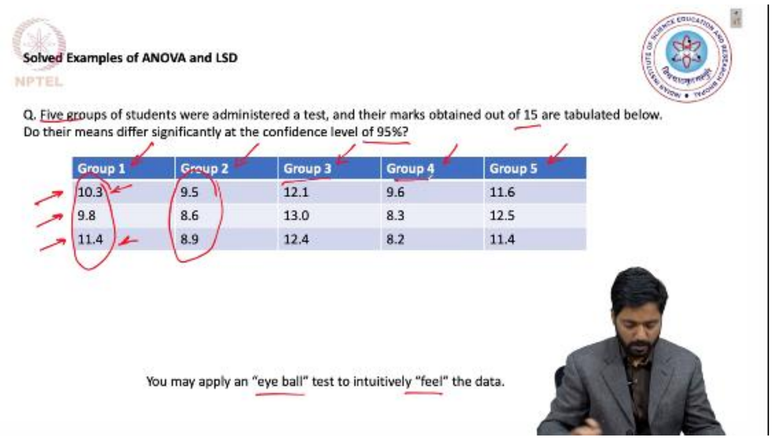
Quantitative Methods in Chemistry

Dr. Aasheesh Srivastava
Dr. Bharathwaj Sathyamoorthy
Department of Chemistry
Indian Institute of Science Education and Research – Bhopal

Lecture - 22
ANOVA and solved Least Significant Difference example

So welcome back to lecture 4 of Week 5 for this course Quantitative Methods in Chemistry. So far we have got introduced to analysis of variants and the least significant difference during this week. Now we will be applying these 2 concepts and protocols on some datasets and see how they can be applied on them.

(Refer Slide Time: 00:56)

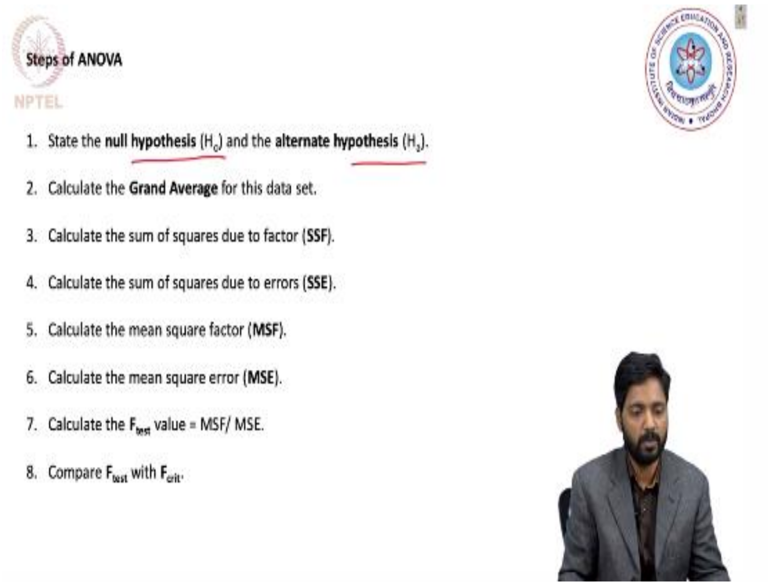


So the question that I have brought up for you is that we have 5 groups of students that were administered a certain test and the marks that they obtained out of 15 are tabulated below and the question that we need to answer is do their means differ significantly at the confidence level of 95%. So this question clearly indicates that we need to apply analysis of variants because we are dealing with multiple groups, group 1 to group 5.

And in each group we have 3 replicates. So before undertaking analysis of variance on this sample I would want to suggest to you that you are free to apply what is known as an eye ball test to intuitively feel the data. In other words you can have a closer look at the datasets and see if you can figure out whether certain datasets will be indeed different from others or not. For example if I look between dataset 1 which is group 1 these 3 data points.

