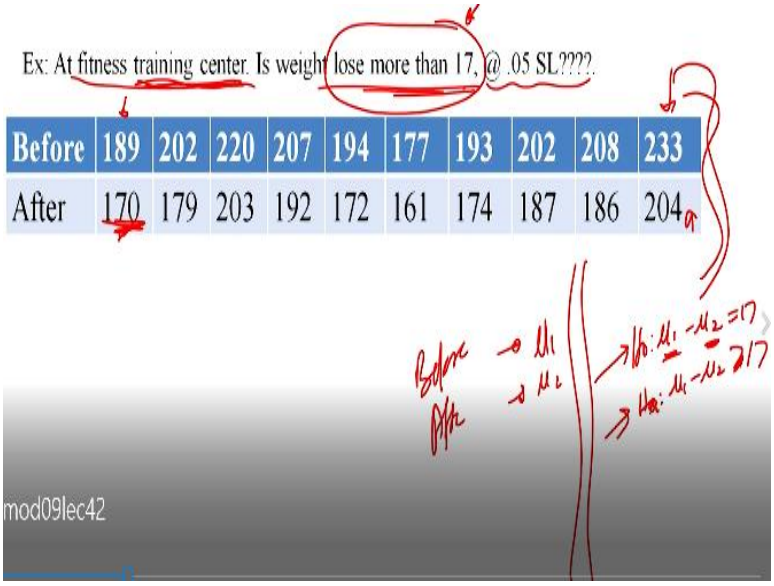
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology- Roorkee

Lecture – 42
Paired Sample Test

Hello friends, I welcome you all in the session as you are aware in previous session, we were discussing about hypothesis testing of few samples and we have worked out a couple of questions on 2 independent samples and we have worked out a question on dependent sample as we will. So you should know what is the meaning of dependent samples a depending on sample is a sample where in you collect data from same sample 2 times.

In previous class, the example which we took we have seen there that we collected the number of complaints data from a salesperson before they went for training after they attended training program is it not and we found that there was no significant difference before and after. So the training program was not an effective training program for those group of sales persons right Let us look at this question which is again an example.

(Refer Slide Time: 01:42)



Paired sample or dependent sample as you may call it variated samples. so at a fitness training centre the manager of the fitness training centre he claims that if you join his fitness centre then weight loss would be more than 17 kg or 17 pounds or whatever is the unit so you need to this

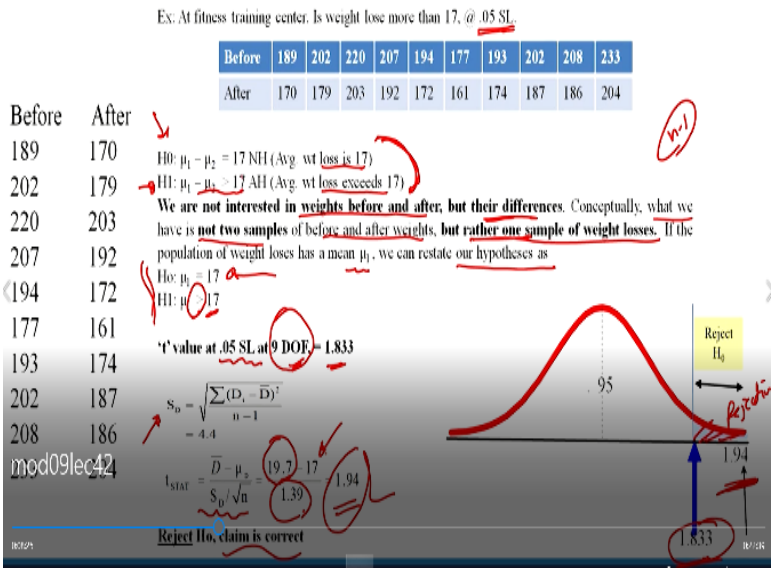
hypothesis at 0.05 significance level. Now he is saying that if you join our training centre then weight loss would be more than 17 kg.

So since you are a student of statistics you have got this data with you and the trainer has given you this set of data and he is saying that weight loss is more than 17 kg right so these are the weights in terms of pounds of different people let us say there is one fellow whose weight was 189 pounds and then he is after attending you know fitness training centre few classes over there the weight loss is there right or the weight is 170 pounds and so on right.

So 233 to 204, now we have to check is weight loss more than 17 ponds. so how would you frame in a hypothesis okay. So there are 2 samples so let us say me μ1 is the weight before right before going for training centre and this after μ2 right weight training centre you have attended sessions in training centre right and the weight loss is more than 17 kg. So how would you write it so first is this and let us you can have this null hypothesis and alternative hypothesis in two ways.

The first is null hypothesis μ1-μ2 so we are saying mu1 is more weightage mu2 is reduced weightage. So this is =17 kg or 17 pounds for example so μ1-μ2 =17 pounds and what is the claim it is more than 17. So alternate hypothesis is μ1-μ2 > 17 pounds so this is the claim and null and alternative hypothesis this Ha alternative hypothesis.

