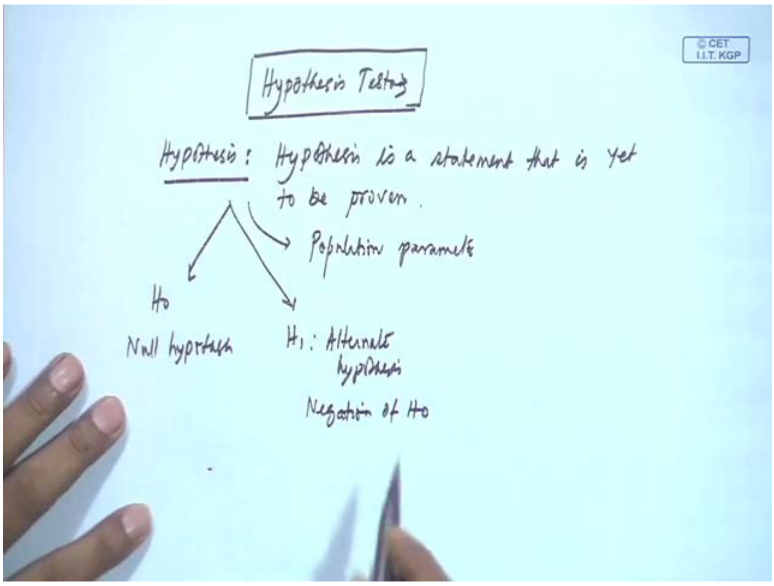
Applied Multivariate Statistical Modeling

Prof. J.Maity
Department of Industrial Instrumentation and Management
Indian Institute of Technology, Kharagpur

Lecture - 7
Hypothesis Testing

Today we will discuss hypothesis testing.

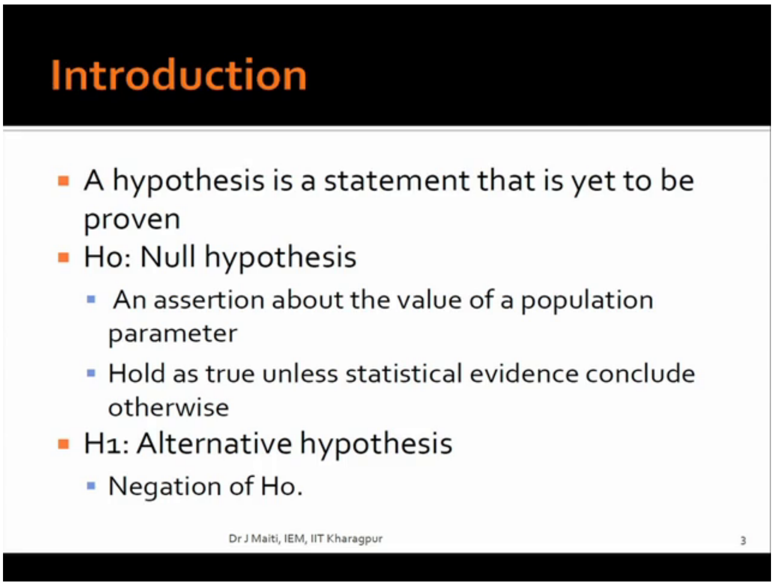
(Refer Slide Time: 00:27)



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.

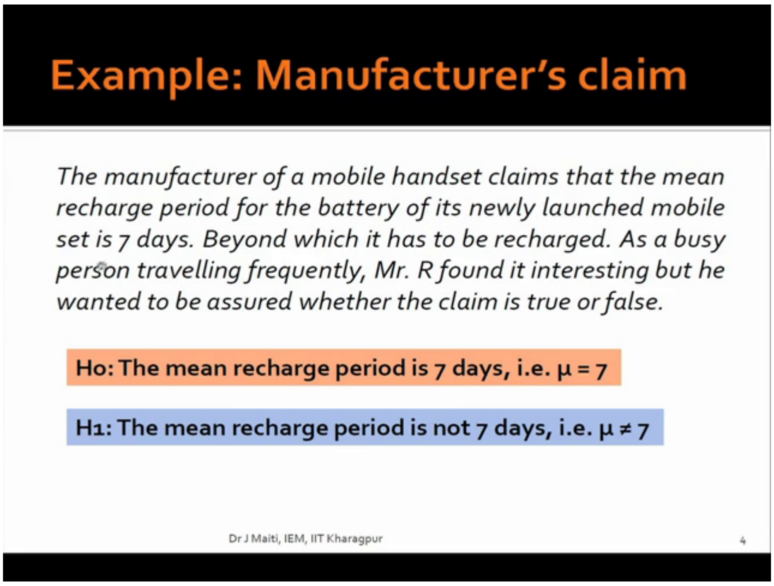
So, if you read the history of science you will find out that n number of hypothesis. There are huge number of hypothesis has been framed by the scientist and proven by experiment or by some other means. So, our hypothesis is limited to the statistical hypothesis testing and before going into detail of this let us see the content of today's lecture. 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: 02:25)



So, come back to this hypothesis testing and I have discuss the hypothesis statement that is yet to be proven and there are we will consider two types of hypothesis. One is H O which is known as null hypothesis and another one is H 1 or H a 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. Alternative hypothesis is the negation of H 0, alternative hypothesis is negation of H 0. Let us see one example here.

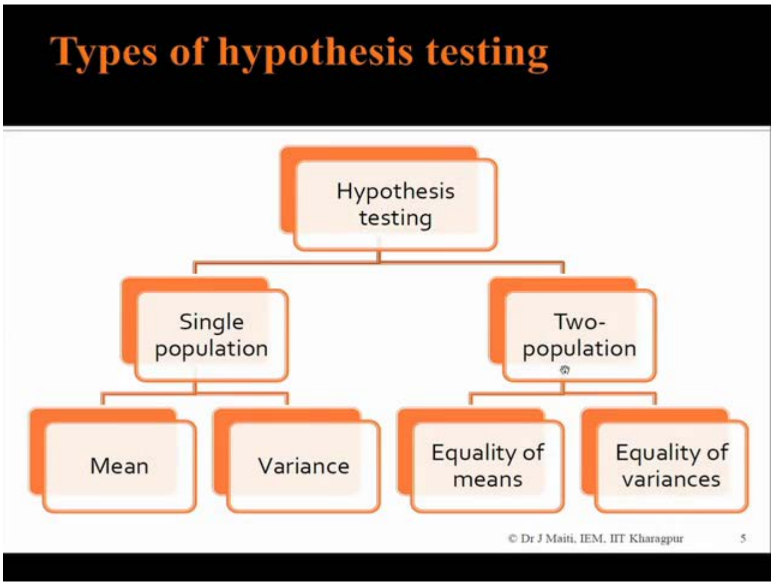
(Refer Slide Time: 03:26)



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.

