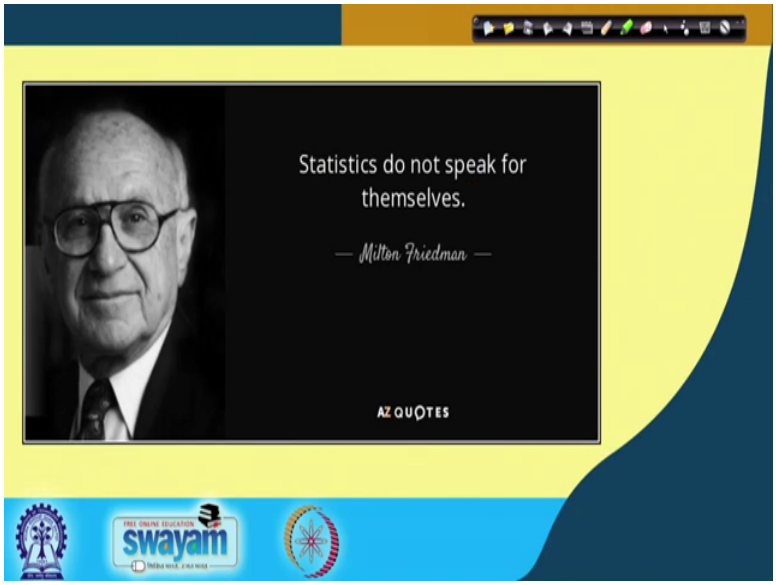
Six Sigma
Prof. Jitesh J Thakkar
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kanpur

Lecture - 35
Two - Way ANOVA

Hello friends, I welcome you to our ongoing journey on Six Sigma. And we have almost reached half way and we are at present discussing the various topics in the analyze phase of DMAIC cycle and as a part of that we have talked about ANOVA one-way ANOVA and its importance and today as a part of lecture 35 we will talk about two-way ANOVA analysis.

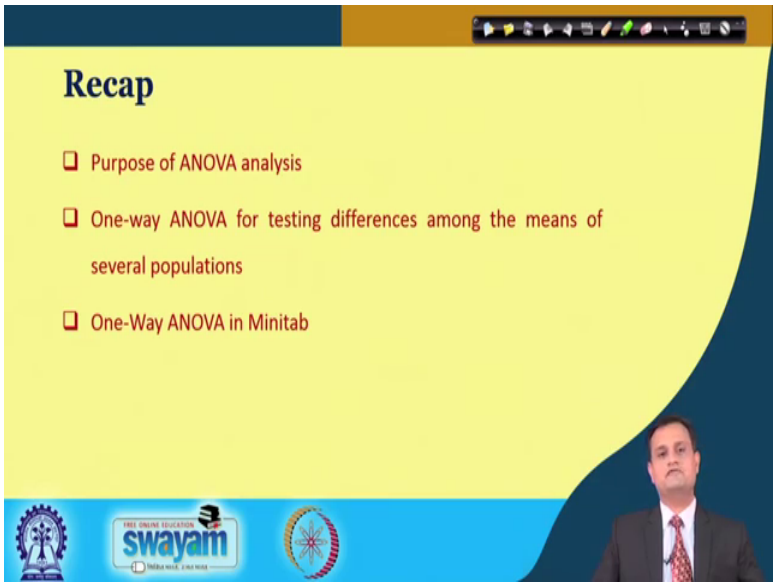
(Refer Slide Time: 00:49)



So, I would like to remind you that statistics do not speak for themselves and we have to interpret them with respect to the real life situation and with respect to the implications of the results on the present and prevailing circumstances.

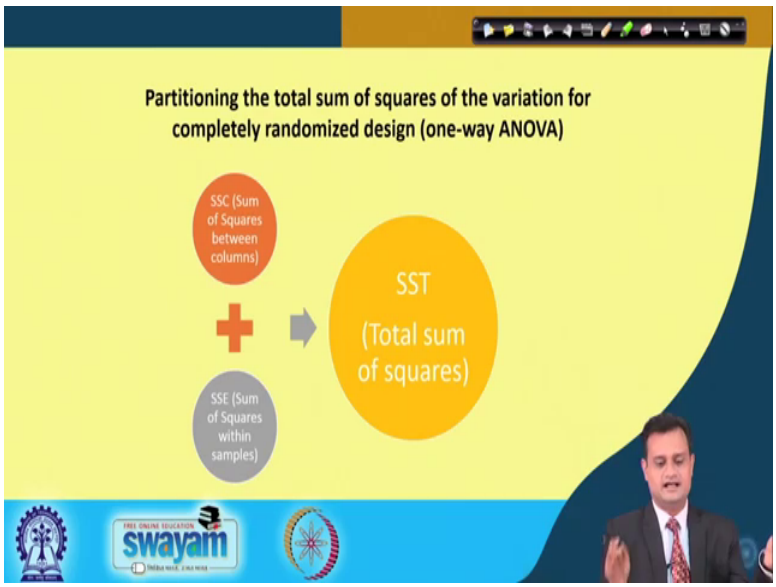
If we can do this then statistics can definitely be used as a competitive weapon by the businesses and manufacturing and service industry.

(Refer Slide Time: 01:18)



So, we have talked about ANOVA its importance One-Way ANOVA, one-way ANOVA with Minitab in the previous lecture.

(Refer Slide Time: 01:28)

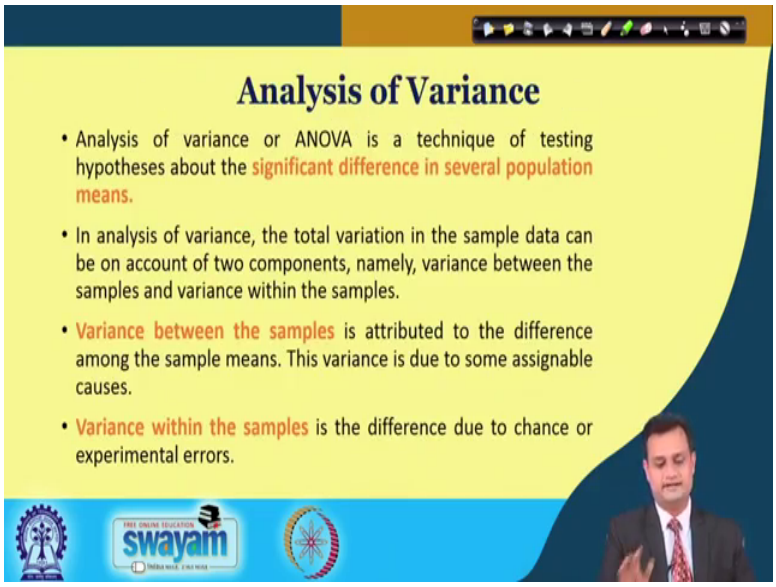


And now in this lecture we will talk about the purpose of two-way ANOVA. Two-way ANOVA factorial design it is called some illustrative examples and then we will see the use of mini tab in conducting two-way ANOVA analysis.

So, if you recall we talked about partitioning the variation and ANOVA as the name suggest analysis of variance, he deals with the variability in the data and this variability

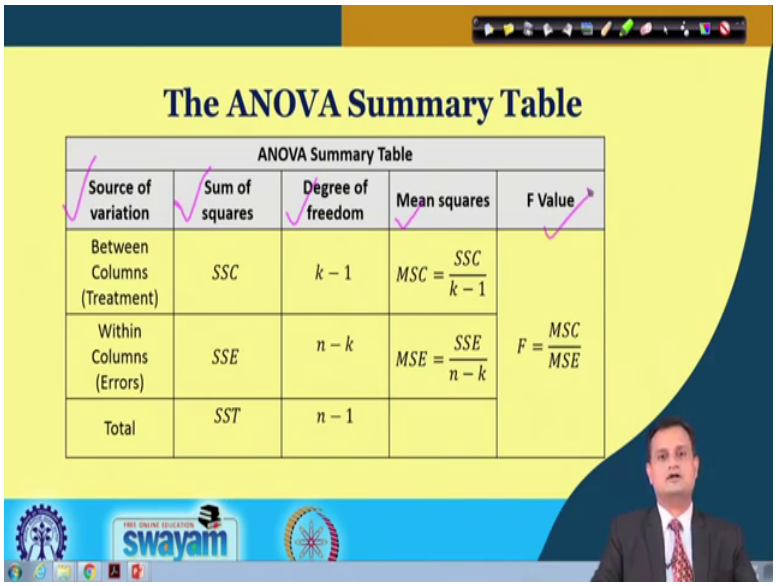
is specific to say among subgroup variability and within subgroup or sample variability. So, you have SSC sum of square between column SSE sum of square within sample so, one is the among and other is the within and this is my total sum of square.

(Refer Slide Time: 02:22)



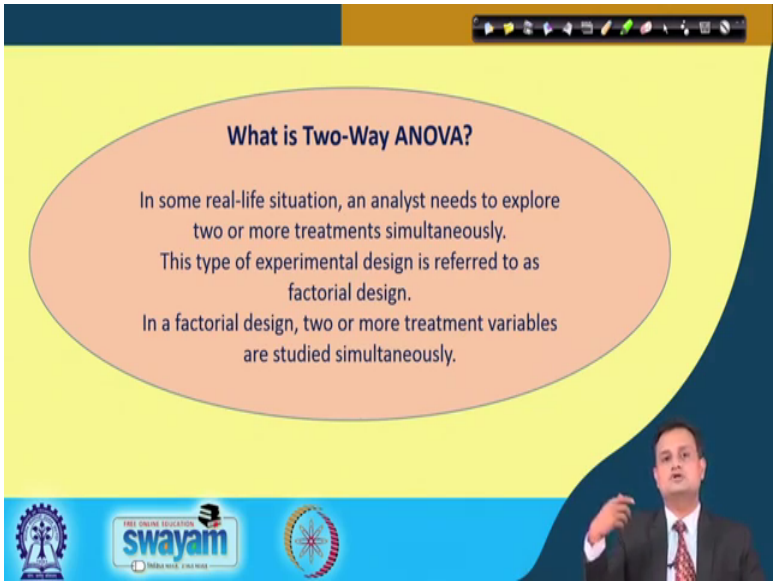
So, typically analysis of variance we conduct, because conducting multiple t test is not and advisable say strategy when you have to compare more than two to three mean for the various population. And if you go for let us say four mean you want to compare you go for four c 2, then such 6 comparison by using t test will have very high type one error and you will end up with the wrong conclusions. So, you make use of analysis of variance with simultaneously analyses the data with respect to within variability as well as among variability and helps us to make the better and sound inferences.

