Marketing Research and Analysis-II
(Application Oriented)

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

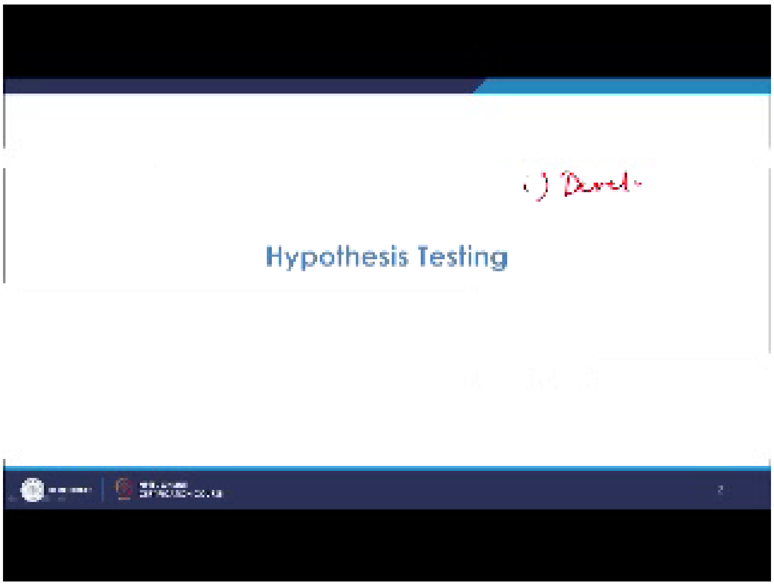
Lecture - 23
Hypothesis Testing – I
(Problem solving)

A very warm welcome to the lecture series of marketing research and analysis second part the last lecture we were discussing about confidence interval Calculation and its importance. So, we understood why confidence interval is to be understood by any researcher what is its importance? So, there we saw if you can calculate the confidence interval then you can know whether your population mean or the sample mean whether it lies within the acceptance region or not.

If it does not really within the region then we say it goes to the; it goes outside it is a part of the rejection. So, there it helps you to test the hypothesis whether it is null hypothesis to be rejected or not to be rejected in one day. Continuing the same we are today getting into hypothesis testing. The hypothesis testing is very, very important for all researchers because whoever does a research in any field of study being it Engineering, Management, Social Sciences or Medical anywhere.

Everybody is interested to test a hypothesis which is an assumption basically that they want to prove right. So, let us see what is; how do you go about testing of hypothesis. So, before we get into the first case let me tell you what are the things, we need to keep in mind while testing hypothesis. So, the first is when you test the hypothesis the first thing is to develop your hypothesis.

(Refer Slide Time: 01:52)

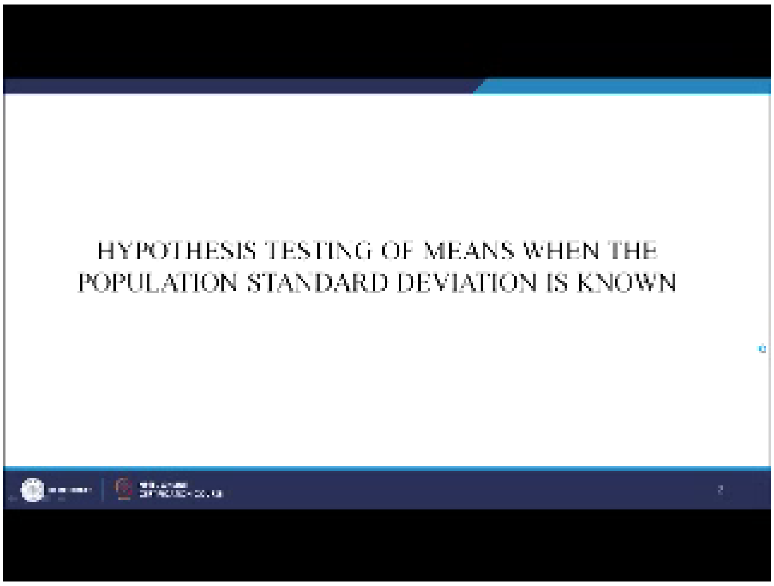


So, develop the null and the alternate ok this is the first step so once you are done with it then the next step is if you need to see the direction of the test. So, what is the direction of the test right the direction of the test we have already learnt that it can be one tailed and two tailed so it can move into any direction or only one direction maybe it is a right tailed test or it is a left tailed test ok. Once you are done with it then you decide then the researcher has to decide what, is the level of significance, at which he wants to check? he or she wants to check right.

Significance level so, this is nothing but the Alpha minus one confidence ok. So, once that is done then the researcher decides the test that it wants to use which test needs to be applied in this case particular case test that needed right test needed which test is needed. So, is it a z test it is t test in the case of you know one tailed 1 sample, 2 sample right what it is what exactly is this or that is a case of Anova whatever Chi square will see in the further classes.

