INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR

NPTEL
National Programme
On

Technology Enhanced Learning

Applied Multivariate Statistical Modeling

Prof. J. Maiti
Department of Industrial Engineering and Management
IIT Kharagpur

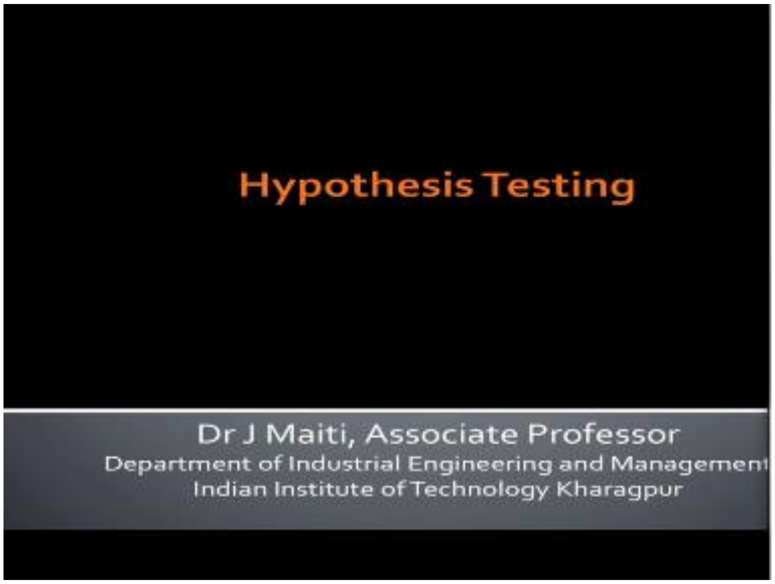
Lecture – 07

Topic

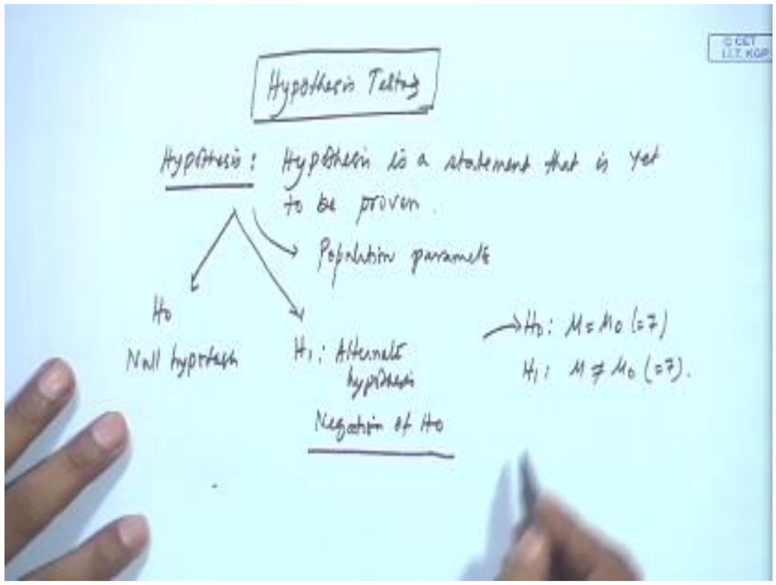
Hypothesis Testing

Today we will discuss Hypothesis testing.

(Refer Slide Time: 00:24)



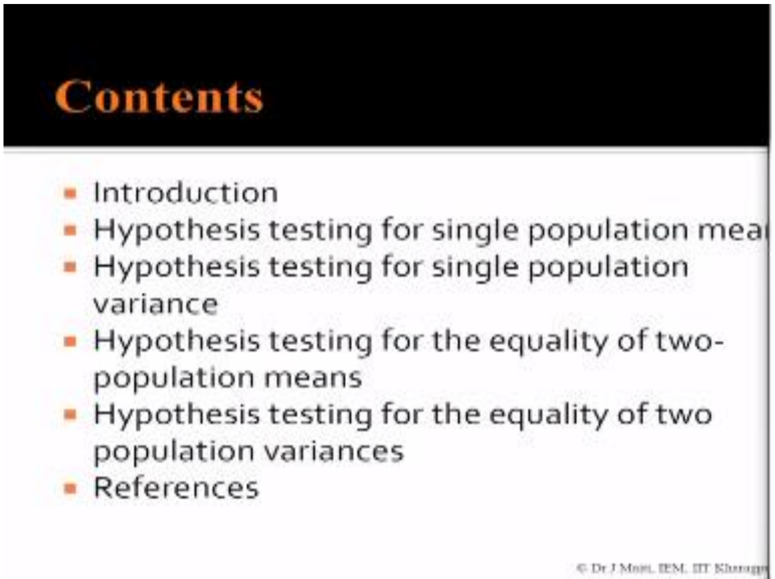
(Refer Slide Time: 00:27)



So what is hypothesis? You have any idea about hypothesis? If I ask you, what is hypothesis? Hypothesis is a statement that is yet to be proven. Usually in statistics we frame hypothesis concerning the population parameter, hypothesis power population parameter. So if you read the history of science you will find out that n number of hypothesis. There are huge number of hypothesis has been framed by the scientist and proven by experiment or by some other means.

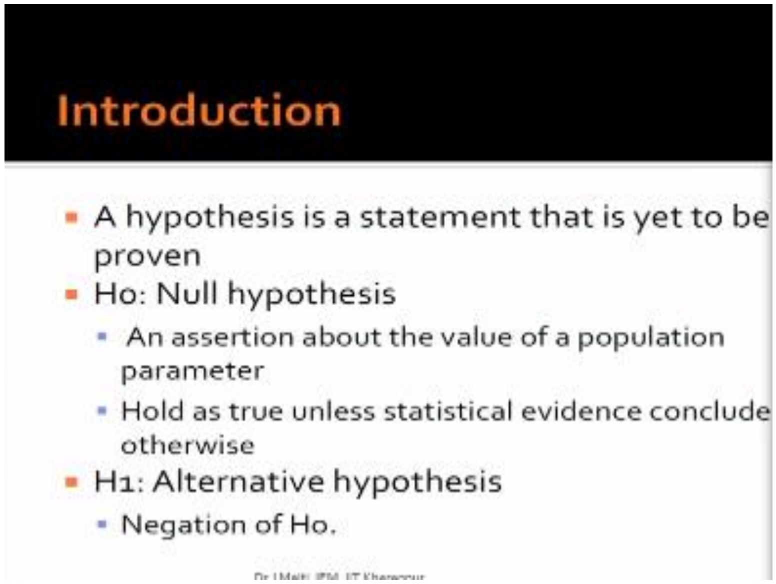
So our hypothesis is limited to the statistical hypothesis testing and before going into detail of this let us see the content of today's lecture.

(Refer Slide Time: 02:09)



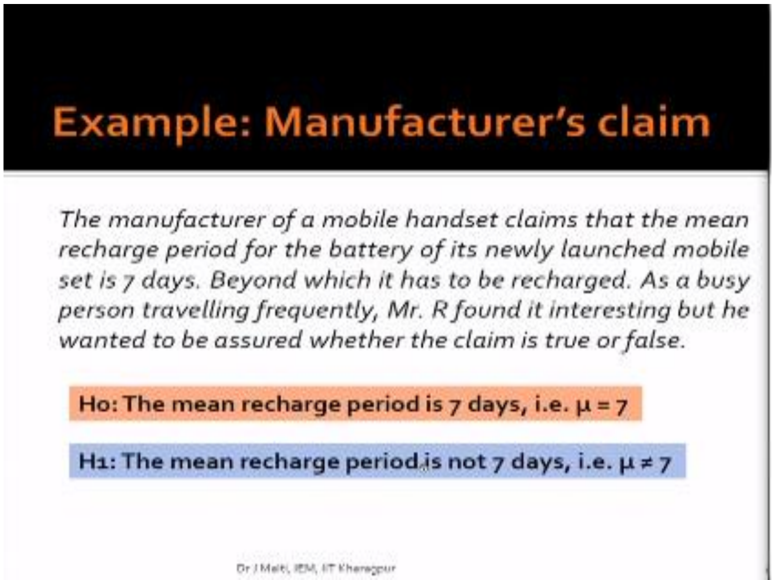
We will discuss the hypothesis testing for single population mean for single population variance for equality of two population means and equality of two population variances.

(Refer Slide Time: 02:29)



So come back to this hypothesis testing and I have discussed the hypothesis statement that is yet to be proven and there are we will consider two types of hypothesis. One is HO which is known as null hypothesis and another one is H 1 or H which is known as alternate hypothesis. Null hypothesis is an assertion about a population parameter and we believe on it unless it is proven statistically otherwise. And alternative hypothesis is the negation of H_0, okay let us see one example here.

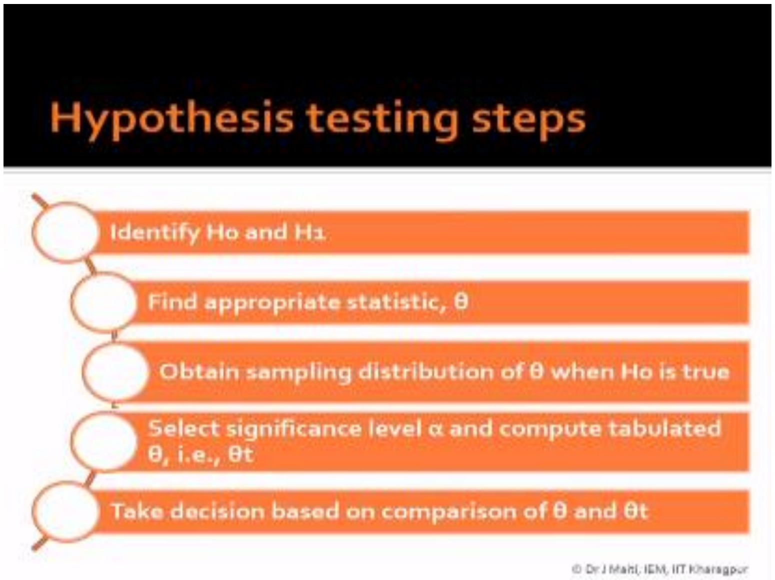
(Refer Slide Time: 03:24)



The manufacturer of a mobile handset claims that the mean recharge period for the battery of its newly launched mobile set is 7 days beyond which it has to be recharged. As a busy person travelling frequently mister R found it interesting but he wanted to assured whether the claim is true or false. So, here our null hypothesis is the mean recharge period is 7 days which is the population mean. We are basically talking about like this that null hypothesis μ is equal to μ_0.

