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

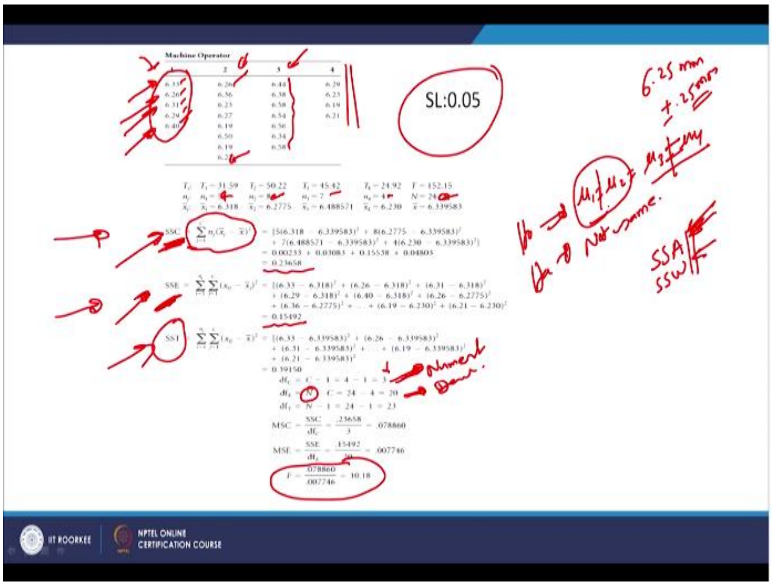
Lecture – 47
Analysis of Variance - II

Hello friends. I welcome you all in this session. As you are aware, in previous session we were discussing about ANOVA or analysis of variance. We have seen that we can use ANOVA for comparing means of two groups or more than two groups. Now let us say there are 4 groups, we can solve the 4 group's example using t-test or z-test as well but we will have to go for let us say 6 combinations or 6 pairs, 4 c2 would be the number of pairs you would be forming.

So when you have got 4 pairs then every time there would be type 1 error in all those 4 pairs. So when you solve a question having 4 groups by let us say t-test then the error probability would be very high. So to minimize that error a new term or new tool has come and that new tool or the technique is called analysis of variance or ANOVA. So if we use ANOVA, we will minimize the type 1 error.

We did workout couple of examples in previous session. Today, we will see some more examples.

(Refer Slide Time: 02:00)



So let us look at this example. In a machine shop, there are 4 operators who are drilling on aluminium plate and they have to drill a hole of let us say a diameter 6.25 mm ±0.25 mm right. So there are 4 operators in machine shop, so when the operator 1 makes hole or when he drills the aluminium plate, the diameter of hole is 6.33, 6.26, 6.31 and finally 6.41. So there are 1, 2, 3, 4, 5; five holes made by operator 1 on aluminium plate.

Similarly, operator 2 first hole diameter 6.26 and 6.22 is the diameter of the last hole. Similarly, third operator these are the diameters of holes, similarly fourth operator. Now the question is can we say that the mean of all these holes is same or does the diameter or the mean of the hole depend upon type of operator. This is the question. So you would be framing null and alternative hypothesis.

And we have to test the hypothesis at 0.05 significance level, so first of all, we will have to frame null and alternative hypothesis and what we have said in null and alternative hypothesis in case of ANOVA, we say that mu 1 = mu 2 = mu 3 = mu 4 right. So these are mean diameters. They all are equal that is our null hypothesis H0. Alternative hypothesis would be at least one of them is not equal to others right.

So all means are not same, this is our alternative hypothesis. Now when we say let us for example if mu 3 is not equal to mu 4, then we will reject null hypothesis or let us say mu 3 is- mu 4 but let us say mu 1 and mu 2 are not equal then also we will reject null hypothesis. So to solve this question, we will first of all find out sum of square among columns and sum of square within columns right.

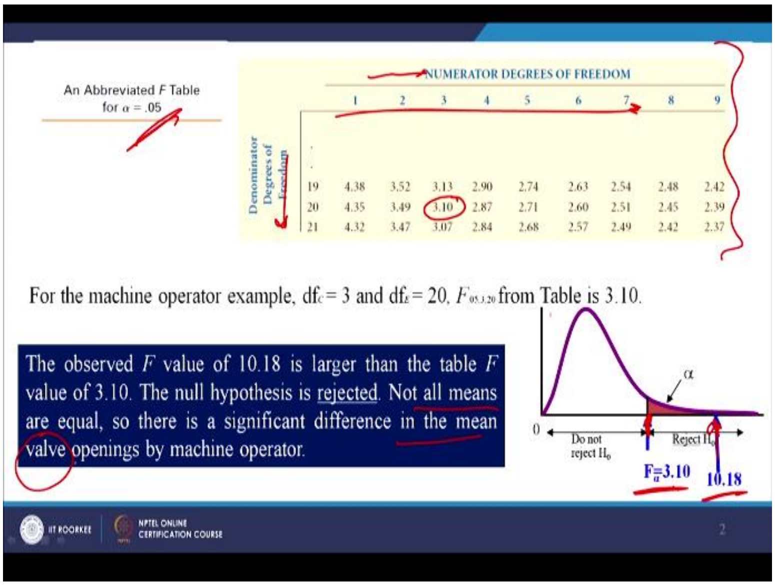
So for that we need to have this process, so this is nothing but sum of squares among columns, sum of square this is within right and this is total sum of square. In other words, this is error term or unexplained term, this is explained term and this is total variance. So if you proceed in this question, we have got 5 units in first sample. So 1, 2, 3, 4, 5 right, 8 holes in second sample or second group or second column, then 7 and 4 right.

