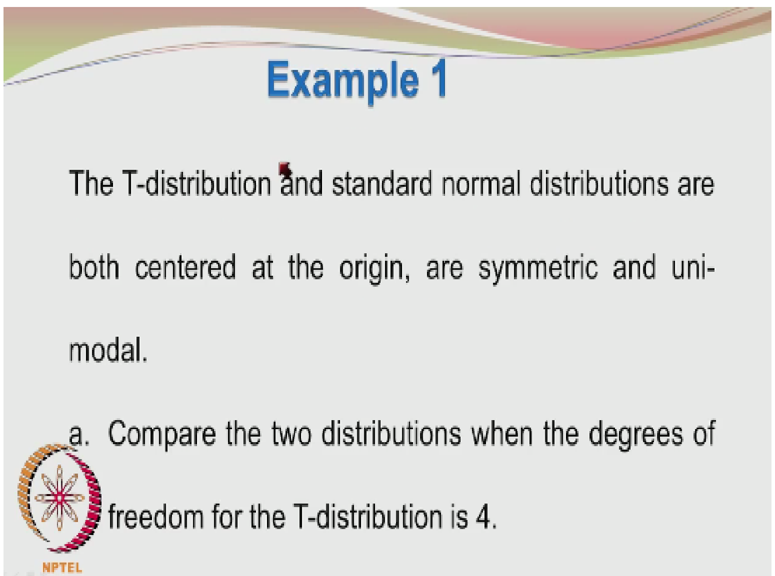
Statistics for Experimentalists
Prof. Kannan A
Department of Chemical Engineering
Indian Institute of Technology – Madras

Lecture – 21
Example Set 5

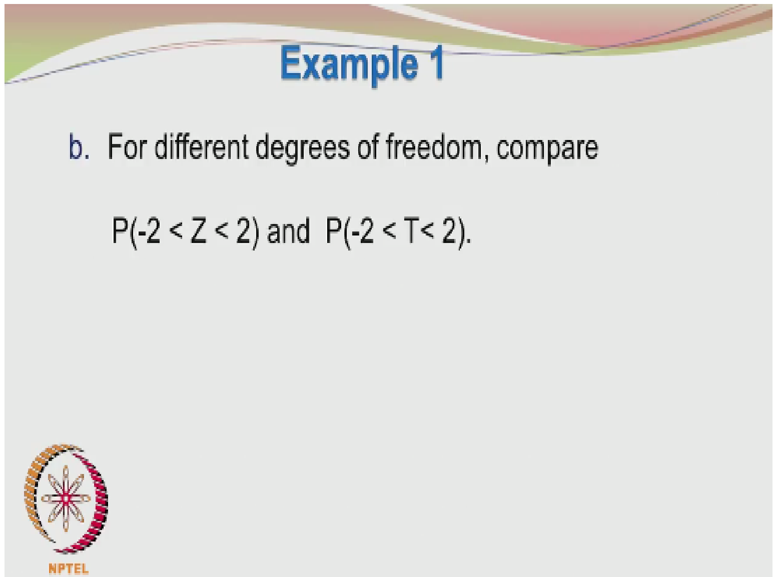
We are going to do a few interesting problems in this lecture. We will be looking at example problems involving the T-distribution, the chi-square distribution and the Fisher F-distribution. We saw that the T-distribution and standard normal distribution had some similarities but they were also quite different. So, the first example demonstrates these differences specially.

(Refer Slide Time: 00:50)



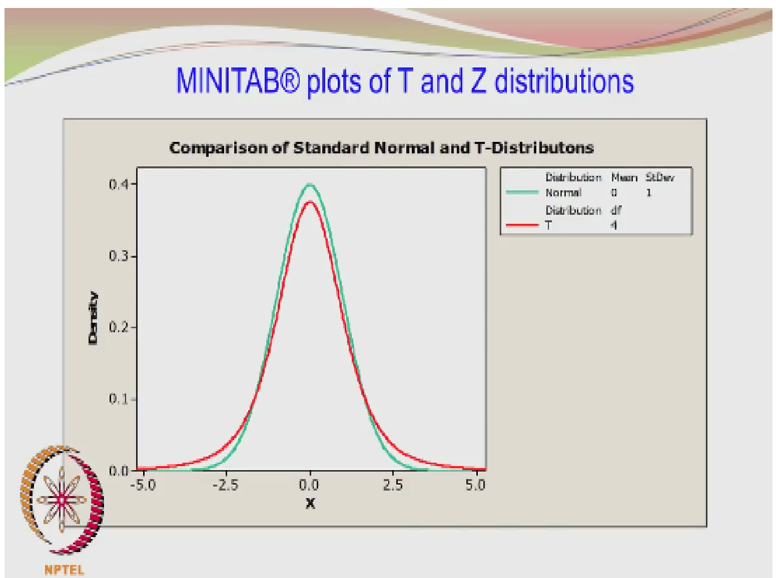
The T-distribution and the standard normal distribution are both centered at the origin. They are both symmetric and uni-model. They have one maximum value. Now, the question is quite simple. Compare the two distributions namely the T-distribution and the standard normal distribution, when the degrees of freedom for the T-distribution is for 4.

(Refer Slide Time: 01:24)



For different degrees of freedom, compare probability of -2<Z<2 and probability of -2<T<2.

(Refer Slide Time: 01:35)



The first question is really asking you about the shape of the two distributions. The standard normal distribution is shown by the green curve and it will have a mean of 0 and standard deviation of 1. Now, when you look at the T-distribution, it is having degrees of freedom as 4 and you can see that it is broader. There is more probability packed in the tail region and then it is also shorter when compared to the standard normal distribution.

(Refer Slide Time: 02:39)

Example 1

b. For different degrees of freedom, compare

$P(-2 < Z < 2)$ and $P(-2 < T < 2)$.

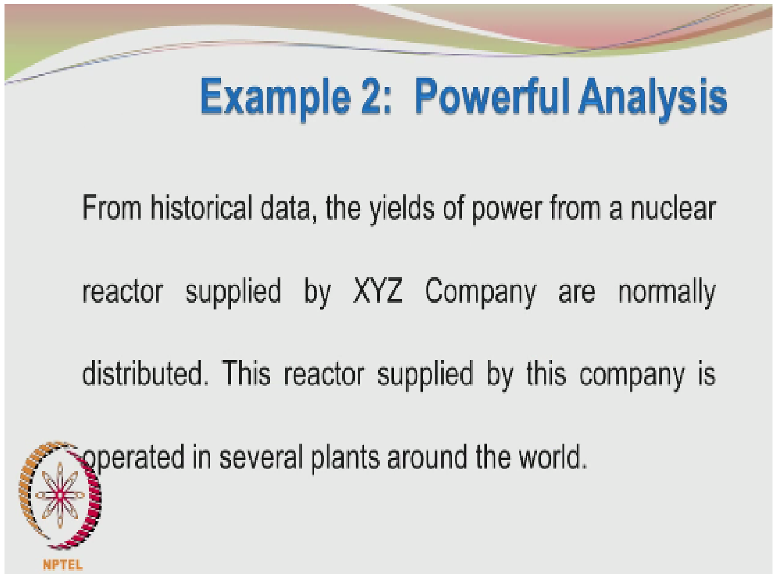
Degrees of Freedom	Standard Normal	T-Distribution
	$P(-2 < Z < 2)$	$P(-2 < T < 2)$
4	0.954	**0.884**
12		**0.931**
36		**0.947**
108		**0.952**
324		**0.954**
972		**0.954**

So, for different degrees of freedom, compare probability of $-2<Z<2$ and probability of $-2<T<2$. So, in this table, we have the first column. It is running from 4 to 972 degrees of freedom through 12, 36, 108, 324 and finally 972. It can be seen that the degrees of freedom have been tripled, 4*3, 12*3, 36*3, 108*3, 324*3; and the standard normal distribution, it is independent of the degrees of freedom.

Degrees of freedom is not a parameter here and hence we have probability of $-2<Z<2$ that is equal to 0.954. If you look at the T-distribution, the probability value is less for 4 degrees of freedom when compared to the standard normal distribution. When it increases from 4 to 12, the probability values also increase from 0.884 to 0.931 and when you go to very large degrees of freedom, the standard normal probability.

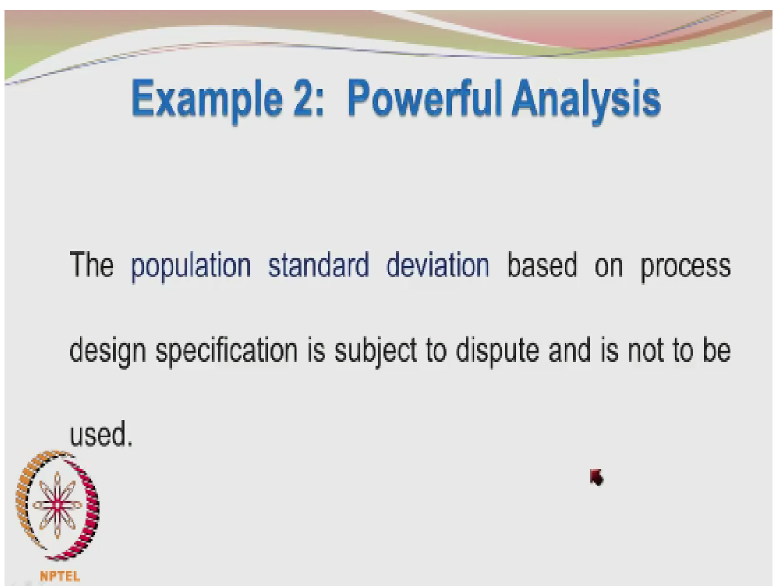
And the T-distribution probability are pretty much the same and this drives home the point that the T-distribution approaches the standard normal distribution when the degrees of freedom tend towards infinity.

(Refer Slide Time: 04:33)



Let us look at the second example. We have seen this example before but now let us look at another version of that example. So, it is known from historical data that the yields of power from a nuclear reactor supplied by XYZ company are normally distributed. So, the power is the random variable and that is normally distributed. So, the reactor supplied by this company is operated in several power plants around the world.

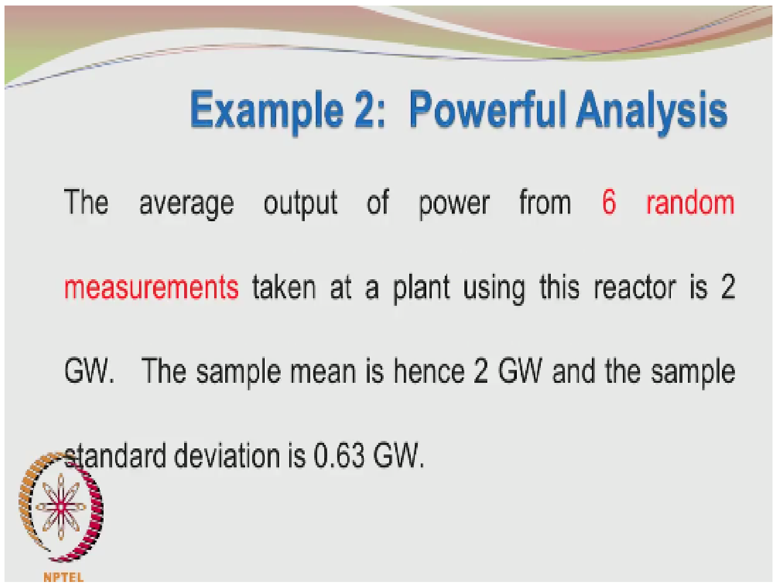
(Refer Slide Time: 05:20)



Of course, this example is strictly fictitious. The population standard deviation based on process design specifications is subject to dispute and is not to be used. We saw why it was so in the previous example set. If the standard deviation given by the company is presumed or taken to be quite high, then the industries may feel that this supplying company is hiding behind this large population standard deviation.

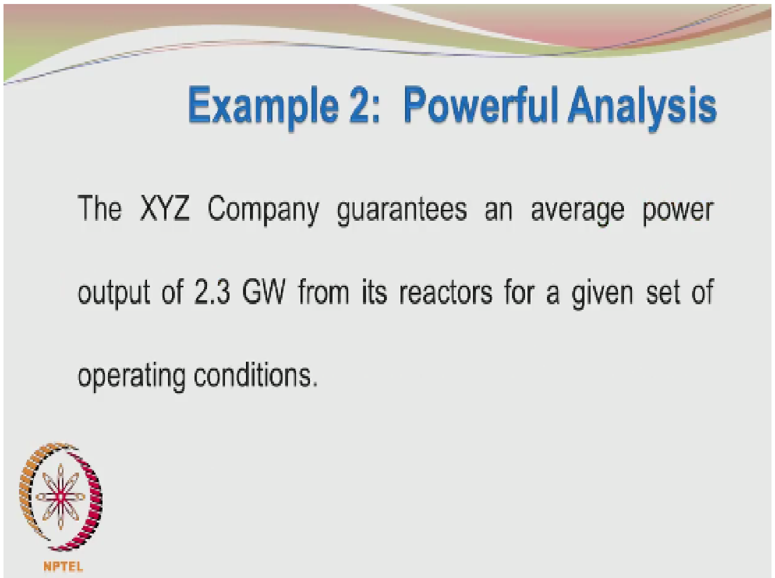
So, it is getting away with supply of less power. So, it is decided not to use that particular value of sigma. Then, it boils down to the case where we are having a situation with unknown standard deviation.

