INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

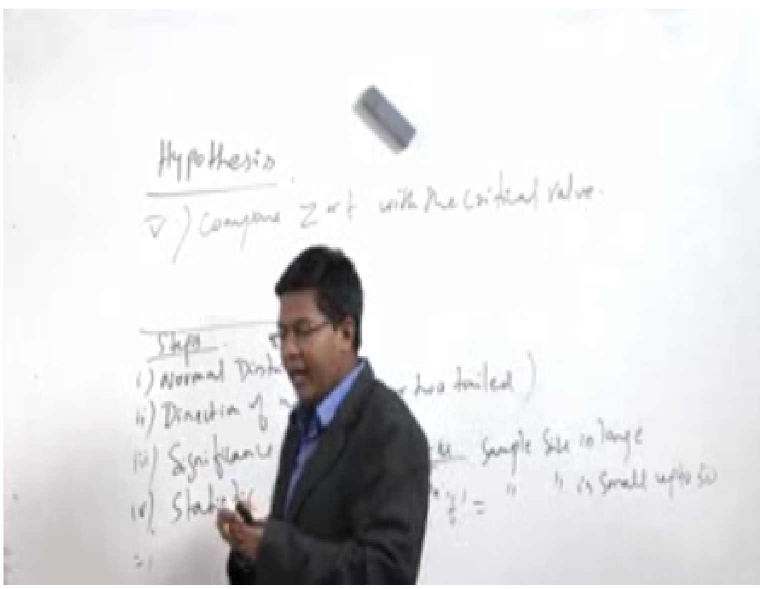
Marketing Research

Lec- 19
Hypothesis Testing: T-Test, Z-Test

Dr. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology Roorkee

Welcome everyone to the session of marketing research analysis. In the last session we have discussed about hypothesis. We had introduced the subject of hypothesis.

(Refer Slide Time: 00:31)



So what exactly is in hypothesis and why it is so important for any researcher. So hypothesis is basically as we understand is an assumption. So normally we say I hypothized that this is going to happen, my hypothesis that today it might rain or I have a hypothesis that this new machine will work better than the old machine. So these are basically that means the thing has not happened and we are trying to predict, we are trying to say something in positive or negative right.

So hypothesis is basically an assumption as we understand. So the question is, so when we did it, we say that there are two types of basically any hypothesis of two types. So the null and the alternate, so the null basically we said is one where the researcher follows the mentis the status-co, that means something would happen what it is normally happening right, or it is a case of n equal to.

That means two things are equal and till we have not proven it. So the null says they are equal to, so it is a case of basically that is why it says it is a case of equal to case. So let us say the mean of two groups is equal to or the intelligence of two people is equal to or same right. So the words same equal to come into the null. On the other hand, alternate is something which is not the null or against the null you can say right.

So suppose the researcher says once to know the intelligent of two people are same or not same, so the null says it is same and the alternate says no they are not same so they are not equal to so this is the case of equal to this is the case of not equal to okay. and the most important thing is that in most of the researchers be say with this want to disprove the null hypothesis we want to disprove or reject the null hypothesis why?

The question is very simple because any researcher if you wants to except with the status co what is happening would happen then what is the point of doing a research? So in order to have an outcome which is significant or effective you know it effective in those cases we always like to check the more important thing the alternate for us right that is why if you read research papers please understand the hypothesis that you see on those research papers on the research papers I have basically they are not the null hypothesis they are the alternative hypothesis right this is the hypothesis that the researcher actually wants to clamed or check okay.

Now let us get in to the subject during hypothesis testing in the last session also I had discuss there are basically sacral steps to check the hypothesis. So what are the steps so first we one has to check for the basic assumptions okay of a normal distribution, so assumptions of a normal distribution right that means the data behaves in a normal manner right it gives similar to a normal distribution right.

Then we said you need to you know check the tail effect is the direction of a test right so whether it is a one tail one or two tail test okay then we sais the researcher has to be sure what level of significant he wants to work right now significant level has to be decided by the researcher now when I am saying the significant level one can understand it as the α right.

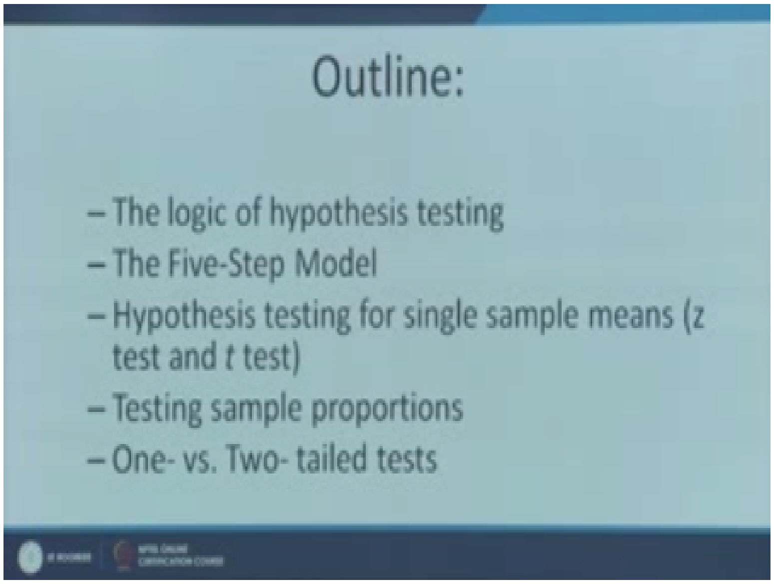
So if you remember α we said this is the chance that the hypothesis which is correct is will be still rejected right the chances of rejecting a true hypothesis is called α. Rather the other side is β which was the chance of accepting a false hypothesis okay so once we do the significant then we calculate the statistic.

Now what is statistic her, now the statistics means we say the z or t statistic right, w calculate the z or t statistic, so that means and how do we calculate, now z I have said if you remember z or t whatever if you see is equal to x which is the x-μ right up on the standard error right this was the formula. So same thing will happen also applicable for t the only difference between a z and a t distribution I will explain is that to start with you have to understand that a z distribution or z test is used when the sample size is large, okay.

t is used when sample size is small or up to 30 okay, 30 if your sample size is small up to 30 then we will use the t test, okay. Let us see, so once you have done with the statistic the final step comes is to compare the compare I think it will not be visible if I write here, maybe I will write at the top this point I am writing here, so visibility will be there. So the fifth point compares the statistic the z statistic z or t statistic with the critical value.

Now critical value is something that you can find from the z or t table at the end of any book or you can just Google out and you can see okay, what is the value.

(Refer Slide Time: 07:01)

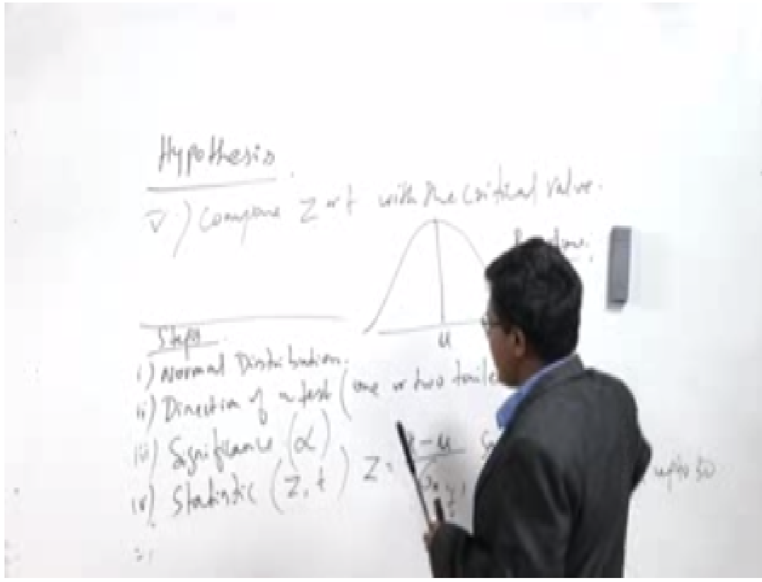


Now this is the outline so let me go through it, so we are getting into the logic of hypothesis testing, so the five step model which I just said, so hypothesis testing for single sample means one sample or it could be more than one sample, sample proportions basically this is the when you talk about z or t there are two ways of understanding. The hypothesis testing is call as a test of means or proportions, right what does it means? It says it is the test of means or proportions that means what, two let us understand, let us go to the basic meaning.