So in this way, we have got total 24 samples. So first of all what you should do? You should find out the mean of each of these samples, each of these columns and then grand mean right. So that can be calculated right, so once you are done with mean and grand mean, calculate column sum of squares or among column sum of square is one and the same thing either you say SSA or SSC right.

Then, when you apply this particular formula to all these values will get SSC as 0.23658. Now you can easily calculate SSC for this question that would be 0.15 right. Now once we have got these two, then you need to calculate F value which is the ratio of among-column variance to within-column variance right. So that F calculated value would be 10.18. Now this 10.18 F calculated.

But you will have to find out critical value from F table as well, so for that you need to have, you need to look at F table at 3 degrees of numerator, 3 degrees of freedom at numerator and 20 denominator right. So why this is 3, because you have got 4 groups right, so number of groups-1 c-1 right this is 3, n-anything but number of items in each of those groups right so this is 20.

(Refer Slide Time: 08:30)

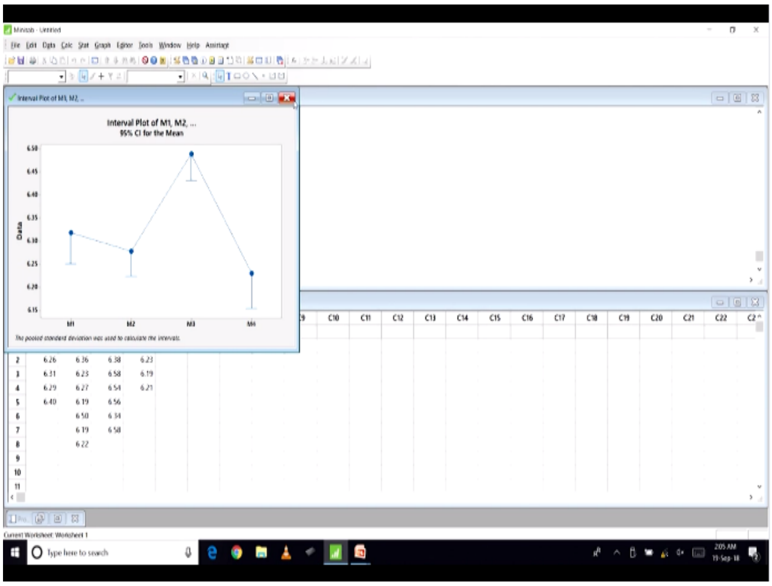


So you need to look at that F critical value from table and this is a partial table F table at alpha is=0.05 right. So that value is 3.10, so 3 degrees of numerator and 20 degrees of denominator let us see what is that value. These are numerator degrees of freedom, these are numerator and these are denominator right. So 3 and 20, so this is this 3.10, so this is your critical value F 3.10.

And as we know that in ANOVA, the F test is always left tailed test or lower tailed test, so this is our calculated F value which is 10.18 which falls in rejection region. So will reject null hypothesis, will reject null hypothesis it means that all means are not same, there is a significant difference in the mean of let us say drilling example you have taken rather than calling it valve opening example.

So will say that the mean diameter produced by these 4 operators, they are not same, they are different right, so will work out the same example using Minitab software. So this is the data we have got right, so will do data entry now.

(Refer Slide Time: 10:18)

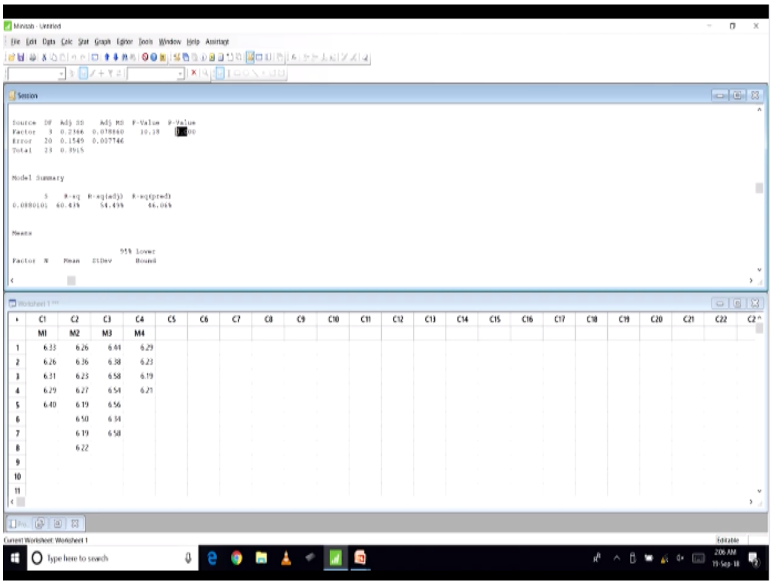


So first of all we can write here M1, M2, M3 and M4 okay. So for first we can look at data from this table so we got 6.33, 6.26 and then cell we have got 6.31. So these data we have got in a separate paper. So that is how we are doing entry. So in first column, there are only 5 units or 5 samples right 6.23, 6.27, 6.19, 6.50. Infact there are 8 samples or 8 units in this sample 6.44.

So this is how you should do data entry in all these 4 columns; 6.56, 6.34, 6.58, you have got 7 samples in column 3 then you got 6.29, 6.23, 6.19, 6.21. After that go to stat first, then ANOVA, one-way ANOVA, now here you have got there are two ways of doing data entry. The first is response data in one column for all factor levels or you can go for response data or in a separate column for each factor level right.