(Refer Slide Time: 04:56)



Let us look at this so this is how you can claim null and alternative hypothesis right. So average weight loss is 17 kg it is more than 17 kg now if you look at this question carefully we are actually not interested infact weight before and after but we are interested in difference right there is the point we try to prove in a dependent samples is there any difference before and after or not right.

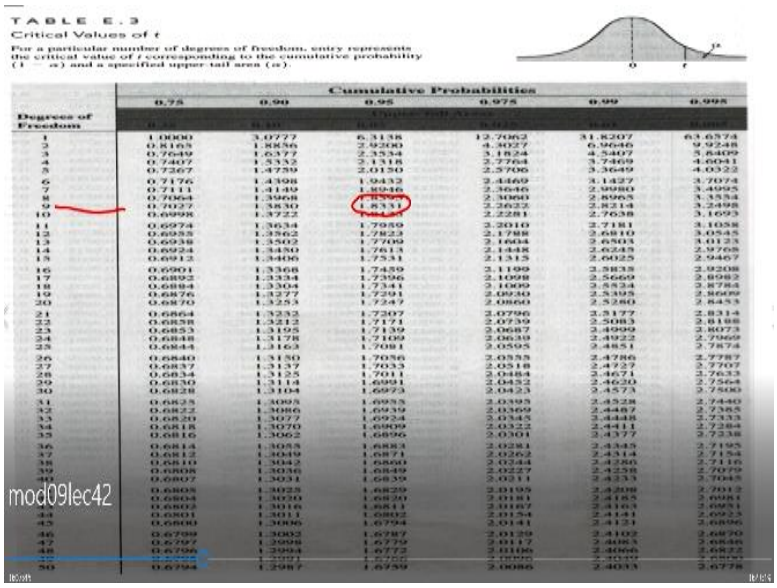
So conceptually what we have is not 2 samples of before and after which but we will say there is only one sample of weight loss and is that weight loss more than 17 so we can rewrite and these hypotheses like this. So if the population of weight loss has mean this we can restate our hypothesis as mean weight loss is 17 and when mean weight loss has gone up and as it is more than 17 right.

So this is a case of an upper tail test this is a right ail test right so this rejection region this is the rejection region okay. So we need to apply this sample standard deviation so D bar is this is 19.7 D bar is nothing but difference of all this and the average the average of differences right - μD this is what is your hypothesized mean we just 17 this is $S_D/\sqrt{n}$ so 1.39 so calculated t value is 1.90.

So this is how you should be calculating t value so t value is 1.90 now you need to since you need to test this hypothesis as 0.5 significance level so the value of t is to be looked at T table so

at0.5 significance level and at 9 degrees of freedom. Why 9 1 2 3 4 5 6 7 8 9 10 so this is n-1 because we have at the end of the day say said that this is just one sample and we are looking whether weight loss is 17 this is more than 17 right. So this is at 9 degrees of freedom so let us look at T value at 9 degrees of freedom.

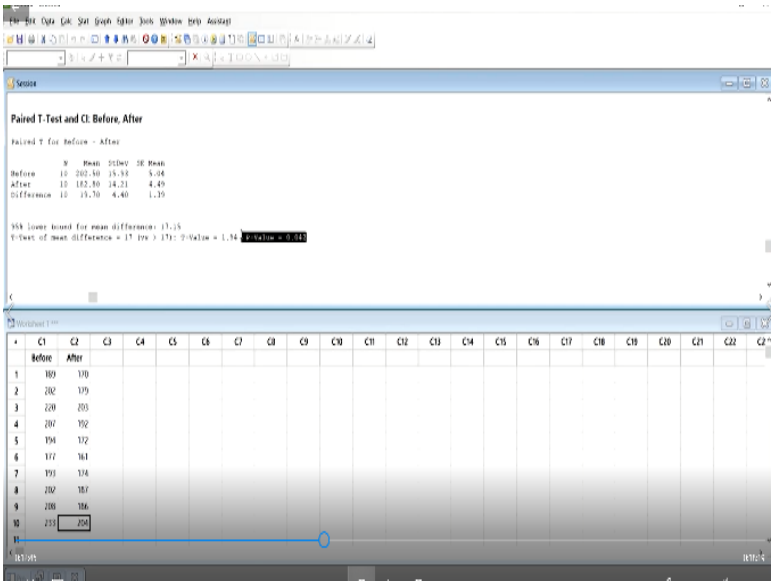
(Refer Slide Time: 07:38)



So this is 9 degrees of freedom 1.83 this okay. so this is your 1.83 this is we are solving this question using critical value approach method. So the critical value is 1.83 but the calculated t value is in rejection region 1.94 so we reject the null hypothesis right. we will reject null hypothesis and we say that the claim of the trainer is correct because when we reject null hypothesis we accept alternative hypothesis.

