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

Lecture - 41
Factorial Design

Hello friends. Once again you are welcome to our Six Sigma journey and if you recall we are presently discussing the improved phase of our DMAIC cycle and in the improved phase, we are talking about design of experiment and the various experimentation strategies I want to improve the design, right. At the design stage I want to improve the process right at the design stage and such kind of analysis that can help me to understand the significance of the factors on the overall system or the product or process can really help me to figure out that what is important and what is not.

So, as a part of lecture 41 we will study factorial design. I have already given some idea of this design when we talked about two way ANOVA analysis, but now we will try to once again appreciate it with some more details.

(Refer Slide Time: 01:25)

So, we have a very beautiful quote experiment, learn, fail, learn and repeat. So, does not matter even if you fail that is ok, but your experimentation will provide you lot of insight, you can learn from that, you can repeat the experimentation with different approach and then you can infer the greater insights.

(Refer Slide Time: 01:53)

So, recap we have talked about minitab application for randomised complete block design.

(Refer Slide Time: 02:01)

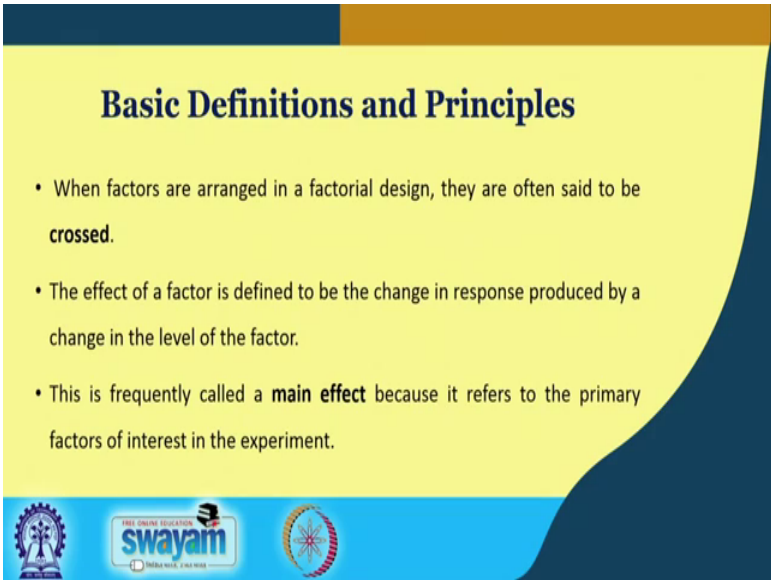
In the last lecture and this lecture we will basically focus on principles of Factorial Design, advantages of factorial design, statistical analysis of factorial design and illustrative example of factorial design. So, more or less the analysis system will remain same, statistical analysis will remain same, but we would be addressing the different conditions under which we are trying to execute the experimentation.

(Refer Slide Time: 02:32)

So, the basic definition in principle as like this that factorial designs are most efficient for this type of experimentation when you have to analyse two or more factors. Factorial design mean that in each complete trail or replicate of the experiment all possible combination of the levels of the factors are investigated. So, I already briefed you that what is the level. Suppose your factor is temperature, then you may have 10 degree, 50 degree, 70 degree three different levels.

If there are a levels of the factor A and B levels of the factor b each replicate contains all a into b treatment combinations.

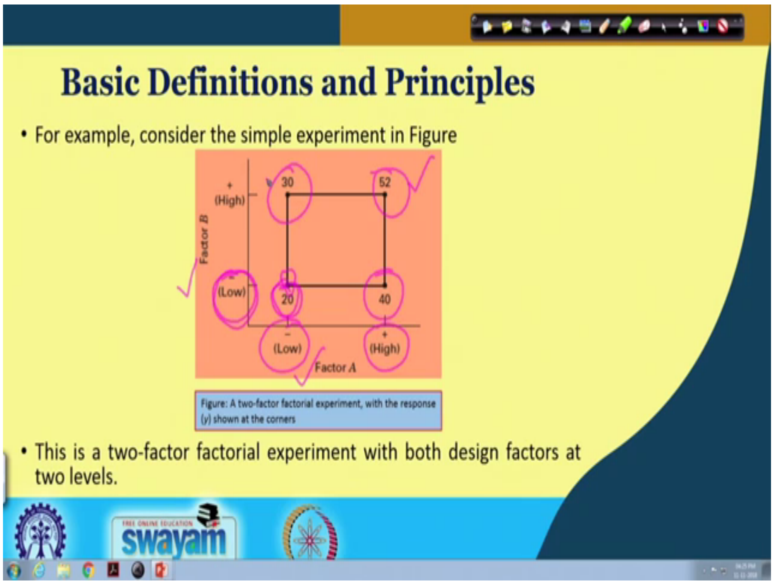
(Refer Slide Time: 03:19)



Now, many times say when you do the design of experiment and when factors are arranged in a factorial design, they are often said to be crossed. So, you have two factors; you are creating number of replicates and often we call them as the cross factors and the effect of a factor is defined to be the change in response produced by a change in the level of factor. So, you change the temperature level, you change the pressure level and then, that will lead to some change in the response may be the productivity, may be the surface finish and this is what exactly we are trying to investigate.

Now, typically when you do this, this is called main effect because it refers to the primary factor of interest in experiment. So, I am introducing one word that is the main effect. So, suppose you have a pressure and temperature are two different factors, I will call individually each factor as the main effect factor and that we try to analyse to our factorial design.

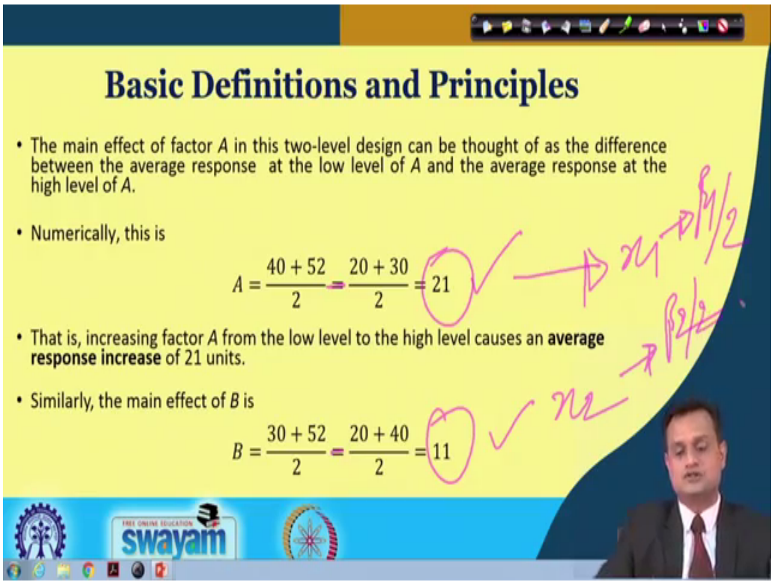
(Refer Slide Time: 04:26)



So, just try to appreciate that you have factor A, you have factor B and you have set the factor A at low level that is 20 and you have set the factor B at the low level. So, this is the condition where both the factors they are set at the low level. Now this is the two factor factorial experiment and the response is shown at the corner. It means you are receiving some response when both the factors are set at the low level.

Now, just see factor A is set at the high level and factor B is set at the low level so, you get the response 40. We are not taking the factors at this stage. Just assume that these are two factors. Now, if you have the factors; factor A at high level and factor B at high level, the response you get is 52 and if factor A is at low level and B is at high level, the response you get is 30. So, this is the two factor factorial experiment.

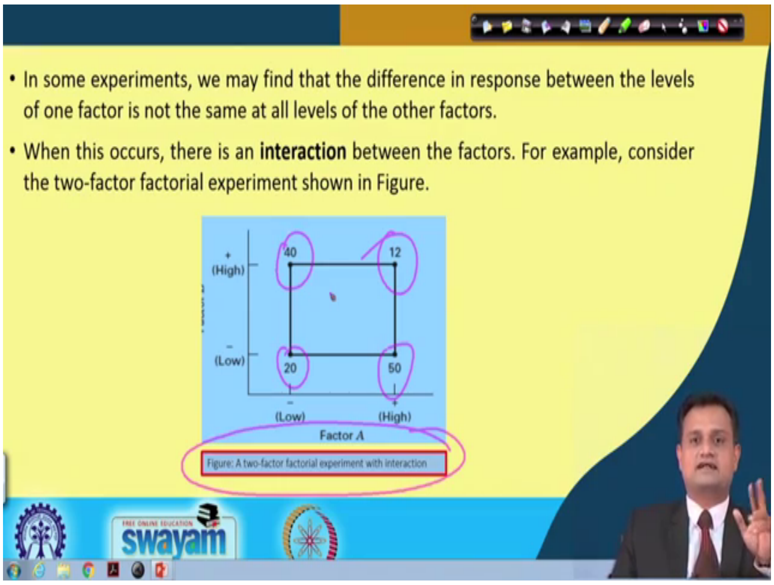
(Refer Slide Time: 05:40)



And we will try to analyse this initially with a very simple understanding and then with our ANOVA analysis. So, typically what you can say that I have A factor A and difference between the average response at lower level of A and the average response at the high level of A if I want to numerically calculate, then just see 40 plus 52 by 2 is equal to minus this should be minus 20 plus 30 by 2 equal to 21.

So, then you can just go back and see that what is happening here. So, I have 40 plus 52. You can see here high level by 2 minus 20 plus 30 by 2. So, this will give me the change in level A because of change in my setting from low to high and same way you can analyse. So, this is also minus, this is also minus and B 30 plus 52 by 2 minus 20 plus 40 by 2, this is 11.

