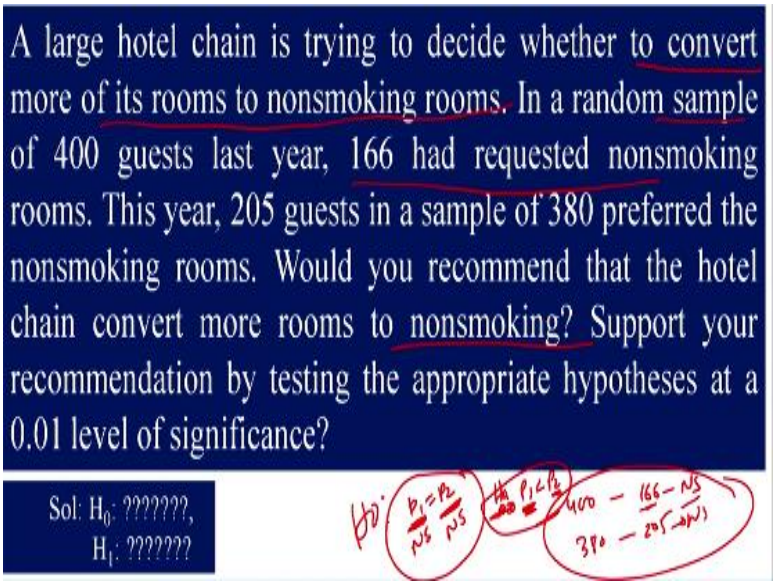
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture - 43
Hypothesis Testing of Proportion

Hello friends. I welcome you all in this session. As you are aware in previous session we were discussing about hypothesis testing of proportion and we have worked out couple of examples as well. In today's session, we look at few more examples on hypothesis testing of proportions 2 samples and we will work out using these example through critical value approach as well as proportion value approach. So let us look at this example quite an interesting one.

(Refer Slide Time: 01:03)



A large hotel chain is trying to decide whether to convert more of its room to nonsmoking rooms. In a hotel you have got smoking rooms and nonsmoking rooms. So the hotel chain is trying to decide whether should it convert its smoking room to nonsmoking room or not right. So in a random sample of 400 guest last year 166 had requested nonsmoking rooms. So out of 400, 166 requested for nonsmoking rooms right.

This year 205 gets in a sample of 380 requested are preferred nonsmoking rooms. Now what would you recommend that the hotel chain convert more rooms to nonsmoking. Support your recommendations by testing the appropriate hypothesis at 0.01 significance level. So this is quite an interesting example and you should know how to frame a null and alternative

hypothesis.

Because in this case these hotels wants to know should it convert its smoking rooms to nonsmoking rooms or not. So there are different ways in which you can frame null and alternative hypothesis. So let us look at it in this way. So let us say P1=P2 right. The proportion of guess who preferred nonsmoking rooms last year and proportion of guests preferred nonsmoking rooms in 2000 let us say this year right.

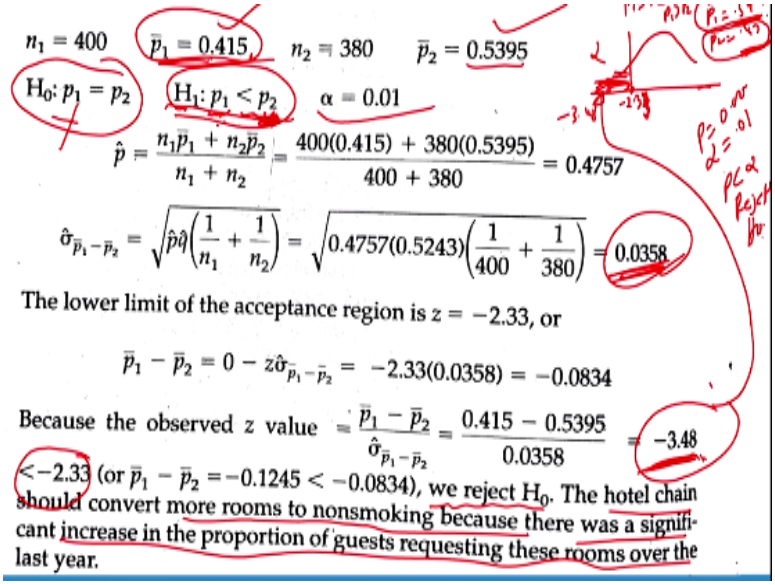
Now when the hotel should convert its smoking room to nonsmoking room there is only one condition the hotel should convert its smoking rooms to nonsmoking rooms when hotel is ensured that there are more and more nonsmokers coming in the hotel right. So if that hypothesis is proved that more and more non-smokers are coming if this hypothesis proved then hotel should convert its smoking rooms to nonsmoking rooms.