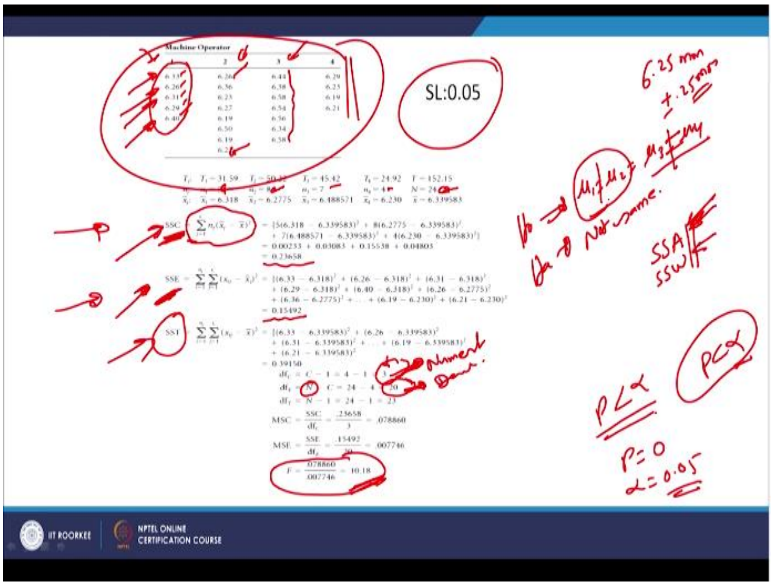
So these are 4 columns, we will select all these 4 values and will choose responses over here right. Then, we will go to option so confidence level is 95%; however, this will be a lower tailed test right. Then, we click okay, again click okay.

(Refer Slide Time: 13:05)



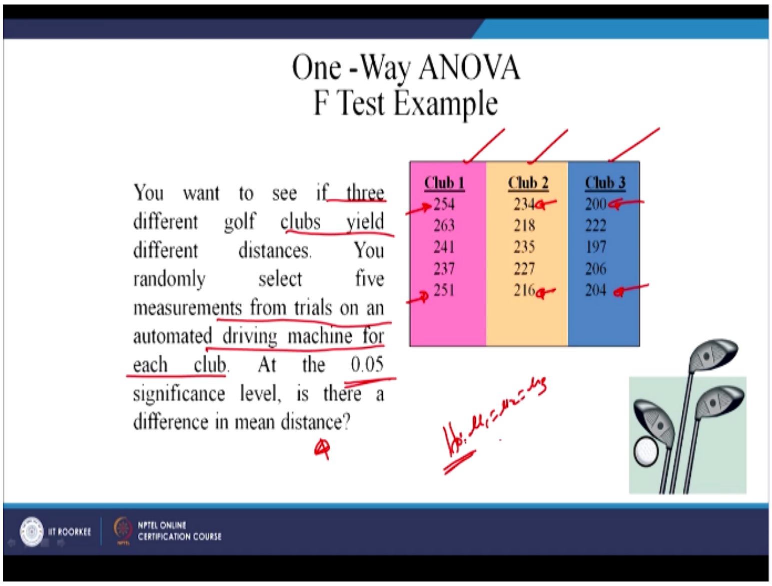
So let us look at you just minimize this yeah so we will look at P value in this. So this is our ANOVA table right, this is our ANOVA table right. Now if you look at ANOVA table, we have got calculated F value which is this right, it is 10.18 just exactly similar to what we have got here right yeah this is the one right 10.18 is calculated F statistics which is here, is not it right? If you look at P value, P value is 0 right, so what is our definition of rejecting or non-rejection of null hypothesis?

(Refer Slide Time: 14:02)



We have always said that if P is<alpha, will reject null hypothesis. In this case, P is=0, alpha is 0.05, is P<alpha? Yes, P is<alpha. So we will reject null hypothesis and we will conclude that the means are not same, they are different.

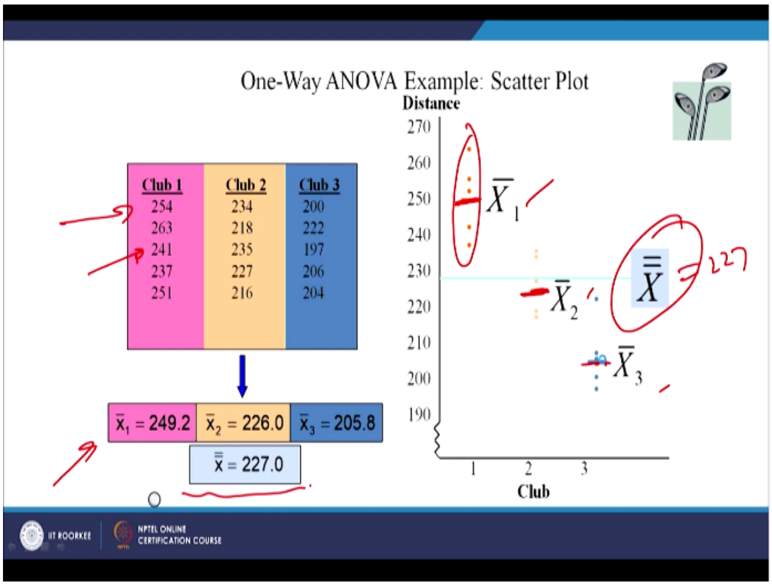
(Refer Slide Time: 14:31)



Let us look at one more example on one-way ANOVA. So you want to see if 3 different golf clubs yield different distances, you randomly select 5 measurements from trials on an automated driving machine for each club right, 0.05 significance level. Is there any significant difference in mean distance? So these are the data points you have been given. So for club 1 the first is 254, last is 251; for club 2, 234, 216; club 3, 200, 204.