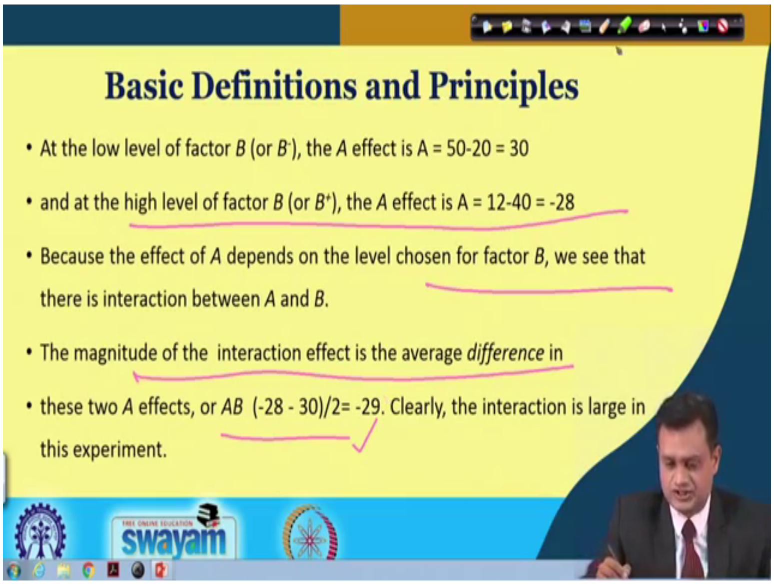
(Refer Slide Time: 07:00)



So, once I can do this, then I have say the system like this which I am discussing and in some experiment we may find that the difference in response between the level of one factor is not the same at all levels of other factors. So, when this occurs you will say that there is something else and this something else is basically called the interaction between the factor.

So, for example you consider a two factor factorial experiment and here I am trying to capture the effect of interaction also. So, you have 20 40 50 and there is something which is 12. So, here I have a suspicion, I have a doubt that these two factors are interacting with each other.

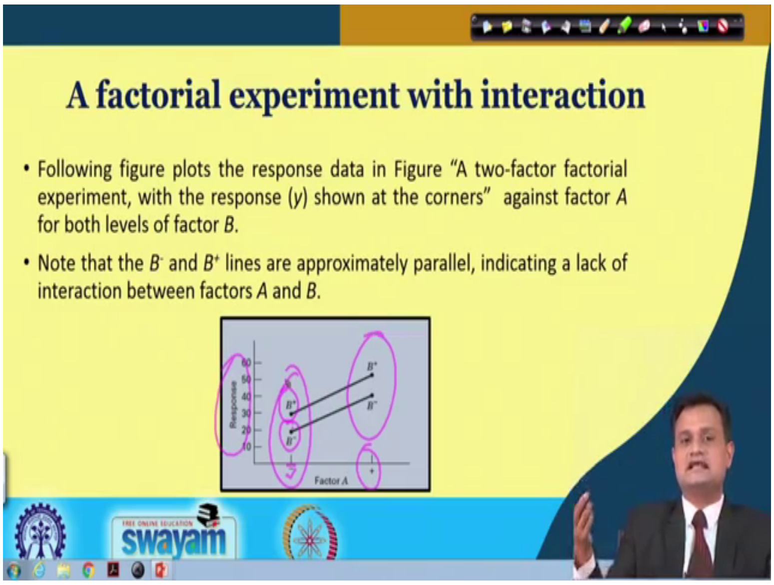
(Refer Slide Time: 07:53)



So, now when you are analysing such kind of case just try to see that when the low level of factor B, that is B minus, the A effect is called 50 minus 20 30. So, we can just go back and try to see what it is. So, 50 minus 20 that is 30, so, you have this 50 and you have this 20, so, 50 minus 20 that is 30. So, now same way you have say level factor B plus and the A effect is 12 minus 40 that is minus 28. So, because effect of A depends on the level chosen for factor B, we see that there is an interaction effect. It is a very important conclusion that effect A depends on the level chosen of B. It means that there is some interaction effect prevailing between A and B.

Now, magnitude of this interaction effect can be easily found by the average difference in these two A effects or AB minus 28 minus 30 divided by 2, so, this will be minus 29. So, this is the large interaction effect whether statistical this is significant or not, I cannot say unless I do the ANOVA analysis, but at least I can say that yes, if I look at the values of main effect and I look at the value of interaction effect, yes there is a significant or say contributing interaction effect, large interaction effect present.

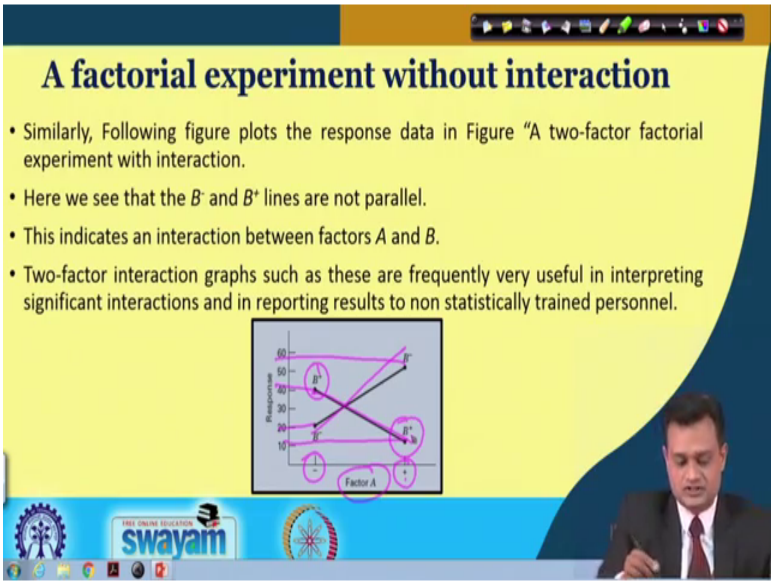
(Refer Slide Time: 09:38)



Now, just see that when you have factorial experiment with interaction, then a two factor of factorial experiment with a response y you can show it as let us say corners and this is my response against factor A. So, against factor A when it is at low level, what is happening in this? So, B plus and B minus and factor A is set at high level, then B plus and B minus.

So, what is the response I am getting when I set factor A at minus low level and I have the factor B at minus same way this thing. So, B minus and B plus lines are approximately parallel indicating lack of interaction between A and B. So, I am just trying to discuss another case that when you plot it and if you find these two lines parallel, then its an indication that the interaction is not significantly present and it is a lack of interaction situation.

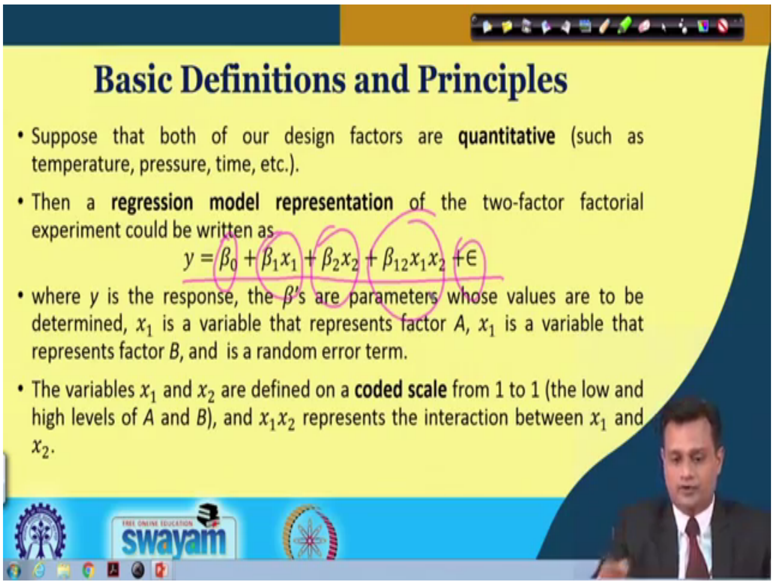
(Refer Slide Time: 10:49)



So, now let us see the different situation where you are trying to say that there is an interaction, it means when you change the level of factor, yes there is a possibility that your response will get reversed or there is an effect impact on the response. So, just try to see what is happening. I have a factor A, this is set at low level and I am checking the value of let us say B plus and then, when I change the factor A to higher level plus and again I am setting B plus then I have change in response.

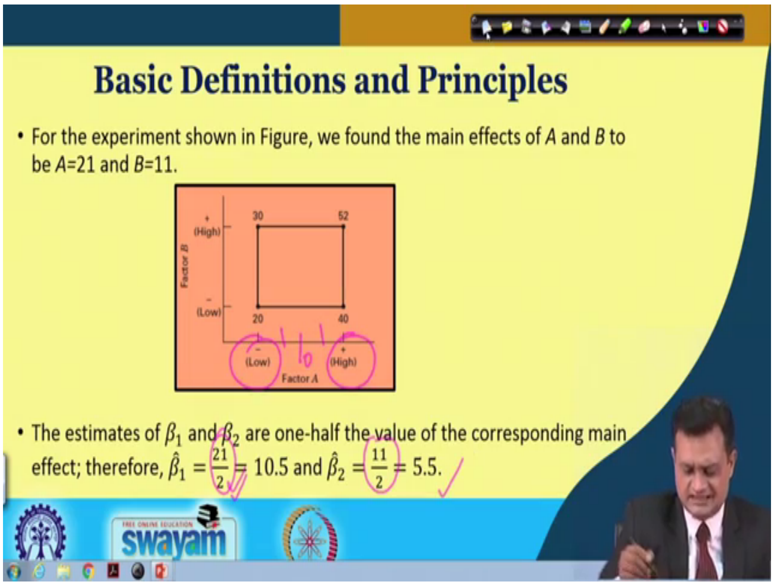
Similar way, when I see this B minus and B minus for this yes there is change in response and you can see that these two lines are not parallel. So, there is something which is called interaction that is prevailing in the system. So, I cannot simply say that suppose my factors are pressure and temperature, these two factors main factors are independent of each other. There is pressure into temperature there is the interaction effect that also prevails. So, this is the concept that we additionally include when we try to analyse the factorial experiment.