In other words, we can say that if let us say if number of smokers are reducing then also the hotel should convert its smoking rooms to nonsmoking rooms. So let us say it is null hypothesis is this that the number of smokers and let us say number of non-smokers are same right. There is no change in this proportion. So alternative hypothesis can be let us say P1<P2 right.

What is P1 the non-smokers who requested rooms last year and non-smokers room non-smokers rooms nonsmokers preferred rooms this year right that is P2. So if we say that the nonsmokers were less previous year and if we accept this hypothesis it means the hotel should convert its nonsmoking rooms to sorry its smoking rooms to nonsmoking rooms once again.

If this hypothesis is true or if we do not reject this hypothesis if we accept this hypothesis it means that the nonsmokers in earlier year were less than this year. So let us look at this approach first.

(Refer Slide Time: 05:46)



$n_1 = 400$ $\bar{p}_1 = 0.415$ $n_2 = 380$ $\bar{p}_2 = 0.5395$

$H_0: p_1 = p_2$ $H_1: p_1 < p_2$ $\alpha = 0.01$

$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{400(0.415) + 380(0.5395)}{400 + 380} = 0.4757$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.4757(0.5243)\left(\frac{1}{400} + \frac{1}{380}\right)} = 0.0358$$

The lower limit of the acceptance region is $z = -2.33$, or

$$\bar{p}_1 - \bar{p}_2 = 0 - z\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = -2.33(0.0358) = -0.0834$$

Because the observed z value $= \dfrac{\bar{p}_1 - \bar{p}_2}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} = \dfrac{0.415 - 0.5395}{0.0358} = -3.48$ < -2.33 (or $\bar{p}_1 - \bar{p}_2 = -0.1245 < -0.0834$), we reject H_0. The hotel chain should convert more rooms to nonsmoking because there was a significant increase in the proportion of guests requesting these rooms over the last year.

So null hypothesis is let us say this both these proportions are same and this P1<P2. What is P1 is the proportion of nonsmokers last year less than proportion of smokers non smokers this year right at alpha=0.01 significance right. So how did you get this one this value 0.415 you just got it by let us say it is 166/400 and 205/380 this is how you would be getting 0.415 and 0.5395. So let us look at this standard error difference of standard error of population proportions it is 0.358.

Now since this is a case of left tailed test right. So this is left tailed test so this is your rejection region alpha okay. So it is value in table or z value in table is -2.33 okay. And the calculated z value is this. So we will reject null hypothesis. It means this is the critical value which is -2.36 sorry 33 -2.33 and calculated value is somewhere here -3.48 let us say here which is in rejection so we will reject the null hypothesis.

We will reject null hypothesis we will reject this. It means we accept this we accept it means the proportions of nonsmokers were lesser in earlier year than this year. It means we should convert the hotel chain should convert the hotel chain should convert more rooms to nonsmoking because there was a significant increase in proportion of guest requesting nonsmoking rooms. So let us work out this example.

In fact, you can work out this example in other way as well. So first let us solve this question using Minitab okay so let us look at this. So this is case of hypothesis testing of proportion 2 samples right. So go to stat basic statistics 2P 2 proportions right. So we have to have summarize data. So number of events and number of trials right. So number of events let us

look at question it is 166 out of 400.

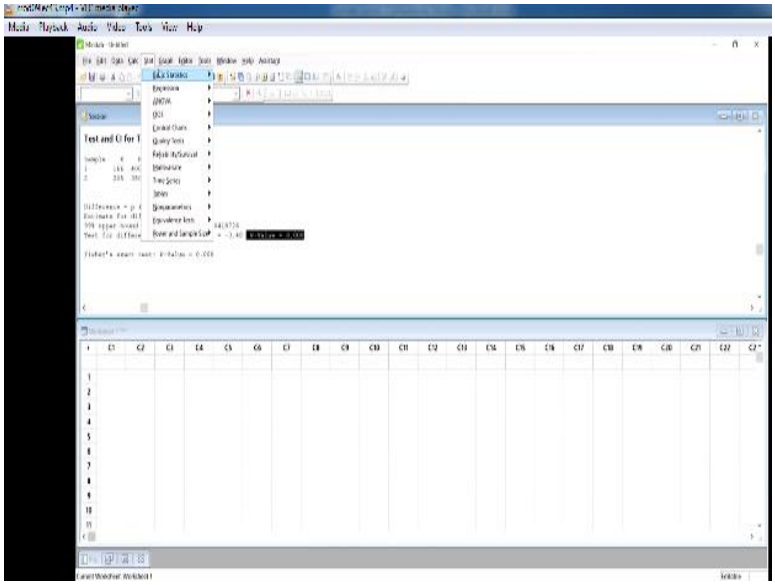
For second sample it is 205 out of 380 options. So this is hypothesis testing is it 99% significance level. Hypothesized difference is 0 and this has to be an alternative hypothesis has to be less than type of hypothesis. So use pooled estimate of the proportion and then click. So this is our calculated z value- 3.48 which is same as this just see -3.48 right and if you look at P value P value is 0 let us compare P value with alpha value.

