will calculate its mean will again find out intervals and like that when we have let us say 100 intervals then out of 100, 95% of the intervals will have population mean ok.

(Refer Slide Time: 06:30)



So we have seen, we can have not only population mean but population proportion population variance as well right. When we talk about population mean we can have 2 situations either standard deviation is known or standard deviation is unknown. In real life we generally in fact we really do not know standard deviation of the population.

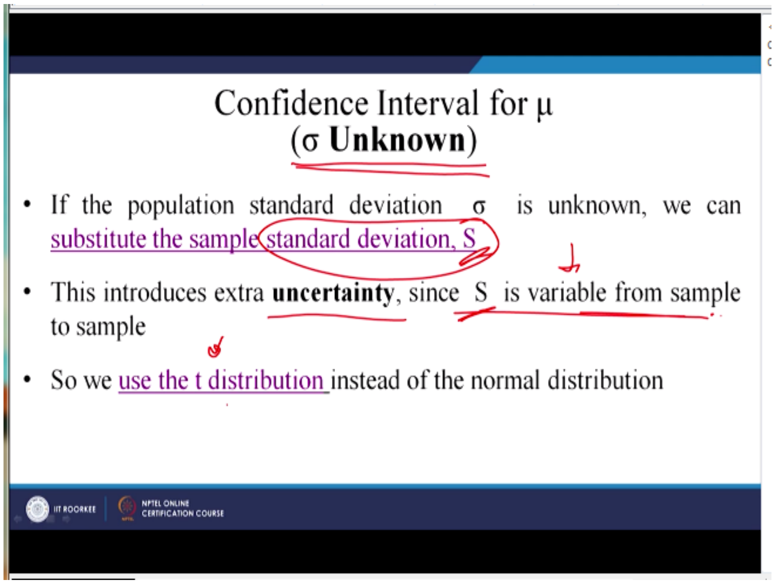
(Refer Slide Time: 06:54)



So do you ever truly know standard deviation no because to find out standard deviation you need to know μ right, population mean. So, once population mean is not there is no need of doing this exercise why are you doing this exercise because we want to know from sample mean you want to know population mean right. So, if there is a situation where this is known then this would be automatically be known.

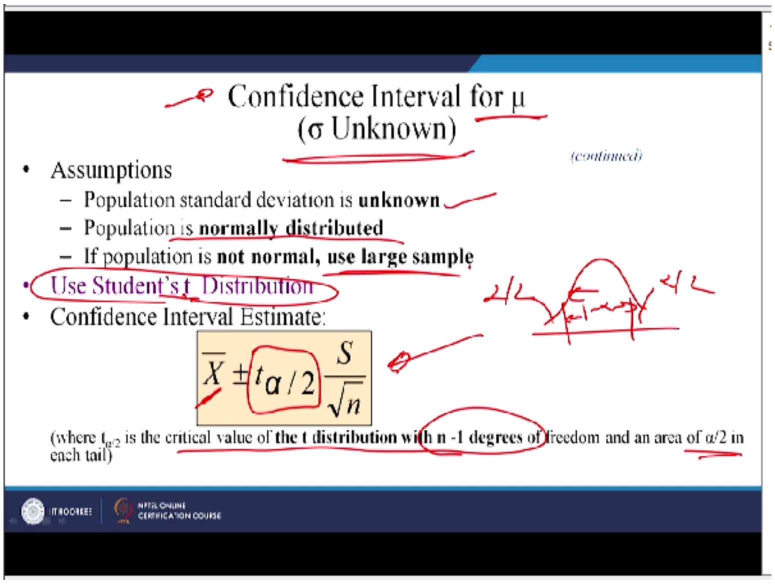
Because standard deviation can be calculated with mu right and if μ is known there is no need of in this exercise right.

(Refer Slide Time: 07:44)



So, you really do not know standard deviation of the population, so what to do in situation like this when you do not know the population standard deviation. In situations like this it is always good to use sample standard deviation, which is S right. So, this introduces an extra uncertainty because S is variable from sample and of course it varies from sample to sample and it is the standard deviation of sample right. So will use a distribution call t distribution rather than z distribution ok.

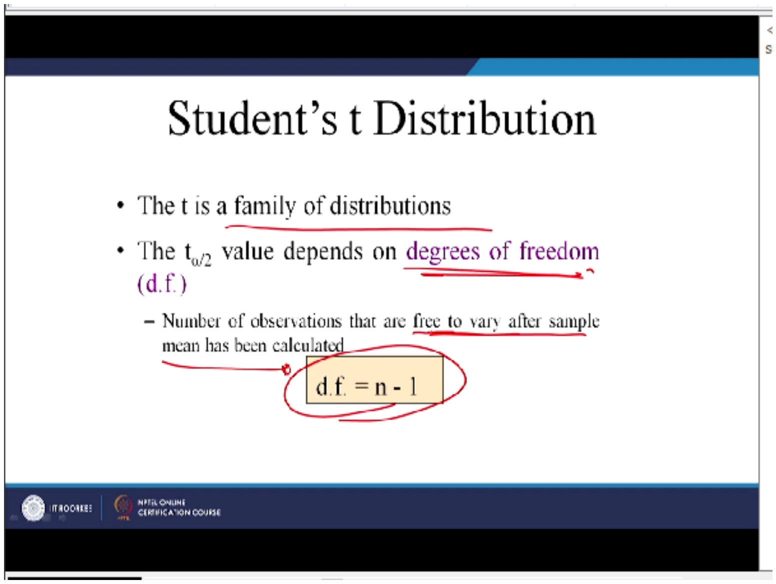
(Refer Slide Time: 08:28)



So, let us look at couple of assumptions for finding out confidence interval of the population mean when this is unknown right. So we will say that standard division is unknown and but the population is normally distributed is our assumption and even if the population is not normally distributed then you need to take the large sample size ok. So either you call t distribution or student's t distribution is one and the same thing right.