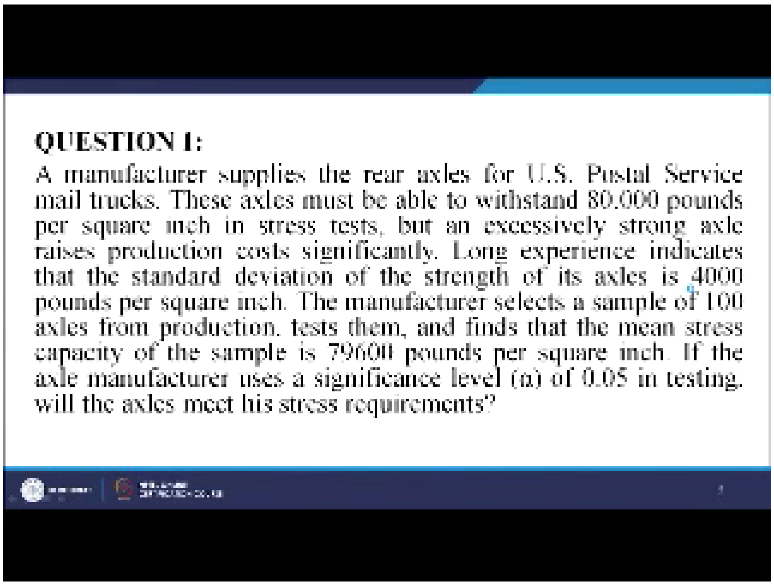
Finally you calculate the test statistic so you once to decide that is needed then you calculate the test statistic right. So, once you decide the test statistic then finally you compare the test statistic with the critical value right. So, compare the test statistic test versus the critical right and accordingly then you reject or accept the hypothesis. So, let us go to the first case the first case today we are going to discuss is hypothesis testing of means.

(Refer Slide Time: 04:00)



When the population standard deviation is known but let me tell you this is a very rare case the population standard deviation is generally unknown right. You do not you are not aware of the population so that is the population parameters are mostly unknown to you. But anyway that let us assume this is a first case and you are aware of the population standard deviation.

(Refer Slide Time: 04:29)

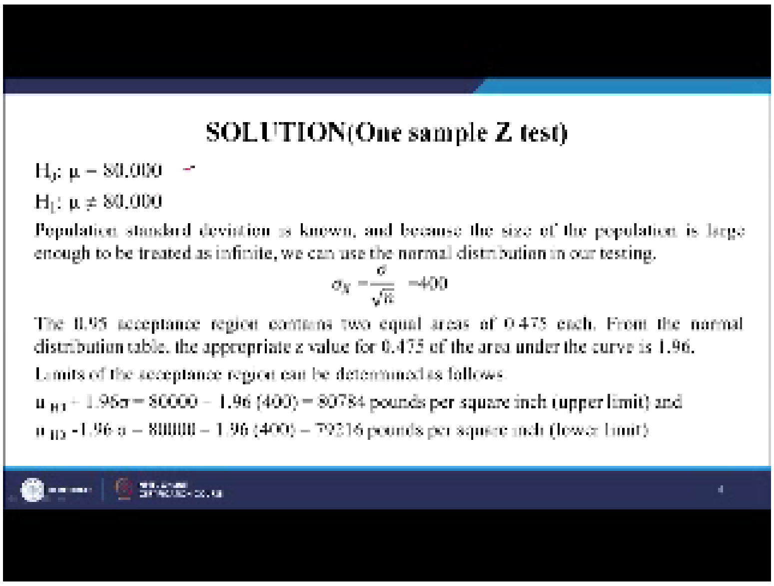


So, this case is saying that a manufacturer you can take any manufacturer who supplies rear axles for US Postal Service mail trucks. This axles must be able to withstand 80000 Pounds now what did he say the buyer things they want that the Excel should be able to withstand 80000 Pounds per square inch in stress test. But if it is excessively strong it raises the production cost ok.

Experience a long experience indicates the deviation of the strength of the axles is 4000 pounds per square inch. So, what is the deviation 4000 pounds per square inch the strength needed is 80000. The manufacturers selects the sample of 100 axles from production test them and finds that means stress of the capacity sample is 79600 so this is the sample value right sample mean per square inch.

If the axle manufacturer uses the stress level of significant level of 5% in the testing will the axle meet its stress requirement? So the case is now very clear to you in a real life situation. Let us assume it is real life situation and you are facing this case how will you attempt. First thing as I said you have to null and alternate hypothesis. So, what you do is when you are going through this exercise kindly start writing a null and alternative.Once you are written then kindly check whether we are doing the same thing or not ok.