What does hypothesis testing actually say, through a hypothesis testing w are actually trying to say that there is a significant difference between the sample mean and the population mean or there is not a significant difference as good as it, that means if I still break its do a more you know elementary level it means if can I say that from a sample mean that this sample comes from particular population or not, is the sample a part of the population or it is some other sample, it is not related to the population.

We are trying to actually test this thing right, in any hypothesis testing so but the question is when we are taking understanding it from the terms of mean right, so we can check it from the terms of mean so what is happening in the normal distribution we said okay, basically we are bothered about the mean right.

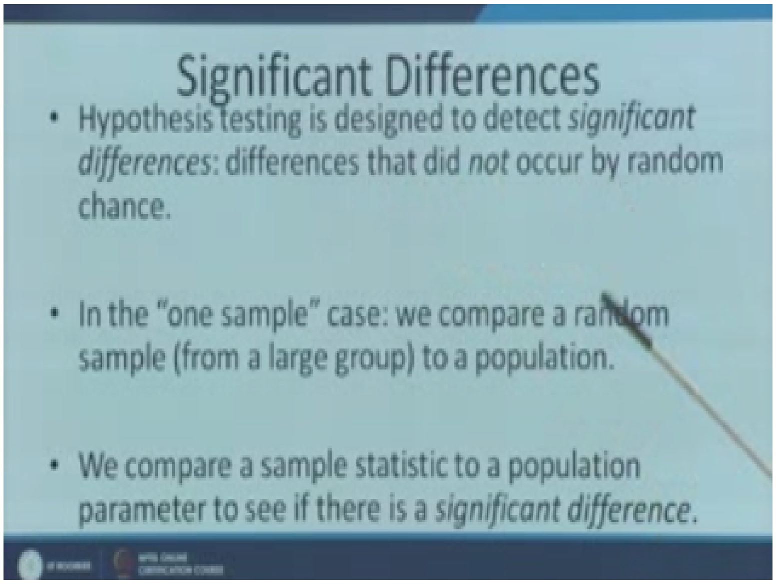
(Refer Slide Time: 08:47)



So that is why if we take the mean of the sample and we take the sample population sample mean and then we say is it significant some relationship is there or nor or if the suppose the reach something is it by chance or it is the sum, it is it would happen again and again as good as that, so one is test of means the other is as I said test of proportions right, so proportions as we understand from mathematics right, so proportions are something which is in a form of p and q right.

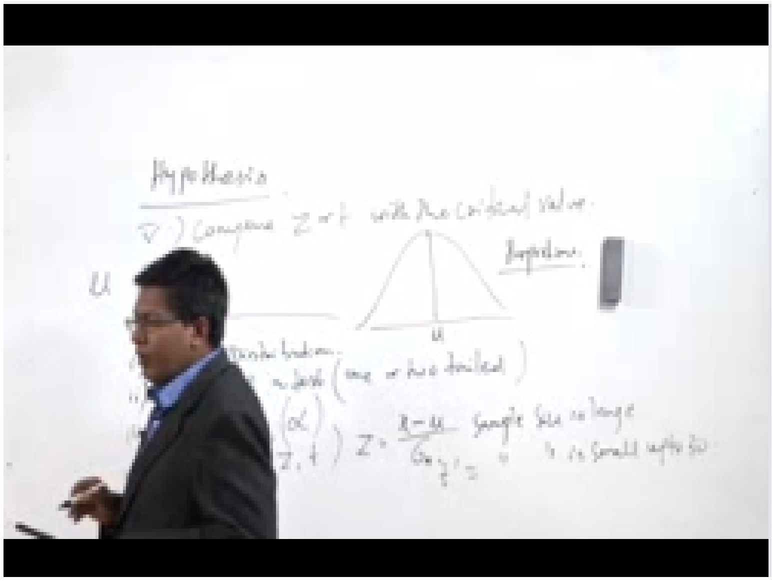
So it is the ratio basically, we say right so what is happening we are trying to see okay whether if there is no mean and we only have a idea okay, what percentage of the population it is there or not there can it have such situations also, can we test hypothesis yes, so in those cases when you do not have the mean you can use the proportion okay. Let us see I have an example I will show you. So let us say.

(Refer Slide Time: 09:49)



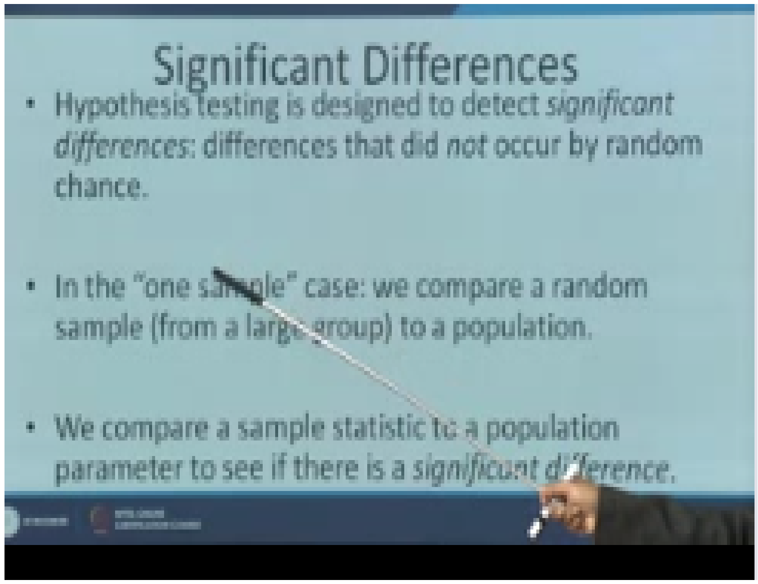
Hypothesis testing is designed to detect the significant differences as I said that did not occur by random chance, so if there is a significant difference we are saying there is a significant difference between the population mean and the let us say sample mean let us say okay.

(Refer Slide Time: 10:06)



Now we are saying that this significant difference that is happened will happen is not one which has happened due to some chance element it is actually it does happened right so to claim or to test this we are doing the hypothesis test right.

(Refer Slide Time: 10:22)



So there are two three types of test basically okay now if you let me rub this off I am not getting space so basically as I said the z and the t these two are more or less the same thing right if you look into any statically software or anything you would not see a z test because that it is that z is nothing but an you know it is extension of the t only right so only thing is that t is small and this is a large right so this is around 30 as I said so what happens is how does the distribution look like now if you take a t distribution the t distribution the curve is something like this okay now what I mean that means what if you look at the curves you know.

The curves of the height the Preakness of the curve the t test the t distribution is more flatter is more flatter than the z distribution the z distribution is more or less it is normal in nature right but the t is more flat and tampering at the end so what happens in this situation what is basically happening is so when a t when you extend increase the number of sample size for example okay.