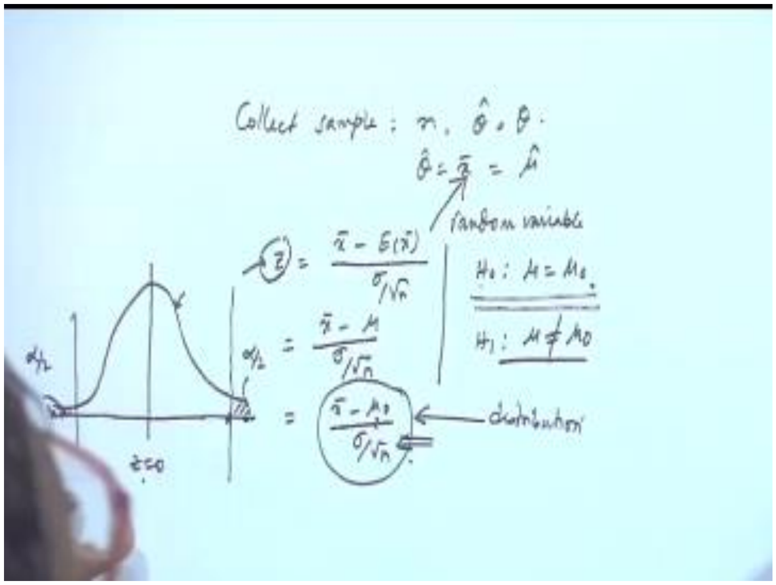
In this case μ_0 is 7 and alternative hypothesis is that μ not equal to μ_0. That means not equal to 7 and if this is the case as I told that we will believe on null hypothesis, so long we are not able to prove that it is wrong. You have to prove it statistically. So what is what is then the, what are the steps? The steps are.

(Refer Slide Time: 04:55)



Identify first null hypothesis and alternative hypothesis, and then you definitely find out the appropriate sampling statistic. Then obtain sampling distribution θ if the statistic is θ, when the null hypothesis is true. Please keep in mind that this is very important concept that when null hypothesis is true that time you are finding the distribution of θ that is the sample statistic.

(Refer Slide Time: 05:29)



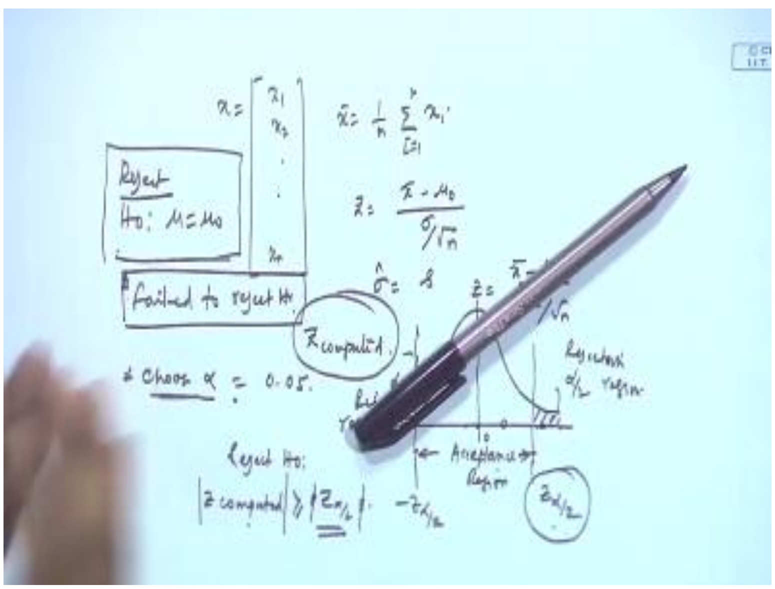
What do you do? You collect sample, collect sample. Let size n and you compute the parameter from the sample and what we are trying to say here. I told that what is basically the statistic is θ instead of cap we have given only θ. For example, this theta will be $\bar{x}$ or $\hat{\sigma}$ cap that is $\bar{x}$ which is basically estimated of μ. Now, this $\bar{x}$ is a random variable and what you require to know. You require framing an appropriate statistics. If it is $\bar{x}$, fine if it is not $\bar{x}$, something else that statistics you find out. For example when we need to talk about $\bar{x}$ we usually frame z that is $\bar{x}$- $/\sigma/\sqrt{n}$ that we usually frame here.

Our null hypothesis is H_0 $\mu = \mu0$. We all know that expected value of $\bar{x}$ is μ, so then we write $\theta/\sqrt{n}$. Now we have assumed that H_0, from H_0 we have we have seen that $\mu = \mu0$. So, that means you can write this one, $\mu0/\ \sigma/\sqrt{n}$. Then if this is my statistic for which you require to know the distribution, what I said here that is obtain the sampling distribution of theta when H_0 is true. That means the sampling distribution of the here your here your $\bar{x}$ which is nothing but we converting it into z and we are writing μo instead of μ. As we are hoping that our null hypothesis is true.

And then what you will do? You find out the critical value. What do you mean by critical value? In this case if it is z distributed, my distribution will be like this. This is z, then we want to see that basically this one is z equal to 0.Now, this quantity $\bar{x}$ - minus $\mu o/\sigma/\sqrt{n}$, this quantity follows distribution, all possible values, all values here they are possible if it is sufficiently away from that z, that mean value z mean value, then we will conclude that μ is not equal to μo.

So, that is why we will frame critical value in this side. Either this may be your, this is your $\alpha/ 2$ and on this side sufficiently away. This side, this is your, $\alpha / 2$.What I mean to say, if the commutate statistics which one is this one. In this case if this value falls in the right hand rejection region or left hand rejection region then we conclude that H_0 is rejected, not true. That mean H_1 is can be accepted, that is μ not equal to μ_0, and then computational what you will do first? You collect data that what you have done in this particular case.

(Refer Slide Time: 09:26)



What we will proceed? You will precede like this, you collect data for x which will be x_1, x_2 like x_n. Then you will compute $\bar{x}$ which is 1 / n i = 1 to n, x_i. Then you will compute the statistic which we are talking about z, that $\bar{x}$ - $\mu o/\sigma/\sqrt{n}$. Now, please remember that root sigma will be given. If σ is not given then $\hat{\sigma}$ cap will be s and in that case if n is sufficiently large then you will be using z distribution like this, $\bar{x}$ - μo/σ/√n. This is the computed z, basically from the data z computed.

Now you have to choose α. So, your next step is choosing α, you have seen the, what is the error. You are going to consume if you say α is 0.05, this is also known as probability level of significance. If we say α = equal to 0.05 then you are this is your point will be like this, your rejection region will be like this. Now, suppose this one is α/ 2 and this side, this is also α /2 then this is your rejection region.

This side, this one is rejection region and in between this is the accepted acceptance region. So, you have your distribution and you know that level of error you are trying. We will consume and then what happen based on this you will find out an interval which is acceptance interval and beyond as it is a two tailed case. So, either right hand or left hand we need to go beyond the critical value. That is z α / 2 right hand side and -z α /2 left side. When you go beyond this you will reject the null hypothesis.

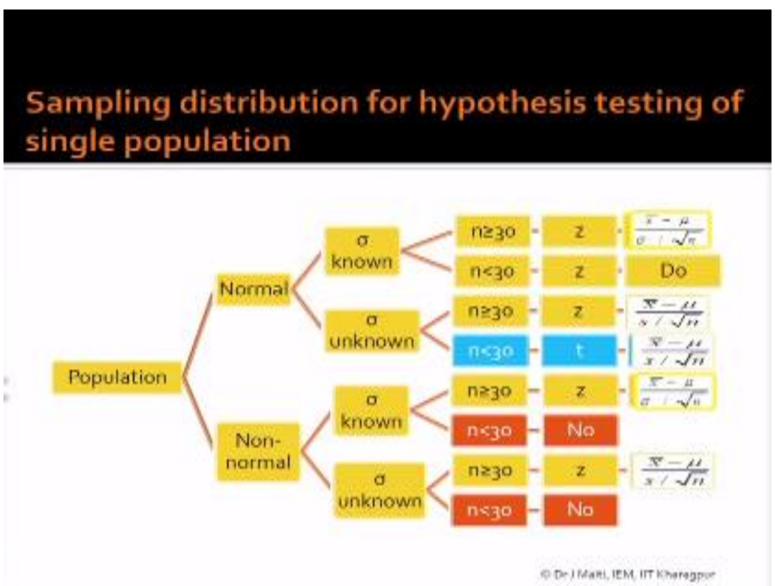
So, what you do? Then you have already computed z, from table you are getting the z α / 2, either it is the mod value we have to take. If z computed, if z computed is greater than mod of z α / 2 absolute value then what will happen that either you will come into this side or this side. Otherwise this maybe negative also, you do one thing you just change little bit that mod of this will also come under this side. I will take mod in both sides.

So, if this one is my zero value. Now, if z value is here this side it will be positive and this side it will be negative and you will be seeing the alpha value from that z α / 2 value from the table. Usually this one will be a positive value. So, you may not take mod here, no problem. So, the absolute value if it is more than z α / 2, more than or equal you can write. So, that means what I am saying here, we are saying that if it falls here or it falls here then it is sufficiently away from the mean value.

So, H_0 mu equal to μ_0 that can be rejected if I reject H_0 that means I am accepting $_{H0.}$ Alternatively I can say I accept H_0, H_1 not 0, you are rejecting this. So, actually you either will be able to reject H_0 or you will be supposing your z value comes here that is within this acceptance zone. Then you will say failed to reject H_0, so this is the procedure. The procedure says first you must know that what is the problem and then the appropriate variable you will find out from the population point of view.

You collect sample, from that sample appropriate statistics you generate, then you create the null hypothesis and alternative hypothesis for the population parameter of interest. Then using the statistic as well as its distribution then you compare the computed value of the statistic and as well as the critical value from the sampling distribution of that statistic when you compare. If you find out that the computed value is more than the absolute value of that computed statistic is more than the tabulated value then you will reject null hypothesis. Otherwise you say failed to reject null hypothesis, okay?