(Refer Slide Time: 06:36)



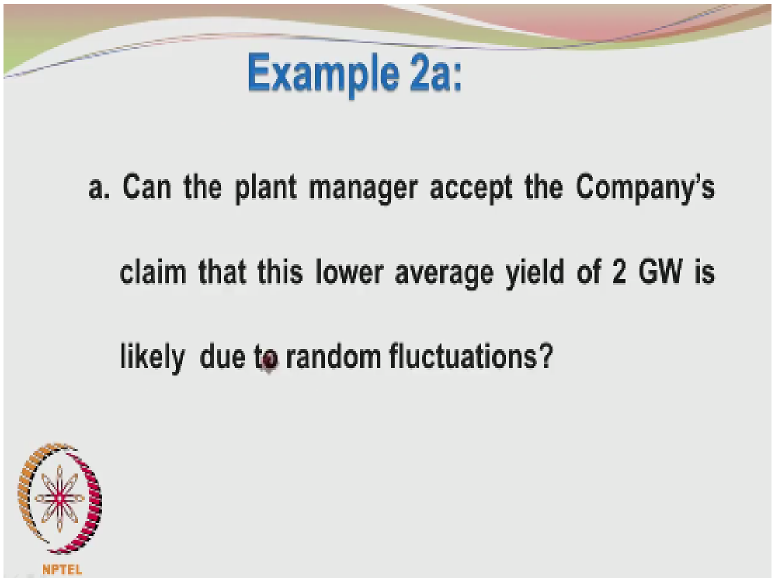
So, the average output of power from six random measurements taken at a plant using this reactor is 2 gigawatts. The sample mean is hence 2 gigawatts and based on the six measurements, the standard deviation from the sample is 0.63 gigawatts. 0.63 gigawatts is quite a larger fraction about 31.5% of the mean value, that is pretty high.

(Refer Slide Time: 07:27)



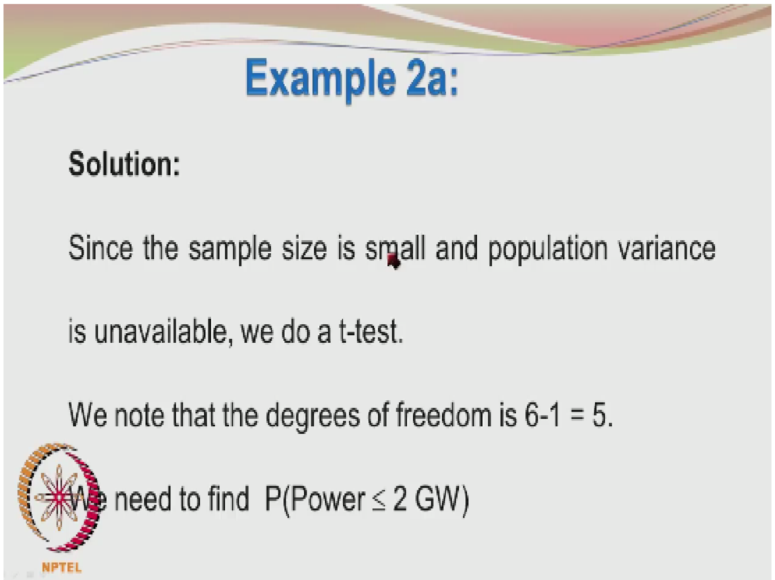
The XYZ company which is supplying such nuclear reactors guarantees an average power output of 2.3 gigawatts from its reactors for a given set of operating conditions.

(Refer Slide Time: 07:42)



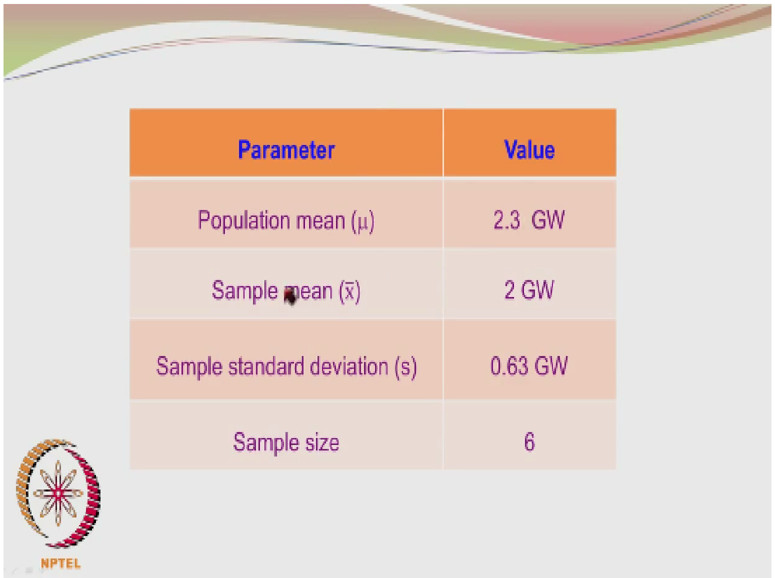
Now, the question is can the client accept the company's claim that this lower yield of 2 gigawatts is likely due to random fluctuations, okay. The sample is indeed taken from a probability distribution centered around 2.3 gigawatts and there is a high probability that from this sampling distribution of the means that you can pick up a sample which is showing only 2 gigawatts.

(Refer Slide Time: 08:26)



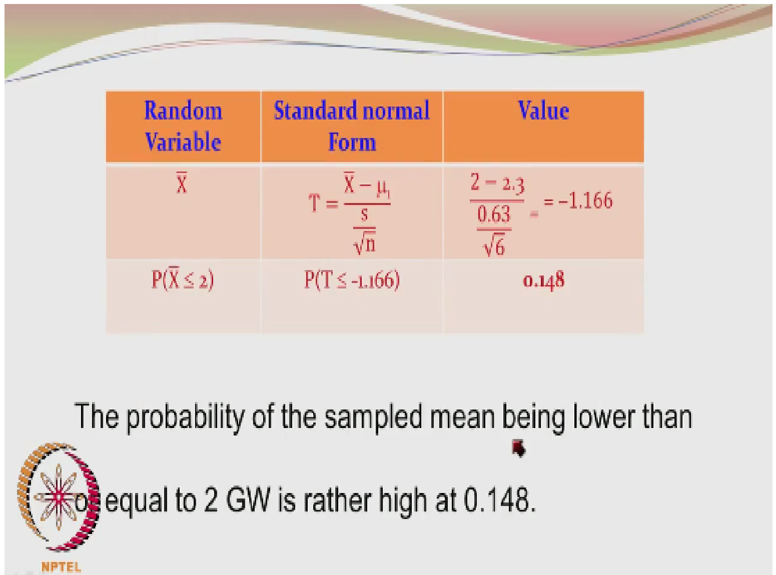
So, the sample size is small and the population variance is unavailable, we do a t-test. Of course, the parent population is normally distributed, that information is given to us. So, the condition for doing the t-test are satisfied. So, the degrees of freedom is 6-1 which is 5. So, we have to find the probability that the average power can be less than or equal to 2 gigawatts even though the population mean is 2.3 gigawatts.

(Refer Slide Time: 09:06)



So, the population mean is 2.3 gigawatts, the sample mean is 2 gigawatts, sample standard deviation is 0.63 gigawatts and sample size is only 6.

(Refer Slide Time: 09:15)



Random Variable	Standard normal Form	Value
$\bar{X}$	$T = \frac{\bar{X} - \mu_1}{\frac{s}{\sqrt{n}}}$	$\frac{2 - 2.3}{\frac{0.63}{\sqrt{6}}} = -1.166$
$P(\bar{X} \leq 2)$	$P(T \leq -1.166)$	0.148

The probability of the sampled mean being lower than or equal to 2 GW is rather high at 0.148.

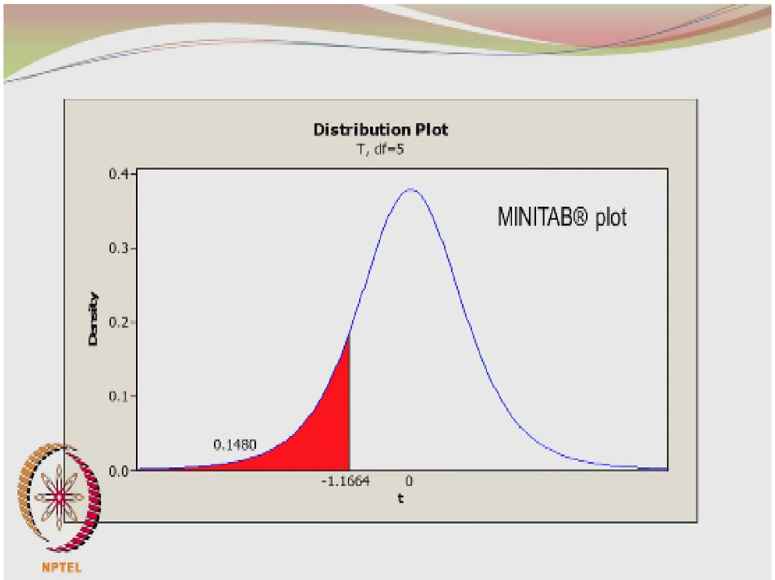
So, we do a t-test here. We define the T variable as X bar-mu/S/root n and there is a typo here. The standard normal form does not apply here because we are not talking about the Z random variable. We are talking about the T random variable, okay, that takes care of the typo. So, T is equal to X bar-mu1/S/root n and that is 2-2.3/0.63/root 6, do not put 5 here, 5 is the degrees of freedom but the sample size is n which is 6 and that number comes to -1.166.

So, the probability of X bar<2 is equivalent to probability of T< or =-1.166 and that probability is 0.148. It not easy to find this probability from the chart because alpha of 0.05, 0.1, 0.025, these are the standard alpha values given in the T-distribution tables. So, how did I find this 0.148. I used spreadsheet to do it. It will be a good idea for you to become familiar with spreadsheets to calculate these statistical probabilities.

Actually, there are also online calculators. I really have not checked into those. If you are having a statistical software with you, that is very good. You can use of Minitab for example to find the probability values. Otherwise, there may be online probability calculators which may give you the T values, the T probabilities, the chi-square probabilities and so on. It is always good to have an independent check for your calculations, so that you can double-check that your reported probability values are correct.

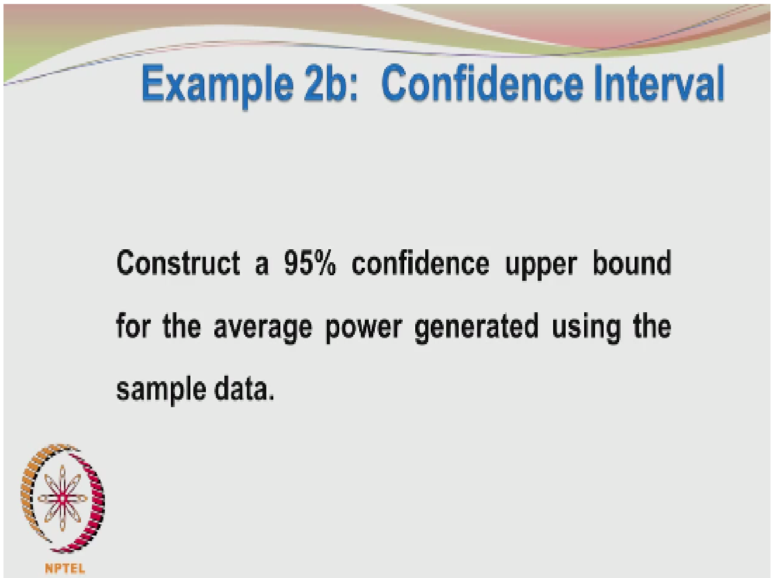
So, the implication is the probability of the sample domain being lower than or equal to 2 gigawatts is rather high at 0.148. So, the supplier can say that the probability of the reactor supplied by me providing a mean output of less than or equal to 2 gigawatts is rather high at about 0.15, okay. So, this is a high probability. So, you have to go with my reactor and you cannot really contradict my statement that the average power output is 2.3 gigawatts, okay.