(Refer Slide Time: 06:10)



So, solution so what is saying the $\mu_1 = 80000$ so the null hypothesis let us go back the null hypothesis let us say must be able to visit 80000 Pounds per square inch let we say; let us assume the mean = 80000 ok. So, what is alternative μ is not equal to alternative so that is what the manufacturer also is concerned if it becomes too strong then it raises the production cost. If it is too weak maybe it will not be acceptable by to the buyers.

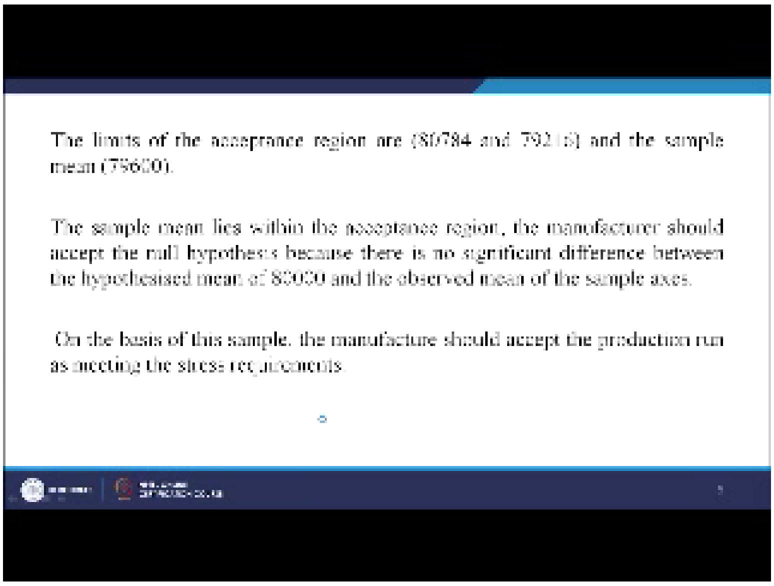
So, μ = 80000, μ is not equal to 80000 it is a clear case of a two-tailed test so that the test can move to either direction this side or this side right the value can fall either way. The population standard deviation is known in this case how much the 4000 it was told to us. So, because a sample is a large sample so we treated as an infinite population. So, we use the normal distribution had did not been infinite would have gone for a finite population check.

So, now let us calculate the standard error as you know is nothing but standard deviation upon the root over of the sample size right. So, in this case it is equal to standard deviation was 4000 / root of 10 how much was the n, 100 correct that is equal to 400. Now we checked at 5% level of significance correct so if you go back of 5%. So, 5% was spread up into two parts so that means the 95% was paid up actual the acceptance region is 95 which is also spread into two parts.

So, how much is 9.5 / 2 so that is equal to .475 this is one. So, from the normal distribution that is find the Z value for .475. So, if you look at the area under the curve for .475 so the value is 1.96. So, now 1.96 so let us calculate the first the confidence interval. So, how much is a confidence interval? So, μ is the is the mean + Z is 1.96 into the standard error so this formula you can right it like this something $\mu \pm Z\sigma_x$ ok. So, that is giving us the upper limit and lower limits so plus and minus.

So, then this value comes to us we can now see that your value of Mu is 80000. Now 80000 surely lies between somewhere here right in between these two right.

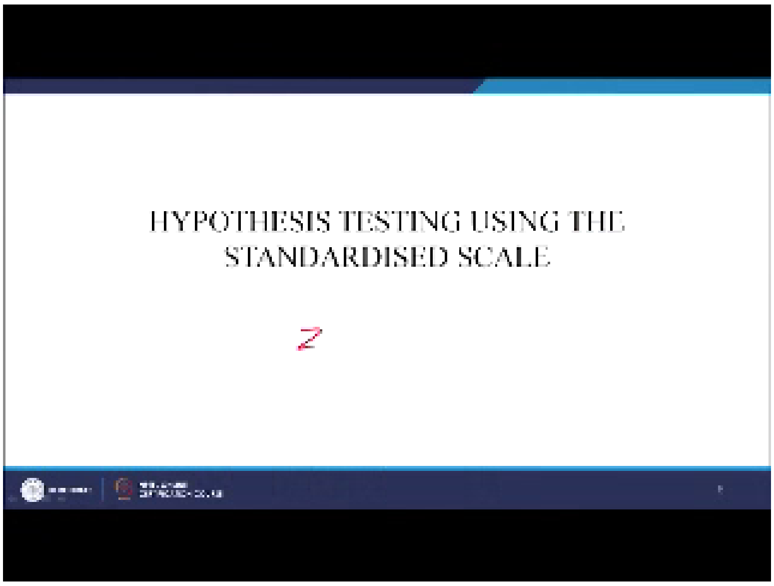
(Refer Slide Time: 08:56)



So, the limits of the acceptance region are 80784 and 79216 in the sample mean is 79600 the population was 80,000 know that. Since it; now the hypothesis mean let us say the sample mean lies within the acceptance region that means the this was 79216 one side and the other side higher upper was 80784 and they are means of the let us say 80000 here ok. The hypothesis mean and your calculated the sample mean is 79600. So, 79600 so the point is, is the 79600 in between the acceptable limits yes it is between the acceptable limit that is the 79216 and 80784.