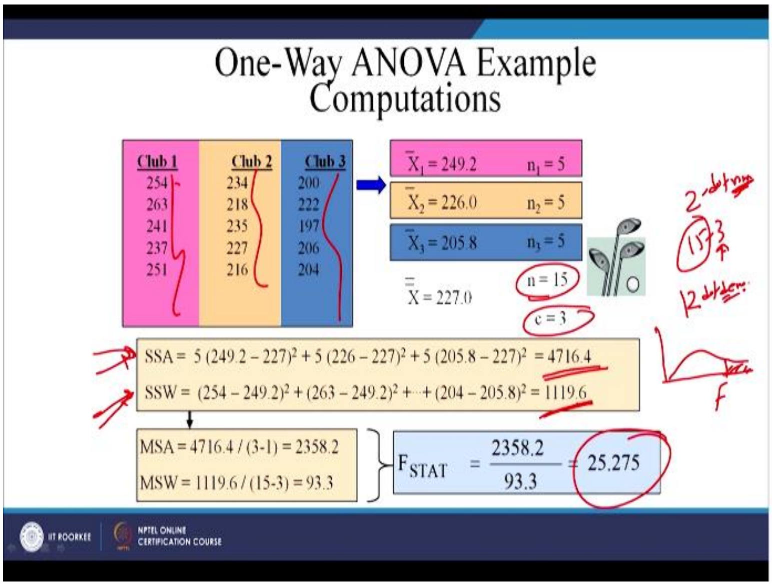
Now here in all these 3 groups, you have got equal number of samples right. So what would be the null hypothesis and what would be alternative hypothesis? So H0 is equal to what, let all these distances are same right, so mu 1, mu 2 and mu 3 are same right and they are not same is alternative hypothesis.

(Refer Slide Time: 15:57)



So these are different data points in this chart, so for first group you have got 254 and the lowest is let us say 241 right. So these are your data points for first group, then for second and for third right. So first of all, calculate mean of each of these groups and then grand mean which is 227 right. So this is nothing but 227 right and X1 bar, X2 bar, X3 bar these are these values right, approximately 250, then this is 226 right and this is 205 right.

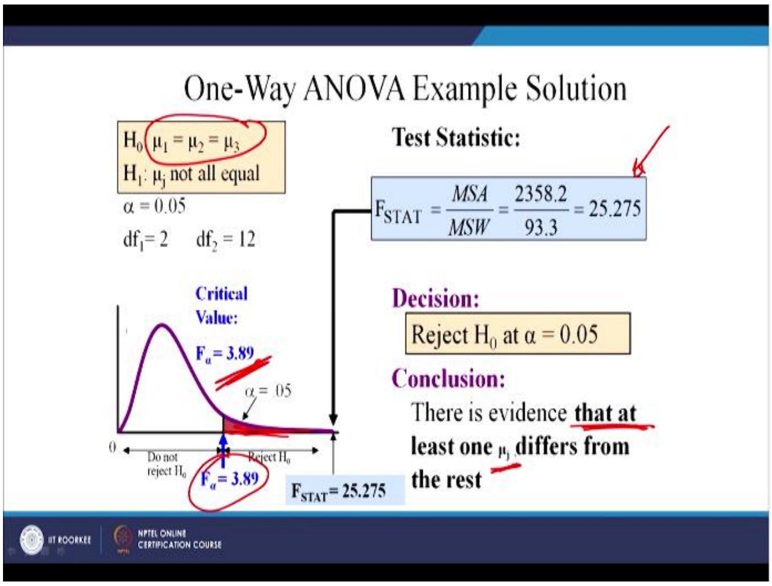
(Refer Slide Time: 16:55)



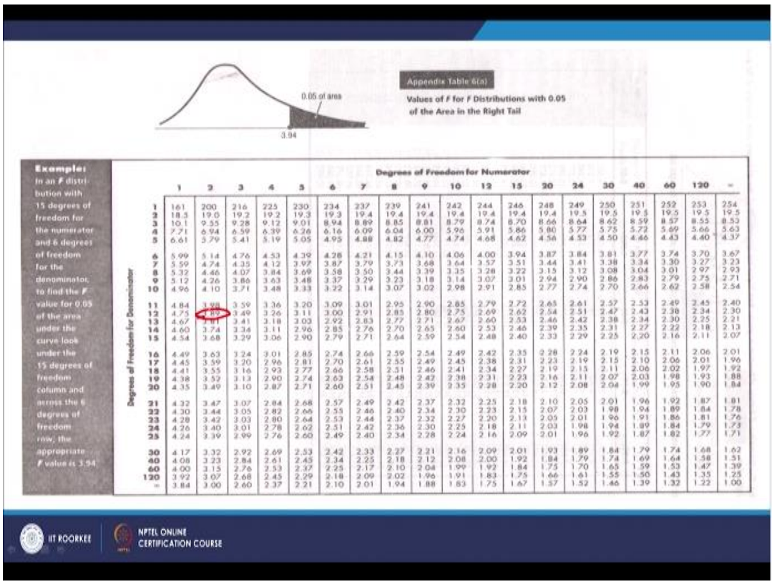
So total n is 15, c is number of groups 3, so among group sum of square is 4716 and similarly within sum of squares or error term is this 1119.6. So F statistics is 25.27, so we have to look at now, we have to look at the F critical value right. So this is your rejection region, so F critical value would be at what degrees of freedom of numerator, so it is c-1 so it is 2 degrees of freedom of numerator and 15-3 right these are 3 is nothing but c, number of groups, 15 total number of points in each of these groups right.