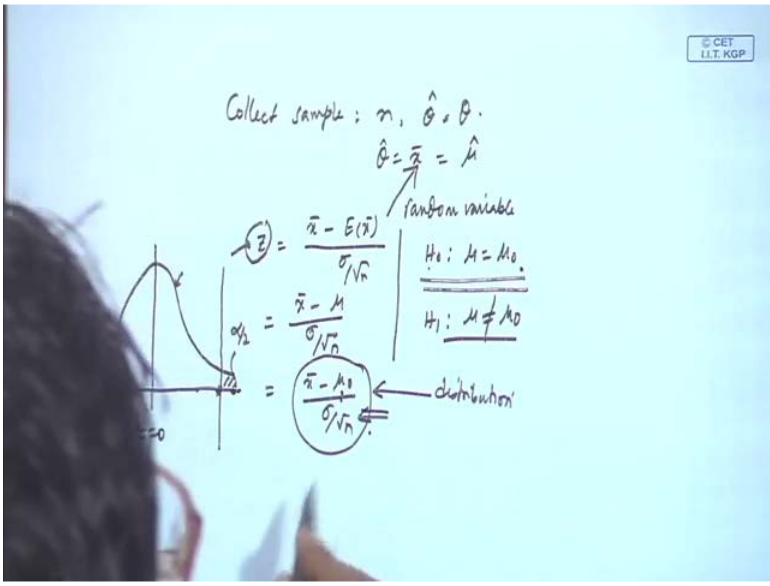
So, here our null hypothesis is the mean recharge period is 7 days which is the population mean. We are basically talking about like this that null hypothesis mu is equal to mu 0. In this case mu 0 is 7 and alternative hypothesis is that mu not equal to mu 0. That means not equal to 7 and if this is 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?

(Refer Slide Time: 04:52)



The steps are identify first null hypothesis and alternative hypothesis, then you definitely find out the appropriate sampling statistic. Then obtain sampling distribution theta. If the statistic is theta, 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 theta, that is the sample statistic. What do you do?

(Refer Slide Time: 05:30)



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 theta instead of cap we have given only theta. For example, this theta will be x bar or theta cap, that is x bar which is basically estimate of mu.

Now, this x bar is a random variable and what you require to know. You require to frame an appropriate statistics. If it is x bar, fine if it is not x bar, something else that statistics you find out. For example, we need to talk about x bar. We usually frame z that is x bar minus expected value of x bar by sigma by root n, that we usually frame here. Our null hypothesis is H 0 mu equal to mu 0. We all know that expected value of x bar is mu, so then we write sigma by root n.

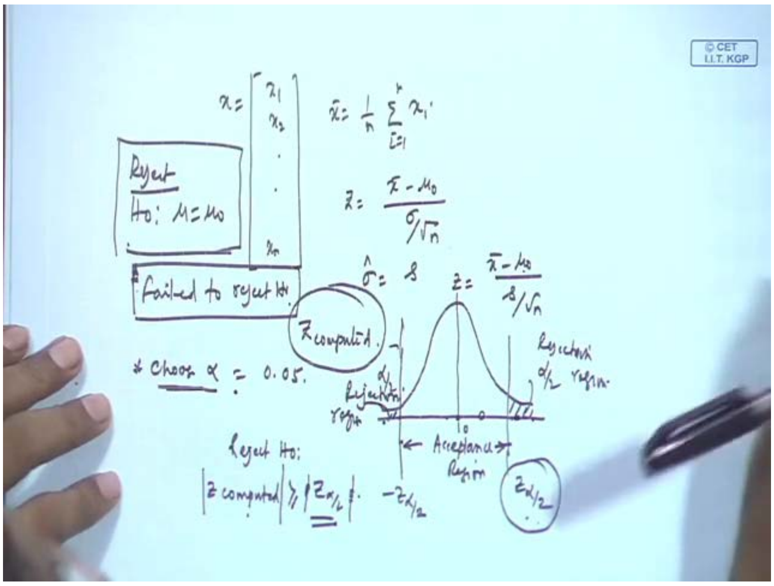
Now, we have assumed that H 0, from H 0 we have we have seen that mu equal to mu 0. So, that means you can write this one, mu 0 by sigma by root n. Then if this is my statistic for which you require to know the distribution, what I said here that is obtain the sampling distribution of theta when H 0 is true. That means the sampling distribution of the here your here your x bar which is nothing but we converting it into z and we are writing mu 0 instead of mu.

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 x bar minus mu 0 by sigma by root 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 mu is not equal to mu 0.

So, that is why we will frame critical value in this side. Either this may be your, this is your alpha by 2 and or this side sufficiently away. This side, this is your alpha by 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 mu not equal to mu 0, then computational what you will do first? You collect data that what you have done in this particular case. What we will proceed?

(Refer Slide Time: 09:28)



You will proceed like this, you collect data for x which will be x 1, x 2 like x n. Then you will compute x bar which is 1 by n i equal to 1 to n x i. Then you will compute the statistic which we are talking about z, that x bar minus mu 0 by sigma by root n.

Now, please remember that root sigma will be given. If sigma is not given then sigma cap will be s and in that case if n is sufficiently large then you will be using z distribution like this, x bar minus mu 0 by sigma by root n. This is the computed z, basically from the data z computed. Now, you have to choose alpha.

So, your next step is choose alpha, you have seen the what is the error. You are going to consume if you say r alpha is 0.05, this is also known as probability level of significance. If we say alpha equal to 0.05 then you are this your point will be like this, your rejection region will be like this. Now, suppose this one is alpha by 2 and this side, this is also alpha by 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 alpha by 2 right hand side and minus z alpha by 2 left side. When you goes 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 alpha by 2, either it is the mod value we have to take. If z computed, if z computed is greater than mod of z alpha by 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.

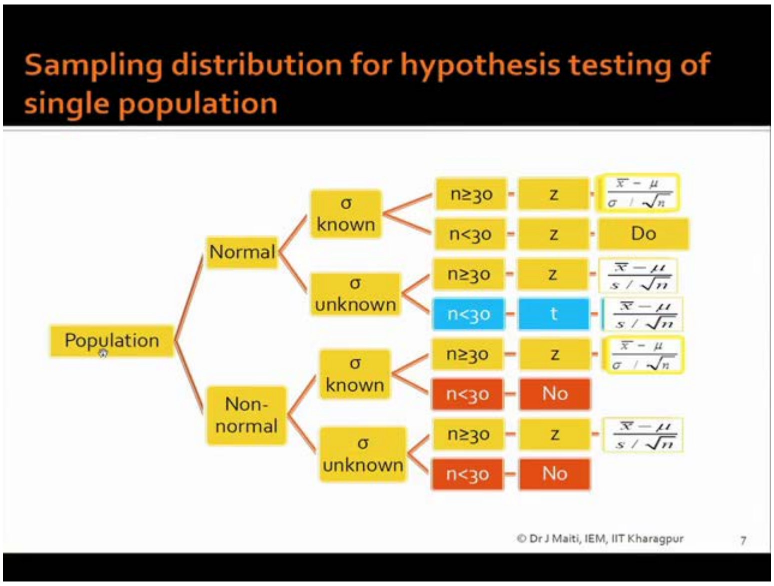
So, if this one is my zero value. Now, if z value is here this side it will be positive and this side it will be negative and you will be seeing the alpha value from that z alpha by 2 value from the table. Usually this one will be a positive value. So, you may not take mod here, no problem.