(Refer Slide Time: 12:44)



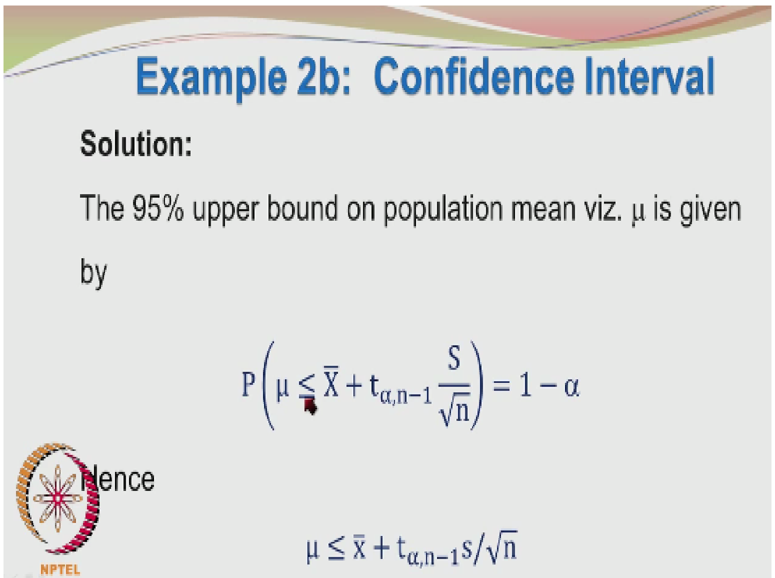
So, looking at the Minitab plot, here we have the T distribution drawn for 5 degrees of freedom and when I am looking at it, I see that when I locate the T variable value -1.1664 here, the area to the left or the left tail region is 0.148.

(Refer Slide Time: 13:22)



Now, the second part of the question is construct a 95% confidence upper bound for the average power generated using the sample data. So, what we are trying to really see here is I am having a pretty low sample average of power output at 2 gigawatts. So, we have the construct a 95% confidence upper bound. Usually, we were looking at the 95% confidence involving the lower limit and the upper limit. Now, we are talking about 95% upper bound only. So, we have two report 95% confidence bound as mu less than or equal to a certain upper limit value. So, let us see how to construct this.

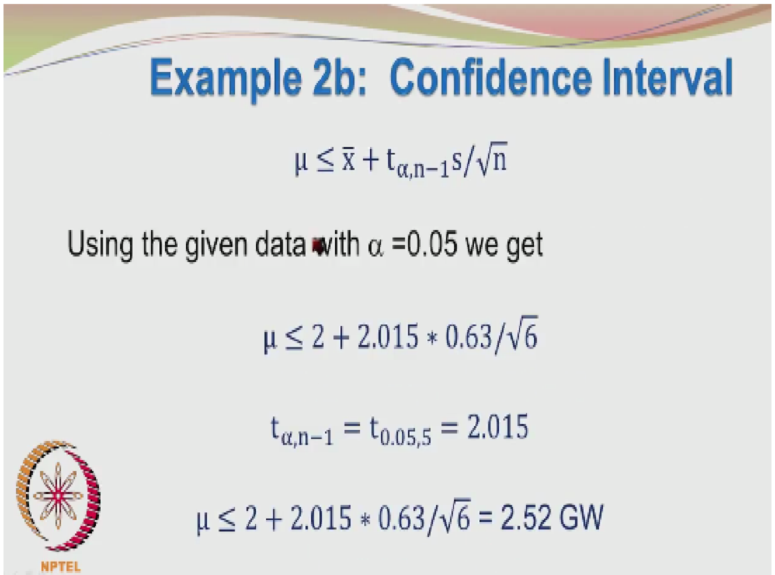
(Refer Slide Time: 14:30)



We know that the definition for the upper bound is mu< or =X bar+T alpha n-1 S/root n is first given, then we take the probability of mu< or =X bar+T alpha n -1 S/root n and that is equal to 1-alpha. Here we do not put alpha/2 because we are talking about one-sided bound and another thing is since we are talking about 1-alpha to be 0.95, alpha value will be of course 0.05. With this definition, we can express the upper bound on mu, the population mean power output as mu< or =X bar+T alpha n -1 s/root n.

So, now a sample has been taken and sample standard deviation is known to us, X bar is also known to us, sample size is known to us. So, this limit can be easily found.

(Refer Slide Time: 15:43)



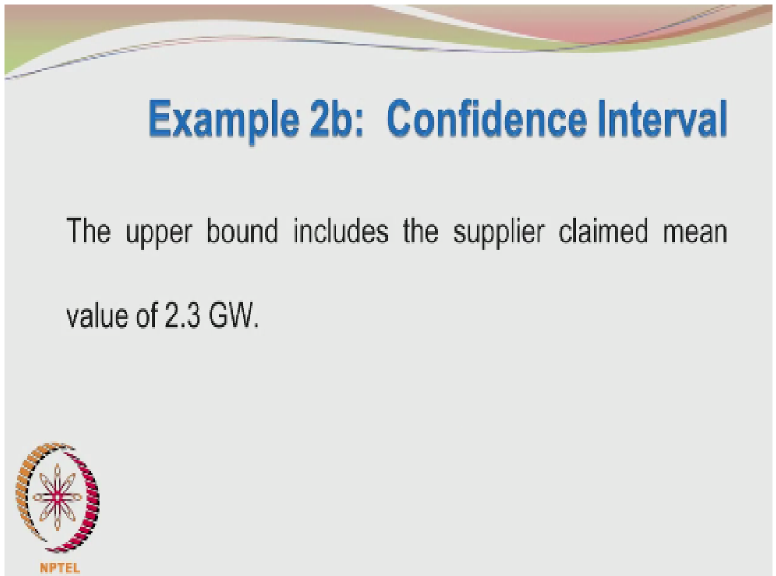
For alpha=0.05, we get t.05, 5 as 2.015. So, mu< or =2+2.015*0.63/root 6. So, this value is 2.015. So, we get mu as 2.52 gigawatts, okay. So, based on X bar, value of 2 gigawatts, we put an upper bound on the mu and that comes to 2.52 gigawatts. What do you interpret from this result. If our raw acceptance or tolerance or penalising criteria is based on 0.05, that means if we can say that probability of the occurrence of the low power is below 0.05, then we can reject the company's claim.

So, we are slowly moving into the hypothesis testing and things will become more clear when we do that. If the population mean is 2.52 gigawatts, probability of the random sample mean taking on values less than or equal to 2 gigawatts will be 0.05. So, only when the power guaranteed by the company is 2.52 gigawatts, the 2 gigawatts can be considered to be unacceptable with our probability limit of 0.05. Only when mu is 2.52 gigawatts, probability of the sample average power falling below 2 gigawatts or equal 2 gigawatts will be 0.05.

If the guaranteed power output is lower than 2.52 gigawatts, then our observed power outputs of 2 gigawatts or lower will obviously have a probability higher than 0.05. So, if our tolerance is 0.05 probability, then until 2.52 gigawatts we have to accept the power yields of 2 gigawatts or lower. So, this might be difficult for some of you to understand at this point but if you think about it, you will appreciate what I said just now.

The same arguments will be presented in hypothesis testing and we can look at it then, but both confidence interval tests and hypothesis tests are giving you the same final conclusion. So, it is also good to think about the confidence interval approach to decision-making.

(Refer Slide Time: 19:57)



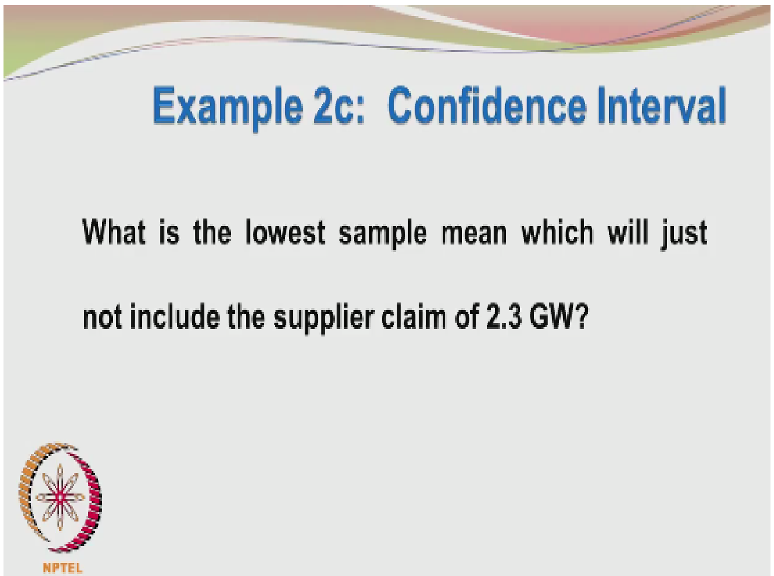
So, we have to see whether the guaranteed power output by the supplier falls within this upper bound of 2.52 gigawatts. Obviously, the supplier is making a claim of 2.3 gigawatts which is lower than 2.52 gigawatts. If the supplier had made statement that I am going to guarantee a power of 2.6 gigawatts or 2.7 gigawatts, then the probability of the observed random sample mean of 2 gigawatts or lower being taken from a population with mean 2.6 gigawatts or 2.7 gigawatts will be lower than 0.05.

Repeating the statement, suppose let us say the supplier is making a claim or guarantee of 2.75 gigawatts and your random sample has taken a value of 2 gigawatts average only. So, the probability of the sample mean taking on values of 2 gigawatts or lower from a population mean of 2.75 gigawatts will be definitely lower than 0.05. So, that sample mean could not be considered to be coming from a population with mean of 2.75 gigawatts.

If the population mean was 2.52 gigawatts, that is the guarantee given by the supplier, then the probability value of the sample mean falling below 2 gigawatts provided the guaranteed mean is 2.52 gigawatts will be 0.05 but the supplier is making the guarantee of only 2.3 gigawatts. So, the probability of random sample taking values less than or equal to 2 gigawatts would be higher than 0.05. So, on this basis, we have to really accept that the deviation from the guaranteed mean value is only because of random fluctuations.

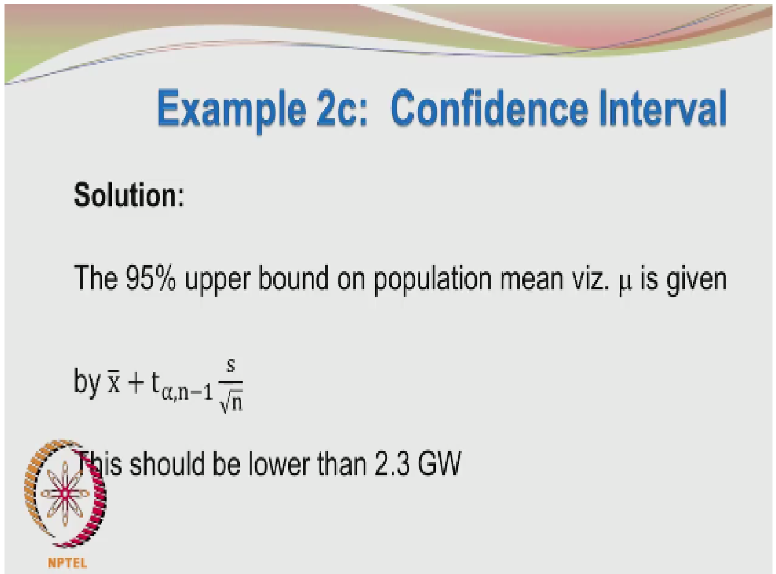
So, if we put the 2.3 gigawatts as the upper bound, what should have been the sample mean which would have led to a probability lower than 0.05, that is an interesting point. So, we are now shifting the upper bound from 2.52 gigawatts to 2.3 gigawatts, then what should have been the sample mean of X bar.

(Refer Slide Time: 23:17)



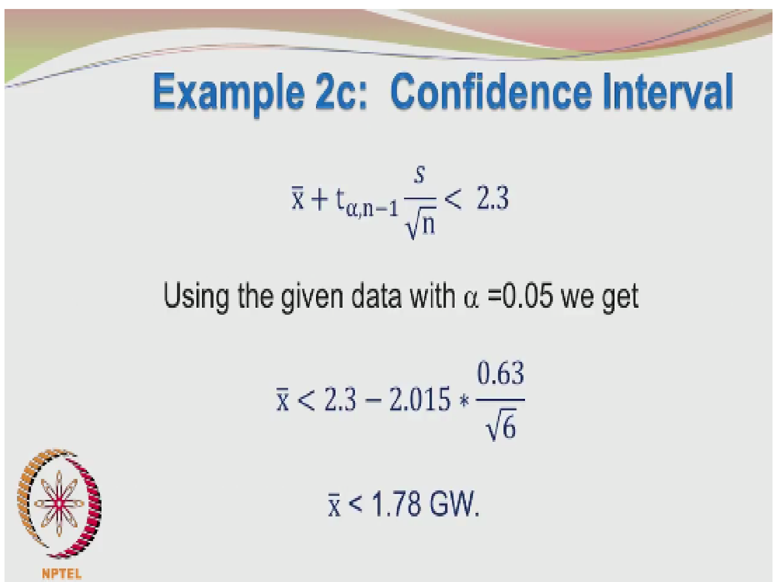
So, what was the lowest sample mean which will just not include the supplier claim of 2.3 gigawatts.