So this is two degrees of freedom numerator and 12 degrees of freedom of denominator right denominator. So let us look at the table value 2 and 12.

(Refer Slide Time: 18:18)



(Refer Slide Time: 18:22)

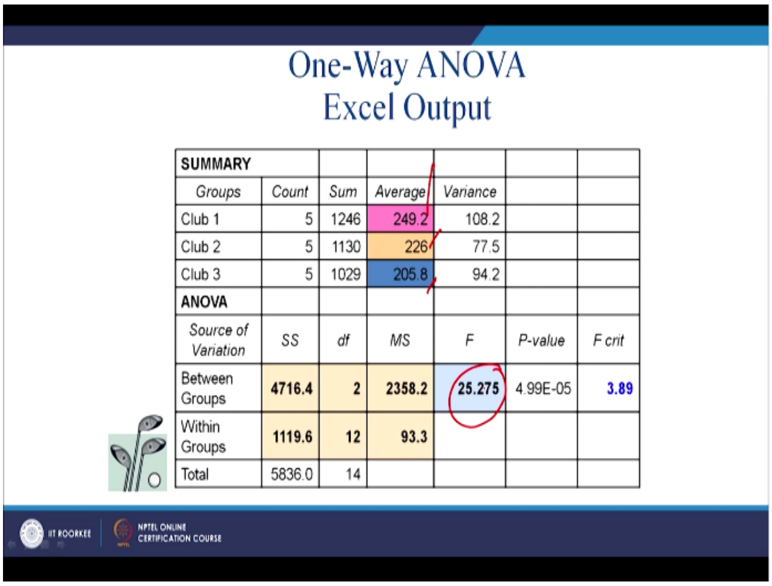


So that value is there in next slide so 2 and 12 this is 3.89. So 3.89 is this right, this is the critical table value, compare it with calculated F value which is in rejection region, so will reject null hypothesis and null hypothesis is that all this mean distances are same. No, they are not same, they are rejecting it right. In fact, if you look at in all these questions which we have worked out in ANOVA, we have rejected null hypothesis so far in all these examples.

So we have always said that the means are not same but we do not know which two means are not same. So that is a disadvantage of ANOVA, it does not tell us which two group means are not same. ANOVA tells only whether there is a significant difference amongst means or not and then there are different methods available to check which two means or which three means or which four means are not same or which different means are different from each other right.

So conclusion is there is evidence that at least one mean differs from the rest of the means okay. So in fact this is the table we have used.

(Refer Slide Time: 20:02)

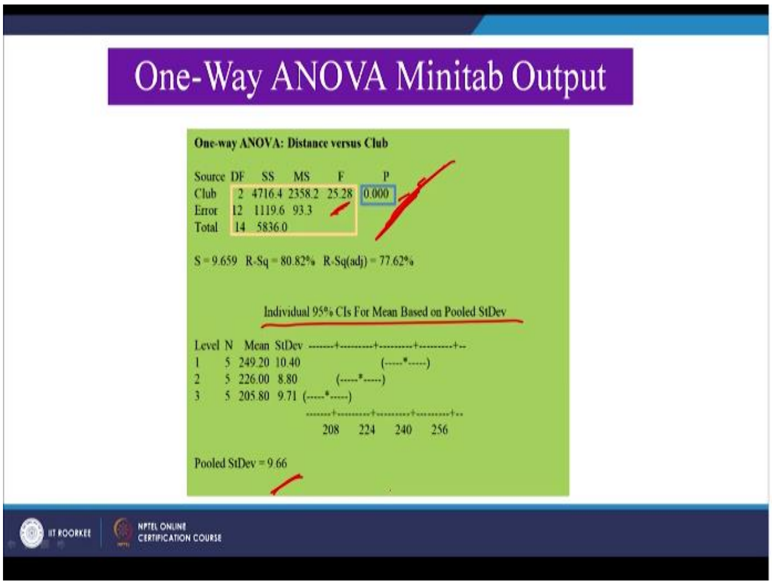


SUMMARY						
Groups	Count	Sum	Average	Variance		
Club 1	5	1246	249.2	108.2		
Club 2	5	1130	226	77.5		
Club 3	5	1029	205.8	94.2		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	4716.4	2	2358.2	25.275	4.99E-05	3.89
Within Groups	1119.6	12	93.3			
Total	5836.0	14				

If you solve this question using let us say Minitab or excel then you should get this answer. So this is the average of first group, second, third at the calculated F value is 25.27. Let us see whether it was there or not. So will wait for will solve this question little later using Minitab. So this is our calculated F statistics 25.27 and the output from excel is same right and F critical is 3.89, this is table value, this is what we have seen.

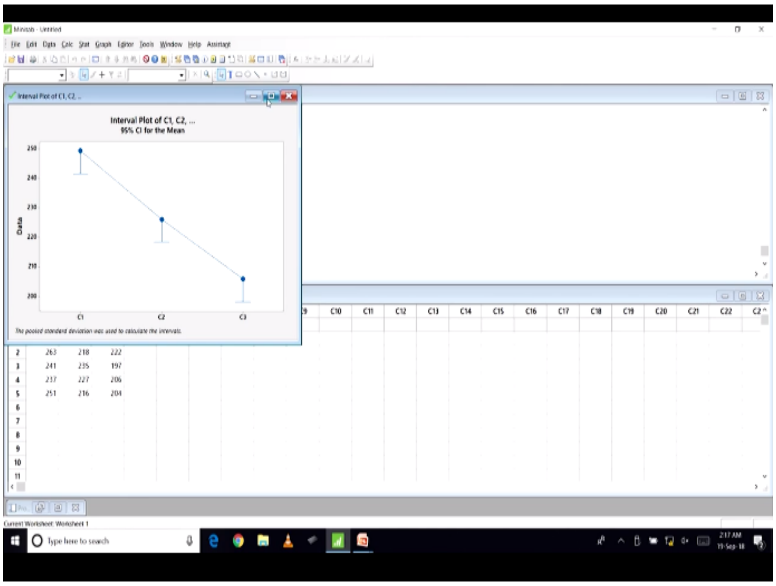
P value, it would be it is approximately 0 right, so 4.99 to the power 10 to the power -5 right. So you will also get let us say this within group and between group variance. Let us check again. So within group variance 1119 and 4716 and this was total 5836.