So, the absolute value if it is more than z alpha by 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 mu 0 that can be rejected. If I reject H 0 that means I am accepting H 0. 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 suppose 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?

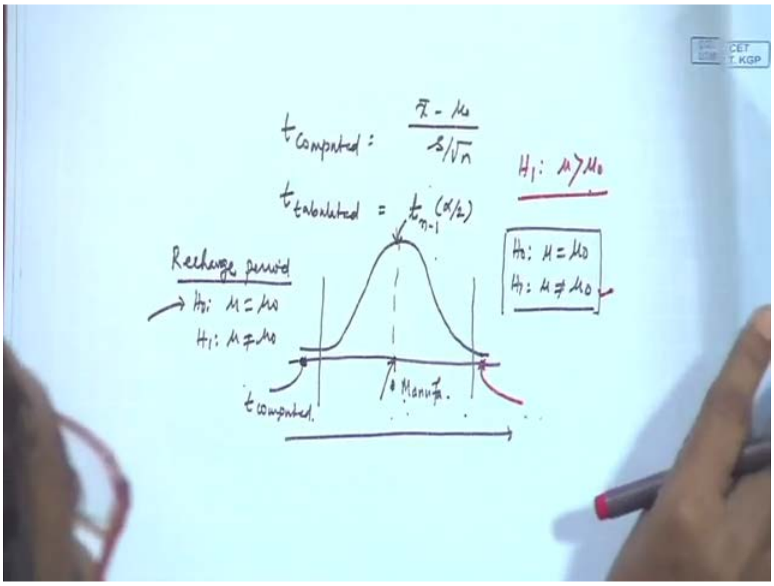
(Refer Slide Time: 15:28)



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 sigma known or sigma 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 x bar minus mu by sigma by root n, that is for the first case, like second case also same, third case your sigma is replaced by s. Fourth case also sigma is replaced by s but sample size small. So, that is why it is t distribution, okay?

(Refer Slide Time: 16:34)



So, if you use the t distribution what is the same thing, you will calculate the statistic that is t computed will be your x bar minus mu 0 by s by root 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 minus 1 and alpha by 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 equal to mu 0 and H 1 mu not equal to mu 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.

Student: 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.

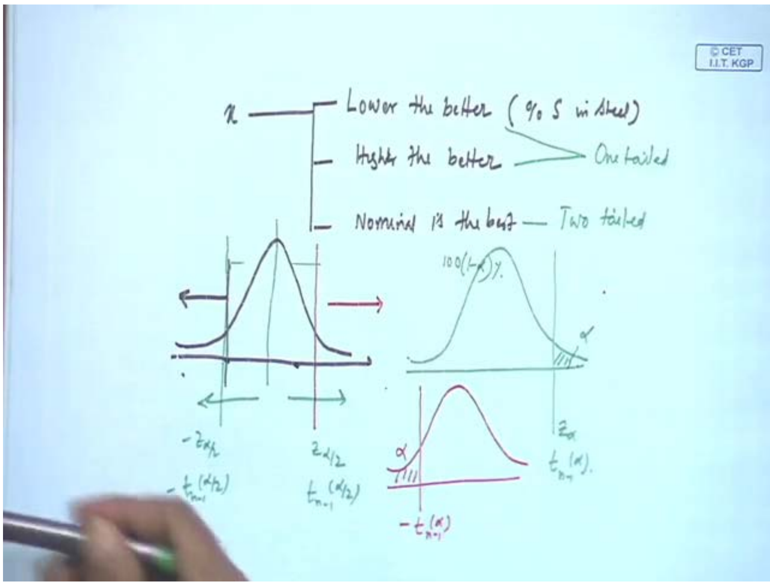
Now, for recharge period if you test like this that H 0 mu equal to mu 0 and H 1 mu not equal to mu 0. Suppose, your that computed value comes here, z value or t value.

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 favour? In your favour if the value lies in other way, that means if your if the computed value comes here that is your favourable 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 greater than mu 0. You may not be interested to test this mu not equal to mu 0, you may be interested to test mu greater than mu 0. Means the manufacturer is claiming it is 7 days, fine 7 days is good.

(Refer Slide Time: 21:24)



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 favour. 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. 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, 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.

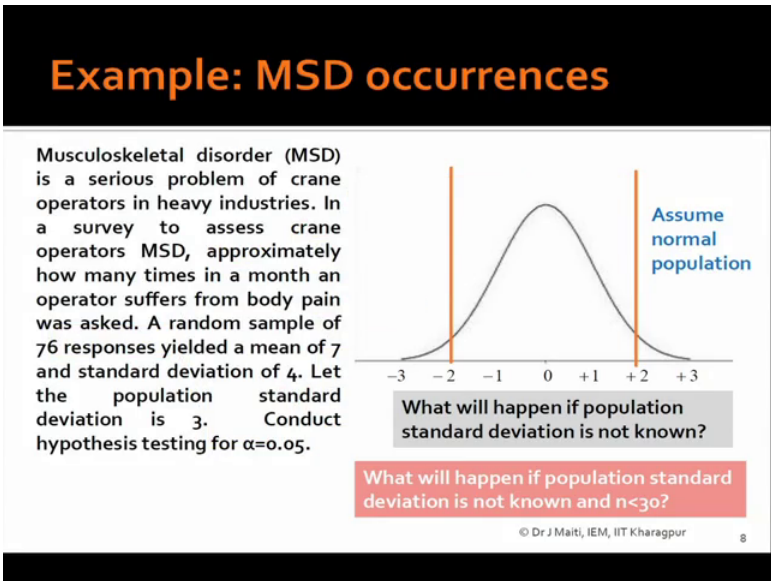
Now, if you plot what will happen, you may find out distribution like this. Let it be like this. So, you will go for the left hand side, for the higher case it will be just reverse. You will fix a value here, somewhere here and you want this side. Now, you can fix here or somewhere here but ultimately 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.