On this basis the manufacturer should accept the production run because it cannot reject the null hypothesis. So, it is rather accept Null hypothesis in simple terms you can say and understand that a σ_{X} = 80000 that means correct. The same thing you can do in terms of a standardized now what is the standard scale.

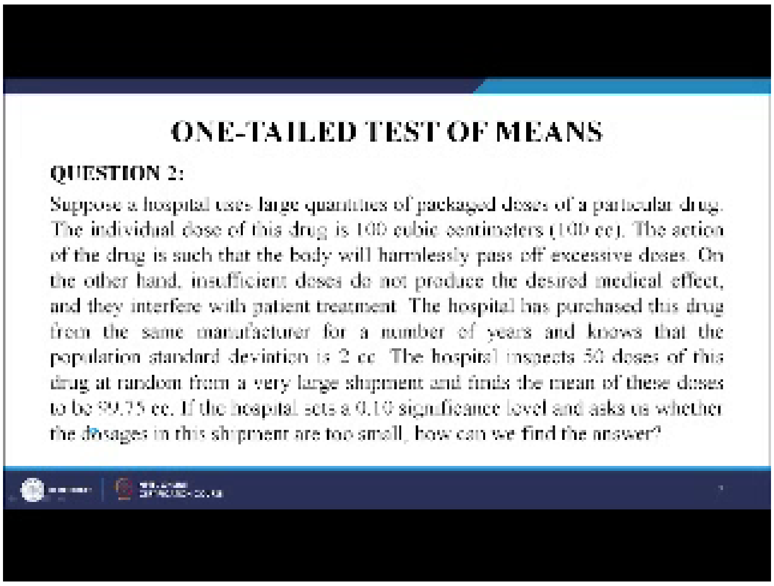
(Refer Slide Time: 10:10)



So, if you remember Z = $\overline{X}$ - μ upon Sigma X so if this Z value is nothing but the standard error in the standardized standard errors from the mean right. So, let us calculate that if you want to calculate let us see this so how much is this 79600 - was 80000 ok and divided by Sigma X was if I am not wrong 400 σ_{X} was 400 so 400 so that is equal to -1 right your Z value should lie between for a 95% confidence level we said -1.96 to + 1.96 ok. So, if you look at it your value somewhere at -1.

Now -1 will be let us say somewhere here so it is within the acceptable limit so again you do not reject null hypothesis ok.

(Refer Slide Time: 11:13)



Let us take the second case; so the second case is the case where is a one tailed test of means. now let us read this case a hospital uses large quantities of a particular drug ok some medicine understand the individual dose of this medicine or drug is 100 cubic centimetres the action the drug is such that if you take excessive it will harmlessly go out from the body. Suppose if you take excessive of drugs it as no implication that means no effect.

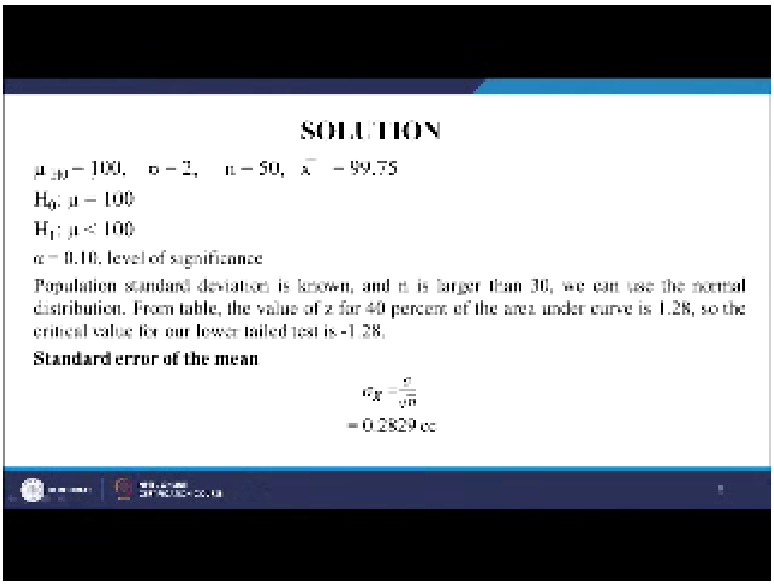
But if you take insufficient dose than it does not produce the desired medical effect rather it is interferes with the patient treatment hospital as purchased this drug from the same manufacturer for number of years and knows that the population standard deviation is 2cc that means 2 cubic centimeters. They take 50 doses of this drug at random and find the mean of dosages are 99.75 cc so which suggest that it is now coming less ok.