And we are saying that the weight loss is no more than 17 pounds right so this is how you can solve the equation like this and we can solve this example using Minitab as we will. So let us work out how to solve this question using Minitab.

(Refer Slide Time: 08:47)



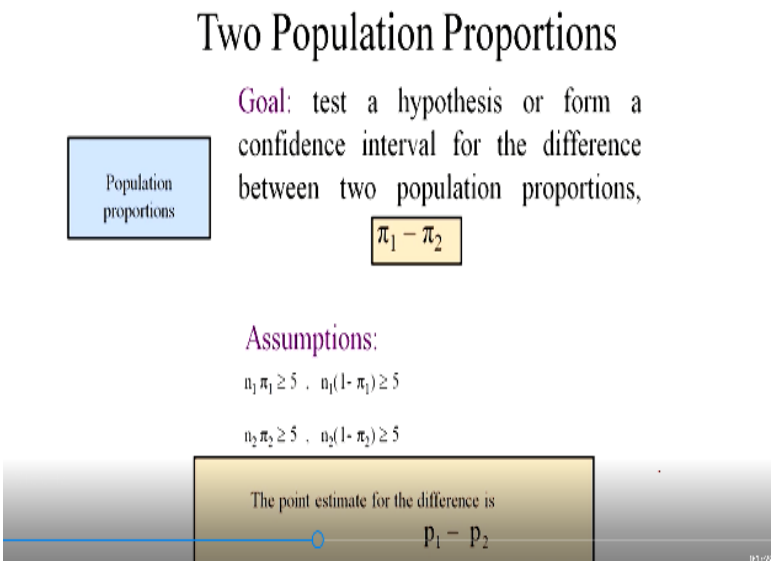
So you have got Minitab software so this part is worksheet and this is your output area so we will have data entry area first. So before weight loss so this date before weight loss or after weight loss after attending fitness sessions weight. So let us look at this so we have got 189 202 220 189 202 220 then we have got 207 194 177 207 194 194 177 then next data set is 193 202 208 193 202 208 and finally 233 till data constant.

Let us look at weight after a group of people attended training sessions right at any centre right at a fitness centre. So after data are this so 170 179 203 170 179 203 right. 192 172 161 192 172 161 174 187 186 174 187 186 and finally 204. So let us work out this question and its a paired sample test so we will go to stat first basic statistics we will go to paired t test right. Now we have got these 2 columns over here so sample 1 is the sample before weight loss so this is how you should select samples then for sample 2 which is after weight loss.

Now go to options now let us check at what significance we have to check this this is at it 95% significance level hypothesize different now this point is to be seen clearly so what we have said that since we are taking these 2 samples so we will say that the hypothesized difference is 17 right. So this is 17 and alternative hypothesis is it is more than 17 is it not this what we have said in fact we can solve this example and just once as a simple t test as well right.

So we will click at okay again okay let us look at output so p value is 0.042 and T value is 1.94 let us check so T value is just what we have got over here 1.94 which you can see here as we have calculated T value P value is 0.042. So P value is 0.042 alpha is 0.05. so P is <alpha, we will reject the null hypothesis P is < alpha we will reject null hypothesis okay so this how you can work out this question using Minitab.

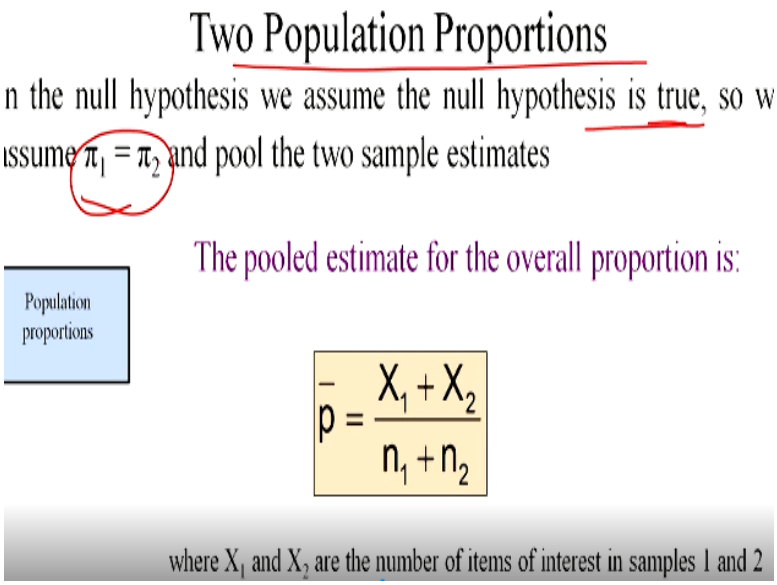
(Refer Slide Time: 13:30)



Let us look at a next topic which is hypothesis testing of 2 population proportion. So far we have seen means of dependent samples in different samples let us compare population proportion of 2 groups or more than 2 groups. So first we will look at a population proportion of 2 groups right so test hypothesis or a form of confidence interval for the difference between 2 population proportions right.