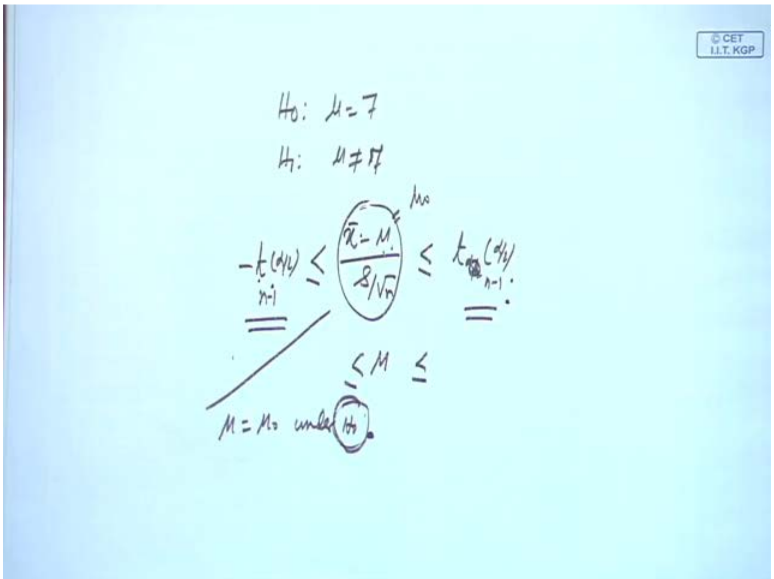
Now, if you use one tailed hypothesis what will happen, then with for the critical value in two tailed case you have taken t r. Suppose, this is the two tailed case, if I say this is my left side, this is right side you have taken z alpha by 2 or you have taken t n minus 1 alpha by 2. This side or this side t r minus z alpha by 2 t minus t n minus 1 alpha by 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 z alpha, if z distribution is applicable or it will be t n minus 1 alpha. In the other side also if you think that no this is my this one is this only, this also alpha, getting me? So, then this is minus n minus 1 alpha and here it is this value is alpha. Now, there is a relationship between this hypothesis and confidence interval, see if this is alpha then this side is 100 into 1 minus alpha that percent. Similar here what is happening? This is 100 into 1 minus the interval, that is why alpha by 2, alpha by 2 you are making, okay? So, this is what our hypothesis testing, for single population mean.

(Refer Slide Time: 26:06)



(Refer Slide Time: 26:49)



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 equal to 0.05. That mean what you require to say that you are basically mu equal to 7, that will be your null hypothesis and what will be your alternate hypothesis?

Your null hypothesis is mu equal to 7. Here let us see the alternate hypothesis mu not equal to 7. Now, here the thing is that as a manufacturer, as a management point of view you want to see that no 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.

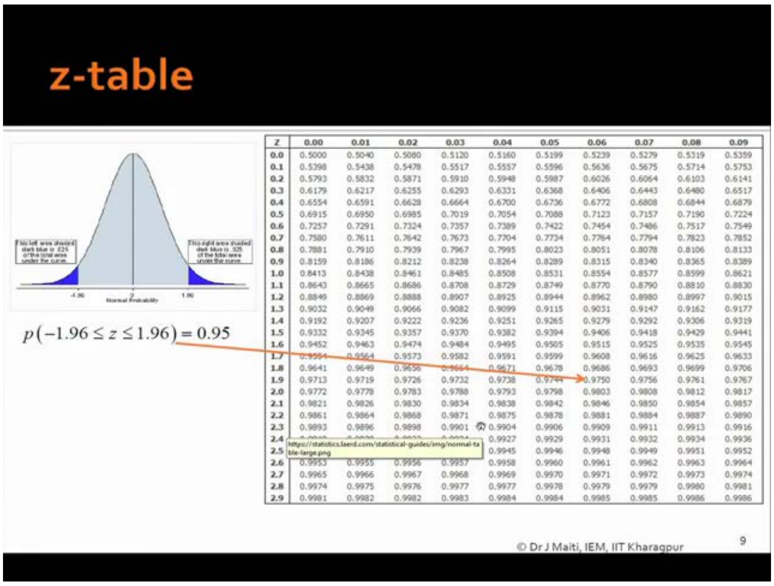
So, then from which side you are testing the hypothesis that is also important. Who is the owner of this hypothesis in that sense, okay? I have given here that what will happen if population standard deviation is not known. You know 76 responses, even if population standard deviation is not known but 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 x bar minus mu by s by root n.

Then you have created like this minus t alpha by 2 n minus 1 t alpha by t alpha by 2 here, n minus 1 that we have created as mu was not known to you. Last class we have seen you have you do not know mu, you created a range for mu. What happened here in this hypothesis testing case mu is given as mu 0 which is under H 0 and we are saying these 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, mu 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.

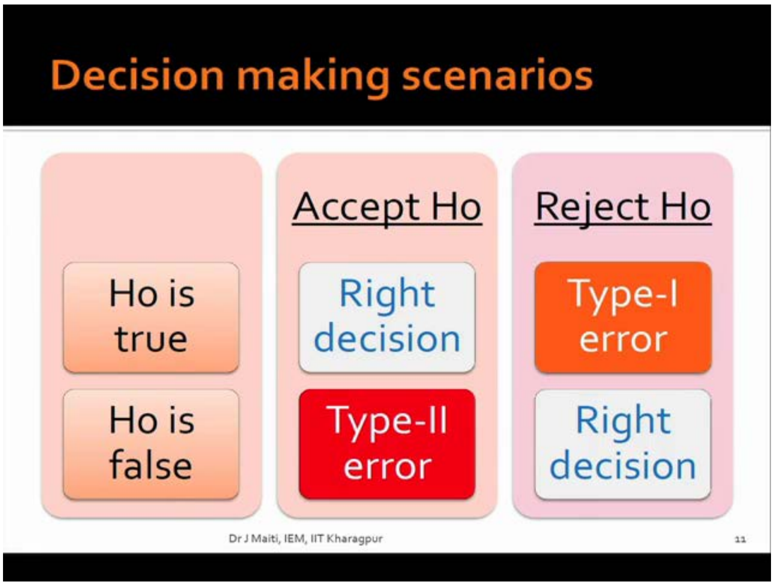
(Refer Slide Time: 30:10)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9031	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	[illegible]	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	[illegible]	[illegible]	[illegible]	[illegible]	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	[illegible]	[illegible]	[illegible]	[illegible]	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

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

(Refer Slide Time: 30:20)

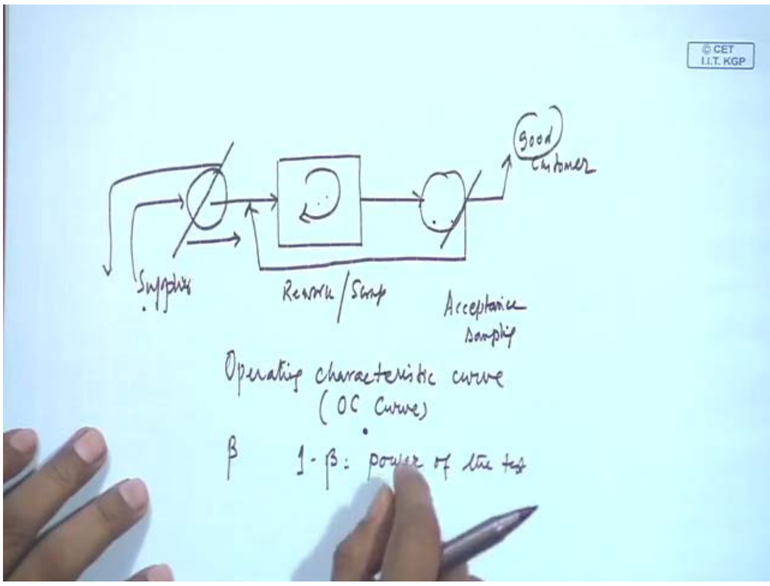


Now, when you make hypothesis testing you basically make take certain decisions. Hypothesis testing is basically a decision making. In confidence interval, estimation that time you are not doing making any decision here. In hypothesis testing you are making a decision based on the sample data. Now, there 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 favour 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 1 error or alpha error and when your H 0 is false and you have accepted as a type 2 error or beta 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?