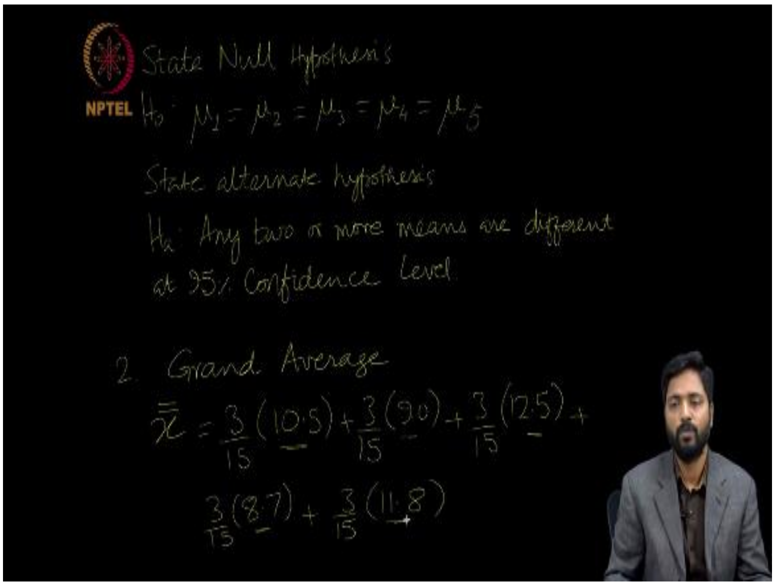
And group 2 what I observe is that in group 1 I have 2 values which are in double digits while in group 2 all the values are in single digits. Similarly, if I compare group 3 and group 4 similar difference is seen. So visually also this data indicates that there may be differences between the mean values of these groups and now we will perform a detailed analysis of variance on this dataset to check whether this indeed holds true or not.

(Refer Slide Time: 03:16)



So the protocol enlisted is here we will have to first state our null hypothesis and also the alternate hypothesis. Let us go to the board and do that.

(Refer Slide Time: 03:32)

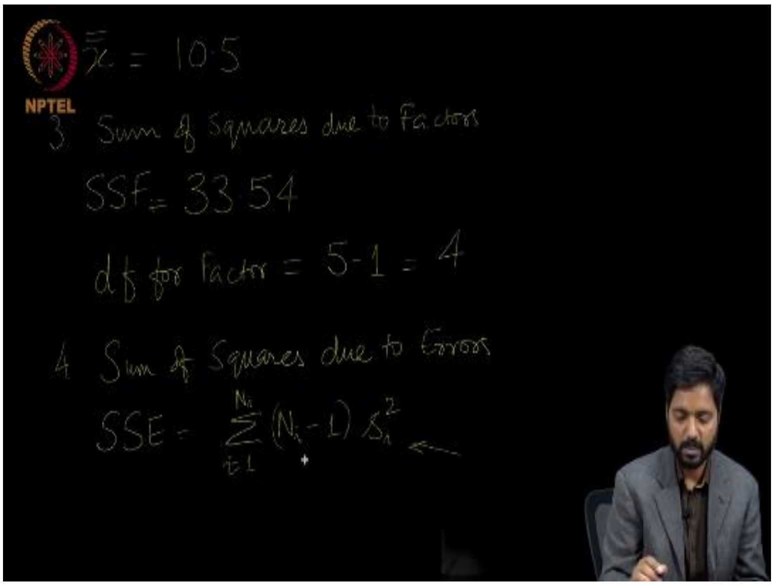


So as step 1 we will state our null hypothesis which we denote as H0 and which indicates that the mean values of group 1, group 2, group 3, group 4 and group 5 are all same. Now we also state the alternate hypothesis and alternate hypothesis we denote as Ha this says that any two or more means are different at 95% confidence level. Now let us see what we need to next, next of course we need to calculate the grand average.

And grand average which we denote as x double bar is the weighted mean of all the data points. So for the first dataset this will be so let me also tell you that I had calculate these numbers which are shown here are actually the means of individual groups and I have tabulated them here. So for group 1 the mean is 10.5 similarly for group 2 it is 9.0, 12.54 group 3, 8.74 group 4 and 11.84 group 5.

And what I have also done is to calculate the standard deviation values for each group and those are also tabulated here for your convenience. So when we use the group means to calculate the grand average.