(Refer Slide Time: 21:30)



This is the output from Minitab, you will get the same answer right P is almost 0, F is 25.28 it is not almost 0 it is 0 right. So these are other values right, so you will also get this the mean and standard deviation of all these groups. So individual 95% confidence interval for mean based on pooled standard deviation right for all these 3 groups right and this is pooled standard division. So let us solve this question using Minitab. This is the question and now let us go for data entry.

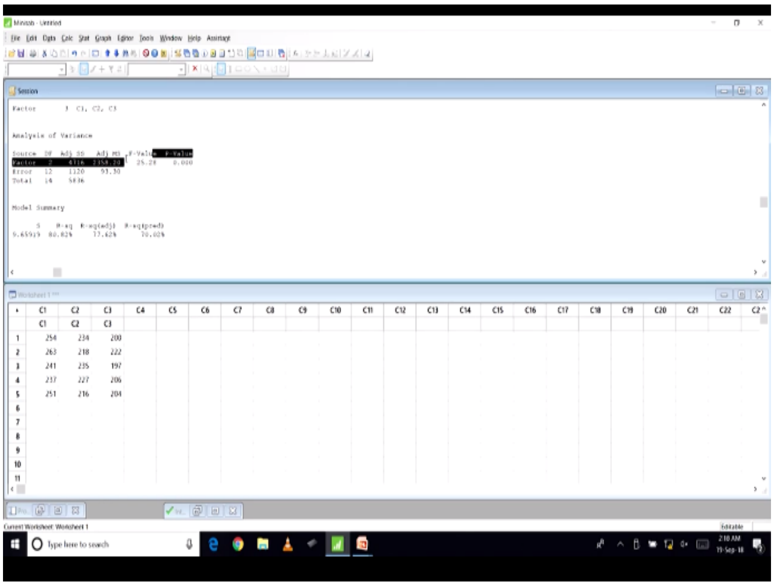
(Refer Slide Time: 22:15)



So will have Minitab software now, so there are now 3 groups right, so we can write club 1, club 2, club 3 or let us say C1, C2, C3. We can delete the remaining values over here. These are the values of previous question. So the first one is let us look at so 254, 263, 241, 237, 251. So these are 5 points for group 1, then for group 2; 235, 227, 216, 200, 222, 197, 206, 204 right.

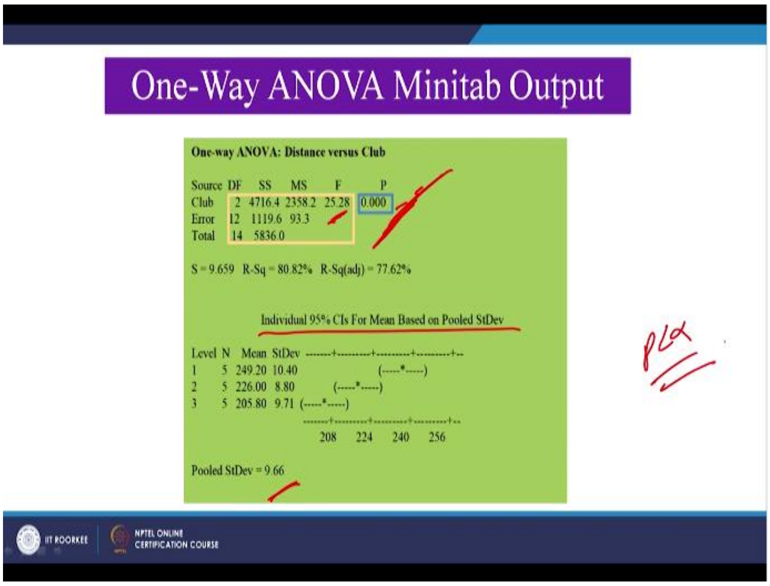
So this is how you should do data entry. Let us look at first we are deleting the output of previous question. So will go to stat ANOVA, one-way ANOVA right, so you can click here, so we will just delete this and select all these three, select option will check at what significance level we have to test it, so we need to test this hypothesis at 0.05 right it means 95, so this is correct which is lower bound, press okay again click okay right.