So P value is 0.000 and so on alpha-0.01. So P is less than alpha so we will reject null hypothesis, we will reject H0. The moment we reject H0 we accept alternative hypothesis and in alternative hypothesis we are saying that the proportions of nonsmokers in previous year were more than this year right. So or in other words the proportion of nonsmokers more than proportion of nonsmokers in last year right.

So we should convert more of hotel rooms to nonsmoking rooms right. Let us work out the same example using different approach. So this is in fact nothing but the percentage of people who preferred nonsmoking rooms right. So when I say these are nonsmokers. So smokers would be let us say smokers would be 0.59 something 0.59 and what about this. It is 0.43 so it is 0.47 something is not it because 1 minus this would be this and 1minus this would be this.

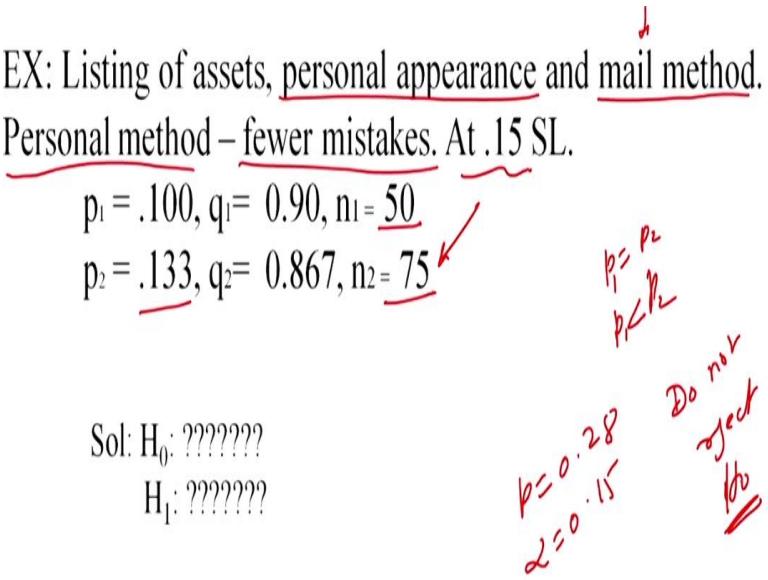
So let us look at this question in this way. Last year these many smokers were there and this year this many smokers were there. So we can frame null hypothesis like this P1=P2 and the P1>P2. The number of smokers in last year were more than number of smokers this year right. Let us look at whether do we reject or we do not reject null hypothesis right. So just keep in mind this 0.59 and this is 0.47 so we will work out this example.

(Refer Slide Time: 12:42)



So we will go to stat basic statistics 2P. So number of events would be just 400*0.49 what it was 0.59 right so this is 234. Similarly, in second case it is 175 right. So we will say that proportion of smokers last year and proportion of smokers this year okay. Go to options and this is let us say > Pi okay. Again P=0 so we will reject null hypothesis. We will reject null hypothesis again here it means we will say that the last year the nonsmokers were more than this year right. So this is how you can solve an example on hypothesis testing.