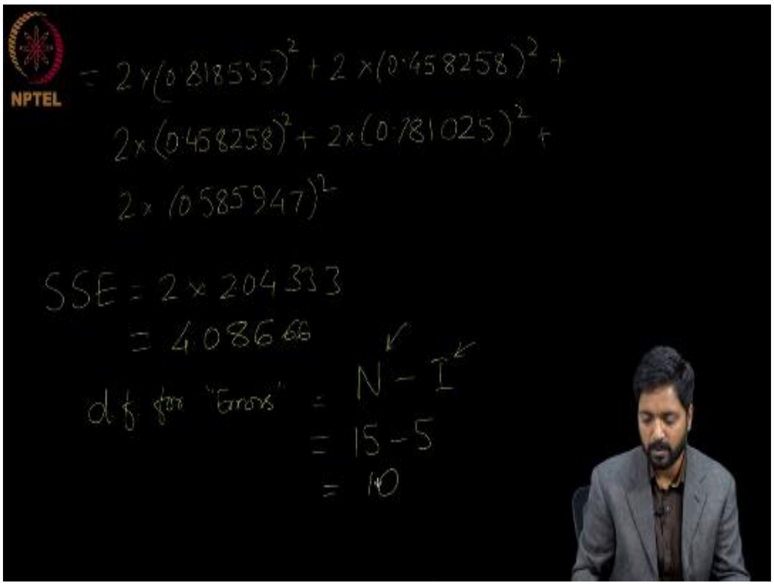
(Refer Slide Time: 07:10)



The grand average value comes out to be 10.5. Now from this as the third step we will calculate the sum of squares due to factors which we denote as SSF and without going into details I will calculate it for you and this is coming out to be 33.54 for this dataset. Now we also know that the degrees of freedom here for the factor is the number of groups which is 5 – 1 and this (()) (08:21) turns out to be 4.

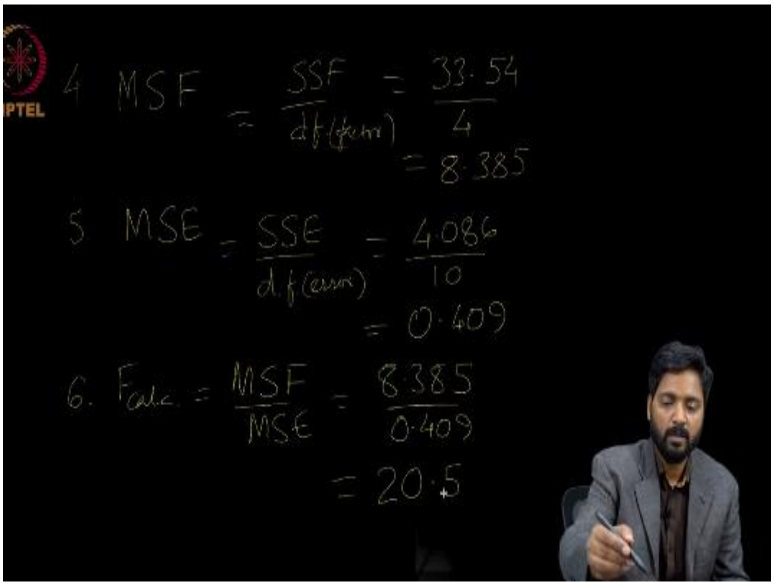
As the next step we calculate the sum of squares due to errors which we denote as SSE and SSE we have figured out is nothing, but the standard deviation squared which is variance for individual groups multiplied by the degrees of freedom within each group.

(Refer Slide Time: 09:14)



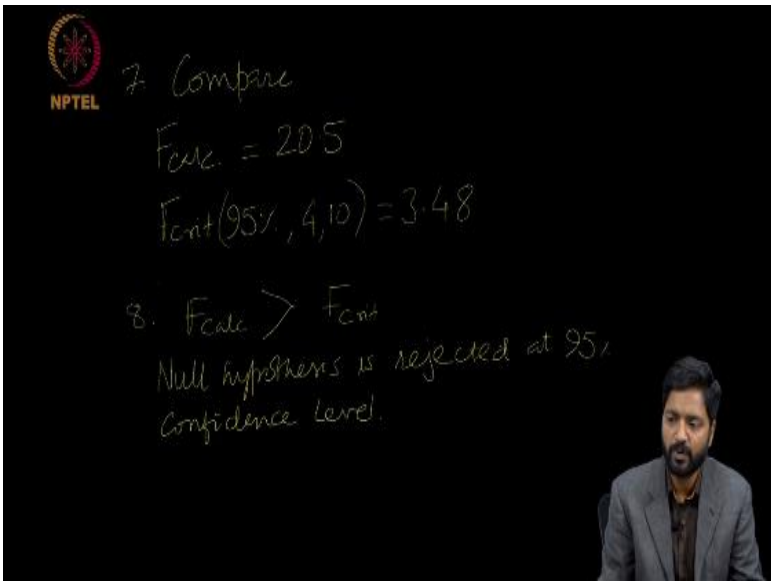
So this if we calculate this turns out to be = 2 into 0.818535 squared + 2 into 0.458258 squared + 2 into 0.458258 squared ++ 2 into 0.781025 squared + 2 into 0.585947 squared. Now when we undertake this analysis the sum of squares due to errors comes out to be 2 into 2.04333 or 4.08666. So for sum of squares due to errors the degrees of freedom for errors is = N – I where N is the total number of data points in our dataset and I is the number of groups present. So this is nothing, but 15 – 5 = 10.