Then the hospital sets at 10% level of significance now if you go back and remember 10% level of significance it means manufacturer is ready to part with 10% of reject 10% of good null hypothesis right appropriate hypothesis good hypothesis this is and ask her whether the dosage in this shipment are too small because it is coming small so he is worried because if it is small then it is interested in the patient's condition.

How can we find answer? Again I would suggest please the first thing is right or null and the alternate. So, what is the null and alternative this is case is first of all is this a one tailed or two tailed test obviously the companies not worried if the dosage is more than 100 correct. So, we are not interested in anything more than 100 so we are only interested in the less than

100 case so obviously it is a one tailed test correct. So, which tail which side obvious to the left side the lower side ok?

(Refer Slide Time: 13:42)



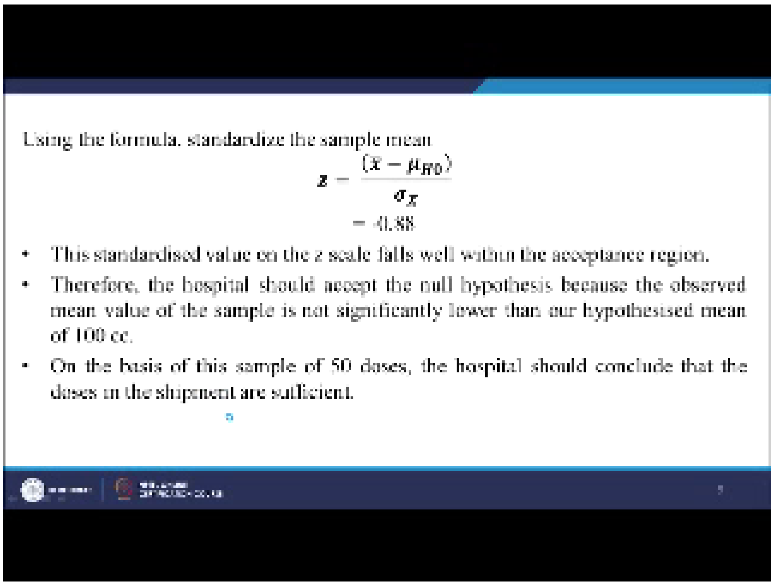
So, let us see in this case μ is equal to hypothesize mean is equal to 100 standard deviation is 2, n = 50 sample size the sample mean is 99.75 right. So, now what is the hypothesis μ = 100 what is a what is of interest is μ is less than 100 if it is less than 100 whether it is less than 100 or not that is what the researcher is interested. He is not interested on the other side. So, first you may do one thing you can just simply right the first one the one that is you are interested and automatically when you are interested in you are writing this then you go for this can be also a very nice way of writing hypothesis ok or framing the hypothesis.

Alpha is given to you so for a 10% level of significance for one tailed test you need to find out the Z value. So, what is this thing population standard deviation is known to us and n is larger than 30 correct n is larger than 30 you have taken 50 we can use the normal distribution. So, the value of Z for the 40% of the area under the curve is 1.28 now let us look at it why this is says 40%? What is the reason it says 40%? Ok let us understand this first.

Let us draw it we are not interested for this zone at all correct the right side we are interested only for the left side. So, the left side so the significance level was 10% so rejection was 10% right so what is left so this entire zone was .5 this was .5 and this was .5 correct so 10% has gone into the rejection so what is left 0.42 correct. So this 40% area under the curve this area under the curve is equal to 1.28 it says so because it is it is left tail so -1.28.

Now first we will calculate the standard error of the mean first calculate this how much it is coming Sigma upon root over of n Sigma was 2 root of 50 ok.

(Refer Slide Time: 15:59)