(Refer Slide Time: 13:59)



So let us look at next question let us say the income tax departments wants to have more and more returns people should file more and more income tax returns. So there are 2 methods the first is either you know either people can go to the office personally or they can send their returns through mail methods right through email right and the income tax departments think that the personal method of filing income tax returns will have fewer mistakes than mail

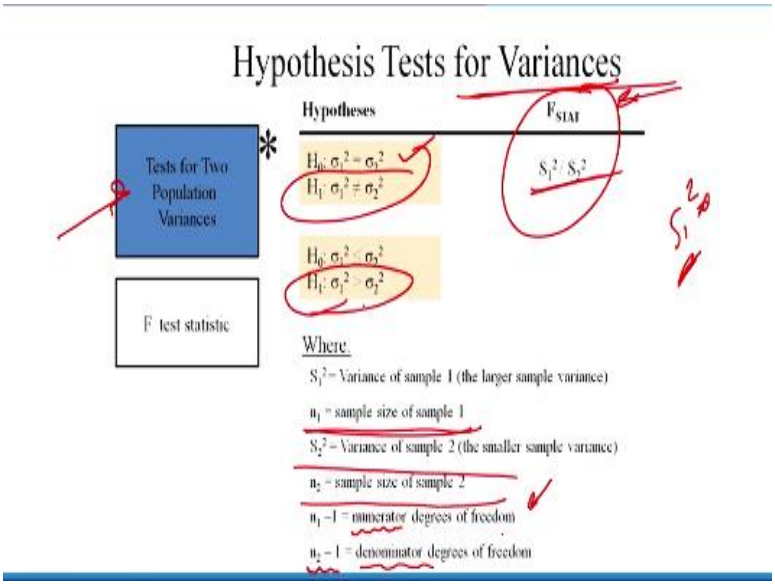
method.

So how would you frame null and alternative hypothesis. So we have got let us say P1=P2 right and P1<=P2 right. So let us work out this example using Minitab. So keep in mind this is 50 and this would be 10 and this 75 and multiply these 2 would be the number of events in this case right. So we will go to stat basic statistics 2P. So this is 50 and this would be 10, 5 right because that is that value is just 0.100 right.

So 75 is the second case so this is 75 and you will have number of events is 10 right. We will go to options. This is at what significance level okay this is 0.15 so this is 85% different hypothesis is < type right. So this is the alternative hypothesis okay again okay right. So P value is 0.28 so P value, 0.28 alpha=0.15 is P < alpha?. No, so we do not reject null hypothesis. So we do not reject null hypothesis it means both these methods are equally good.

The personal method does not decrease mistakes right so they are equally effective whether its personal appearance or mail method.

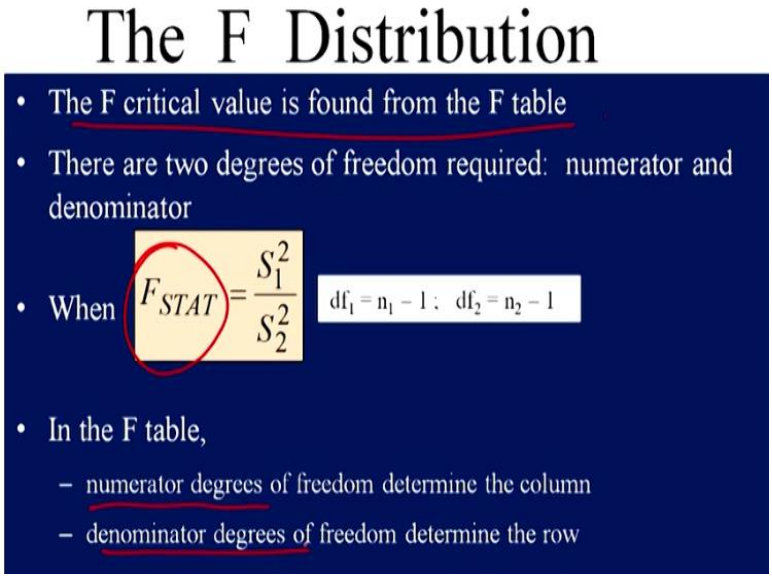
(Refer Slide Time: 17:00)



Let us look at hypothesis testing for variances. So far we have seen hypothesis testing of means and hypothesis testing of proportions right. Let us look at hypothesis testing of variances. So you have got 2 populations and you want to test whether their variances are same or not right. So for this we will use something called F statistics which is a ratio of variances of 2 samples sizes right.