(Refer Slide Time: 12:07)



And then my statistical model would look like this where y is equal to beta 0 that is the intercept. You have beta 1 x 1 that captures the main effect, beta 2 x 2 that captures another factor effect, main effect, beta 1 2 x 1 x 2 and you have epsilon that is your error component. So, this is how my model will look like.

(Refer Slide Time: 12:34)

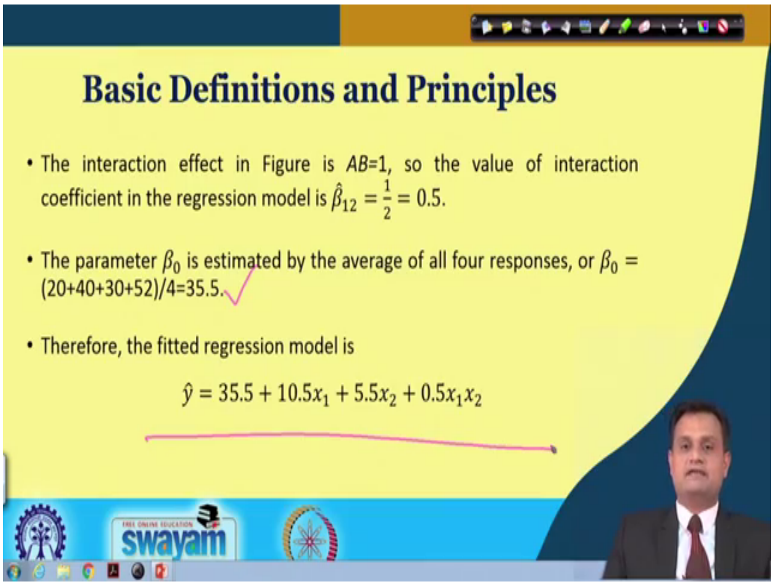


Now, let us try to go little bit deeper and investigate the values of the slope. So, what I will say that beta 1, this is the system I have already explained and beta 1 is basically estimate beta 1 hat is 21 by 2 and beta 2 hat is 11 by 2. So, from where this 21 and 11 has come you can just see I can just go back to help you.

So, you have computed this 21 and 11. So, this 21 is computed, this is minus; this is minus when the factor a can be thought of as difference between the average response at low level of A high level of A. So, this is the estimate of factor A, this is the estimate of factor B. So, here when I am setting the regression model, this is x 1, this is x 2 and the slope pertaining to this will be beta 1, this will be beta 2.

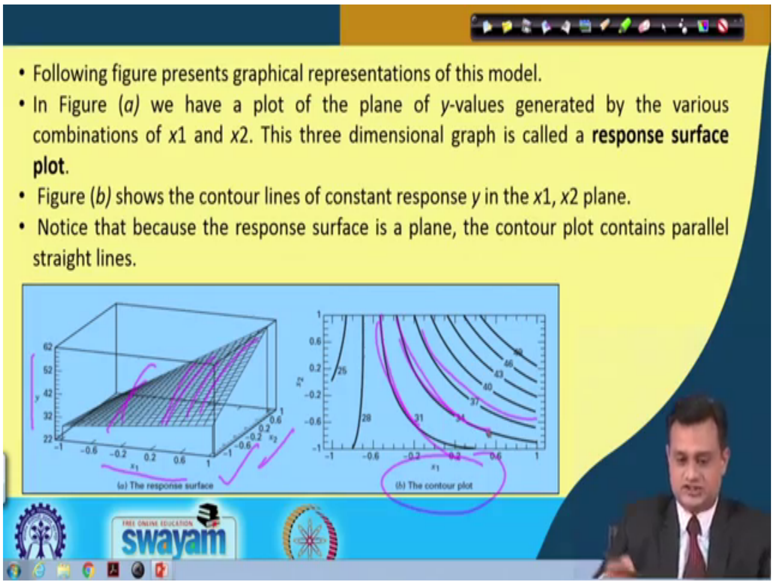
Now, the question comes that why I divide this by 2; divide this by 2. So, now let me go back to the slide which we were discussing. So, now why do you divide this by 2 and this by 2? So, its very simple. I move from low to high level. In between there could be 0 and hence, I am moving two steps. So, my average value will be 21 by 2, this is one unit; this is one unit. So, I am you I am moving by two step. So, I am taking the say average of this that is 21 by 2, this is 11 by 25.5. So, this is the simplest way, in fact not much statistical approach.

(Refer Slide Time: 14:27)



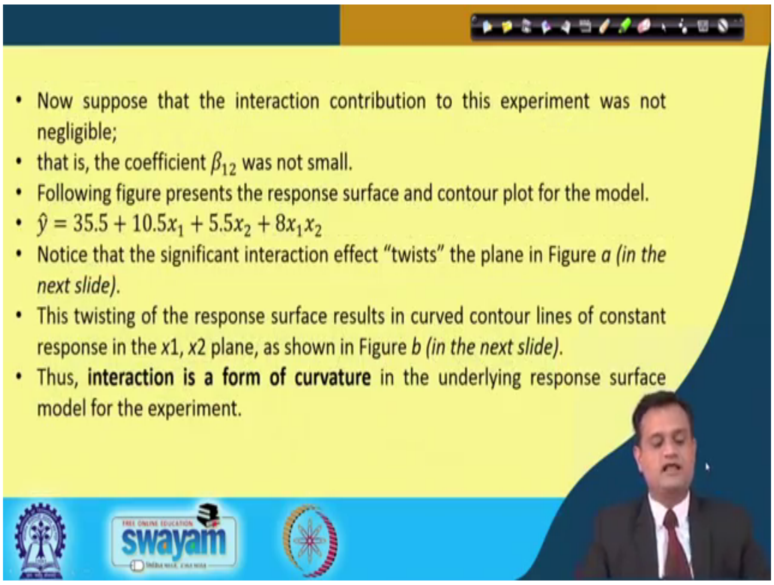
But I can estimate my beta 1 beta 2 and same way I can do it beta 1 2. So, in figure A, B is equal to 1, so, I just divided by 1 by 2 and I have beta 0 that is the intercept. So, I have this core at each corner and this is summation of these divided by 4. So, what I get here is basically 35.5. So, my regression model would look like this 35.5 plus 10.5 x 1 plus 5.5, x 2 plus 0.5, x 1 into x 2.

(Refer Slide Time: 15:06)



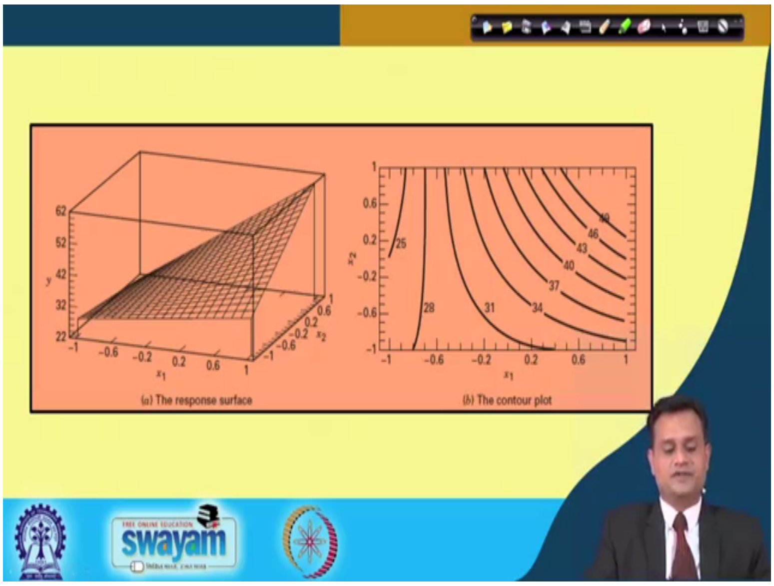
So, now when I have this model what I can say that this model is typically can be plotted as a response surface and you can see that your response surface is basically curved and this curvature is mainly because of the interaction effect and this is your contour plot which also says that you have the curvature when you try to plot your responses on x 1 x 2 and y. So, this is my factor 1, your factor A x 1 factor B x 2 and this is my response when I am trying to put it, I get the curvature; I get the curvature and it is mainly because interaction effect x 1 into x 2 is present. So, I will have a quadratic function, so, my response function would be say curved in the nature.