So confidence interval estimate is this is X bar is the sample mean ± is t alpha/2 rather than $Z_{\alpha/2}$. Because we are using t distribution, so this is the where $t_{\alpha/2}$ was the critical value of t distribution with n-1 degrees of freedom. And an area alpha/2 in each tail right, so it is similar to what you have seen. So, this again alpha/2 and this is alpha/2 and this 1-alpha right, this area 1-alpha ok.

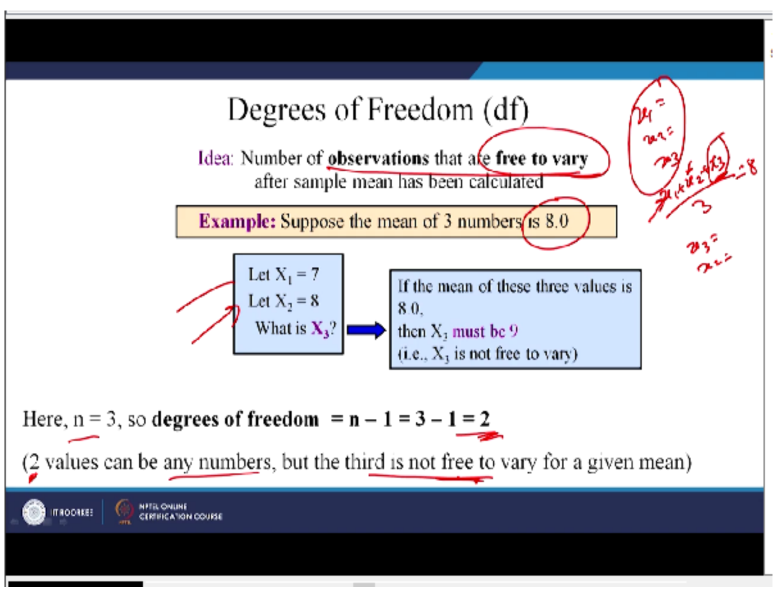
(Refer Slide Time: 09:47)



So, let us look at what is meaning of degree of distribution and before this t distribution is nothing but a family of distributions a family of curves. And the shape of each curve depends on the degree of freedom, so the number of observations that are free to vary, so what is degree of freedom, the number of observations that are free to vary after sample mean has been calculated, so this degree of freedom n-1.

So, whatever is your sample size let say if it is 10 then degree of freedom would be 19 and it is **is** meaning is that the variable which are free to choose.

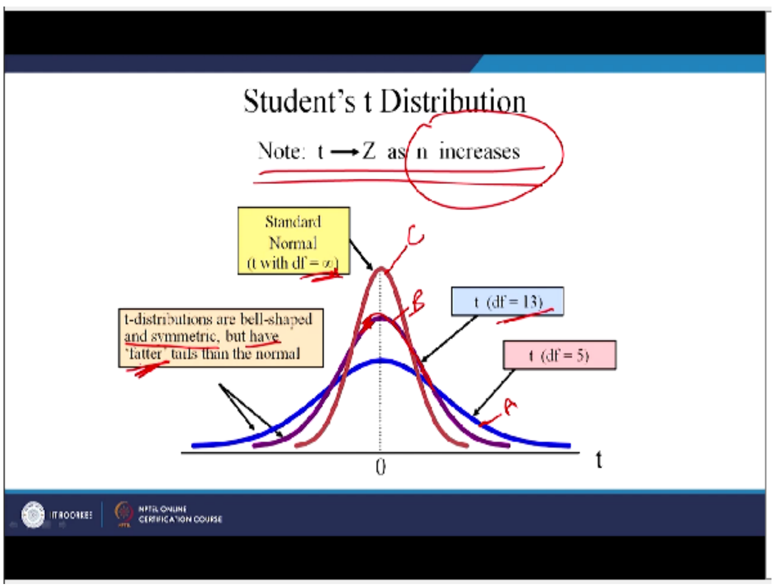
(Refer Slide Time: 10:31)



Let us take an example, so let us say there are 3 numbers X1 X2 X3 and the mean of those 3 numbers is let say 8. So, you have got X1, X2 and X3 right and their mean is this let say X1+X2+X3 divide by 3 = 8, so this 8 is given. And for example let say X1 is 7 and X2 is 8. If you put X1=7 here X2=8 then X3 has to be 9, you cannot have any other value of X3.so, let say if there are 3 variables then you can freely choose values of X1 and X2 not of X3 or in other words if there are 3 variables let say X1, X2, X3 say if X3 is chosen X2 is chosen the X1 one is fixed.