(Refer Slide Time: 24:23)



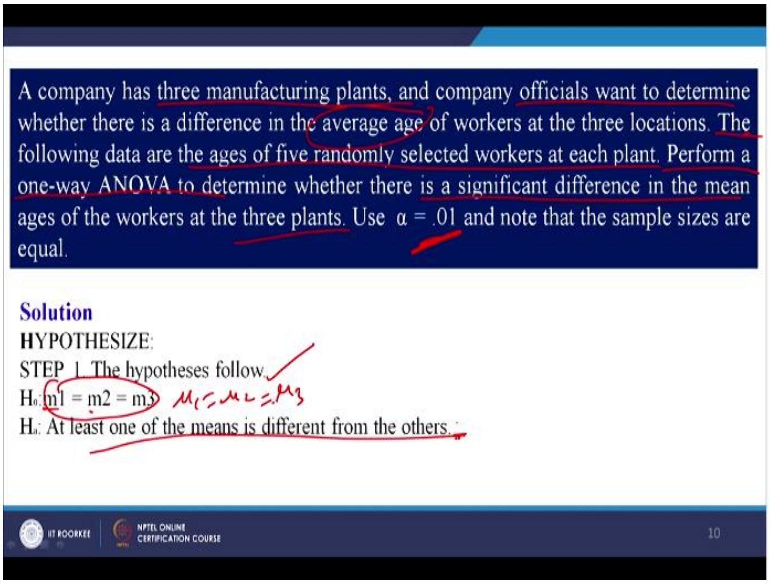
So let us look at ANOVA table which is there up, yeah this is our ANOVA table right. So we just see this table is exactly similar to what I just showed. Let me show once again. This is the ANOVA table and our answer is this right. So this is you have got sum of total square, error variance and explained variance right.

(Refer Slide Time: 25:12)



So this is P is 0 at values 25.28, so will reject null hypothesis because P is <alpha, so will reject null hypothesis right.

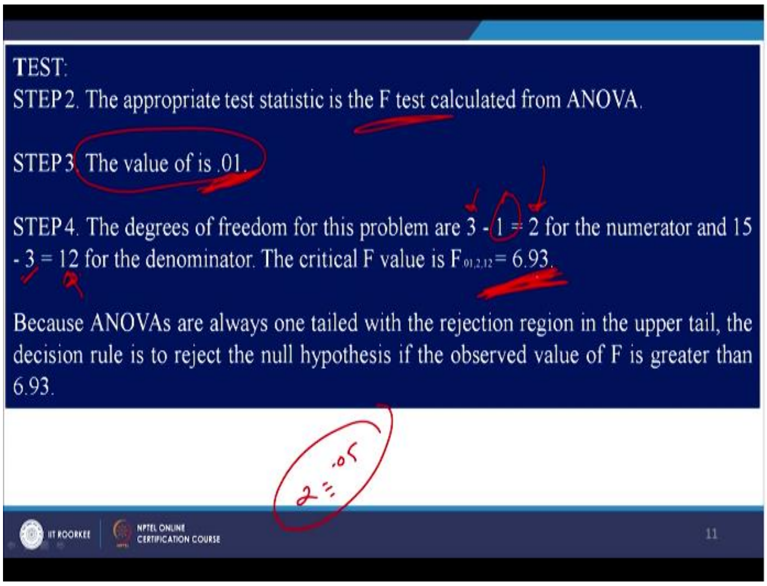
(Refer Slide Time: 25:18)



Now let us look at one more question. Now this is the question on the difference in average workers. So a company has got 3 manufacturing plants, so the company officials want to determine whether there is a significant difference in average age of the workers of those 3 locations where plants are there right. So the following data are the ages of 5 randomly selected workers at each plant.

So we have selected total 15 workers and we noted down their age. Perform a one-way ANOVA to determine whether there is a significant difference in mean ages of the workers at 3 plants given alpha is=0.01 and note that the sample sizes are equal right. So what we will do, step 1 it is mu 1, mu 2, mu 3 right, so this is in fact you can write it like this, mu1= mu 2 = mu 3 right, at least one of the means is different from others right.

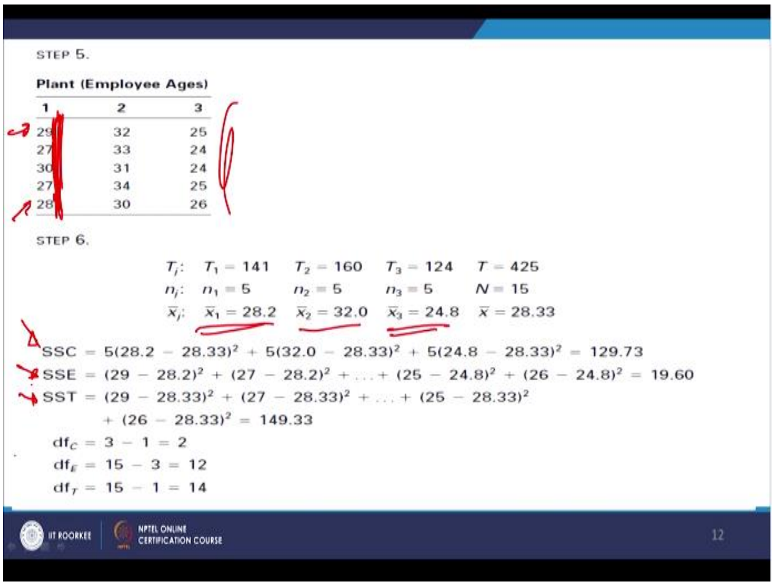
(Refer Slide Time: 26:46)



Calculate F value; find out table value at 0.01 significance level. Now this is someone different from remaining questions we have solved on ANOVA. Here alpha is 0.01, in all earlier cases it was 0.05. So here you will have a different table, different F table for alpha is=0.01. So will look at that table and that value at 2 degrees of freedom numerator and 12 degrees of freedom denominator right.