(Refer Slide Time: 16:07)



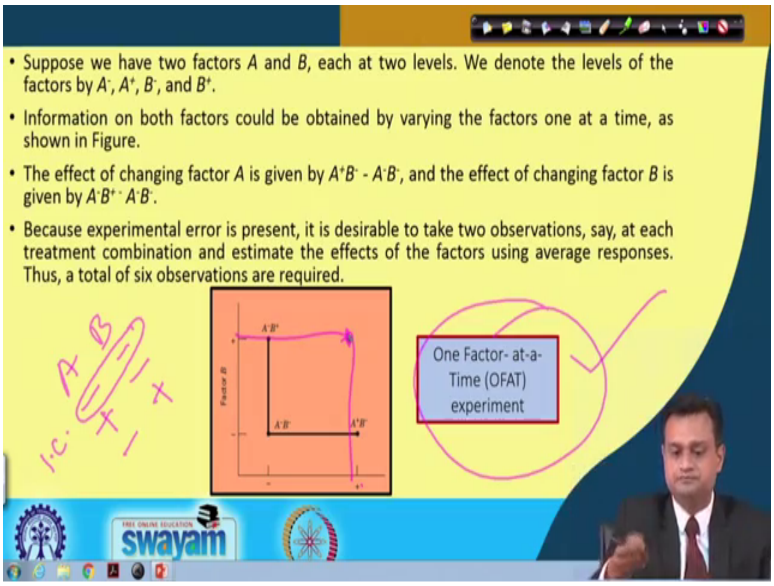
So, when interaction effect is present, you will see that your response function will be curved.

(Refer Slide Time: 16:12)



And this basically indicates the interaction effect. So, this is what exactly; now the advantages of factorial that it helps me to analyse the interaction.

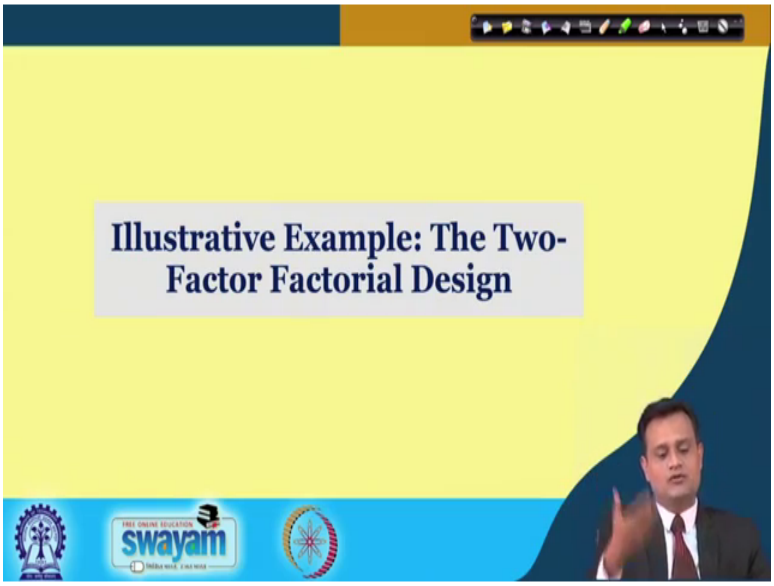
(Refer Slide Time: 16:20)



Now, you just see in order to appreciate this you just see this is one factor at a time experimentation strategy. So, one factor at a time means suppose I have factor A and B and both the factors are set initially at low level. Now, I am changing factor A to positive, I will not change the level of factor B. It will be set at the initial condition. So, I will call this as the initial condition IC let us say initial condition and when I will change factor B positive, I will keep A as the initial condition. So, I will not have plus plus means both set at plus and n that is why this particular point is missing.

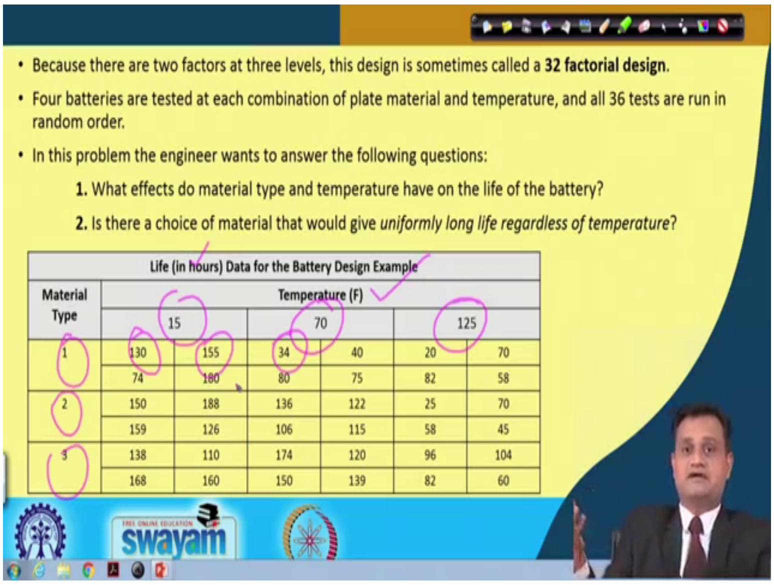
So, when I conduct the experiment one factor at a time, I am missing this particular say point and this is the additional information I get when I conduct the factorial experimentation when I am allowing the factors to be changed simultaneously and there is nothing like setting or keeping a particular factor constant at the initial condition.

(Refer Slide Time: 17:46)



So, this is what exactly we have to appreciate about the factorial experiment compared to the one factor at a time, I do not say one factor at a time is an inferior strategy. That is that is a good strategy in some of the conditions when you feel that simultaneous changing of the factor or the interaction effect are not really our interest of importance and we just want to be happy with the one factor change at a time. Now, let us try to see some illustrative example of two factor factorial design.

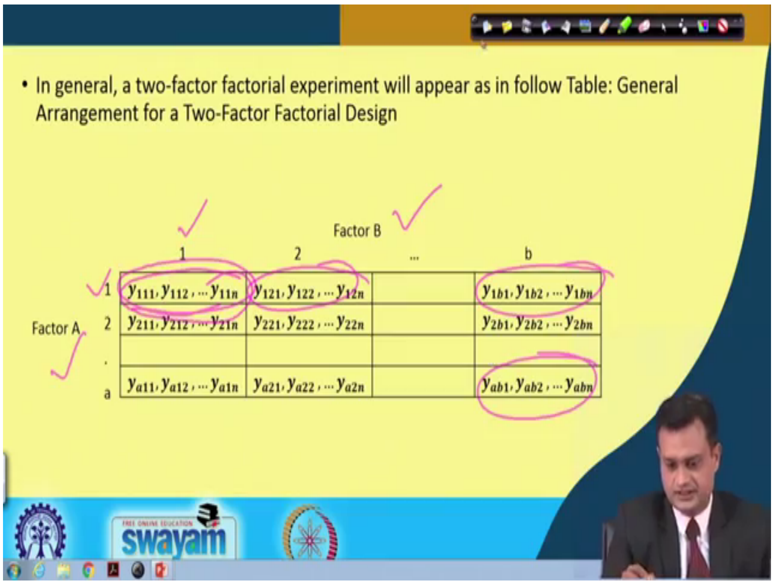
(Refer Slide Time: 18:25)



So, my design is like this. I have the temperature setting and let us say 15 70 125, I have the material type 1, 2 and 3 and this is typically the problem referring from Douglas Montgomery design of experiment, life data of the battery design example. So, these are the life data may be in hours, here it is hours. So, I want to see that suppose I am using a particular material and suppose there is a particular temperature, then will it have an impact on the life of the battery.

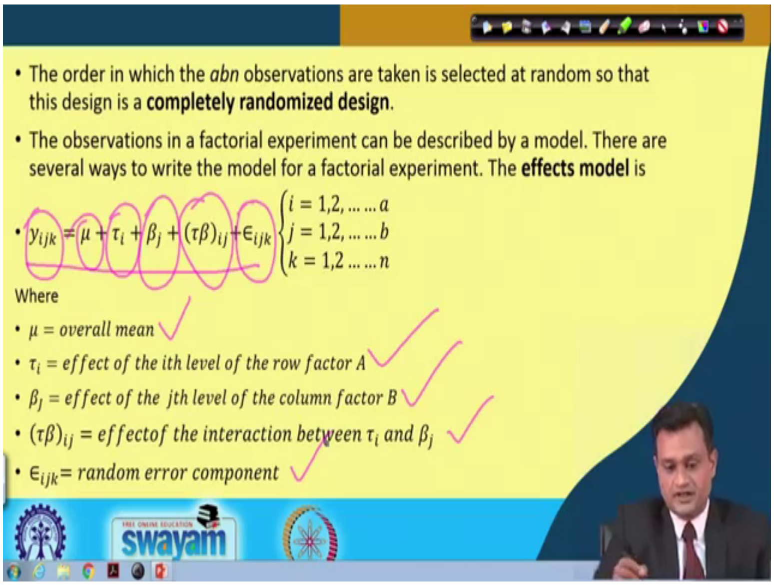
So, you are using battery in automobile and many other say appliances vehicles. So, now we have battery operated vehicles also. So, this is a very important problem that the temperature at which the battery is getting exposed and the material you have used really it has some impact on the life of the battery or not. So, this is what I am trying to analyse.

(Refer Slide Time: 19:40)



So, now let us try to appreciate the basic structure of the problem. So, my structure is like this; you can see that I have some readings in a particular block and not block particular cell rather. So, factor A and factor B, I am taking y ijk reading for each particular cell and this is how my data is organised. So, we will see the example it would be better, clear.

