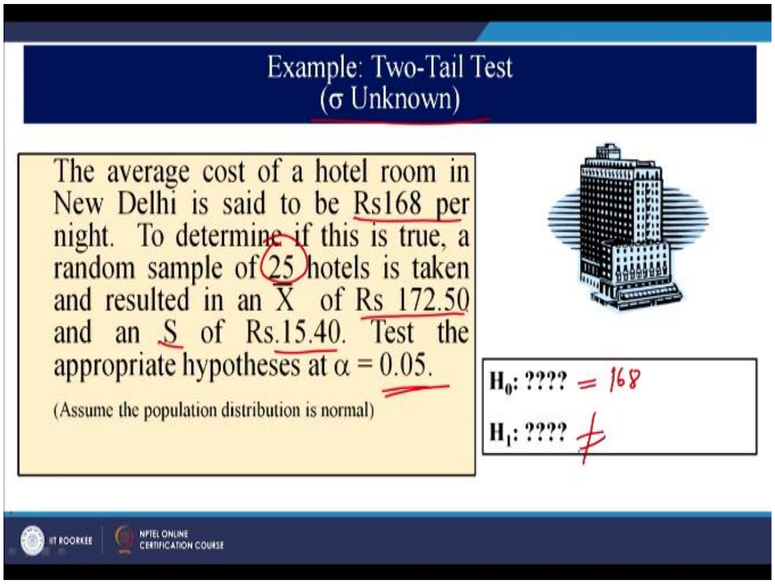
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-34
Hypothesis Testing of Proportions-I

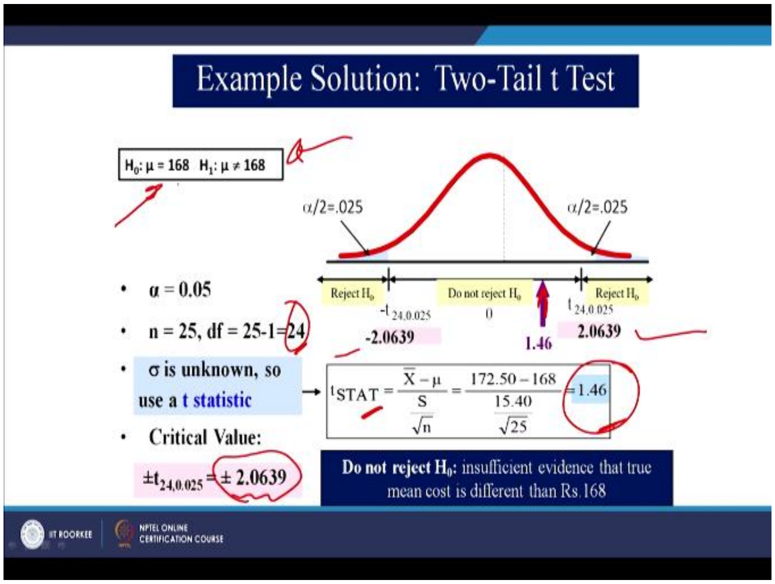
Hello friends, I welcome you all in this session, as you are aware in previous session we were discussing about hypothesis testing and we have seen several methods of testing hypothesis, one of them was critical value approach, the other one was P value approach and we have seen other approaches as well. So in previous session we were discussing about this question wherein the standard deviation of population is unknown.

(Refer Slide Time: 00:53)



So this is the case which we have seen in previous session, so average price of hotel room in Delhi is 168, when a sample of 25 hotels were taken, the average was 172.5 and sample standard division was 15.4. So we tested the hypothesis that alpha is equal to 0.05 and we framed then hypothesis that the then hypothesis that the price is 168 and alternate hypothesis is not 168 is not it. This what we have framed in previous session.

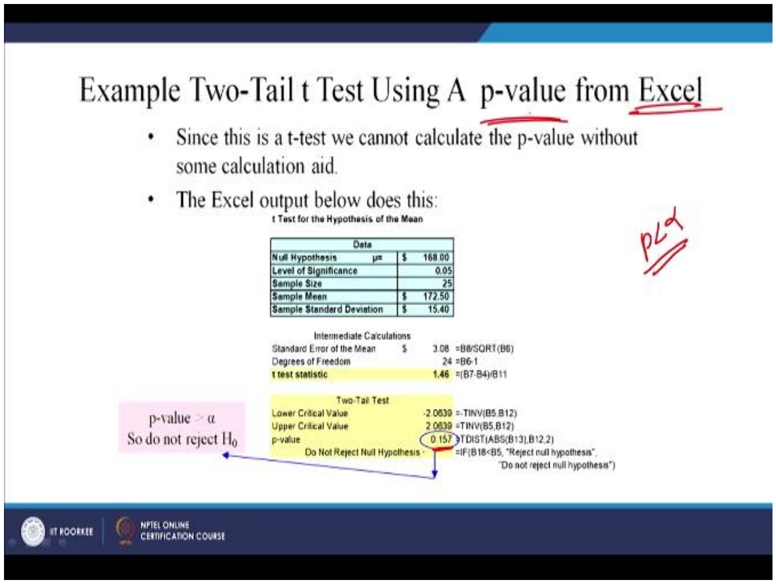
(Refer Slide Time: 01:36)



And this how we worked out, this example, 50 statistics which we have which we have got over here is 1.46 and stable value is at 24 degrees of freedom is not it, because this is a case of t-test why t-test because sample standard deviation is the population standard deviation is unknown and sample size is less than 30 right. So we are using t-distribution over here and the value of t at 24 degrees of freedom is 2.0639 right so + and - 2.0639.