(Refer Slide Time: 32:36)



There is a production process, producing something. 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 rework 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.

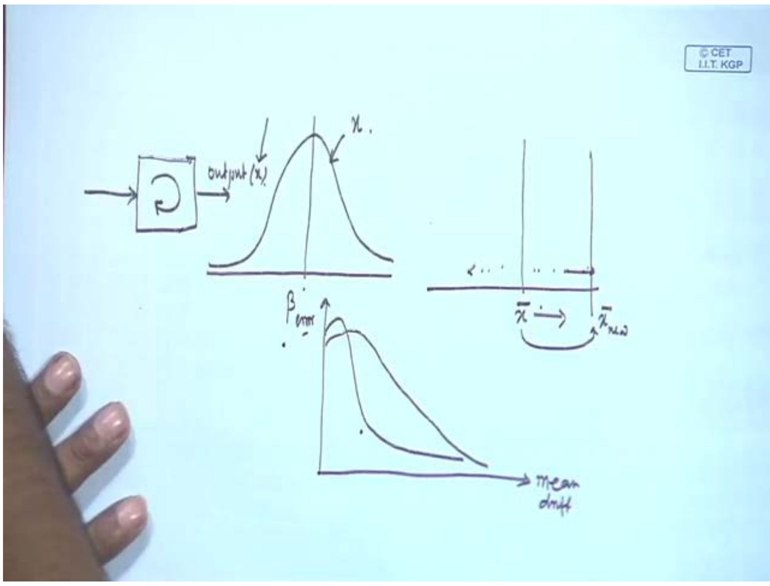
In the same this case also the supplier is sending something the manufacturer is testing it and according to the test either returning to the supplier or accepting it. Now, in this case what will happen? Suppose, the manufacturer produces the right one but test says it is wrong one. So, you will your error is type 1 error because manufacturer produces the right one, manufacturer produces the right one and you have not accepted it. Type 1 error or alpha 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 accepts 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 manufacturer 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 alpha error and what is the beta 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 beta error and 1 minus beta is the power of the test, beta is the power of the test. This operating characteristic curve can be understood. Suppose, you see I am giving one example here.

(Refer Slide Time: 37:12)

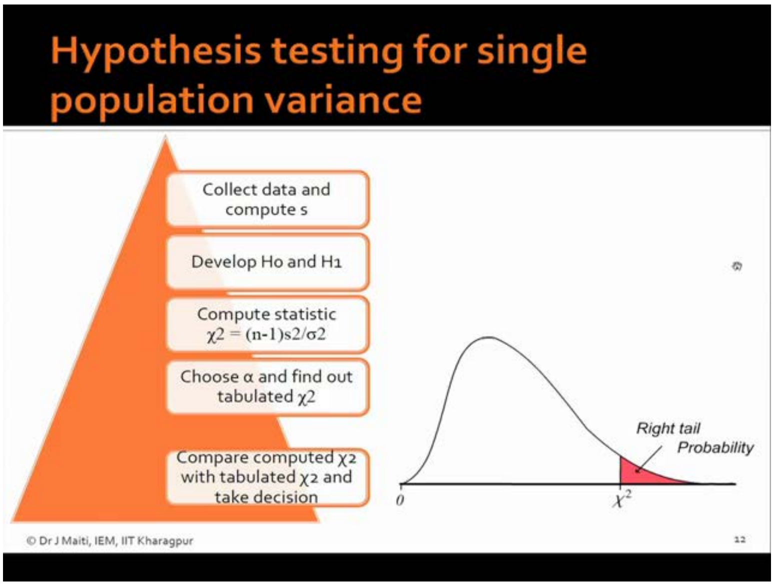


That example is that you think of a quality characteristic and that characteristics suppose follows this normal distribution. This is my quality characteristic, let it be some variable x. Now, its mean value will be here but over this is nothing but the process which is producing something, some output which is measured through x. What happened this overtime because of maybe v r and t r or because of your 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 on to this level, this is the mu 1, new mean then you find out that much change has taken place.

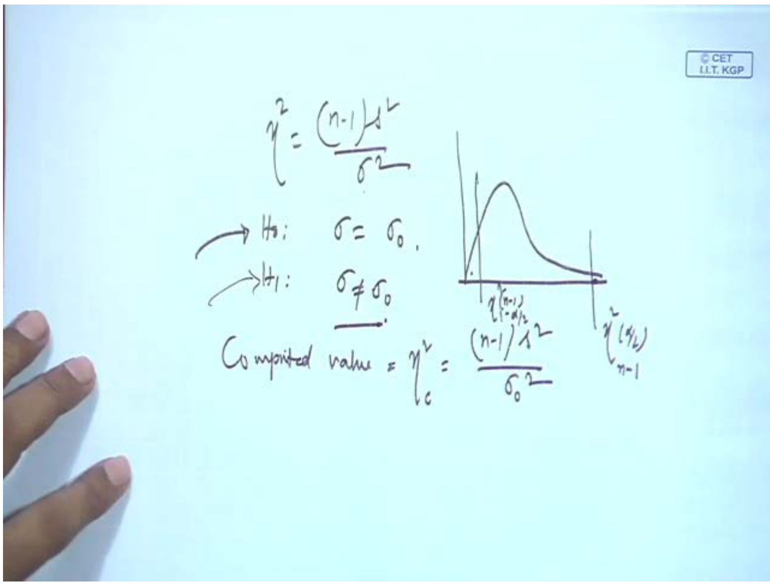
So, that type also if I write in this side mean change or mean drift and this side suppose the beta error beta value what will happen, you get a certain curve when the beta is what type 2 error. There is change but you are not able to detect so long the change is small, you are not able to detect, that 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.