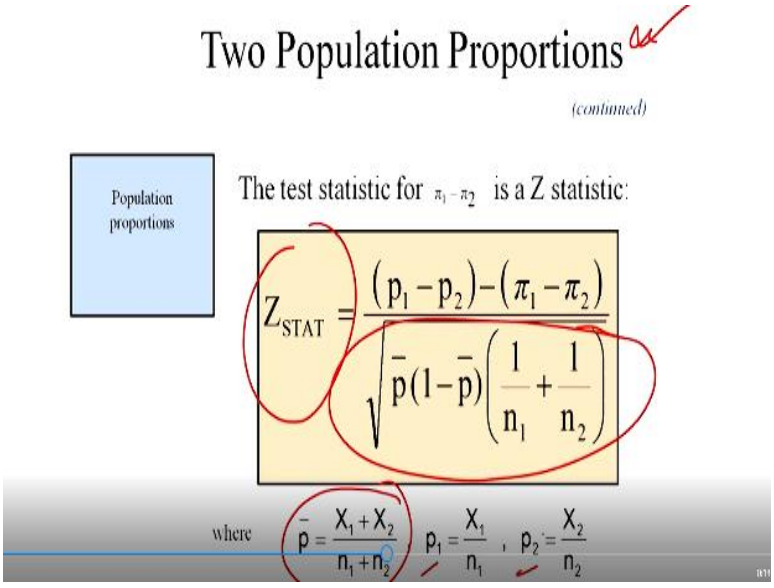
So just like μ1 in previous case in previous example we have seen difference between mu1 and mu2, so we can say here difference between pi1 and pi2 right so these are population proportions of 2 different populations right and of course assumptions will remain same nπ * 1- π both have to be >=5 and we will use this point estimate for the difference which is this we will use this to hypothesized population proportions.

(Refer Slide Time: 14:48)



So these are this is the case of 2 population proportion then in hypothesis we assume that null hypothesis is true o we assume that mu1 sorry π1 and π2 are equal right. So this will be a null hypothesis and this is in regard will calculate pool estimation for the overall proportion so this p bar its called pooled estimate of overall proportion which is nothing but the sum of values in sample 1 and sample 2 and these are sample sizes of these 2 samples right so these are number of items of interest in these 2 samples right.

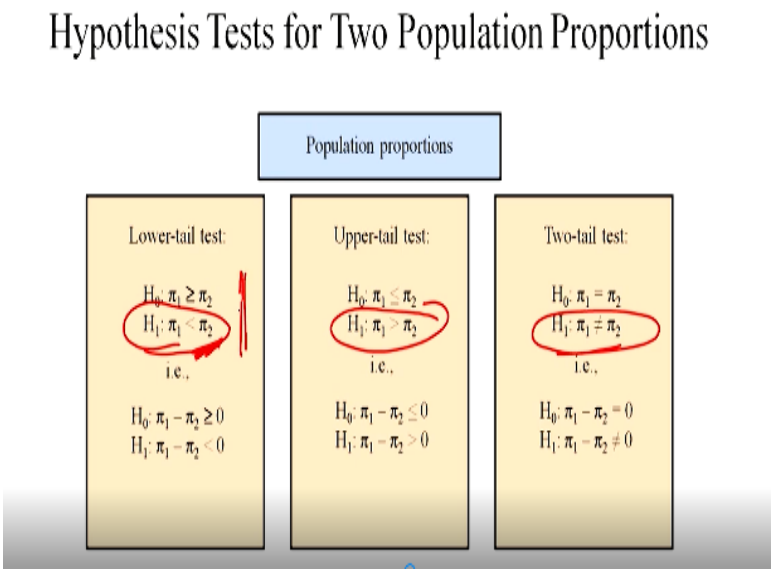
(Refer Slide Time: 15:39)



This is how you should be calculating Z statistics when you calculate hypothesis proportion of 2 samples right so this is almost similar to what we have seen earlier only the differences is this is