(Refer Slide Time: 11:19)



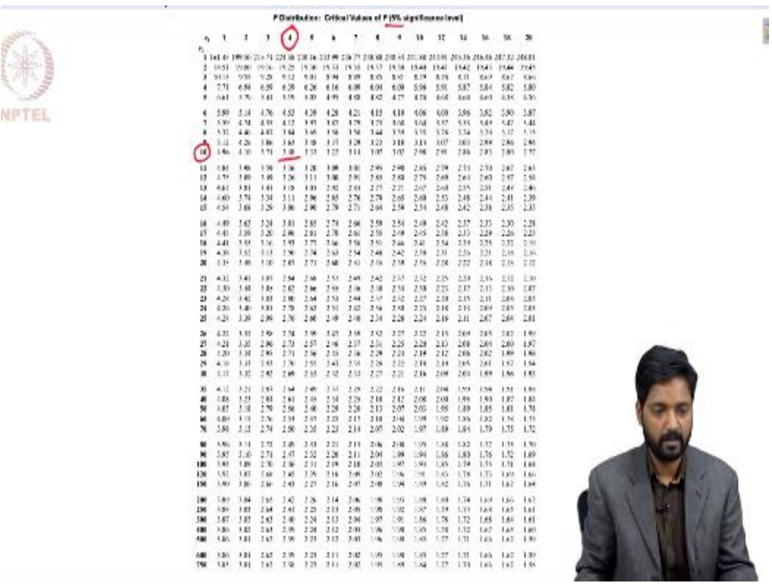
So from these we will be able to calculate in the next steps mean square factor and the mean square error. The mean square factor is nothing, but SSF/degrees of freedom for factor which is 33.54/4 equivalent to 8.385. Similarly mean square error is sum of squares due to errors//degrees of freedom for the error which is 4.086/10 and this turns out to be 0.409. Now from these we will finally calculate the F value which is nothing but MSF/MSE = 8.385/0.409 and this turns out to be a large number as 20.5.

(Refer Slide Time: 12:59)



In the final step we will compare the F calculated which is nothing, but 20.5 with F critical at 95% confidence level and the degrees of freedom being 4 and 10.

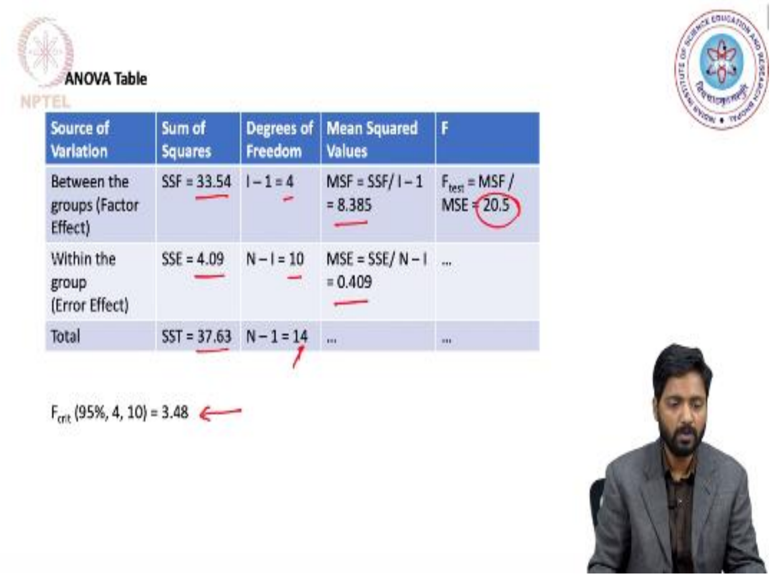
(Refer Slide Time: 13:42)



So let us quickly go to the F table at 5% confidence level and see what this number is for our data point. So we are talking about 5% significance level and 4 degrees of freedom in the numerator and 10 degrees of freedom in the denominators. Our F critical value comes out to be 3.48. So obviously what we infer here is that F calculated value is > F critical value and hence null hypothesis is rejected at 95% confidence level.