(Refer Slide Time: 39:55)



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)

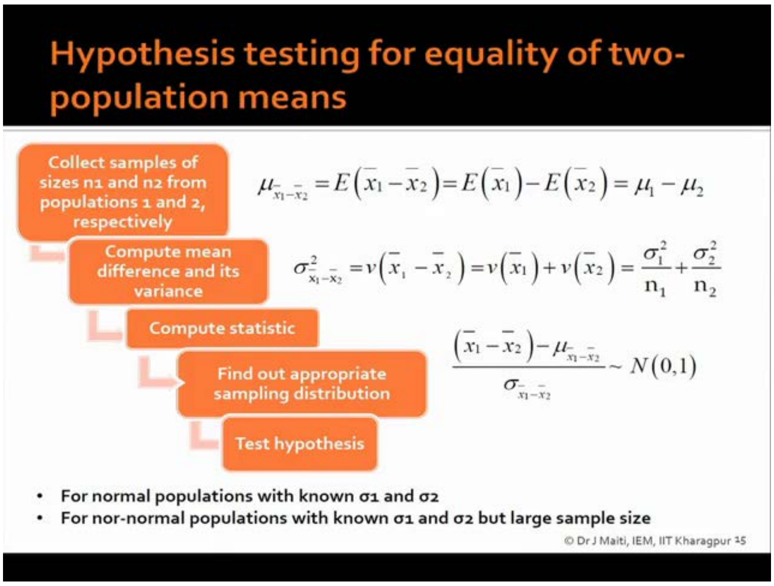


We say the ki square. Ki square is n minus 1 s square by sigma square. What is your null hypothesis sigma is sigma 0, what is your alternate hypothesis sigma not equal to sigma 0. So, then your computed value will be ki square computed which is n minus 1 s square by sigma 0 square. So, as you have taken two tailed condition here also, what you will do? You will just find out here it is ki square alpha by 2 and here also ki square 1 minus alpha by 2 and all of you know that ki square distribution. One parameter of ki square distribution is the degree of freedom.

So, here you are required to write degree of freedom n minus 1 here also, it is n minus 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.

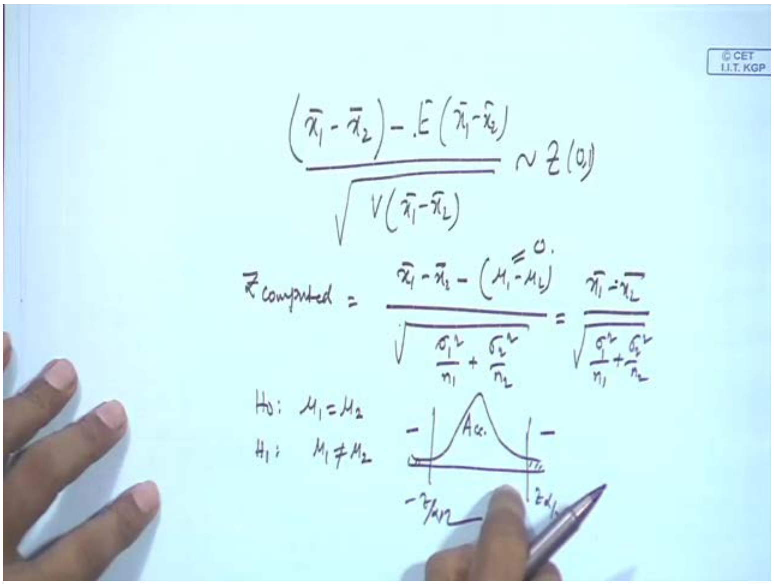
So, in that case alternate hypothesis if you have created it like this, there is no problem but I have given you some example. Based on that example if you go, that mean from the problem to hypothesis then I think got the better. Higher is the better nominal is the best, that particular concept you must use and again if I go for the next one, that also that one also easy for you. This is one example I have given and anything, no problem at all.

(Refer Slide Time: 42:54)



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

(Refer Slide Time: 43:11)



x 1 bar minus x 2 bar. So, this x 1 bar minus x 2 bar, that will follow normal or z, that is t distribution. Depending on the conditions and all of you have seen that x 1 bar minus x 2 bar minus expected value of x 1 bar minus x 2 bar divided by that variance of x 1 bar minus x 2 bar. 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 x 1 bar minus x 2 bar minus, this is mu 1, minus mu 2 by this one. We have seen last class n 1 plus sigma 2 square by n 2, that we have seen what is your null hypothesis here H 0 mu 1 equal to mu 2, that I mean this quantity becomes 0.

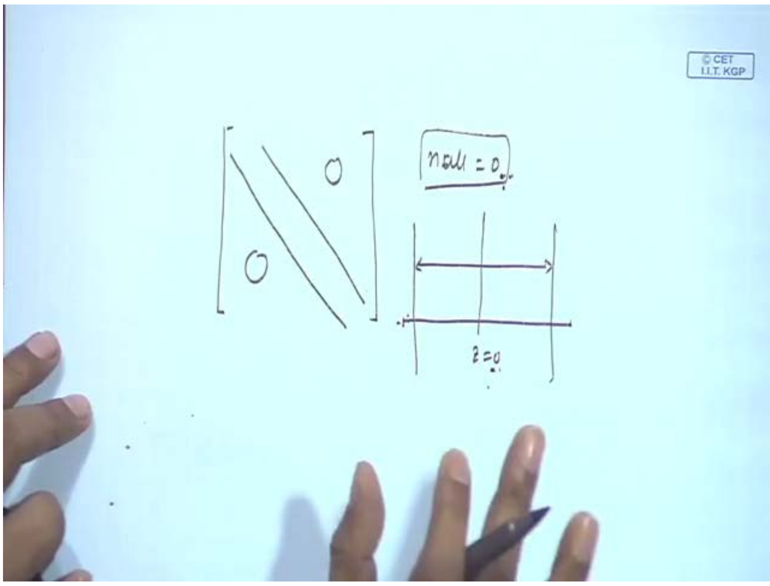
So, your z computed becomes x 1 bar minus x 2 bar by square root of sigma 1 square by n 1 sigma 2 square by n 2. You know n 1, n 2 and sigma 1, sigma 2, compute this and then again find out. You set your alternate hypothesis mu 1 not equal to mu 2, that is two tailed, two tailed case.

So, z alpha by 2 minus z alpha by 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 matter is 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.

Student: 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 modelling.

(Refer Slide Time: 47:38)

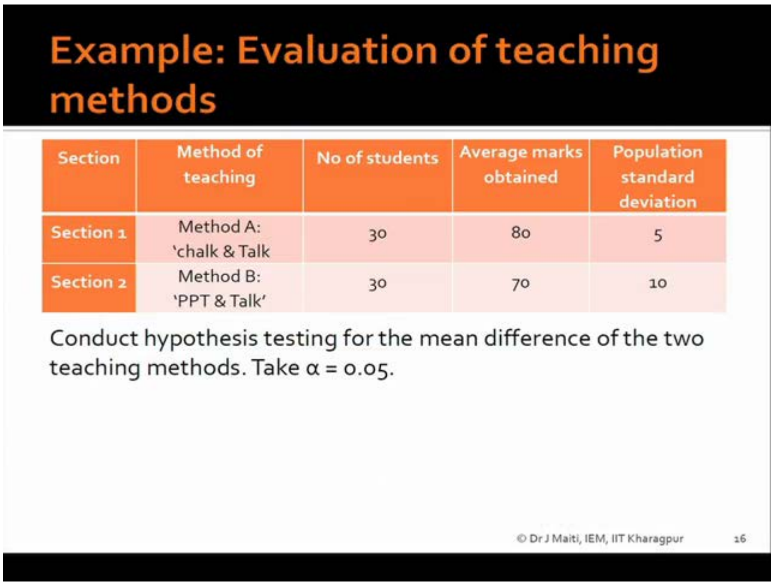


There are they have taken co variance matrix and then only the variance component they consider and other component they put it to 0. So, I may not be 100 percent sure that null basically is devoid or 0, getting me? So, if this is the case, so then when we test you will find out you create. Suppose, 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.