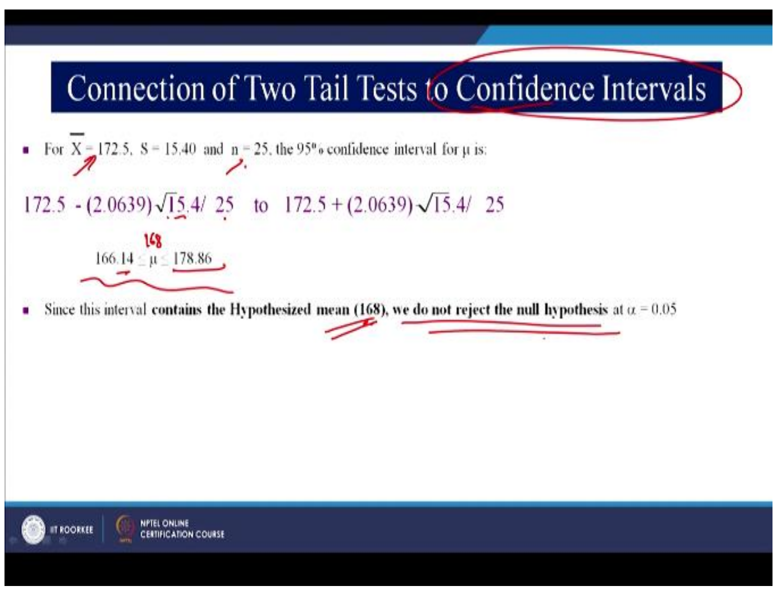
And our calculated value is in this range which means non rejection region, so that is why will not reject null hypothesis and what is our null hypothesis, let the average price per room in Delhi is 168. So will say yes this statement is true or this assumption is true right. This what we have seen in previous classes. You can find out the solution using P value approach in fact using excel sheet right.

(Refer Slide Time: 02:58)



So here P values is 0.157 which is more than alpha, so will not reject null hypothesis, we reject null hypothesis if P value is less than alpha is not it, but in this case we will use more than alpha will not reject null hypothesis.

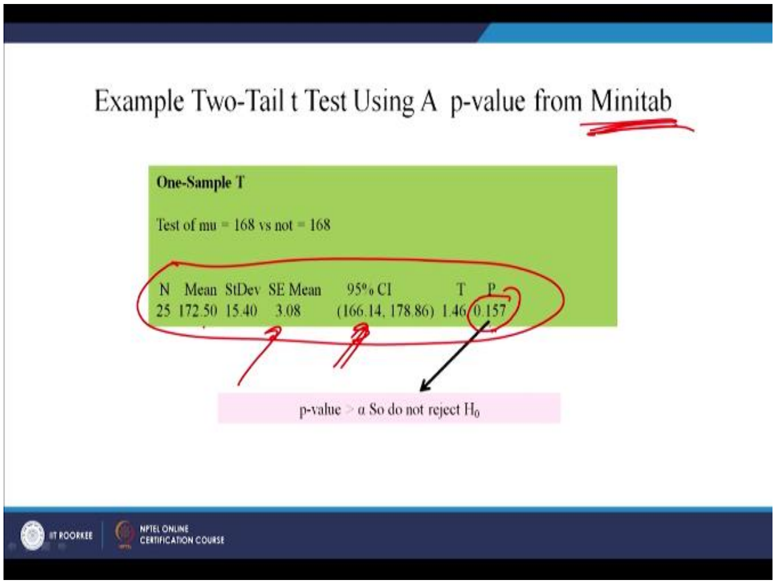
(Refer Slide Time: 03:19)



Now this one more way in which you can take a call whether you need to reject or you need not reject null hypothesis and this is known as is a confidence level or confidence interval approach. Now in this case since we know sample mean, sample standard deviation, sample size in confidence level, then we can easily calculate confidence interval. So this sample mean ± P under root of s divided by n.

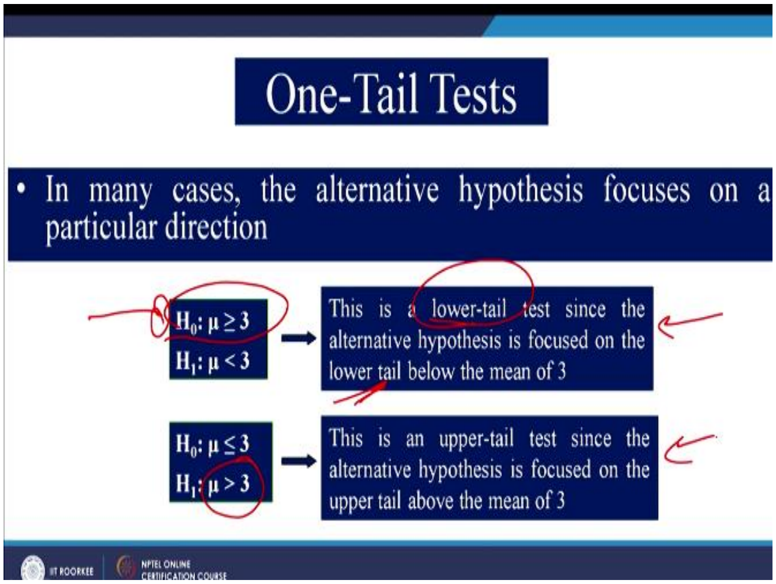
So you will get upper limit as 178.86 and lower limit as 166.14. Now you to check whether your hypothesized population mean, which is 168 rupees, does this fall in this range, no yeah it is in this range right. Since the interval contains this hypothesized mean 168 is in between these 2, is not it. So will not reject null hypothesis, otherwise we would have rejected the null hypothesis.

(Refer Slide Time: 04:47)



So there are multiple ways in which you can test any hypothesis in fact this is just for example I have taken this slide, this is an output from Minitab software for this question. So you will find all these details in output like this. So you have got sample size mean standard deviation, standard error of mean, these are our limits right 166 and 178 right, this t value and P value. So we are interesting in p value only right, so P value is less is not less than alpha in fact P value is more than alpha so we will not reject null hypothesis.

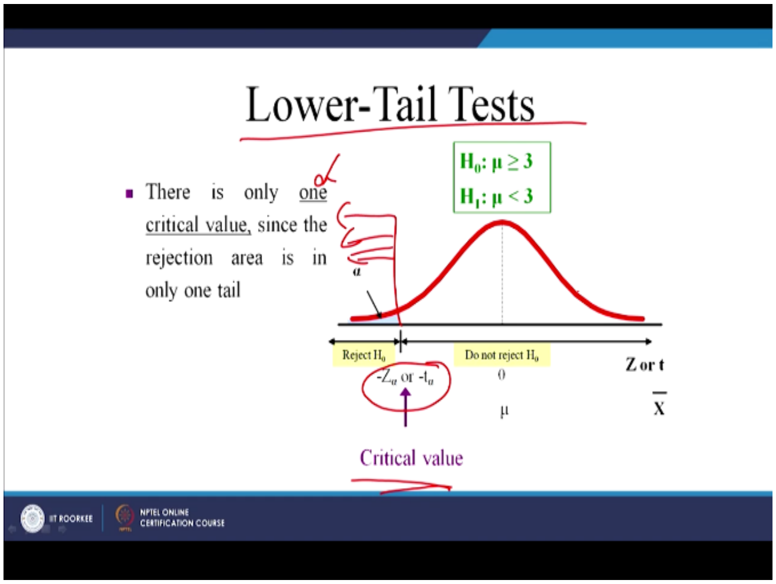
(Refer Slide Time: 05:28)



Let us look at some more points about hypothesis testing, we know that the sign of alternative hypothesis will decide whether you would be using two-tailed test or one-tailed test is not it. So let us look at this example where your null hypothesis is mean or whatever it is let say average number of TV sets in Indian homes is greater than or equal to 3 and the alternative hypothesis is less than 3.