(Refer Slide Time: 23:23)



This should be lower than 2.3 gigawatts.

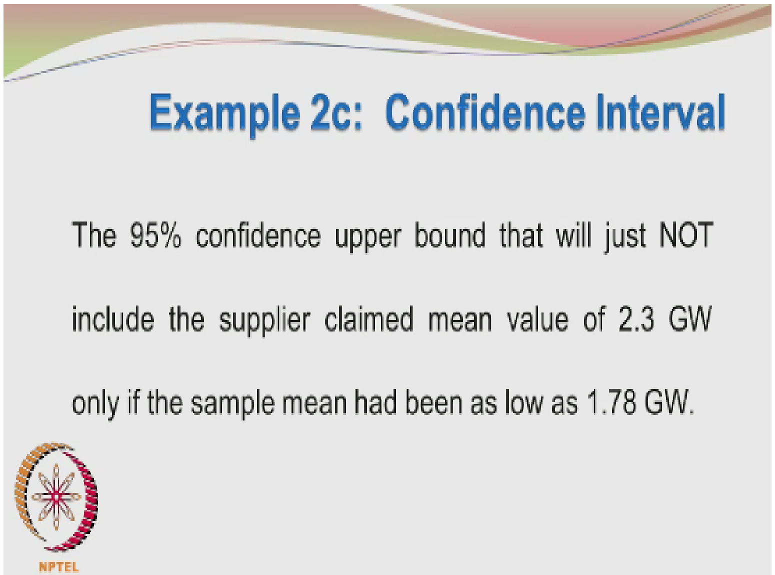
(Refer Slide Time: 23:28)



So, we know the value of t alpha n-1 T.05 and 5 degrees of freedom, it is 2.015 and we know the standard deviation S and we know the sample size and X bar should be less than 1.78 gigawatts. Only if we had observed the average power output from the samples to be 1.78 gigawatts or lower, we can question his claim of 2.3 gigawatts saying that my sample mean is as low as 1.78 gigawatts or lower and the probability of the sample mean falling below 1.78 gigawatts if the population mean 2.3 gigawatts is less than 0.05.

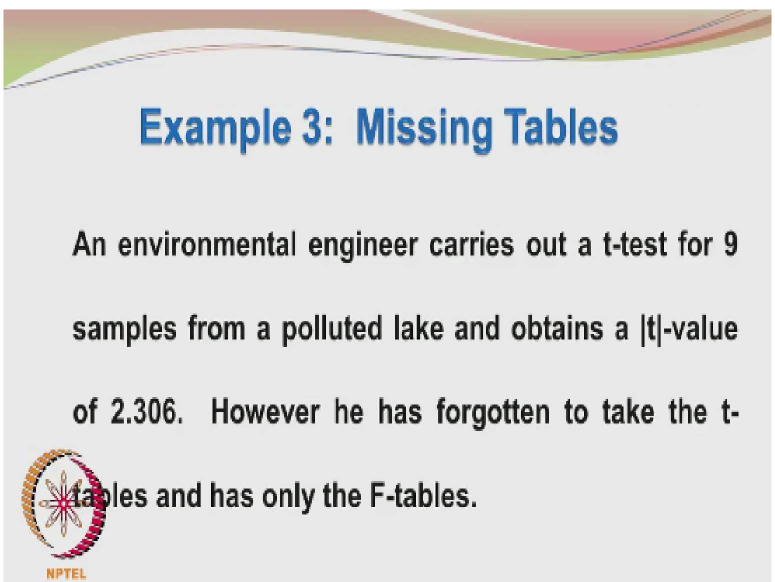
So, what you are saying is not correct. This cannot be because of random variations. Something is faulty with your reactor. So, we were actually getting 2 gigawatts. So, the probability value then was 0.148 or 0.15 which was a very high probability. Only if the samples showed a mean of less than 1.78 gigawatts, can we really tell the reactor supplier look your supplied reactor is not performing up to its stated performance.

(Refer Slide Time: 25:10)



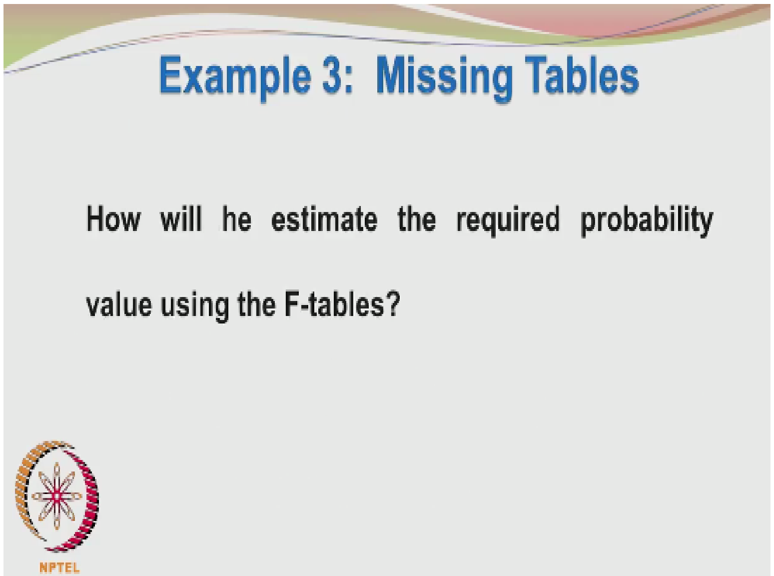
So, concluding the 95% confidence upper bound that will just not include the supplier claimed mean value of 2.3 GW only if the sample mean had been as low as 1.78 gigawatts.

(Refer Slide Time: 25:28)



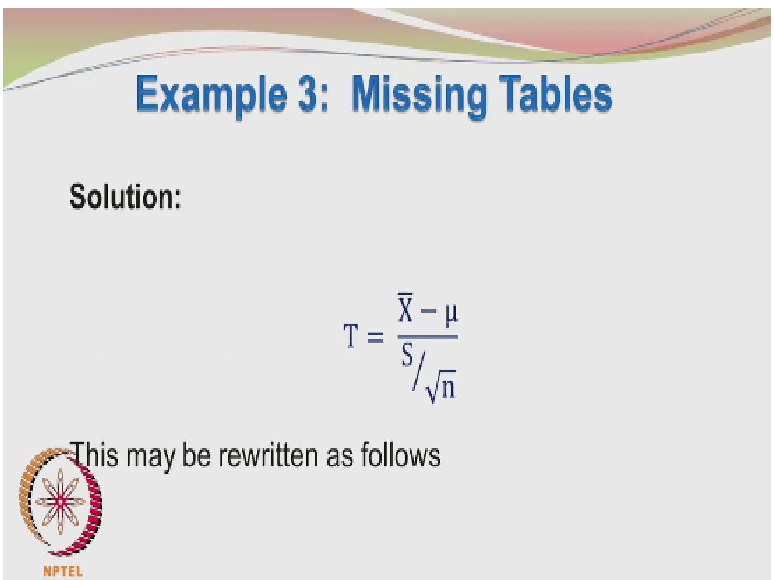
Let us now go to the third example. Normally, when any professional goes out for site visits or field tests or conferences or even vacations, they usually take laptop which has most features like spreadsheet, PowerPoint, etc. But let us imagine a situation, let us say about 20 years back, an environmental engineer carries out some field measurements. He carries out t-test for nine samples from a polluted lake and he obtains modulus t-value of 2.306. So, he has forgotten to take the t tables and he has only the F tables with him.

(Refer Slide Time: 26:30)



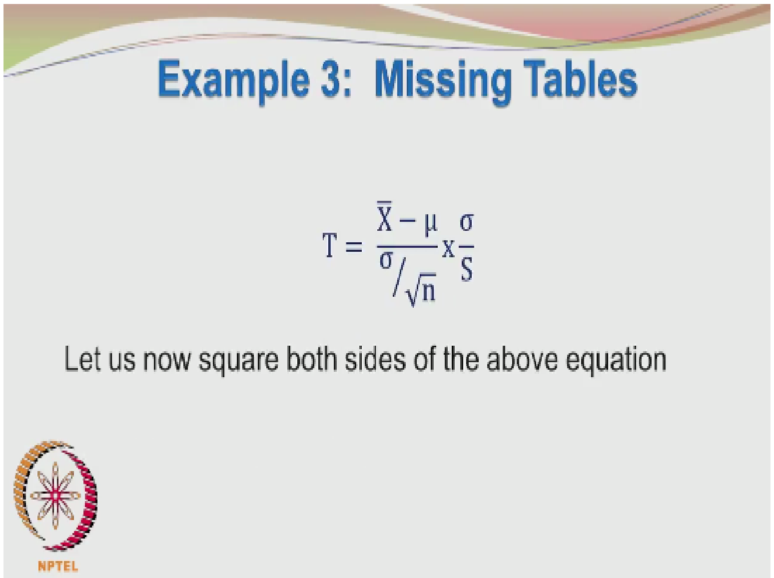
So, how will he find the probabilities using the F tables, that is an interesting problem. Essentially, we have to find the relationship between the T-distribution and the F-distribution.

(Refer Slide Time: 26:48)



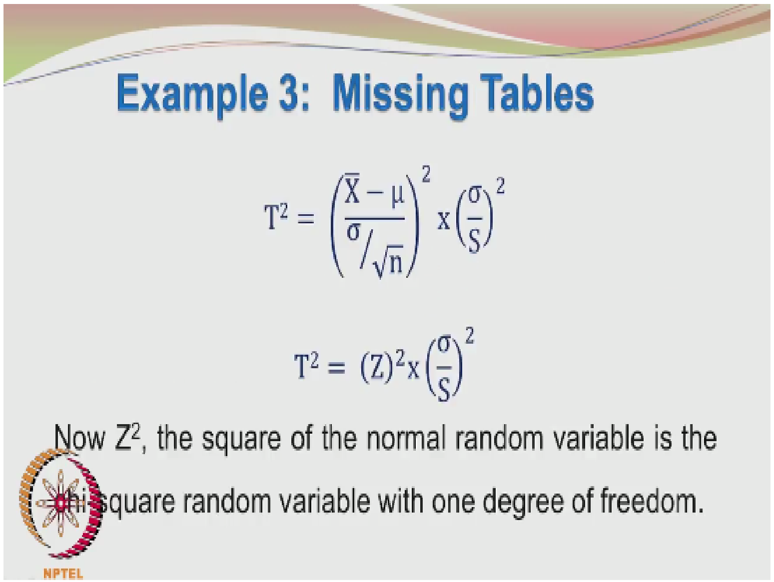
It is rather elegant. We know that the T random variable is given by X bar-mu/S/root n.

(Refer Slide Time: 26:56)



This may be rewritten as T=X bar-mu whole/sigma/root n*sigma/S. So, the sigma here and the sigma here will cancel out and essentially you are having S/root n which is our original definition of the t random variable.

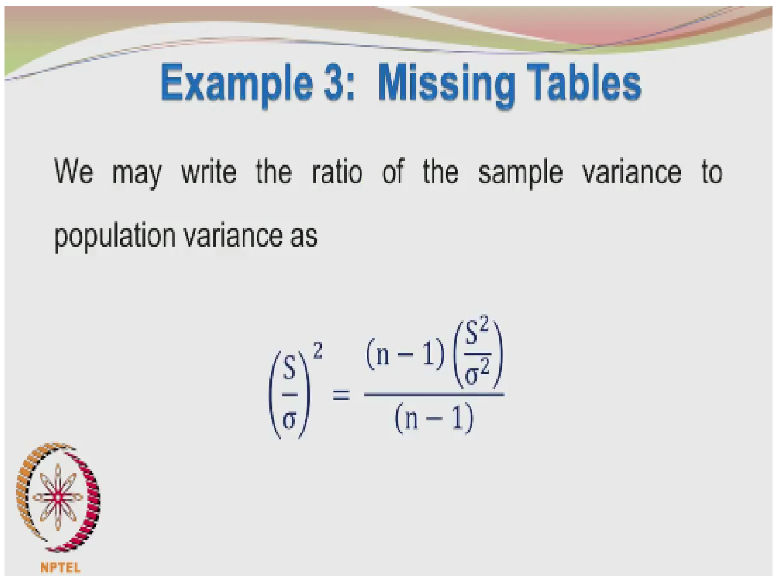
(Refer Slide Time: 27:23)



So, now we can take square on both sides of the above equation and we get T square=X bar-mu/sigma/root n whole square*sigma/S square. So, T square is equal, this becomes the standard the normal variable defined for the sample mean. The sample mean has mean mu and standard deviation sigma/root n. So, when you write X bar-mu/sigma/root n, we get Z. We also assume that the parent population from where the samples have been taken are normal, that is the implicit assumption made when we are using the t-test.