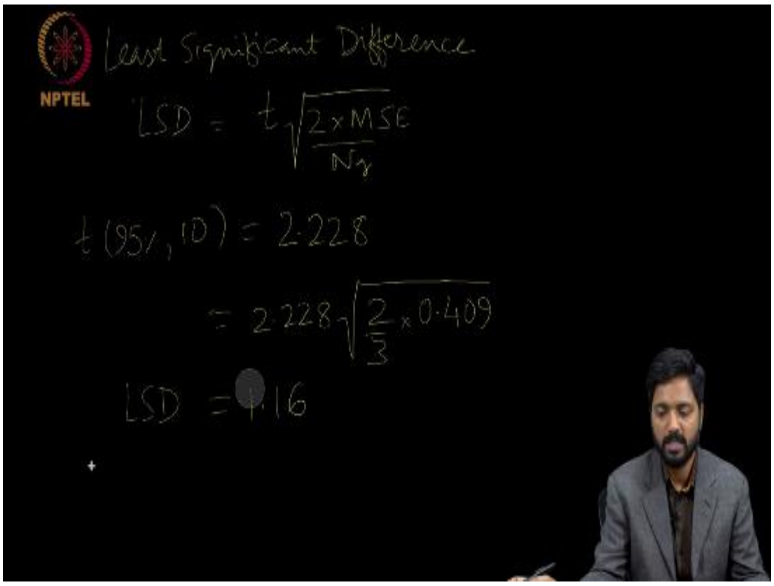
Now since the null hypothesis has been rejected we will be next calculating the least significant difference here.

(Refer Slide Time: 15:08)



And the analysis of variance table has been provided here based on the values that we obtain. So the SSF here is 33.54, SSE is 4.09 sum of squares total is also enlisted in the degree of freedom for the factor and error are enlisted and the degrees of freedom total which is 15 – 1 of 14 is also shown. The MSF and the MSE values are provided and the F calculated value is given F critical value is also denoted later for comparison.

(Refer Slide Time: 15:58)

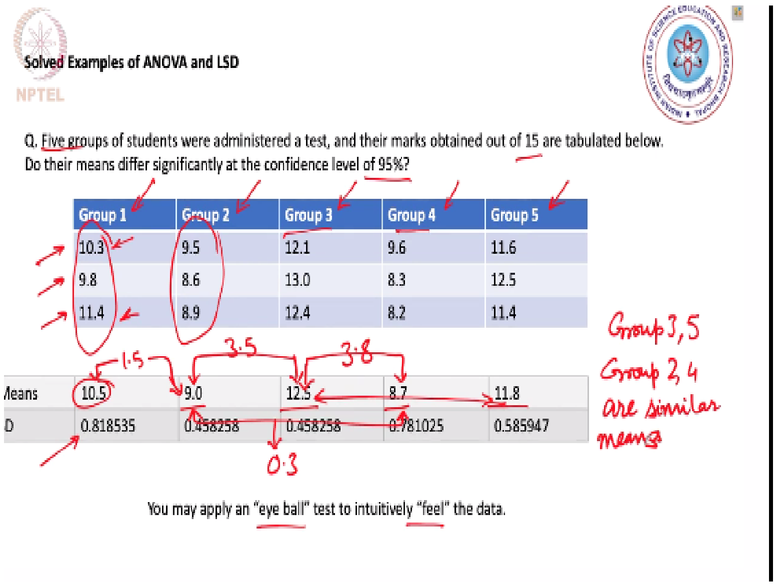


So going back the board we observe that we need to now apply the least significant difference method or protocol. So the least significant difference for this dataset will have. So the t value will be calculated again at 95% confidence level and at 10 degrees of freedom which is 15 –

5. So let us go to the t table and see what this value would be. So t table at 5% significance level and 10 degrees of freedom gives us a t value of 2.228.

Now inserting this in our LSD calculation we get the mean square error here was 0.409 and this results in a LSD value of 1.16. Now any group or sample means which we compare if the means are > 1.16 they are considered to be statistically significant.