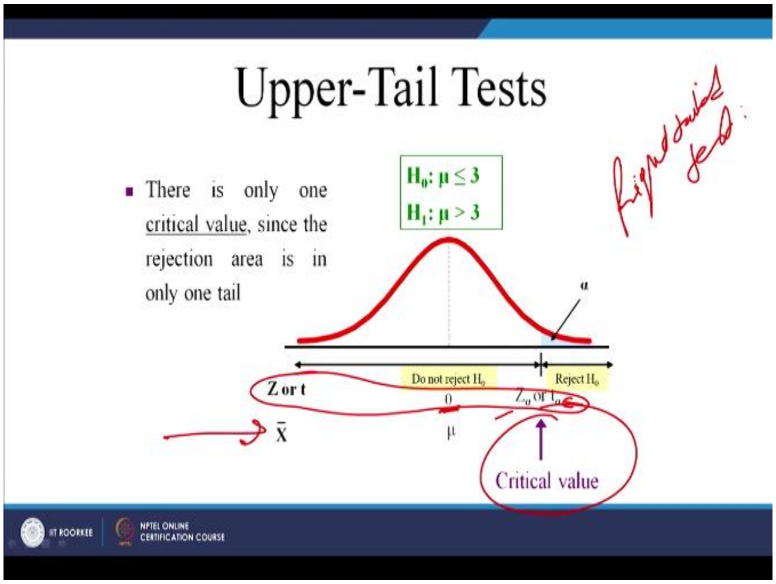
So this is the case of lower tail test or left tail test right, since an alternative hypothesis is focus on lower tail, we know the mean of 3 right, similarly the other way around right, so we alternative hypothesis is that mean is more than 3. So you can have this yes lower-tail test and upper tail test right,

(Refer Slide Time: 06:36)



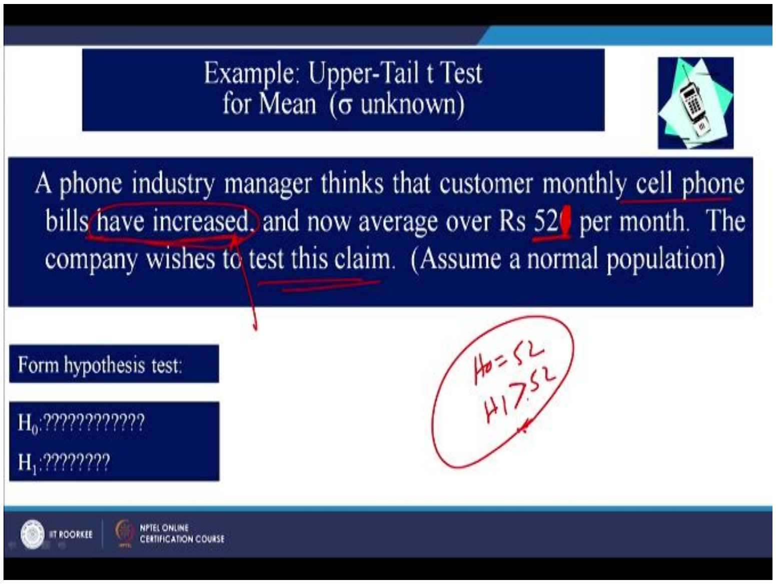
So this is lower tail test example, so this is lower tail test it means like that it is one-tail test right because there is only no rejection reason. So you need not have two critical values right you just need 1 critical value either of p or z whatever it is right. So there is only one critical right and this is your rejection region right this region, this is rejection area right and this one is non rejection area. So this is how you can have distribution for lower tail test right.

(Refer Slide Time: 07:16)



This is an upper-tail test example so again there is one critical value which is wither z value or P value from table which we will find out right and this is your mean 0 right this is your standardized skill right and this is your broad scale right, so this is a case of a upper tail test or this also known as right tail test right, right tail test is not it.

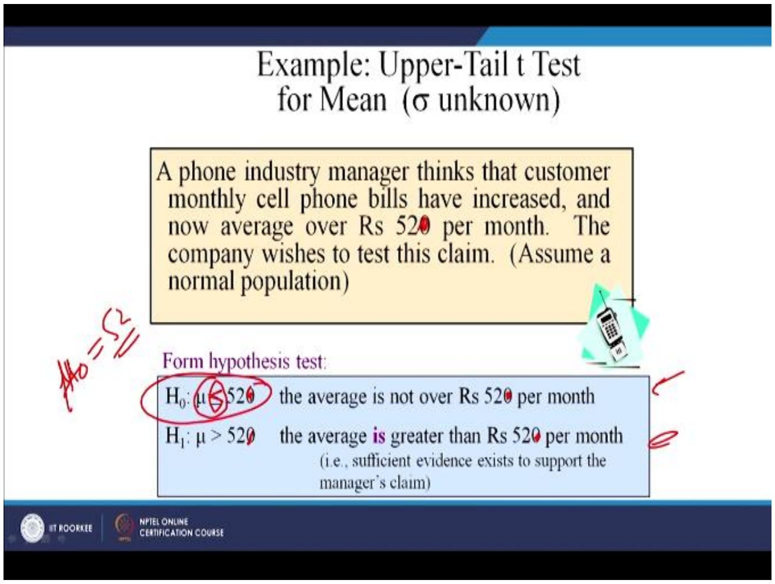
(Refer Slide Time: 07:51)



Now let us look at this question, a phone industry manager thinks that customer monthly cell phone bills have increased and now average over 52 per month, in fact this is 52 for simplicity you can just avoid 520 right. So and now, average over 52 per month. The company wishes to test claim. Now the company has worked on the average monthly bill of different customers.

And now the company thinks that average bill is no more than 52 rupees per month, so what kind of hypothesis test would be specially alternative hypothesis would be, so this is a case wherein a null hypothesis is that the average or the monthly cell phone bill is 52 rupees or alternative hypothesis test is more than 52 rupees right because it is written in the question itself, what is the question the phone industry manager thinks that customers monthly cell phone bills have increased is not it.