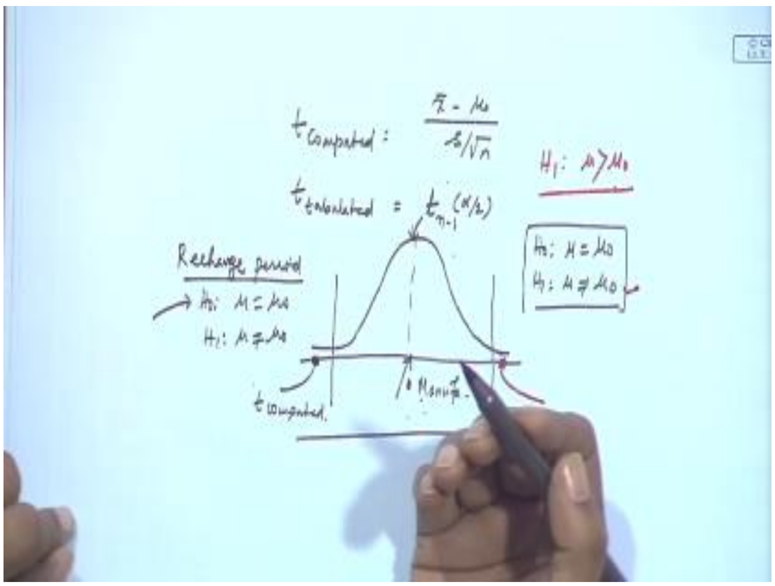
(Refer Slide Time: 15:29)



You see this figure and I know that last class also you have seen the same thing. What we are talking about here, if it is population from normal or non normal whether σ known or σ unknown and the condition of the sample. That means the size of the sample whether it is large or small and depending on this all of we have seen under interval estimation that your quantity that statistics may follow z distribution may follow t distribution or we may not get any distribution, parametric distribution.

So, this one $\bar{x} - \mu/\sigma/\sqrt{n}$, that is for the first case, like second case also same, third case your σ is replaced by s. Fourth case also σ is replaced bys but sample size small. So, that is why it is t distribution, okay? So, if you use the t distribution what is the same thing, you will calculate.

(Refer Slide Time: 16:33)



The statistic that is t computed will be your $\bar{x} - \mu_0/s/\sqrt{n}$, and then you will be finding out t tabulated. The t tabulated means basically you are required to know the degree of freedom, n - 1 and α / 2 because t distribution is also a two tailed distribution and your hypothesis also two tailed hypothesis. The way you have created hypothesis, you have created hypothesis like this. H_0 $\mu = \mu_0$ and H_1 μ not equal to μ_0 that mean both side is open for you.

So, if it goes this side or this side, beyond this level you are not accepting the null hypothesis or rejecting the null hypothesis. So, this is your t distribution and the mean value will be definitely 0. Here is one important point when you go for two tailed distribution, when you go for one tailed distribution. Sir for mean distribution, means only positive values are there. Is it true that for all the time that for mean you will go to two tailed or there is something different. Suppose, here in this example, we say that recharge period that is our example recharge period.

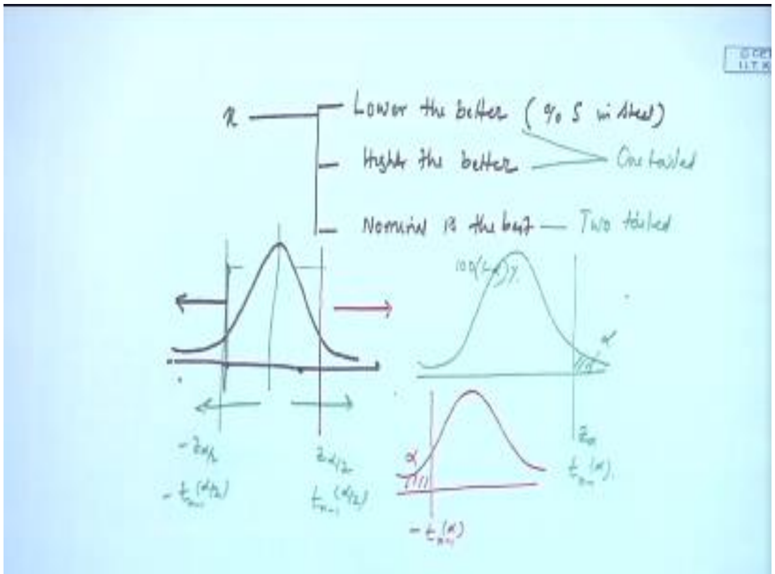
As a user what you want? You want the more the recharge period it is the better. Now, for recharge period if you test like this that H_0 $\mu = \mu_0$ and H_1 μ not equal to μ_0. Suppose, you are that computed value comes here, z value or t value. Suppose, t computed is here. Now, this is our, this is our smallest to largest. So, the manufacturer is claiming that it is here mean value, it is the manufacturer claim. He is saying that this is the mean value.

Now, you are testing that whether the manufacturer claim is true or wrong. Now, you have collective data and you have found that t computed or z computed depending on the distribution. It is in this case, we are talking about t distribution, it is falling here. So, null hypothesis is rejected fine, that means alternative hypothesis is accepted. So, in this case do we go for the mean set, we will not go because the mean value is much lower than claimed. What will be in your favor?

In your favor if the value lies in other way, that means if you are if the computed value comes here that is your favorable case, you are rejecting so long it is in between. This it is nothing but what is the manufacturer is claim, if it comes here this side it is more than that. So, what you may be interested to create null hypothesis like this, $\mu > \mu_0$.You may not be interested to test this μ not equal to μ_0, you may be interested to test $\mu > \mu_0$. Means the manufacturer is claiming it is 7 days, fine 7 days is good.

You may be happy but you are thinking that if it is more than 7 days that is better. So, in this case it is mu only but if I go by this two tailed distribution. Your hypothesis, your alternate hypothesis and you cannot, you may reject H_0 but it may not go in your favor. So, as a result what it is said that it is always greater. If you do like this first you understand what is the variable.

(Refer Slide Time: 21:26)



If suppose x is your random variable, what type of variable is it, lower the better type or is it higher the better or it is nominal is the best, getting me? For example, suppose the sulphur content in steel percentage, sulphur in steel. We do not want that sulphur content, more sulphur content, and zero sulphur. So, it is lower the better in this case what you want, suppose this is the sulphur content.

So, ultimately you will be looking for the lower side. So, that means you will create a value that is the maximum value, what you want your acceptance region. Your zone of working will be region will be this, you do not want higher one. Now, if you plot what will happen, you may find out distribution like this let it be like this. So, you will go for the left hand side, for the higher case it will be just reverse.

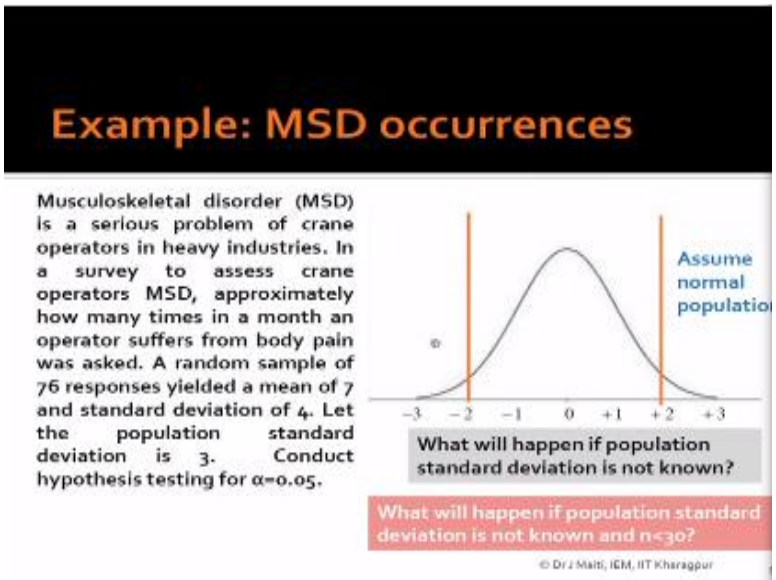
You will fix a value here, somewhere here and you want this side. Now, you can fix here or somewhere here but ultimately you see lower the better and higher the better case, mean the one tailed test is better only when nominal is the best mean. There is a target and if you deviate from the target, if you deviate from this target this side or this side, this is not desirable. Then your two tailed hypothesis is better. So, for both the cases this is one tailed and for this two tailed, getting me?

When you frame the hypothesis, null hypothesis, there is absolutely no problem because from what you are doing you are basically telling some value for that hypothesis statement. You are making it is something like this; usually we say it is something like this. In statistics hypothetical testing but what will be your alternate hypothesis. Now, if you use one tailed hypothesis what will happen, then with for the critical value in two tailed case you have taken t r. Suppose, this is the two tailed case, if I say this is my left side, this is right side you have taken z α/2 or you have taken t n -1 α/2.

This side or this side t r - z α/2t - t n - 1 α 2 but when it is one tailed, suppose so long it is within this you are accepting H_0. When it is going this side you are rejecting H_0, then it is one tailed. In that case this will be zα, if z distribution is applicable or it will be t n - 1 α. In the other side also if you think that no this is my this one is this only, this also α, getting me? So, then this is minus n - 1 α and here it is this value is α.

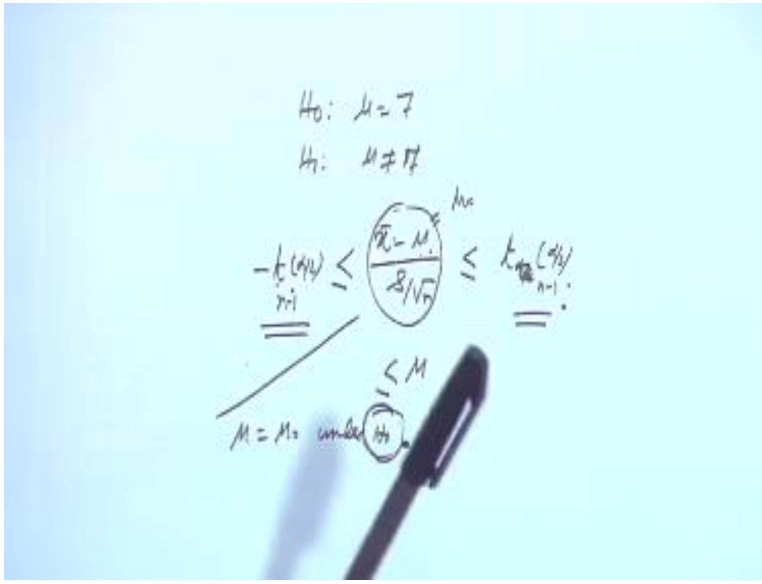
Now there is a relationship between this hypothesis and confidence interval, see if this is α then this side is 100 x1 minus α that percent. Similar here what is happening? This is 100 x 1 –the interval that is why α/2, α/2 you are making, okay? So, this is what our hypothesis testing, for single population mean.