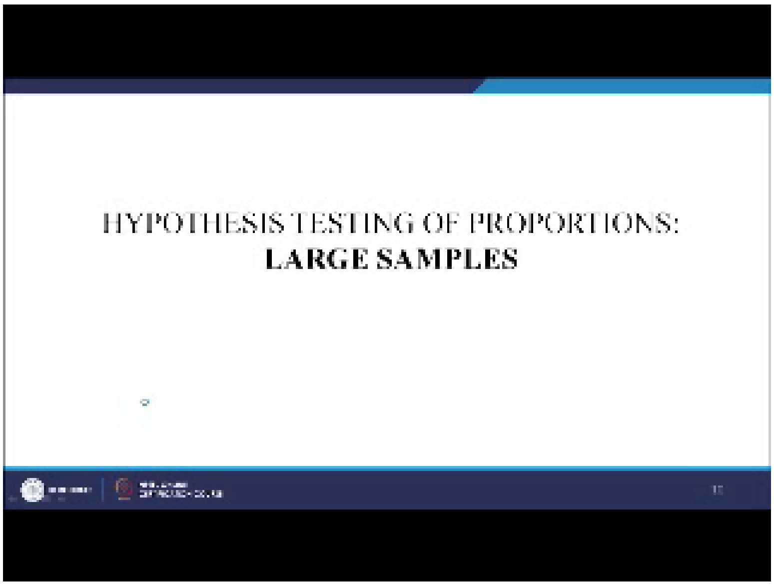
Using the formula standardized the sample mean now Z is equal to how much $\bar{x}$ - μ hypothesized divided by the standard error which is giving as a result of -0.88 so where is -0.88 what was your critical value in a -1.28 where is your estimated value -0.88 so if I see if we also it that means it is coming within the range ok so on basis of this you can say that we cannot reject the null hypothesis.

So, otherwise some will say we accept the null hypothesis. This standardised value on the Z scales will fall well within the acceptance region and therefore the hospital should accept or again I am repeating we should never say accept we should always say we should not reject actually anyway so this is normally this is used but you should write in your any paper publication or any thesis it is not rejected by the null hypothesis because the observed mean is not significantly lower than hypothesis mean of 100 cubic centimetre.

On basis of this sample of 50 doses hospitals should conclude that the doses in the shipment are sufficient right because it is not below the lower level right of the; which is there in alternate axis so this is the second case where we did. Now let us go to the third one more case you can do you can think of real life situation where you can such face such kind of situations where you are dealing with one tailed.

And let us for example as given you several examples maybe for example manufacturers try to make products and then there are worried that the quality of the product should not be too less or the in the cost of the product should not be too high what is the objective of your research accordingly you decide the null and alternative ok.

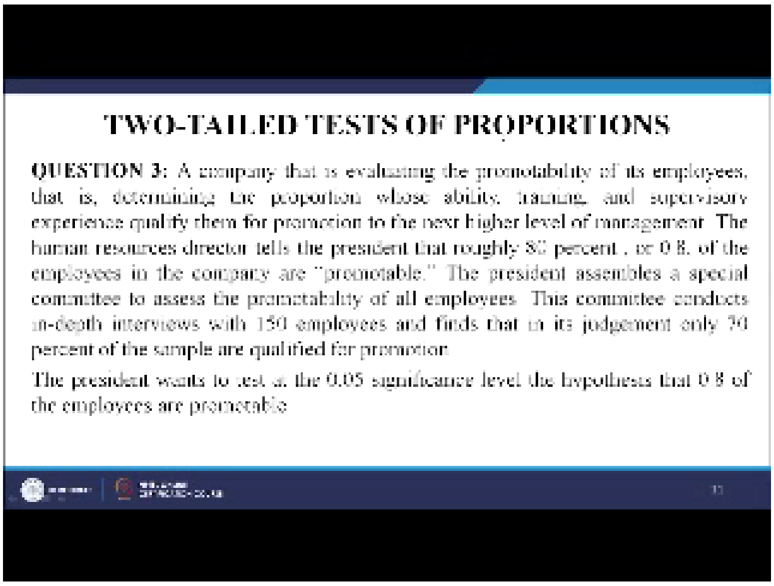
(Refer Slide Time: 17:59)



Now coming to the third, the third case is case of hypothesis testing of proportion know what are the proportions? Proportions are nothing but part of the whole so it is a part of the whole it is some significant it is the portion of the whole. So, what we are doing in this case we are dealing with large samples in this case. What do you mean by large samples? So, we have said in statistical terms anything that is beyond 30 or larger than 30 we say it goes into the category of large sample right.

So, below 30 with a small samples and we use the t-distribution large samples and greater than 30 and if it follows a Z distribution.

(Refer Slide Time: 18:45)