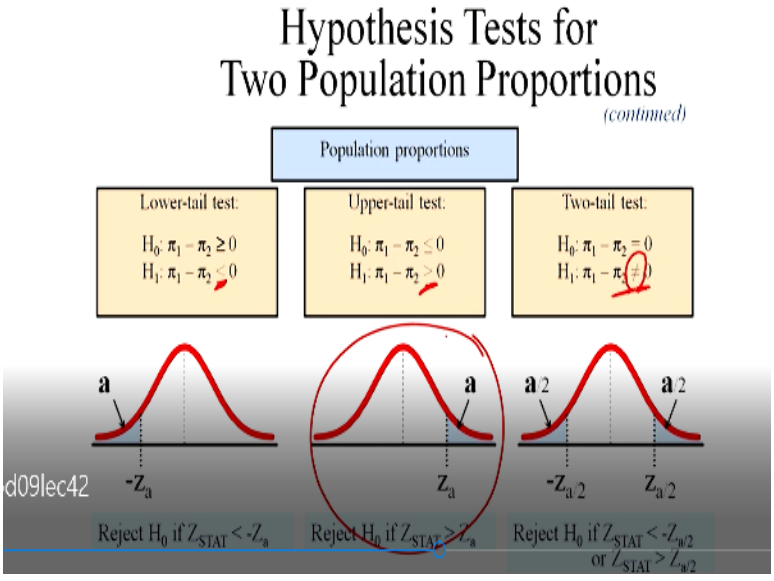
now there in this equation. So p bar of course we know how to calculate this and p1 is this p2 is this.

(Refer Slide Time: 16:10)



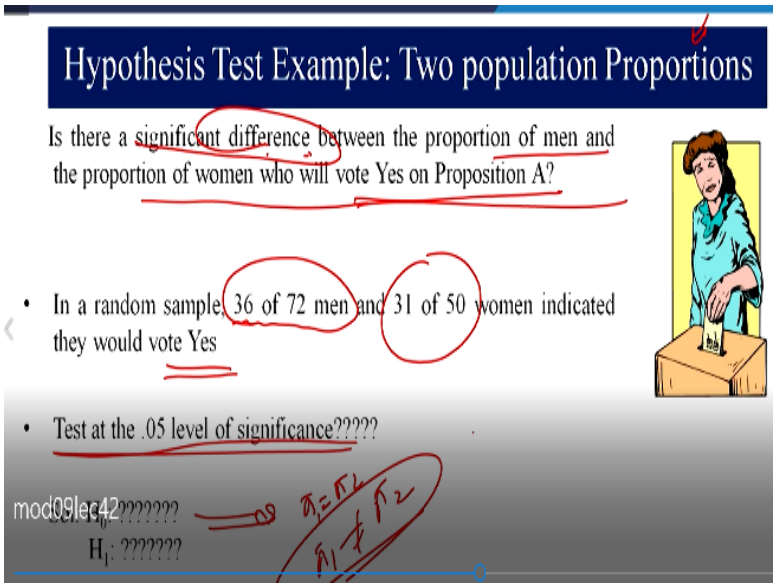
So we will work out a couple of examples so far as the distribution is concerned or how the hypothesis the distribution of hypothesis testing would look like. We have again this thing so this is a case of left tail test is it not this is an upper tail test and this is 2 tail test right so you can write null and alternative hypothesis either in this form or this form during the same thing.

(Refer Slide Time: 16:43)



This is how the distributions would look like so this is a case of left tail test because this symbol is < type this is the right tail test or an upper tail test this is the symbol and this is 2 tail test you got 2 rejection region because this symbol is not equal to type.

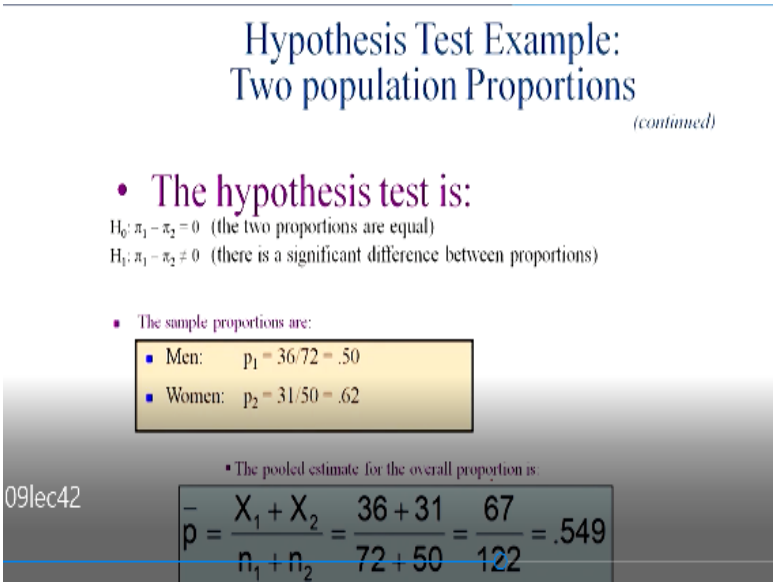
(Refer Slide Time: 17:09)



So just look at an example is there a significant difference between proportion of men and proportion of women who will vote yes on proposition A. Question is there a significant difference between the proportion of men and the proportion of women whom will vote yes on proposition A. In a random sample of 36 of 72 men and 31 of 50 woman indicated that they would vote yes on proposition A.