(Refer Slide Time: 26:06)



This is one example. Last class we have described, we said that MSD is a problem with industrial workers, particularly the crane operators in heavy industry. Now a random sample of 76 responses yielded a mean of 7 and standard deviation of 4. Population standard deviation is also given. Conduct hypothesis testing for $\alpha=0.05$. That mean what you require to say that you are basically $\mu=7$, that will be your null hypothesis and what will be your alternate hypothesis?

(Refer Slide Time: 26:50)



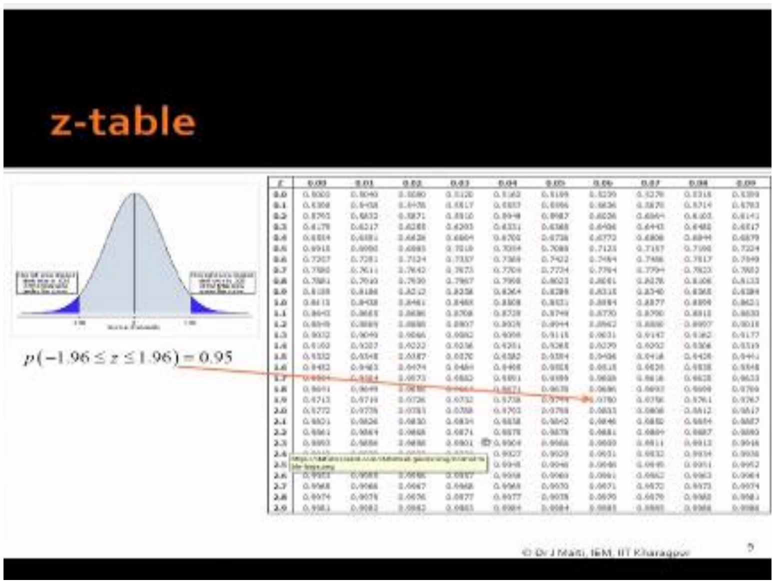
Your null hypothesis is $\mu = 7$. Here let us see the alternate hypothesis μ not equal to 7. Now, here the thing is that as a manufacturer, as a management point of view you want to see that now it is not 7 it is below 7. From operators point of view they may say no, it is not 7 it is more than 7.The operator who are using the crane they will be saying we are suffering more. That the management who are basically maintaining the system they will be saying no, it is less than 4, less than 7.So.

Then from which side you are testing the hypothesis that is also important. Who is the owner of this hypothesis in that sense, okay? I have given here that what will happen if population standard deviation is not known. You know 76 responses, even if population standard deviation is not known but sample standard deviation is known, z distribution will still be applied. What will happen if population standard deviation is not known and in less than 30 t distribution, you will be getting the critical value for competition from t distribution are you getting similarity with confidence interval?

There you have created the same way. You created this first, you find out the statistic then in that case what happened in when you have found out some statistic like $\bar{x}$ - μ /s/$\sqrt{n}$ then you have created like this -t α/2 n - 1 t α/2 here n - 1 that we have created as μ was not known to you. Last class we have seen you have you do not know μ, you created a range for μ. What happened here in this hypothesis testing case mu is given as mu 0 which is under H_0 and we are saying this particular quantity follows t distribution when H_0 is true.

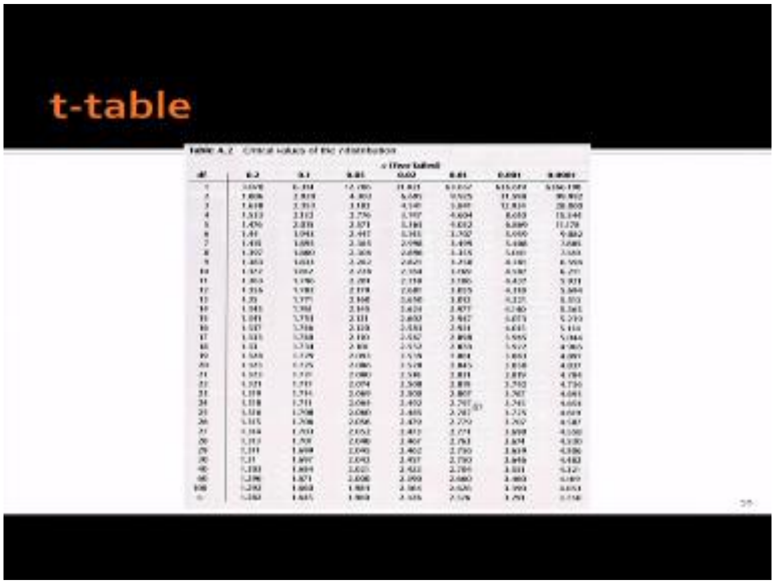
That is why I told you in the beginning that you please keep in mind that whatever statistics you will generate and the statistical sampling distribution. You will consider that is true when H_0 is true, then now instead of mu z, μ_0 this value you know, entire value you know and you also know that the value is critical values. This is left side critical value this is right side critical value. Both values you know whether this value is less than this value or this value is more than this value that you are finding out and accordingly you are rejecting null hypothesis.

(Refer Slide Time: 30:12)



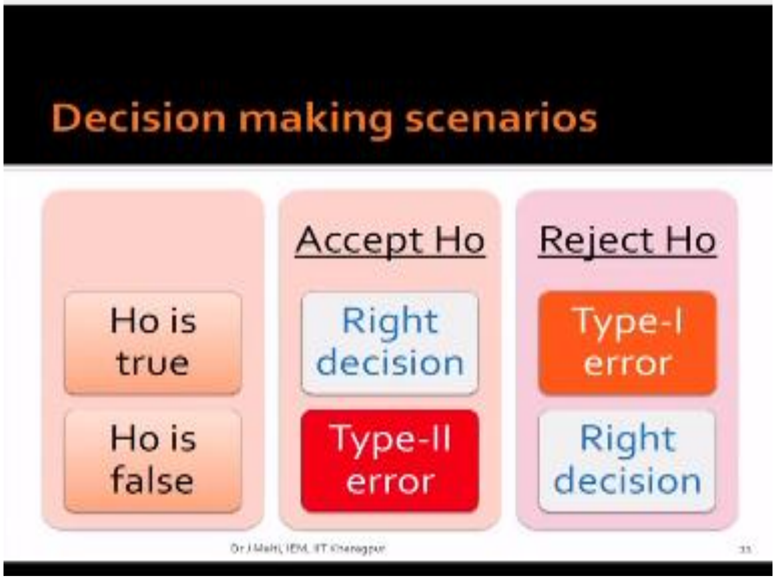
I know how to use z table that is very much known to you.

(Refer Slide Time: 30:18)



T table also.

(Refer Slide Time: 30:20)

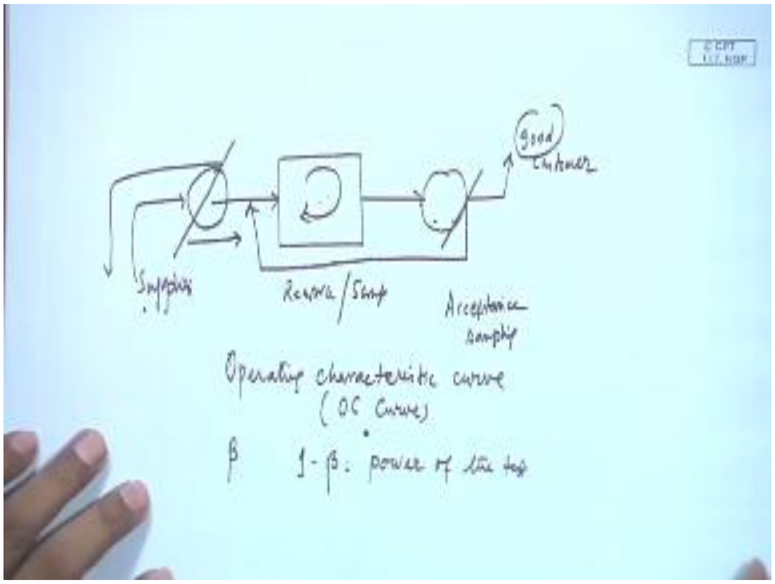


Now when you make hypothesis testing you basically make take certain decisions. Hypothesis testing is basically a decision making. In confidence interval, estimation that times you are not doing making any decision here. In hypothesis testing you are making a decision based on the sample data. Now, there is possibility that there are four scenarios. First scenario is you just think like this that you have considered hypothesis H_0 as something, some statement you made in favor of H_0 that statement may be true, may be false.

Now, if statement is true that mean the hypothesis null hypothesis is true then depending on that based on your analysis of the based you may accept it you may reject it. If your null hypothesis is true and you have accepted it, that it is the right decision again if null hypothesis is false and you have rejected it based on your data analysis that is also right decision. Problem comes when H_0 is true but you have rejected it or H_0 is false you have accepted it.

When H_0 is true you have accepted it. That is known as type 1error or α error and when your H_0 is false and you have accepted as a type 2 error or β error, what is the physical interpretation of this, how do you get the physical meaning of decision making if you see that manufacturer risk versus consumer risk, getting me? There is a production process.

(Refer Slide Time: 32:39)



Producing something you are producing and there is suppose some testing like this and it will be good or it will be bad depending on the test here. Let it be if it is good, it is going to the customer, bad it is going back means re work or scrap or something like this. This is of my production process. So, when you are what do we do? Basically, here we do certain sampling here in quality terminology it is acceptance sampling, acceptance sampling.

Suppose you are doing this one as a, this is manufacturer and you are the customer or other way also you are purchasing raw materials. There is supplier, supplier supplying raw materials supplier and here is a quality control, acceptance sampling. Again if it is bad product, good product going to you but if it is bad it is again coming back to supplier. All these two places it is an acceptance sampling case.

In the first case here manufacturer is producing something and there is a sampling, acceptance sampling scheme for the customer. Based on the data he may accept or reject, that means good or bad is coming based on the data. In the same this case also the supplier is sending something the manufacturer is testing it and according to the test either returning to the supplier or accepting it.