(Refer Slide Time: 03:14)



So, this is typically ANOVA table to remind you that I have sources of variations, I have sum of squares degree of freedom, when you divide sum of square by degree of freedom mean square and then you compute or calculate the F value you have the tabulated F value for a given degree of freedom. And then you compare it and as per the decision rule, you either except the null hypothesis or you reject it and you try to verify your claim. So, this is what we had seen.

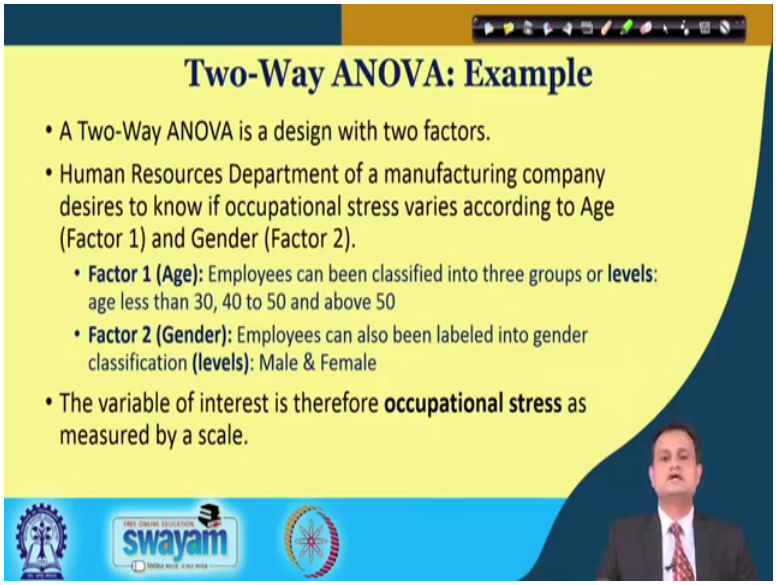
(Refer Slide Time: 03:53)



Now, let us see what is two-way ANOVA analysis so, in one-way you had seen that I am only interested to explore maybe one treatment, one particular say effect of one particular factor on my response. And that is where I conducted the one-way ANOVA analysis. Just think that your response variable or dependent variable is productivity and I am considering the machinery technology has one of the variable and I am trying to analyze its impact on the productivity it is ok, and I may do it by conducting one-way ANOVA analysis.

But you can definitely think that productivity is not solely dependent on the machine, it also depends on the skill level it also depends on this shift supervision and many other factors. So, many a times you are interested to analyze the two or more treatment simultaneously and this is where your two-way ANOVA or factorial design comes in picture.

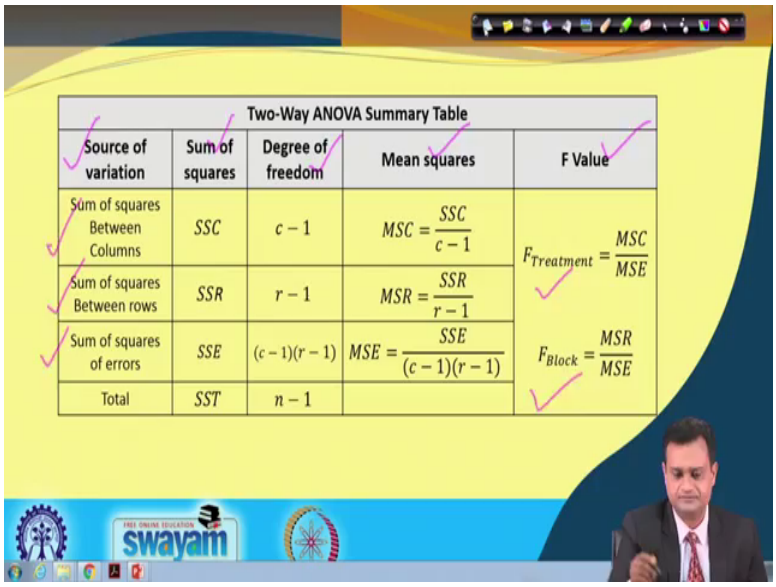
(Refer Slide Time: 05:08)



So I can give you the example to make it better clear, let us say there is a human resource department of a manufacturing company, and they want to analyze that what is the impact of two factors 1 is let us say your age and factor 2 is gender on occupational stress it is a very important topic. And let us say they consider 2 factor so, young people they may have be able to accept more say stress or able to digest, but maybe with the age that capacity may go down and as well as gender could be the factor.

So, we want to check our claim that to what extend say this factor 1 and factor 2, they contribute to the occupational stress. So, what I have done that factor 1 I can divide it further into some subgroups or categories. So, I will say 30 40 to 50 and above 50 these are 3 my say groups levels and same way I can say factor 2 that is gender male and female. And then I can conduct the ANOVA analysis to check the impact of this factors on the occupational stress.

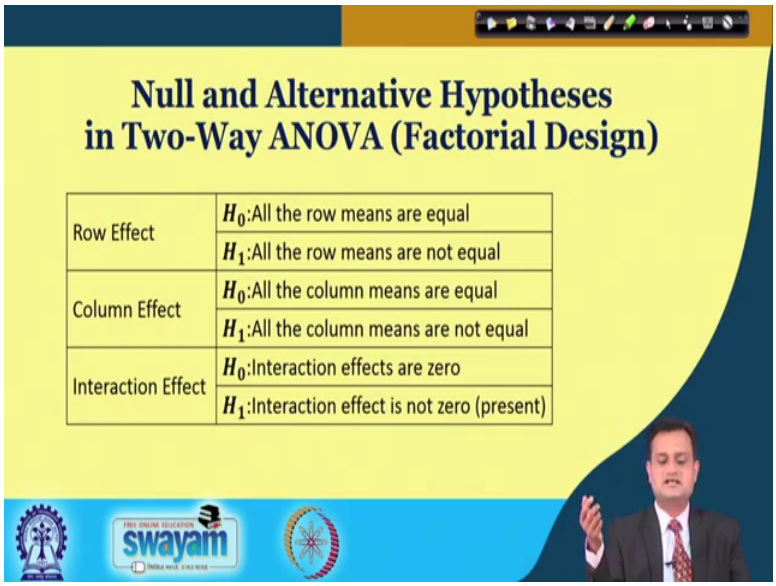
(Refer Slide Time: 06:24)



Now, typically when you refer the two-way ANOVA table, it is same like what we have done in the one-way ANOVA sources of variation sum of square degree of freedom mean square and F value, but here because you have 2 factors to be analyze simultaneously you have row effect as well as column effect.

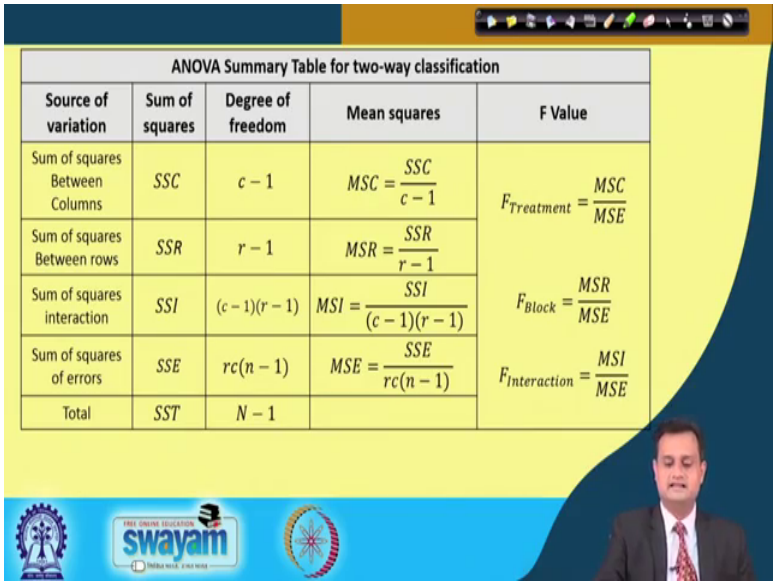
So, you will have sum of square between columns and sum of square between rows. So, you are interested to analyze among as well as within variability and you have sum of square errors. So, basically you will have F treatment F block and this you would like to analyze for conducting your two-way ANOVA analysis. So, let us see how this can be done ok.