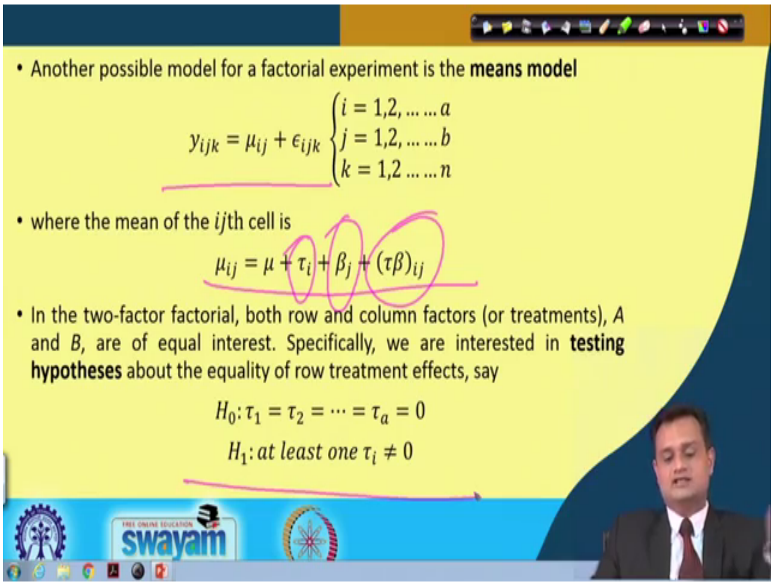
(Refer Slide Time: 20:10)



But before that let me try to understand my model, what is my factorial design model? So, my model looks like this y i jk this is the effect model I am trying to analyse the effect, y i jk is the response. So, i pertains to your particular factor and j pertains to your another factor, you have temperature as well as you have material type and k you are referring to a particular say cell because within a cell you are taking number of replicates and in order to improve the accuracy of your experimentation.

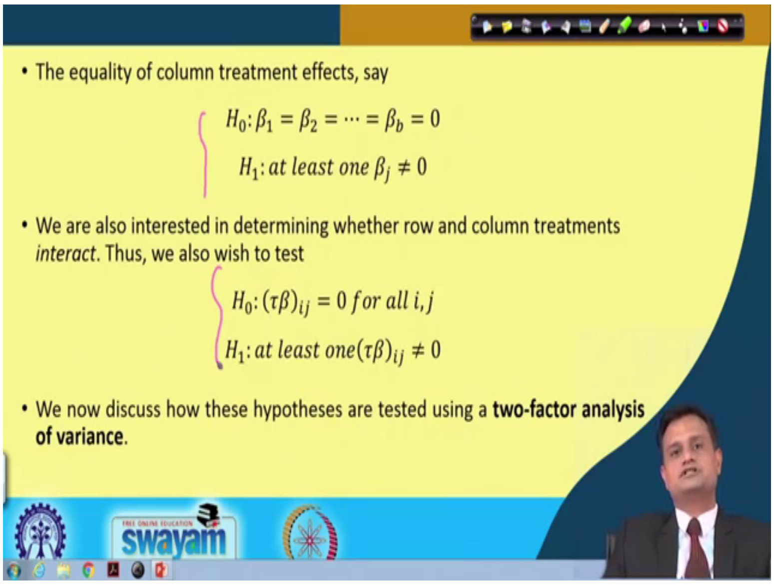
So, mu is the overall mean, tau i is the effect of the level of the row factor A, then beta j is the effect of the jth level of the column factor B tau, beta i j is the effect of the interaction between tau i and beta j and this is your random error component that is epsilon i jk.

(Refer Slide Time: 21:19)



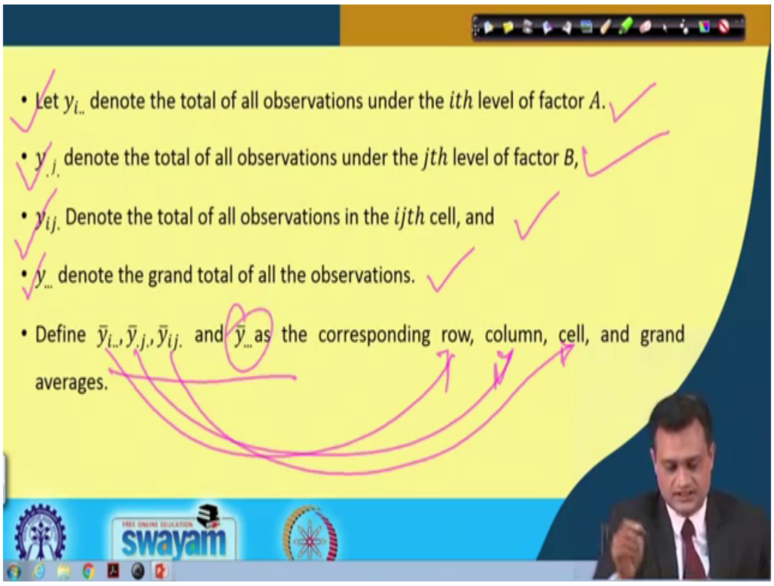
So, when you appreciate this basic model what I say that y i jk basically is the mean specific to i and j plus epsilon i jk and mu ij. The way its represented in a statistical mathematical term mu plus there is tau i effect, row effect, column effect, interaction effect and I can have a null hypothesis that tau i is equal to 0 means tau 1 is equal to tau 2 is equal to tau a is equal to 0 and I can say that at least one of the treatment tau a or tau i is not equal to 0.

(Refer Slide Time: 22:04)



So, now I think you are familiarise with such kind of treatment because you have two factors row and column. Similar way you can set the null hypothesis alternate for the column factor, similar way you can set the null and alternate hypothesis for the interaction and this will basically help you to investigate the two main effects and the interaction effect.

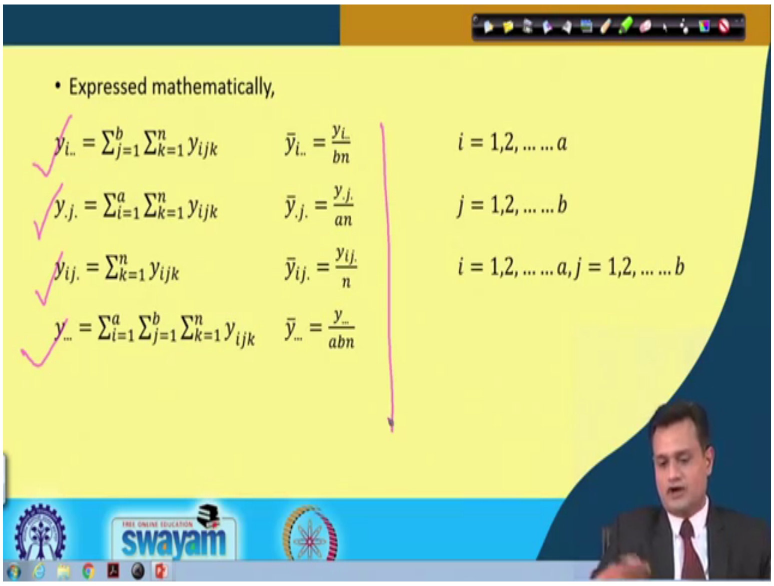
(Refer Slide Time: 22:27)



So, statistical analysis of the model is like this you have yi triple dot basically denote the total of all the observations. You take the summation under the ith level of factor A, you have y dot j dot denote the total of all the observations under jth level of factor B, you have y i j dot denote the total of all the observation in say particularly i jth cell and y triple dot denote the grand total of all the observations.

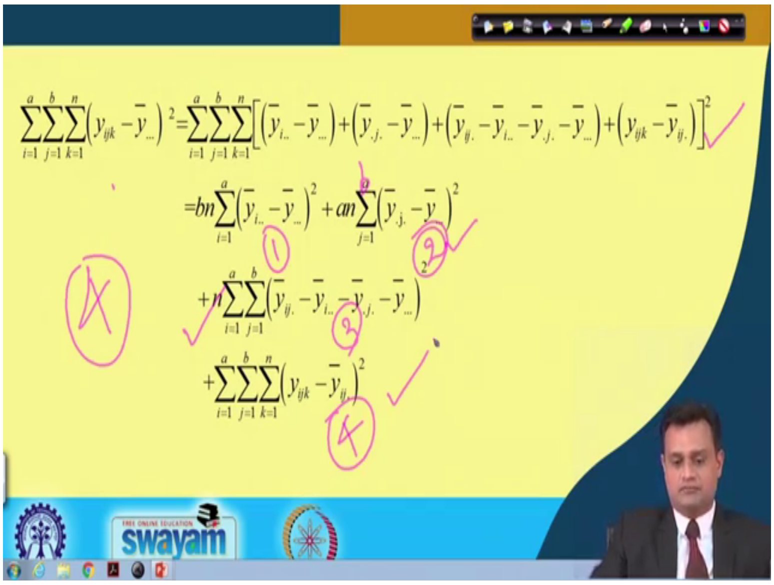
So, these are the mean values corresponding mean values yi dot refers to row y j dot refers to column yi j dot bar refers to cell and then, you have the grand averages.

(Refer Slide Time: 23:20)



So, these are the values you need to compute and then you can say find the total of y i double dot y dot j dot using this expression y ij dot and y triple dot and this is the averages of each particular y i double dot bar y dot j dot bar y ij dot bar and this. So, I hope now we are comfortable because we have done this many times.