Now, in this case what will happen? Suppose the manufacturer produces the right one but test says it is wrong one.

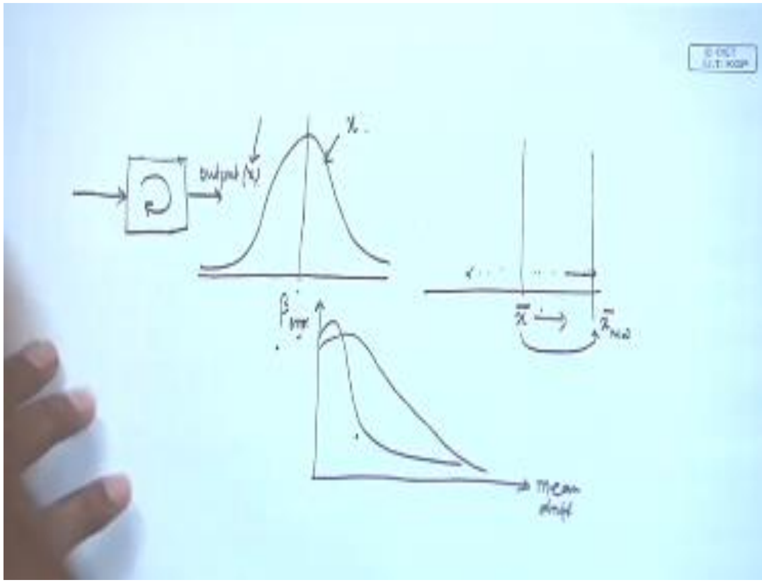
So, you will your error is type1 error because manufacturer produces the right one, manufacturer produces the right one and you have not accepted it type 1 error or α error, fine? This is good because the produced item is gone back to the manufacturer. Now, suppose what will happen if manufacturer produces the bad one and your scheme accept it. Then you will get a bad product, the second one is more problematic because the bad product goes to the market. It is problematic for the customer, problematic more problematic for the manufacturer also.

Because next time what will happen if someone else wants to purchase this as a customer. You will say do not purchase this, it has lot many problems. So, as a result what happen here actually? Customer will not do the acceptance sampling at the market level but the manufacturers himself do certain sampling here. Once the production is completed, how many to be sent to the market based on certain test, getting me?

So, hypothesis testing is nothing but a decision making and many times these decisions are very crucial decisions and you must know what is the α error and what is the β error. Another issue you will go through if you find time, that is operating characteristic curve, popularly known as OC curve. We say the type 2 error is β error and $1 - \beta$ is the power of the test, β is the power of the test.

This operating characteristic curve can be understood. Suppose, you see I am giving one example here. That example is.

(Refer Slide Time: 37:14)

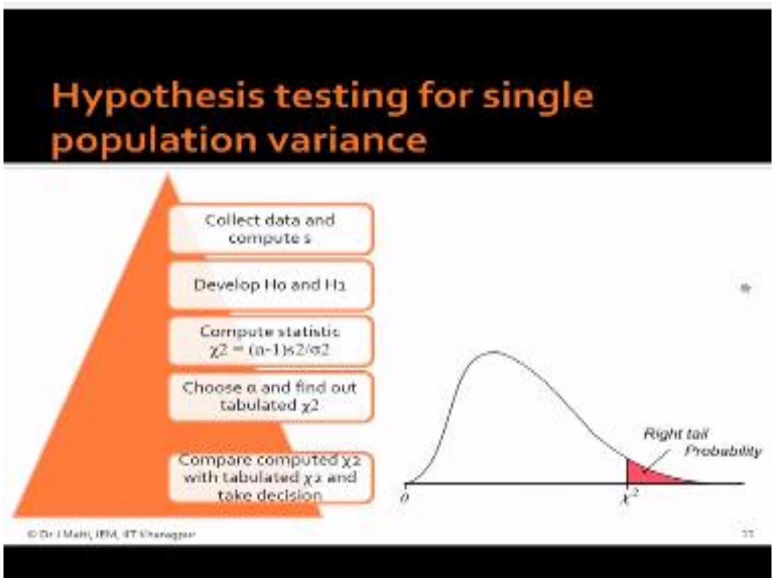


That you think of a quality characteristic and that characteristics suppose follows this normal distribution. This is my quality characteristic let it be some variable x. Now, it is mean value will be here but over this is nothing but the process which is producing something, some output which is measured through x. What happened this overtime because of maybe v r and t r or because of what can I say that not perfect maintenance.

These ultimately the process bin that means the x value, the mean of x, it will basically sipped. Suppose, initially this was my mean value but because v r and t r of maintenance, this mean value may shift in this direction or in this direction depending on the quality. What is happening in the process parameters level? So, in that case suppose I want to take if there is but it is very difficult to change that. This mini shift has taken place immediately, you will not find out the change, when the change has taken substantially then only.

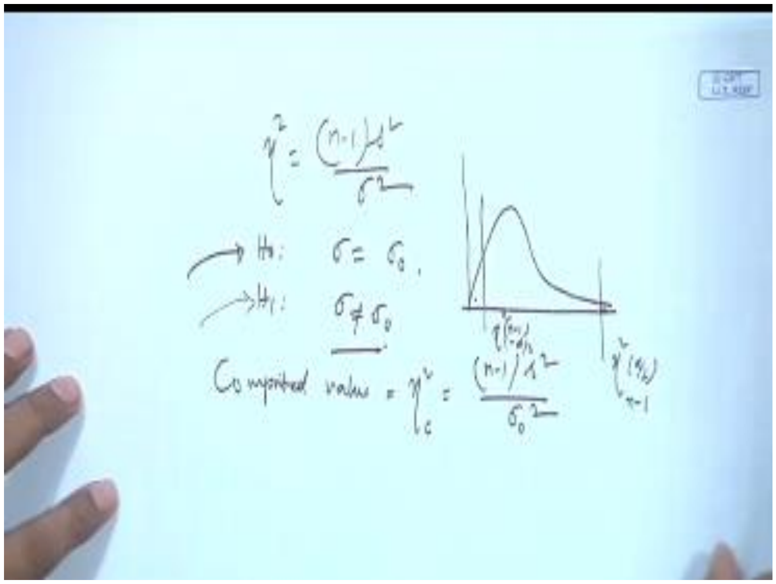
Suppose, if the mean has gone onto this level, this is the mu 1, new mean then you find out that much change has taken place. So, that type also if I write in this side mean change or mean drift and this side suppose the beta error β value what will happen, you get a certain curve when the β is what type 2 error. There is change but you are not able to detect so long the change is small, you are not able to detect, that means you will accept. So, your things may be like this it may be like this, it all depends on which type of it will come under the operating characteristic. It is very important because when you are able to change the find out the change has taken place. That is very critical for if not only manufacturing or other system also it can very critical.

(Refer Slide Time: 39:56)



Now, we will go for the population variance and I am sure that you will appreciate one thing here that you know how to go about hypothesis testing. For population variance also what is the required, what you are required to know, you are required to know the statistics. What will be the statistic in the case of population variance?

(Refer Slide Time: 40:22)

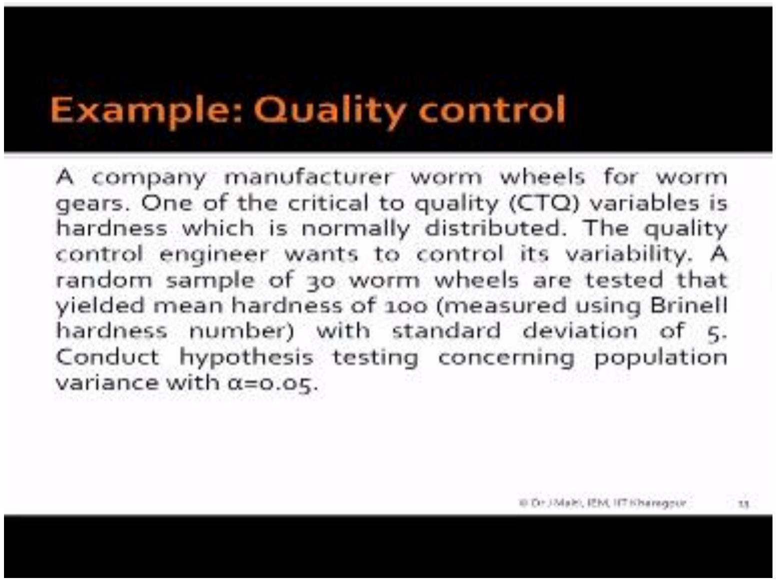


We say the Chi^2 is n -1 s^2/σ^2. What is your null hypothesis σ is 0, what is your alternate hypothesis σ not equal to σ_0. So, then your computed value will be Chi^2 computed which is n -1 s^2/σ^2. So, as you have taken two tailed condition here also, what you will do? You will just find out here it is Chi^2 $\alpha/2$ and here also $Chi^2 1 - \alpha / 2$ and all of you know that Chi^2 distribution one parameter of Chi^2 distribution is the degree of freedom.

So, here you are required to write degree of freedom n - 1 here also, it is n - 1. Then if you find out that if your, this value will fall here or here you will reject null hypothesis. So, that means you have to compute and find out where it is falling and as you know the variance is also a proper such a what can I say parameter. We do not want more variance, basically we want less variance.

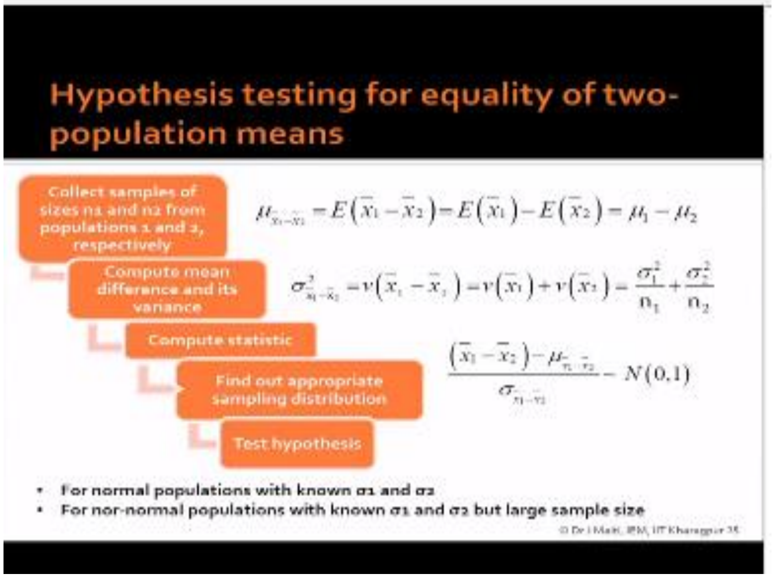
In that sense if you are interested then maybe you can go for one tailed also but here what is the issue is that one null hypothesis is given, you are trying to prove that null hypothesis is true or false. So, in that case alternate hypothesis if you have created it like this, there is no problem but I have given you some example. Based on that example if you go, that mean from the problem to hypothesis then I think got the better. Higher is the better nominal is the best, that particular concept you must use and again if I go for the next one that also that one also easy for you.