(Refer Slide Time: 07:17)



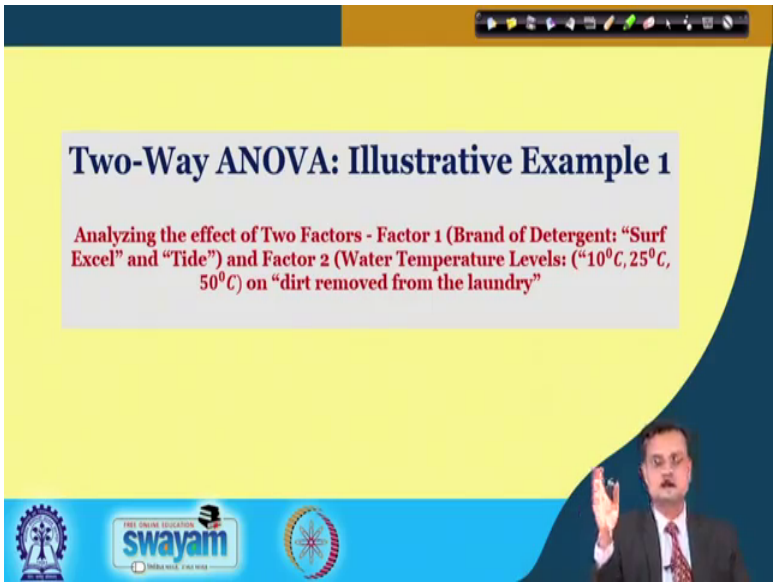
So, here my null and alternate hypothesis would be like this row effect I would say null hypothesis all the row means are equal, alternate hypothesis I will say all the row means are not equal column effect same way, I will say all the column means are equal alternate all the column means are not equal or at least one of them is not equal. Interaction effect there are 2 factor so, there could be interaction between these 2 factors possible.

(Refer Slide Time: 07:53)



So, interaction effects are 0 interaction effects is not 0. And this is how I can construct my two-way ANOVA table and compute F treatment F block and the F interaction.

(Refer Slide Time: 08:03)

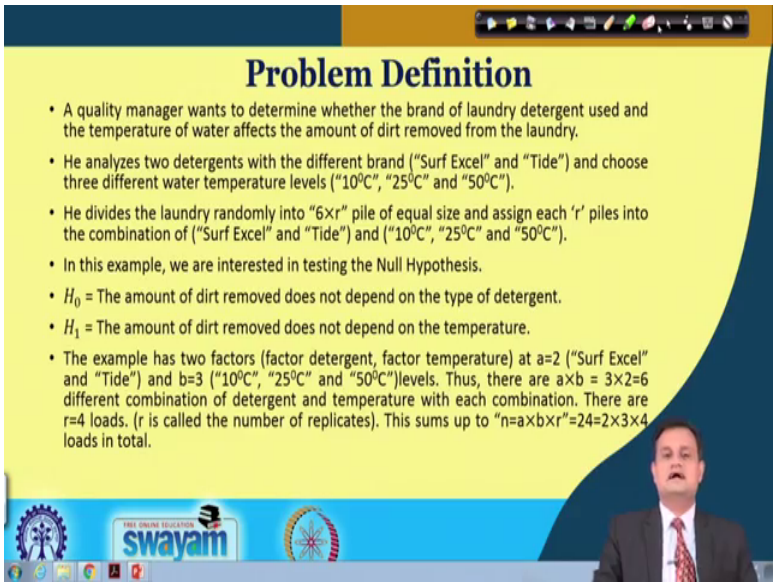


So, now let us see the illustrative example and that will help us to appreciate, that how the two-way ANOVA analysis can be conducted manually with the use of software and mainly we are interested in interpreting the results and making the necessary recommendations for the case organization. So, here the situation is like this that I want to analyze 2 factors number 1 factor 1 brand of detergent Surf Excel and Tide. So, many a times you get confuse that should I purchase Surf Excel or Tide and you compare the price and what not.

Factor 2 what a temperature levels 10 degree 25 degree 50 degree and dirt removed from the laundry. Now, say here this is not only specific two factor 1 that is the choice of your detergent, but nowadays you can have the washing machine, where you can set the temperature of the water and here you would like to really understand that if water temperature is varied say I set it at the normal or maybe medium hot or maybe hot, then will it have any impact.

So obviously, when you are purchasing a technology with advanced feature, you have to pay more and similar way when you go for a particular brand that is factor 1 you may have to pay more. So, here I would like to first understand through ANOVA analysis, that to what extend factor 1 factor 2 and their interaction is significant for a given level of significance and if it is not, then I will ignore that particular source of variation.

(Refer Slide Time: 09:51)



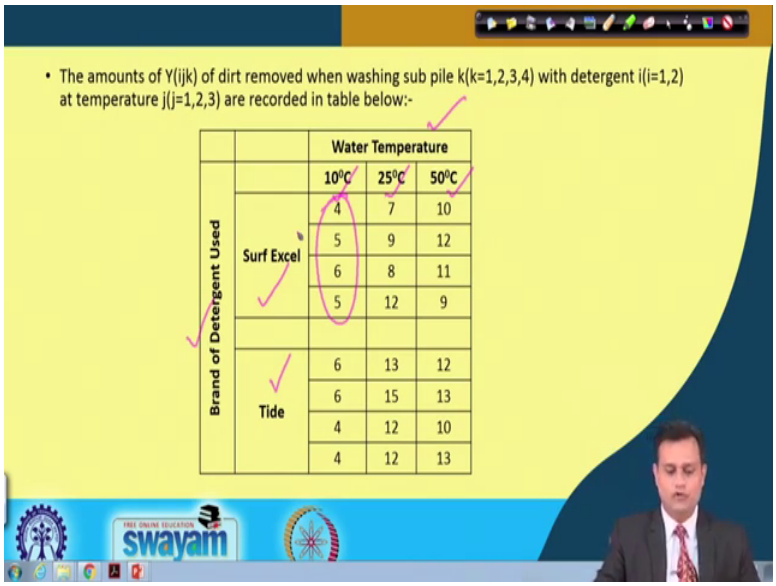
So, as I mention the situation is like this, you have 2 detergent and you as a quality manager want to analyze the impact of factor 1 that is brand of detergent Surf Excel versus Tide and the temperature I am selecting here three temperature levels one is water temperature levels, 10 degree 25 degree 50 degree. So, 10 degree may be more or less room temperature normal tap water 25 degree centigrade maybe your medium hot and 50 degree centigrade hot temperature.

I have null hypothesis that the amount of dirt removed does not depend on the type of detergent, alternate hypothesis the amount of dirt removed does not depend on the temperature. So, you can verify that whether my temperature is important or my detergent is important or for each particular factor, you can separately set null and alternate hypothesis and then you can conduct the ANOVA analysis.

So, here just try to appreciate some notations that I will say a is equal to 2, because I have to brand detergent Surf Excel and Tide, I will say b is equal to 3 10 degree 25 degree 50 degree centigrade and I can have a into b that is 3 into 2 is equal to 6 and there are say r is equal to 4 loads.

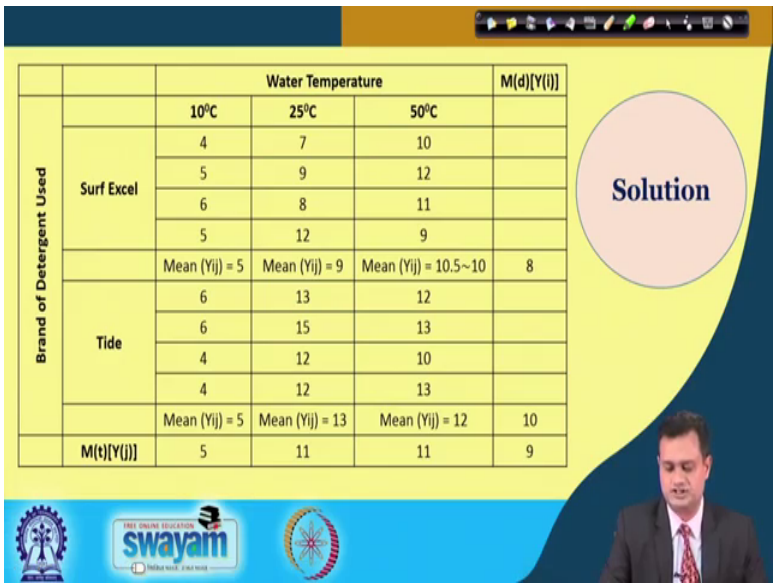
So, I am trying with the different combination and you can have total number of observation that is a into b into r so, 2 into 3 into 4 that is 24. So, I would just like to say that this is just an hypothetical example and you need not to say make any strong claim about Surf Excel or Tide it just a learning exercise. So, let us see what happens actually.

(Refer Slide Time: 11:53)



So, what you can see here that I have the water temperature 10 degree 25 degree 50 degree and I have the brand of detergent Surf Excel and Tide. And for each particular combination I have selected 4 readings. So, I have 4 readings for each particular say combination and this is how I have 4 into 6 that is 4 into 3 into 2 and this is 24 readings data set in my table.

(Refer Slide Time: 12:28)

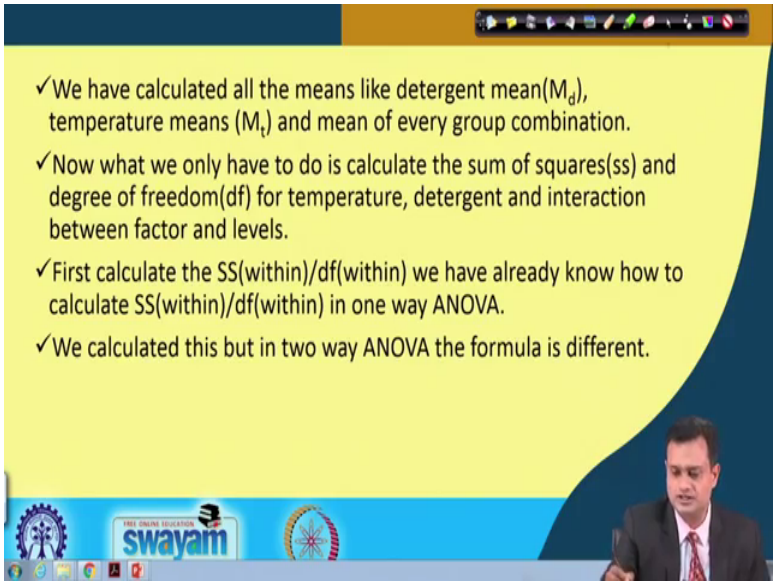


So, now I am just doing little computation in order to show you the manual calculation of two-way ANOVA analysis, we have already seen the manual calculation of one-way

ANOVA analysis. And now here say I am just extending my table by considering the mean of this that is at 10 degree centigrade Surf excel mean and then I have mean of this mean of this and mean of this.