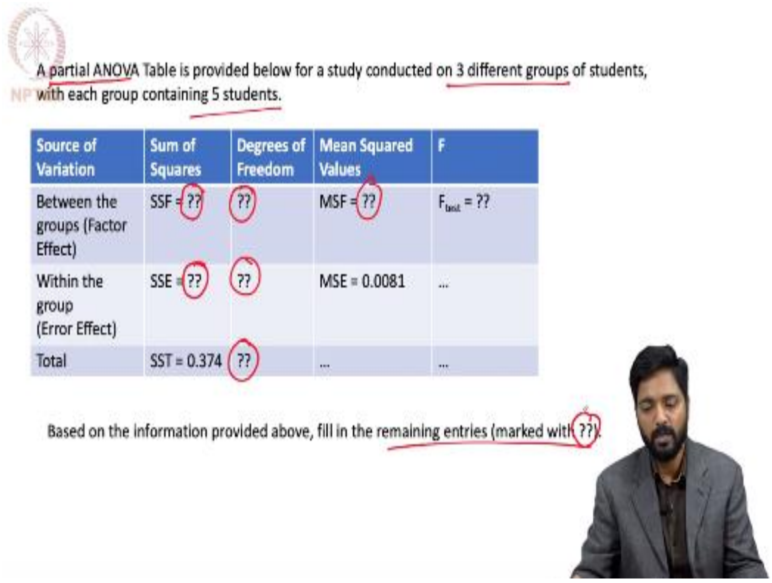
(Refer Slide Time: 18:12)



Let us go back to the desktop and see which one of these means are to be considered statistically significant. Now we have already calculated the means here. So the difference of these two is 1.5. So these two are statistically different, these two are also considered to be statistically different because the difference here is in fact 3.5 and these two are also statistically different because the difference here is about 3.8 so on and so forth.

So what we see is that only group 2 and group 4 have a mean difference of 0.3. So these two groups have statistically similar means rest all of the others will turn out to be statistically different (()) (19:31) again for example the groups 3 and 5. So group 3 and 5 group 2 and 4 are similar means or have similar means rest others have statistically different means.

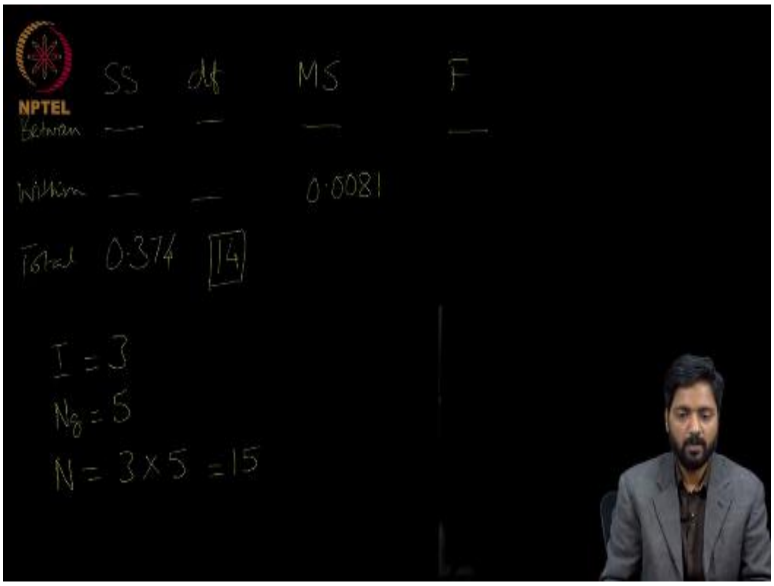
(Refer Slide Time: 20:07)



Now another way in which you can be asked this question of analysis of variance is provided here. So what one can do is to provide you with a partial analysis of variance table and ask you to fill out the rest of it the information which is not provided which I am circling here. So the only information that is provided to you is that there are 3 different groups here and each of these groups contain 5 students.

So based on this information we need to fill in the remaining entries which are marked with question market and I have also encircled them. So let us see what all information do we have here and how do we proceed on this problem.