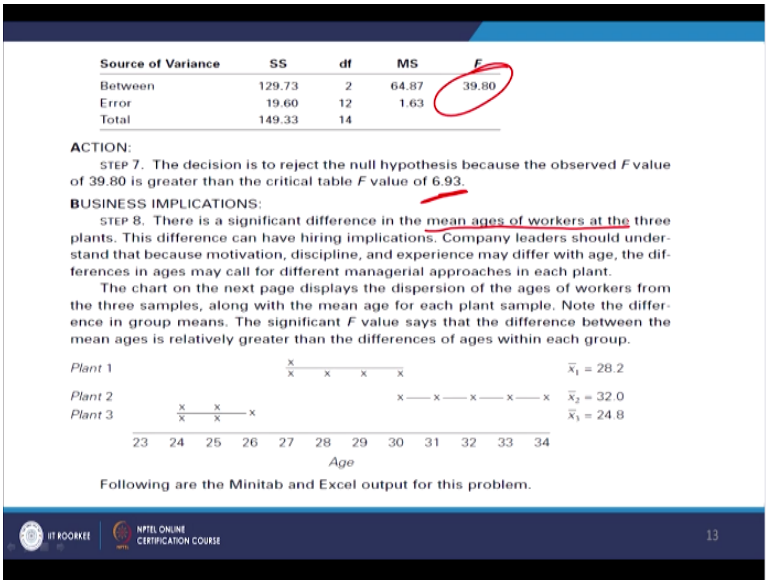
Why 12? This 15-number of groups, this is number of groups-1 right 6.93 and will reject the null hypothesis if calculated F value is more than this right.

(Refer Slide Time: 27:45)



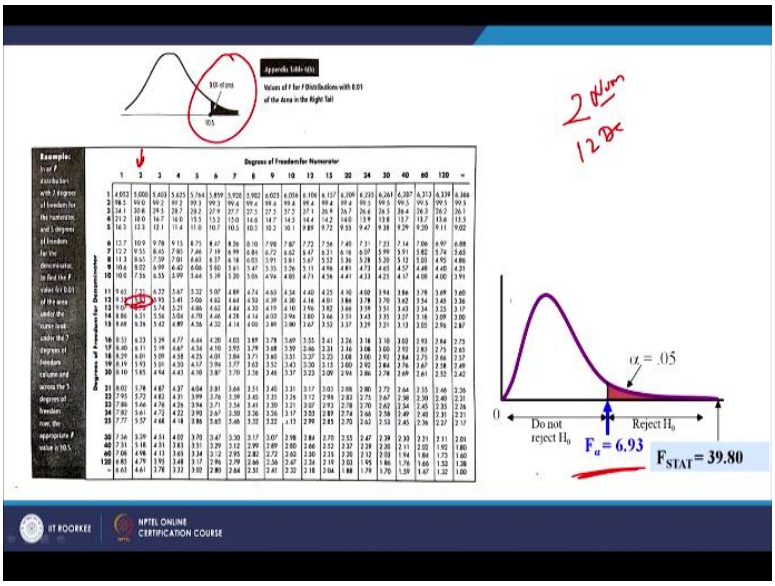
So this is how you can solve this question. So you have got let us say age of first worker in first plant, age of fifth worker in first plant and so on right. So you can calculate mean of these 3 groups right, then grand mean, again this is sum of the square among columns, this is sum of square within columns, this is total sum of square.

(Refer Slide Time: 28:29)



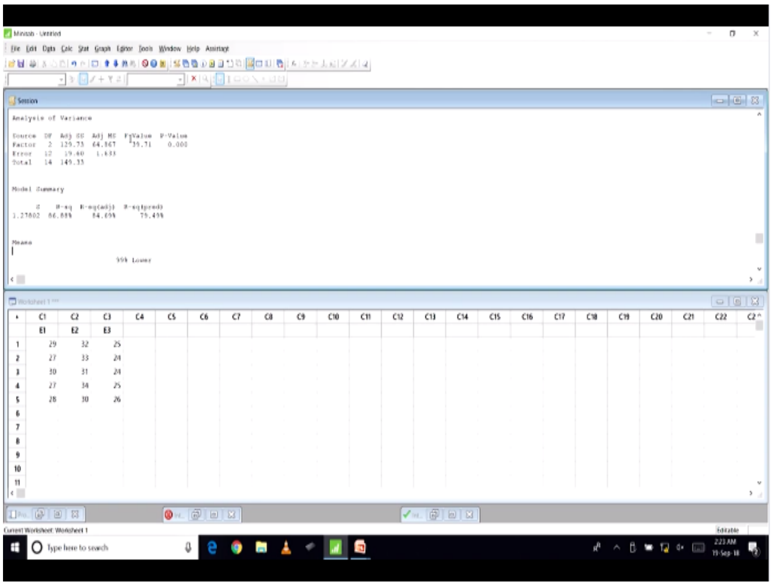
So the calculated F value is this, it is 39.8 and the table value 6.93. So there is a significant difference in the mean ages of workers at these 3 plants right.

(Refer Slide Time: 28:49)



So this is the table when alpha is=0 01, so we have to look at F critical value at 2 degrees of freedom numerator right this column and 12 degrees of freedom at denominator right, so this is the one 6.93 is this right. So this is how you can solve the question like this again using Minitab.

(Refer Slide Time: 29:19)



So will solve this question again quickly, so will delete all these data related to previous question. So let us say these are 3 ages let us say for plant 1, 2, 3. So it is 29, 27, 30, 27, 28. So these 5 points, next 5 points, these are different again ages, workers of plant 2; 25, 26 right. So we will go to okay first we will delete this. Let us look at output so stat, one-way ANOVA.

So select all these 3 and then go to option, here it is 99% right, click okay. So look at ANOVA table so which is here right. So this is what is our calculated value F value is 39.71 and P value is 0. So P is<0 so will reject null hypothesis. So this is how you can work out examples on ANOVA using Minitab. Let me stop here for the time being. In next class, will see some more points related to ANOVA. Thank you.