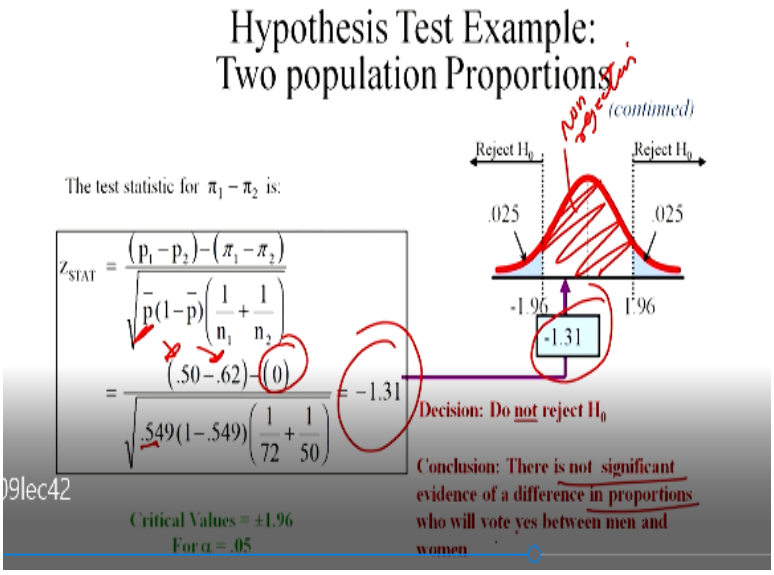
So we have to test is there any significant difference between these proportions so you need to frame null and alternative hypothesis. So how would you write this is a this is a question on hypothesis testing of proportion indirect 2 populations right. So null hypothesis would be what it would be pi1=pi2 is it not that they are same and what we have to see is there any significant difference is that impression right so there is no significant difference. So this is your alternative hypothesis this is the case of 2 tail test is it not these two are rejection regions.

(Refer Slide Time: 18:45)



Let us work out this question p1 is this 0.5 and p2 is 0.62 so this is very simple 36/72 and 31/50 calculate pooled estimate of overall proportion so we have got number of items in sample 1 X1 which is 36 X2 31 and sample sizes that there right 72 and 50. So this is 0.549 right is p bar pooled estimate of overall proportion right now since you got p bar you have got p1 and p2 n1 and n2 so you just put this value these values over here and calculated Z statistics. Of course, you will have to pi1-pi2= 0 because they are same.

(Refer Slide Time: 19:48)

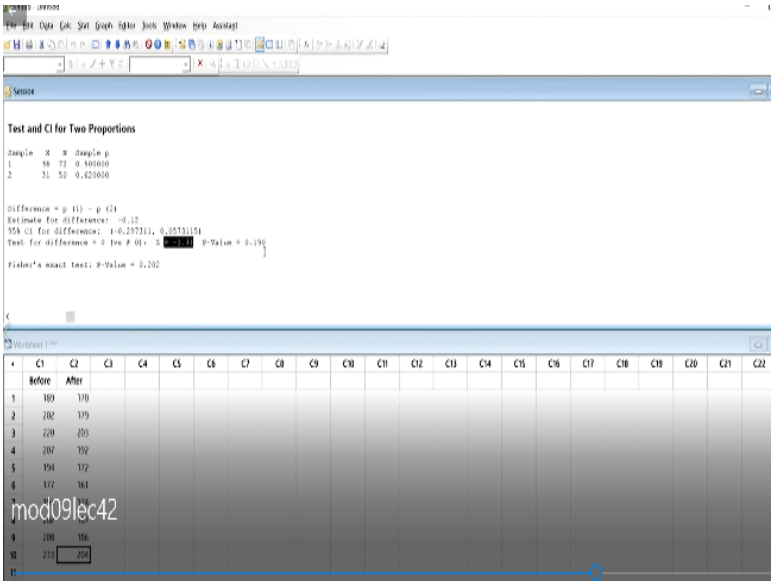


So this is what you have done pi1-pi2=0 so p1 -p2 is given over here is what we calculated pooled proportion 0.549 and this how you will you will be getting Z statistics which is -1.31 right. Now is this t value in rejection region or non-rejection region yes this is in non-rejection

regions this region is non-rejection region right non-rejection region right. So we will not reject null hypothesis and we will say that we both men and women vote equally on proposition right.

So there is no significant difference in proportions who will vote yes between men and women right so men and women equally vote on proposition A right. so we would not say that we would not say that woman vote more compared to men right but if you look at just this p2 value you will conclude that woman would prefer preposition more than men but it is not like this statistically they are one and the same thing they vote in same direction right. So let us work out this example using Minitab software.