(Refer Slide Time: 21:13)

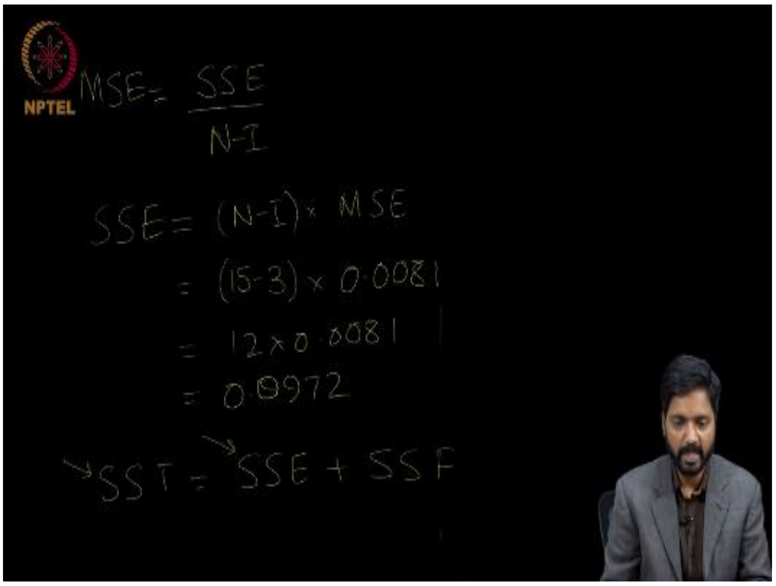


So let us again go back to the board. So what we have is the sum of squares between the groups and within the groups the degrees of freedom values, the mean square and finally the

F value and then we have our total values. So the total is provided as sum of squares total is 0.374 and the only other information provided is the mean square error which is given as 0.0081 okay and rest of the details are unfilled or needs to be provided by us.

So what we see here is that there were 3 groups that means I = 3 and each of this groups had 5 measurement so Ng or the number of replicates = 5. So the N value the total number of measurements is 3 into 5 = 15. So the degrees of freedom here total would be 14 as has been filled here. Now we have the mean square error value which is given to us and mean square error is 0.0081.

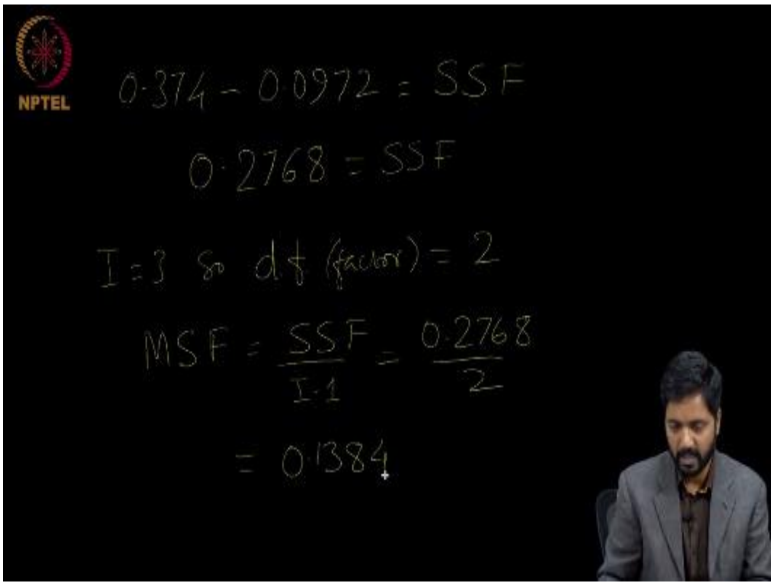
(Refer Slide Time: 22:54)



Now what we understand is that the mean square error = sum of square due to error/N- I which is the degrees of freedom for error. Now this gives us sum of square due to error = N – I into the mean square error. When we put in these numbers this is 15 – 3 into the mean square error value is 0.0081. Finally as we solve this, this comes out to be 0.0972. Now another thing that is given to us as we go up.