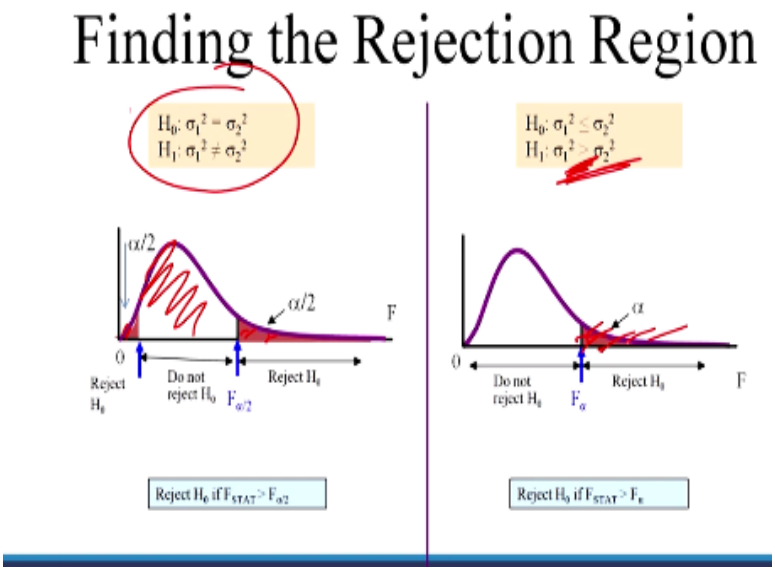
So you can have test for population variances you can have hypothesis testing like this alternative like this and you can have alternative hypothesis like this is not it. So where we all know that S1 square is variance of samples per sample n1 is you can have number of units in first sample or sample size for first sample and n2 sample size for sample 2. S2 is variance for second sample n1-1 is numerator degrees of freedom and n2-1 is denominator degrees of freedom right.

(Refer Slide Time: 18:18)



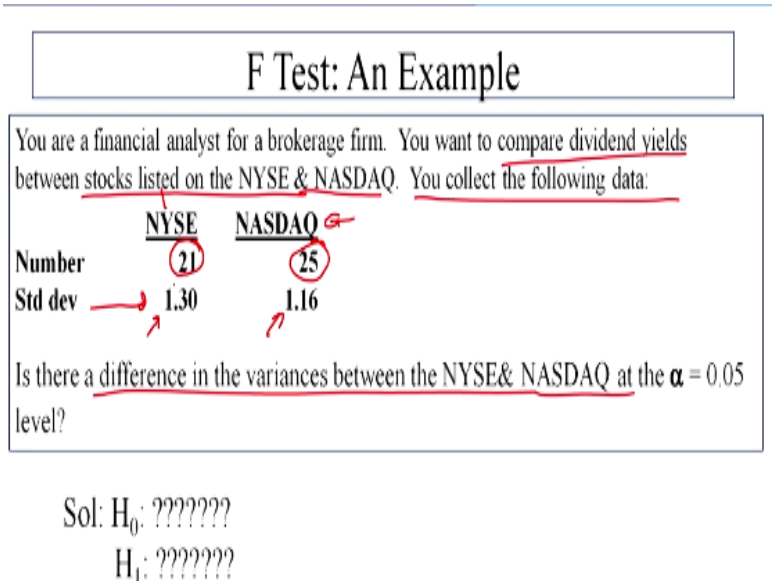
So let us look at this test so we have to find out critical value from table F table and we need to compare those F table values with the F statistics which we calculate using this formula which is ratio of these 2 variances, but the F table we get value at numerator degrees of freedom and at denominator degrees of freedom right.

(Refer Slide Time: 18:55)



So this is how the distribution looks like so null hypothesis is that these 2 distributions are equal and they are not equal. So this would be rejection region right. So this is 2 tailed test right this is your non rejection region. Similarly, this is an upper tailed test we have got just one rejection region because alternative hypothesis is this sign is > type. So this is your rejection region.

(Refer Slide Time: 19:28)



So let us look at this question. You are a financial analyst for a brokerage firm. You want to compare dividend yields between stocks listed on the New York Stock Exchange and NASDAQ you collect the following data. So 21 stocks from here and 25 stocks from NASDAQ and standard deviation 1.3, 1.16. Is there a difference in variance between these two? Though, we have been given standard deviation we can easily convert a standard deviations into variances right.

So we just want to know is the variance of these 2 samples same? So you need to frame null and alternative hypothesis right.

(Refer Slide Time: 20:25)

F Test: Example Solution

- Form the hypothesis test:

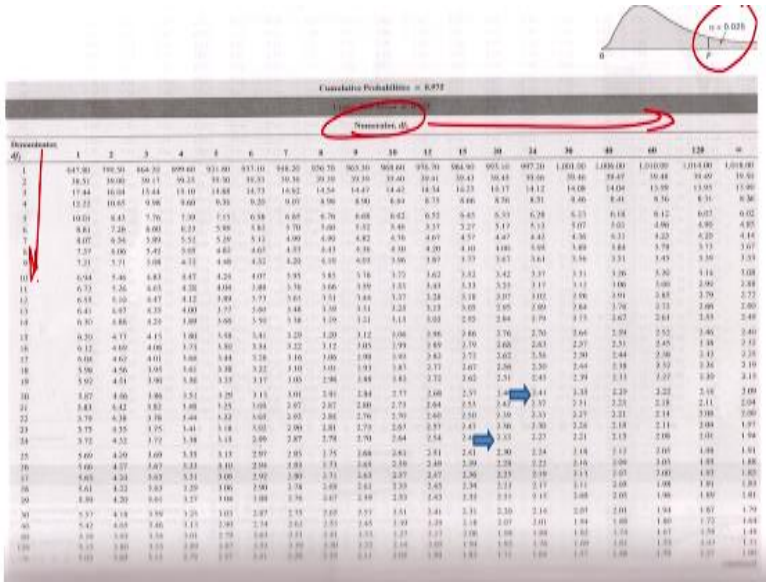
  $H_0: \sigma^2_1 = \sigma^2_2$ (there is no difference between variances)

  $H_1: \sigma^2_1 \neq \sigma^2_2$ (there is a difference between variances)

- Find the F critical value for $\alpha = 0.05$:
- Numerator d.f. = $n_1 - 1 = 21 - 1 = 20$
- Denominator d.f. = $n2 - 1 = 25 - 1 = 24$
- $F_{\alpha/2} = F_{.025, 20, 24} = 2.33$

So let us look at this. So this is your null hypothesis where there is no difference and there is a difference so it is a case of 2 tailed test and its significant level 0.05. So at numerator degrees of freedom and at denominator degree of freedom we need to look at F value from table right. So let us look at the table value.

(Refer Slide Time: 20:52)

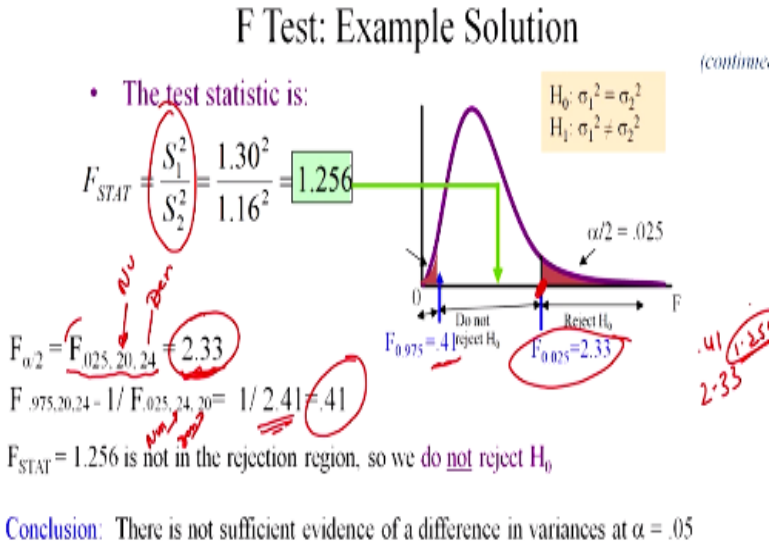


So this is the table your alpha is 0.05 so look at this table okay. Now how to look at F value in table right. So first of all you have got a numerator degrees of freedom which is for sample 1 right. So sample 1 of size 21 so 20 numerator degrees of freedom and sample 2 is denominator degrees of freedom. So n2 is 25 okay. So 25-1, 24. So what you want to know we want to find out F value at 20 degrees of freedom of numerator and 24 degrees of freedom of denominator right. So 20 and 24.

So these are numerator degrees of freedom this row wise and this column is denominator right. So what we want 20 numerator and 24 denominator right 20 numerator so this numerator is 20, 20 numerator and 24 denominator 24 is this so this is the value 2.33. So we will get because this is a 2 tailed test you will have to find out 2 values of F, 2 critical values of F.

It is not like hypothesis testing of means or proportions where just one value would be sufficient for left side and right side. Here for left and right side you will have to have 2F critical values. So the first F value is 2.33 right.

(Refer Slide Time: 22:57)



Now this is what we have calculated 2.33 which is this. So this is the upper limit upper critical limit. Now you need to find out lower critical limit as well. So for that what you need to do is this it is very simple. Since you have already calculated this right. So to calculate lower critical value what you have to do is this 1/this you need to just change the degrees of freedom just see.