So, if there are 3 variables you are having only 2 variables for which you can freely choose their values right. So, that is the point written over here, number of observations that are free to vary after sample mean has been calculated right. So this is degree of freedom. So, n=3 means degree of freedom right, if n=7 degree of freedom would be 6 right. So, 2 values can be any number of course but the third value is not 3. Because third value would be determined by these 2 values, so this is degree of freedom, so if you look at the difference between t and Z test.

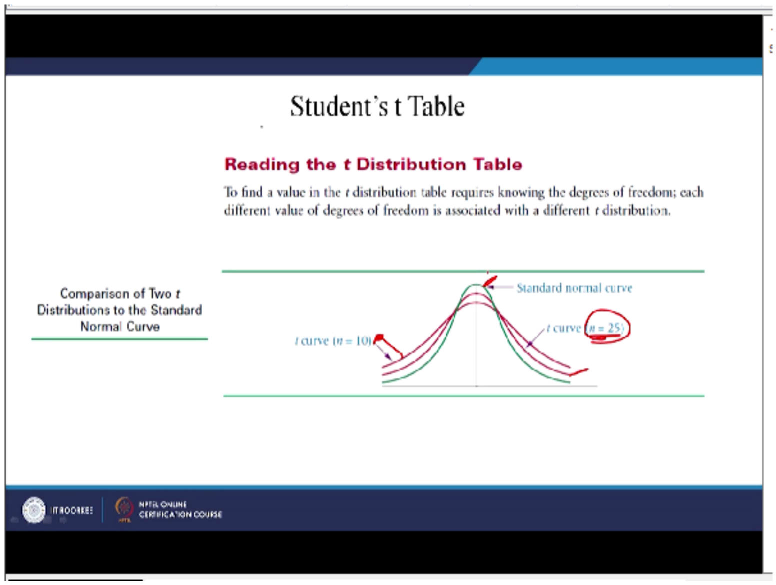
(Refer Slide Time: 12:33)



Then you will see that S be increase sample size the t distribution becomes Z distribution. So, let us look at these are different t distribution curves. So, look at this one this curve right, so it has got degree of freedom 5 right, let us call this A right. This curve this one, this one, this curve let us call this as B right, so, it has got degree of freedom 13, so what is the difference between curve A and B. curve A is flatter right which has got mode dispersion right compared to curve B.

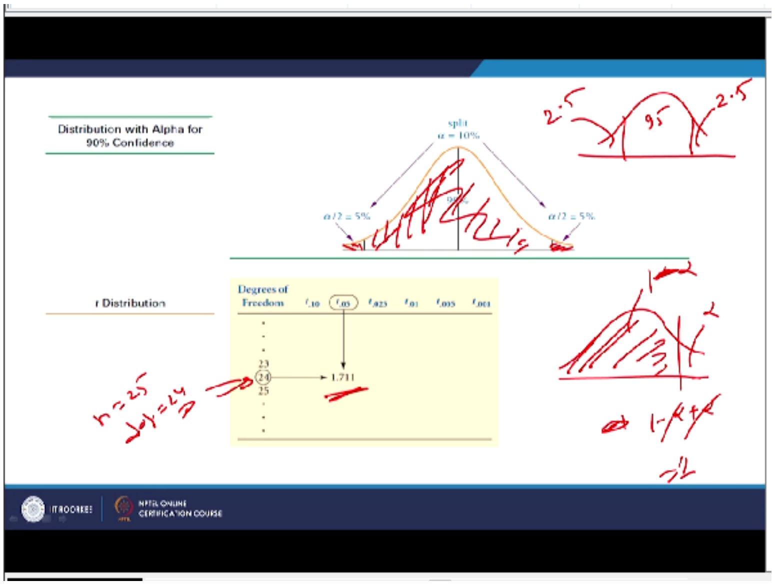
And let us look at this, this is curve C with degree of freedom infinite, so this at t is equal to infinity, sorry if degree of freedom equal to infinity, t distribution becomes Z distribution right. So, t distribution are bell shaped and symmetric but have fatter tails than normal curve this the difference between these two.

(Refer Slide Time: 13:49)



This is one more example, so you have got this curve with n=25 and this curve with n=10 right, so the curve which has got higher value of n is more towards normal curve right. So, this is comparison 2 distribution and to t distribution and standard normal curve and this one is normal curve right.