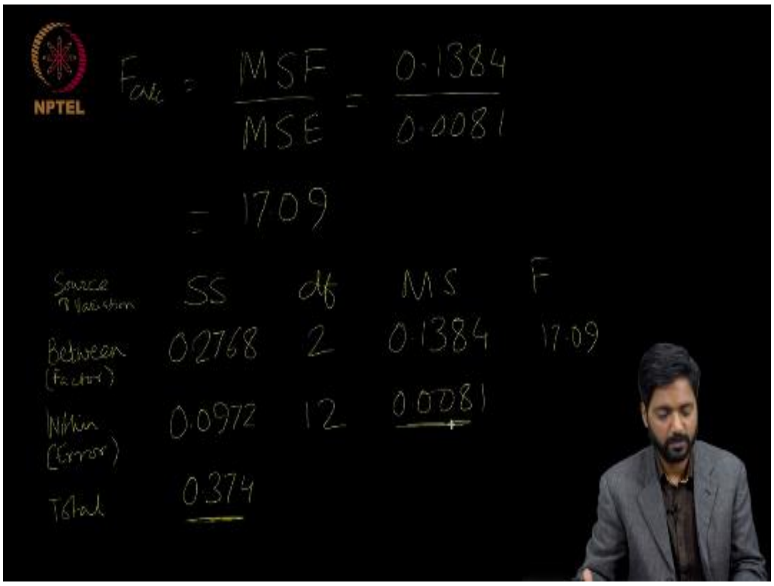
We see that the sum of squares total is provided to us and we know that sum of squares total = sum of squares due to errors + sum of squares due to factors. So since SST as well as SSE is now know to us it would be straight forward to calculate the SSF. Now let us do that.

(Refer Slide Time: 24:34)



So the SST is provided as 0.374 and from that if we subtract the 0.0972 this will give us the sum of squares due to factors and this turns out to be 0.2768 as the SSF. Now we have the SSF value given and we also know that I = 3. So the degrees of freedom for factor is 2 so this gives us the mean square factor = SSF/I – 1 = 0.2768/2 which turns out to be 0.1384. Okay now I presume we have everything available with us.

(Refer Slide Time: 25:49)

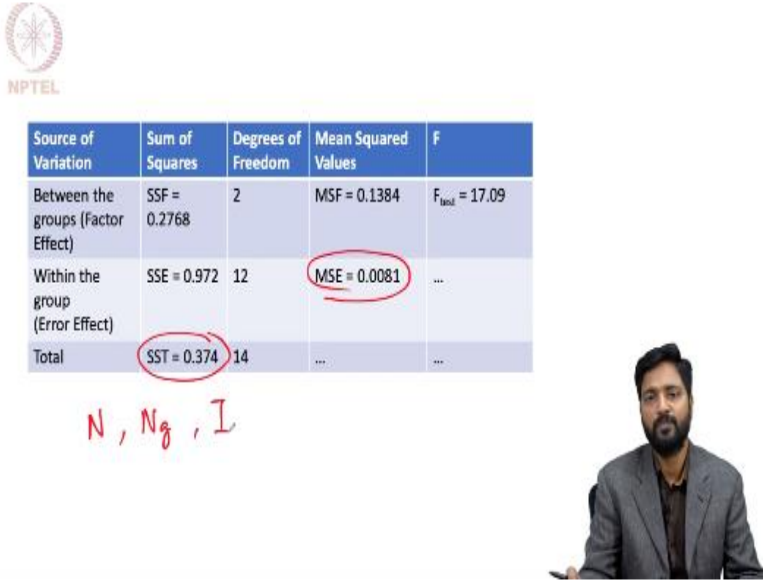


To calculate the F calculated value or F value which is nothing, but MSF/MSE and this is 0.1384/0.0081 and the F calculated turns out to be 17.09. So we have all the details now available with us to create the ANOVA table. Let us do that so the ANOVA table would look like as source of variation sum of squares, degrees of freedom, mean square and the F value when we say the source of variation is between the groups we have the sum of squares due to factor which is 0.2768 degrees of freedom.

Here are 2 because there are 3 different groups that are being compared the mean square factor will be 0.1384. Similarly, the source of variation within that is the error in that case the sum of squares due to errors was calculated as 0.0972 the degrees of freedom is 15 – 3 = 12. The mean square error is 0.0081 this is already given and we have already calculated the F value as 17.09.

So this completes our ANOVA table with only a very few information that was provided to us in the form of sum of squares total which is 0.374 and mean square error which is 0.0081.

(Refer Slide Time: 28:21)



So you can see that we can complete this ANOVA table when only very little information based on the mean square error and the sum of squares due to total has been provided. The N value and the number of replicates and the I value which is the number of groups was also provided to us that allowed us to complete this ANOVA tablet. So with this we come to an end of this week lecture.

I hope you got a fair idea of how analysis of variance is to be used both manually and through excel and you can also utilize the least significant difference to figure out which of the groups have a statistically different means at a certain confidence level. Thank you.