(Refer Slide Time: 21:26)



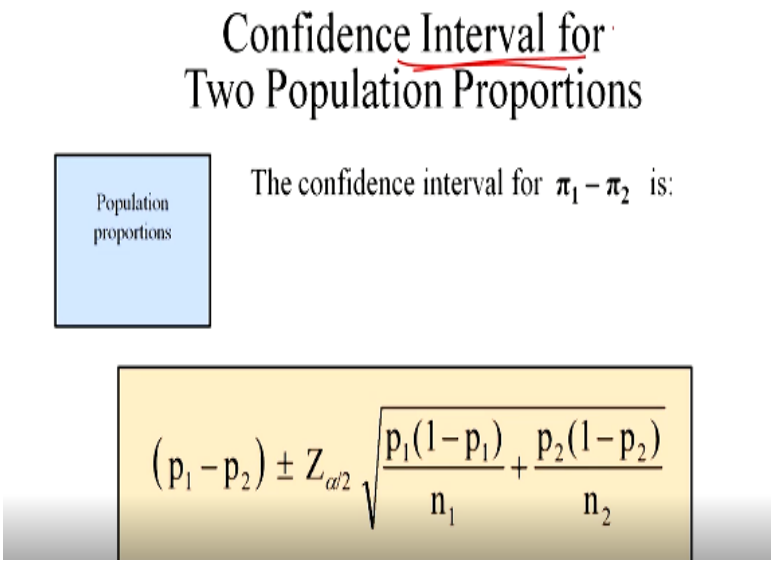
Go to stat, basic statistics 2p right 2p so both samples in 1 column no we do not want to have that type of data entry system. So let us just look at summarised data so sample 1 sample 2 number of events and number of trails so let us look at for sample 1 it is 36 and 72 we will 72 trials and 32 are events. Similarly, for sample 2 we have 31 and 50, 31 and 50 right options at what significance level we are testing it at 0.05 right.

So this is 95% hypothesis difference is 0 now alternative hypotheses is that alternative hypothesis is over here is not equal to type right so difference is not=hypothesis difference estimate proportion separately no we do not want to we want to use the pooled estimate

proportion right then click okay again click okay. so P value is 0.190 okay Z value is -1.31 let us check yes we are getting same t value sorry Z value -1.31.

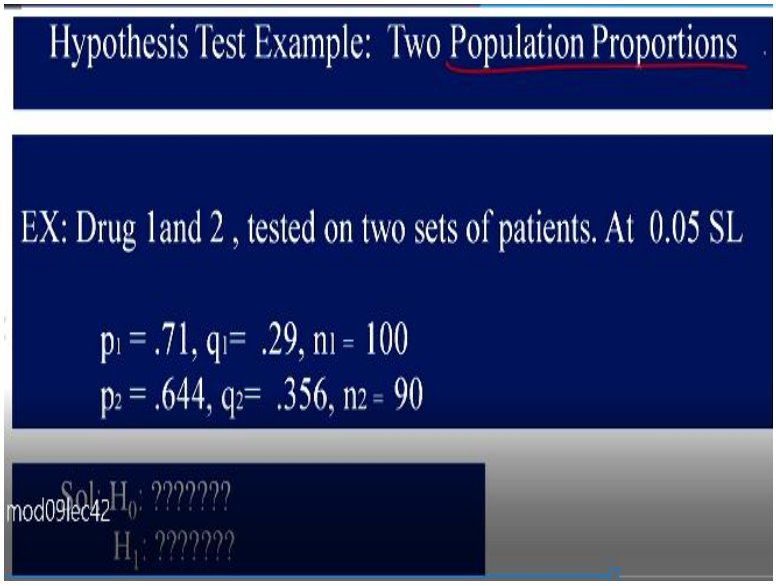
As far as P value is concerned P value is yes 0.190, P value is 0. So the P value which we have calculated using Minitab software is 0.190. Now just compare this P value with alpha value which is 0.05 right so is P value <alpha? No, P value is > alpha so we do not reject null hypothesis is it not. So this the conclusion we are getting using P value approach and the same what is the conclusion using critical value approach. As I have told you that there are multiple approaches.

(Refer Slide Time: 24:07)



And one of them is confidence interval approach. So if you wish you can apply confidence interval approach for testing hypothesis testing of two population proportion. So this is the formula for that which we will not be doing any question using this particular method.

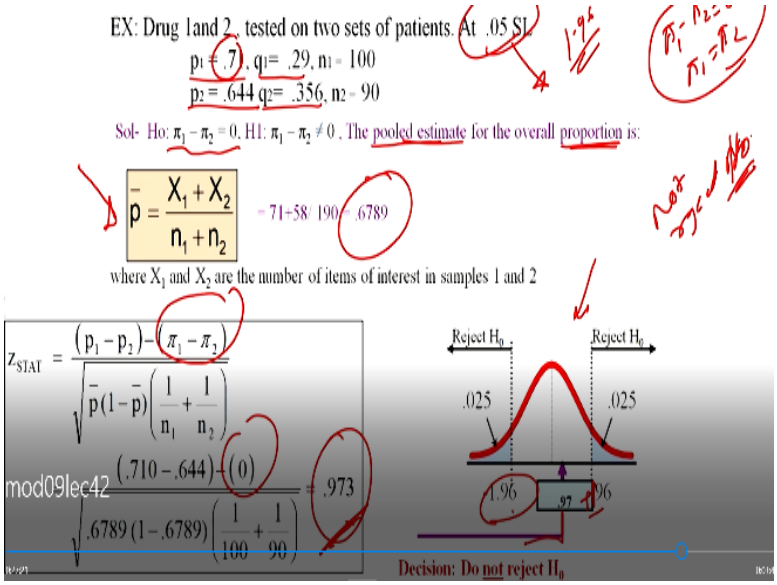
(Refer Slide Time: 24:28)