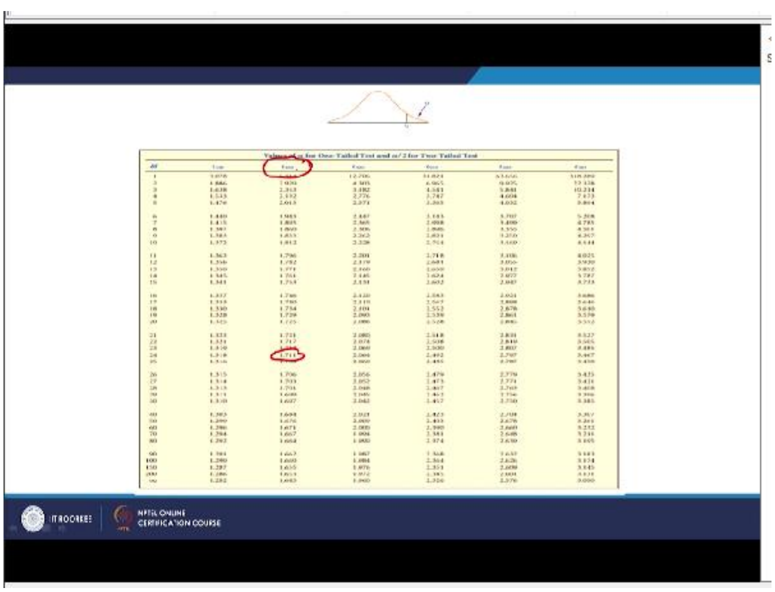
(Refer Slide Time: 14:25)



So, how to look at value of t distribution in t table, so let say this is your this value is your 1- alpha this 90%, so this value is 1- alpha it means remaining 2 sides has to be alpha. So, we are equally dividing this area, so this portion would be alpha/2 this would be alpha/2 right. So let us look one more example let say this entire area is 1-alpha let say 1-alpha then this becomes alpha because this total is 1, so 1-alpha, 1- alpha + alpha this becomes 1.

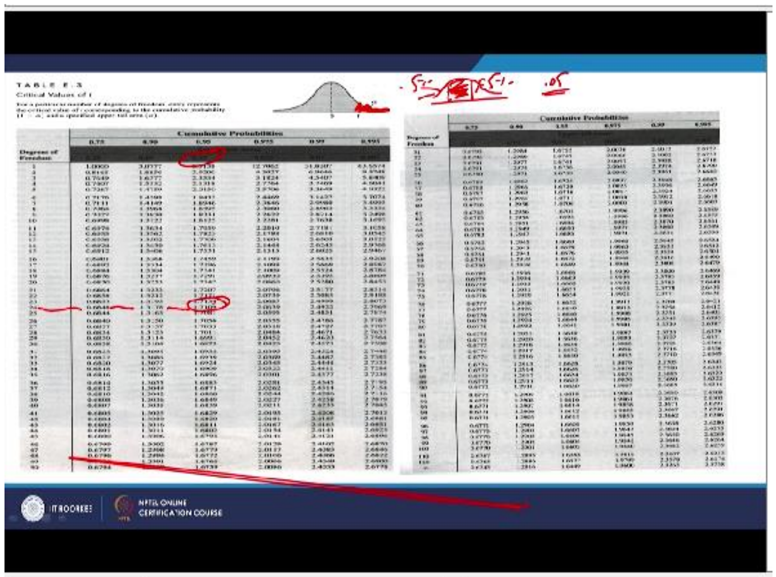
Let say if this is let say 1-alpha is 95 for example then this becomes 2.5 and this becomes 2.5. So, let say degree of freedom is 24 it means n=25 at degree of freedom. So, the degree of freedom would be 24 and t value is just look at t table.

(Refer Slide Time: 15:53)



So, this is t table at t=24 this is 1.71, 1.711 is $t_{0.05}$ right, so this is $t_{0.05}$ right ok.

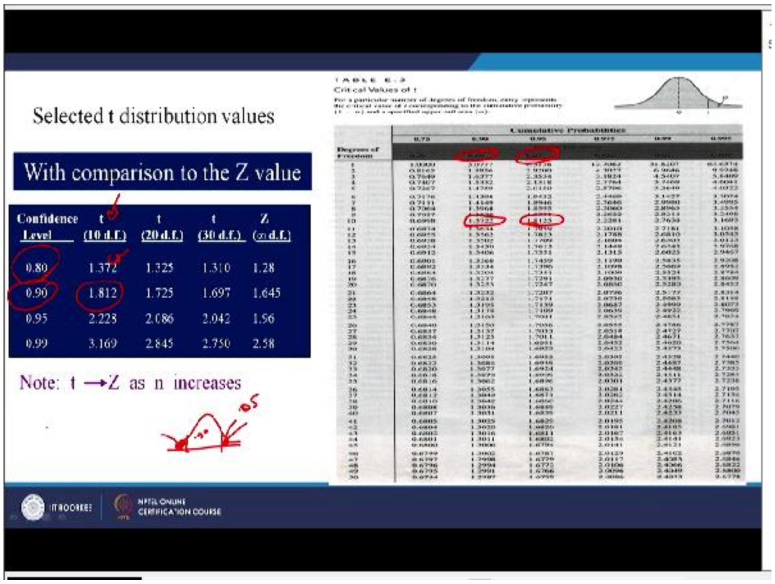
(Refer Slide Time: 16:17)



Let us look at this, in fact you can have different ways of representing t tables in different books right. So this is one of the t tables is quite standard table. So if you look at this table this area is alpha right this one this is alpha and this remaining is 1-alpha ok. Now let us see what is the value of t let degree of freedom=24, so what was the value of 1-alpha here it was 90% right, so this is 90%, so will go to this table once again.

So this is 90%, this is point this is again 5% this is 90%, this area is 5% this area is 5%. So, 5% means 0.05, so look at 0.05 in this table which is here this is 0.05 at degree freedom=24 right this is here right 1., so this is the answer 1.7109 which you can see here 1.71 ok.