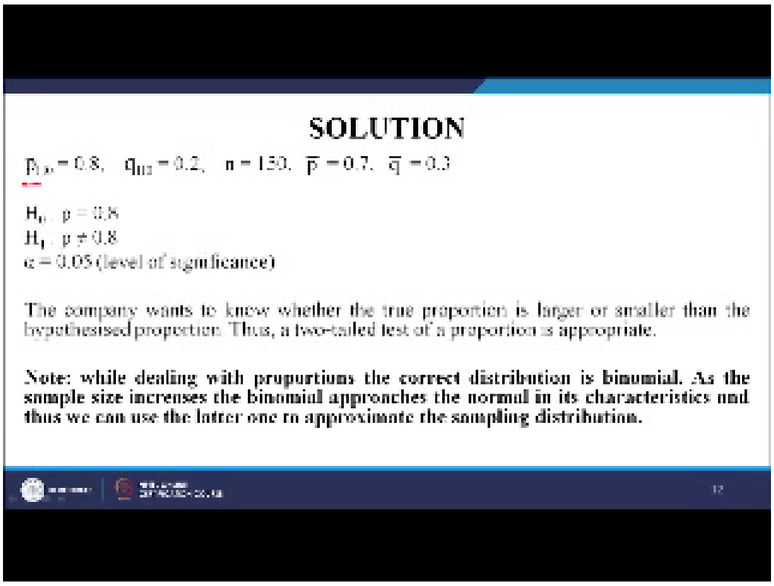


So, now the first case is of a two-tailed test of proportion Noe let us see that a company is evaluating the promotability of it employee's it wants to promotes its employees. It is trying to determine the proportion of people whose ability training and supervisory experience qualified them for promotion to the next higher level of management. The HR director tells the President that roughly 80% how much 80% of the whole total of the employees are promotable correct.

The president; see this kind of decisions will have an impact on the company. Now how will it have any impact? Because suppose it is not 80% it is 70% so the extra 10% people are unnecessary given extra money. So, the president assembles special committee to asses this promotability. The committee conducts in depth interview with 150 employees and finds that it is judgement only 70% are actually promotable.

So, out of the samples that they collect that they found that only 70% promotable and not 80% but the president does not want to take a risk he says he wants to test at a 5% significance level that the hypothesis that 80% of the employees are promotable is it true or not true. So, in such a condition how would you proceed. So, this is not a case of means it is a case of proportion it is a clear case of proportions right. So, first again I would suggest write the null and alternative.So, I am giving up pause you just you may write your null and alternative by the time right.

(Refer Slide Time: 20:36)



So, in this case what is the data? Hypothecs proportion is equal to .8 so this is a proportion of success correct the failure is automatically 1 – p. So, that is equal to q which is equal to 0.2 for this one, n = 150. Now in the case of the sample because this indicates the sample this is .7 so q = .3 what is your null hypothesis 80% of the people are promotable but is that what he is interested. If he was would have interested he would not have asked for checking he would have directly given promotion.

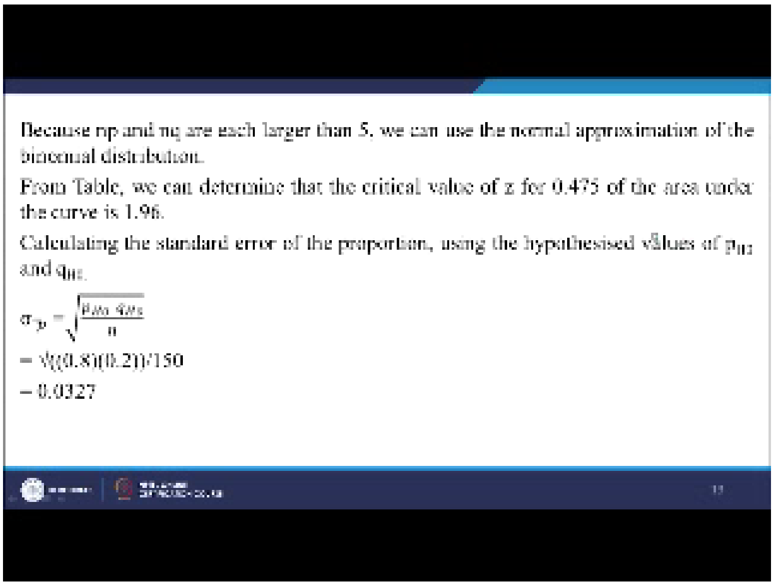
But he is not doing that because he knows that adds to the cost of the company so that is why the manager has taken the; the president has taken interest in checking this one. So, this is what is the interest to him right. So, he says p is not equal to .8 but he is not telling at this moment that p is less than 80% because if he touches that may be the staff in his company might start thinking that this president is a very negative man right.

Negative person he is forcefully thinking that we are not promotable or we are not efficient people so he does not take any risk he says well let us leave it to the study we will say it can be more than 80 also if there are more than 80% people then we will give it to everybody what is the problem or 80% out of this. Suppose 90% of the people promotable but at least 80% will give surely.

But if it is 70% but actually heart of heart that is what is interest is right α = 5% we have said. So, again alpha that will be spread up correct. Company wants know whether the true proportion is larger or smaller than the hypothesed proportion. So, thus tail of proportions is appropriate. Here you note please this is important when you deal with proportions the correct distribution is always binomial in nature binomial distribution should be used because it is a fraction it is always a part of the person it is not absolute number you are taking a fraction so that is why the Binomial Distribution is a more appropriate and as the sample size increases.