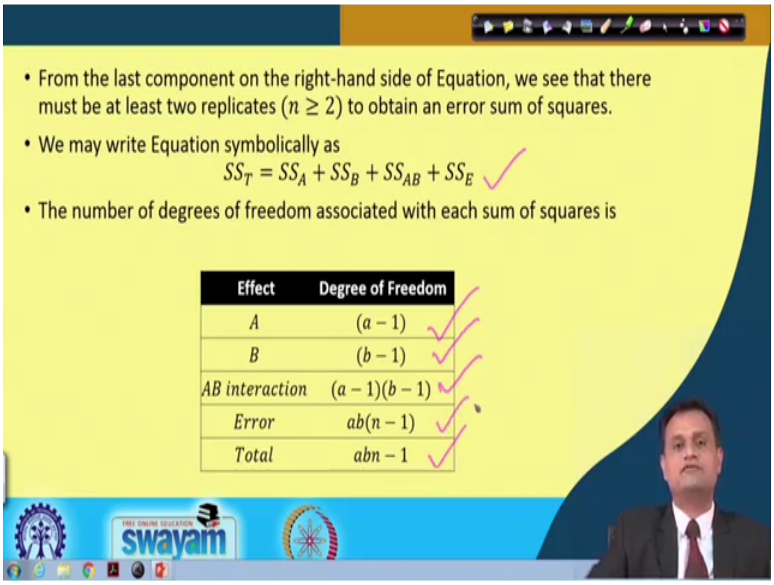
(Refer Slide Time: 23:50)



And now I want to basically find the sum of square. So, my approach will remain same, I will try to minimise the error component and I am checking the expression that y i jk minus y bar triple dot. So, I am trying to take the difference between individual value and the grand mean and square it. So, that will help me to find the total sum of square what is the total variability in my data and when I just expand this expression, then I will end up with say this, this and this.

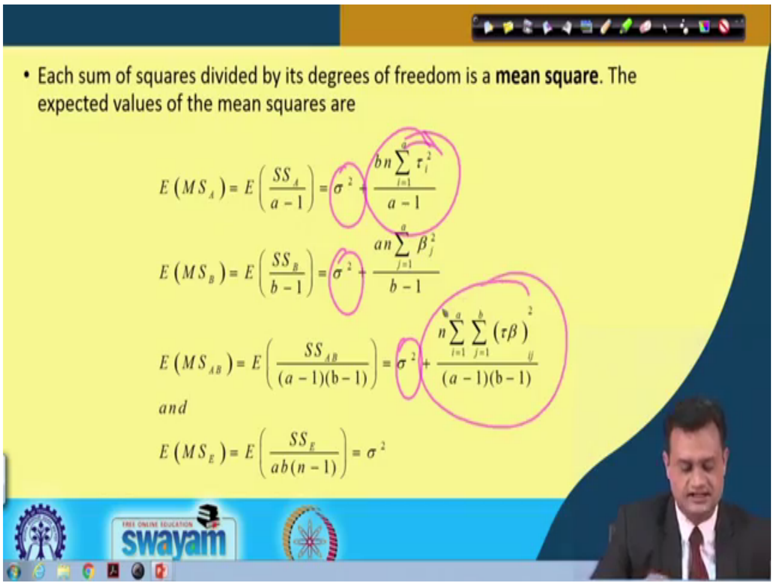
So, basically you have number 1 and this expansion then when you expanded, you have number 1 component, number 2 component and number 3 component. So, you can very well understand that you have basically say this is number 1, this is number 2. I will rewrite this is your basically say rearrangement of the terms you have number 1, number 2, number 3 and number 4. So, you have derived total four components out of these and these four components are basically nothing you have the row effect factor one effect, second you have the column j is equal to 1 to b and you have say interaction effect and you have the error component.

(Refer Slide Time: 25:32)



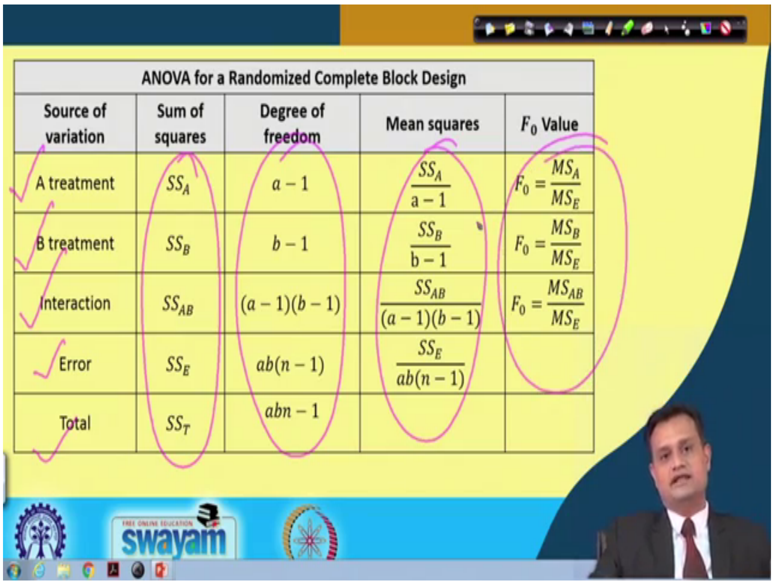
So, this is what you basically try to do. So, I have the expression like SS T is equal to SS A plus SS B plus SS AB plus SS E and these are the corresponding say degree of freedom.

(Refer Slide Time: 25:51)



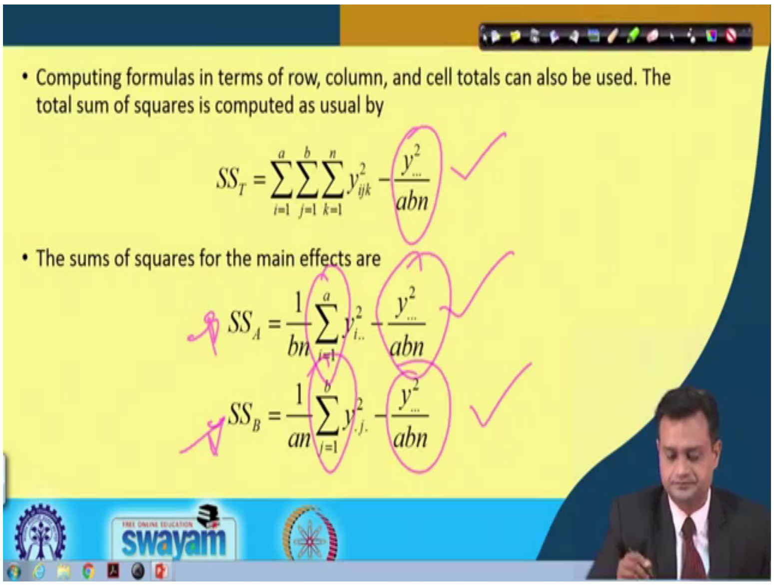
So, once you have done this, then you can compute the degree of freedom and you can also estimate the mean square error just by including the treatment effect to the population, variance population variance and this will help you to appreciate the basic logic behind the ANOVA analysis.

(Refer Slide Time: 26:13)



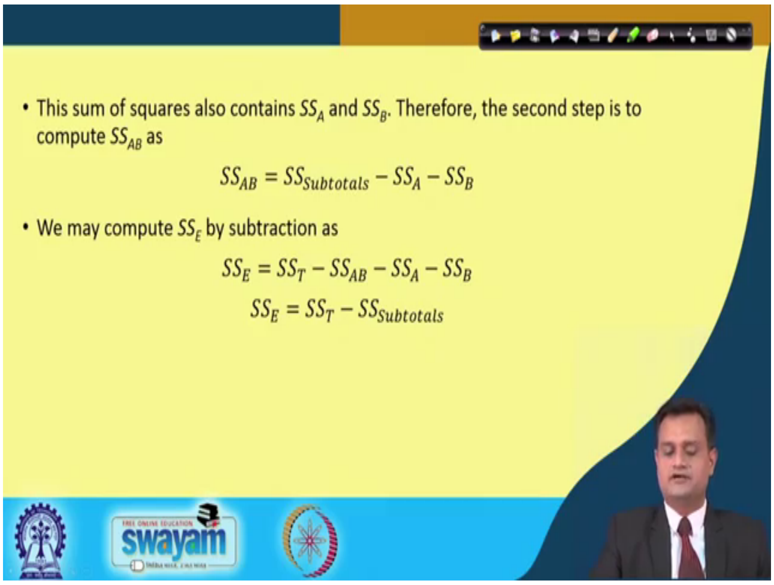
So, now if I look at my ANOVA table, then I have A treatment B treatment interaction error and total I have all the sum of square, I have degree of freedom, I have mean square and I can find the F 0 values. So, once you have done this, then its very easy to analyse the significance of treatment A treatment B and interaction.

(Refer Slide Time: 26:39)

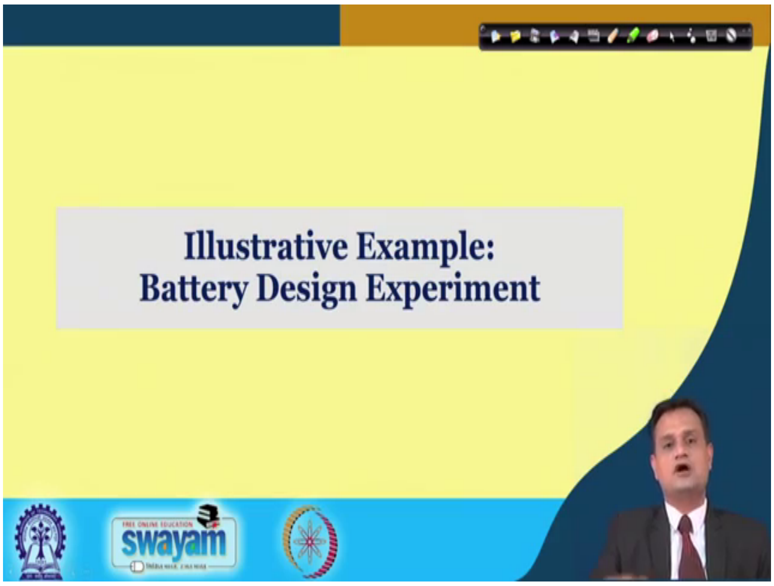


And then you can use easily this expressions SS T, SS A here. It is with respect to grand mean square and because this is your row effect, so i is equal to 1 to a j is equal to 1 to b column effect and you can easily find this value. So, interaction effect in error.