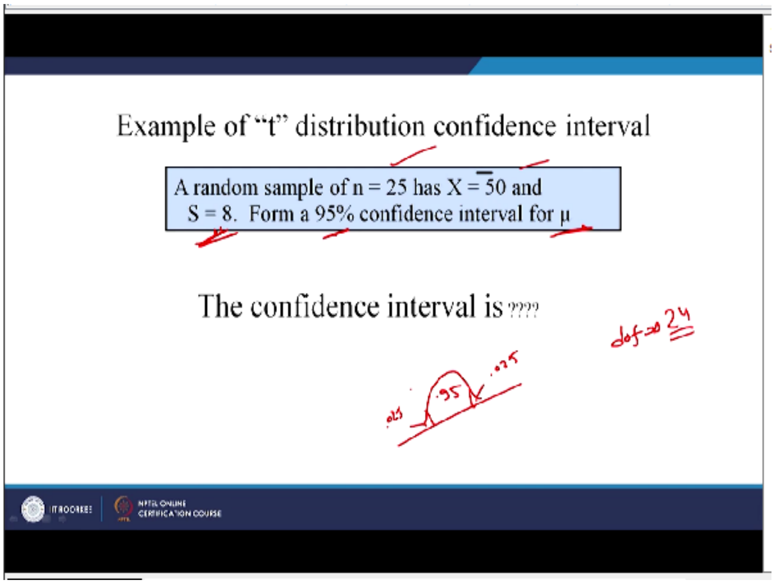
(Refer Slide Time: 18:07)



Let us look at one more table because you need to understand how to find out t value in t table. So let say confidence level is 80, so when I say confidence level is 80 it means this area is 10% right, this area is 10% and this area is 10% right. So, 10% means you just divided by 100 right 0.10, so 0.10 is here in this table ok. Now that degree of freedom=10. So, this becomes 1.3722 which is here is not if I ask you to calculate to find out t value at 10 degrees of freedom when confidence level is 0.90.

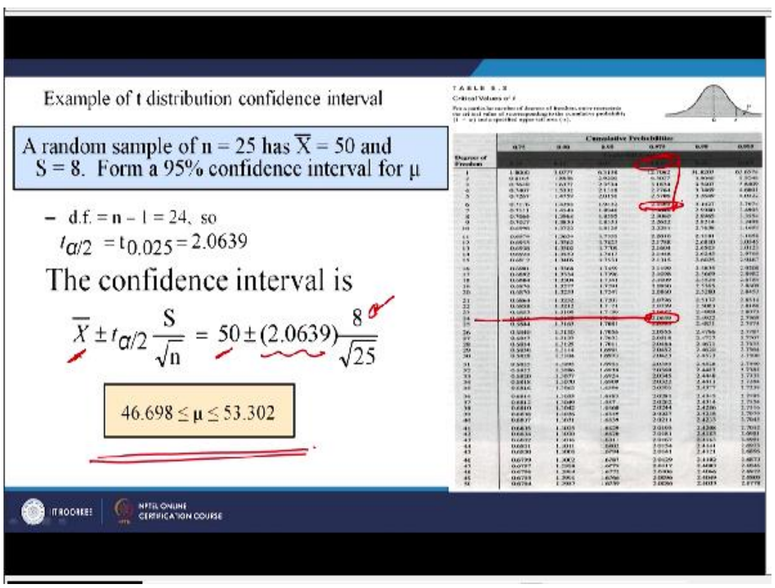
So, what would you do, this is 0.90 right, so this is 0.90, so this side would be 0.05. Because this is in otherwise 0.1/2 and this also 0.1/2 right, so 0.05, this 0.05 is here right and 10 degrees of freedom this is here 1.8125, 1.812 and so on right.

(Refer Slide Time: 19:55)



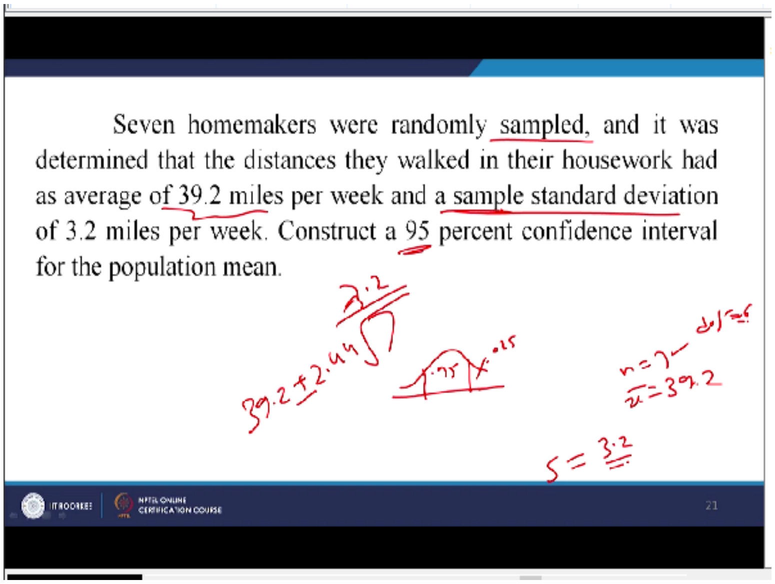
So, this how you should be looking at the t for solving a question, so let us look at this example a random sample of 25 has mean=50 and standard deviation is sample standard deviation 8 from 95% confidence interval for population mean right. So we have to look at t value in t table at degree of freedom=24 and this 0.95 right, so this is 0.95. So, this is 0.025.