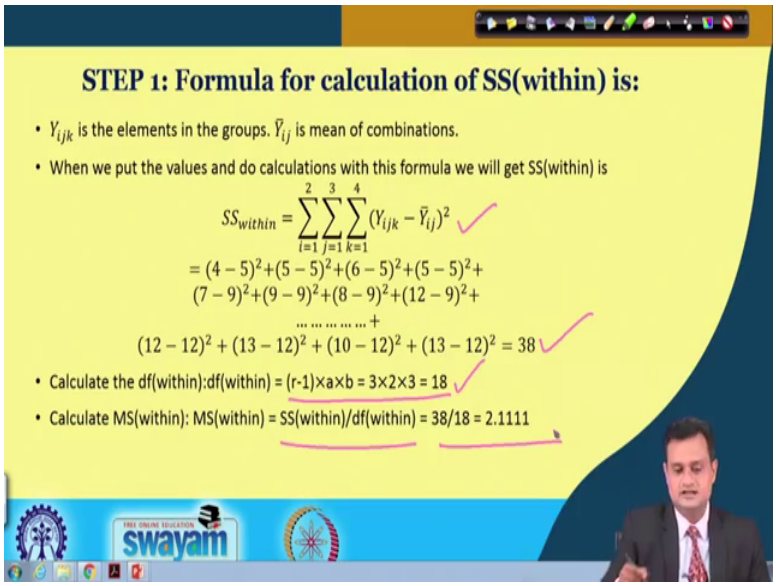
So, I am just computing mean for all and I am also computing the mean for particular say column. So, 5 plus 5 divided by 2 11 plus 13 divided by 2 so, it is 11 10 plus 12 divided by 2 it is 11, I am also computing the average of this that is 9 plus 5 plus 10 divided by 3 so, it is 8 and this is 10 so, I have just extended my table with the computation of mean.

(Refer Slide Time: 13:36)



Now, we have calculated all the means like detergent mean M d and temperature mean M t. So, I am just using this particular say suffix to make you easy it make the understanding easy that M d and M t stands for mean for different temperature, now that we only have to do it calculate the sum of square as we do in ANOVA analysis. And then calculate SS within divided by degree of freedom and we have already known the calculated SS within degree of freedom in one-way.

(Refer Slide Time: 14:23)

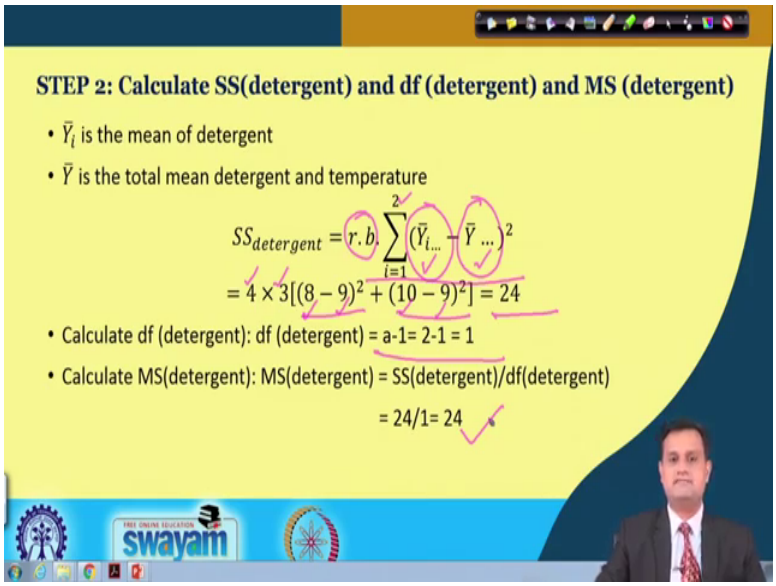


So, similar kind of approach we will use to compute, the various SS within SS between or among and then the other possibilities. So, here what I say that I have Y i j k and you will see there is a summation of 3 sigma.

So, i j k means you have say detergent you have particular brand of detergent you have particular temperature. So, temperature columns are there, you have detergent specified in the row and you are referring to particular cell or particular say reading so, I will say i j k. So, I have i is equal to 1 to 2 j is equal to 1 to 3 and k is equal to 1 to 4, because in each particular setting let us say Surf Excel and temperature 10 degree I have 4 reading so, k is equal to 1 to 4 I have 2 brand so, i is equal to 1 to 2 and I have 3 different temperature levels or setting so, j is equal to 1 to 3.

So, this is how i interpret once you have interpreted this then just plug in the values into this particular expression and that will give you say 38. So, calculated degree of freedom is r minus 1 into a into b, I have already explain how to compute the degree of freedom and this is my 18 so, calculate MS within this is SS within divided by degree of freedom and this comes out to be 2.11.

(Refer Slide Time: 16:09)

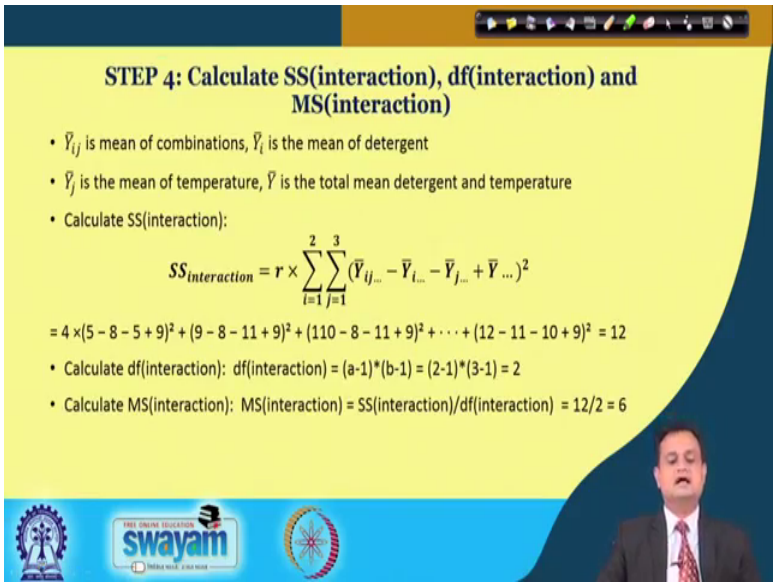


So, now, same way we can go for the analysis of SS that is detergent and I will say r into b sigma i is equal to 1 to 2 because there are 2 detergent Y i triple dot bar minus Y bar triple dot or Y triple dot bar square. So, here I will consider the grand mean as well as the individual mean. So, you just see this is my 4 this is my 3 and I have taken 8 minus 9 10 minus 9 and this will give me 24.

So, for this you can just go back to that slide our table and see that how these Y i bar and Y Y triple dot bar, they are basically inserted in this particular equation. So, you have degree of freedom a minus 1 we have 1 less possibility to maneuver that is the degree of freedom we have seen and 24 by 1 is 24 so, now I have calculated MS detergent that is my 24.

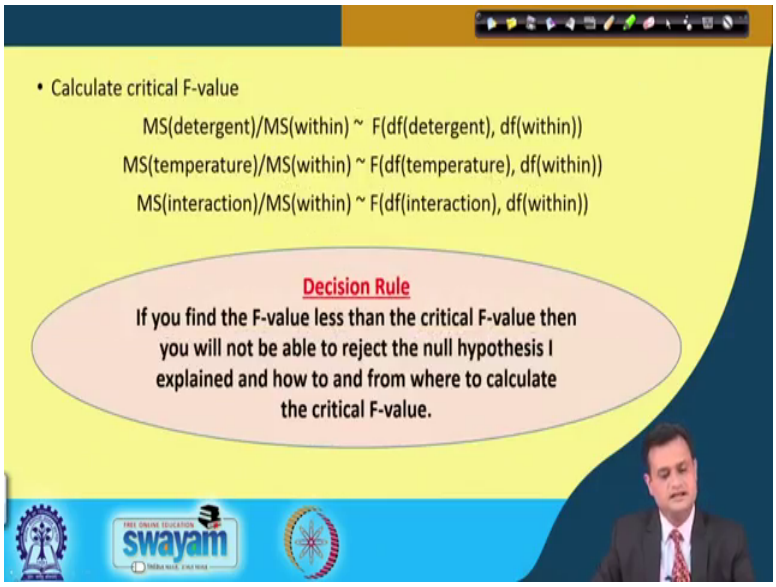
So, now, with this particular calculation we can go ahead and we can also compute SS temperature. So, SS temperature previously we took Y i triple dot now temperature varies in columns. So, I will have Y j triple dot bar and this will be same Y triple dot bar that is the overall say mean your grand mean, this is 9 and now you have 192 divided by 2 comes out to be 81.

(Refer Slide Time: 17:57)



So, you have basically degree of freedom for SS.