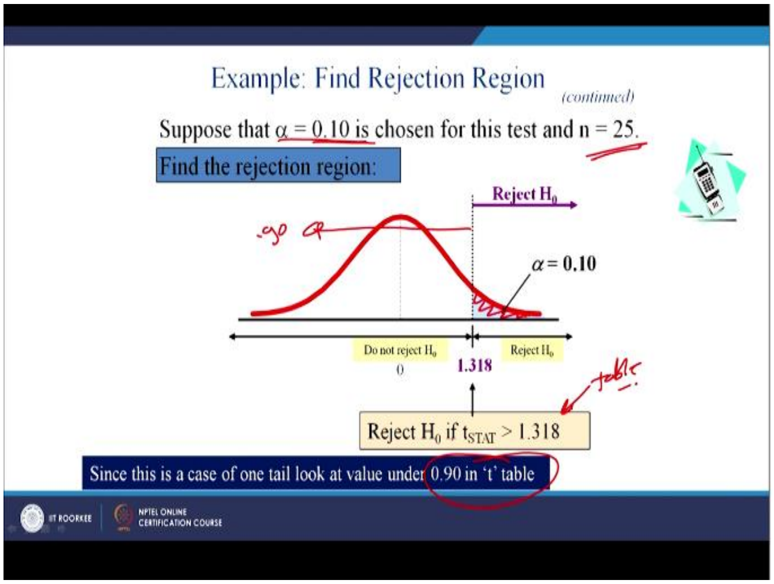
(Refer Slide Time: 09:35)



So this is this point will give you an idea about whether alternative sign would be greater than or less than right, so this is a case of an upper test or right tail test right. So in fact you have all these values are 52 right you can just keep for calculation purpose right, so this strike 52 rupees right. So you have got alternative and null hypothesis. In fact you can refine your null hypothesis because in the question what we have said that the company is trying to find out whether the average cell phone bill is more than 52 or not or it has gone up beyond 52 or not.

so though there is a possibility that the monthly phone bills would have gone down but since as I said hypothesis is a is something which researcher is trying to prove. So here the manager of the company or any other person is trying to prove his alternative hypothesis. So he is testing whether monthly bill has gone up or not. So there is no need of infact writing less than our there in fact if you have written like this.

(Refer Slide Time: 11:09)



Let us say H0 = 52 that is also equally correct right, but there is nothing around in writing less than or equal to type. So this is a case of one-tail test and upper tail test or right tail test, so alpha value is given which is 0.10 right, so this is your rejection region, this is your alpha, 0.10. So this is 0.9 centre right this entire region is 0.90 ok, sample size is 25. So we need to calculate its t statistics t value and t value is 1.318.

(Refer Slide Time: 11:57)

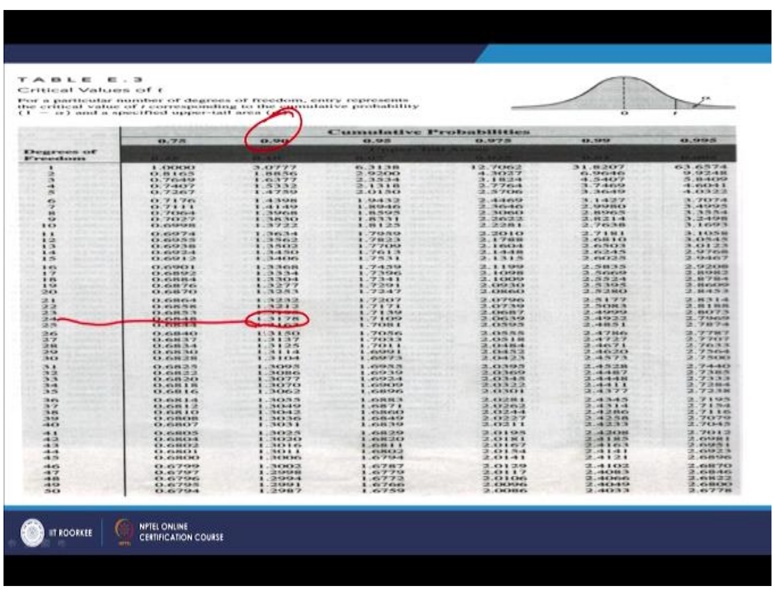
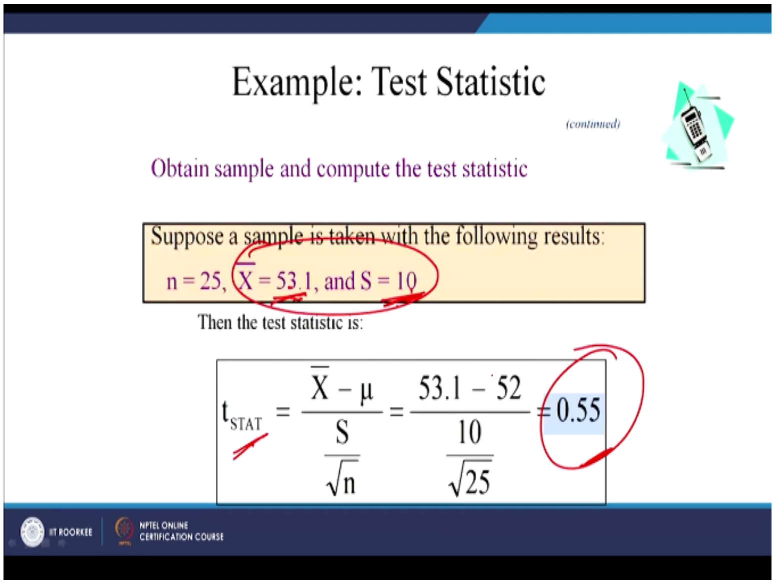


TABLE E.3
Critical Values of t
For a particular number of degrees of freedom, entry represents the critical value of t corresponding to the cumulative probability (1 − α) and a specified upper-tail area (α).