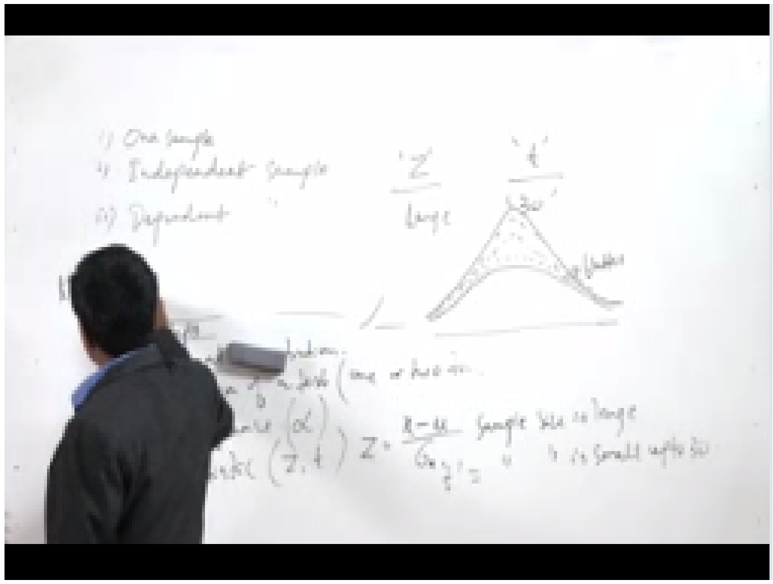
So as you go on increasing the sample size the t tends to become a z so that means the t and z hardly there would be any difference because when you increase the sample size from 30 to let us say 40, 50, 70, 80, 90, 100 whatever then automatically the t and the z would more or less look same right so there is a basic understanding.

So okay now I was saying so that is why we basically when we talk about the t test right what we are doing is t test is of again is of three that t basically test one is called one sample t test okay the second is i8ndependent sample t test okay the third is the dependent sample t test okay so the t test has been you can say there are three types of test right the t test can be explain in terms of the one sample t test the two sample or independent sample t test and the pair sample t test right now what does it mean what is that one sample when you have only one sample When you have only one sample.

And you want to compare you want to compare again sampling right what will you compare so you will compare this one sample the mean of this sample against what you will compare against some hypothetical some hypothesized mean right so that means let us say when you have got a group of sample let us say the intelligent or the score of a group of people right of one section or one class you want to check the you are checking the mean right.

So you found something the mean is let us say 60% or something okay now this 60% is significantly different from the population mean or not how would you know and to know that is such cases what we do is basically we compare it against the hypothesized population mean right so the hypothesized value.

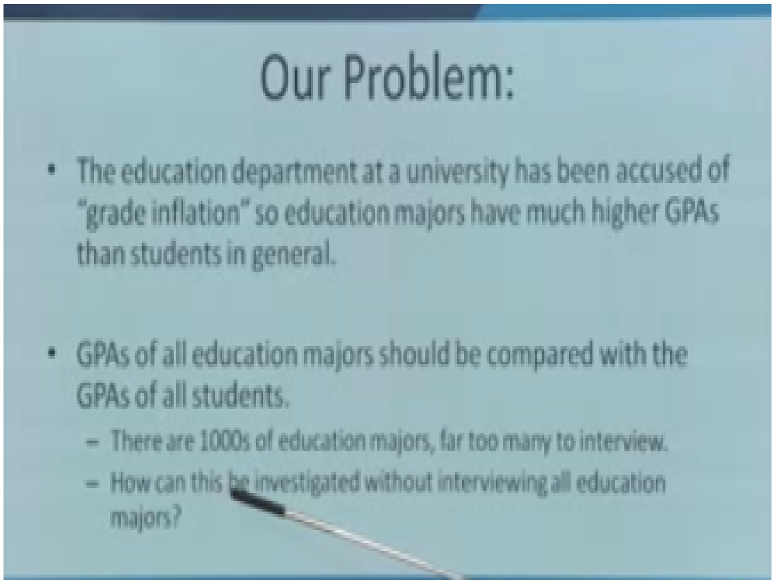
(Refer Slide Time: 14:17)



That we use is basically something that we compare and this value must have come may be from some past record or past experience so we know that the people in this class generally in a class of let us say marketing research score around let us say 70% marks okay so now this is a something that we have hypothesized from the because of the past records and now whatever we have calculated from that one sample right.

Now we will compare this mean with that hypothesized mean value okay so we compare what is saying we compare a random sample okay from a large group to a population okay and this population value is the hypothesis value second we compare a sample statistic to a population parameter to see if there is any significant difference or not.

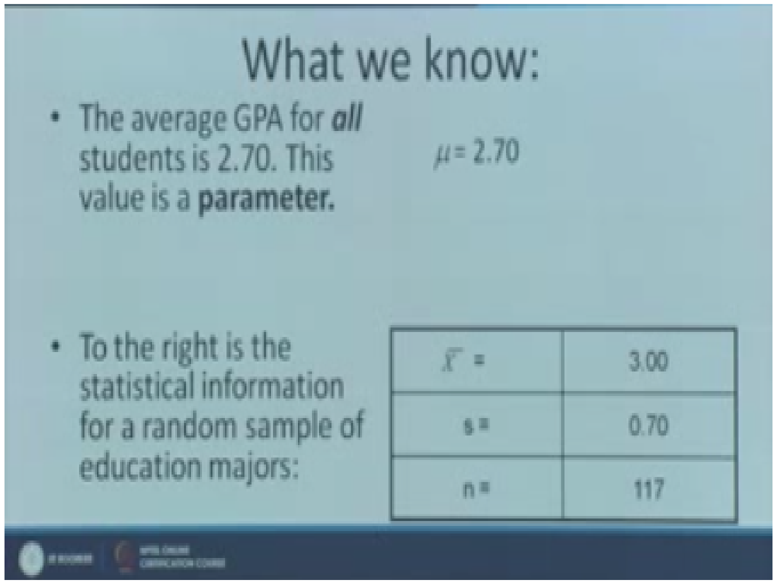
So this is highly useful for those studies like example in industries in manufacturing industries where you know they are trying to find they are making suppose some kind of products and they want to check the strength of the products, so they will take the ample and they will compare it again some earlier value that they know atleast the share should be able to take the take 150 weight at one times suppose. Now that 150 k is something they are comparing against the sample mean against this 150 okay, let us say this is the problem we have taken.
(Refer Slide Time: 15:47)



The education department at a university has been accused of grade inflation they will accuse the grade inflating the grade so the education majors have much higher GPAs than students in general. Now people who have taken education as a major subject right have found to be having higher GPAs right then the students in general okay GPAs of all the education major should be compared with the GPAs of all students there are generally if you see so we have to compare the GPAs of all the people who are having education as a major and the non ones and check them.

So there are thousands of education majors right there are thousands of subject which are where people have majors right and which is too many to interview it is very difficult to work on such a large sample okay large group how can this be investigated without interviewing all the majors so you have around thousand majors or more than 1000 majors now if I am going on if I go on checking then it is like checking the whole population and that is not wise and that is not advisable because of the lakh of time and money so in such a condition what we will do now what we know the data says.

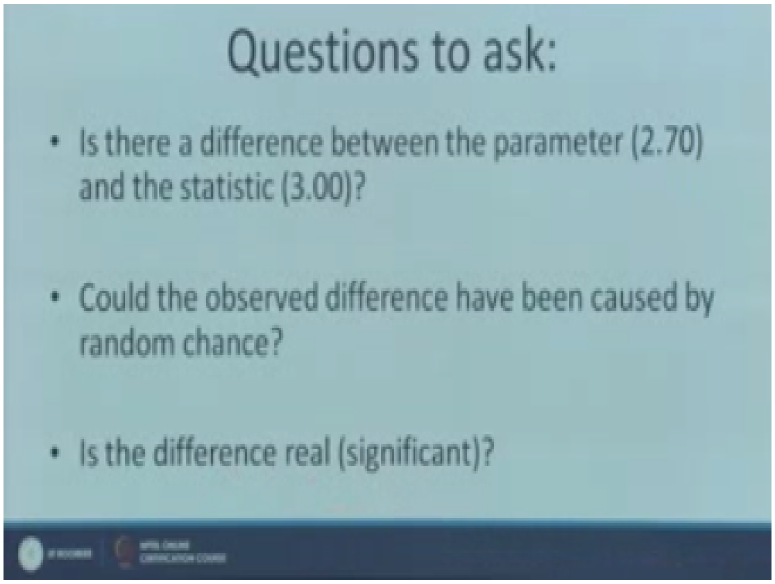
(Refer Slide Time: 16:59)



The average GPA for all the students is 2.7 okay now this is the population parameter that means that is the population statistic okay. So μ = 2.7 if you remember I told you μ is the sign symbol you assure population now if you look at this is the sample values right the x bar or the sample means sample mean means to the people who had some kind some majors education majors right some subjects let us say history bio technology or anything right.