(Refer Slide Time: 20:35)



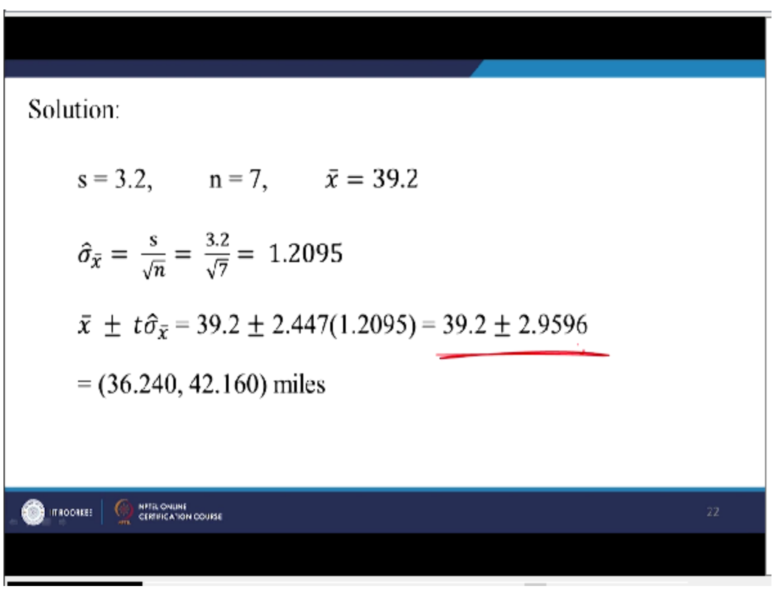
So, look at table 0.025 is this is right this called ok, so 0.025 and what degree of freedom 24 right, so 24 degrees of freedom is this is 2.0639. So, sample mean is 50 ± t value * S upon under root n right. This is sample standard deviation, so the confidence interval is 46.692, 53.30 right, so what does it mean, it means that the 95% of the confidence intervals, so framed would have population mean in this range ok.

(Refer Slide Time: 21:40)



Let us look at this question 7 homemakers were randomly sample and it was determined that the distance they walked in their house work had an average of this much miles. So sample size is 7, X bar is 39.2 miles per week and sample standard deviation is 3.2, S is this construct 95% confidence interval, so how would you do it 39.2 ± you have seen the value of t in previous example. So, but it was at 24 degrees of freedom here we will have to check it 6 degrees of freedom. So, n=7 means degree of freedom=6, so this is 0.95 and this is 0.025 right.

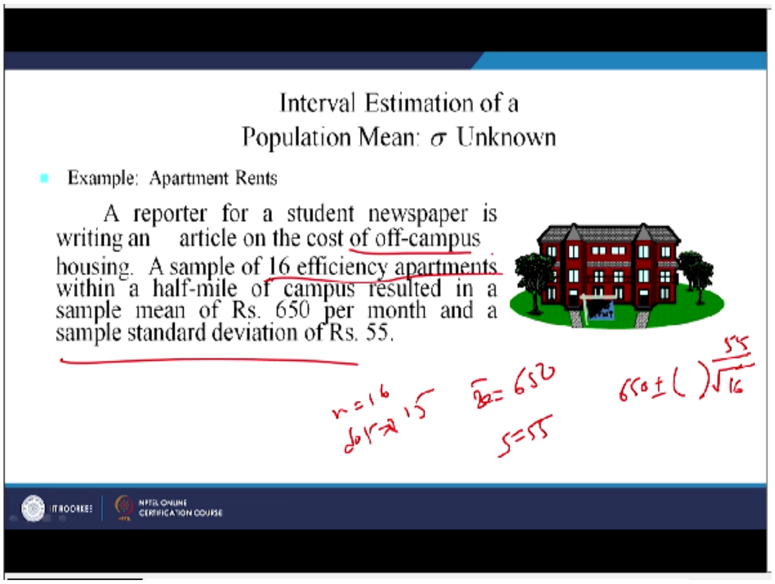
(Refer Slide Time: 22:51)



So, look at table will go back to table, so at 6 degrees of freedom 0.025 right, so this is 0. 025, 6 degrees of freedom 2.44 right. So, is 2.44*S, S is 3.2/sample size right under root of sample size.