(Refer Slide Time: 18:02)



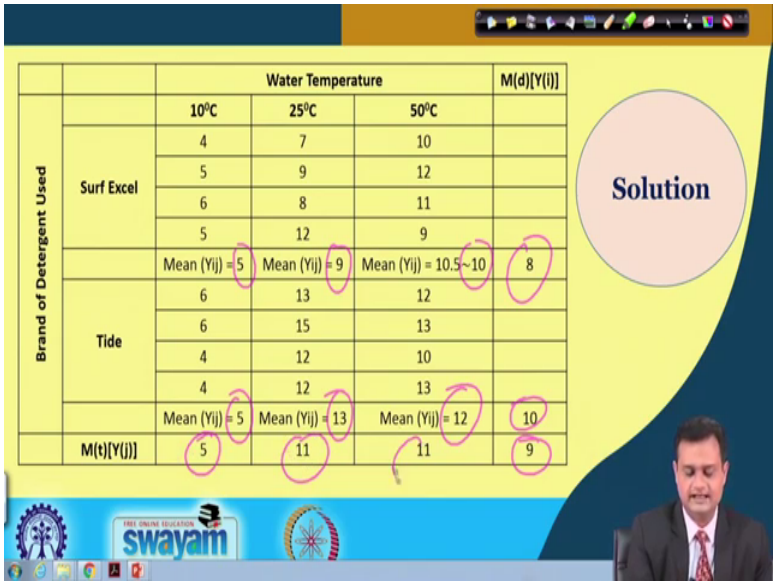
I can just go back to help you SS temperature, SS detergent, now you have SS interaction.

So, here I cannot neglect because a particular temperature a particular detergent, they can have better effect on dust removal capability and I am also trying to see, that whether this combination of a particular detergent and temperature it really helps me to have better

cleaning better dust removal. So, I will find SS interaction you can use this particular expression.

So, r and you will say that is i and j 2 sigma i is equal to 1 to 2 I have 2 brand of detergent 3 level of temperature, you will say put your expression like this Y i j triple dot bar minus Y i triple dot bar minus Y j triple dot bar plus Y bar. So, you will have here the values of you can plug in the values here and you will have the calculated value 12 degree of freedom 2 12 divided by 2 will be 6. So, please refer the table where I have put all the mean and values and just try to say see that what is this Y i j bar, this is 5 here it is nine here it is 110 and likewise you are just trying to say put the different values in this expression.

(Refer Slide Time: 19:47)



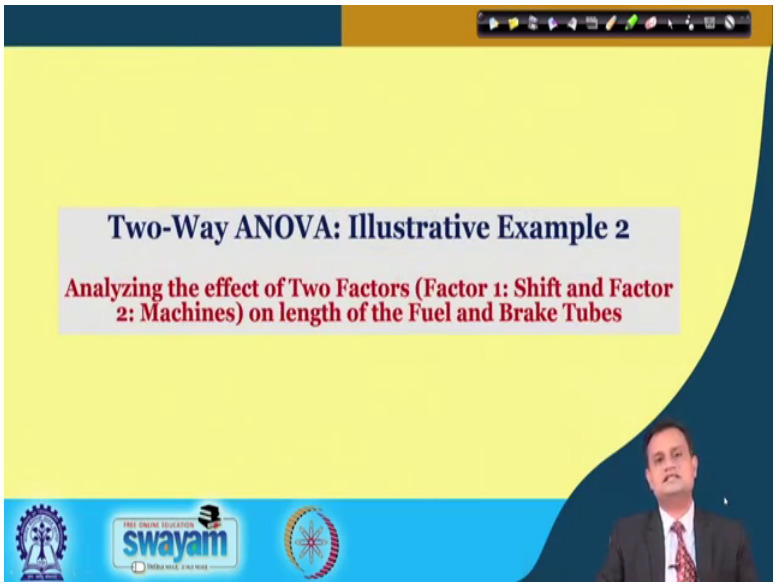
So, if we just a little bit go back then so, now you can see very well here that you have the values this 5 sorry this is your 5, you have 9 you have 10 you have 5 you have 13 you have 12 and you have 8 as well as you have 10 for this is your 9 that is the overall and you have 5 11 and 11. So, you please make use of this and then you try to plug in the values in the given equations so, that you can easily carry out the computations for this particular say analysis. So, I think this is 11 so just correct it ok.

Now, let us go ahead MS detergent by MS within is basically F degree of freedom detergent degree of freedom within, this is how you will find the critical value MS temperature or a MS within is basically you will have degree of freedom temperature

degree of freedom within, you find the F value from the table and similar way for interaction MS interaction divided by MS within you find the F value now.

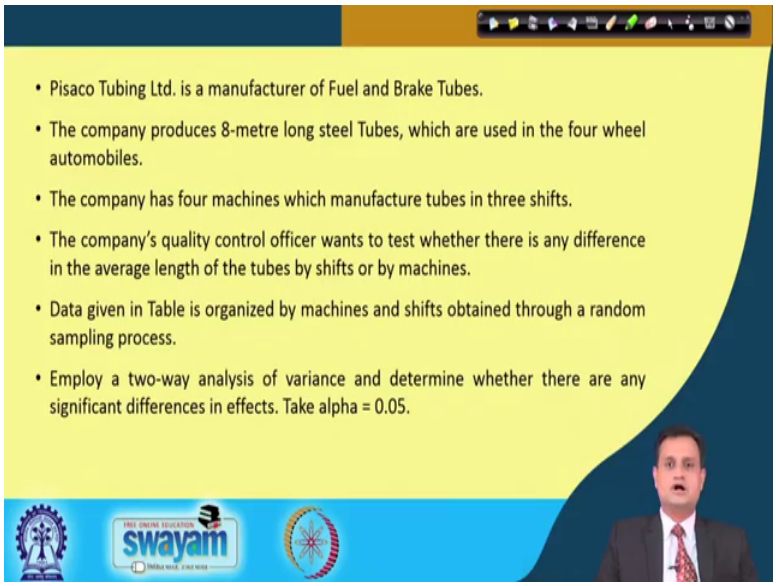
The decision rule we all know it is very simple, if you find the F values less than the critical value, then you will not be able to reject the null hypothesis and this is what we discussed previously also and how to and from where to calculate the critical value that is also discussed.

(Refer Slide Time: 21:34)



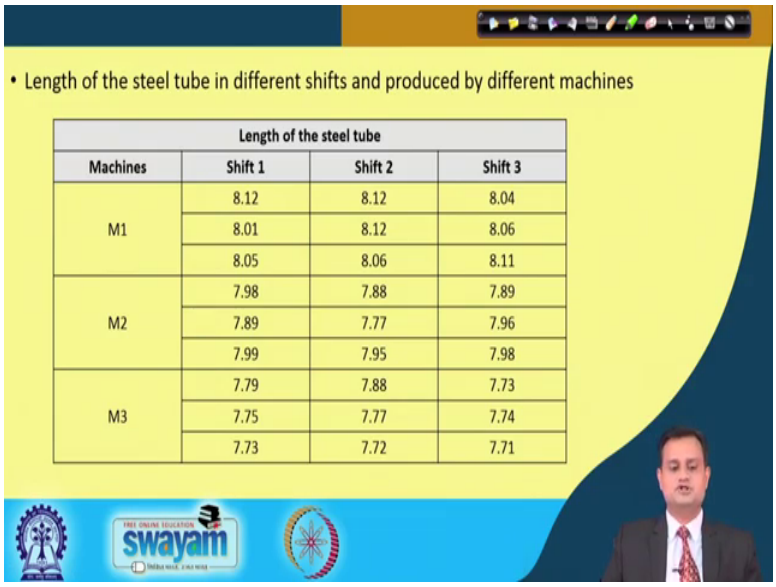
Now, let us see the illustrative example two here I will not do the manual calculation, but I will try to interpret the results and before that I will explain the situation. So, that you can appreciate the context and where you can apply the two-way ANOVA analysis. So, here the situation is like this I have 2 factors factor number 1 shift and factor number 2 machines and I want to see the effect of these on the length of the fuel and brake tubes.

(Refer Slide Time: 22:08)



So, there is a company called Pisaco tubing limited and this company manufactures fuel and brake tube and they feel, that there is could be impact of factor 1 that is your factor 1 and there could be impact of factor 2.

(Refer Slide Time: 22:29)



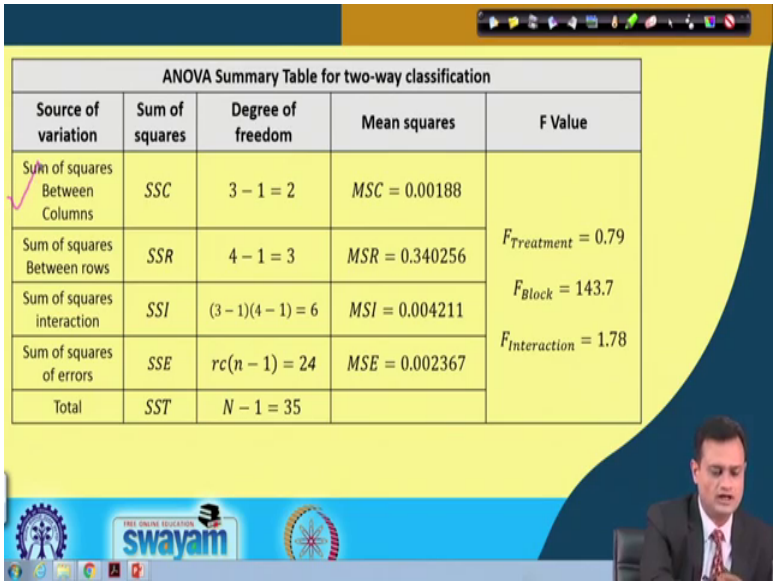
Length of the steel tube			
Machines	Shift 1	Shift 2	Shift 3
M1	8.12	8.12	8.04
	8.01	8.12	8.06
	8.05	8.06	8.11
M2	7.98	7.88	7.89
	7.89	7.77	7.96
	7.99	7.95	7.98
M3	7.79	7.88	7.73
	7.75	7.77	7.74
	7.73	7.72	7.71

So, you have basically factor 1 shift and factor 2 machines. So, they want to analyze the impact on this, now they have collected some data about the length of the tube and what you can see here that you have shift 1 2 and 3 and you have machine 1, machine 2, and

machine 3 and this is basically length of the tube this tube is used for say braking and fuel system in the four wheeler automobile.