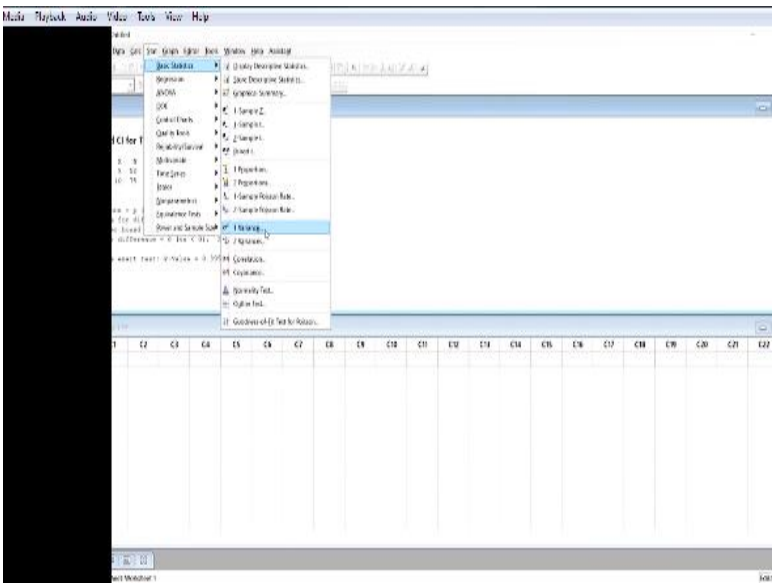
Here it was 24 degrees of numerator this was numerator this was denominator. Now change this 2 this is now numerator and this is denominator. So we will just change the degrees of freedom. Now let us look at the same table when 24 degrees of numerator and 20 degrees of denominator. So 24 numerator = same table right 24 numerator and 20 denominators so this is the value 2.41. So 1/ 2.41 will give you lower critical limit okay.

So now you have to see whether this ratio is in this limit or not. If it is in this limit we do not

reject null hypothesis otherwise we will reject null hypothesis. So in this case if you look at our lower limits is 0.41 upper limit is 2.33 and calculated F value is 1.256 which is in between. So we will not reject null hypothesis. We will not reject null hypothesis it means what it means what it means that the variances of these 2 stocks are same one and the same thing right. There is no difference in variances.

However, if you look at this standard deviation you will say that variance of this would be higher no it is not like that. Statistically these 2 variances are exactly same there is no difference. So we will work out this example using Minitab.

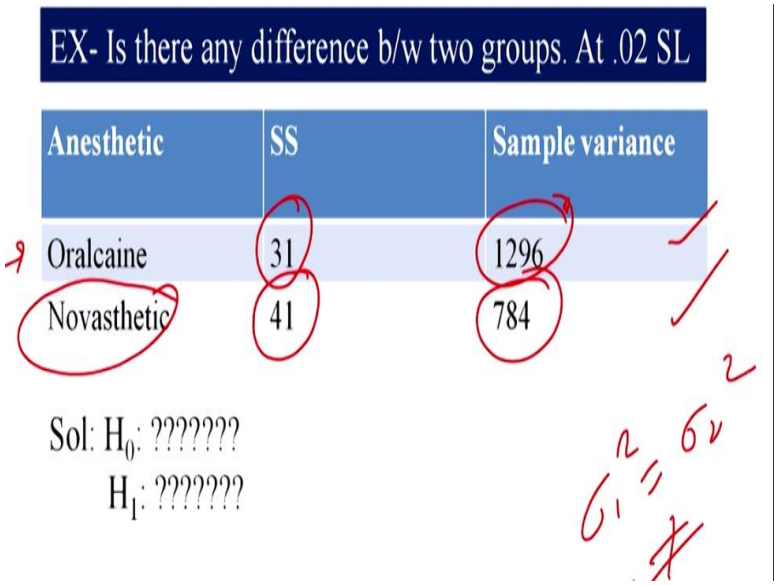
(Refer Slide Time: 25:46)



So we will go to stat, basic statistics, go to 2 variances right. So we will have sample okay sample standard deviations. So we have sample size it is I think 21 just check we will check it 21 1.3 and 25 1.16. We will go to options at what confidence level we have to test at 95% significance level. So it is 95 hypothesized ratio is 1 right. We want this to be 1. If this ratio is 1 it means, we will not reject null hypothesis.

So let us look at this ratio is not equal to hypothesized ratio we use okay again okay. So just look at P value over here this is P value right. In this case if you look at. So but 21 degrees of numerator 24 degrees of denominator so P value is 0.58. So this P value is if you look at carefully P value is 0.589 and here it is 0.05. So P value is it is more than alpha so we do not reject null hypothesis. We do not reject null hypothesis. We do not reject null hypothesis it means they are one and the same okay.