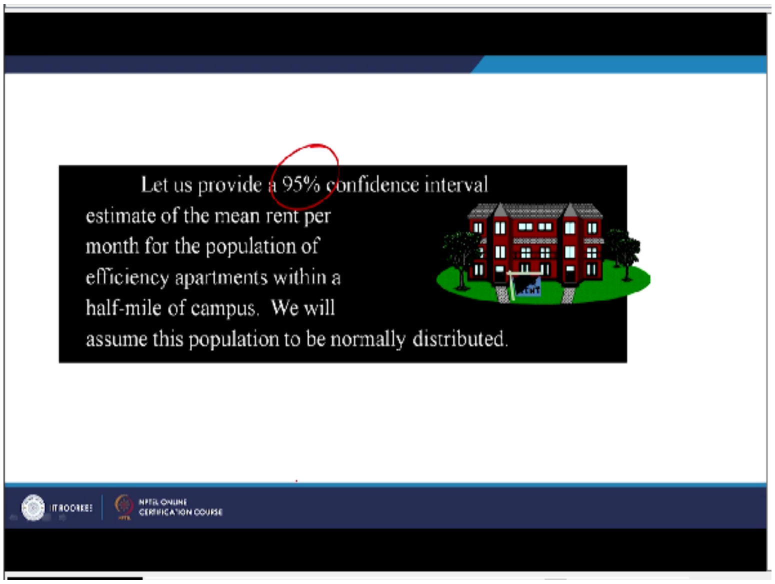
So, this how you should be calculating this and the answer is this in this ok, so this how you should be calculating the confidence interval range when the population standard deviation is unknown.

(Refer Slide Time: 23:35)



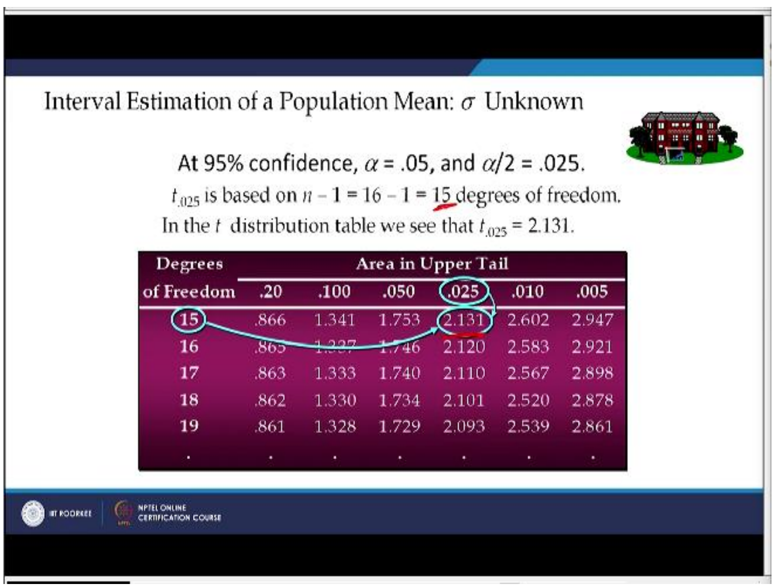
Let us look at one more example which is a sample one a reporter from a student newspaper is writing an article on cost of off campus housing. So, he selected 16 apartments, so n=16 degree of freedom=15 and within half mile of the campus resulted in a sample mean this is 650 and sample standard deviation is 55 right. So how you should be solving at 650 ± whatever is your t value into S is standard deviation is 55/factorial n what is n, n is 60 right. So, you have to just look at t value for of course confidence interval is there in next slide.

(Refer Slide Time: 24:33)



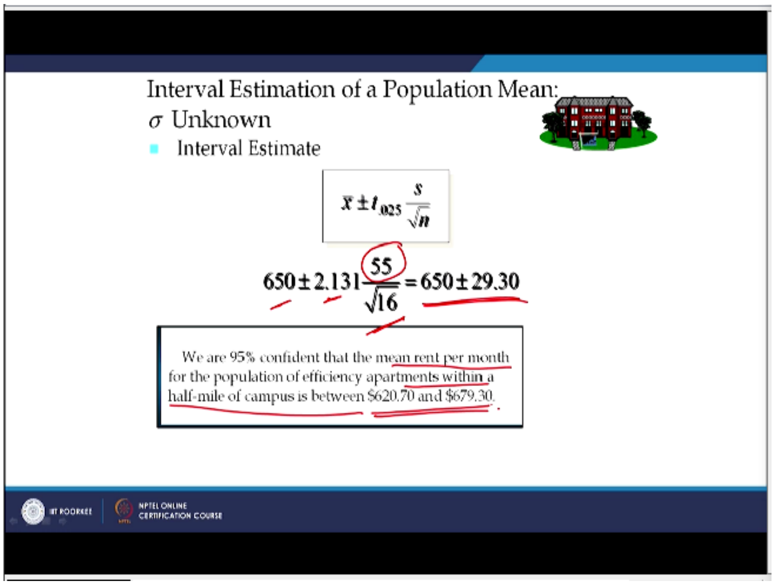
So, at 95% confidence interval you have to find out t value ok.