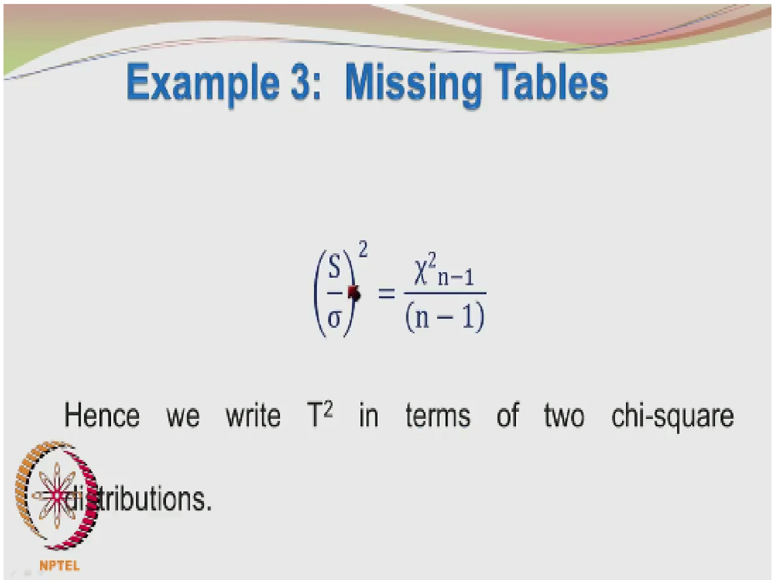
So, the distribution of the sample means also is a normal distribution and so we are able to standardise it in this fashion. So, we get T square=Z square*sigma/S whole square and we know Z square, we are having a single standard normal variable which has been squared is chi-square random variable with 1 degree of freedom. So, we have a chi-square random variable with 1 degree of freedom.

(Refer Slide Time: 28:41)



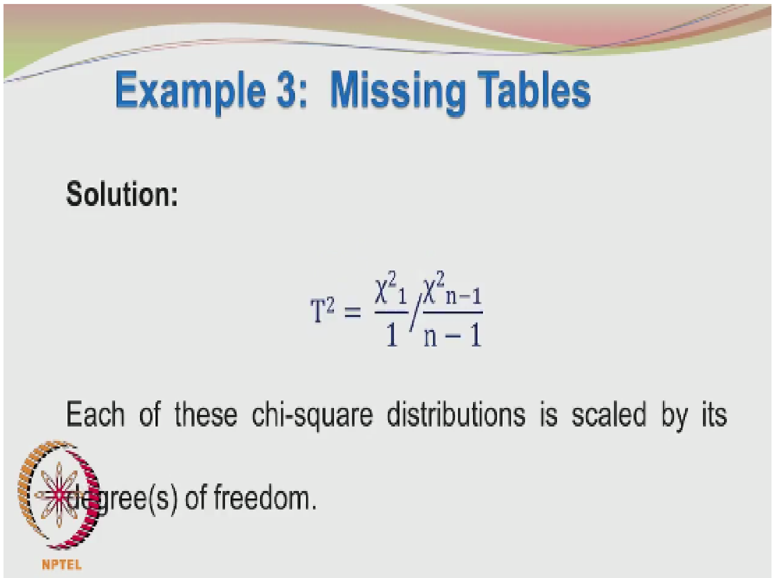
If you are looking at S/sigma square, I can write it as n-1 S square/sigma square/n-1. Of course, n-1 will cancel out in the numerator and denominator. So, we are writing it in this form, so that n-1 S square/sigma square may also be related to a chi-square distribution with n-1 degrees of freedom. So, this is the definition for the chi-square distribution.

(Refer Slide Time: 29:08)



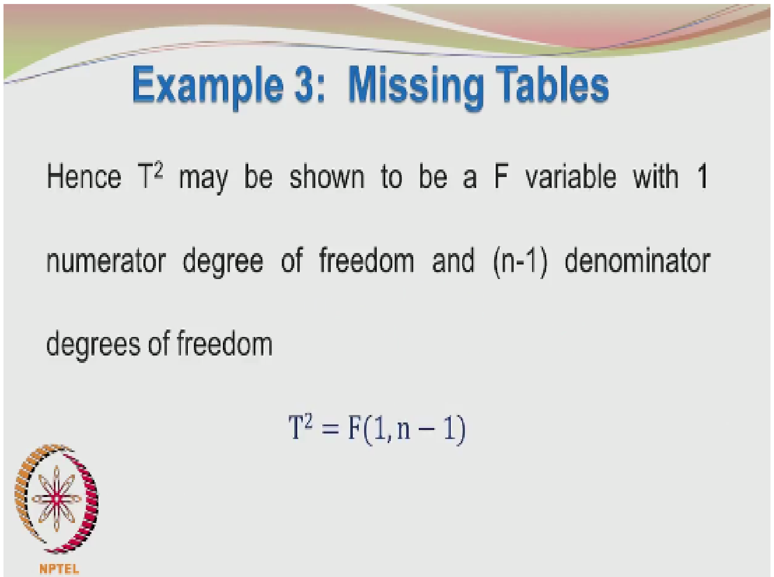
So, we have S/sigma whole square as chi-square n-1/n-1. So, we can write t square in terms of two chi-square distributions.

(Refer Slide Time: 29:22)



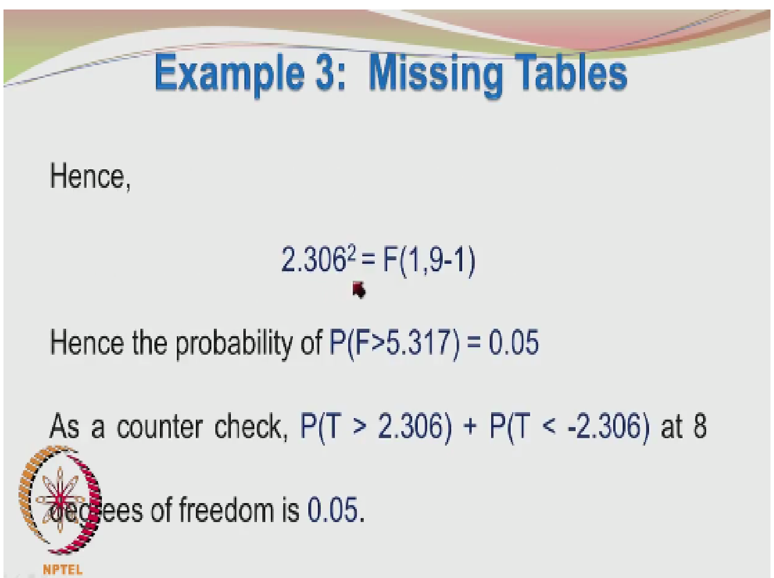
T square is chi-square 1 corresponding to Z square/chi-square n-1/n-1. This corresponds to S square/sigma square. This corresponds to Z square. So, we showed that T square may be written as Z square/S square/sigma square. Z square is written in terms of chi-square distribution with 1 degree of freedom and the S square/sigma square is written in terms of a chi-square distribution with n-1 degree of freedom.

(Refer Slide Time: 30:00)



So, we are having T square like this and the ratio of two chi-square distributions with numerator 1 degree of freedom and the denominator n-1 degree of freedom may be expressed in terms of F random variable with 1 and n -1 as the parameters. So, we can show that T square is equal to F1, n-1.

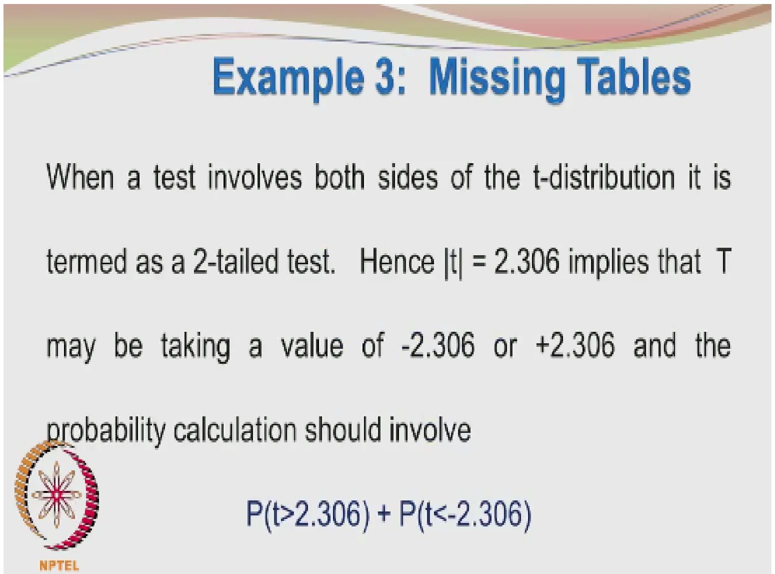
(Refer Slide Time: 30:29)



T is 2.306, so T square is 2.306 square=F 1, 9-1. How did you get this 9, the sample size is 9, so we are going to have 9-1 as the degrees of freedom. So, this leads to probability of F>5.317, we are looking at the upper tail probability obviously, probability of F>5.317 as 0.05. So, if you want to countercheck later when the T tables are available, we can find what is the probability of T>2.306+ probability of T<-2.306 at 8 degrees of freedom and if you add these two, you will get

0.05 again.

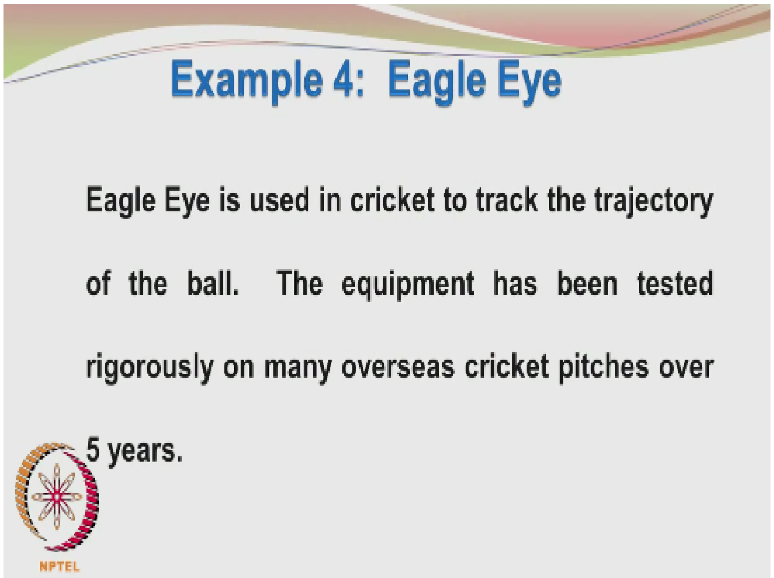
(Refer Slide Time: 31:40)



When a test involves both sides of the T-distribution, it term as a two-tailed test. Since, modulus of 2 is 2.306, it implies that T may take a value of -2.306 or +2.306 and the probability calculation should actually involve probability greater than 2.306, probability of T>2.306+ probability of T<-2.306. So, if you compute this, you will get 0.05 which is same probability value obtained from the F-distribution.

Let us look at the next example. I do not know how many of you are cricket fans or how many of you even know cricket. Anyway, that is a separate story altogether. Let us now look at the actual example even for those people who do not follow cricket. You can just try to identify the main parameters and then work out the problem. So, I think people who were finding certain examples not in their fields or not in their knowledge domain should not get intimidated.

(Refer Slide Time: 33:03)

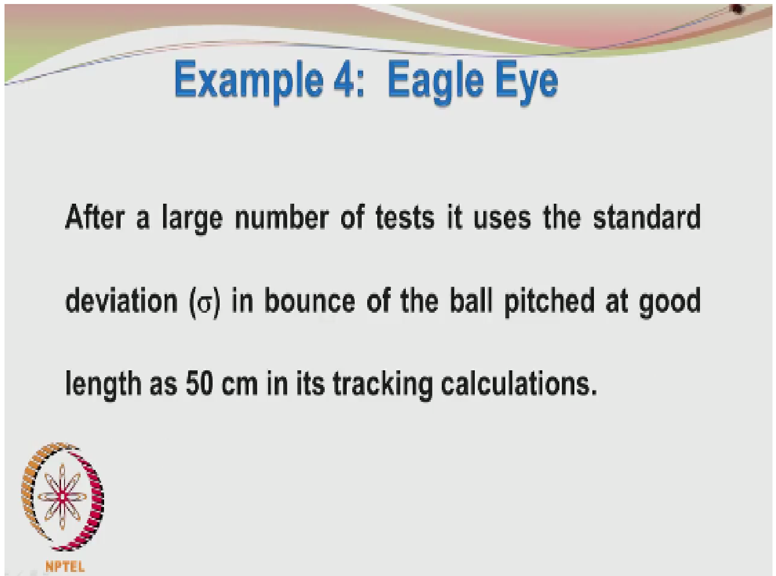


So, coming to the example 4. The title of the example is Eagle Eye. It talks about cricket. There may be example problems which may not be in your field of research or in your field of specialisation but I request in such cases the students and viewers of this particular course should not feel intimidated. The more important thing is to correctly identify the parameters. You have to make sure that have written down the degrees of freedom correctly, that is number 1.