(Refer Slide Time: 42:50)



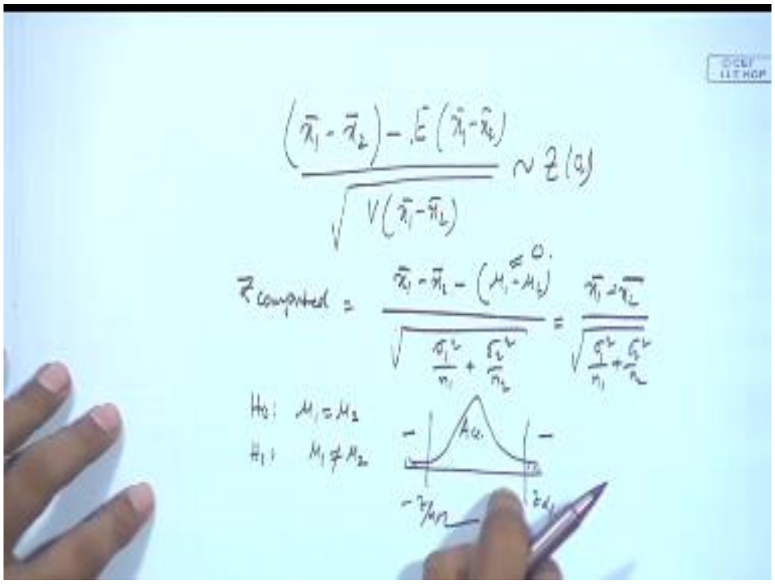
This is one example I have given and anything, no problem at all.

(Refer Slide Time: 42:56)



What is our next topic? Next topic is you want to test whether two populations are equal or not equality. Two populations mean what you will do here, what is the random variable here?

(Refer Slide Time: 43:10)

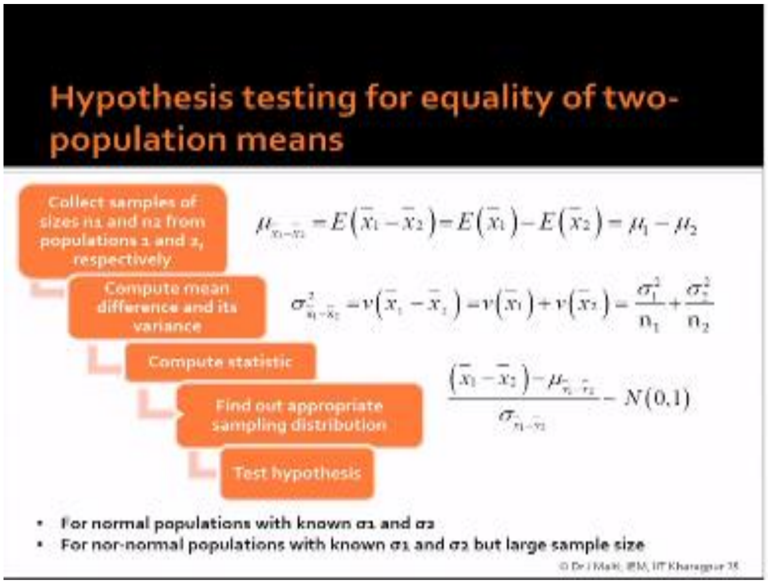


$\bar{x}_{1/}$-$\bar{x}_2$. So, this $\bar{x}_1$ - $\bar{x}_2$ that will follow normal or z that is t distribution depending on the conditions and all of you have seen that $\bar{x}_1$ - $\bar{x}_2$ - expected value of $\bar{x}_1$ - $\bar{x}_2$ divided by that variance of $\bar{x}_1$ - $\bar{x}_2$. This follows z, we are assuming that sigma and other things are known. You know z or t that is that again I no need of discussing further. So, 0, 1 then what will be your here that z computed. If I use z value then z computed is $\bar{x}_1$ - $\bar{x}_2$, - this is μ_1, - μ_2 by this one.

We have seen last class n_1 + $\sigma 2^2$ / n_2, that we have seen what is your null hypothesis here H_0 μ_1 = μ_2, that I mean this quantity becomes 0. So, your z computed becomes $\bar{x}_{1/}$ - $\bar{x}_2/\sqrt{\sigma i^2/n_1 \sigma 2^2/n_2}$. You know n_1, n_2 and σ1, σ2, compute this and then again find out. You set your alternate hypothesis μ_1 not equal to μ_2 that is two tailed, two tailed case. So, z α / 2 - z α /2, find out where it is falling. Is it falling in this region or it is in between this two?

This is acceptance this is rejection for whom for null hypothesis, correct question here, anything clear? Basically, the crux of the matrix knowing the appropriate statistics and its distribution why that is required because then only you will be using the table otherwise you cannot frame anything. It will not be a parametric one it will be different kind for parametric case. It is the must you must know the statistic appropriate statistics and its distribution then things are very simple.

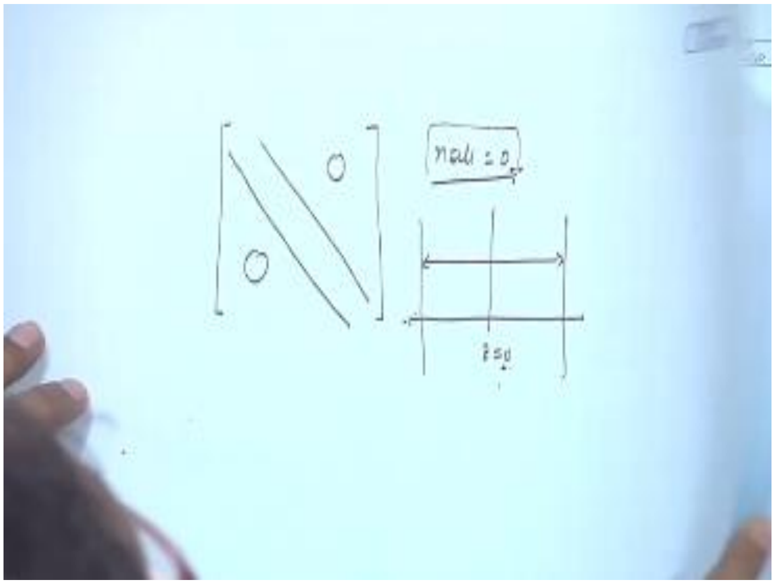
(Refer Slide Time: 46:05)



Sir, why we call it null hypothesis because our target is good. By contradiction is it like we always try to negate it? This is what I can say, this way it is developed one to negate it. Definitely you want to negate it null hypothesis, the word null I have no idea about the word null, mean why they are writing null but I can say that as you are saying that alternative hypothesis.

There is a purpose of alternative hypothesis that you are to reject the null hypothesis, you will claim in such a manner that you will do in such a manner that possible to test null hypothesis as such why the word null is coming difficult for me. So, I may serious one otherwise this is a general way, this is a very good question. Basically, I think yes why the word null is used. I can give you one explanation as sometime what happened I have seen in structural equation modeling.

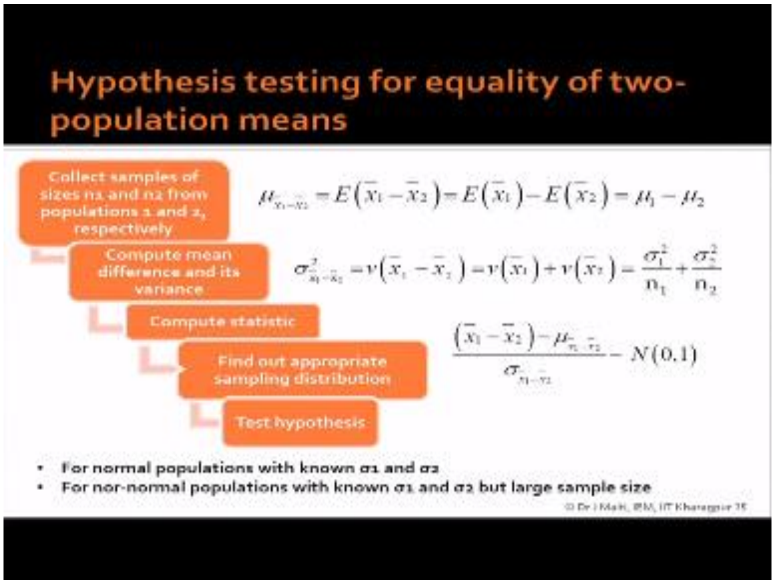
(Refer Slide Time: 47:38)



There are they have taken co variance matrix and then only the variance component they consider and other component they put it to 0. So, I may not be 100 percent sure that null basically is devoid or 0, getting me? So, if this is the case, so then when we test you will find out you create? Suppose, you are making z test then that z test, that z value is 0. If z value is 0 so long your statistic is close to 0, you are accepting this, you are when you are rejecting you are saying if that my computed value is sufficiently away from 0, getting me?