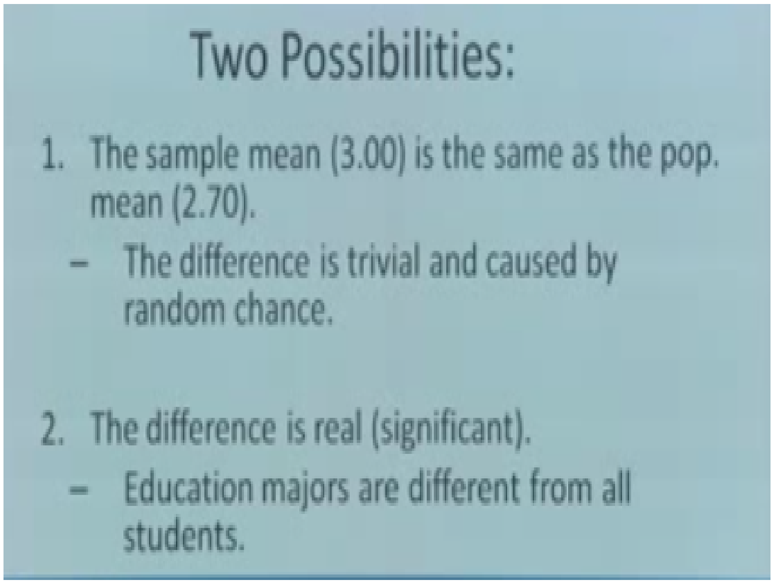
There scores were taken to found to be the sample mean was 3 right the s is the sample standard deviation there is a population standard deviation which we denote by let us say σ okay. Now this is the sample standard deviation s is 0.7 n = 107 so they have taken 117 candidates to took their interview to the score and they wanted to do the test okay.

(Refer Slide Time: 18:04)



The question is there a difference between the parameter the population mean let us say and the sample mean if I am asking is there a difference between the population mean and the sample mean yes or no, so to do that what we are saying could the absorb different if suppose there is a different we are finding 3 – 2. 7 is 0.3 but is this difference actually really there is a difference or it has is there a chance that by chance it has happen by those samples which we are taken so is there a difference real we want to check okay.

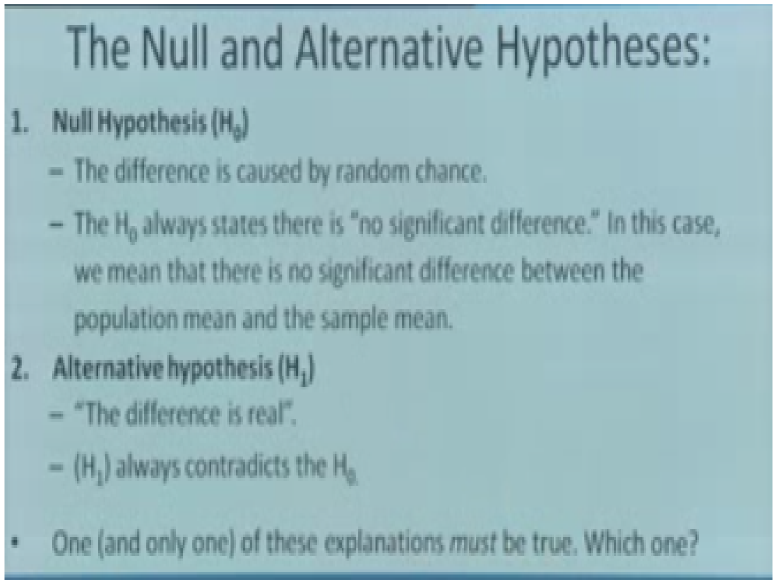
(Refer Slide Time: 18:42)



Now it saying the sample mean is the same as the population means two possibilities there are two possibilities actually the sample mean is the same as the population mean that means it is only by chance it has happen is time the difference is trivial and caused by random chance okay or the difference is actually significant the difference is real the education majors are different from all students that means the people who have taken education some education majors there mean that they have derived the scores are actually different from the other students okay.

Now what is the as I said if you remember you have to first ride the null and alternative hypothesis.

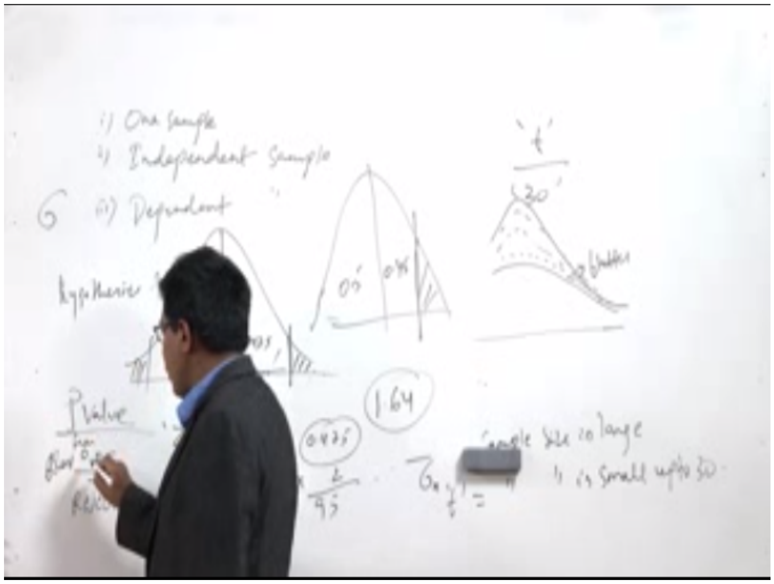
(Refer Slide Time: 19:25)



So the null hypothesis is what is this the difference is caused by random chance so it states there is no significant difference what does it say that whatever is a happened if there is a difference of 2. 03 or something this is due to a chance okay and there is no significant difference between the two groups the ample and the population in this case we say that there is no significant different between the population mean and the sample mean right but as a researcher are you interested to find that no so what are you interested to find?

Now we are interested to find to see that no the difference is actually real now what is it mean now it says that there is difference that means the population mean the difference between the population and the sample is significant in nature right So and if you see both the explanation cannot be true the true possibilities cannot be true null and alternative at the same time cannot be true only one either it is a difference or there is not statistical difference okay. So now which one is true let us see?

(Refer Slide Time: 20:36)



So to assuming that the null hypothesis true right we always test the null hypothesis we always we will although we will interested to have the alternate but we will check the null hypothesis what is saying what is the probability of getting the sample mean 3 if H0 is true and all education majors really have the mean of 2.7 in other words the difference between the means is due to random chance null hypothesis right.

Now what is the probability? Now he is taken if the probability associated with this difference is less than 0.05 reject the null hypothesis now we should remember I had told you in the last session also how do you except or reject a hypothesis so if I said the suppose you can find out the z value right now if the z value that you have found out you compare it against the table value now table value at what confidence level that you have to decide earlier now why earlier.

Now the point is if you do not decide their confidence levels earlier from the beginning such let us say 95% or 99% then the researcher might change his mind if he does not get the decide result so that is why it is always you have to decide it fix it from the beginning okay. though z value let us say at 95% for a also said for a two tail test and tail test both the values would differ for a two tail test may be the regression roles are spreaded two ends right so at 95% it becomes 1.96 right.

So the area is basically if you how do you check it now you can go to the normal distribution the table and look at the value of 0.475 now why it is pose it upon 0.475 now 0.475 x 2 is basically nothing but 95% so if mu two tail or taking 0. if you how do you check it now you can go to the normal distribution the table and look at the value of 0.475 now why it is pose it upon

0.475 now 0.475 x 2 is basically nothing but 95% so if mu two tail or taking 0.25, 0.25 so I have 0.475 here.