Degrees of Freedom	0.75	0.90	0.95	0.975	0.99	0.995
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	[illegible]	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	[illegible]	[illegible]	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800
50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778

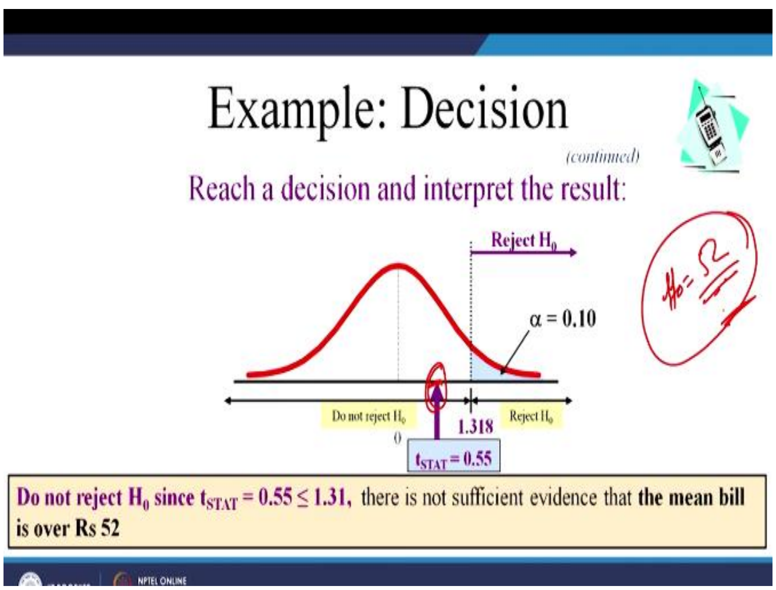
And this is table value right ok, so since this is a case of t test and one-tail test, so you need to look the value of t under 0.90 column. So this 0.90 column is this right, and let what degrees of freedom 24 is not it, this is the value, this is 1.317 right, so is the point you need to keep in mind that this is a case of one-tail test and finds 0.10. So you need to look at under column 0.90 right ok. So t value you from table is 1.3178 right.

(Refer Slide Time: 12:32)



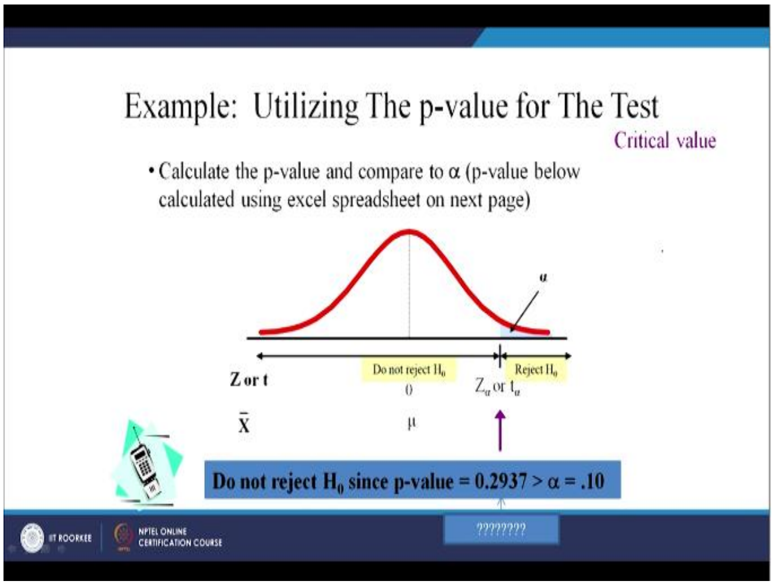
And what is our calculated value calculated value is this, since we know that the mean is sample mean is 53.1 and standard deviation is 10 right. These are of sample right, not a population right. So you can easily find out the t statistics over here which is 0.55. Now you need to compare this 0.55 with critical value right with this value right. Now if you look at this question, now this is what is your calculated value here is not it.

(Refer Slide Time: 13:10)



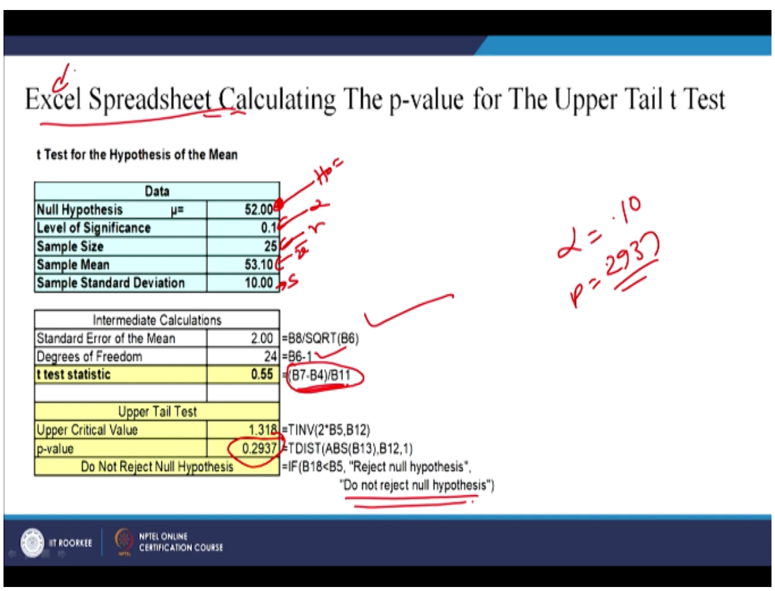
So is this in rejection region or non rejection region, so we calculated t values in non rejection region, so will not reject null hypothesis, we will not reject null hypothesis and what was our null hypothesis when the monthly cell phone bill is 52 rupees and we are not rejecting this right, it means monthly cell phone bill is still 52 rupees is not it is not increase right So this how you can conclude this particular question right.

(Refer Slide Time: 13:44)



Now you can solve this same question using P value approach as well, we have seen this kind of example earlier as well. So let us look the excel sheet calculation for this question.