You should also know which is mu1, which is X bar, which is sigma square, which is S square, which is sample statistic and which is a population parameter, that is very important and once you have written these things correctly, you can use the formulae to get the final T value or the chi-square value or the F value and then you have to make sure that you estimate the probability values correctly.

So, the problem statement goes on like this. Eagle Eye is used in cricket to track the trajectory of the ball. The equipment has been tested vigourously on many overseas cricket pitches over five years.

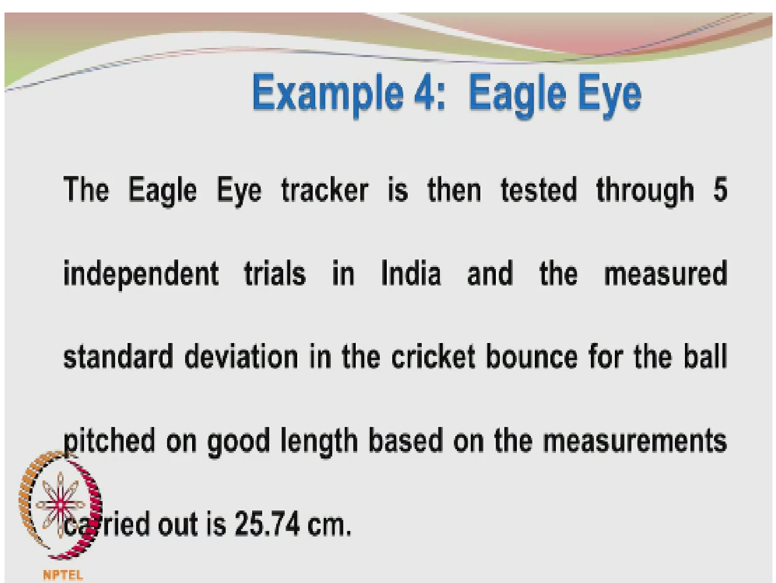
(Refer Slide Time: 34:24)



After a large number of tests, it uses the standard deviation sigma in the bounce of the ball pitched at good length as 50 cm in its tracking calculations. Please note that I am not talking about the average bounce of the cricket ball on overseas pitches. I am only talking about the variability in the bounce of the cricket ball in terms of its standard deviation. So, even the standard deviation is quite high.

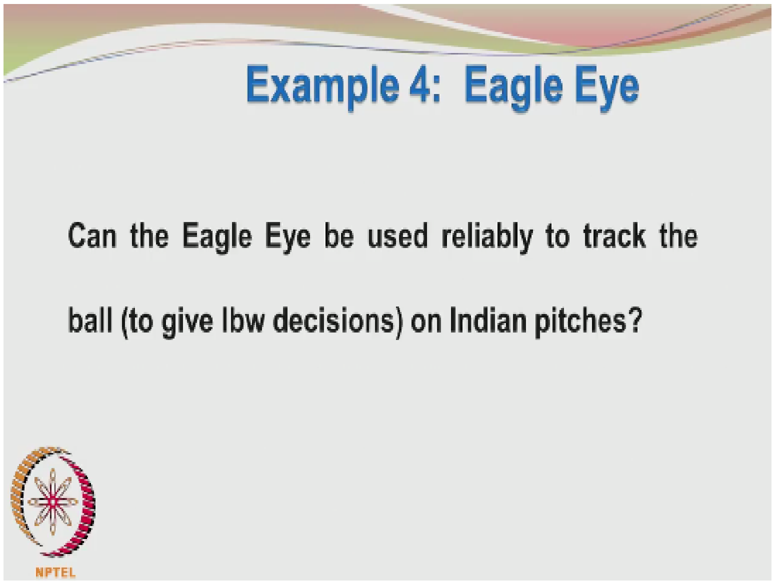
For a ball which is being pitched at good length, the variability in the bounce expressed in terms of a standard deviation is 50 cm. We really do not know what is the average height of the ball pitched at good length. Obviously, it must be better than 50 cm.

(Refer Slide Time: 35:33)



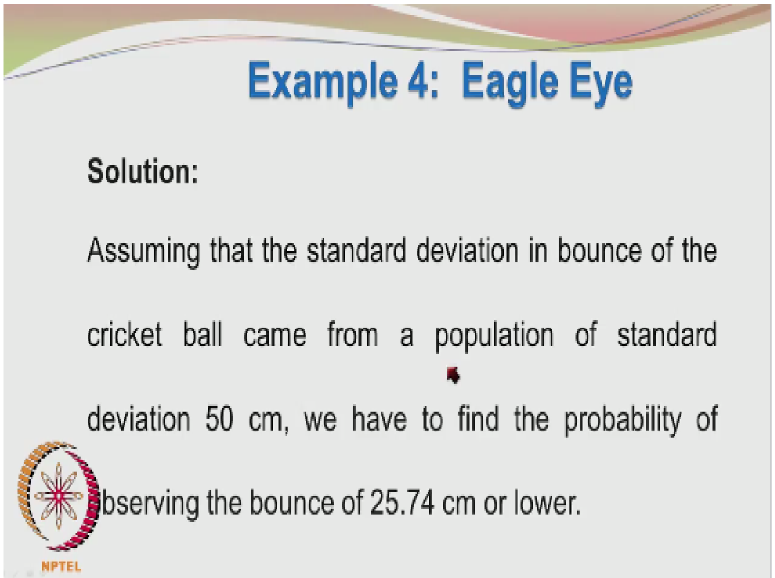
Now, this Eagle Eye Tracker is brought to India and then tested in five independent trials, may be the five major cricketing centres in the country and based on the five trials, the measured standard deviation in the cricket bounce for the ball pitched on good length is only 25.74 cm. Probably the Indian pitchers have more consistent bounce and hence the standard deviation in the bounce of the cricket ball pitched at good length is smaller at 25.74 cm. As far as this problem statement is concerned, we are having a sigma value of 50 cm and the measured standard deviation which obviously is S from the random sample is only 25.74 cm.

(Refer Slide Time: 36:38)



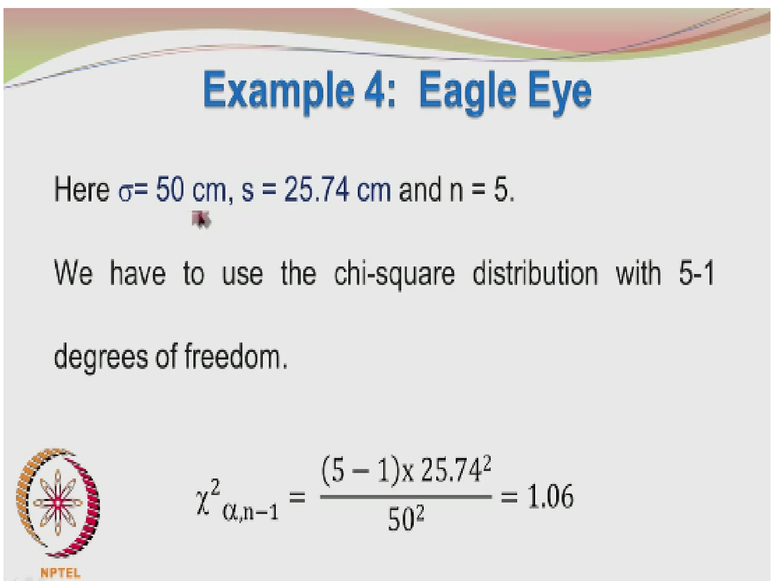
So, can the Eagle Eye be used reliably to track the ball to give LBW decisions on Indian pitches because the LBW or leg before wicket decisions in cricket is usually made on the trajectory of the ball after pitching. Anyway, let us not go too much further into the details.

(Refer Slide Time: 37:07)



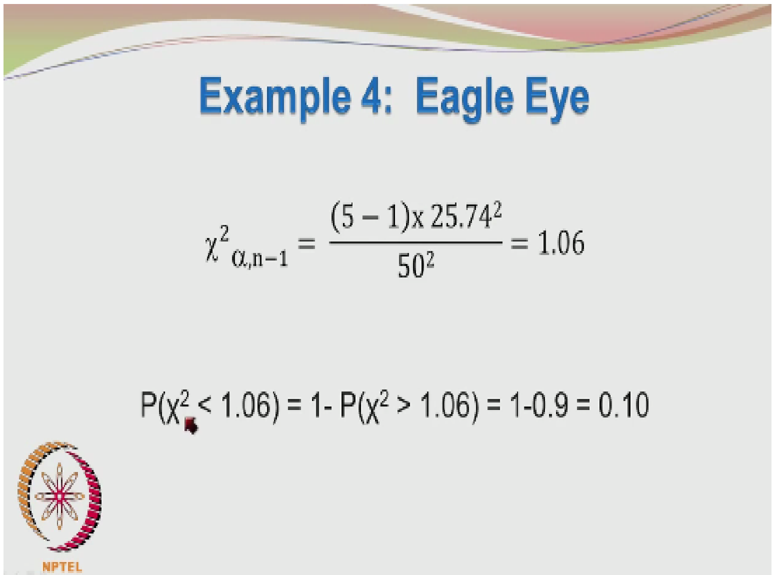
So, the solution is, so we have to assume that the standard deviation in bounce of the cricket ball came from a population of standard deviation 50 cm. So, we have to find the probability of observing the bounce of 25.74 cm or lower when the population standard deviation is 50 cm. So, if I am taking a random sample from a population of 50 cm, what is the probability of occurrence of the random sample statistic value being 25.74 cm or lower.

(Refer Slide Time: 37:49)



So, sigma is 50 cm, S is 25.74 cm and n is 5. So, the chi-square distribution with 4 degrees of freedom is used and chi-square alpha n-1 is n-1 S square/sigma square n-1 is 5-1 and S square is 25.74 square and 50 square is the sigma square. Here sigma is known. So, this chi-square value comes to 1.06. So, we have to find the probability of the chi-square random variable taking on values 1.06 or lower.

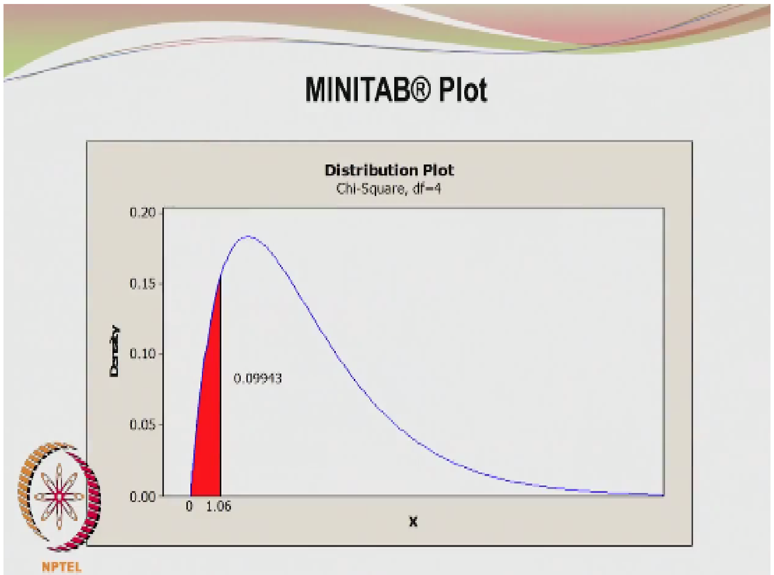
(Refer Slide Time: 38:41)



So, probability of chi-square less than 1.06, that comes as 1-0.9 which is 0.1. So, there is a 10% chance that these sample with the standard deviation in the bounce as 25.74 cm could have indeed come from a population with the standard deviation of bounce of 50 cm. Whether this is a low probability or a high probability, it is up to decision-makers. Normally, we specify alpha value to be 0.05.