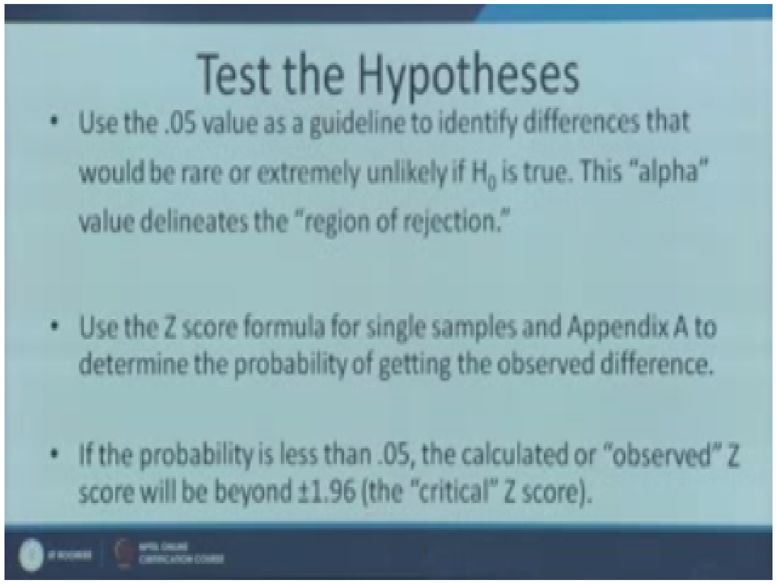
But if it is only a one tiled or one tile test it will look something like this okay, that this side may be suppose I am not getting the directions, suppose I am not interested in the left I am only interested in the right, so the rejection will lie, all the 5% will lie here okay so this portion is 0.56 as it is this become 0.45. So to do this if you want to check the area under the curve, so if you look at 0.45 at the table you will find the area under the curve or the Z value sorry, is not 1.96 now, it is only 1.64 okay.

So once you have calculated then you see if this is the acceptance zone right, now whether your value falls this side to this cutoff value or this value, if it is falling this side your null hypothesis is accepted right. But if it is falling somewhere this side away from the cutoff value then it is rejected okay. This is one; I also said if there is something called a P value, now a P value in the last session only I told you P value is the probability value.

If this P value is less than 0.05 in your case right, what is this P value basically it says what is the chance of the value of your calculated value falling at a extreme zone or the other extreme ends okay. So if it is less than 0.05, then you reject the null hypothesis right. If it is less than 0.05, if the P value is less than 0.05 you reject, if it is less than 0.05 you reject the null hypothesis. But suppose it is more than 0.05 at a 95% confident level please remembers this, if it is 95%, if it is 99 the value will change.

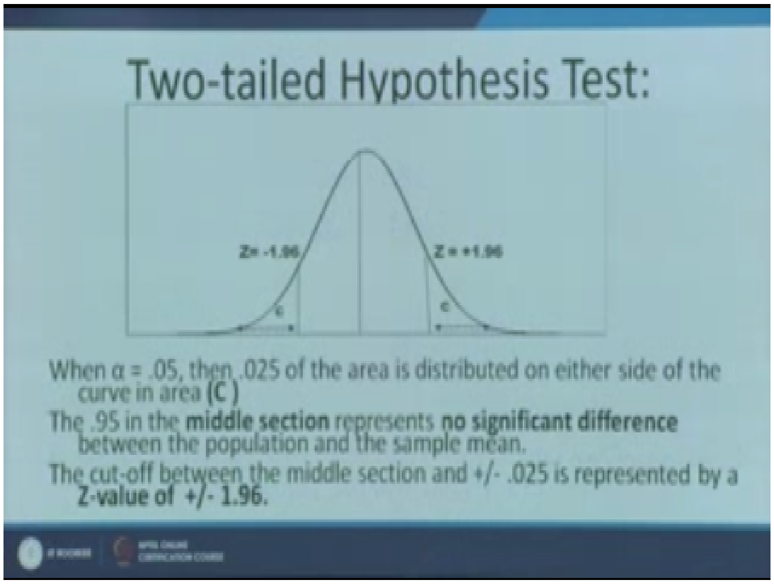
Then you will accept the null hypothesis right; accept the null, so it is a probability. So is it the probability of falling is within the 5% or is less than that, if it is less then reject it okay, as good as that. Now let us look at this, so I am not getting into this, so you have to calculate I have already told you so right, if the probability less than 0.5.

(Refer Slide Time: 25:06)



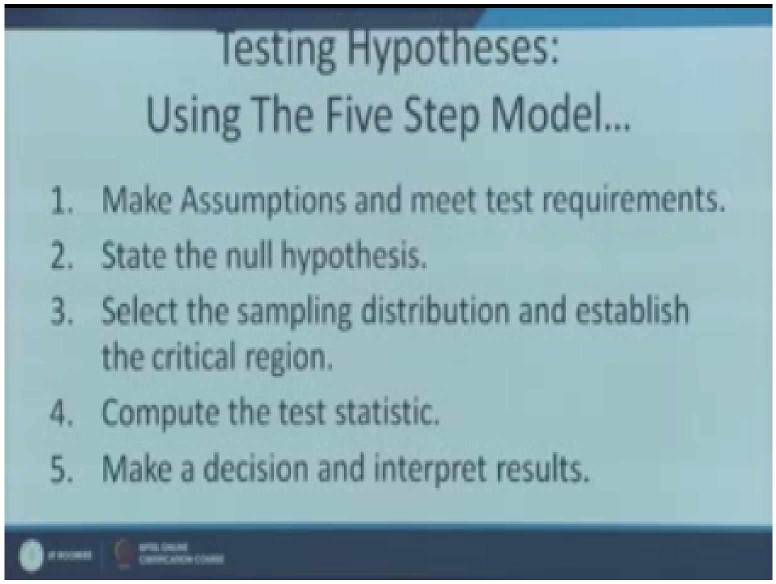
The calculated observed z will be beyond + or -1.96 as I said.

(Refer Slide Time: 25:11)



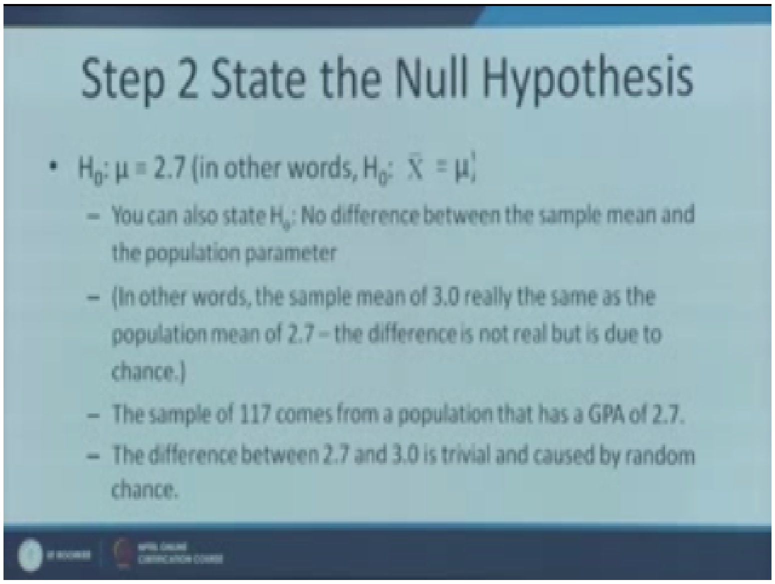
Now this is how it looks the cut off value right, so this is then area where you are talking, we were just talking about okay.