But as a sample size increases the Binomial Distribution approaches the normal distribution in its characteristics and thus we can use the normal the latter one is the normal approximate the sampling distribution correct this is very important kindly note. If yo u have small sample size then it will not approach the normal thus you will use the Binomial Distribution. And in case of Binomial Distribution for example the you know the formula was slightly different for example the mean, mean in case of a Binomial Distribution is equal to mean is given as is equal to np right. The standard deviation is given as root over of npq ok so it is changing so it is right similarly the variants obviously will becomes square of this one right just npq so this changes.

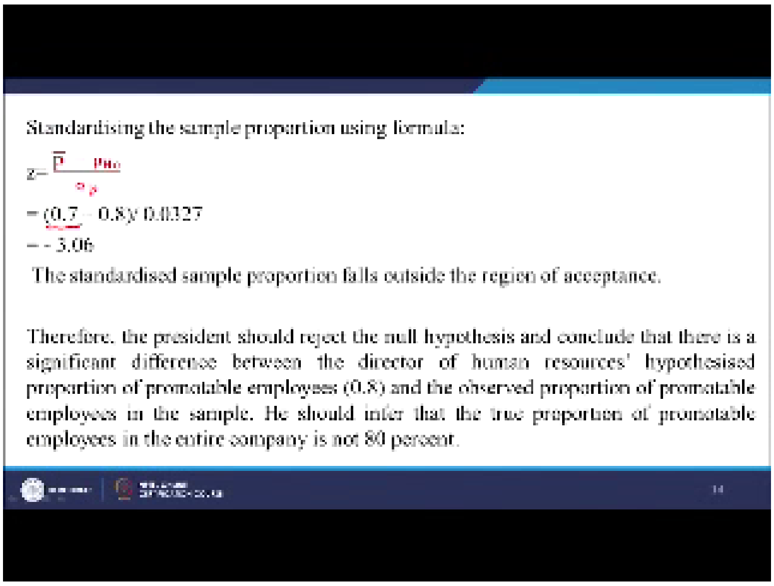
(Refer Slide Time: 23:51)



Because np and nq this is another important thing that you need to remember are larger than 5 which is taken as the minimum number we use the normal approximation of the binomial distribution. So, from table one and Z value for an area under the curve is .475 because 95% so 95% / 2.475 is the Z values 1.96. Now let us calculate the standard error first thing is you calculate the standard error of the proportion this time one tailed test.

This formula it has changed earlier when you are using standard error in terms of means we were saying standard deviation by root over of n correct but here the formula has slightly changed it has become p into q upon n right so pq by n, so that means if I say p*1- p/ n right root over. So, this is equal to earlier case and our case p was .8 the hypothesed proportion q was 0.2 / n which was 150, so 0.0327 so the standard error of proportion is .0327 right

(Refer Slide Time: 25:17)



Let us calculated Z value from this now for to do that what they will use is the sample proportion mean which is the sample proportion not means sorry sample proportion which .7 minus the population proportion right hypothesed proportion which is .8 divided by the standard error so this gives us a value of -3.06 what was the acceptable value 1.96 right 1.96 both sides 1.96 but now we are getting a value which is 3.06 right.

So, automatically we were allowing acceptance was up to this level right this was the acceptable but this is falling beside the acceptable limit. So, since it falls the proportion falls outside the region of acceptance. So, in this case the null hypothesis is rejected. So, if null hypothesis is rejected the alternative is accepted so what was the alternate the president should reject the null hypothesis and conclude that there is a significant difference between the director of human resources hypothesed proportion of .8 that means 80% and the observed proportion of promotable employees in the sample.

So, observed was 70% so there is a difference right you should infer that the true proportion of promotable employees in the inter company is not 80% it could be either way but it is at least not 80% right. So, I will stop this class here lecture here and e I will try to although in the next session also we will continue with hypothesis testing because you see it is very important that in real life you understand the how to calculate and how to interpret the values they are very important.

Nobody is going to tell us in real life what values to be taken and all. You need to be very, very clear these things and then only you can reject or accept your alternative reject the null hypothesis whatever it is or accept your null hypothesis whatever. So this decision can only be taken by the researcher when you understand this. I hope this lecture was clear if there is any doubt you can always put in your questions later on also.

But I am sure this must be clear to you and in the last so I will also explain to you how do you calculate the probability the p-value also right that was also very simple and but very important for any research to understand because many times when you use software you will not be given sum of the values. So, at least the description how they have arrived so there you might face confusion and that is a reason maybe I have seen many a times researches use statistical tools through the help of software’s.

But the clarity is not there, so this clarity has to be there and it is very, very important otherwise your understanding of the study is questionable right. So, I hope you have understood for today all the best thank you so much.