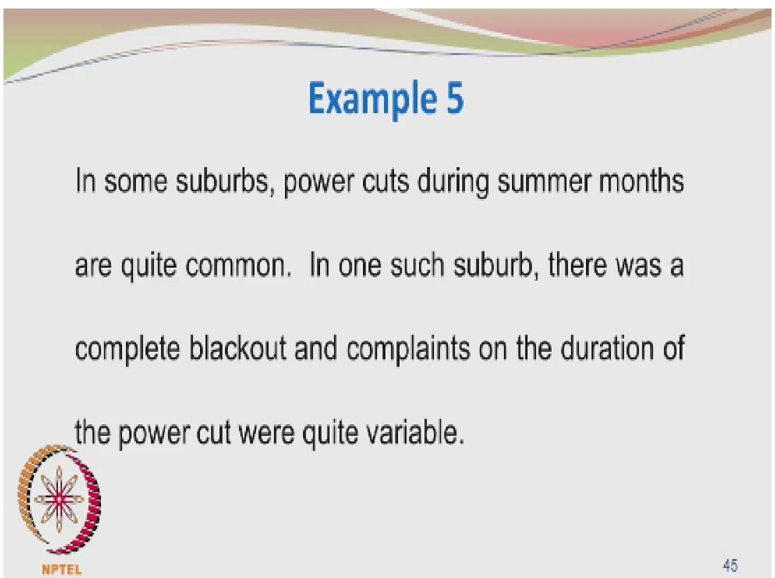
So, here we have got 0.10. So, this probability value of 0.1 is low or high is left to the administrators of the sport. We will just report the value and then move on.

(Refer Slide Time: 39:39)



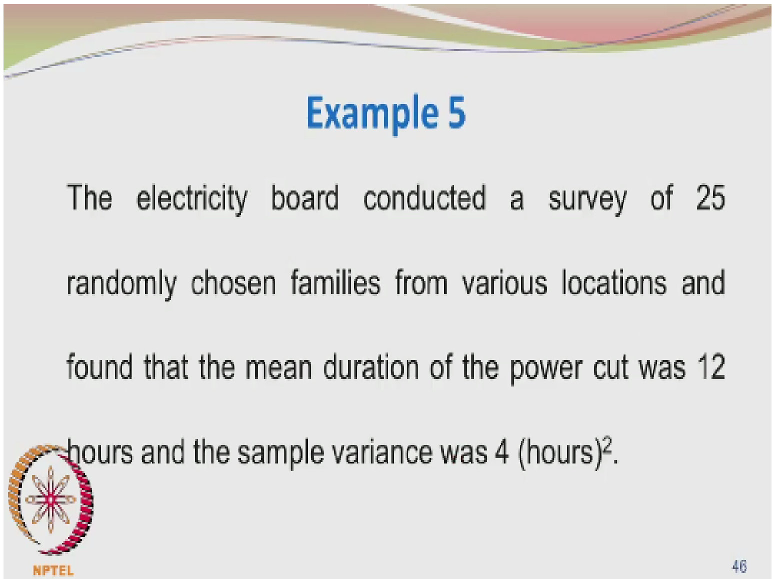
So, plotting this using Minitab, we can see that the chi-square variable was 1.06. The value taken by the chi-square random variable was 1.06 and the probability below that is 0.099 or pretty much 0.1. Please note that the degrees of freedom is equal to 4 and this is a chi-square plot. You can see that it is skewed to the left.

(Refer Slide Time: 40:12)



Going to the next example, well I do not know how many of you are living in cities which experience lot of power cuts during summer. According to this problem statement, in some suburbs, power cuts during the summer months are quite common. In one such suburb, there was a complete blackout and complaints on the duration of the power cut were quite variable, okay. When there are a lot of power outages or shutdowns, different localities will experience different lengths of power failure, so the complaints on the duration of the power cut were quite variable.

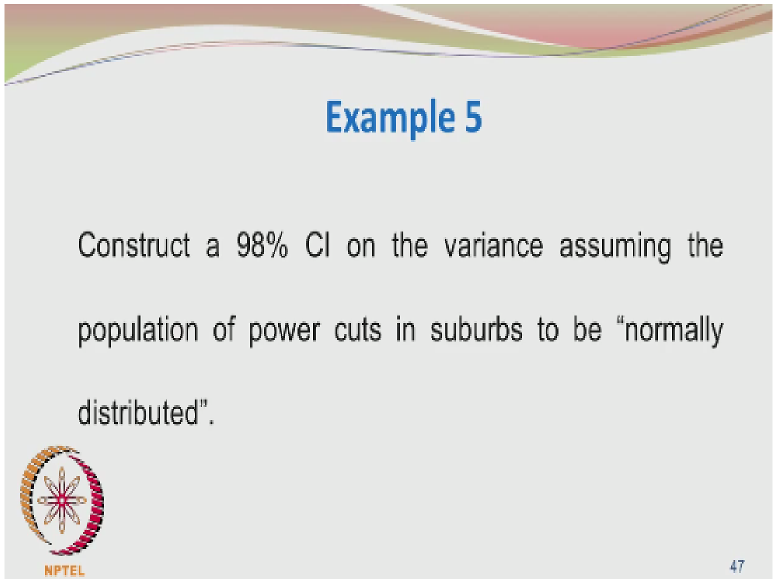
(Refer Slide Time: 40:58)



So, the Electricity Board conducted a survey of 25 randomly chosen families from various locations and found that the mean duration of the power cut was 12 hours and the sample variance was four hours square, right. So, the sample variance is four hours square and the sample size is 25 and the associated degrees of freedom would be 25-1 which is 24. The sample mean is of course 12 hours. We will not really use it.

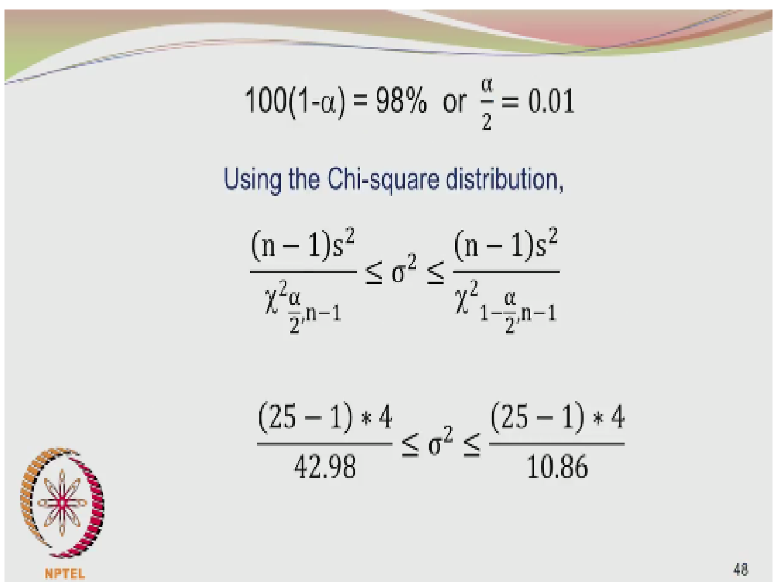
If actual data had been given on the length or duration of the power cuts, we could have used those actual data and then the sample mean of 12 hours to find a sample variance. But in this particular case, the sample variance is directly given to us, so we really do not use the sample mean.

(Refer Slide Time: 42:11)



Moving on, what do we have to do. We have to construct a 98% confidence interval on the variance assuming that the population of power cuts in suburbs is normally distributed. Again, this example is completely fictitious.

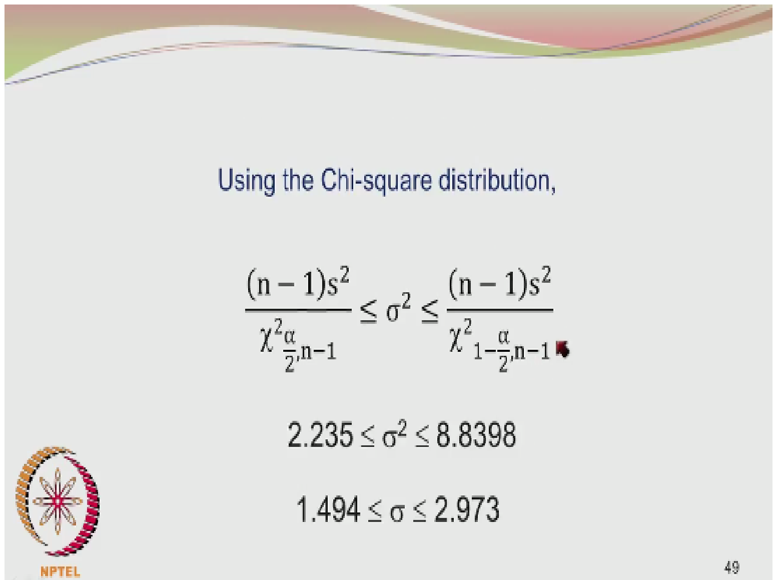
(Refer Slide Time: 42:32)



Let us see how we go about constructing the 98% confidence interval. So, we have 100*1-alpha=98% or 1-alpha is 0.98. So, alpha would be 1-0.9*(()) (42:53) 0.02, so alpha/2=0.01. So, we can use the chi-square distribution confidence interval to find upper and lower bounds for sigma square. So, we have n-1 S square/chi-square alpha/2 n-1 < or =sigma square< or =n-1 S square/chi-square 1-alpha/2 n-1.

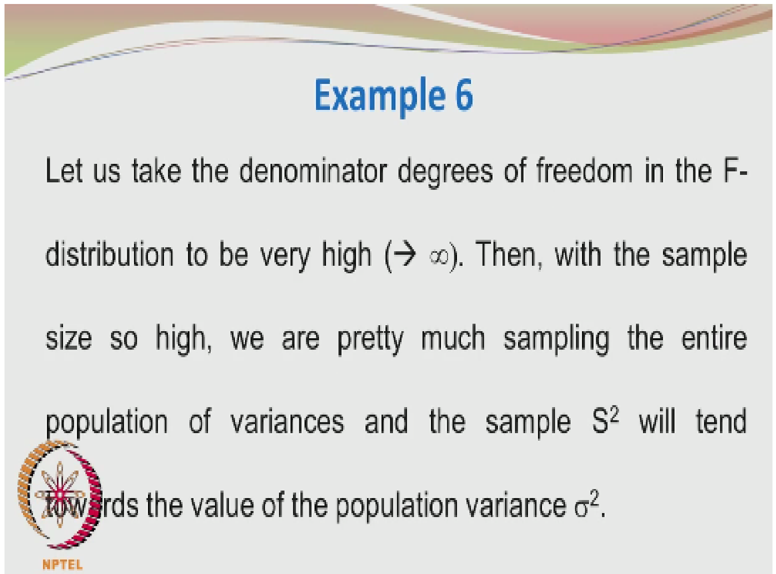
So, sample size was 25, 25-1*S square was four hours and that divided by chi-square 0.01, 24. That value we can see from the tables as 42.98 and similarly we do the same thing here, 25-1*4 and chi-square 1-alpha/2 n-1. Here, we use 1-alpha/2 please note. So, we have to find out chi-square 1-0.01. So, we have to find out chi-square 0.99 24.

(Refer Slide Time: 44:06)



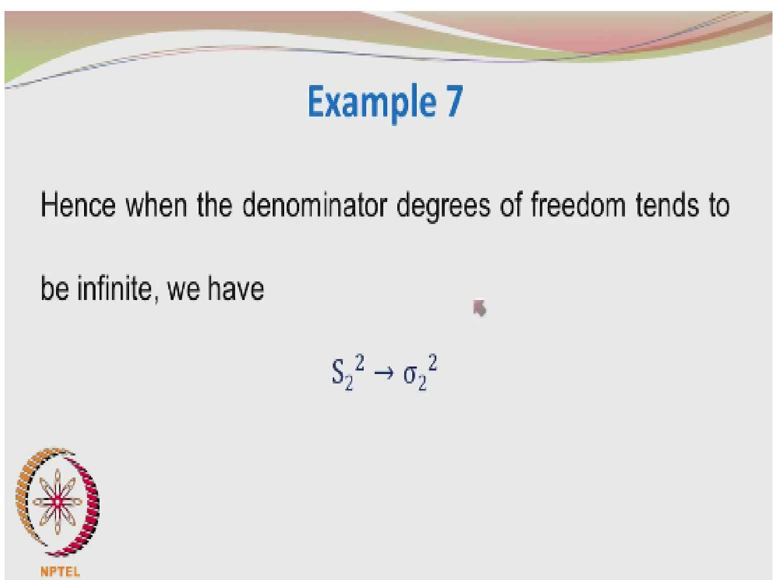
So, reading the numbers from the tables, we get to 2.35< or =8.8398. So, the standard deviations obtained by taking the square root. So, 1.5 nearly to 3. So, the standard deviation of the power cuts is falling between 1.5 hours to 3 hours. Let us go to example number 6. How will you estimate the probability in distributions involving the chi-square using the F-distribution tables. Earlier, we saw how to relate the T-distribution with the F-distribution. Now, we are trying to relate the chi-square distribution with the F-distribution.