(Refer Slide Time: 25:19)



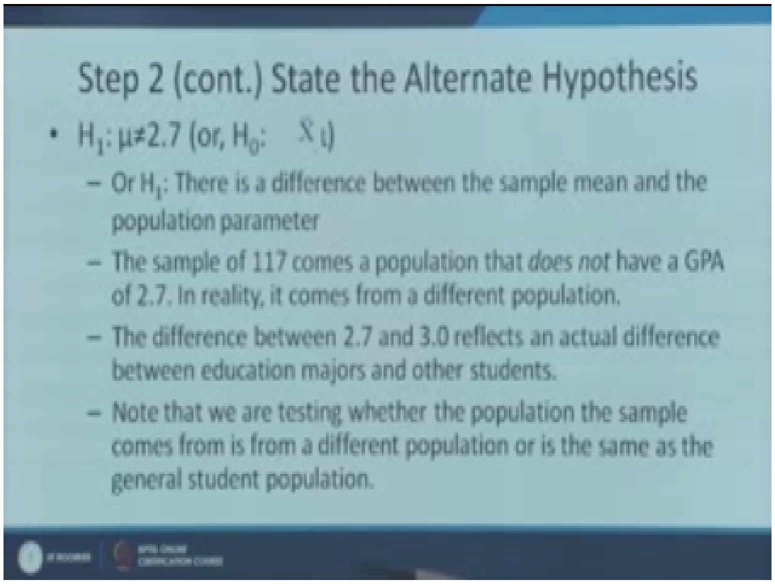
Now this is the five steps which I told you at the beginning right, now let us see what has been done, let us go to the calculation.

(Refer Slide Time: 25:28)



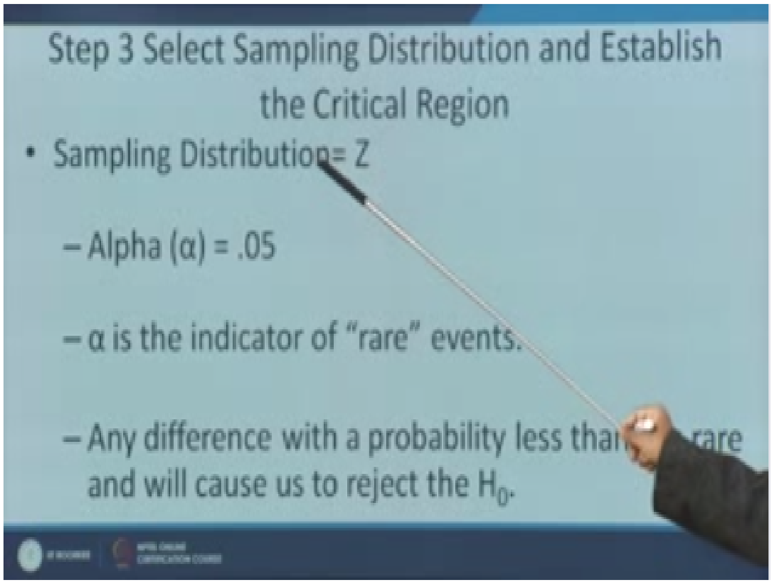
Now μ=2.7 in other words in null hypothesis we are saying that the population means and the sample mean are same equal, right so there is no difference okay, now the sample of 117 comes from the population that has a GPA of 2.7 right, the difference between 2.7 and 3 is trivial and caused by random chance this is what we have to prove okay.

(Refer Slide Time: 25:54)



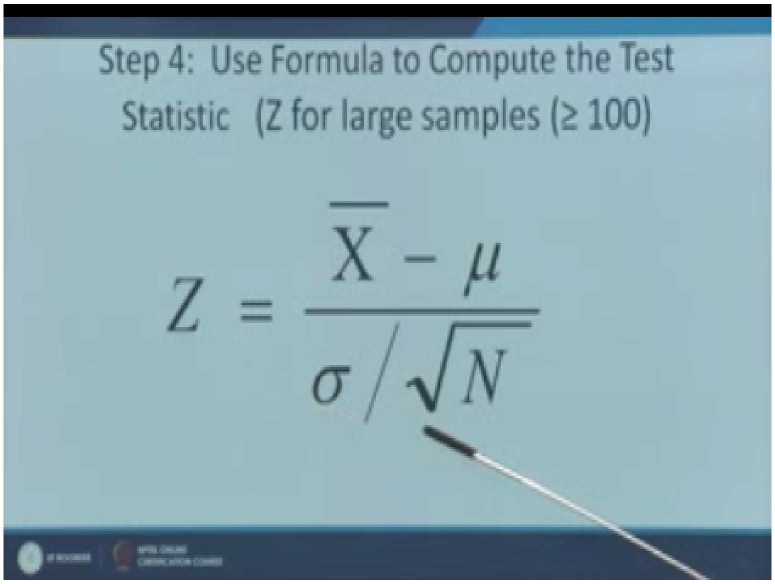
And what is an alternate hypothesis, now μ is not equal to 2.7 right, okay now let us look at the calculation.

(Refer Slide Time: 26:03)



Now what he is doing, the sampling distribution is Z right, now okay one thing you have to understand now whether it will be a two tailed test or one tailed test. So the μ is equal to the population or not equal to the population there will be two tailed right, because it can be less than it can be greater than. So any difference with the probability less than the α is rare and will cause us to reject the null hypothesis, okay.

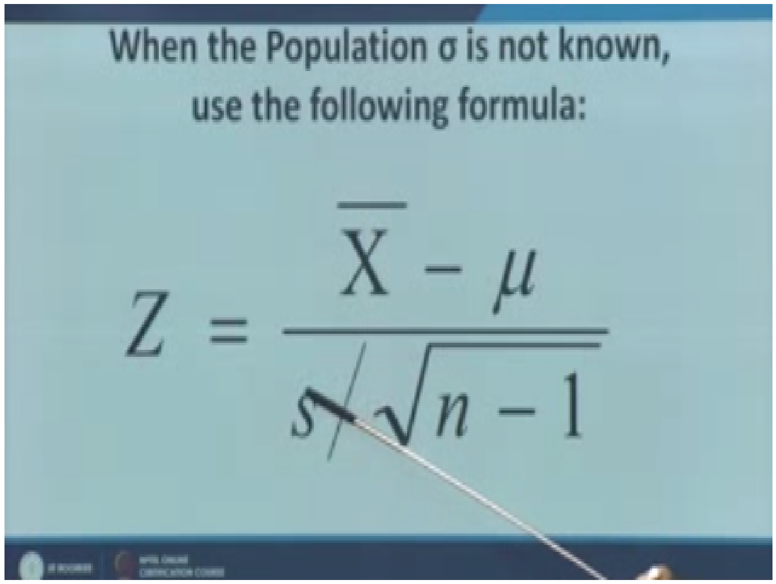
(Refer Slide Time: 26:31)



So let us go to the what is the formula as I have already done this formula many a times the Z for large samples which is greater than or equal to 100 if Z=$\bar{X}$ sample μ – the population mean up on the standard error or $\sigma/\sqrt{N}$, N is the sample size, so through also you can calculate the sample size as I told earlier.

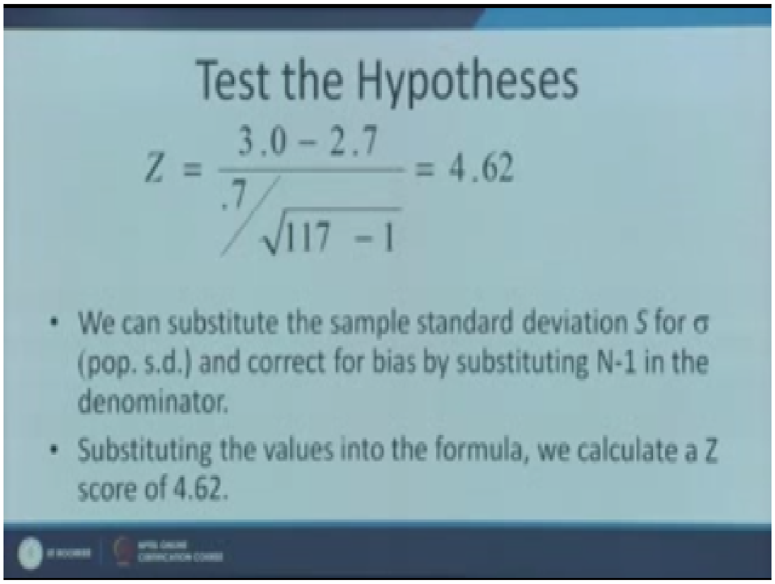
But suppose your sample deviation is not known, suppose the sample deviation, the standard deviation of the population is not known in that case your formula will slightly change that means if you do not have the population standard deviation you have to take the sample standard deviation.