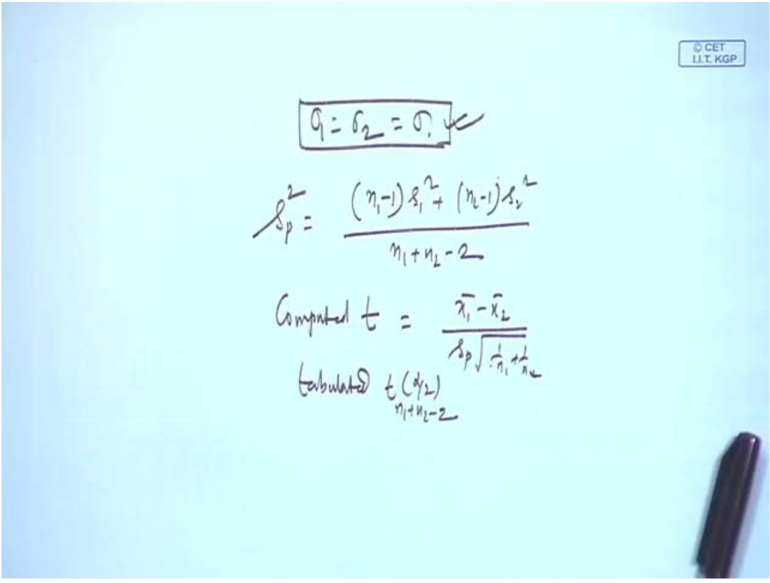
So, it is sufficiently away from 0 so long it is close to 0. That means so long within this region say it is in the null region, if I say null equal to 0, null means void. Void means 0, that sense I think this maybe but one of the explanation but not 100 percent sure that whether this I think this can be thought of.

(Refer Slide Time: 49:32)



This is the example. 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 is 5 and method B it is 10. In your confidence interval you have taken this 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. No, here we are basically doing the difference between 80 and 70, that is mean difference we are taking and considering population that 5 and 10, not that what I said that ki square distribution here.

(Refer Slide Time: 50:56)



Now, you come to the special case, what is the special case? Sigma 1 equal to sigma 2 equal to sigma. This time you will use s p square, 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 equal to mu 2.

So, if I use t computed, t will be x 1 bar minus x 2 bar minus 0 because mu 1 minus mu 2 equal to 0 divided by s p square root of 1 by n 1 minus 1 by n 2 and go for tabulated t tabulated t n 1 plus n 2 minus 2 alpha by 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.

Student: This is just a special case of the original?

Original one.

Student: Original one? Original one was the.

Sigma 1 square by n 1 plus sigma 1 square by n 2.

Student: That special case would hold for the original one also using others cases.

Okay.

Student: What is the requirement of the computation of the several type of convolution.

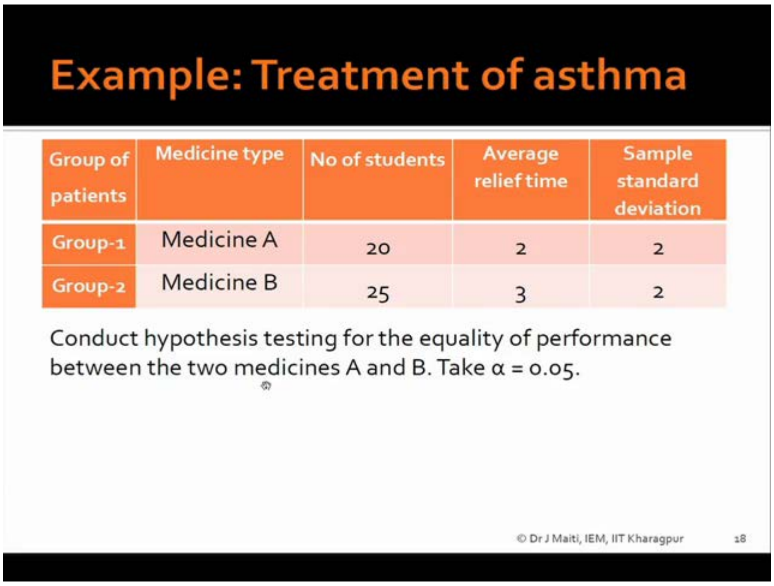
Correct, correct. Now, because under this special case scenario what happened actually all this that whatever the distribution you are finding out this distributions are governed by the number of parameters. As the size of the sample and all those things are involved here. Now, what happen when this special condition arise that sigma 1 equal to sigma 2 sigma, then if we convert this s p, if we find s p square 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 alpha error or beta error is there. When the power of the test will be better that is why this 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.

Student: That this is better case?

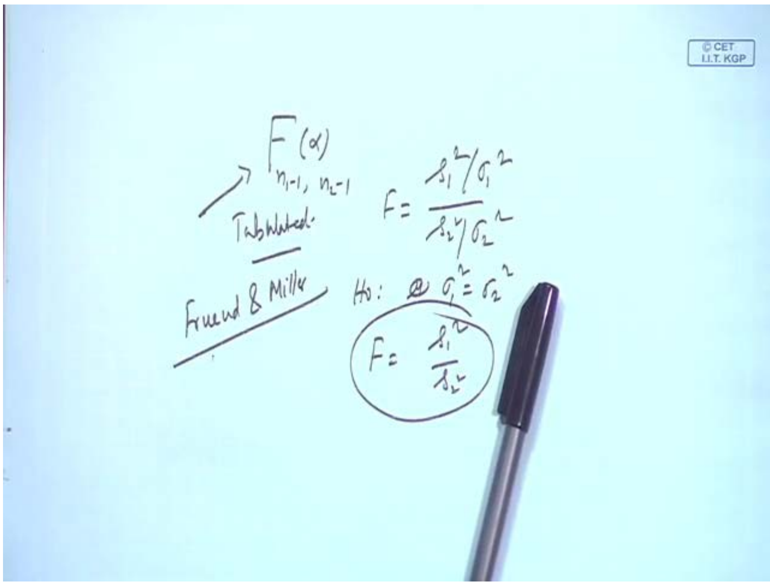
Always better.