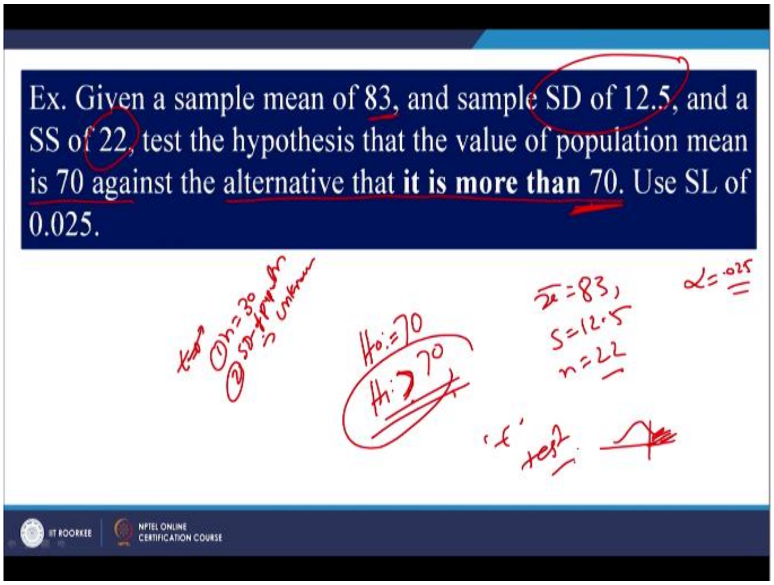
(Refer Slide Time: 14:03)



So mean of your null hypothesis this alpha is 0.10 sample size n is given sample mean is there x bar and alpha is not it, this is your null hypothesis is not it, this is your sample standard deviation S right. So just calculate first of all sample the standard error of mean then degrees of freedom and t statistics by just writing this particular formula right in excel. So this 0.55 and then you can easily calculate P value right.

So p is 0.29 alpha is 0.10, P is 0.2937 is alpha is P less than alpha no, so will not reject null hypothesis, so do not reject null hypothesis, so the same question you can solve using excel sheet as well right.

(Refer Slide Time: 15:10)



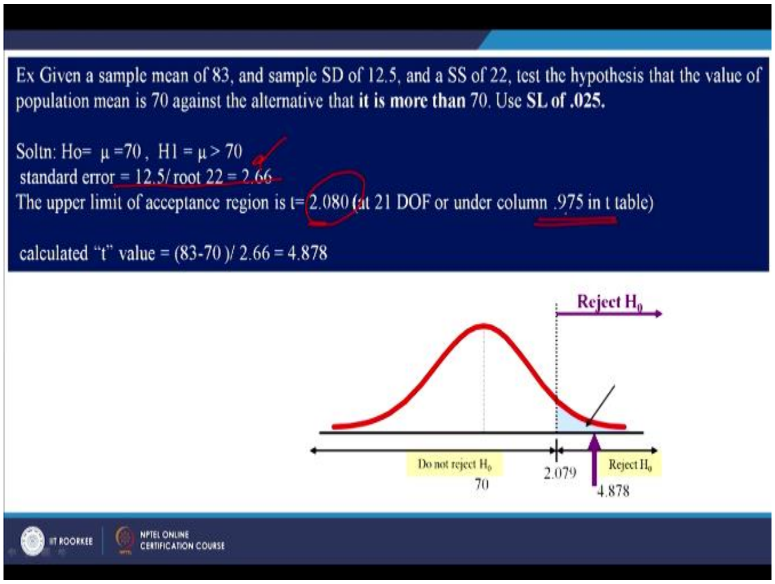
Let us look at this particular question, given a sample mean of 83, so this sample mean X bar 33, sample standard deviation is 12.5 and sample size of 22, test the hypothesis that the value of population mean is 70 against an alternative that it is more than 70, what we have to test the population mean is 70 against the alternative that it is more than 70 is alpha = 0.025. So what kind of question is this.

So you should go step by step, so the first is you need to frame null hypothesis and alternative hypothesis and you need to correctly formulate these 2 hypothesis then only you can take correct decision. So what would be the null hypothesis, this is a case wherein what null hypothesis is population mean is 70 right. So null hypothesis that mean=70 and it is now the alternative that is now more than 70.

So H1 is more than 70 is not it, so this is a case of an upper tail test is not it, because the sign is greater than type here. So this is an upper tail test like this is not it, this is your rejection region, what about would you be applying t test or Z test, just think for a while we will be applying t test or Z test in this question, when do we apply t test keep in mind, there are two conditions right, first is in sample sizes when sample size is 30 or less than 30 and when standard deviation of population is unknown.

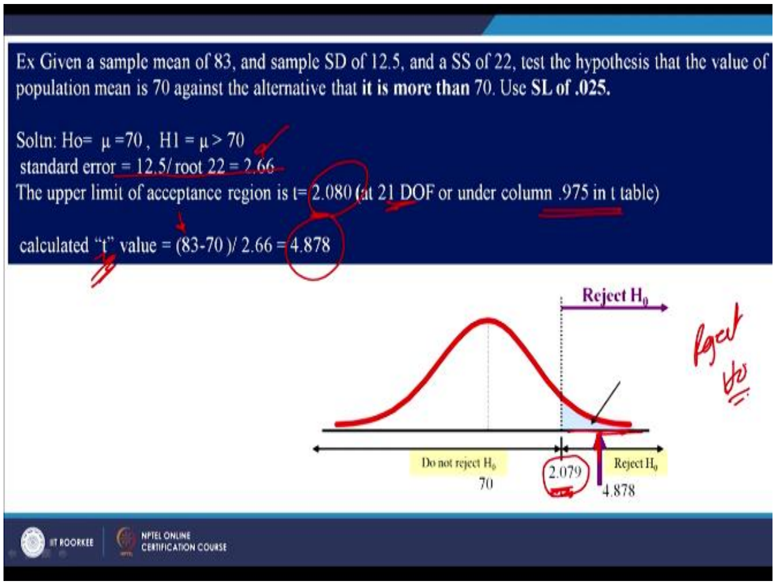
These are the conditions for applying r test, in all other situations we apply Z test right, so here you have been given sample standard deviation so population standard deviation is unknown, sample size is 20, so will be using t test over here right.

(Refer Slide Time: 18:03)



So standard error you need to calculate is s by root n, this is 2.66 since this is a case of an upper tail test, you will be requiring just one critical value, so t value from table is under column 0.975 is this is not it.