(Refer Slide Time: 27:11)



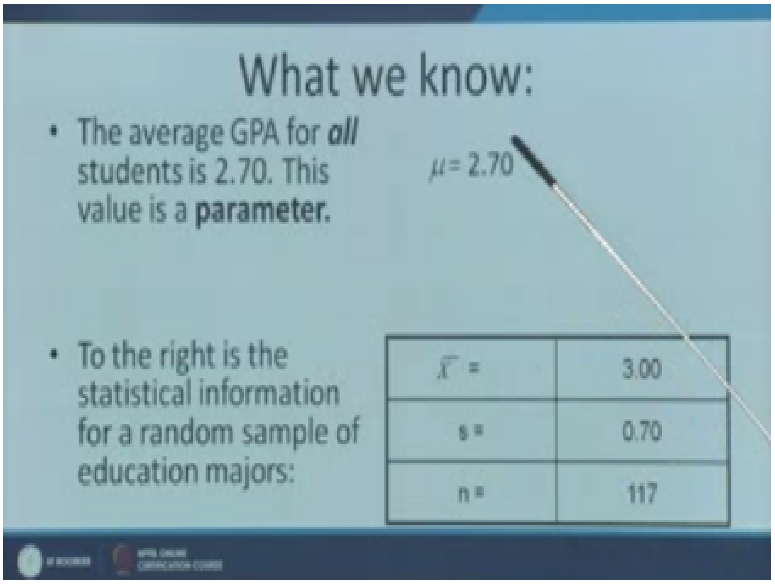
And when you take the sample standard deviation which was 0.7 if I am not wrong if I remember it was 0.7 you have to divided by a degree of freedom of not N by n-1 right, √ not N by n-1 but, so this is the only change right.

(Refer Slide Time: 27:27)



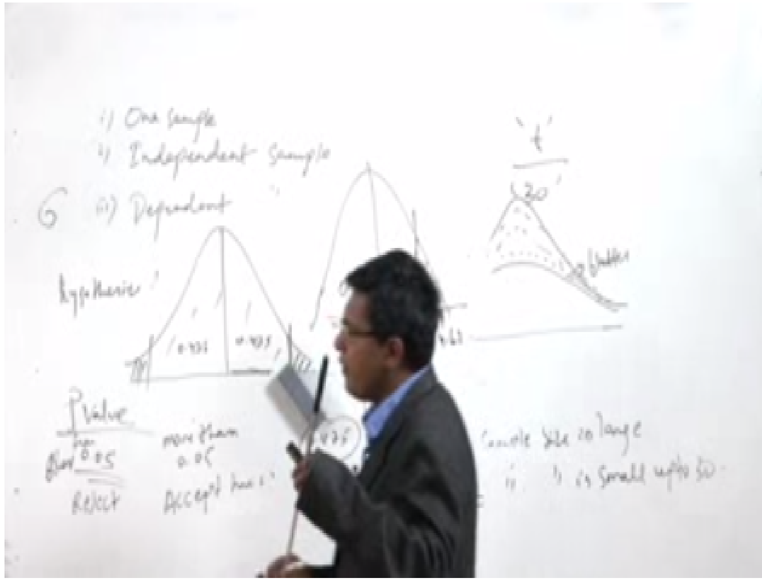
Now in this case let us go back and see so what it has done, so to test the hypothesis he has taken 3-2.7 divided by sample because you do not know in this case in our case with the population let us go back if you have forgotten I will show you, I think I show you okay, so if you see we did not have the population standard deviation.

(Refer Slide Time: 27:52)



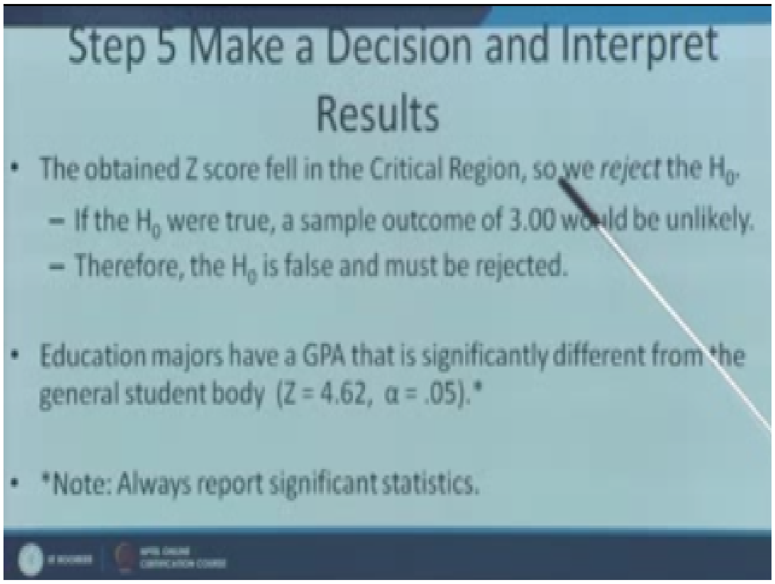
It was not given to us, we had the sample right, so we are using this okay, so 117.07/√117-1 so how much is the value in a 4.62. Now this 4.62 is obviously our 95% value was 1.96.

(Refer Slide Time: 28:13)



So 4.62 would come obviously this else somewhere here right, 4.62 so if it is there automatically you can understand that it is to be the null hypothesis would be rejected, okay.

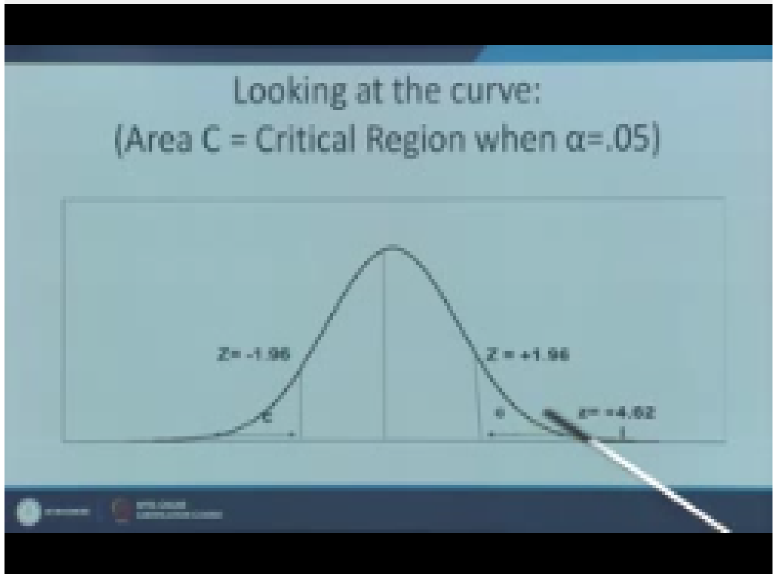
(Refer Slide Time: 28:29)



Now the obtained Z score fell in a critical region so we reject the H0, if the Ho where true sample outcome of 3 would be unlikely therefore the H0 is false and must be rejected. Now what is the conclusion education majors have a GPA that is significantly different from the general student body, so earlier hypothesis was okay, there is no difference between the education majors and the normal students.