(Refer Slide Time: 53:58)



This is another example that we want to equality of performance between two medicines. Confidence interval case you have seen. 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: 54:36)

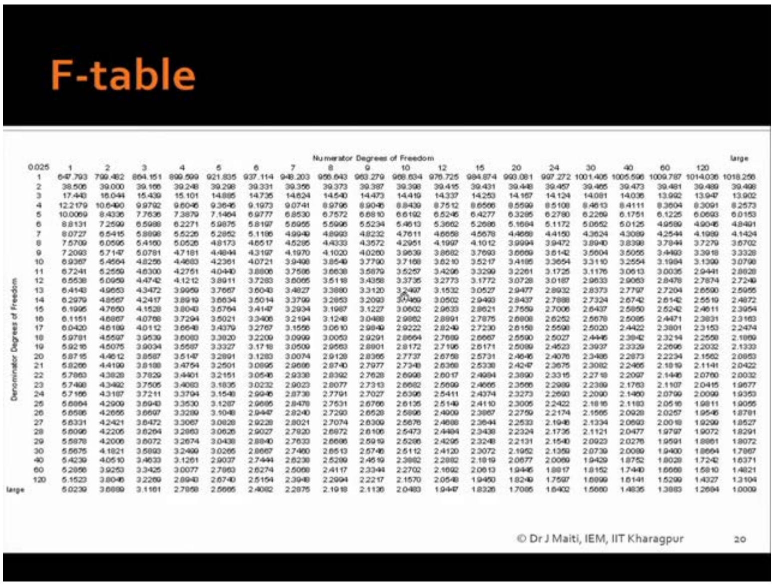


So, your F n 1 minus 1 and n 2 minus 1 and definitely you will be getting certain alpha values and that will be your tabulated one and what was the statistics. That time statistics we considered F equal to s 1 square by sigma 1 square by s 2 square by sigma 2 square. What is our hypothesis? Here H 0 that sigma 1 square is equal to sigma 2 square.

So, sigma 1 by sigma 2 whole square is basically this is 1. So, your F is basically resultant F, basically s 1 square by s 2 square. 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. Here some special cases also occur basically depending on the sigma 1, sigma 2 that which one is more, which one is less.

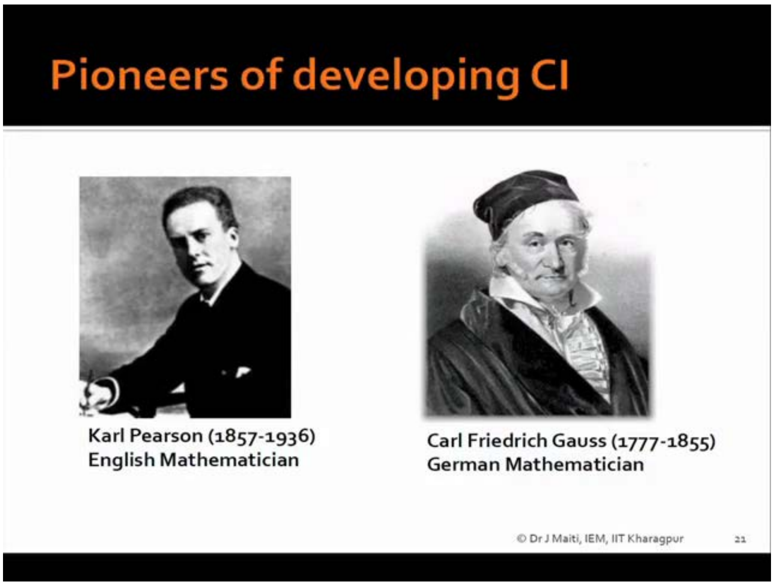
You go through by this Frueud and Miller book. Frueud and Miller basic statistics that book basically engineering I think this is basically introduction to statistics. That book you go through, you will be finding out many more cases are there..

(Refer Slide Time: 56:13)



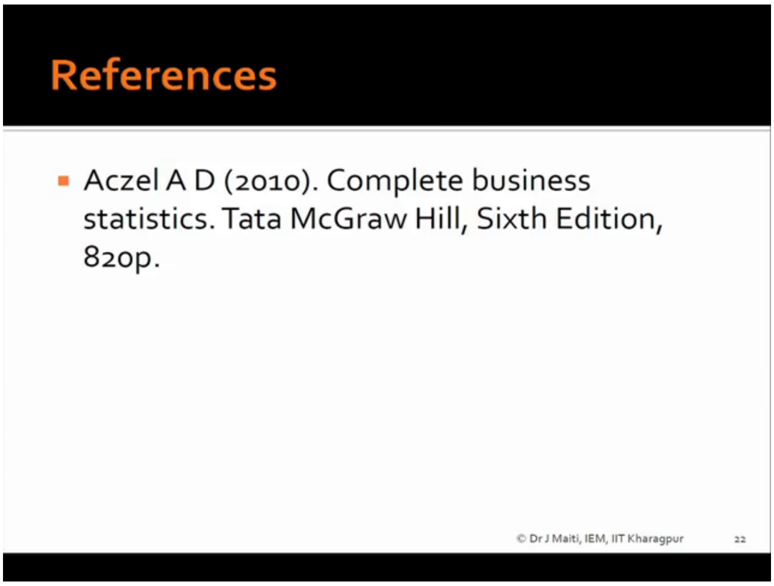
This is F table.

(Refer Slide Time: 56:16)



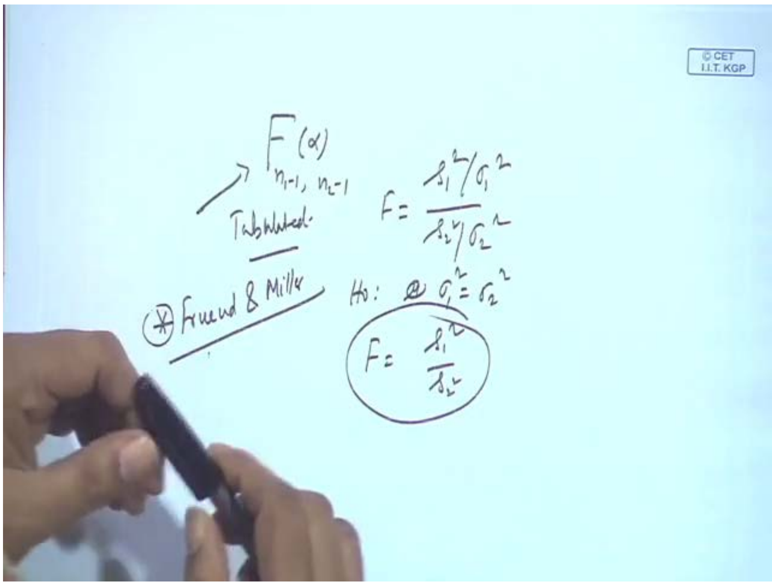
I told you last class that when the confidence interval as well as this hypothesis testing.

(Refer Slide Time: 56:26)



These are the some pioneers hypothesis. Fischer has tested so many hypothesis in each lot many experiments. What he have 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.

(Refer Slide Time: 57:05)



In addition as I told you that Frueud 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 univariate case, multivariate books are very limited.

So, next class what we will do? We will basically talk about multivariate descriptive statistics. 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 univariate. 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 univariate. 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 univariate basic statistics.

Then thank you.