(Refer Slide Time: 24:45)



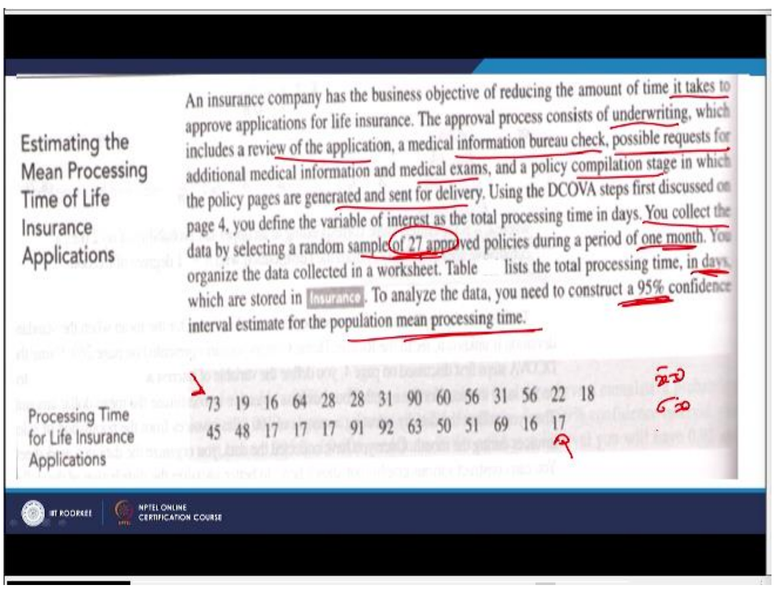
So what was the answer of course the answer is of course you can calculate answer and the answer would be first of all look at t value this is 2.131 at 15 degrees of freedom ok.

(Refer Slide Time: 25:04)



And then you can solve this question right 650 ± t value sample estimated standard deviation divided by under root of sample size, so this is the range right 650 ± 29.30 which would be this right. So, we are 95% confident that the mean rent per month for the population of efficiency apartment within half mile from the campus is this right.

(Refer Slide Time: 25:39)

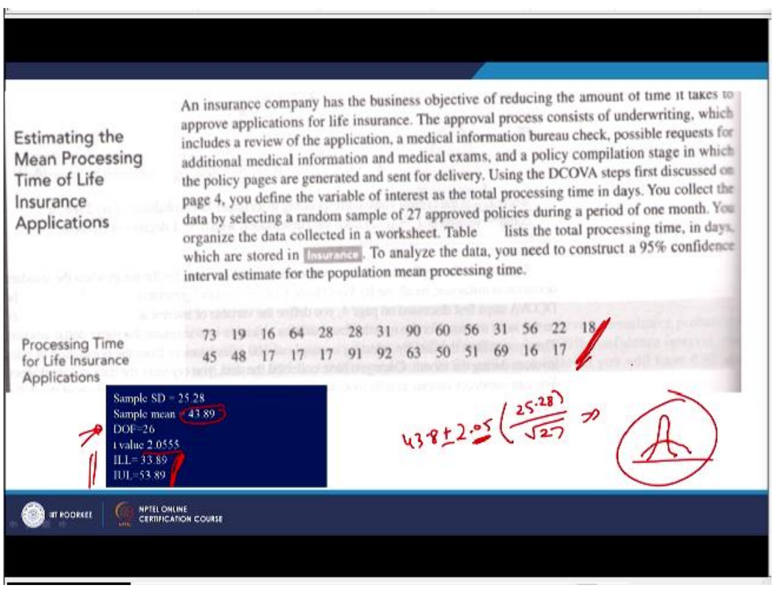


You can work out this example of course I will give you solution but the question is like this an insurance company has the business of objective of reducing the amount of time it takes to approve application for life insurance. So, this approval processes got several steps which consists of underwriting which includes review of application, medical information about the client right, possible request for additional medical information and medical examination and policy come compilation stages in which the policy pages are generated and sent for delivery right.

So you have collected data by randomly selecting a sample of size 27 approved policies during period of 1 month. Now you organize the data collected in the worksheet and this table consists of this time estimates in days right. So, processing time was 73 for 1 application for some other applicant it was 17 and so on right. So, you have been given 27 samples right, you are suppose to construct a 95% confidence interval estimate for the population mean processing time.

So this is a **c**ase of raw data right, you are not being given mean, you have not been given standard deviation. So, first of all what you should do calculate mean and standard deviation and then put those values in the formula right.

(Refer Slide Time: 27:30)



So, this how you should be calculating sample mean just read all these values divided by 27, so you will get mean. Calculate sample standard deviation you know how to calculate sample standard deviation then at 26 degrees of freedom t value is 2.05 right. Now once you are done with this you are supposed to calculate lower limit and upper limit right, so hot it would be this 43.89 ± 2.05 let us take only values up to 2 decimal points right, into S is sample standard deviation is 25.28, 27.

So, if you solve this question you will get these 2 answer right upper limit and lower limit, so before moving onto next slide let me tell you what we have done today. In today's class we have look at how find out confidence interval for a given question we will seen 2 cases especially over here. We have calculated confidence interval for those wherein population standard deviation was given and when population standard deviation was not given.

So, when standard deviation was unknown what we did we just used sample standard deviation to solve the question and instead of Z distribution we have used t distribution. So t distribution is not be the probability table or just like Z table because Z table gives you that the probability that within which a population parameter will fall right. The importance of t table is that it approaches towards Z table as we increase degrees of freedom.

So lower degrees of freedom it would be quite a fatter curve just like this curve and if you increase sample size it becomes a curve like this right. So, before moving on to next topic this what I wanted to summarize, in next class will have some more questions on let say how to find out confidence interval of population proportion, how to find out sample size. So that you can come up with accurate answers, so with this let me finish here, thank you very much.