(Refer Slide Time: 45:11)



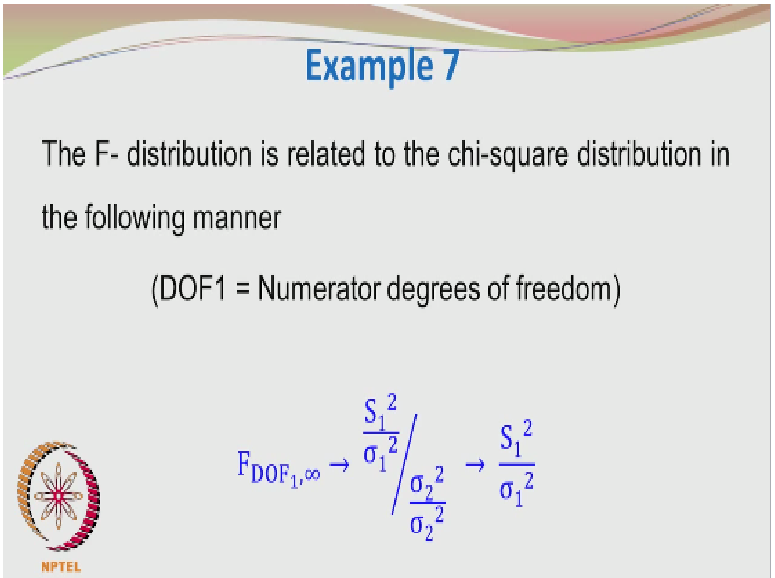
If we take the denominator degrees of freedom in the F-distribution to be very high tending to infinity, then there is a simplification possible. When the sample size is so high, we are pretty much sampling the entire population of variances and the sample S square will tend towards the value of the population variance sigma square. When you take a larger and larger sample, it is as if you are sampling the entire population or as if you are finding out sigma square itself directly. So, S square will approach sigma square and that helps us to simplify a few things.

(Refer Slide Time: 45:53)



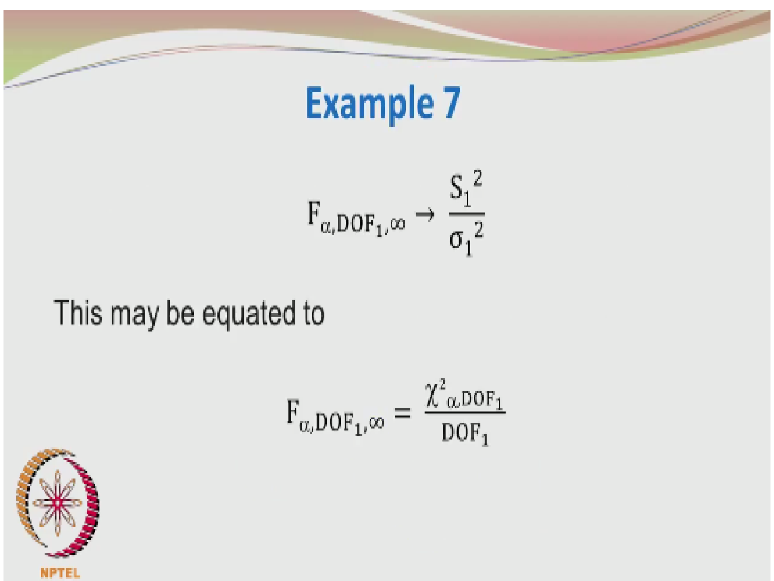
So, when the denominator degrees of freedom tend to infinity, we have S2 square tending to sigma 2 square, what happens then.

(Refer Slide Time: 46:03)



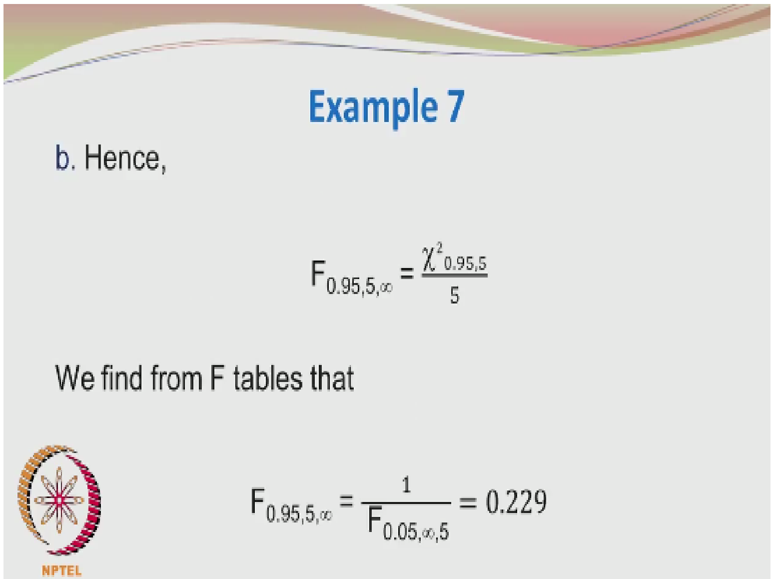
We know that the F-distribution is given by S1 square/sigma 1 square/S2 square/sigma 2 square. Since, S2 square approach sigma 2 square it will become 1. The ratio of S2 square/sigma 2 square will become 1. So, we are only left with S1 square/sigma 1 square, that we represent as F degrees of freedom in the numerator, infinity. So, DOF1, infinity tends to S1 square/sigma 1 square only.

(Refer Slide Time: 47:06)



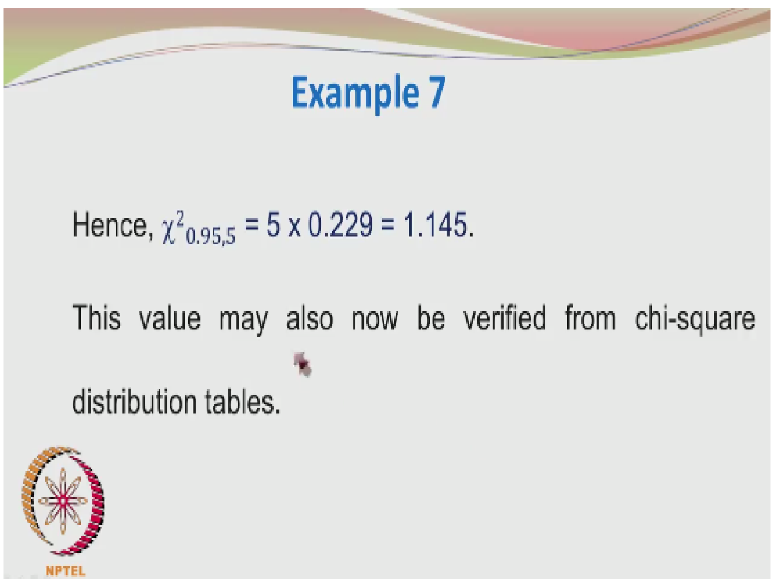
So, we may write this as DOF1 S1 square/DOF1sigma 1 square. We also know by definition of chi-square alpha DOF1 as DOF1 S1 square/sigma 1 square. So, when we do that, this term becomes chi-square alpha DOF1/DOF1. So, F alpha DOF1 numerator degrees of freedom, infinity may be written down as chi-square alpha DOF1/DOF1. It is very simple. You may want to work it out on a piece of paper.

(Refer Slide Time: 48:03)



So, suppose we want to find F of 0.955 infinity, we have to find chi-square 0.955/5. If you just look at this, numerator degrees of freedom used here, used here and then here, right. So, F of 0.955 infinity is 1/F of 0.05 infinity 5 which is 0.229.

(Refer Slide Time: 48:47)



Hence, chi-square 0.95,5 is 0.229*5 which comes to 1.145. Independently, you can use the chi-square distribution chart to verify that the value of the chi-square random variable which has an upper tail probability of 0.95 for 5 degrees of freedom is 1.145.

(Refer Slide Time: 49:30)

Percentage Points of the T-distribution for different Probabilities and Degrees of Freedom

DOF	Probability									
	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.30	636.61
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073

NPTEL

So, these are tables of T-distribution.

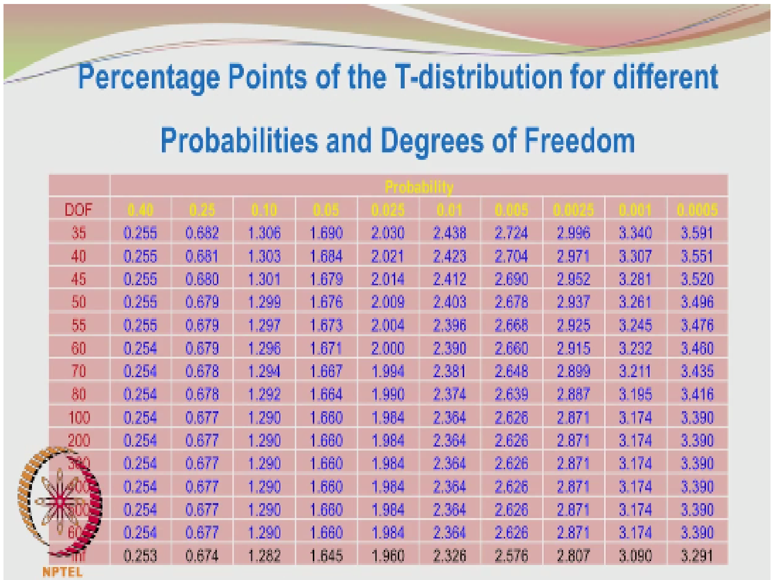
(Refer Slide Time: 49:32)

Percentage Points of the T-distribution for different Probabilities and Degrees of Freedom

DOF	Probability									
	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646

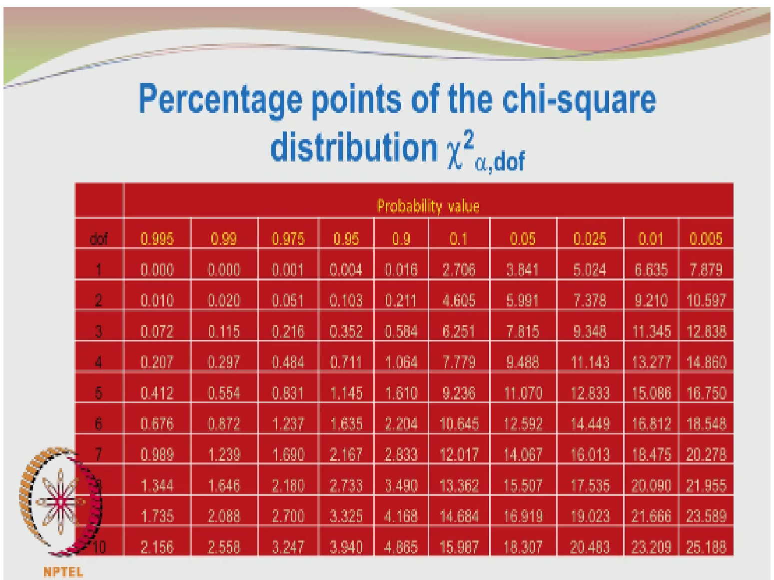
NPTEL

(Refer Slide Time: 49:33)



	Probability									
DOF	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
35	0.255	0.682	1.306	1.690	2.030	2.438	2.724	2.996	3.340	3.591
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
45	0.255	0.680	1.301	1.679	2.014	2.412	2.690	2.952	3.281	3.520
50	0.255	0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
55	0.255	0.679	1.297	1.673	2.004	2.396	2.668	2.925	3.245	3.476
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	0.254	0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	0.254	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
200	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
[illegible]	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
[illegible]	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
[illegible]	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
[illegible]	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
Inf	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

(Refer Slide Time: 49:34)



	Probability value									
dof	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
[illegible]	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
[illegible]	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

So, when we look at the chi-square, so we want a probability of 0.95 with 5 degrees of freedom. So, you can see that for probability of 0.95 with 5 degrees of freedom it is 1.145 and that is matching with the value given when using the F distribution. So, this completes our discussion on the T, chi-square and F distributions, okay. So, we have seen a few illustrative examples.

We also showed that these distributions may be elegantly related to each other. There are several textbooks available on probability and statistics. Essentially, we have only covered the distributions, the random variables so far but these are most important in the design of experiments and analysis of experiments. Without this background, it will be impossible for you to really appreciate the various results that are reported in design of experiments.

To restate it in a different way, if you have a good appreciation and understanding of the normal distribution, T-distribution, chi-square distribution and F-distribution, you will have a firm grip on the concepts involved in design of experiments. Now, there is an important bridge linking these are distributions with the design of experiments and that will be done through the hypothesis testing, that will form the basis for our future lecture or maybe a couple of lectures.

What I would like to emphasize at this point is, please try to solve as many problems as possible and it is not only enough if you get the answer correctly but also try to understand what the answer is telling to you and how you will interpret and apply the result in real world situations. Thank you. So, we will continue on hypothesis testing in the next lecture.