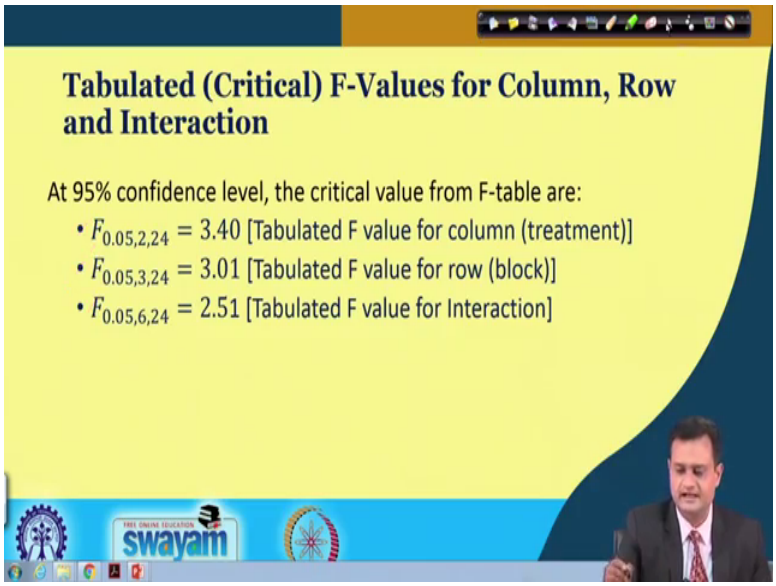
So now, based on the variability analysis I want to see that what is happening with respect to sum of square between column sum of square, between row sum of square interaction and sum of square error.

(Refer Slide Time: 23:14)



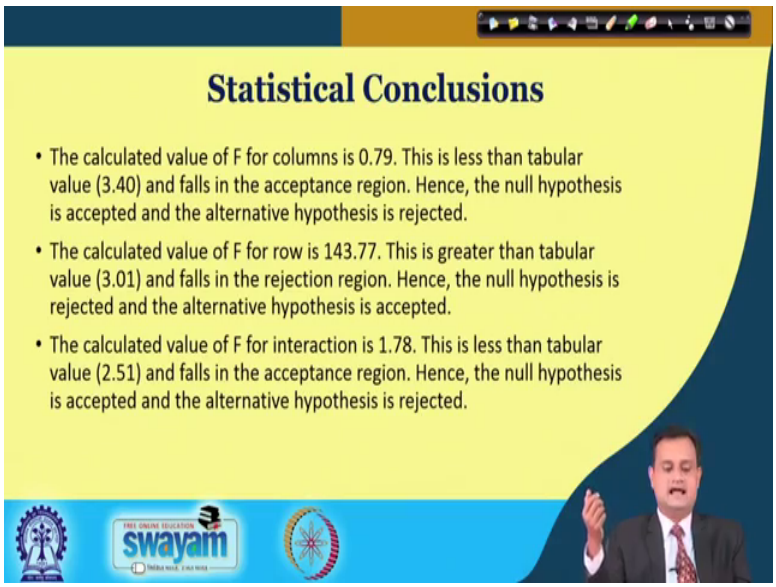
So, this is how you compute the degree of freedom and you have basically MSC you are dividing SSC divided by degree of freedom, so here I am not going for the manual calculation I am just trying to interpret the results and then you have computed F treatment F block and F interaction.

(Refer Slide Time: 23:47)



So, now with this what is to comment so, I can have the interpretation, once I have the critical value taken from the table for a given level of significance. So, I am taking fall alpha significance level 0.05 5 percent. So, for 95 percent confidence level I have found the tabulated value, that is F 0.05, 2, 24 this is 3.4 F 0.05, 3, 24 2.51 F 0.05, 6, 24 this is 2.51. So, these are the tabulated values.

(Refer Slide Time: 24:25)



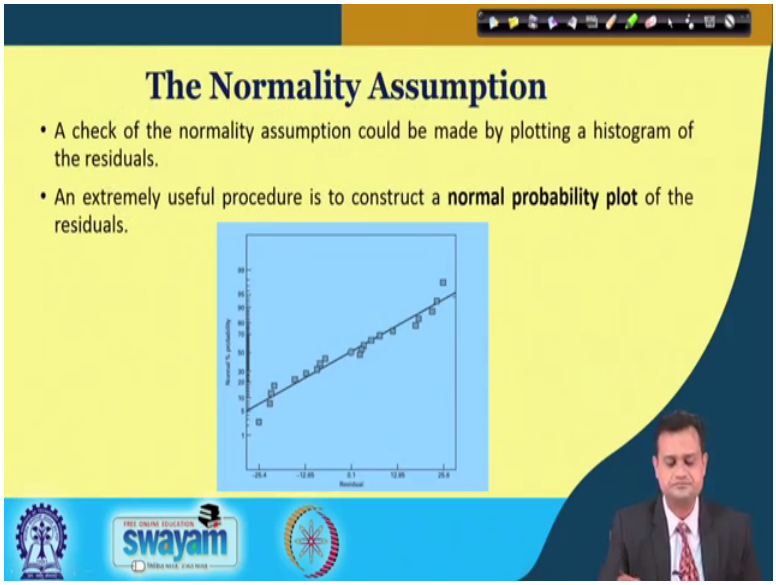
And I already have the calculated value so, the analysis is like this the calculated value of F for columns. So, here your columns are shift is 0.79 this is less than the tabulated value

and falls in the acceptance region so, null hypothesis is accepted, it means that the length produce in shift 1 shift 2 and shift 3 are equal and there is no difference.

Then calculated value of F for row is 143.77 and this is greater than tabular value 3.01 so, it falls in the rejection region it means I am accepting the alternate hypothesis. So, null hypothesis is rejected and alternate hypothesis is accepted here your row is basically machines. So, yes machine to machine there is some difference.

And calculated value that is F for interaction is 1.78 so, this is less than the tabular value and falls in acceptance region. So, I will say that no interaction is possible or prevailing and I accept the null hypothesis. So, now as I said that you cannot accept the model just like that and the hence the results you have to check the model for its adequacy.

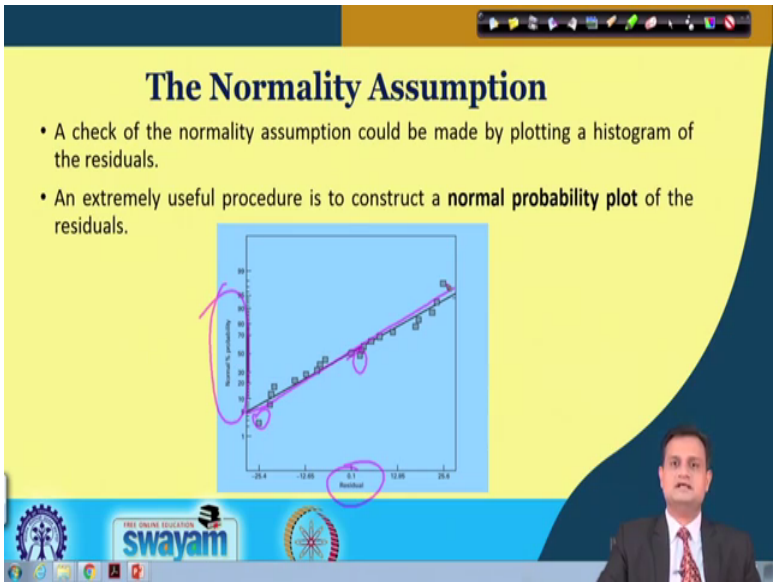
(Refer Slide Time: 25:42)



So, in order to check the model adequacy we do the analysis of error or residue, we have we have done it in detail in the regression analysis. So, here e i j is basically the difference between the observed value and the predicted value y i j is the observed value minus y i j high at is the say estimated value or the predicted value and I will try to check the difference between these two so, when I put it into my y i j. Basically is mu hat plus tau i hat. So, tau i hat is basically refers to my treatment effect and this could be say further re written as mu hat could be estimated as y double dot hat and tau i hat could be expressed at y i dot hat minus y double dot hat.

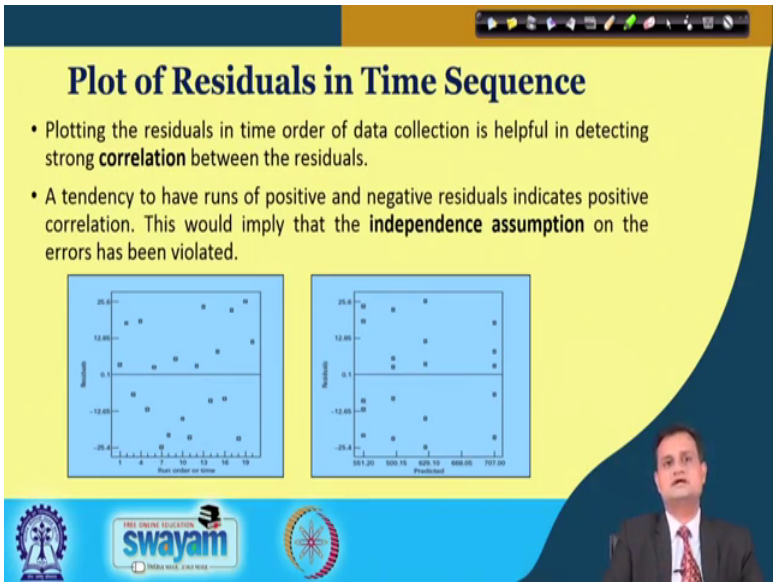
So, I will have the y i j hat is equal to y i dot hat and I am just trying to prove that my estimate about the y i j hot hat is basically y i dot bar. So, this is how we try to just do the statistical analysis.