So we will move on to next question so let us say one more question on population proportion so let us say there are 2 drugs and drug 1 and drug 2 and you have to see how effective these drugs are when you test these drugs on 2 different sets of patients. So when you give drug 1 to let us say a 100 patients it is effective on 71 patients right or p1 value is 0.71 or 71%. drug 2 when you test it on 90 samples are 90 patients it is effective on 64% or in other words 64.44 percentage of patients have positively reacted to this particular drug 2.

Similarly, 71% patients positively reacted to drug 1 so we have to check are these 2 drugs equally effective that is the question. Are these 2 drugs equally effective so null hypothesis will be of course pi1=pi2 and of course it would be not equal, right? So this is pi1-pi2 not equal to each other right. So you can work out this question.

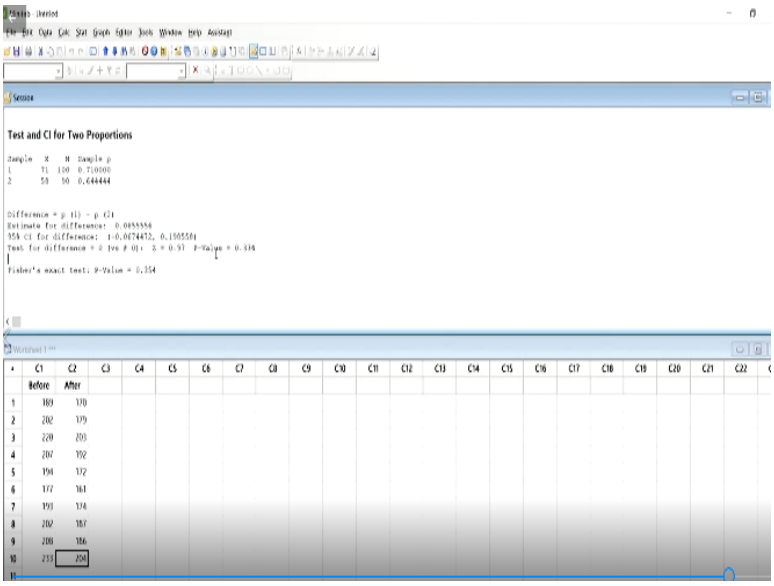
(Refer Slide Time: 26:05)



So we have got p1 then q1, 1-p1 is q1 p2 is this q2 is 1-p2 right so this is our null hypothesis that they are equal right. So you can have pi1-pi2 =0 in other words pi1=pi2 one and the same thing. Pooled variants or pooled estimate of overall proportion is this p bar which is 0.67 and Z value is this the difference of these two is 0 right pi1-pi2=0 in this case so when you calculate Z value at this point .973 and the critical table value at 0.05 significance level.

So the Z value would be 1.96 okay so +1.96 so this +1.96 and this is -1.96 and Z value is this which is in non-rejection region so we will not reject null hypothesis right H0 we will not reject null hypothesis it means we will say the drugs are equally effective right it means pi1=pi2. Now let us workout this example using Minitab.

(Refer Slide Time: 27:35)



So we will go to stat basic statistics 2p so let us look at the number of trials and number of events so for the first one it is 100 and this 0.71 so 71 would be the values over there. So this 71 events and number of trails 100 right that is why p1 we will get as 0.71 now tis is 90*0.644 so you just do this calculation 90*0.644 you will get some value. So this is 58 so you have to just take the downward of value right.

You cannot enter 57.5 or 57.8 or whatever right so out of 90 it has 58 is the number of events right. Now go to options we will have a confidence level as this 95% alternative hypothesis will remain same we have to use pooled estimate of the proportion right so we will click it okay again okay yeah so Z value we are getting a here is 0.97 which is same as here right when we calculated it right.

Now the P value is 0.334, 0.334 is the P value P value is 0.334 alpha is 0.05 is alpha. Is P <alpha? No, P is >alpha so we do not reject null hypothesis so using P value approach we are getting the same answer which we were getting using critical value approach. So these are a couple of examples we have worked out on hypothesis testing of proportion. You should keep in mind this is quite an interesting topic on hypothesis testing of proportion and whenever you are given two population proportions.

You can compare whether there is a significant difference exists or not. So with this, let me finish today's session, we will have some more examples on this particular topic in next session. Thank you.