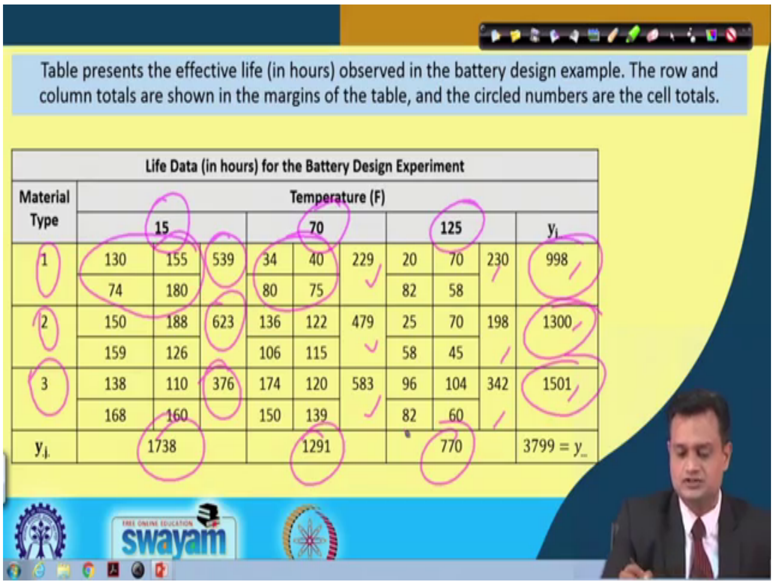
(Refer Slide Time: 27:02)



(Refer Slide Time: 27:05)

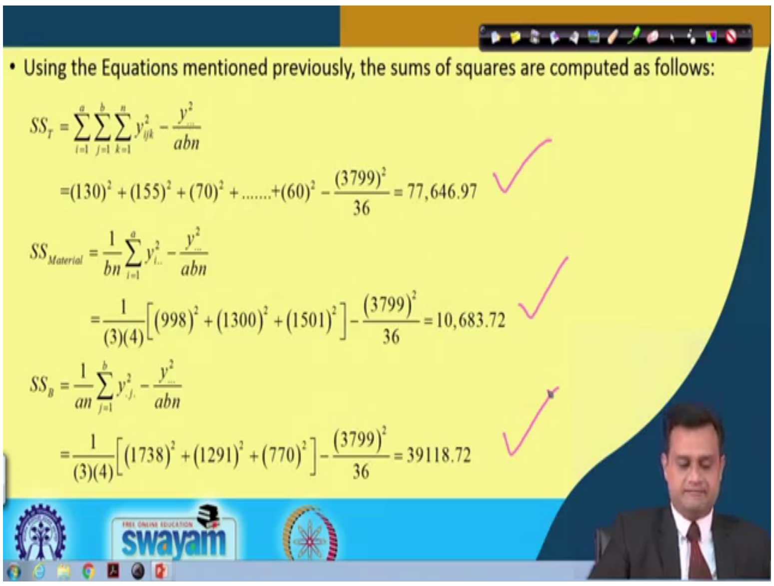


(Refer Slide Time: 27:08)



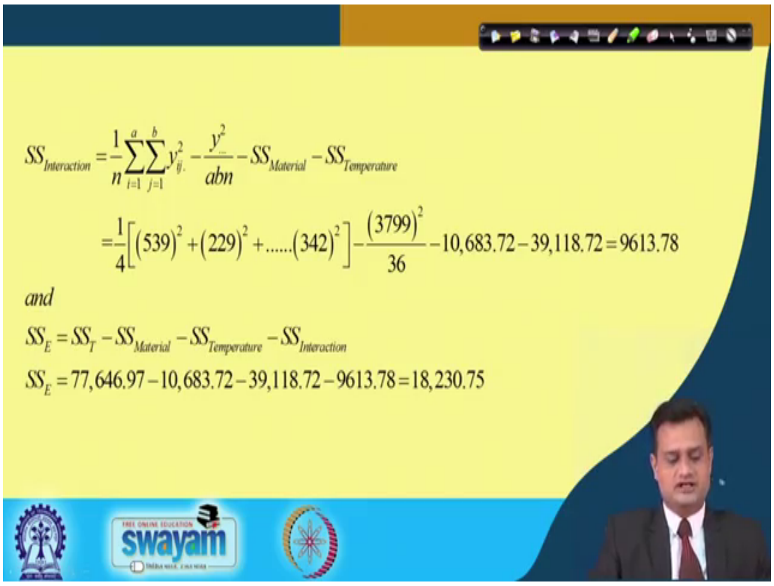
So, now if you go back to our battery design experiment, this is the data and what you can see here that there is a factor called temperature, there is a factor called material type and you have the four readings in each particular cell. This is the total of all the four readings, total of all the four readings, total of all the four readings and same way this is the total of all the and this is the total of all the four readings and this is the total of my particular row and this is the total of my particular column. So, this is what exactly you can do.

(Refer Slide Time: 27:43)

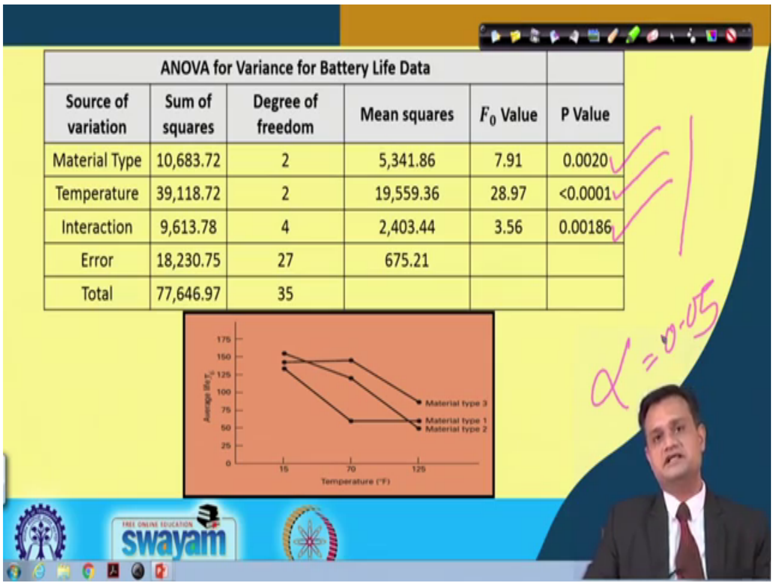


Now, you have the expression. So, just plug in the values you will get SS T, SS material, then you have SS B that is the your temperature that is the another factor and when you do this you have SS interaction.

(Refer Slide Time: 27:59)

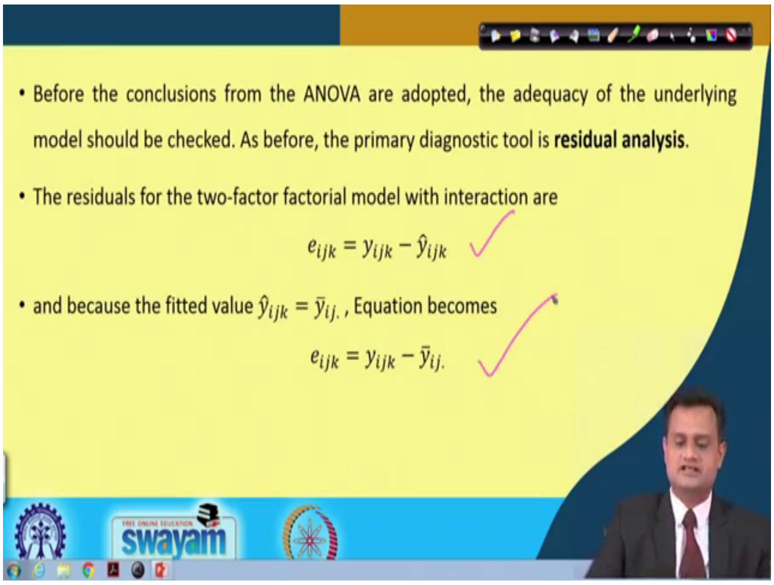


(Refer Slide Time: 28:03)



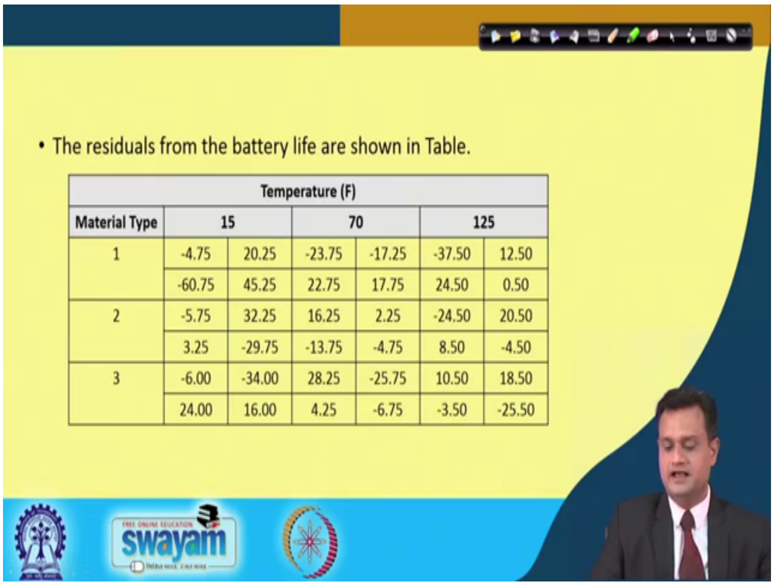
So, typically you will have the results like this. So, what you get here is basically the P value for material type its 0.002. This is quite less than 0.001. So, you can say it is 0.00 this is your interaction effect. So, by referring these three, you can say that if I am checking at level alpha is equal to 0.05 all these are falling in the rejection region. So, my null hypothesis that there is no main effect A, there is no main effect B, there is no interaction effect is rejected and hence, there is a significant impact of temperature on the battery life, there is a significant impact of material on the battery life and interaction effect is also significant in terms of battery life.