(Refer Slide Time: 27:10)



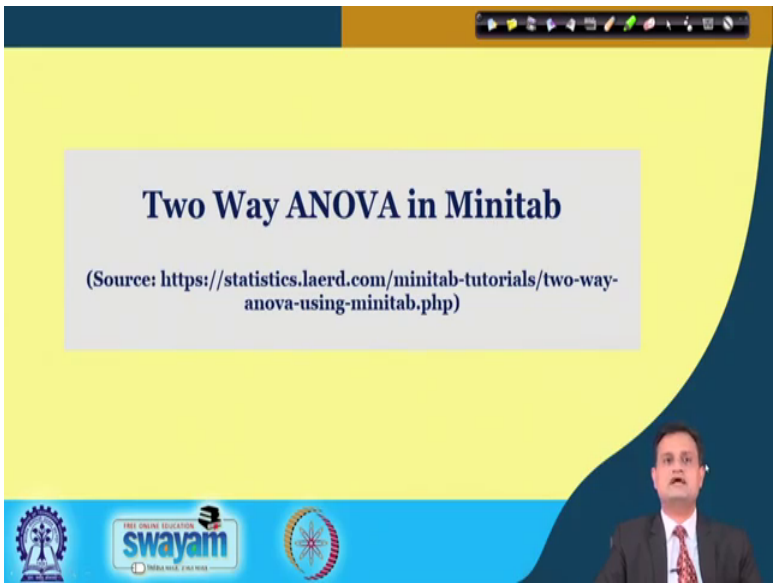
And set the parallel among different parameters here, the most important test is my assumption verification about normality, you can plot the normal probability plot and here you are trying to put the residual and the corresponding say probability value referring the z table, if these values basically from z table they pass through one line, or if you put a pencil and almost all the points are covered, then you will say that my assumption about normality is true. So, this is my normality assumption verification.

(Refer Slide Time: 27:54)



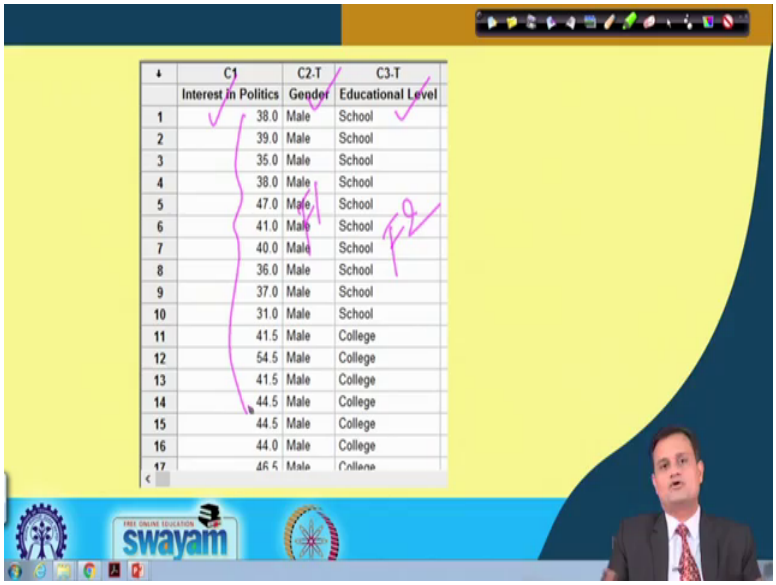
And similar way you can also see that is there any correlation between residuals. So, you can have run order time in you can plot the residual, if they are scattered randomly you do not bother much, if you see some pattern increasing decreasing and this then there is some reason to say that, because of some machine problem setting problem or the skill level of the supervisor or worker, there is some autocorrelation or some other factor which is external to your system maybe humidity and other. And then you try to set it correct before you actually add up the ANOVA analysis results.

(Refer Slide Time: 28:37)



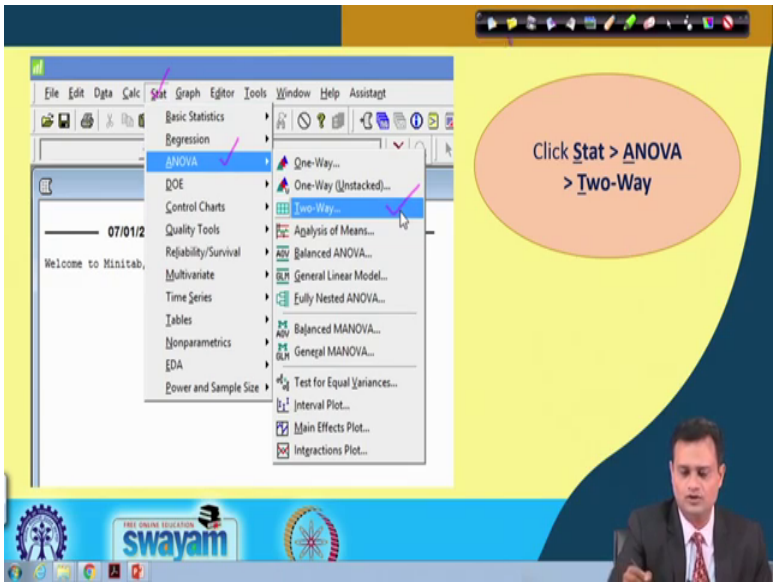
Now, let us see quickly what is to be done in two-way ANOVA in Minitab how it can be conducted.

(Refer Slide Time: 28:43)



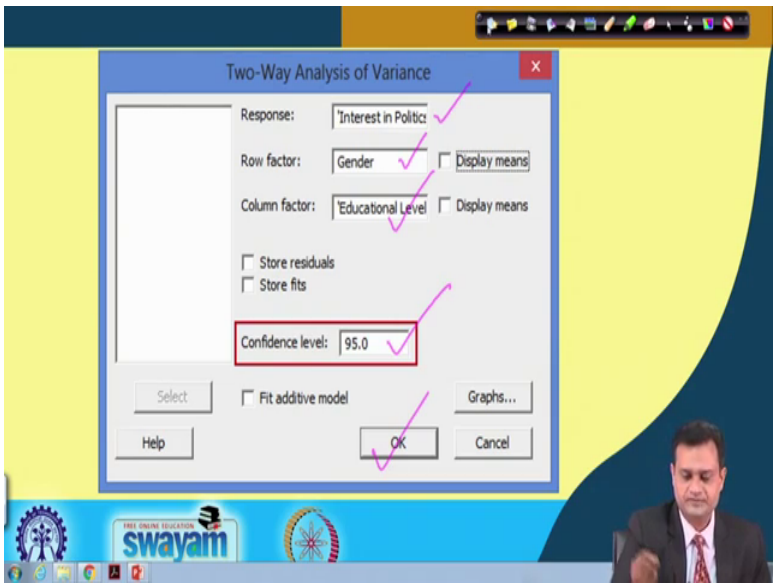
So, let us say you have a data interest in politics gender and educational level, you have two factor 1 is gender factor 2 is educational level and I want to just check or measure for some score that interest in politics. So, I may has them couple of questions and then try to catch their interest I may sum it up and find the score maybe out of 50 or whatever. Now, with this score I want to do the ANOVA analysis to see, that whether factor 1 gender and factor two educational level are the responsible or the interaction of this two is having the influence on the interest in politics.

(Refer Slide Time: 29:26)



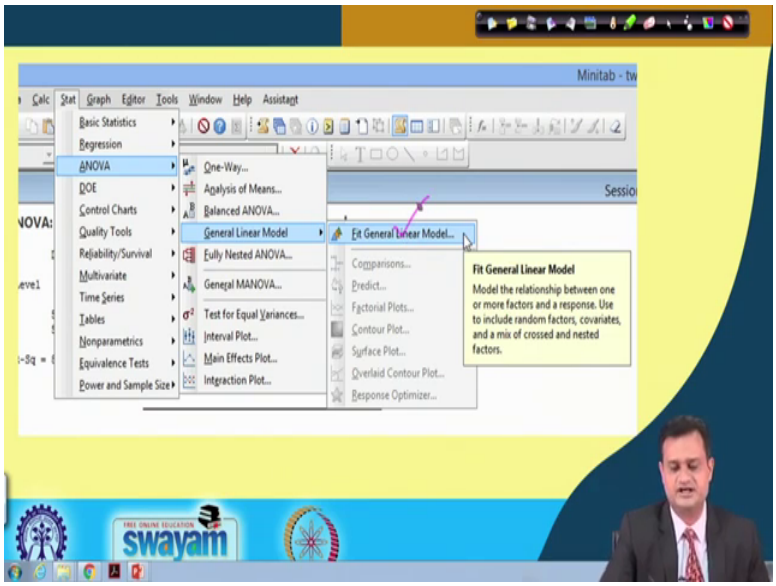
So, basically you need to do something very simple you go to statistics, you go to ANOVA two-way ANOVA this is the link you follow.

(Refer Slide Time: 29:38)



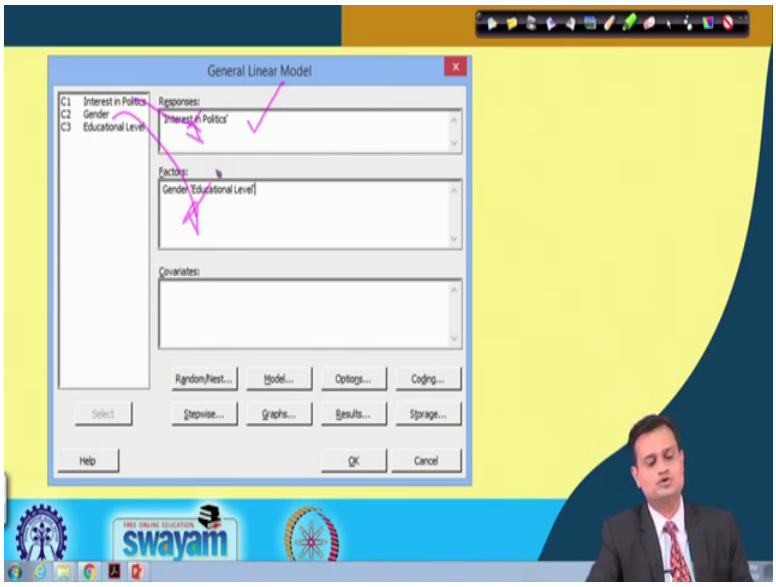
And then you open this window where you put the response variable interest in politics row factor, column factor, confidence level press ok.