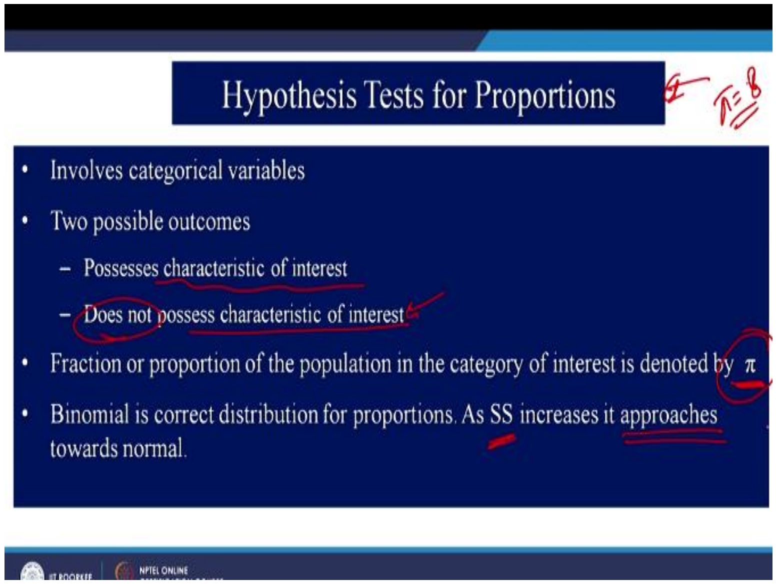
(Refer Slide Time: 18:29)



Earlier what under which column we saw the t value in previous case, it was under 0.90 column, because alpha was 0.10 here alpha is 0.025 ok. So, at what degrees of freedom right 21 degrees of freedom let us look at this 0.975, 22 degrees of freedom which is this 2.079 right and this is our table value right, critical table value and what is our calculated P value this 4.87 right. So which is here in this rejection region.

So will reject null hypothesis is not it, what is null hypothesis that these the null hypothesis that the population mean is 70. So we say that population mean is not 70 it is more than 70 why more than 70 because the sample mean is 83 right. So this how you can work out questions on one tail test.

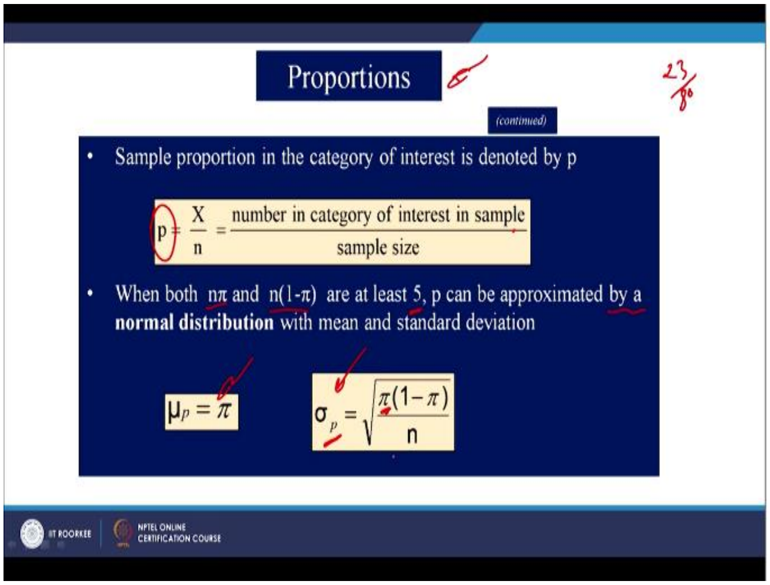
(Refer Slide Time: 20:03)



Let us look at hypothesis testing for proportions. So for you have seen population mean, population let us say variance and let us look at hypothesis testing for population proportion, in fact when we talk about population proportion, in fact there are two possible outcomes. So we are interested in the characteristics of a particular phenomenon and does not possess characteristic of interest.

So you will have let us say probability of left handed in class is 40%, the other one would be 60% is not it or vice versa right. So fraction or proportion of the population category of interest is denoted by pi, so here will say let us say in a class 60% students are right handed. So is it pi=0.6 is not it or 80%, so pi=0.8 is not it. So binomial distribution is the correct distribution for this type of proportion.

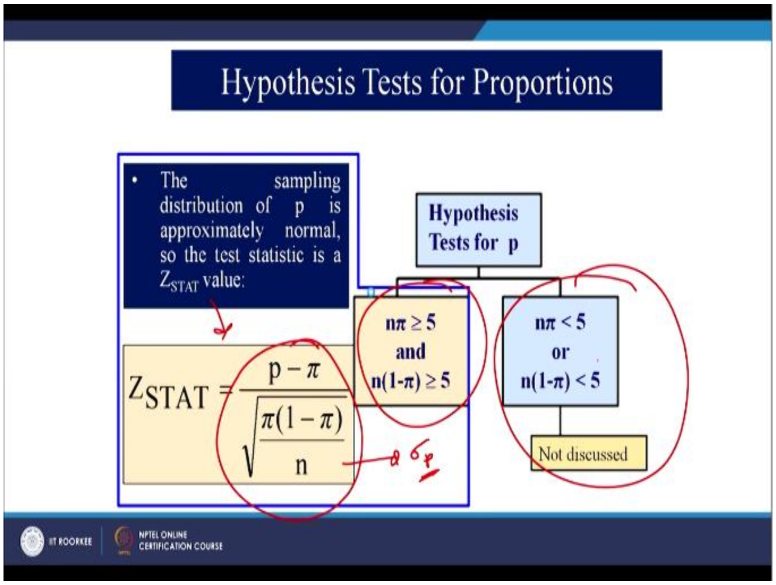
(Refer Slide Time: 21:41)



So we also know the test some increase sample size it appears towards normal distribution or the distribution approaches towards normal distribution as we increase sample size. When you look at this proportion, we need to calculate first of all sample proportionate right, so this X/n number of categories of interest in the sample divided by the sample size, let us say if I say in a class of a P 23 or left is right.

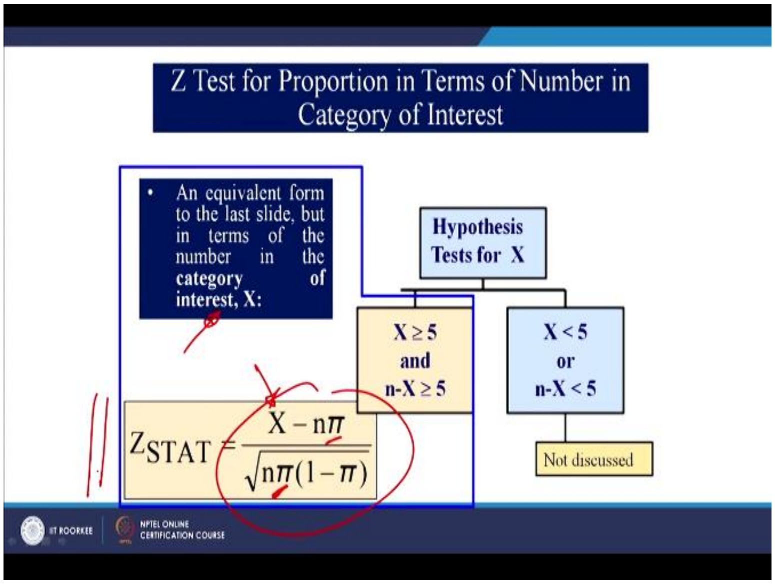
So what it would be 23/80 is not, so that will be P value when both nπ and n*1-π are at least 5 P can be approximated by normal distribution with mean this and standard division this right. So this mean value and this is standard deviation of proportion right, we have already said what P is, what is pi here pi is population proportion right, n is sample size.