(Refer Slide Time: 28:57)



You can check the model adequacy as usual by having the residual component.

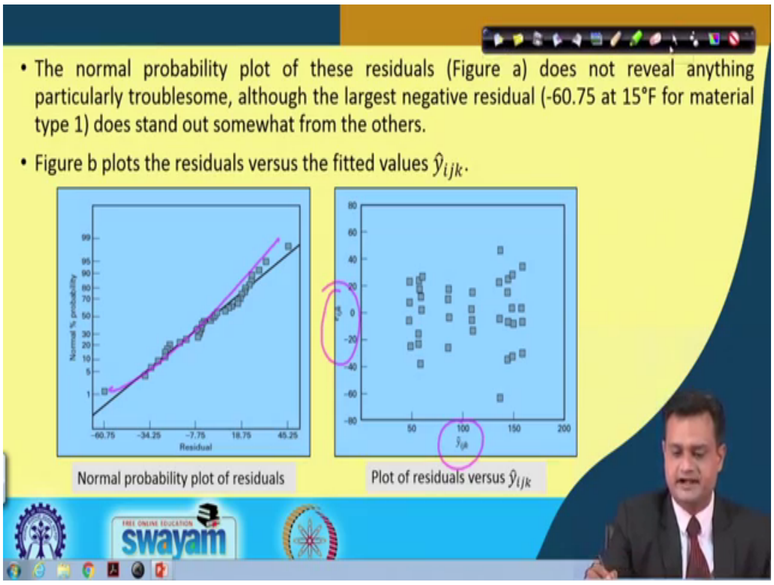
(Refer Slide Time: 29:04)



	Temperature (F)					
Material Type	15		70		125	
1	-4.75	20.25	-23.75	-17.25	-37.50	12.50
	-60.75	45.25	22.75	17.75	24.50	0.50
2	-5.75	32.25	16.25	2.25	-24.50	20.50
	3.25	-29.75	-13.75	-4.75	8.50	-4.50
3	-6.00	-34.00	28.25	-25.75	10.50	18.50
	24.00	16.00	4.25	-6.75	-3.50	-25.50

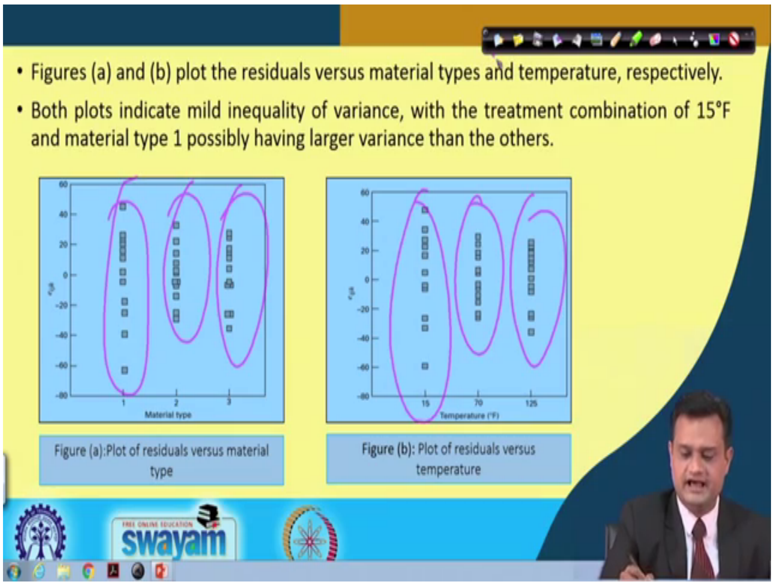
And you can find the residual for each particular material by subtracting it from the grand mean.

(Refer Slide Time: 29:12)



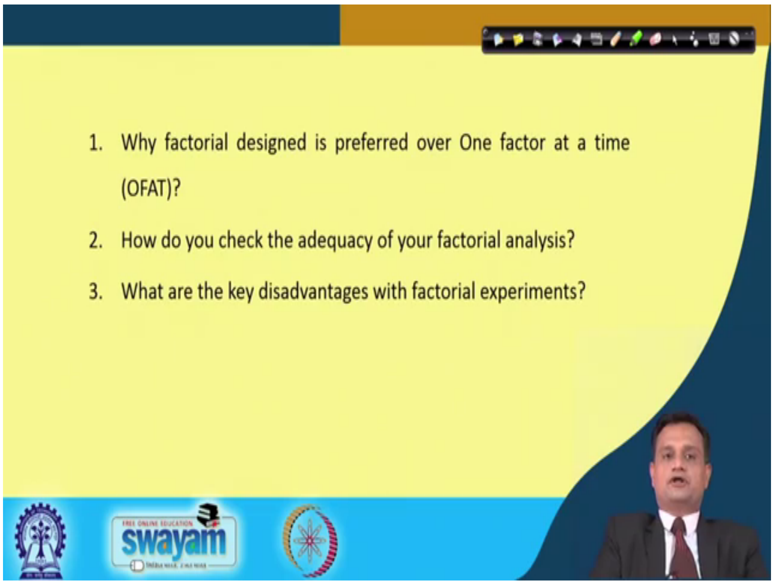
And you can plot it. So, more or less this is going through the line and you have the normality assumption varied say valid and this is when you plot, you can see the scattered net. So, there is an independence which is also verified.

(Refer Slide Time: 29:32)



Now, these are the plots that shows the individual variability for a particular material type. So, more or less there is nothing great to say observe more or less the variability is there, not that too less too high and we can comfortable with the equal variance also.

(Refer Slide Time: 29:56)



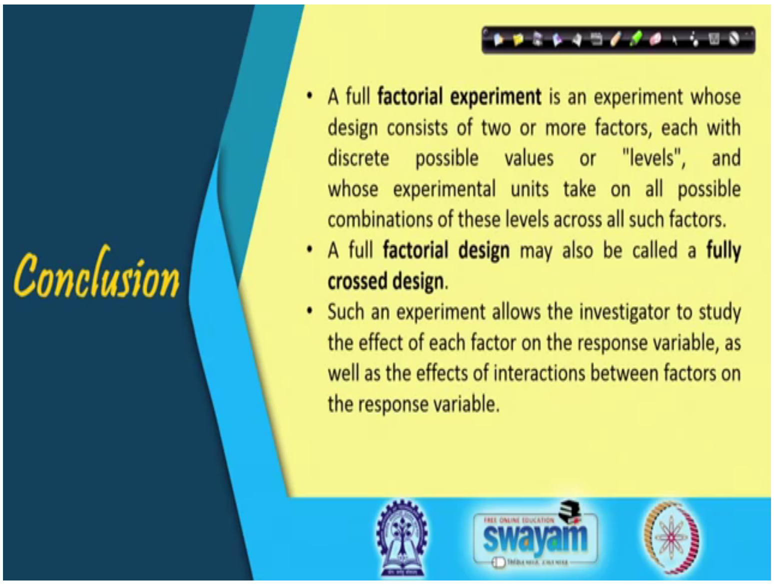
So, now with this let me plot couple of thing for your understanding, for your introspection. Why factorial design is prepared over one factorial at a time OFAT? How do you check the adequacy of your factorial analysis and what are the key advantages with factorial experiments? So, please try to go through the concepts covered in this particular lecture and this will really help you to understand the concept of factorial design and how it helps us to analyse the factors.

Simultaneously consider the interaction effect and basically it is all about revealing more information about the factors and their interaction effects, so that I can design my processes product with greater accuracy and that can help me to have robust product and design when it will be delivered to the customer.

(Refer Slide Time: 30:54)

So, I am mainly referring Montgomery, D. C you can also refer this particular book.

(Refer Slide Time: 31:00)



And conclusion is that a full factorial or factorial experiment is an experiment whose design consists of two or more factors which with discrete possible values or levels and experimental units take on all possible combination of these level across all such factors. So, this is called fully crossed design and this helps investigator to study the effect of each factor on response as well as the interaction effect.

So, thank you very much for your interest in learning this particular session and you would really be benefited if you solve a couple of example or you collect some real life data and conduct the experimentation, then you would be able to internalise this concept. So, we will advance in our say DMAIC cycle typically, right now, we are discussing the improved phase of DMAIC cycle. We will advance in this and then, you will have the better filling about the complete phase. So, keep revising. Be with me, enjoy.