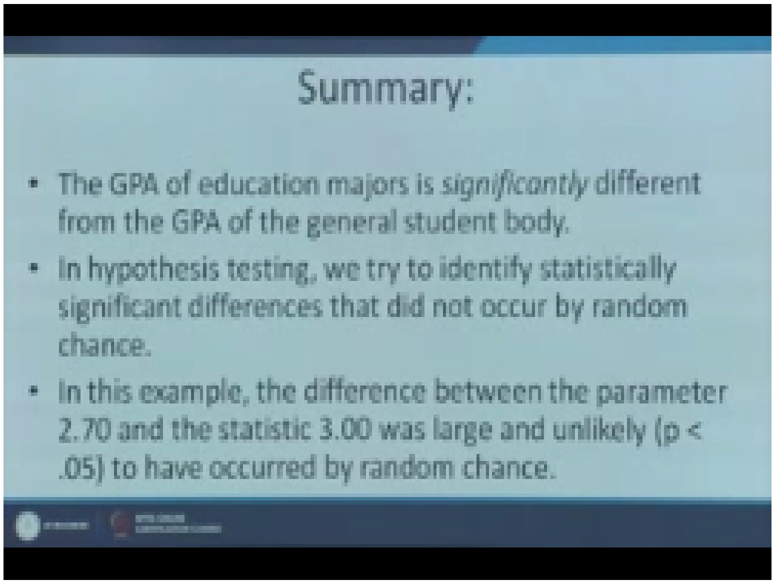
But now you are saying okay, no there are null hypothesis has been rejected and actually this is what we wanted that null hypothesis should be rejected and there is a difference between the two groups okay. So this is how it looks like 4.62 so this is 1.96.

(Refer Slide Time: 29:13)



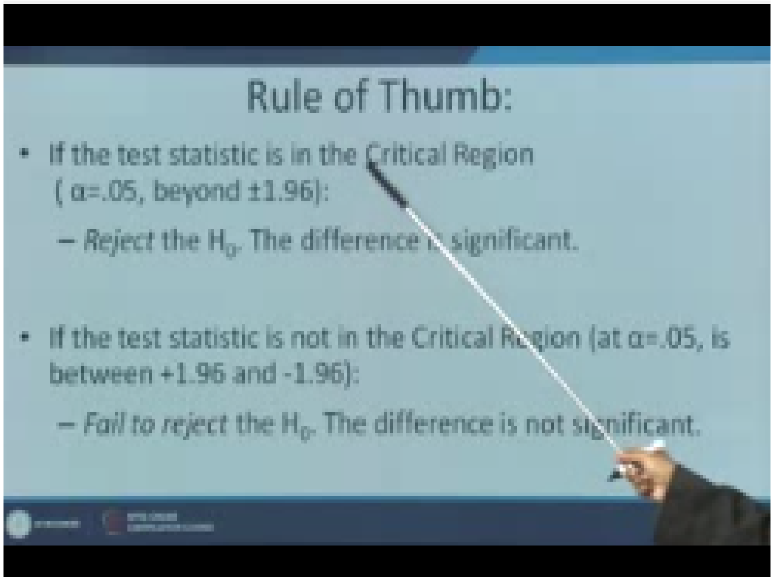
So we are saying it is falling somewhere here okay.

(Refer Slide Time: 29:16)



So summary is already I have explained the gp of education is significantly different from the general body so this is all right we are going to do so we rejected this 0 and concluded that the differences was significant right okay fine now this is the rule of thumb.

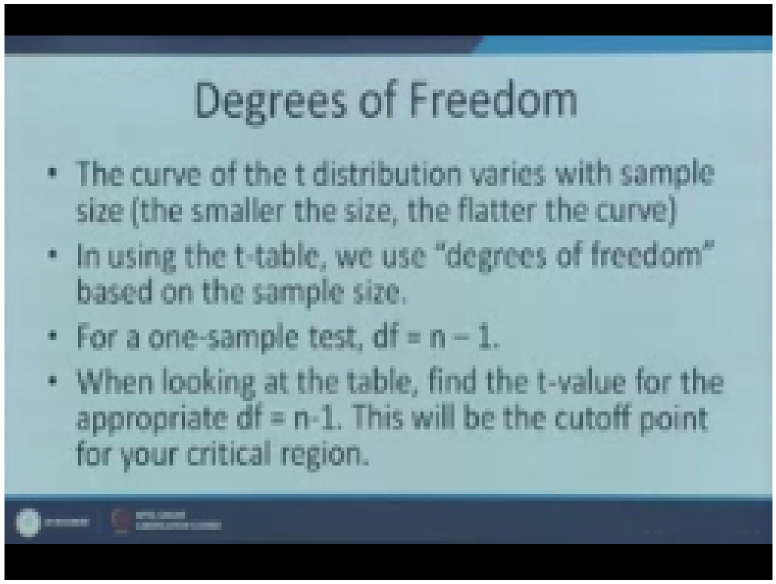
(Refer Slide Time: 29:38)



If the test statics is in the critical region α is 0.5 it is beyond reject the height 0 the difference is significant right suppose it falls in the critical region that is in between 1.96 to -1.96 that means what please if you remember I told you they never say we never ever should say that we have excepted the null hypothesis we always say we failed to reject the null hypothesis so understand the differences in our test.

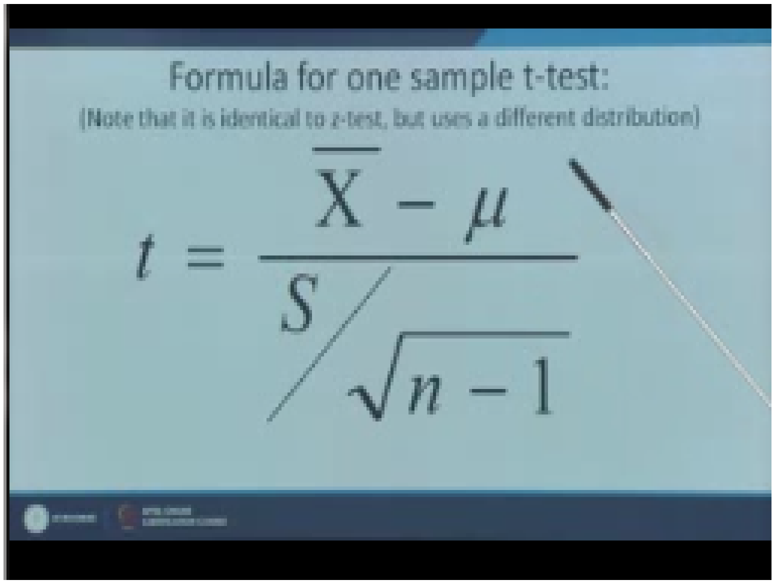
We generally use we say for a normal you know interpretation we say we except the null hypothesis but that is the wrong interpretation we should say that we fail to reject the null hypothesis so here we will say if something falls in between +1.96 and -1.96 right so we will say that the difference is not significant and it is only a matter of chance that this time it has happened okay so this is the students distribution for small samples also there.

(Refer Slide Time: 30:38)



I can show you this is what we have done right.

(Refer Slide Time: 30:42)



So this is the formula for a t test which is like which was similar to the one when you did not have the population standard deviation right so this is all this are the some other problems I have brought but I think you can do it later on right so may be in next class or something we will continue this session okay thanks for this session we will meet in the next session where we will continue with the way the t and z will just formulate.

And we get into a third condition where we have more than two right till now we have only worked with one sample we are not been able to even do the two sample and other things so we will may be continue in the next session by hope that you are clarity has been there what is the null what is the alternate and how do you check the null so this is what in further session thank you so much.

For Further Details Contact

Coordinator.Educational Technolgy Cell
Indian Institute of Technology Roorkee
Roorkee 247 667
E-Mail Etcellitrke@gmail.com etcell@itr.ernet.in
Website: www.itr.ac.in/centers/ETC. www.nptel.ac.in

Production Team
Sarath Koovery
Mohan Raj. S
Jithin. K
Pankaj saini
Graphics

Binoy. V.P
Camera
Arun. S

Online Editing
Arun.S
Video Editing
Arun.S

NPTEL Cooridinator
Prof B.K Gandhi

An Educational Technology Cell
IIT Roorkee Production