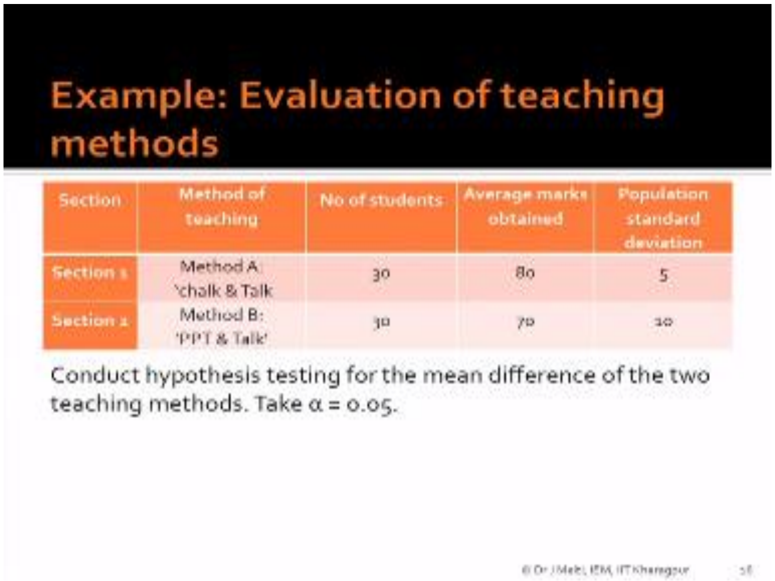
So, from that analogy I can tell you because of this 0, you are getting me, because of this 0 what I am doing? Basically I am computing the z value. The statistics value I know this will follow certain z, that z distribution where the mean value is 0 because you are testing for mean. So, it is sufficiently away from 0 so long it is close to 0. That means so long within this region say it is in the null region, if I say null equal to 0, null means void. Void means 0, that sense I think this maybe but one of the explanation but not 100 percent sure that whether this I think this can be thought of.

(Refer Slide Time: 49:30)



This is the example.

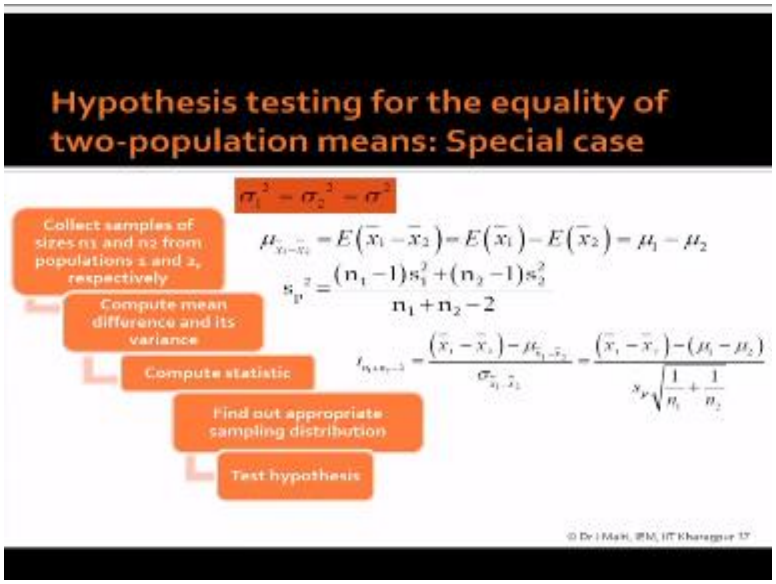
(Refer Slide Time: 49:33)



As I told you in last class that the method of teaching is tested and two population standard deviations are given for section one, that method A it is5 and method B it is 10. In your confidence interval you have taken these values but in when you have s pooled, that time what happened, that time you have assumed that these two values are not different. We are here, we are interested to test whether this two values are really different or not and that is the difference between two teaching methods.

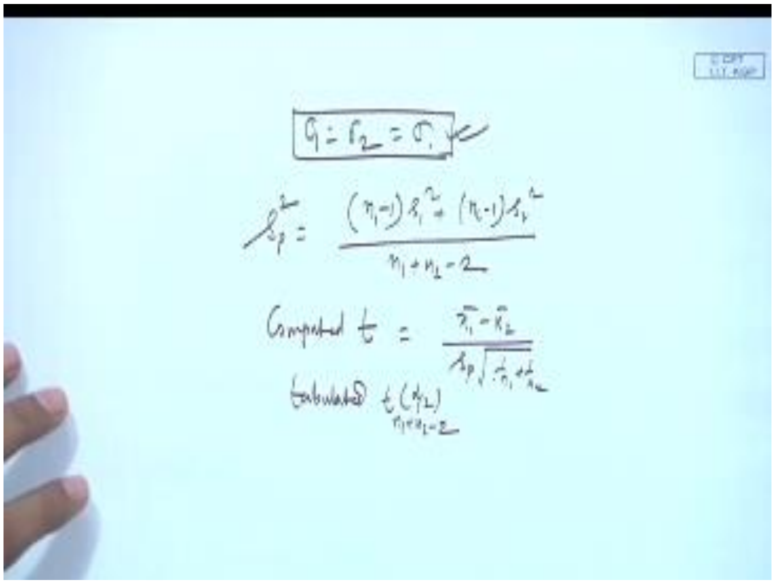
So here we are basically doing the difference between 80 and 70, that is mean difference we are taking and considering population that5 and 10, not that what I said that chi^2 distribution here.

(Refer Slide Time: 50:44)



Now, you come to the special case, what is the special case?

(Refer Slide Time: 50:57)

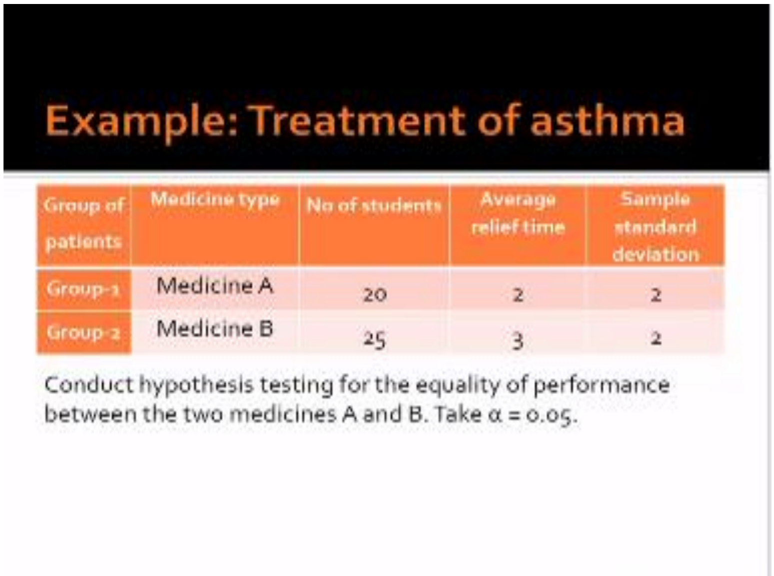


$\sigma_1 = \sigma_2 = \sigma$. This time you will use s p^2, that is full variance, rest of the things remain the same and you statistic that basically computed value will be under H_0 $\mu_1 = \mu_2$.So, if I use t computed, t will be $\bar{x}_1 - \bar{x}_2$ -0 because $\bar{x}_1 - \bar{x}_2$ / sp^2 $\sqrt{n_1}$-/+n_2 and go for tabulated t tabulated t n_1 +s n_2 - 2 $\alpha/2$. If my tabulated t is more than or less than depending on which side that it is that the tabulated t, you will reject the null hypothesis.

This is just a special case of the original? Original one. Student: Original one? Original one was the $\sigma 1^2/ n_1 + \sigma 1^2/ n2$.That special case would hold for the original one also using others cases. Okay. What is the requirement of the computation of the several type of convolution? Correct, correct. Now, because under this special case scenario what happened actually all this that whatever the distribution you are finding out this distributions are governed by the number of parameters. As the size of the sample and all those things are involved here.

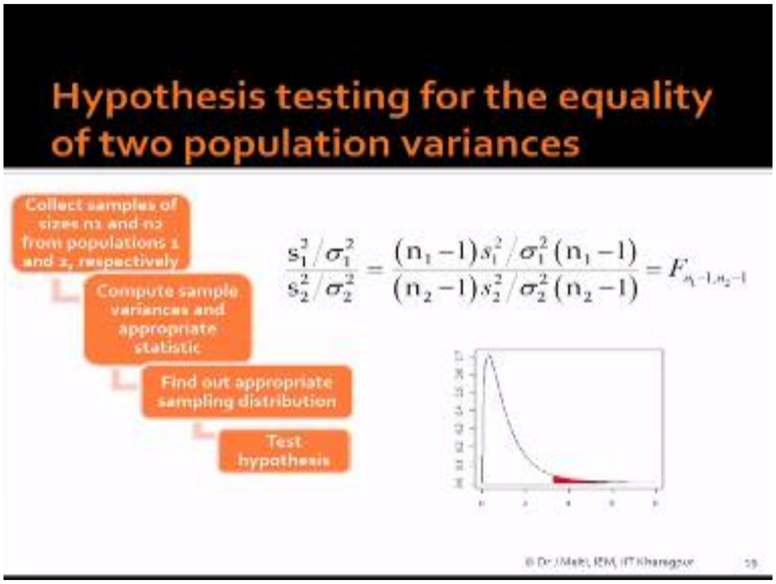
Now, what happen when this special condition arise that $\sigma 1 = \sigma$, then if we convert this s p, if we find sp^2 and accordingly we will go that power of the test is much better, you are getting me? Always there is a possibility of error, there is α error or β error is there. When the power of the test will be better that is why these special cases are found out and they are also reported and most of the times you use this. That is why what we said if we find out that the two population variances are equal, that you first find out then you go for the special case. Do not use the general case. General case is many a times a general case because general case will give you some result but if this special case will occur. That this is better case? Always better.

(Refer Slide Time: 54:01)



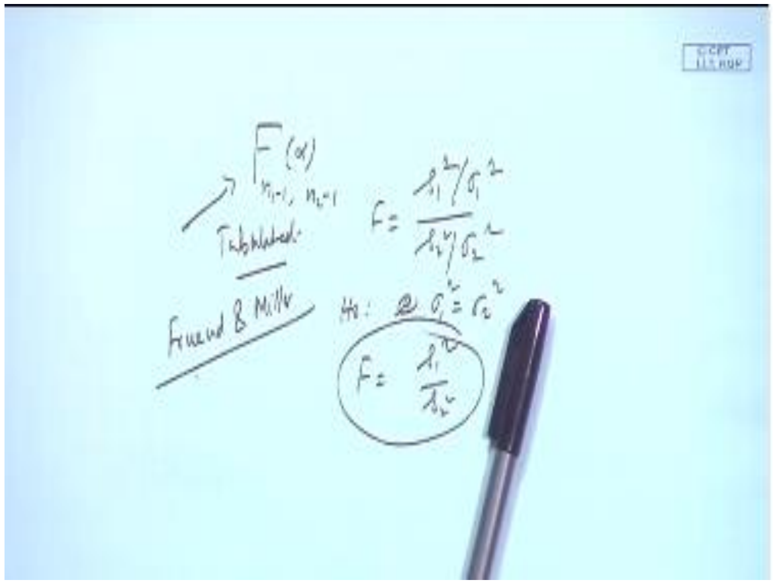
This is another example that we want to equality of performance between two medicines. Confidence interval case you have seen.