(Refer Slide Time: 22:39)



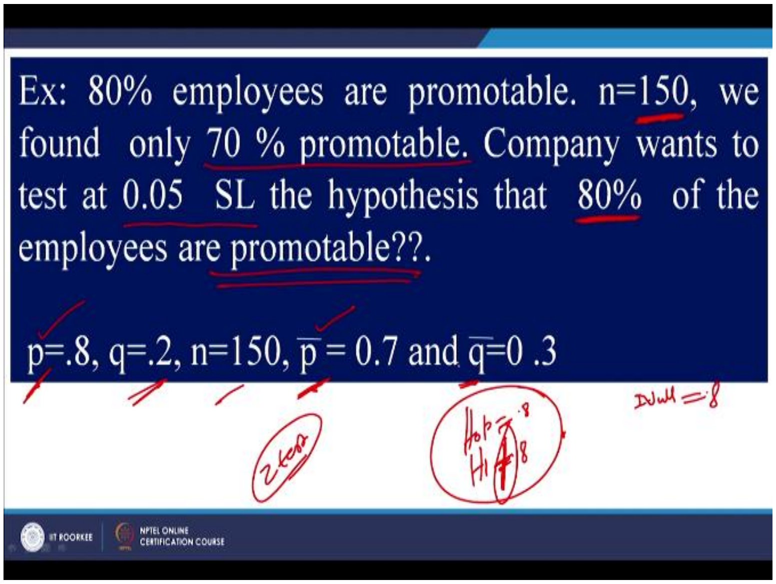
Let us look at the formula for this how to find out t statistics or Z statistics. So this is how you should be calculating z statistics P-π divided by this nothing but standard deviation of proportion mean same formula which is there in previous slide is not it. So P-π/standard deviation of proportion ok. So this is the test statistic for proportion. Now will not be discussing this type of hypothesis where in pi is less than 5 or 1 – nπ is less than pi right. So these are different cases. So we will look at only this particular case ok.

(Refer Slide Time: 23:43)



In fact an equivalent form of the previous slide in terms of X variable X is this instead of P we are writing X and instead of π we are adding nπ right. So in fact this not of much use for you, but you should be using this particular formula for solving questions.

(Refer Slide Time: 24:18)



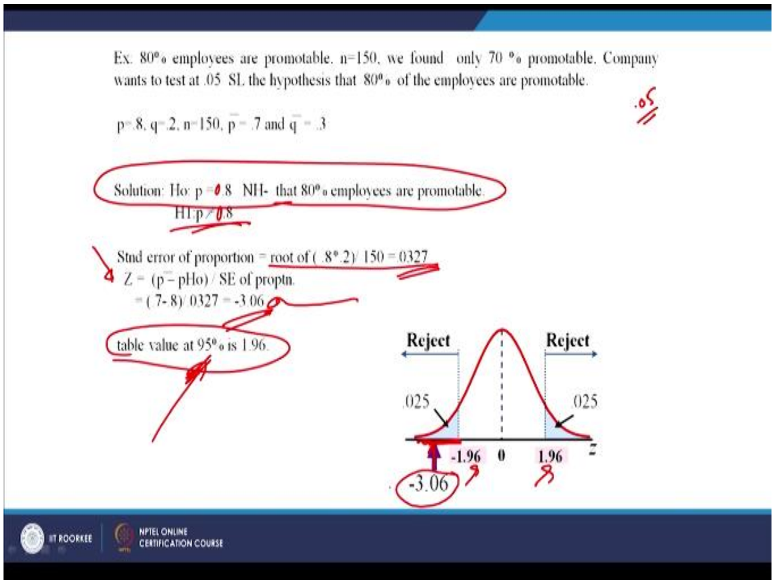
Let us say there is a manager in a company and he is claiming that 80% of employees are promotable in his company. So you want to test whether his claim is correct or not. So what you have done we just took a sample of 150 employees and we found that only 70% of them are promotable. So company wants to test it 0.05 significance level, the hypothesis this is that 80% of the employees are promotable.

So first of all what you should do, we need to frame null hypothesis and alternative hypothesis. The first is what is null hypothesis, how to frame null hypothesis. So what we have to test company wants to test at this significance level that I would I did 80% of the employees are promotable. So then I will that 80% employees are promotable right. So p=0.8, and this is your null hypothesis and alternative hypothesis is not equal 0.8 right.

So this is a case of one tailed test or two-tail test you need to answer couple of questions before solving this question. So is this a case of two-tail test or one-tail test, just look at this symbol carefully, this is not equal to type so this is case of two-tail test right, so there would be two rejection regions right and what about test statistics, we would be using t table or Z table. I have told you in previous question when to use t test.

Here sample size is 150 right and there is nothing like sample standard deviation, so we will use Z test over here right, because for applying t-test we need 2 conditions and those conditions are not there in the question itself. So will use Z test right. So P is this 0.8 or you can caller it pi as well right. So $\pi = 0.8$, this is $1-\pi$ is -0.2 and this P bar right, this sample proportion right, this Q bar right.

(Refer Slide Time: 27:24)



So since you know P and P bar are π and P bar you can easily calculate q and q bar right, so how to proceed for this question, so this is your null hypothesis right, that 80% employees are promotable, and they are not promotable in this, this is 0.8, 0.8 right, standard error of proportion just calculate first this is under root of π into $1-\pi/n$ is not it, so this 0.0327 right. So when you calculate Z value it comes out to be -3.06.

And table value will be what 0.05 significance level, the table value of Z or Z value would be 1.96 in of course ±1.96. So you need to compare your calculated Z value with table Z value which is in rejection region. So whatever the value calculated is now in rejection region, so you reject the null hypothesis is not it. So rejection of null hypothesis means the employees 80% employees are promotable.

No rejected that means either more than that percentage or employee are less than that percentage are promotable. So in this session we have worked out couple of examples on hypothesis testing on you know one-tail test and two-tail test and we have seen hypothesis testing of proportion as well, so thank you very much.