(Refer Slide Time: 29:53)



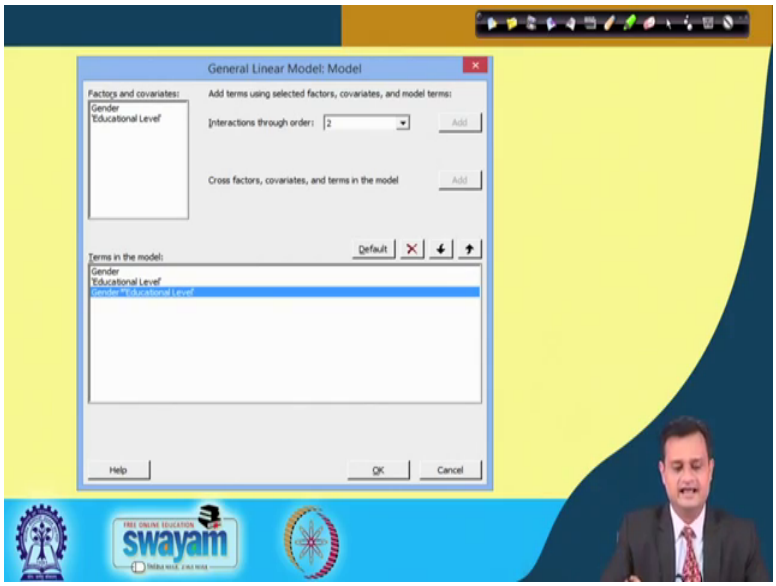
And this way say you can go it you also fit the general linear model and this general linear model, you go to fit general linear model and I also need to capture the interaction effect.

(Refer Slide Time: 30:05)



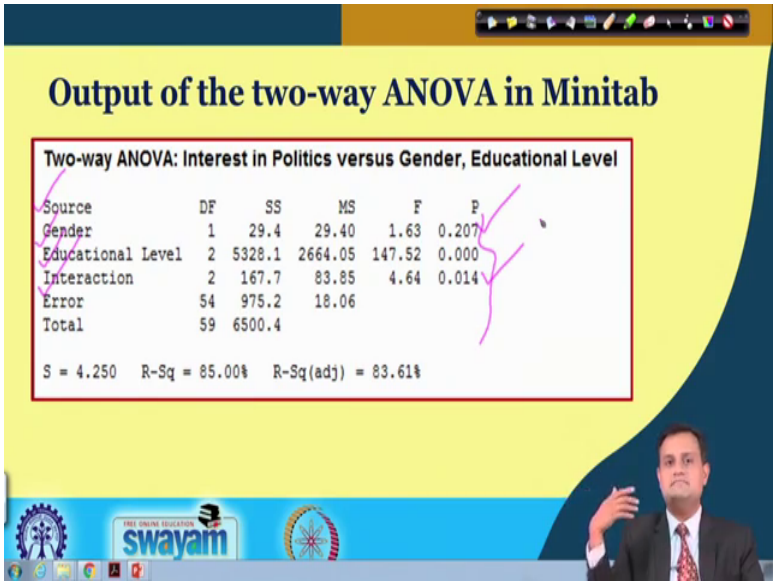
So, you say that in general linear model window you pull this particular c 1 into this, interest in politics this is your factor 2 that is gender.

(Refer Slide Time: 30:24)



And you can define the interaction through order that is 2, because I want to check the interaction between gender and the educational level.

(Refer Slide Time: 30:30)

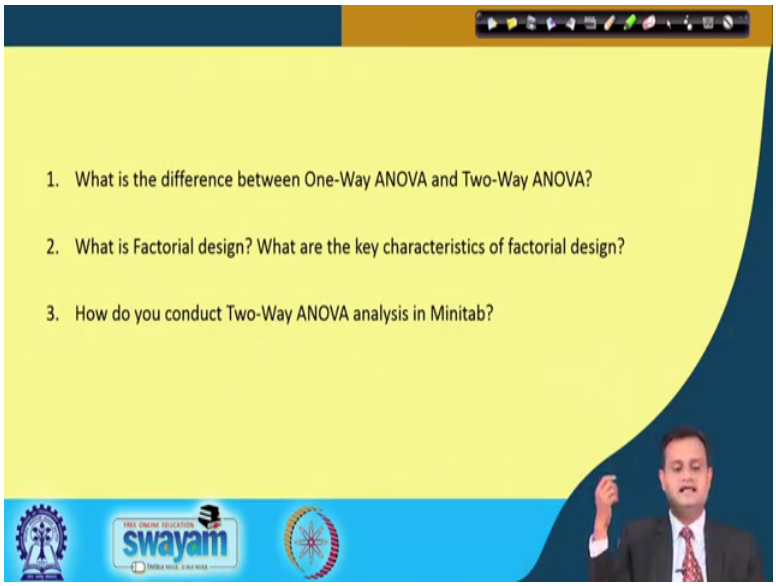


And then you can find your ANOVA. So, similar table you will get that source gender education and interaction between these two, if you just refer the p value easily you can make the inference for example, say educational level the p value is almost 0 or 0 so, it falls in the rejection region and I will say that gender has no effect on interesting politics, but if you look at 0.207 that is the p value for gender, then you can say that gender yes it

has some impact on the interest in politics and educational level it does not have because it falls in the rejection region, when I look at the interaction, then suppose I have 0.05 level of significance this is 0.014.

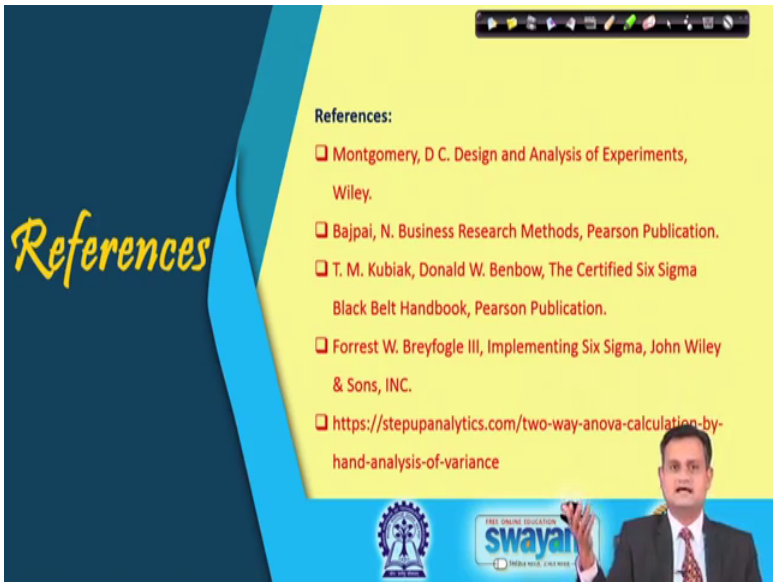
So, this will fall in the rejection region and I will say that there is no interaction prevailing between gender and the educational level and it can have any impact on the interest in politics. So, I hope you have appreciated this particular lecture.

(Refer Slide Time: 31:41)



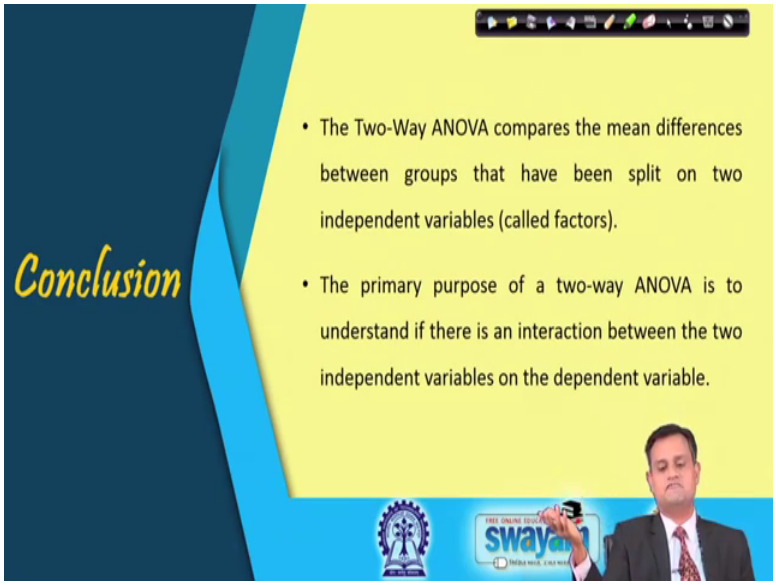
Before I end I would just like to float couple of think it for your introspection, what is the difference between one-way ANOVA and two-way ANOVA, what is factorial design what are the key characteristics of the factorial design and how do you conduct two-way ANOVA analysis in Minitab.

(Refer Slide Time: 31:58)



So, you can go through couple of references and try to appreciate the different examples solved using two-way ANOVA analysis.

(Refer Slide Time: 32:09)



So, typically two-way ANOVA analysis compares the mean of different differences between groups that have split on two independent variables called factors and the primary purpose of two-way ANOVA analysis, is to understand if there is an interaction between the two independent variables on the dependent variable.

So, thank you very much for your interest in learning two-way ANOVA analysis, I hope you are revising the concepts time to time and trying to appreciate the our advancement in the journey of six sigma, from define to measure, measure to analyze, analyze to improve and so on we will go ahead. So, try to focus on various concepts internalized through application be with me enjoy.