(Refer Slide Time: 54:09)



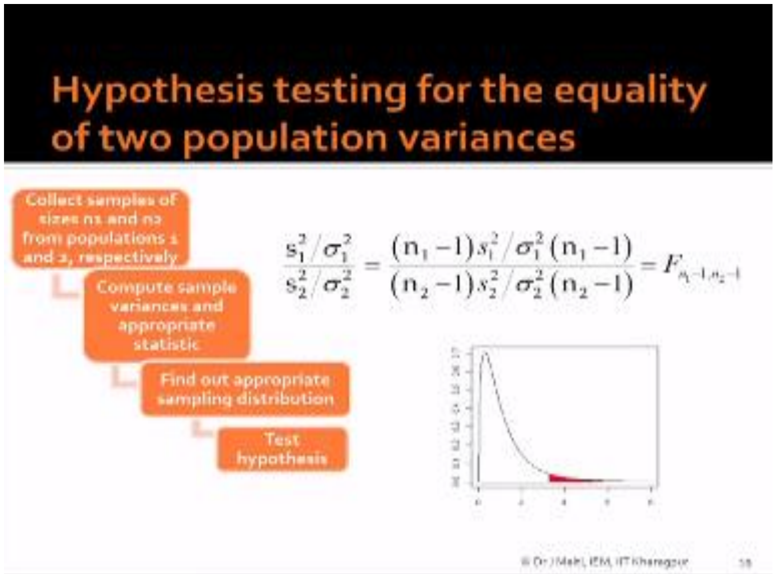
Now, equality of two population variances should I go for this? I think you will all be able to find out this when you say that the equality of two population variance is you must know the distribution of the ratio of two population variances. Last class we have proven it that it will be basically F distribution.

(Refer Slide Time: 54:38)



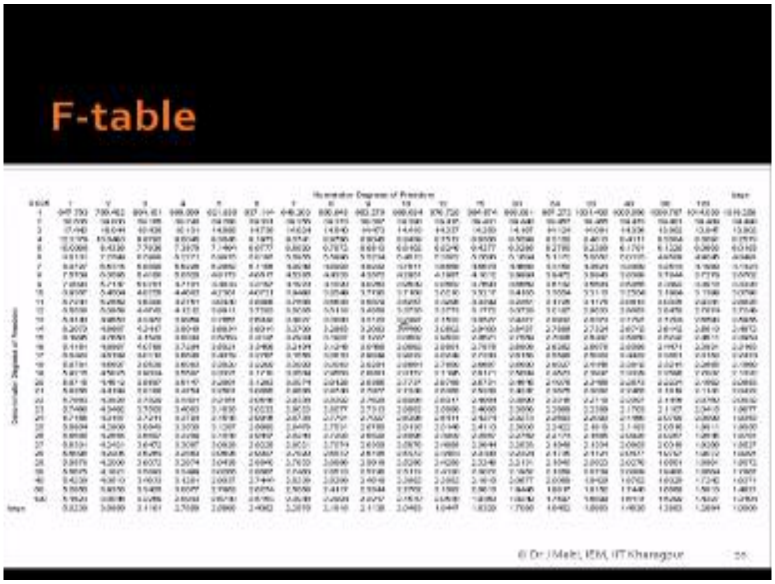
So, your F n_1 -s 1 and n_2 - 1 and definitely you will be getting certain α value and that will be your tabulated one and what was the statistics. That time statistics we considered F =to s_1^2 / $\sigma1^{2/}s2^2/ \sigma2^2$. What is our hypothesis? Here H_0 that $\sigma_1^2 = \sigma_2^2$. So, $(\sigma_1 / \sigma_2)^2$ is basically this is 1. So, your F is basically result ant F, basically s_1^2 / s_2^2. So, you want to test with the computed tabulated value. If you find that the tabulated value is less than the computed value, reject the null hypothesis.

(Refer Slide Time: 55:45)



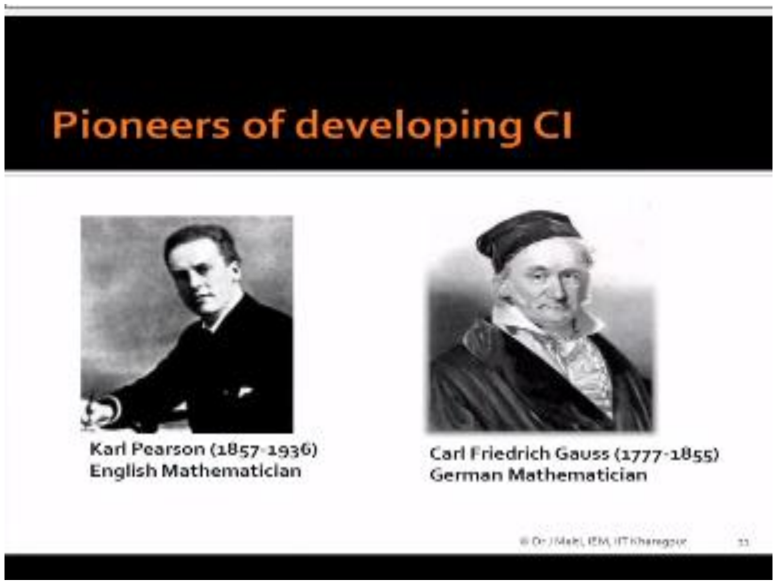
Here some special cases also occur basically depending on the σ_1,σ_2 that which one is more, which one is less. You go through by this Freud and Miller book. Freud and Miller basic statistics that book basically engineering I think this is basically introduction to statistics. That book you go through, you will be finding out many more cases are there.

(Refer Slide Time: 56:14)



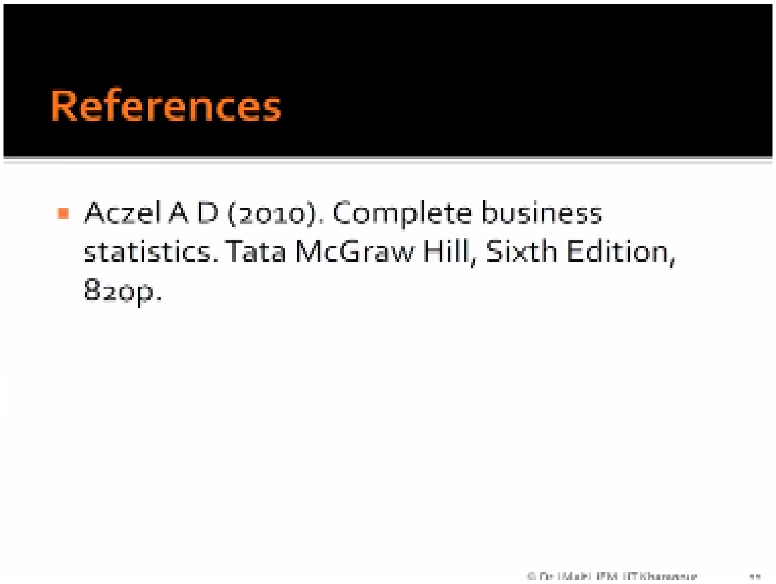
This is F table.

(Refer Slide Time: 56:18)



I told you last class that when the confidence interval as well as this.

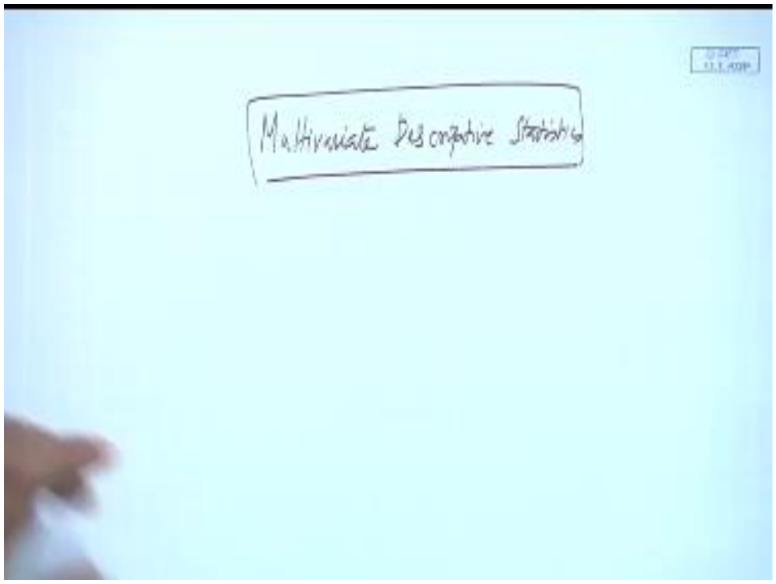
(Refer Slide Time: 56:26)



Hypothesis testing these are the some pioneers hypothesis. Fischer has tested so many hypothecs is in each lot many experiments. What he has conducted and for hypothesis testing this book is good book. This is Aczel A D, complete statistics. I have seen this book is a very good book, you can go through book, it is a compact one in addition as I told you that Freud and Miller, that statistics book that also you go through, you will find out that fantastic.

So, many good books are there and particularly for basic statistics is concerned, universal statistics is concerned, multivariate statistics books. Yes, good books are there and some more books are there but compared to uni variate case, multivariate books are very limited. So, next class what we will do? We will basically talk about.

(Refer Slide Time: 57:44)



Multivariate descriptive statistics okay so this is the end of today, this one, this is the end of our basic or prerequisites, remember what I have told you? Under this basic statistic is very miniscule, I have told you very simple just for to get some idea that what is happening, what will be is there but in general one uni variate. That basic statistic book is itself 1000 pages, getting me?

So, if you really want to get your fundamentals very strong, you have to have some good books on basic statistics also and you have to go through and I have given you some of the things what the concept. This concept we will be using in the multivariate and that lectures also that you can easily grasp, means that first I will tell you in uni variate. You got this one see how we are converting the same concept to multivariate. So, those many portions only I have taken into consideration, it is not the totality of uni variate basic statistics, okay. Then thank you.