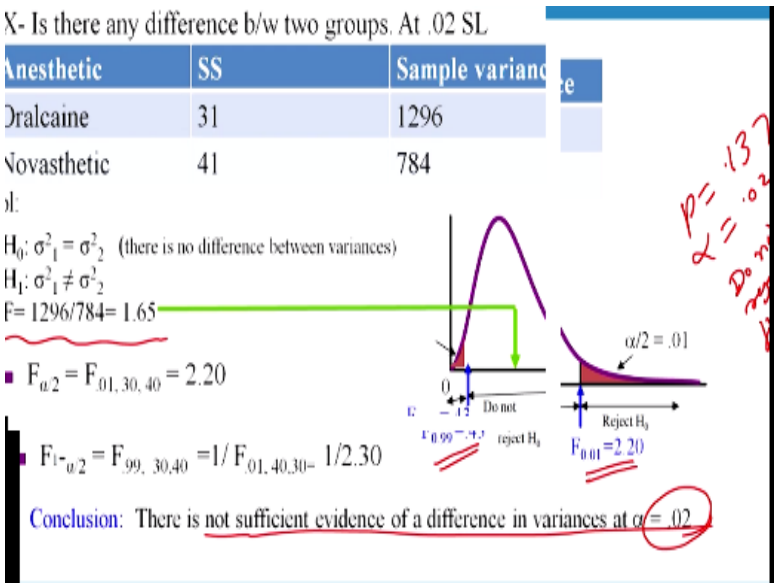
(Refer Slide Time: 27:58)



We will try one more example and let us say there are 2 types of anesthetics and we want to find out the effectiveness of these 2. So when anesthetic this one oralcaine was given to 31 patients the sample variance was this. This was the sample variance for second anesthetic sample size was 41 and sample variance was 784 right. So can we say that these 2 anesthetics are one and the same thing are they equally effective.

In other words, can we say that sample variances are same. So this is your null hypothesis and this would be alternative hypothesis right.

(Refer Slide Time: 29:00)



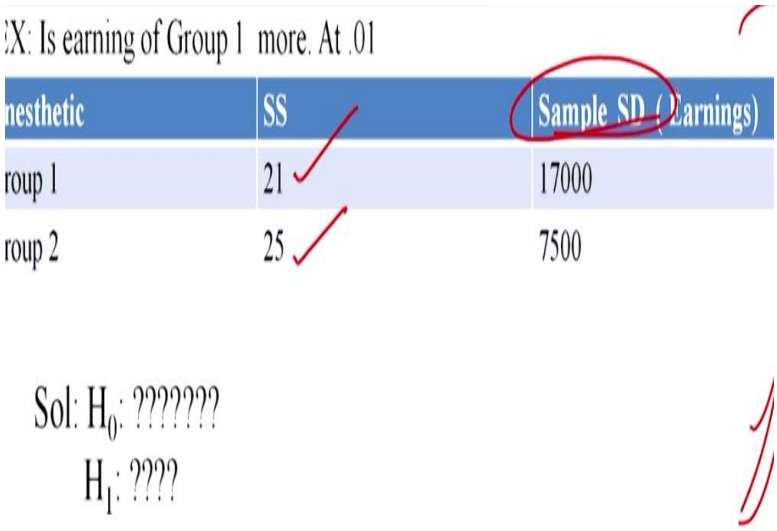
So similar to previous example we can work out upper limit and lower limit for this as well. And since the calculated F value is in non rejection region so we will not reject null hypothesis and we will say that there is no sufficient evidence of difference in variances at

0.02 alpha. So let us look at this question using Minitab. So go to stat basic statistics, 2 variances right.

So sample size is 31 and 1296. In fact, rather than this we can change this variance right a simple variance right. Otherwise you will have to just square it and write down their values right. So variance for first sample is 1296 and for 41 it is 784 right. We will go to options it is 0.02 yeah so it is 98% significance level and of course alternative hypothesis will remain same it is not equal to type so we will press okay and again okay.

Let us look at P value P value is 0.137, P value is 0.137 and alpha=0.02. So P value is > alpha. So we do not reject null hypothesis. Now this is the way in which you can work out examples in fact the lower and upper critical limits can be obtained from this table.

(Refer Slide Time: 31:05)



Similarly, you can work out this example as well. So you have given let us say 2 groups sample size 21 here sample size 25 and you have been given sample standard deviation. So we just want to know is earning of these 2 groups same. Is there any difference between earning of these 2 groups right? So you can work out this example using Minitab or by critical value approach. So with this let me complete today's session. Thank you very much.